THE VALIDITY AND VALIDATION OF MATHEMATICAL MODELS: METHODOLOGICAL, THEORETICAL, AND PRACTICAL STUDIES WITH EMPHASIS ON THE MODELLING OF COMPLEX BIOLOGICAL SYSTEMS

Mark S. Leaning

Thesis submitted for the Degree of Doctor of Philosophy

Department of Systems Science
The City University

December 1980

ORIGINAL COPY TIGHTLY BOUND

To Pat

CONTENTS

ACKNOWLEDGEMENTS

I offer my sincere thanks to my supervisors, Dr. Ewart Carson and Professor Finkelstein, for their tremendously helpful guidance and support.

I also thank Maha Al-Dahan, Nazar Sarhan, and Ranjeet Uttamsingh for their close collaboration on the biological case studies, Keith Webb for his help with Chapter 9 (and for showing me how little I know about the social and political sciences), and Phil Edwards for his stimulating comments over the last three years.

Finally, many thanks to my typists, Aline Bowers, Audrey Cackett, Donna Fowler, and Gill Lambert.

ABSTRACT

In recent years there has been a tremendous growth in the development and application of mathematical models in all areas of science and engineering. Aided by the advances and availability of computers, models have been used in many new areas, such as biology and the social sciences, and applied to increasingly complex systems. At the same time, model validity and validation have become correspondingly more problematic yet received little attention. The aims of this thesis are to clarify the meaning of model validity, to develop a range of procedures for model validation, and to consider in depth the validity of a number of specific models. The main focus is the use of models in systems science and in biology and medicine.

A review of the scientific literature of model validity and validation is made which reveals many techniques for empirical validation, but exposes the lack of a consistent conceptual approach towards model validity. In reviewing the philosophy of science with reference to validity and validation, the importance of regarding models and validation as part of an evolving research programme and of heuristic considerations in assessing model validity are emphasised.

A new and innovative theory of model validity is proposed which explicates model validity as a multidimensional concept closely related to modelling objectives. The different modelling objectives and types of data are classified and the various concepts of validity are expressed as validity criteria. The general relationship between modelling objectives, data, and validity criteria is explained. The theory is then used to devise a range of validation methodologies suitable for models in research areas at different stages of development.

Models of the human cardiovascular, renal, and respiratory systems are used as case studies for validation. Extensive use is made of the conceptual framework of the theory of model validity and the validation methodologies. The results are a precise delimiting of the validity of the models, the areas of uncertainty, and the potential for future development. This indicates the critical value of the theory and the appropriateness of the methodologies to complex biological models. Further support for the theory and its wide applicability is obtained in using it to consider aspects of validity and validation of models in the social sciences.

Finally, the implications of the work for modelling and validation in systems science and in biology and medicine are examined. In both areas it is shown that the theory of model validity leads to an improved understanding of the nature of modelling and validity, and that the validation methodologies are suitable for the critical and effective validation of a wide range of models. In biology and medicine specific recommendations are made for the types of model appropriate to different modelling objectives and for suitable techniques and methodologies for validation.

This thesis contributes to an improved understanding of the concept of model validity and offers a repertoire of validation methodologies. On another level, it is a broad methodological study of the kind urgently required in systems science. More practically, however, much of the thesis is concerned with the detailed validation of three specific biological models.

CHAPTER 1

INTRODUCTION

Since the 1950's there has been a tremendous growth in the development and application of models[1] in all areas of science and engineering, made possible largely by the advances and availability of digital computing; models are the fashion of contemporary science. This growth is most clearly reflected in the development of systems science, a model-based science par excellence. However, as models have dealt with ever more complex subjects and have been applied to new areas, such as biology and the social sciences, their validity has become increasingly problematic. In some cases, for instance the "World models", model validity is so doubtful that their use has been strongly criticised. Whilst conceptual devices and practical techniques for model formulation and simulation have proliferated, the problems of model validity and validation have received relatively little attention. Thus, having formulated a model, it is often very difficult to assess the confidence that can be placed in the conclusions drawn from the model or to decide whether or not the model is suitable for some practical application. There is an urgent need therefore for research into the nature of model validity and the ways in which models may be validated. In this thesis, an attempt is made to consider the whole range of the validation problem, from the philosophical and methodological aspects of the meaning of validity to the practical validation of some specific models.

In very simple terms, validity refers to the correctness, the adequacy, or even the truth of a model. Often, model validity is determined by comparing the model with data (observations and measurements) from the subject or system which is being modelled. There are many factors that may make validity problematic and validation difficult, of which four major ones are outlined below:

(1) The difficulty of acquiring sufficient data to validate the model (e.g. in biological modelling it is often very difficult to make measurements of the internal states of a biological system). For some models (e.g. political models) not all model variables are presently measurable.

---

[1] Throughout this thesis, the term "model" refers to theoretical models, which may be verbal, symbolic, procedural, mathematical, or formal.

(2) Inadequate theory or understanding on which to base the model and against which to test it (e.g. for models of neural control in biology, or for models in the social sciences in general).

(3) The complexity of the systems which are modelled. The large number of interrelationships and possible behaviour modes may make models uncertain and very difficult to validate fully, even in areas where there are few fundamental theoretical inadequacies or data problems (e.g. models of complex physical or technological systems).

(4) The limited resources available for modelling leave insufficient time for validation which is usually (and incorrectly) only considered after model formulation.

Furthermore, there is much confusion over the meaning of the term "model validity". The various usages of the term include: a close match between model and data; agreement with accepted theories, models, or understanding; practical usefulness; and potential for scientific discovery. Unfortunately, these are all used more or less independently and there is no satisfactory explanation of their interrelations or why some should be important for some models but not for others.

There are two main aims of the thesis. The first is to identify and define the various aspects of model validity, and to investigate their relations to each other and to other factors such as modelling objectives, the availability of theories and data, etc. The second aim is to devise and test a variety of methods for validating models in practice. These two aims may be expressed more precisely:

(1) To explicate the concept of model validity, and to provide a conceptual framework for explaining model validity.

(2) To develop and appraise methodologies for the critical and effective validation of models in specific research areas.

The two aims are simply an attempt to answer the two main questions of model validity and validation, namely, "what does model validity really mean?", and "how can models be validated?". In order to answer these questions a wide-ranging research programme was required. This was pursued along three fronts: extensive reviews of previous work and validation methodologies; detailed validation case studies; and the development of a conceptual framework, or theory, of model validity and of specific validation methodologies. The reviews covered the scientific and engineering literature in areas in which the issues of model validity and validation

have been seriously considered. In addition, because the concept of model validity raises many philosophical issues, a review was made of the major schools in the philosophy of science.

The models used as case studies for validation were three mathematical models of biological systems which have been developed in the Department of Systems Science. The area of biological modelling was chosen for two reasons: firstly, there are both theoretical inadequacies and data difficulties, associated with biological models, which expose many of the problems of model validity and validation; and, secondly, the models themselves were in need of extensive validation. An additional aim of the thesis is an assessment of the potentialities and limitations of biological modelling. As well as the results of the detailed case studies of biological models, the results of some more general investigations on the validity and validation of models in the social sciences will be presented in the thesis.

1.1 Structure of the Thesis

The thesis has three distinct parts: (1) scientific and philosophical reviews; (2) theoretical and methodological development; and (3) case studies.

(1) Scientific and philosophical reviews

A critical review of the scientific literature of model validity and validation is presented in Chapter 2. In the philosophy of science validation is usually referred to as verification, confirmation, corroboration, acceptance or, conversely, falsification. A general review of the major schools in the philosophy of science, with reference to their views on validity and validation, is made in Chapter 3.

(2) Theoretical and methodological development

In Chapter 4 a theory of model validity is developed which acts as a conceptual framework for model validity and validation. Where possible, the theory makes use of the recommendations of the review chapters. The theory is used in Chapter 5 to devise validation methodologies and techniques appropriate for a wide range of modelling objectives and types of model. In one methodology, the range of techniques for comparing a model with data, many of which are taken from Chapter 2, is presented in a systematic form.

(3) Case studies

Firstly, the results of the detailed validation of the three biological models are reported, and then some more general considerations of the validity and validation of models in the social sciences are made. In all the case studies extensive use is made of the conceptual framework of the theory of model validity and the techniques and methodologies for model validation previously developed in Chapters 4 and 5. The case studies are therefore a test both of the individual models and of the applicability of the theory of model validity and the validation methodologies. The three biological models are mathematical models of the human cardiovascular system (Chapter 6), the human renal-artificial kidney machine system (Chapter 7), and the human respiratory control system (Chapter 8). The areas of modelling in the social sciences which are considered (Chapter 9) include econometric modelling, world modelling, and the modelling of bicommunal political conflict.

In Chapter 10 the overall implications of the work for methodological, theoretical, and practical aspects of modelling and validation in systems science and biomedicine are examined. Finally, the conclusions to the thesis are made in Chapter 11.

The appendices contain listings of the full mathematical models for the biological case studies. In addition, Appendix I is a historical study in model validation based on William Harvey's discovery of the circulation of the blood.

1.2 A Note on Technical Terms and Symbolisms

Throughout the thesis, many fairly common words or terms, such as "scientific", "utilitarian", "objectives", or "range of application", are used. Because of the nature of a methodological study it is important to give these words more precise, technical definitions, many of which are given in Chapters 4 and 5. Unfortunately, in their common usage, such words often have a wider meaning or have other connotations. To help with the reading of the thesis, an index of key terms has been provided at the end, and this should be consulted if there is confusion.

Owing to the wide range of topics and models considered, it is impractical to devise a single symbolism. Symbolisms are consistent within each chapter and are defined at their first occurrence in each chapter.

CHAPTER 2

THE SCIENTIFIC LITERATURE OF MODEL VALIDITY AND VALIDATION

Review 1

2.1 Introduction

The past two decades have seen a tremendous growth in the use of mathematical models in all areas of science and engineering as well as in the application of scientific techniques to business problems and at various levels of public decision making. For this reason, the issues of model validity and model validation have a general scientific and practical importance. In this chapter a review of the literature of model validity and validation across a broad range of sciences is made. The motivation for this extensive review was the demonstrable lack both of understanding of the nature of model validity and satisfactory validation methodologies in specific areas (such as biological modelling and systems science, the main areas of interest in this thesis) as well as in general. The aim was to search for a common core of meanings for the concept of model validity and to catalogue the various techniques for model validation.

Throughout the review it is apparent that there is much confusion about the meaning of model validity and that many approaches to model validation are over-simplistic. Nevertheless, in the concluding section (Section 2.9), a common core of concepts of model validity is identified. It is interesting that this core is drawn from a wide range of application areas. In Chapter 4, it forms a partial basis of the theory of model validity which is developed. The various techniques for model validation are systematised in a range of validation methodologies presented in Chapter 5.

An important reason for the growing interest in model validity and validation is the widespread use of mathematical modelling in new application areas, such as biology, geography, the social sciences, world modelling, etc. In these areas, there is often a lack of established quantified theory and severe measurement difficulties compared with the areas from which mathematical modelling originates (physics and engineering), and this makes model validity problematic. Furthermore, the availability of powerful computing facilities has led to the simulation of models of highly complex systems which, in all application areas, adds an extra

dimension of difficulty to the validation problem.

The structure of the chapter is as follows: in Section 2.2 some formative ideas on model validity and validation in the 1950's and early 1960's are reviewed. The development of a validation methodology in simulation modelling (one of the few coherent well-developed methodologies) is outlined in Section 2.3. Some aspects of model validity in the social sciences are considered in Section 2.4, with reference mainly to empirically-based models. In Section 2.5, work on the validation of biological compartmental models is reviewed. Some of this work is related to system identification and parameter estimation which are considered in Section 2.6. The concept of model adequacy, which was offered as an alternative to model validity, is critically reviewed in Section 2.7. In Section 2.8, the cross-validation of statistical models, the MIT energy laboratory assessment programme, and the nature of model validity in the physical sciences and engineering are considered.

2.2 Some Early Ideas on Model Validation

2.2.1 Introduction

In this section some of the early work on model validation which preceded the major growth of mathematical modelling (in the 1960's) is reviewed. Although some of the references are philosophical rather than scientific, they form a good basis and introduction for the later sections.

2.2.2 Model testing in Operations Research; Churchman et al., 1957

"Introduction to Operations Research" by Churchman, Ackoff, and Arnoff, is a classic work on the philosophy and methods of Operations Research. In it, the problems of model testing and data acquisition are considered at length: "Part IX Testing, Control, and Implementation"; Chapter 20, "Data for Model Testing".

> "In testing a model we ask, 'What are the possible ways in which a model can fail to represent reality adequately and hence lose some of its potential usefulness?'" (p. 577)

The authors suggest four ways in which the adequacy of a model may be questioned:

(1) "... the model may fail by including variables which are not pertinent." (p. 577)

(2) "The model may fail to include a variable which does have a significant effect." (p. 577)

(3) "The model may inaccurately express the actual relationship which exists between the measure of effectiveness (E) and one or more of the pertinent independent variables (x_i, y_i)." (p. 578)

(4) "It may fail to yield good results because of incorrect parameter values."

This approach distinguishes between (1), (2) and (3) which are concerned with the validity of the model structure, and (4) which is concerned with the accuracy of parameter values. Such a classification has been made often, and in diverse areas (e.g. Carson and Finkelstein, 1977, in biological models; Chrostowski et al., 1978, in mixed dynamic systems). Note that the emphasis by Churchman et al. is on the "adequacy" of a model rather than validity. Adequacy can be defined by operational tests,

whereas validity requires a conceptual definition.

The importance and problems of the process of collecting empirical evidence are stressed:

> "The design of the process of collecting evidence [for model testing] consists of the following parts: definition (including measurement), sampling (including experimental designs), data-reduction, use of data in the test, examination of the result, and possible redesign of the evidence." (p. 581)

The last phrase indicates that if a model fails the test against the evidence, it may be the evidence that is at fault, not the model. In other words, the data itself may not be a valid representation of the system. This important consideration throws doubt on purely falsificationist accounts of model validity (e.g. Popper, 1935; Braithwaite, 1953).

Although the book is clear and detailed on the collection and reduction of data, there is little on the actual confrontation of the model and evidence in testing, and how the four questions of adequacy may be answered. Possibly, the operational philosophy and lack of a clear concept of model validity makes this problematic.

2.2.3 Validity as isomorphism

In his stimulating essay, "Models and Archetypes", Black(1962) argues convincingly for the use of models in science, both for the interpretation of theory and as a tool for discovery. This was largely an argument against the philosopher Braithwaite who put forward strong criticisms of models (the tendency of positivist philosophers of science to prescribe methodologies is discussed in the next chapter). Black describes two aspects of model validity:

> "We can determine the validity of a given model by checking the extent of its isomorphism with its intended application. In appraising models as good or bad, we need not rely on the sheerly pragmatic test of fruitfulness in discovery; we can, in principle at least, determine the 'goodness' of their 'fit'." (p. 238)

The notion of a model as an isomorphism, or "partial" isomorphism, of a system was also developed in the early cybernetic literature by Ashby (1956) (although, in general, model validity has received very little attention in cybernetics). In equating the concept of validity with the

formal concept of isomorphism it must be remembered, however, that a model contains propositions of a general kind (e.g. mathematical equations), whereas knowledge of the intended application (the system) is in the form of data (measurement and observations) which are records of specific events. Thus model and data have a different logical status.

Reasoning along these lines, Suppes (1962) sets out to show in a formal way that:

> "... an exact analysis of the relation between empirical theories and relevant data calls for a hierarchy of models of different logical type." (p. 25) Such an approach is also "... closely connected to the statistical analysis of the empirical adequacy of theories." (This paper is closely related to Suppes' work on axiomatic measurement theory.)

2.2.4 Kaplan's methodology for the behavioural sciences; the effect of model purpose on model validity, 1964

Kaplan's "The Conduct of Inquiry" (1964) is a careful analysis of scientific method in the behavioural sciences from an instrumentalist perspective. This view considers the methods of science in relation to the purposes or functions they serve in inquiry. The book contains chapters on experiment (IV), measurement (V), and theory (VII), which are of particular interest.

In Chapter V, Kaplan firstly gives an account of the logical basis of measurement (i.e. non-instrumental), but makes the switch back to an instrumental viewpoint in considering the validity of measurements and problems of measurement uncertainty (there is particular reference to psychological measurement). Here, the validity of a measurement is assessed on the basis of its function in inquiry. Perhaps this indicates that neither a positivist nor an instrumentalist philosophy is alone capable of a satisfactory analysis of science.

§36 of Chapter VII, "The Validation of Theories", considers the validity of a theory in relation to its purpose. To Kaplan, the validity of a theory depends upon its ability to satisfy its intended purposes as well as not failing empirical tests, although he stresses that the methodological strength of methods of falsification should not be under-estimated.

Some other aspects of this book are reviewed in Section 2.4.2.

2.3 The Development of a Coherent Validation Methodology: A Case Study - Simulation Modelling

2.3.1 Introduction

Simulation modelling is a technique for solving stochastic modelling problems using digital computers. Its largely statistical nature stems from its origin in Monte-Carlo methods. These methods, which concern experiments with random numbers, began their systematic development during World War II when they were applied to problems related to the atomic bomb. The work involved direct simulation of probabilistic problems concerned with random neutron diffusion in fissionable material. In the 1950's simulation experiments became feasible on digital computers, allowing the solution of problems associated with the complex interactions of system components.

The technique has been applied to a wide variety of systems, especially nuclear and missile engineering, and economic and political systems. The main application, however, is in the modelling of operations, or events, within a business or industrial framework. A typical example is the two-stage arrival/service-activity queuing model which may be simulated stochastically in a simulation model.

A simulation model is composed of a logical structure which specifies the discrete units of behaviour (events) and how these units are combined to represent system behaviour. Stochastic units and interactions are represented by statistical distributions. The model is simulated by applying the appropriate inputs (often stochastic) and by random sampling from the statistical distributions. Repeated runs of the simulation (as in the Monte-Carlo method) build up a statistical picture of the model's behaviour. The use of a simulation model in this way is known as a "simulation experiment".

The problems of model validity and validation arose as a practical one associated with the need to justify the use of such a technique. However, the first approaches to these problems were from a methodological basis (Naylor et al., 1966, see Section 2.3.2; and Hermann, 1967, see Section 2.3.3). In the late 1960's attention was focussed on the statistical adequacies of simulation models (e.g. Fishman and Kiviat, 1968), such as investigating the true randomness of the so-called "random-number generators".

Simulation models are used primarily as black-box models, i.e. to predict outputs for certain inputs. Consequently, validation centred around the statistical comparison of model and system outputs.

By the end of the 1960's, a coherent methodology for the validation of simulation models had been developed, and the 1970's saw the refinement and application of the various techniques (see, for example, Ignall, 1978, Section 2.3.10).

In the review of the literature which follows, it will be seen that the methodology is essentially statistical, and that many aspects pertinent to a model's validity (e.g. structural uncertainty, or measurement problems) are largely neglected. Although the methodology is probably meaningful in the context of simulation modelling, it would be wrong to apply it uncritically elsewhere, or, indeed, to suggest that it is a general validation methodology, as Mihram has done (Mihram, 1972, Section 2.3.6).

2.3.2 Four methodological positions - Naylor et al., 1966

"Computer Simulation Techniques" by Naylor, Balintffy and Burdick (1966) deals with simulation models of economic and business systems. Chapter 8, "The Problem of Verification", discusses the validity of such models by considering four distinguishable methodological positions. They point out that their analysis goes beyond simulation models,

> "The question of verification of simulation models is in reality no different from the question of verification when applied to any type of hypothesis or model, whether it be expressed as a verbal model, a physical model, a mathematical equation, or a computer program." (p. 310)

The authors use verify and validate interchangeably, yet give an impossible definition:

> "To verify or validate any kind of model means to prove the model to be true." (p. 310)

However, it is not possible to prove a model to be true, but it may be possible to disprove it. Naylor et al. go on from this definition to question the concept of truth, leaving the reader wondering quite what the definition means. Fortunately, it is not central to the main argument of the chapter, which continues by describing four methodological positions on validation (verification) that have been taken in economics.

(i) Synthetic apriorism

This holds "that economic theory (or for that matter any kind of theory) is merely a system of logical deductions from a series of synthetic premises of unquestionable truth". (p. 311)

The validity of a theory follows automatically and no empirical test is necessary.

(ii) Ultraempiricism

This is an extreme form of logical positivism which asks that we begin with facts, not assumptions.

(iii) Positive economics

Milton Friedman, in his essay "The Methodology of Positive Economics", "argues that critics of economic theory have missed the point with their preoccupation with the validity of economic assumptions. According to Friedman, the validity of an economic model depends not on the validity of the assumptions on which the model rests, but rather on the ability of the model to predict the behaviour of the endogenous variables that are treated by the model." (pp. 313-314)

However, this is a single criterion for validity which asks that we accept the model, regardless of how implausible the assumptions are. (Blaug, 1962, discusses this criticism more fully.)

(iv) Multistage verification

The three preceding methodological positions suggest yet a fourth possible approach which is a three-stage procedure incorporating apriorism, ultraempiricism, and positive economics.

First stage

Formulation of postulates or hypotheses.

Second stage

Submit postulates as tentative hypotheses:

"Wherever possible we will insist on applying Karl Popper's criterion of falsifability to our postulates." (p. 314). The postulates may be tested subject to the limitations of existing statistical tests, such as the t-test, F-test, chi-square test, distribution-free tests, etc.

So after getting into a philosophical muddle by talking about truth at the beginning of Chapter 8, the problem becomes reframed statistically in a manner which characterises later work on simulation validation:

"Although we cannot solve the philosophical problem of

'What does it mean to verify a postulate?', we can apply the 'best' possible statistical tests available to us to these postulates." (p. 315)

Third stage

Testing of the model's ability to predict system behaviour.

Finally, a consideration is given to the distinction in validity for normative and explanatory models. The validity of the latter is determined by the refutation or confirmation by empirical observations, whereas normative models are to be judged by their efficacy in achieving certain policy aims.

2.3.3 Five criteria for model validity - Hermann, 1967

Hermann's 1967 paper considers problems of validation in games and simulation, with reference to models of international politics. Rather than stumble at the question, "What is it for a model to be true?", as do Naylor et al. in Section 2.3.2, Hermann adopts the approach, "In what ways can a model be compared with empirical data/evidence in order to check it?". Firstly, he notes that the purpose of the model affects the way in which a model is validated.

Secondly, he considers five types of validity criteria, which are applicable to certain purposes. The five criteria are:

(1) Internal validity
(2) Face validity
(3) Variable-parameter validity
(4) Event validity
(5) Hypothesis validity

(These criteria are discussed in more depth in Section 2.4.3.)

The paper concludes with an examination of the special problems of validation of models which involve human participants (gaming models). Reference is often made to this paper, but the ideas put forward in it have received little further development in the area of simulation modelling.

2.3.4 Validity as a statistical problem - Naylor and Finger, 1967

In their paper, "Verification of Computer Simulation Models", Naylor and Finger (1967) review the philosophical and methodological

positions on validation and then describe in detail a variety of statistical tests which may be used to validate simulation models.

2.3.5 The distinction between "Verification" and "Validation" - Fishman and Kiviat, 1968

Fishman and Kiviat (1968) introduced the distinction between verification and validation for simulation models:

> "Verification determines whether a model with a particular mathematical structure and data base actually behaves as an experimenter assumes it does. Validation tests whether a simulation model reasonably approximates a real system." (p. 186)

Since the simulation invariably consists of applying random inputs and random sampling of the model's statistical distributions (data base), "verification"largely consists in testing the quality of the random number generator,

> "The most important hypothesis to test is that the pseudo-random number generator creates sequences of independent random variables." (p. 188)

This stage of verification is called "data verification". The next stage is examining the "substructure outputs and determining whether they behave acceptably" (p. 189), and is "structure verification". For example, a simpler analytic or simulation model may be found to be behaviourally equivalent to a more complex one, and the replacement leads to a better understanding of the system, as well as improving computational efficiency.

In simulation models, output functions usually satisfy covariance stationarity assumptions, and the theory of covariance stationary processes provides a good framework within which to study the nature and extent of autocorrelation, the principal form of intertemporal dependence. Autocorrelation is given by :

$$R(\tau) = \varepsilon\left[X_t\, X_{t+\tau}\right] - \{\varepsilon\left[X_t\right]\}^2 \qquad \ldots\ldots\ldots (2.1)$$

although, for statistical reasons, Fishman and Kiviat prefer the spectrum:

$$g(\lambda) = \frac{1}{\pi} \sum_{\tau=-\infty}^{\infty} R_\tau e^{-i\lambda\tau}, \quad 0 \leqslant \lambda < \pi \qquad \ldots\ldots\ldots (2.2)$$

In validation, a similar distinction exists between data and structure validation. Data validation basically consists in checking that the numerical data conform to some theoretical distribution, which is then used for sampling to expose the simulation to:

> "the universe of possible stimuli rather than those that have occurred in the past. Often, graphical methods suffice to judge the validity of theoretical distributions." (p. 191)

"Structure validation" is usually done by an autocovariance or spectrum comparison of system and simulation outputs. However:

> "Validation, while desirable, is not always possible ... Despite its difficulty, effort must be expended on model validation - first, to give credence to results within the validated range of model operations, and, second, to instill confidence in the extrapolation beyond the range of model experience." (p. 192)

Once the model has been verified and validated, Fishman and Kiviat discuss a third stage - Problem Analysis. This is using the model to help collect and analyse data, to make inferences.

This important paper contributes to the validation methodology of simulation modelling in two ways. Firstly, it reinforces the relation between validation and the classical body of statistical theory, and, secondly, it introduces the distinction between verification and validation. However, the use of the word "verification" implies a proving of the model in an exact way, which is not meant by Fishman and Kiviat. A more appropriate phrase might be validation of the model in simulation vis-à-vis its intended mathematical, logical, statistical and algorithmic form.

2.3.6 Textbook methodologies - Naylor, 1969 - Mihram, 1972

The book, "The Design of Computer Simulation Experiments", edited by T. H. Naylor (1969) is a standard work on simulation models (with particular reference to economic models) rather than a research text. This is a good indicator that, by 1969, the paradigm of validation for simulation modelling had achieved a coherency and acceptability by its theorists and practitioners. Consequently, there is little new conceptual advancement recorded in the book, although the statistical content is more sophisticated than in the earlier work.

The importance of validation in simulation modelling is recognised, and Chapter 8, "Validation", by R. Van Horn is devoted solely to the issue. This draws its basis from Naylor et al. (1966) and Fishman and Kiviat (1968). A new idea suggested by Van Horn is an equivalent to the Turing test for models. If a policy maker is unable to tell whether events predicted by the model are from the model or the real system, then the model is valid. However, this is unacceptable since it would limit the scope of the model to the knowledge of the policy maker and disturbingly threaten the objectivity of the model.

Mihram's "Simulation: Statistical Foundations and Methodology" (1972) is another book in which previous results of simulation modelling are brought together in a statistical framework. Although not indicated in the title, the book shows Mihram's belief that simulation methodology is the correct one for the "system sciences". This is clearly so in two of his 1972 papers, "The Modeling Process" and "Some Practical Aspects of the Verification and Validation of Simulation Models". The latter is a reasonable review paper which traces the development of simulation validation. In this validation is split into the two aspects of validation and verification (cf. Fishman and Kiviat, 1968) and is regarded as an essentially statistical problem.

2.3.7 Fitting, calibration and validation - Wigan, 1972

Wigan's 1972 paper, "The fitting, calibration, and validation of simulation models", follows Naylor et al. (1966) in a methodological way, yet does not refer to any of the later work in simulation validation (Naylor and Finger, 1967; Fishman and Kiviat, 1968, etc.). This results in a rather different approach, outside the "coherent methodology" described in the rest of this section.

In modelling, Wigan describes a "natural hierarchy" of five stages, in which each stage is dependent upon those above it:

(1) Postulates
(2) Fitting
(3) Calibration
(4) Identification
(5) Validation

The dependence of the final model validity upon the prior stages is what, according to Wigan, makes validation a difficult problem. For this

reason he considers the fitting and calibration stages as well as the validation stage.

In the "Fitting" stage, parameterised functions are fitted to the data, and the "Calibration" stage adjusts the interdependence of the functions.

"Identification" is a difficult stage, and aims at "assuring that the detail of the calibrated model is justified by the available data" (p. 190). The problem here is that different objective functions often produce different results.

The final stage, "Validation", is the "process of discriminating between different sets of postulates by reference to fresh data not used in the setting up, fitting,and a calibration process" (p. 191).

This approach is then applied to a model for traffic flow assignment in a road network (Wigan works for the Ministry of Transport).

Although Wigan does not base his approach in a statistical framework, he does consider the most important aspect of model validity - the degree to which the structure of the model can be justified prior to the testing of the aggregate behaviour. This point often gets submerged in a statistical swamp in the mainstream simulation methodology.

2.3.8 Deciding between competing models - Schaeffer, 1975

In "Model Validation using Simulation and Sequential Decision Analysis", Schaeffer (1975) regards model validation as a problem of choosing between a set of alternative models. Models are compared with data in order to produce order statistics. Once enough data have been obtained, a sequential decision algorithm chooses the most valid model. (For a similar order-theoretic approach see Reggiani and Marchetti, 1975, Section 2.7).

With this approach, it must be assumed that there is no a priori reason for preferring some models to others in the model set.

2.3.9 "Four Questions regarding the Credibility of Simulation" - Mihram, 1976 and "Four Further Queries concerning Similar Credibility" - Mihram, 1976

In the first of these papers, Mihram pedantically proposes that the words "verification" and "validation" should be replaced by "scrutinisation" and "confirmation". This is probably justified in the case of "verification"

which is too strong for its intended meaning (see Section 2.4.2), but not for "validation" which is a totally appropriate word.

The four queries are as follows:

(1) Can data be split for calibration and validation?

(2) How much effort should be spent on each stage of a model's development?

(3) What specific procedure should be followed so as to ensure the credibility of models reported in the literature?

(4) What is a suitable format for model publication?

The latter two have been discussed in depth by House (1974, see Section 2.5.5).

The second of these papers is more contentious, and concerns the appropriate methodology for the modelling of the dynamics of social and world systems. In this Mihram elevates the methodology of simulation modelling into a philosophical status which is "in accordance with the established Scientific Method" (p. 1233).

A more practical and sensible point is made, however, in the reporting of simulation models,

> ".... algorithmic simulation models, unlike models written in a (first person) natural language or the (third person) language of mathematics, are authored in a computer-directed (second person) language. Thus their publication in the printed medium is not particularly helpful ..... Thus the credible computerised models of systemic scientists should be reported in the printed medium of natural language." (p. 1233)

2.3.10 Recent developments

The methodology of simulation validation has to date been an essentially statistical one. The papers on the subject have either been concerned with applications (e.g. Ignall et al., 1978, in developing and validating analytic models of New York City's fire and police operations; Kheir, 1976, A validation procedure for missile system simulation models), or with technical statistical problems (e.g. Rowland, 1978; on the problem of sparse data, which is solved using Theil's inequality coefficient).

2.4 The Validity and Validation of Models in the Social Sciences

2.4.1 Introduction

This section reviews some literature in which the problems of model validity and validation in the social sciences have been considered. The main areas in which these problems have received attention include economic modelling, business/management modelling, behavioural science, and simulation modelling. These areas are characterised by an empirical methodology, and applications to practical problems.

In other areas of the social sciences, such as sociology or political economics, there is a lack of coherent methodologies and, indeed, there is often controversy between theoretical and practical schools. An important aspect of this divide is the difficulty of finding close relationships between theoretical concepts and the empirical process of measurement. On the other hand, in the aforementioned areas (economic modelling, etc.) most conceptual variables are measurable (e.g. money, time, commodities), although practical problems may exist.

It seems correct, therefore, to draw a distinction between measurement and non-measurement methodologies, and it is in the former, in which the models used are subject to the empirical notions of validity and validation, which is the concern of this review. (Models are being increasingly used in the social sciences for purposes such as the comparison and/or integration of competing theories, the development of new theories, and for structuring experimental research (e.g. in conflict dynamics; Bowers, Mitchell, Webb, 1979). These uses of models are effectively a bridge between the two types of methodology to date; there is no distinguishable literature on the problems and techniques of model validity and validation for this type of model and hence the review is unfortunately limited to measurement-based methodologies in which these issues have received a reasonable degree of attention. In Chapter 9, however, the theory of model validity (Chapter 4) is applied to the problems of model validity and validation in the social sciences and an appropriate range of techniques for the validation of such innovative models which do not yet have a complete empirical base is proposed.)

The work of Naylor and his colleagues led to the development of a coherent validation methodology for simulation modelling, and this was discussed at length in Section 2.3. This area of modelling is considered more generally below, in Section 2.4.3. Others include: the validation

of dynamic economic models (Section 2.4.4); the concept of model credibility (Section 2.4.6); and critical factors affecting the validity of models in the social sciences as opposed to the physical sciences.

2.4.2 Behavioural science - Kaplan, 1964

Kaplan's book, "The Conduct of Inquiry" (1964), concerns methodology for the behavioural sciences, and was reviewed in Section 2.3.6. From his "instrumentalist" perspective, a theory or model is valid if it both satisfies its role in scientific inquiry (i.e. its purpose) and does not fail empirical testing. Kaplan, however, does not detail how a model is to be validated vis-à-vis its purpose, nor the actual confrontation of a model with empirical data. (In many ways, these two issues were answered by Hermann, 1967, see Section 2.4.3).

In Chapter V, on measurement, Kaplan considers the validity of measurements. The approach is that taken by the psychological measurement theorists, where three types of validity are distinguished:

(i) Content validity: This concerns the specified domain over which the measurement or test is made (e.g. the range and sampling of a subject on which a student is examined).

(ii) Predictive validity: This compares the results of the test/measurement with those of other tests, or actual experience (e.g. do school entrance exams select children suitable for their first year in the school?).

(iii) Construct validity: This concerns the validity of construct measures, to which there is not a directly observable dimension. The validity depends upon the degree to which the construct depends upon measurable components, and other concepts. It is clearly related to the validity of the functional relationships.

A good reference on this subject in psychological measurement is J. C. Nunnally (1970).

Most variables used in the social sciences are not directly observable (i.e. there is no single empirical attribute that can be used as a measure of the variable). However, many variables are reflected in a cluster of directly measurable attributes, and measures of these may be used to determine a value for the variable in question. In the social science literature (e.g. Blalock, 1974) these component attributes are

referred to as "indicators", and the indirect variables as "unobserved" variables. There are three distinct methods of measuring unobserved variables by indicator techniques:

(1) Single Indicator. A single indicator is used as a measure of the variable (c.f. "pointer property" in physical measurement).

(2) Indices. A group of indicators are combined to build a single summary score or index.

(3) Multiple Indicators. In this method the indicators keep their separate identity.

All three methods are subject to considerations of content and predictive validity. The validity of an index also depends upon the conceptual framework (theory/model) which determines the indicator components and the mathematical way in which they are combined (construct validity). A recent paper by de Neufville (1978) illustrates some of the problems of validating policy indicators in political science.

These considerations demonstrate two important points: firstly, that measurements or observations are not simply facts about the world, but depend upon theories or models; and, secondly, when comparing a model with empirical data, the validity of the model depends not only on the closeness of the model and data, but on the degree to which the data are valid measures of the empirical system.

2.4.3 Gaming and simulation modelling - Naylor et al., 1966 Hermann, 1967

Naylor et al. in their book, "Computer Simulation Techniques" (1966), distinguish four methodological positions on verification (i.e. validation) in Chapter 8, "The Problem of Verification" (pp. 310-320). These are synthetic apriorism, ultraempiricism, positive economics, and multistage verification. These positions, and the subsequent development of a validation method in simulation modelling, were discussed in Section 2.3. Unfortunately, their choice of a multistage verification, in which the validation procedure examines the assumptions, their validity, and overall predictive ability of the model, was not heeded in this development, and the methodology is essentially statistical, missing many important aspects of validity.

Hermann's paper (1967) on the problems of validation in gaming and

simulation models of international politics goes far beyond that immediate application. The paper is divided into three sections: the first deals with the effect of a model's purpose upon its validity, the second with five criteria for validity, and the third is devoted to an example, in which there are particular problems related to the use of human participants in a gaming model.

In the "Effects of Purpose on Validity", Hermann states that a model (or a game or simulation) frequently has

> "The purpose to explain or predict the behaviour of B [a selected reference system]. Not all games or simulations, however, have that purpose. When the primary objective for a ..... model is not to replicate aspects of some system, then the model's validity is affected." (p. 217)

He outlines some other possible objectives:

(i) "Alternatives and their Consequences"

Models may be intended to explore policy problems associated with the consequences of alternative courses of action (p. 217).

(ii) "Relative Predictive Ability"

"... the ability of a model to predict certain outcomes as compared to the projections of other methods of prediction." (p. 218)

(iii) "Instruction"

In this the objective is to maximise the learning experience of the students,

> "... the validity criteria have shifted from the observable universe to the effects on the cognitive and affective systems of those individuals whom the operating system is intended to instruct." (p. 219)

(iv) "Hypothesis and Theory Construction"

For instance, a model may be used to test the completeness of a theory. Comparisons with the observable universe are still required, although,

> "the final validity criteria are in terms of the heuristic payoff from the simulation for hypothesis and theory building." (p. 219)

(v) "Nonexistent universes"

The reference system to a model may not always be the observable universe,

"The fact must be recognized that some simulations are concerned with nonexistent universes rather than observable ones." (p. 220)

For instance, the investigation of,

"'What-would-have-happened-if?' worlds".

The next section of the paper examines five validity criteria which relate ways in which a model can represent a reference system,

"In all but the first approach we are asking to what features of the observable universe can we extrapolate or generalize operations occurring in a game or simulation." (p. 220)

(i) "Internal Validity"

"With regard to operating models, the critical requirement for reliability or internal validity is that the variations [between repeated runs of the simulation] be accounted for by identifiable relationships within the game or simulation." (p. 221)

(This corresponds to Fishman and Kiviat's use of the term "verification of the computer simulation", 1968, see Section 2.3.5.

(ii) "Face Validity"

"... is a surface or initial impression of a simulation or game's realism."

There is a danger, however, in applying this too far, for an experimenter may not be aware of some actual behaviours of the system after a limited experience of observing the actual phenomena.

(iii) "Variable-Parameter Validity"

"Comparisons of the simulation's variables and parameters and their assumed counterparts in the observable universe." (p. 222)

(iv) "Event Validity"

An "event" is defined to include patterns of behaviour as well as isolated occurrences. One critical problem, however, is the level of generality at which events should be compared; event validation is useful for checking the composite set of interrelations in a model, but may be less useful for locating the exact parts of a model responsible for incongruities between simulation events and those in reality to which it

is being compared.

(v) "Hypothesis Validity" (p. 223)

A hypothesis makes a relation between two or more variables and this should be confirmed empirically. As in hypothesis testing, this is often done with the help of statistics. Hermann distinguishes two kinds of relationship:

(a) Those which are an integral part of the model, and can be stated as "researchable hypotheses" or from which "empirically verifiable" propositions can be made.

(b) Independent of relationships contained in the operating model. A hypothesis can be investigated "even though the relationship is not required for the operation of the model." (p. 224)

The final section of the paper investigates the validity of games with human participants. Often humans are used in gaming models to represent the behaviour of people in a particular situation. In this kind of model there are other questions concerning model validity,

> "How can college students behave like experienced political leaders? How can American participants represent the cultural values of other societies? ..... The issue raised by these inquiries affects model validity as directly as do those of purpose and criteria." (p. 226)

Hermann argues that one of the main reasons for using human participants is to "reduce some validity problems" (p. 228), but one must still be certain that the game players are representative of those in the reference system. This "representativeness" is the requirement for validity of such gaming models, although "the particular criteria will depend upon the kind of situation we attempt to replicate." (p. 230)

In the conclusion, Hermann notes that validity has received little attention, and it is premature either to accept or reject the value of gaming and simulation models in the behavioural sciences.

2.4.4 Validation of macro-economic models - Young et al., 1973

Keynes' economic analysis (1936) and Tinbergen's pioneering work (1937) stimulated the development of dynamic models of aggregated capitalist economies. In the paper by Young et al. (1973), macroeconomics is treated as a case study in the modelling and identification of a

dynamic system. The paper divides econometric modelling into four main headings: Model Specification, Data Collection and Treatment, Identification and Estimation, and Model Validation.

Model Specification deals with such aspects as "What is the purpose of the model? What are the important variables and relationships? How should the model be represented?", etc.

In "Data Collection and Treatment", Young et al. point out that there is a tradition of assessing and dealing with errors of measurement in the physical sciences that has not been sufficiently considered in the social sciences. The problem with economic data is not just its low quality,

> " ... if the judgement of the quality of economic data creates headaches, the paucity of the data makes things even worse." (p. 156)

Once the value of the data has been assessed, the modeller should aim at a model which explains the data consistent with the level of confidence associated with the data. (The problems of data uncertainty in economics first received prolonged treatment in Morgenstern's classic work, "On the Accuracy of Economic Observations", Princeton, 1950.)

In "Model Identification and Estimation", a model structure is first sought, and then the parameters of the equations are estimated. There is a need to remove non-stationary disturbances from the data, and to use detailed and comprehensive noise models. (N.B. The use of the word "identification" has a different meaning from that in control engineering, in which the system is identified from I/O measurements.)

On "Model Validation", Young et al. do not go into an analysis of the concept of model validity, but rather discuss the inadequacies of some existing methods of validation in econometrics,

> " ... existing procedures - mainly ex-post forecasting - appear to be rather inadequate." (p. 156)

Too much emphasis is placed on statistics such as the correlation coefficient, R^2, and the Durbin-Watson statistic. Even quoting of the standard errors on the coefficients (i.e. the parameters) is insufficient because they are likely to be highly correlated; consequently, it is essential to quote the variance-covariance matrix of the estimation errors of the coefficients.

Young et al. recommend more sophisticated tests such as those based on residual correlation analysis, though these do little to "assess the sensitivity of the model to parametric uncertainty" (p. 157). Their experience with Monte-Carlo simulation techniques suggests that it has a considerable practical potential.

An example is given in which the technique is applied to the Livesey model of GDP. At each point in time, a probability density was obtained by making a histogram based on the stable trajectories generated in a 500-run Monte-Carlo simulation and then fitting this histogram with a probability distribution using an Edgeworth-Charlier distribution. From this it was shown that the "model is fairly sensitive to parametric uncertainty" (p. 157).

This method offers a way of validating the parameters in an econometric model, particularly when they are uncertain owing to the scarcity and uncertainty of the data. It does not deal with the problem of validity of model structure, or the hypotheses involved in the model construction, and for this reason the validity of the model structurally may still be in doubt.

2.4.5 The need for standardisation of model reporting in the social sciences - House, 1974

In this paper ("Diogenes revisited - the search for a valid model"), House considers the problems of validating social science models for use by policy makers. He proposes that in using a model, the policy maker should act as though buying a crystal ball which has a probable but not exact accuracy. House's criterion for validity is, therefore, an acceptable level of predictive ability.

The distinction between verification and validation in simulation modelling is discussed (see also Section 2.3.5). Verification is more or less a "mechanical" procedure to test the "design consistency" of the operating model - although it does establish some credibility - whereas validation is the comparison of model outputs with real world data (mainly in a statistical manner), and is more problematic.

It is impossible to validate a model completely for a variety of reasons, including: the complexity of the model and the problems of measurement (data may be of low quality, and are not available at all for validating future predictions).

From this position, House forms two hypotheses:

(i) In the formal sense, a model can be validated historically (i.e. in the past and present) if there is a "data set of sufficient scope" (p. 121). However, for policy makers the real problem is the validity of the future forecasts,

> "What is needed .... is more intensive research on the dynamics and modelling of turning points in policy situations . Possibly this can be done by more intensive and higher calibre study of futures forecasting." (p. 121)

(ii) It is not meaningful to attempt to validate a model completely, as the real world is too complex.

If futures forecasting becomes more successful, for instance by examining the "turning points" that cause qualitative changes in the system, House notes that it is quite likely that historical validity may conflict with futures validity, and consequently this throws doubt upon the value of historical validation for models used by policy makers.

House lists six points which are pertinent to the validation of present-day social science models. These include: problems of historical validation; difficulties with standard statistical testing - these require "a definition of reality after the fact" (p. 122); problems of measurement of present and past, let alone of variables pertinent to the future; complexity of models; lack of any general laws concerning social systems.

In using models to help make decisions,

> "... the real questions are therefore not whether man should place confidence in such devices, but how good they are compared to what else is available." (p. 123)

One constituent of model validity is therefore relative predictive ability of a model versus other models. The "mental models" of policy makers could also be included,but, at present,the validity of a policy maker is assessed ex post facto (if his judgements prove to be invalid, he probably loses office), whereas the validity of models is often determined prior to their use.

The paper concludes with a proposal for a standardised model reporting format,

> "A good reporting format would at least inform people with an interest as to where a particular model fits in the field

and would allow them to evaluate the model on the basis of actual performance rather than on promised results." (p. 124)

Such a format would typically consist of:

Basic description of model
Subject matter of model
Modelling technique
Computer aspects of the model
Validation of the model
Model use and transferability.

House gives a good treatment of the problems of validity for models that are required to assist policy makers. These models need good predictive ability, and this cannot be simply assessed by validating over past data and extrapolating into the future with statistical limits. The credibility of a model can be increased if it includes parts which represent changes, or turning points, in the system relevant to policy.

The idea of validity that House puts over is one of output validity (cf. positive economics), where the assumptions and form of the model are not directly tested. If these, as well as the outputs, are validated, then confidence in applying the model outside the validation interval is considerably increased. This deductive aspect of model validity is implied when House requires the turning points of behaviour to be modelled as well, but he does not seem to be aware of the different methodological position (see, for example, Naylor et al., 1966, Section 2.3.2).

2.4.6 Model credibility for large-scale systems - Kahne, 1976

In this paper, Kahne separates model "credibility" into two distinct issues: validity and value; and studies the evaluation of large-scale models from the viewpoint of the buyer-seller interaction in the market place. The types of model Kahne has in mind are those that may be of interest to governments, for instance, World Models (Forrester, Meadows, etc.) or Leontief I/O models. By credibility is meant,

" .... capable of being believed." (p. 587)

The validity of a model is given by the closeness of its output vector to the system's "natural" output vector under an equivalent input. The value is assessed by the model's use in a particular situation (e.g. in making profit) and some direct considerations (e.g. computational efficiency).

Whilst this is a good point, it is part of a still larger one which Kahne ignores. Both validity and value must be assessed by the degree to which certain objectives or purposes of the model are satisfied. For validity we may define certain criteria of theoretical and empirical representation, and for value particular performance criteria or measures, and both sets of criteria are interdependent. There may be cases in which there are conflicting objectives (cf. computational simplicity versus precision), and so it is not meaningful to talk of a single "value", or to separate the issues of validity and value.

There are many points to disagree with in this paper and thereby refute Kahne's credibility hypothesis, yet he pre-empts this by stating,

> "This approach is not offered as a theory to be proved or disproved." (p. 587)

and, instead, recommends that it should be assessed by applying it to a number of modelling situations. Of course, any empirical methodology should be tested out practically - this is its purpose; on the other hand, with issues such as validity, the methodologies must receive a considerable degree of critical conceptual and theoretical evaluation as well.

2.4.7 A critical factor for model validity - Karplus, 1977

Walter Karplus, in his paper "The Spectrum of Mathematical Modelling and Systems Simulation" (1977), discusses the methodology of modelling in the physical, life, and social sciences. Models, Karplus states, in the physical sciences tend to be derived structurally from basic laws and insights in a process of deduction, whereas those in the social sciences are formulated often from the basis of system input-output measurements (i.e. induction). From these two extremes, Karplus proposes a spectrum of mathematical modelling according to the degree of induction/deduction required in model construction.

A critical factor determining model validity is the derivation of a model on a hypothetico-deductive basis rather than a purely inductive basis. This appealing approach also finds support in the analysis of measurement theorists.

In his early work on system identification and pattern recognition (1972), Karplus suggests a technique whereby important structural information about the system can be determined by applying a pattern recognition

algorithm to system measurements. A very similar approach has been taken by biological modellers (Bali et al., 1976; Carson and Finkelstein, 1977; see Section 2.5).

Although Karplus (1976) is not concerned with an analysis of the concepts of model validity (and their criteria), but rather with the reasons for the differences in mathematical modelling in a variety of sciences, he provides a good, but perhaps over-simplistic, framework in which to do so.

2.4.8 Berlinski's critique - 1976

In considering questions of the validity and validation of mathematical models, it is important at some stage to consider the validity of general techniques and concepts used in modelling. This is especially so for the social sciences, where many of the modelling techniques have been borrowed from the physical or engineering sciences. Berlinski's "On Systems Analysis" is a consistent negativist critique of the methods of systems analysis and mathematical modelling as applied to the social, political, and biological sciences, and attempts to delimit the domains in which the methods are valid. Although the book appears to be a polemical attack against systems analysis, it contains many well argued points of criticism which should be answered. It is divided into three sections:

I. General Systems Theory

This concentrates on such aspects as the definition of a system and theories associated with such definitions. Berlinski focusses on the lack of logical rigour that occurs in systems theory, and the difficulty of the conception of "system" (e.g. as something which transforms inputs to outputs).

II. Dynamical Systems

This section discusses the limitations of the modelling of dynamical systems by sets of ordinary differential equations. Berlinski's general comments include: problems of existence theorems for solutions, desirability of analytic solutions, and geometric properties.

Berlinski argues that if models are to be a tool for understanding and explanation, then we should first try to understand the analytic properties of a model. If this is not possible, then simulation should be used to validate qualitative aspects of the model vis-à-vis the system before uncritically accepting their quantitative results. On this point,

he gives an example of the Klein-Goldberger model of the U.S. economy:

> "The Klein-Goldberger model of the United States economy consists of twenty-five difference equations ..... Its qualitative properties were not perfectly understood when it was first presented; analytical appraisals were restricted to artificial cases. Then the Professors Adelman simulated the system and ran it forward into the future to learn whether the model depicted the oscillations of a modern industrial economy, i.e. the so-called business cycles."

But with no additional constraints or shocks, the model was monotonic, and far from looping in cycles it surged ever forward,

> "So much the worse for the Klein-Goldberger model: a gap of such qualitative tractability between theory and reality is evidence, if anything is, that a model is wrong." (pp. 76-77)

(The validation of a model by comparing qualitative features derived analytically or by feature extraction, subsequent to simulation, is very important and discussed elsewhere in this review; see, for example, Section 2.5.4, and also in Chapter 5.)

In this section, there is also a long analysis of the "World Dynamics" type of model (Forrester, Meadows, etc.); both the form of representation and the lack of empirical control are questioned.

III. Mathematical Systems Theory

Amongst other areas, this section deals with control theory and its application to particular models. Because of the complex nature of political systems, Berlinski concludes that control theory is misapplied here, and it is wrong to assume that humans and their societies behave in such a way as machines have been designed to do.

Berlinski raises many substantial criticisms to mathematical modellers, particularly in the social and political sciences. If these modellers have meaningful concepts of validity and apply corresponding validation tests to their models, they will have gone more than far enough to answer them.

2.5 Validation of Biological Compartmental Models.

2.5.1 Introduction

In this section, some papers concerned with the validation of biological compartmental models will be reviewed. The use of this type of model in biology stems from the analysis of isotopic tracer kinetics in living organisms. Three main advances in this area were:

(i) Hevesey, 1923

The use of radioactive lead to demonstrate the uptake and loss of lead ions by the roots of Vicia Faba.

(ii) Zilversmit, 1943

The first quantitative treatment.

(iii) Sheppard, 1948

Introduction of the concept of a "compartment".

Together with theoretical advances in applying mathematical techniques, a greater range of isotopes became available for making tracer tests on biological systems. These two factors are the main ones contributing to the early developments in biological compartmental modelling, and demonstrate the relation between the growth of theory and concepts and the empirical process of measurement.

In the 1950's and 1960's the introduction of computers (Analogue: Hickey and Brownwell, 1954; Digital: Worsley and Lax, 1962) gave a considerable impetus to compartmental modelling, and the ability to simulate more complex systems, thereby achieving more realistic representations of biological systems.

In recent years, compartmental analysis has adopted many of the techniques of state variable dynamic analysis and has been applied to other biological systems (e.g. metabolic processes, circulation and respiration, pharmacokinetics, etc.). These models have become increasingly concerned with representing control aspects of biosystems in maintaining their internal environment (i.e. homeostasis), and for this the mathematical theory of control has been very useful. For an introduction to the concepts and techniques of biological compartmental modelling consult Atkins (1969) or, for a more detailed account, Jacquez (1972). Gold (1977), and Finkelstein and Carson (1979) consider mathematical modelling in biology more generally, and both contain short sections on model validation.

Questions of model validity and validation in this area are particularly interesting, because the models are a combination of known insights and theories, and of structure and parameters inferred from the data. Biological theory often provides only a partial basis for developing a model which may contain much theoretical innovation, and data, if available, are usually highly uncertain. Very little, however, has been written specifically on this subject. One of the papers reviewed below (Bellman and Åström, 1970, Section 2.5.3) is not about validity or validation, but about system identification, and is included because of the importance identification has assumed in the appraisal of compartmental models in biology (system identification is reviewed in more detail in Section 2.6).

As pointed out many times in this first review, model validity is related to the purpose of the model. In compartmental modelling the purpose may range from a convenient description of the data in mathematical form, to prediction for medical therapy, to educational purposes, to hypothesis testing, and to a general increased understanding and explanation of biological systems.

2.5.2 Testing of models - Berman, 1963

This paper ("The formulation and testing of models", 1963), which is concerned with aspects of the formulation and testing of compartmental models, is drawn from Berman's experience in the use of such models for interpreting tracer kinetics. Berman discusses both the qualitative and quantitative phases: the choice of model type, number of compartments, number of parameters, least squares fitting of the data, judgement of the fit, and judgement of the model.

Model testing is implicitly regarded as a problem of parameter estimation with the aim, for a valid model, to be a unique, consistent, estimate of the parameter values. To achieve this the difficulties of non-uniqueness and inconsistency in the estimates are considered.

For a model whose structure is known, parameters can be estimated using a least squares fit to input-output data. If the data contain inadequate information, then only a class of adequate models (and not a unique solution) can be determined. Berman proposes two methods to cope with this difficulty:

(i) Fix some parameters

(ii) Assign a presumptive precision.

(In the first it is often difficult to choose which parameters should be assumed.)

An "Inconsistent solution ..... occurs when the model chosen does not have sufficient freedom to adjust to the data" (p. 186), and may result in the model equations becoming singular.

An acceptable solution for the model is a local or global minimum of the least-squared error residual. (This approach is used by Bellman and Åström, 1970, in defining the concept of "structural identifiability", see Section 2.5.3).

For models of unknown structure, the simplest form compatible with the data should be chosen, and can be derived from an exponential fit.

In an effort to overcome the problems of non-uniqueness, Berman proposes two kinds of experiment:

(i) From the same initial conditions make measurements from other compartments

(ii) New and independent initial conditions may be introduced.

Another approach is to change the parameters in the system and repeat the experiment. Choose the model in which a minimal number of parameters need changing to give adequate responses in both cases. This is known as the "minimal change postulate".

Berman's approach to model validity is one of achieving a unique set of model parameters by estimation from the data, and this characterises the prevalent attitude towards validity in compartmental modelling in general. There are problems with this approach, one of which is that of data uncertainty,

> " ... the uncertainty in the data must be reflected in the parameters of the fit regardless of the method." (p. 192)

Similarly, the data uncertainty may lead to the inability to discriminate model structure.

A further problem is associated with the implicit assumptions of fitting a model. The major one (in most cases) is that of linearity, and many biological phenomena are essentially non-linear, leading to insoluble problems of non-uniqueness if the model is to be evolved from the data alone. Attempts to resolve some of these difficulties have been made by Carson and Finkelstein (1977, see Section 2.5.5), and Mehra (1980), for example.

2.5.3 Identifiability and model validity - Bellman and Åström, 1970

In 1970 Bellman and Åström published a paper in which they introduced a new concept, "structural identifiability". The concept is useful in answering such questions as: To what extent is it possible to get insight into the internal structure of a system from input-output measurements? What experiments are necessary to determine the internal couplings uniquely?

The emphasis is on compartmental models, and the approach, like Berman's (Section 2.5.2), is to produce a unique solution to minimising a loss function, $V(\theta)$, between the theoretical model response and the data.

Bellman and Åström define an "identifiable structure" thus:

> "Assume that the measured output is generated by a system $S_o \in \mathcal{S}$ with parameters θ_o. The structure $\mathcal{S}$ is called (locally) identifiable if the function $V(\theta)$ has a minimum at $\theta = \theta_o$. If the minimum is global, the structure is said to be globally identifiable." (p. 332)

A sufficient condition for a structure $\mathcal{S}$ to be identifiable is,

> "that the matrix of second-order partial derivatives with respect to the parameters, $V_{\theta\theta}$, is positive definite for all $\theta \in \tau$." (p. 322)

They then consider this criterion applied to linear structures characterised by:

$$\frac{dx}{dt} = Ax + Bu, \quad y = Cx \qquad \ldots\ldots\ldots (2.3)$$

where the matrices A, B and C are assumed to have constant coefficients and $A \in \mathcal{A}$, $B \in \mathcal{B}$, $C \in \mathcal{C}$, where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are classes of matrices, specifying all internal couplings of the system. It is assumed that the system is observable and controllable (Kalman, 1963).

Since the input-output relation of a system initially at rest is uniquely given by the impulse response, "identification can be achieved from an impulse response measurement." (p. 333)

Some specific linear structures are discussed: Diagonal, Companion, all states observable. The results are then applied to compartmental structures. They show that a system of n cascaded compartments is identifiable if a tracer is injected into compartment n and if measurements

are taken from the same compartment. However, it is not possible to determine the parameters (rate coefficients) if the experiment is arranged in a different way.

If measurements are taken from all compartments, then it follows that,

> " ..... if the tracer's injection is chosen in such a way that the system is controllable, then the corresponding structure will be identifiable no matter how complex the interactions are." (p. 338)

Implications for validity

Identification is an "inverse problem" - making inferences from system data. Difficulties arise when using this to test or validate a model (as with Berman's, see Section 2.5.2). These include:

(i) The model is assumed to belong to a certain class of structures before identification (linear time invariant)

(ii) No consideration of measurement problems (e.g. scarcity and uncertainty).

To include the results of "structural identifiability" a meaningful validation procedure, the assumptions (as to model class, etc.) and the limitations imposed by data uncertainty should themselves be investigated.

However, Bellman and Åström's paper is very important in that it defines a class of soluble identification problems and shows the limitations of the purely inductive approach to modelling.

This was also argued by Carson and Finkelstein (1973):

> "Good models must be based on 'a priori' knowledge rather than the'black-box' approach." (p. 201)

by considering the problems of non-unique models identified from noisy data (using an example of Albumin metabolism). Further aspects of system identification and its relation to model validation are considered in Section 2.6.

2.5.4 Feature space and pattern recognition techniques for validating biological compartmental models

An alternative method of gaining insight into a system's internal structure is to examine general features or patterns of its behaviour. If these are found to occur in the model as well, this may be an indication of structural (or broad) validity.

Karplus (1972) proposed a pattern recognition method of system identification on the grounds that conventional methods break down when the model is not close enough to the data, or the data are of low quality, and that combining all the data points into one objective function may smooth out the effects of important features in the data.

A practical application of the use of feature space techniques is that by Pullen (1976). In validating a dynamic model of the human cardiovascular system, he proposes a systematic procedure for validating non-linear models based on the comparison of features in a number of important tests. This model is used in the first case study in model validation (Chapter 6) and feature comparison techniques are used extensively.

An example of the use of pattern recognition techniques in model validation is that by Bali et al. (1976) in the validation of a model of the human respiratory system. The features chosen were sample values of the ventilation rate response to percentage step changes in inhaled CO_2. A linear classifier was used to classify model responses according to the input level. The data provided a training set to construct the classifier, which was then applied to the model responses under the same inputs. This method was capable of discriminating between three candidate sub-models for respiratory control, only one of which was found to produce correct model response classification for all inputs. (There were additional theoretical and empirical reasons for the choice of controller submodel and so the pattern recognition results provided additional confirmation rather than a sufficient test.)

The rationale behind these methods of validation is that particular feature sets, or patterns of response, are associated with certain classes (and modes) of system. Theoretical work on this form of representation for dynamic systems is an essential precursor if they are to be adopted in criteria for model validity (Leaning, 1979).

2.5.5 Validation of control system models in physiology - Carson and Finkelstein, 1977

This paper is an attempt to resolve some of the problems of validation of mathematical (control system) models in physiology. These problems, which are regarded as those of achieving unique model structures and physiologically feasible parameters, reflect the degree to which the methodology of compartmental modelling is reliant upon the techniques of system identification. A model is built up from a priori knowledge as much as possible; the rest must come from identification and parameter estimation.

The purposes of modelling physiological systems are discussed, and this leads to a definition of model validity:

> " ..... [a] model can be said to be valid if it not only describes those aspects of structure and behaviour which entered into its formulation, but also predicts correctly all relevant behaviour of the system under future tests." (p. 2)

The combination of inadequate structural knowledge with bad initial estimates of parameters and low quality and insufficient data available from the physiological system results in a non-unique model solution to the identification problem. A valid model is equated with a unique solution, and consequently the method of identification is inadequate for model validation.

Carson and Finkelstein propose two solutions:

(a) Further studies should be made on the subsystems (i.e. increasing the deductive validity of the aggregate model, see Section 2.4.7)

(b) Using feature space and pattern recognition techniques to compare system and model.

The first approach is one which emerges time and time again in work on model validity in the social and biological sciences (see, for example, Berlinski, 1976, Section 2.4.8, and Karplus, 1977, Section 2.4.7) and indicates that a good modelling procedure is one based both on a priori understanding of the system as well as inductive identification and parameter estimation techniques. The second approach was discussed in Section 2.5.4 as a method for validating compartmental models of cardiovascular and respiratory systems. For example, Attinger et al. (1977)

use cluster analysis in the validation of a hierarchical biological control model of the respiratory system.

A point is made by Carson and Finkelstein (1977) that, "..... having constructed a mathematical model, the temptation exists to attempt simultaneously to validate the model structure and estimate its parameters. This cannot be achieved". However, from the covariance matrix of parameter estimates conclusions on the structural validity of the model can often be made. Another way to check structural validity is through consistency of parameter estimates (discussed in Chapter 5). Nevertheless, it is true that greater confidence can be placed in the validity of a model if it is estimated from one set of data and validated against a different set. If only one set of data is available, the date should be split, " ..... so that some are available for validating structure, whilst others are used to estimate the parameters of a given model". An objective method of splitting the data (e.g. as in statistical cross-validation, Section 2.8.2) should be used, so that the validity of the model could not be suspect on this account (Mihram, 1976).

2.5.6 The validity of an arterial system model - Ohley et al. (1980)

Ohley et al. (1980) derived a mathematical model of an arterial system based on a finite difference solution to the Navier-Stokes equations which retains some nonlinear features. To validate the model, a series of functions was used to define the difference between the dependent variables of the model and corresponding haemodynamic variables of the animals. The variables chosen were mean diastolic pressure and cardiac output. These difference functions were displayed as a family of three-dimensional surfaces which related the function value to parametric changes of the independent variables (heart rate and peripheral resistance). Experimental data from dogs were compared with model predictions. By examining the difference surfaces, Ohley et al. were able to determine regions of validity for the independent variables.

They found that validity in certain regions in one animal does not necessarily guarantee model validity in these regions for other animals. Furthermore, validity for one dependent variable in a given region does not necessarily imply validity in this region for another dependent variable. This suggests that the model requires modification.

Visual techniques of comparing model and data are very good for

model validation, and many features can be noted which would be missed by a standard statistical test. The use of families of curves or, even better, three-dimensional surfaces, can therefore be used as critical tests of model validity. However, in large models, there is a complex interdependency between variables and it is not generally possible to determine a subset of only two major dependent variables (maximum allowable for graphical display). There are also further complications in models of human physiological systems - it is not always possible, for ethical reasons, to vary systematically parameters or variables and hence build up complete graphs. This leads to further uncertainty, and the need for a range of different validation tests (e.g. consideration of the theoretical and empirical validity of the elementary submodels, assumptions, etc.).

2.6 System Identification and Model Validation

2.6.1 Introduction

System identification refers to the procedure of determining a model, and its parameter values, of a system on the basis of input-output data. The usual method, known as parameter estimation, is to adjust the parameters until the model response is as close as possible to the data, whilst assuming a certain model structure. The structure can be varied also until a "best fit" is obtained. The origins of the technique were in the estimation of economic variables or parameters, which could not be directly measured, using econometric models. The use of system identification for biological compartmental models was briefly discussed in Section 2.5. These models are usually formulated using understanding of the basic unit physical and chemical processes in the biological organism. However, data for such models are often only available from experiments on a limited number of input-output ports, and therefore system identification offers a method of determining parameter values and resolving structural uncertainty in the model. There is an immediate problem, which is essentially theoretical, concerning the extent of the correspondence between a given model structure and set of parameters and a certain input-output transformation. This problem is a general one in the application of system identification and stems from the lack of theory and direct measurability of a system required to formulate a model completely (other classic areas of application are aircraft dynamics and nuclear reactor physics). The problem was considered by many workers (e.g. Berman, 1963, Section 2.5.2), but it was not until 1970 that it was well-posed by Bellman and Åström (Section 2.5.3). They introduced the concept of structural identifiability: a model is said to be structurally identifiable if there is a global minimum of the loss function (between model response and data), i.e. the structure and parameters of the model can be determined uniquely from input-output data. This was applied to the class of linear compartmental models, and it was found that only a limited number of experimental designs (i.e. input/output sites) produce structurally identifiable models.

In this section, some of the concepts and methods of system identification will be reviewed (using DiStefano and Cobelli's, 1980, distinction between a priori and a posteriori aspects) followed by a discussion of the relationship between system identification and model validation.

2.6.2 A priori (or theoretical) identifiability

Since Bellman and Åström defined structural identifiability (1970) there has been a proliferation of many different concepts and new definitions in the system identification literature accompanied by a considerable degree of confusion. The paper by DiStefano and Cobelli (1980) provides a new set of simple definitions which includes most important points. A priori identifiability is concerned with the uniqueness of the correspondence between a model and locations for input and output regardless of considerations of system data. In fact, the system does not need to exist at all. (The term "theoretical identifiability" is perhaps a better one than "a priori identifiability".) One of the formal difficulties in a priori identifiability has been the inclusion of a priori information, for instance constraints on parameters, but this is avoided by DiStefano and Cobelli (1980) by using an extended model formalisation known as a "constrained structure". A constrained structure, with parameter vector $\underline{p}$, on the observation interval $t_o \leqslant t \leqslant T$ is defined by:

$$\dot{\underline{x}}(t, \underline{p}) = \underline{f}[\underline{x}(t, \underline{p}), \underline{u}(t), t; \underline{p}] \quad \ldots\ldots\ldots (2.4)$$

$$\underline{y}(t, \underline{p}) = \underline{g}[\underline{x}(t, \underline{p}); \underline{p}] \quad \ldots\ldots\ldots (2.5)$$

$$\underline{x}_o = \underline{x}(t_o, \underline{p}) \quad \ldots\ldots\ldots (2.6)$$

$$\underline{h}[\underline{x}(t, \underline{p}), \underline{u}(t), \underline{p}] \geqslant 0 \quad \ldots\ldots\ldots (2.7)$$

where $\underline{x}$, $\underline{y}$, $\underline{u}$ are the state, output and input vectors, respectively; $\underline{f}$ is a nonlinear vector function which defines the known input-state couplings between $\underline{u}$ and $\underline{x}$, parameterised by $\underline{p}$; and $\underline{h}$ represents all additional and independent algebraic constraints known a priori relating $\underline{x}$, $\underline{u}$ and $\underline{p}$.

Definitions:

1. The single parameter p_i of the constrained structure is said to be "unidentifiable" on $[t_o, T]$ if there exists an infinite number of solutions for p_i from these relationships. If one or more p_i is unidentifiable, then the model is said to be "system unidentifiable".

2. The single parameter p_i of the constrained structure is said to be "identifiable" on $[t_o, T]$ if there exists a finite number of solutions (> 0) for p_i from these relationships. If all p_i are identifiable, the model is said to be "system identifiable".

3. The single parameter p_i of the constrained structure is said to be "uniquely identifiable" on $[t_o, T]$ if there exists a unique solution for p_i from these relationships. If all p_i are uniquely identifiable, the model is said to be "parameter identifiable".

(from DiStefano and Cobelli, 1980.) System identifiability (2) and parameter identifiability (3) correspond to Bellman and Åström's local and global identifiability (1970).

Linear models have general analytic solutions and there are a number of tests for checking the uniqueness of the parameter solutions (such as the transfer function and Markov parameter matrix methods). Although these can involve complex algebraic manipulation, it is always possible to determine the a priori identifiability for linear models.

For nonlinear models, although the above definitions still hold, there are no general analytic solutions and therefore the a priori identifiability cannot generally be determined. However, if the linearised (first order Taylor expansion) model is system unidentifiable then it follows that the full model is also system unidentifiable. In practice, in identifying nonlinear models it has to be assumed that the model is sufficiently constrained (by a priori knowledge) that the parameter estimation will yield a global minimum.

If a model has satisfied a priori parameter identifiability, then it can be used in parameter estimation algorithm with data obtained from the system. The model is then subject to considerations of a posteriori identifiability.

2.6.3 A posteriori (or practical) identifiability

There are many methods of system identification. Usually these involve combining model outputs and system data in a loss function (sometimes a likelihood function or Bayesian probability estimate) and optimising this function in the model's parameter space (a standard reference is Eykhoff, 1974). If the model structure is uncertain, then this too can be varied until the optimum model structure and parameter values are obtained. Since more complex higher-order models can also exhibit the same behaviour as lower-order models (degenerate behaviour) the minimal order model is usually accepted.

In practice, the uncertainty of the data (e.g. sampling errors, noise disturbances, etc.) and of the model structure as a representation

of the system (e.g. the number of state variables) leads to uncertainty in the parameter estimates, which may not be unique even if the model is a priori parameter identifiable. Carson and Finkelstein (1973) showed in simulation studies that uncertain data (with as little as 6% noise), combined with inaccurate initial parameter estimates and structural uncertainty, can lead to the identification of an incorrect model structure and incorrect parameter values, even though the model was theoretically a priori identifiable. A similar examination was conducted by Cobelli and Salvan (1976) on the various factors affecting the reliability of parameter estimates; these included: the order of the model, the effects of sampling and random errors, the type of estimation algorithm, and initial parameter estimates. Brown and Godfrey (1977) tackled the problem analytically and illustrated that a model which is a priori parameter identifiable can become system identifiable or even system unidentifiable if the data have a finite uncertainty interval. They termed the problem of a posteriori identifiability and model validation as "determinacy". Even quite simple linear models can become indeterminate and Brown and Godfrey suggested that the only solution is to obtain measurements from other locations in the system. The requirement for greater understanding of subsystem dynamics was stressed by Carson and Finkelstein (1977) who also suggested the use of feature space techniques in order to extract additional information from the data.

When a parameter estimate has been made, it is possible to determine a measure of confidence in the estimated values, even though the estimate may only be a local as opposed to global minimum. The most informative measure is a Bayesian probability, but it is rarely possible to acquire sufficient information to evaluate this. The Cramér-Rao lower bound gives the maximum achievable accuracy on a parameter estimate in terms of the likelihood function (which is a substitution of the model/data error residuals into the probability distribution for the data noise. In the maximum likelihood estimator, the model parameters are adjusted until the likelihood is maximised.) The Cramér-Rao inequality gives a lower limit on the covariance matrix of the parameter estimates if the estimator is unbiased (Eykhoff, 1974) equal to the inverse of the Fisher information matrix. The information matrix depends on the likelihood function which, in turn, is a function of the input and therefore the attainable accuracy of parameter estimates can be maximised by optimal input design.

If a least squares error estimator is used, an estimate of the uncertainty of the optimum parameters can be obtained by sensitivity analysis, and the assumption that the error residuals are Gaussian. The variance-covariance matrix of the parameters is very useful for model validation.

2.6.4 The relationship between system identification and model validation

The role of system identification in model validation is a function of the purpose of the particular model. A model may be a priori parameter identifiable, fit closely to the data, and yet may not represent the internal structure of the system at all. For instance, the model may be linear whilst the system contains significant nonlinearities. However, if the purpose of the model is simply to reproduce the input-output characteristics of the system (e.g. for a limited range of inputs) then for this purpose it may be valid, and identification equals validation.

If the model is to be used for future prediction, then validation requires an estimate of confidence in the predictions as well as comparisons of the predictions with new data (i.e. beyond the identification interval $[t_o, T]$. The uncertainty covariance of the predictions can be transmitted through the model from the variance of the parameter estimates, and for validity a critical maximum acceptable level can be specified. However, if the model also represents the internal structure of the system, then much greater confidence can be placed in its predictions. If a significant discrepancy is found between the predictions and new data then it can be concluded that the parameter estimates are only a local minimum, or, if they are a global minimum, that the model structure is incorrect.

If the model is intended as a representation of both internal and input-output characteristics, then the additional validation tests after identification are much more stringent. For instance, with linear models tests of linearity should be made (e.g. by using different inputs and testing for linear superposition). Important qualitative and quantitative features of the data should be reproduced in the model; even in linear systems much information on system structure (and parameters) can be revealed in this way (Leaning, 1979), for nonlinear systems the range of qualitative behaviour is very varied (e.g. limit cycles, jump effects, saturation, bifurcation phenomena and the loss of structural stability; Thom, 1975) and provide a good starting point for nonlinear system identification (Mehra, 1980). However, care should be taken in inferring a

mathematical model from qualitative aspects alone; Sussman and Zahler (1978) demonstrated the misuse of applied catastrophe theory models in biological and sociological sciences on the basis of this type of behavioural data.

The accuracy of the parameter estimates can be used to test further the validity of the model's structure. For instance, if repeated parameter estimates are made under different conditions (e.g. inputs) and the estimates are inconsistent (i.e. finite local solutions) then, if the model is a posteriori identifiable, it must be concluded that the structure is invalid. Sometimes, however, the inconsistency may be simply due to the unidentifiability of the model and the estimator locating various local solutions. The variance-covariance matrix provides yet more information and can be used to check both a posteriori identifiability and model structure, i.e. validation (e.g. Cobelli and DiStefano, 1980).

On the whole, if a model has a good fit to the data (i.e. small residuals), has small parameter variances and, especially, covariances, and also exhibits interesting qualitative aspects, this is good evidence for the validity of the model as a representation of the system. The importance of model validation in system identification is widely recognised in the recent literature; in the application to metabolic modelling (Carson, Cobelli, and Finkelstein, 1980) and in theoretical aspects of nonlinear system identification (Mehra, 1980), for example. Mehra (1980) distinguishes between "theory-based" and "data-based" validation tests and recommends that "it is best to apply a whole battery of tests to each model to reveal its theoretical and statistical deficiencies". Theory-based tests include simulation tests, sensitivity analysis and modal, eigenvalue and stability analysis. Miller (1974) gave a method of validation based on parameter sensitivity analysis which requires that a deviance measure on the estimated values is less than a previously defined critical value, but this is too simplistic. Sensitivity tests can clearly reveal the limitations and may be applied with respect to parameters, inputs, structural changes and the data.

2.7 The Concept of Model Adequacy

2.7.1 Introduction

The concept of model adequacy was conceived with the aim of avoiding philosophical problems associated with validity. It is usually defined with respect to a given set of models; the most adequate model being the one which is closest to a given data vector and also satisfying a pragmatic objective, such as simplicity.

2.7.2 Vector Approach to Model Adequacy - Reggiani and Marchetti, 1975

Model adequacy is treated by Reggiani and Marchetti (1975) as the inverse problem:

> "The question consists in finding in $\{\mathcal{M}_t\}$ [a set of models] the most adequate $\mathcal{M}$ to represent Θ [a given object]." (p. 322)

They describe a procedure which is based on the definition of a vectorial distance between models. It allows one to order models from the standpoint of their adequacy and apply to the problem many results of ordered sets.

A "single output can only rarely be considered sufficient to define the adequacy of a model" (p. 323) and so "the distance between two models (or else between a model and a given object) must be vector valued.". The concept of a vector distance is supported by the mathematical study of pseudometric spaces.

Once the vector distance between two models has been defined, it is possible to define the distance $\overline{\mathcal{M}\Theta}$ between a model $\mathcal{M}$ and the object Θ as well. This distance is known as the "modulus" $|\mathcal{M}|$ of the model (it is a vector). A definition of model adequacy follows:

> "..... we shall consider $\mathcal{M}_1$ more adequate than $\mathcal{M}_2$ in representing Θ if the following relation
>
> $$|\mathcal{M}_1| \leq |\mathcal{M}_2| \text{ holds."} \quad \text{(p. 324)}$$

(n.b. the inequality holds component-wise)

This inequality orders the set $\{\mathcal{M}_t\}$ and may be used to select the most adequate model.

The paper concludes with an application of the method to an example from transistor theory. The particular problem is:

" ..... for each transistor in a circuit C, find the model of least complexity that will give acceptable accuracy." (p. 326)

and is based upon the work of Lindholm (1971).

The vector distance $|\mathcal{M}|$ has n+1 components, where there are n transistors in the circuit. The first n components are assigned the value 0 or 1 according to whether a particular transistor model is valid or not in its particular location i, $i \in [1,n]$ (under the i^{th} operating conditions). The $n+1^{th}$ component is an integer giving the total number of parameters required in the whole model.

The requirement for model adequacy is that the first n terms of the distance $|\mathcal{M}|$ are zero, and the most adequate model is the one which requires least parameters (i.e. minimum $|\mathcal{M}|$).

The example requires a slight modification of Reggiani and Marchetti's framework, but on the whole establishes its applicability.

Implications for validity

To use the method in model validation two further requirements are necessary:

(i) The functions which comprise the model distance $|\mathcal{M}|$ must be shown to be a proper representation of the system. This will be related to the model's purpose. For instance, if the model is required to demonstrate and explain dynamic behaviour, then the vector distance should contain pertinent dynamic components (if possible, those which can discriminate model structures and be used in an estimation algorithm).

(ii) The data obtainable from a system or object Θ will contain uncertainty, making the exact determination of $|\mathcal{M}|$ problematic. If the theory is recast in a probabilistic form, this may be overcome.

In conclusion, the paper places the question of model adequacy in a formal framework and is a valuable contribution to the work on model validity.

2.7.3 A methodology for the evaluation of model adequacy - Argentesi (1978)

As discussed in Section 2.7.2, the approach of Reggiani and Marchetti (1975) is rather simplistic, and the elements of the vector, V, characterising the model can be given any interpretation. Argentesi (1978) extends the concept of model adequacy by classifying these elements of V

into the following types:

1. Global fitting index
2. Partial fitting indices
3. Simplicity index (e.g. number of parameters)
4. Sensitivity analysis and error analysis indices
5. "Relevant and characteristic indices that are related to the occurrence in the model of an important feature of θ [the measurements/observations of object θ_o]".

If the indices are appropriately chosen, all elements of V for the data θ will be zero. Thus minimising the vector V is equivalent to minimising $|V(M) - V(\theta)|$, as in Section 2.7.2. Essentially, this requires that an adequate model has both a good overall (or macro) fit (index 1), a good fit of its submodels (index 2), is economic (index 3) and nevertheless reproduces key features observed in the data (index 5) as well as satisfying some error criteria (index 4). Far from avoiding the problems of the concept of validity, this forms an implicit and fairly reasonable theory of model validity.

It can be seen that, in modelling complex phenomena, there will be a conflict between accuracy and simplicity of a model and it is possible that there may not be a unique model which satisfies the criterion of minimising V. In general, there will be a trade-off between minimising any two types of indices, and a multiobjective decision function is required if the problem is to be solved. This would involve a great deal of formalisation and the need for this can be questioned. The choice between competing models is best made by devising critical theoretical or empirical tests. Frequently some alternative models are merely contrivances in order to demonstrate the validity of the main model. Anyway, it is a good maxim in model validation to keep close to the model and data and to introduce as little formalisation as possible. Perhaps the major contribution of Argentesi is to stress the multidimensional as opposed to unidimensional nature of model validity.

2.7.4 Model set compatibility - Baram, 1977

Baram (1977) proposed that a set of models could be represented by an information distance measure, where the distance between two models is the information required to tell them apart in parameter space. If the distance between models is "long" the model set is said to be (a priori) "separable". The information distance between models and the data is

defined on the output space (in a posteriori validation). The model set may be considered incompatible if the shortest distance between the actual system (data) and the set is greater than the span of the model set. The closest model to the actual system may be considered valid, and the model set compatible, if its distance from the actual system is of the same order of magnitude or smaller than its distance from close models in the model set.

The concept of an information distance is both intuitively and mathematically appealing. However, the development of the theory is limited as yet to linear models and system data which is stationary and ergodic. A further application of the information distance would be in devising critical experiments in which the distance between competing models was maximised.

2.8 Cross-Validation, Energy Model Assessment, and the Validity of Models in the Physical Sciences

2.8.1 Introduction

In this section, three additional aspects will be considered which do not fall clearly into any of the previous sections: the "cross-validation" of statistical regression models (Section 2.8.2), the MIT energy model assessment program (Section 2.8.3), and some general comments on the validation of models in the physical sciences (Section 2.8.4).

2.8.2 Cross-validation of statistical regression models

The method of cross-validation is to split the data into an estimation set and a prediction (or evaluation) set. The coefficients of the regression model are determined from the estimation set, and the predictive validity of the model is determined by comparing the model predictions with the data values in the prediction set. An early paper is by Mosier (1951), later work (e.g. McCarthy, 1976) contains algorithms for splitting the data in ways which ensure that the two sets have similar statistical properties (e.g. Snee's "DUPLEX", 1977). Although cross-validation is essentially concerned with the predictive validity of statistical regression models, similar data-splitting algorithms could be used to resolve the problem of a single data set for both model structure identification and parameter estimation (Carson and Finkelstein, 1977, Section 2.5.5; see also Section 2.6.4).

2.8.3 The MIT energy model assessment program

In recent years, the energy laboratory of MIT has set up a model assessment program concerned with the validation of the complex models they have developed in energy systems research (Wood, 1978). These models are used to predict the future state of the energy system given certain policies and are intended as aids to decision makers. A recent paper by Gruhl (1979) outlines their general conceptual and methodological approach to model validation, and is summarised below.

Validation is described as the "formulation and resolution or evaluation of questions and hypotheses about the appropriateness, with respect to a specified set of applications, of the model's logical structure and the validity of its implementation". Energy models cannot be validated by repeatedly comparing their outputs with data from designed

experiments, since they are based on a single past historical tract. "Instead of the ultimate and easy comparison of end results, such models must be assessed by checking all of the steps in their evolution and use". The first focal point for assessment activities is "points of potential corruption" in passing between development stages in model construction. The second step is the assessment of the sources of potential inaccuracy in model utilisation. Gruhl gives two flowcharts for model development and use and lists alongside the appropriate assessment techniques as well as suitable documentation.

An array of validation techniques is presented for use in assessing the extent of uncertainties introduced at the points of corruption. These techniques involve two steps: first, some piece of the model is examined or changed, and then this action is evaluated with respect to some basis for comparison. A typical change would be a parameter or structural variation with the basis for comparison, "examination of reasonableness and accuracy" (the standard sensitivity analysis). Gruhl makes the point that it is important to assess the potential inaccuracies in a model with respect to a specific application.

Finally, the strategy of the assessment process is considered. Although many techniques are available for assessment, in practice it is constrained by available manpower resources and funding:

> "The development and choice of the most cost-effective validation techniques is virtually a new field of research. It seems as though systematic techniques and procedures may well be the most appropriate, but they will be difficult to develop as they are likely to be both model and application specific."

2.8.3 Validity of models in the physical sciences and engineering

Mathematical models are used extensively in the physical sciences and engineering, both in research and industry, yet the literature of modelling in this area is virtually devoid of detailed treatments of the nature of model validity. Why?

Models are used to design chemical plant, to test theories in plasma physics, and many other important applications. Most are successfully used for their intended applications, and there is no higher measure of validity.

The validity of these models arises from the methodology used to

formulate them. They are based on a well-established body of theories and models, and applied to particular systems whose physical and geometrical properties are well-known. This a priori input of deductive certainty is the critical factor for model validity (see Karplus, 1977, Section 2.4.7).

The questions of validity and confirmation (in physics, chemistry, etc.) are usually dealt with at the level of theory and hypothesis and the comparison (with empirical data) relies upon a great deal of additional theory and experimentation. The relation between theory and experiment is consequently complex, and many philosophers of science have grappled with this question (e.g. Popper, 1935; Braithwaite, 1953; Bunge, 1973; see Chapter 3).

The modeller has all this theory and knowledge at his disposal and constructs, on a deductive basis, a model of high validity. The problems he faces are essentially those of "top-up" validity, or of polishing of the final article. Here the techniques of system identification to resolve structural uncertainties and estimate parameters have been invaluable, and applied to an incredibly wide range of problems - aeronautical, nuclear, chemical, process, etc. The literature is vast (see, for example, Eykhoff, 1974; or any of the IFAC Symposia on Identification and System Parameter Estimation). Occasionally, these techniques have been referred to as validation (or "verification", Leal, 1977), but the real validity comes from the deductive base, not the estimation of a few parameters.

There are some exceptions to the above general remarks. One is in the area of artificial intelligence and robotics. An example of the latter is given in a paper by Siklossy and Roach (1975). They devised a computer program (DISPROVER!) to test models of robots to see if certain robot tasks are possible according to various conservation laws.

A paper which considers validation in more depth is that by Chrostowski et al. (1978). This deals with the validation of models of "mixed" physical systems (e.g. containing fluidic, mechanical, electrical, etc. components), and divided validation into two parts:

(i) checking of the model configuration,
(ii) parameter estimation,

and relates the closeness of model and system responses to the data uncertainty. This is the identification method again (see Section 2.6).

The methodology of validation in simulation modelling has recently been applied to missile system models in the USA (Kheir, 1976; Kheir and Holmes, 1978).

2.9 Conclusions

The review has exposed the general confusion surrounding the nature of model validity and the wide range of techniques available for model validation. Despite the confusion, there is a central core of meaning of, and conceptual approach towards, model validity that emerges in the better work in many different application areas: model validity is a multi-dimensional concept closely related to the modelling objectives, the nature of available data, and the overall process of modelling. This is explained in detail below:

Kaplan's approach to model validation (in the behavioural sciences, 1964; Section 2.4.2) related model validity to the instrumental purpose a model was intended to serve. This was elaborated in depth by Hermann (for political modelling, 1966, Section 2.4.3) who distinguished various types of modelling objective (scientific, practical, educational, etc.) and the various ways in which a model may be compared with empirical data, which were expressed as validity criteria. Each criterion may be related to a set of modelling objectives and represents a different concept of model validity. The importance of modelling objectives on validity was also stressed in biological modelling by Carson and Finkelstein (1977, Section 2.5.5); and the variety of empirical validity criteria, or bases for comparison, by Gruhl (in energy system modelling, 1979, Section 2.8.3). In comparing a model with empirical data, considerations of the validity and scope of available data are very important (behavioural sciences, Kaplan, 1964, Section 2.4.2; psychology, Nunally, 1970, Section 2.4.2; biological modelling, Berman, 1963; Section 2.5.2; Carson and Finkelstein, 1977, Section 2.5.5; statistical treatment in simulation modelling, Naylor and Finger, 1967, Section 2.3.4). The extent to which a model is based on previously validated theories, models or data and the influence of the overall modelling process has a profound effect on model validity but has received less attention (in systems science, Karplus, 1977; Section 2.4.7; in energy system modelling, Gruhl, 1979, section 2.8.3).

The similarity of the conceptual approaches towards model validity in such a range of application areas suggests that it may be possible to develop a theory of model validity which has general applicability. However, there are serious omissions in the literature reviewed which should be rectified before developing such a theory. These include: a complete classification of modelling objectives; the relation between

modelling objectives and the intended range of application of the model; other types of validity criteria, e.g. theoretical and heuristic (concerned with the understanding or discovering potential of a model), which are extremely important; the precise relation between modelling objectives and validity criteria; and a full analysis of the role of the available knowledge (theories, models, data, etc.) in a particular area. In Chapter 4, a theory of model validity is developed which is based on the common core of the review and which attempts to fill the omissions. This theory is used in Chapter 5 to devise a range of validation methodologies which systematically incorporate many of the techniques for validation that have been reviewed.

Throughout the review, model validity has been mainly treated as a simple correspondence with data. However, there are many scientific and philosophical reasons why this is an inadequate account, including the importance of theoretical coherency and heuristic potential. In addition, there are many philosophical questions which have been raised, such as "is it possible to prove the validity of a model?", or problems associated with epistemology. In the next chapter, therefore, a review is made of the philosophy of science with particular reference to the issues of model validity and validation. This will provide further insight into the nature of model validity (and also some aspects of systems science and measurement) and provide a philosophical basis for the theory of model validity which is developed in Chapter 4.

CHAPTER 3

A REVIEW OF THE PHILOSOPHY OF SCIENCE WITH REFERENCE TO MODEL VALIDITY AND VALIDATION

Review 2

3.1 Introduction

The scope of this chapter is an outline of the philosophy of science with reference to model validity and validation. In particular, the major historical developments, including the rather dramatic recent challenges to traditional views, some aspects of "systems philosophy", and the philosophy of measurement will be considered. The reasons for the inclusion of this chapter are as follows: firstly, investigations into the philosophy of science have been very fruitful in developing the approach adopted in this thesis; secondly, many issues of a philosophical kind are raised throughout the thesis which philosophical techniques and devices help to analyse; and, thirdly,the weaknesses in the philosophy of systems science. The aim, therefore, is a general review of the philosophy of science highlighting, where relevant, aspects related to the current topics of model validity and validation and systems science.

As a prelude to the review some basic philosophical terminology and ideas will be presented here. Philosophy has four main branches: logic, the theory of reasoning; epistemology, the theory of knowledge; metaphysics, the theory of concepts and their relations; and ethics, the theory of values (particularly moral evaluation). The philosophy of science is concerned with the logic, epistemology and metaphysics of science (although some systems workers have introduced ethics as well). In other words, the philosophy of science examines the reasoning processes of scientists and the logical properties of scientific knowledge, the way in which science provides knowledge about the world, and the nature of scientific concepts (particularly those higher-order ones called metaphysical presuppositions).

Scientific theories or models can be regarded as sentential in form - that is as sets of symbolic propositions which may be natural language sentences, mathematical equations, or, in general, well-formed strings from a symbolic system. Propositions are located in logical space by delineating their: (i) Syntax, (ii) Semantics, and (iii) Epistemological status. There are two distinct kinds of proposition. If a well-formed proposition Q is such that not-Q entails a contradiction in the symbolic

system, then it is: analytic (syntax); necessary, since its negation leads to a contradiction (semantics); and must hold a priori (epistemological status). On the other hand, if both Q and not-Q are allowed in the symbolic system, it is: synthetic (syntax); contingent or vulnerable, since its truth depends on something else (semantics); which is determined a posteriori by experience (epistemological status). Ordinary empirical propositions are synthetic, contingent and a posteriori. Analytic, necessary propositions are sometimes referred to as tautologies. Some classic problems in the philosophy of science arise from the fact that some scientific propositions (such as general laws) appear to have an a priori epistemological status (the famous Kantian "synthetic a priori" or self-evident facts) which contradicts the above dichotomy.

The review begins by considering the dominating influence that logical empiricism has had on the philosophy of science for the first half of this century (Section 3.2). Logical empiricism is a combination of the empiricist tradition in British philosophy and the positivism of the Vienna circle, and it attempts to reduce knowledge to the basic units of sensory experience or, at least, physical observation. It has a formal logical approach and tends to regard theories as axiomatic deductive systems. The focus of logical empiricism has been on the individual theory, its structure and validation, and therefore offers a static view of science. In logical empiricism great emphasis was placed on the principles of confirming (validating) or falsifying, theories, and it offered many prescriptive methodologies for this purpose. More recently, attention has been shifted from individual theories to series of theories and even to overall views of science as a dynamic process - the so-called "Weltanschauungen analyses". These new developments are an attack on basic positivist epistemology, and are outlined in Section 3.3, together with a formal approach to the structural dynamics of theories, and the latest historical realism in the philosophy of science which aims at demonstrating the rationality and reasoning patterns used in scientific practice. These latest approaches have little to say about the individual theory or model and how it should be validated. This is partly because the unit of analysis has become larger (groups of theories) and partly because of some radical new epistemologies. Despite this, it is these approaches which have been most fruitful in the work reported in this thesis, and model validation can only be correctly understood by regarding it as part of an overall research programme.

In Section 3.4 some aspects of systems philosophy and the philosophy of systems science are considered. The earlier sections (Sections 3.2 and 3.3) will have provided some critical philosophical devices that expose weaknesses in this philosophy. There are nevertheless some important philosophical problems associated with systems science and these are discussed. It is important, however, to distinguish between systems-inspired philosophies of the world, and philosophical analysis of the actual scientific practice of systems scientists.

In the theory of model validity developed in the next chapter, detailed attention is paid to the nature of system data, that is the measurements and observations of the modelled system. As an introduction, Section 3.5 is a review of the philosophy of measurement, in particular the history of measurement theory, and some recent trends and problem areas.

3.2 The Static View of Science - Logical Empiricism

3.2.1 Introduction - the origins of logical empiricism

The logical empiricist view of science originates in the British tradition of empiricism in the 18th and 19th centuries and the logical positivism of the Vienna circle (1920's - 30's). Hume's epistemology (1739, 1749) was an attempt to reduce all knowledge to elements of subjective sense perception, and he denied that any general empirical knowledge was possible. Other philosophers such as Locke adopted Hume's empiricism - Berkeley pursued the sceptical argument to a pure form of subjective idealism which even denied the existence of an external reality. The British empiricist tradition was continued in the early 20th century by Russell, who introduced a new form of symbolic logic - the propositional calculus ("Principia Mathematica", 1913) which was to prove vital in the later developments of logical empiricism.

In Europe in the 19th century Mach drew a distinction between theoretical and observational terms in science and insisted that all scientific concepts should be reducible to primary sensations (1886). His paper of 1868 on the concept of mass forms a historical basis for the theory of measurement. Hertz (1894) demonstrated that Newtonian mechanics could be formalised in an axiomatic system with very few axioms and only one fundamental law.

These are the two major developments leading up to the logical positivism of the Vienna circle of the 1920's and 1930's (e.g. Schlick). The positivist programme consisted in showing how knowledge could be built up logically from the basic units of sensory experience - the basic facts. It was an attempt to develop a metaphysics-free epistemology. The Verification Principle asserted that for a statement to be meaningful it is necessary to specify what kind of sensory facts would disclose it as certifiably true. Logical empiricism grew out of the effort of positivists such as Carnap and Hempel to apply positivism to the philosophy of science. Sensory data and the verification principle could not yield scientific theories and so Carnap introduced a weaker "testability criterion" which only required that propositions are capable of verification and a "protocol" language for the description of basic facts - essentially a physical observation language (1936).

A standard logical empiricist approach to the philosophy of science is first to consider facts as neutral observations or measurements of the world, and then to show how such singular facts can lead to general empirical knowledge (usually by probabilistic induction) or the confirmation (or falsification) of scientific theories which have a deductive logical structure. As will be seen in Section 3.3, all these aspects have been challenged in the contemporary philosophy of science.

3.2.2 Hypothetico-deductive systems, explanation and reduction

Logical empiricism regards scientific theories as hypothetico-deductive systems. A hypothetico-deductive system consists of a small set of general empirical propositions together with rules for deducing more specific empirical propositions from them. It is only these specific propositions, at the "bottom of the page", which confront experience and by which the system is confirmed or falsified. In a hypothetico-deductive system, even the highest-level propositions are synthetic. The structure conforms well to that of some real scientific theories which have been highly developed, as Hertz's axiomatisation of Newtonian mechanics illustrates, and was advocated by philosophers such as Carnap (1936), Hempel (1945), Popper (1935), and Braithwaite (1953). However, as Hanson points out (1971), it does not describe the structure of theories in development nor how theories evolve. Furthermore, there are alternative logical and linguistic structures which are equally rigorous but not axiomatic and there is no reason why an axiomatic form is preferable (the theory survives or falls as a whole, and the notion of "synthetic a priori" axioms is unjustified).

A central concern in the philosophy of science has been the way in which scientific theories explain phenomena in the world. In logical empiricism the Hempel-Oppenheim thesis (1948) provided a model for deterministic explanation. An event E is explained by showing how it follows deductively from certain factual conditions (C_i) under a set of scientific laws (L_i) - the deductive-nomological model (D-N) of scientific explanation. Later, Hempel (1962) developed the inductive-statistical model (I-S) of statistical explanation where it is required that the event E should follow from the conditions C_i with a "high degree of probability". The two models, which are also known as "covering-law" models, are summarised below:

D-N model		I-S model	
$L_1, \ldots, L_n$		$L_1, \ldots, L_n$	
	Explanans		
$C_1, \ldots, C_m$		$C_1, \ldots, C_m$	[with high probability]
E	Explanandum	E	

The D-N model equates explanation with prediction, such that questions about explanation are reduced to questions concerning deductive connections between events and initial conditions. However, some theories are used for prediction which do not offer an explanation. Hanson (1971) gives some classic cases, and he also argues that the D-N model is only a necessary condition, not a sufficient condition, for explanation and that the additional aspects of explanation are the familiarity and understanding ("conceptual grip") which a theory brings. The I-S model has received even more criticism since it does not even fit most statistical theories - an event can be "explained statistically" even if its probability as an outcome of some initial conditions is low (e.g. probability of particle emission in radioactive decay).

Another feature of the logical empiricist view of science is the notion of inter-theoretic reduction. This is supposed to occur when a new theory is introduced which explains all the phenomena that the older theories did. In this way the onward cumulative progress of science occurs. Whilst reduction of this form undoubtedly does occur, it is not the whole case. As Suppe (1977) points out, reduction often takes place not between theories but between domains (bodies of scientific knowledge). In extreme cases, favourites of Feyerabend (1958) and Kuhn (1962), the change of theory is often accompanied by a complete meaning change and

the new theory is incommensurable with the older ones.

3.2.3 Induction and confirmation

Induction is concerned with the epistemological problem of what type of knowledge can be obtained by inductive inference. Hume set out to discover under what conditions induction was capable of proving the truth of a generalisation G on the basis of the examination of instances of G. Such a proof is only possible if there is an "induction hypothesis" in the premisses of the argument. This is effectively a proposition about the regularity of the world and is therefore a synthetic contingent generalisation which, in turn, has to be justified by introducing yet another induction hypothesis. This leads to an infinite regress, and Hume concluded that no general empirical knowledge is possible.

The positivists believed that, whilst it was not possible for every instance of a general proposition to be examined, nevertheless such a proposition could have a high probability of being true. This leads to theories of probabilistic induction and involves showing that for a statement S and set of data D, the conditional probability $P[S|D] \approx 1$. However, this requires a probabilistic induction hypothesis and leads to an infinite regress. Reichenbach (1951) developed a "self-corrective" probabilistic induction method based on a frequency interpretation of the probability operator, but this, too, requires a probabilistic induction hypothesis. In an attempt to avoid the induction hypotheses associated with establishing the truth of general synthetic propositions, Carnap (1950) introduced a system of logical probability such that P operates on an analytic statement. The statement is in terms of a set of objects, a set of properties, state descriptions and a set of complex properties. Unfortunately, as Popper demonstrated, the probability of a generalisation in Carnap's system is always zero. Hintikka (1965) modified Carnap's system such that probabilities of generalisations are non-zero; however, this involved a priori partitioning of the world into kinds of individuals, a metaphysical presupposition of the kind positivists wanted to avoid. Such probabilistic induction can be valid against a background domain (of scientific theories and knowledge) which can be regarded as limiting induction to a certain class of systems ("local" as opposed to "global" induction - Suppe, 1977), and provides a philosophical basis for techniques of statistical inference and system identification, for example. Induction plays only a small role in the contemporary philosophy of science.

Confirmation of scientific theories is closely related to the problem of induction in logical empiricism. To confirm a scientific theory requires a set of correspondence rules linking empirical terms to observational terms, and a set of co-ordinating definitions linking higher order and lower order theoretical propositions. The positivist epistemology requires that the observational terms are theory-neutral facts, and stems from the very old "correspondence theory of truth". The theory cannot be confirmed in every possible instance (since there is an infinity of cases) but receives more confirmation each time it agrees with a fact. Some attempts were made by logical empiricists to formalise this probabilistically but the problems of an induction hypothesis arise. Popper (1935) and Carnap (1936) used measures of confirmation (or "corroboration") that were non-probabilistic.

Popper's doctrine of falsification (1935) proved to be a radical one in the philosophy of science and was based on the trivial observation of the logical asymmetry between verifiability and falsifiability; that whilst it is not possible to confirm a theory fully, a single fact that disagrees with the theory will falsify it. Thus Popper suggested that scientists seek to disprove their theories in a critical fashion rather than to confirm them conservatively. This perhaps marked Popper's divergence from logical empiricism and can be regarded as a basis of Lakatos' much later work on criticism, rationality and the growth of knowledge (1970). This doctrine also allowed Popper to formulate his "Criteria of demarcation": "it must be possible for an empirical scientific system to be refuted by experience" (p. 41, 1972, orig. 1935). It demonstrates Popper's closeness to the positivistic aim of delineating meaningful statements, but, unlike the verification principle, does not eliminate the whole of natural science.

Logical empiricism maintains a clear distinction between theoretical terms and observation terms expressed in a neutral language as facts about the world. Scientific knowledge about the world comes from the correspondence between theoretical propositions and observation statements (positivist epistemology). However, in the 1950's the idea of a neutral observation language was challenged. Hanson (1958) showed that facts are essentially "theory-laden" and that they change depending on the theory which is held. This undermined the objective epistemology of logical empiricism. In logical empiricism, the way in which new theories are created or modified is placed outside the context of philosophical enquiry and regarded as a matter of psychological or sociological fact. Recent work in the philosophy of science shows that

it is the good reasoning patterns in the evolution and selection of theories which probably constitute the rationality of science, and emphasise the dynamic processes involved, unlike logical empiricism which offers a series of static pictures. Logical empiricism forces science into its formal straightjacket and is not based on historical studies. Work by philosophers such as Kuhn (1962), Feyerabend (1975) and Shapere (1977) is much more concerned with drawing out general principles of the scientific process by very detailed studies of past and contemporary science. These are discussed in the next section.

3.3 The Dynamic View of Science - Contemporary Philosophy of Science

3.3.1 Introduction - the challenge to logical empiricism and new directions

In the last section, the challenges to the basic principles of logical empiricism were introduced: the denial of the theoretical-observational distinction; criticism of the emphasis on the formal structure of theories and on induction and confirmation; inadequate treatment of scientific explanation and inter-theoretic reduction; and the lack of consideration of the evolution of scientific knowledge. Since the 1950's there have been dramatic changes in the character of the philosophy of science. Firstly, there was the strong reaction against logical empiricism in the historical approach of Kuhn, Feyerabend and others of the so-called "Wentanschauungen" school which focussed on the dynamics of the overall process of science rather than the individual theory (Section 3.3.2). Kuhn's famous concept of a "paradigm" was a great stimulus to many scientists as well as philosophers. However, the extreme position of the Weltanschauungen analyses entailed a subjective idealist epistemology which meant, in effect, that science did not have to say anything about the world. Consequently, these analyses have received a great deal of criticism. Constructively, emphasis since the late 1960's has been on the study of the rationality of scientific development by examining actual scientific practice and attempting to extract the general reasoning principles ("good" reasoning patterns) which scientists use. This is typified in the work of Lakatos, Toulmin and Shapere (Section 3.3.4). Suppe (1977) described this movement as one of "historical realism", and it is characterised by a belief that science does yield knowledge about the world and the principle that the objectivity of science comes from the rationality (critical reasoning) of science.

Although the confirmation or validation of theories has only a small role in these newer philosophies of science they offer an understanding of the overall scientific process ("theory of theories" - Harré, 1972) which is a crucial input into the theory of model validity (Chapter 4) and which is absent from logical empiricism despite its emphasis on confirmation theory.

3.3.2 Weltanschauungen analyses

In his famous "Patterns of Discovery", Hanson (1958) showed how seemingly neutral observations of fact do contain theoretical aspects - "theory-laden" observations. This attacked the old Machian distinction of theory and observation central to positivism. Theories no longer correspond to the facts, but to the facts as seen by the theories, and this weakens considerably the objectivity of science. Hanson's view on explanation was that of a D-N covering-law model (see Section 3.2.2) which additionally conceptually organises phenomena so as to render them comprehensible; however, he did not elaborate on how this function is fulfilled.

The early work of Feyerabend (1958) was concerned with establishing a new kind of observation language which was capable of dealing with the influence of theory on observations. Reports of observations agreed on by scientists through common consent are "uninterpreted" sentences. Their interpretation depends on theories, and it is only "interpreted"sentences which can be true or false in a correspondence sense with the theory. Thus there is no connection between truth and the acceptance of interpreted observation sentences. Feyerabend's doctrine on meaning was that no terms common to different global theories have the same meaning and that all propositions in a theory are analytic. Therefore, different global theories are incommensurable and theory reduction is not possible. Another feature of Feyerabend's philosophy was the insistence on theory proliferation in scientific practice as opposed to holding on resolutely to the same old theory. However, this requires a technique for the comparative evaluation of incommensurable theories which presupposes a common language, and this is incompatible with Feyerabend's doctrine of meaning. Some implications of Feyerabend's early work are that the rejection of a global theory becomes an irrational process, knowledge yields to belief, and there are no reasons for the proliferation of theories (since under Feyerabend's epistemology there is no guarantee of converging to the truth).

In "Against Method" Feyerabend (1970) attempted to correct these difficulties. The basic method was to find cases in the history of science where

people held theories in the face of "disconfirming observational tests". Feyerabend concluded that there are no firm methodological rules and that, in science, "anything goes" and put forward a principle of "counter-induction" (that there will always be a clash between the theory and the facts). This contradiction drives science forward, thus exemplifying Hegel's dialectical method. However, Feyerabend did not give any means for resolving contradictions, and the new views are compatible with his earlier epistemological views and subject to the same criticisms.

Kuhn's "Structure of Scientific Revolutions" (1962) was a formative book in the modern philosophy of science. Kuhn conceived of science as periods of relative stability - "normal" science - interspersed with "revolutions", periods of dramatic change. During periods of normal science, scientists share a "paradigm", a rather nebulous concept referring to the background knowledge, world-views, techniques, theories, models, etc. and the paradigmatic (standard) ways of applying theories to experience. As normal science progresses, there is a build up of "anomalies" (data the theory cannot explain) and unresolved problems which eventually lead to a crisis in which the paradigm is discarded and there is a temporary flux of many new and competing theories ("revolution"). Soon, however, a new paradigm is instated and science continues in this cyclic process. This was illustrated with many cases from the history of science. At this stage, Kuhn's views on epistemology and meaning were identical to Feyerabend's and he was subject to the same criticisms, in particular the charge that science becomes an irrational process and a matter of "mob psychology".

These criticisms were taken very seriously by Kuhn (unlike Feyerabend) and he has recently changed his position considerably (1970 and 1977). The old concept of a paradigm is replaced by the new notion of an "exemplar" together with the more structured notion of a "disciplinary matrix". An exemplar is a learned resemblance relation which determines how symbolic generalisations are applied to nature. The disciplinary matrix is a structured notion referring to the set of knowledge, putative facts, and theories in a particular discipline. With the use of exemplars Kuhn maintains that the interpretation of observation sentences does depend on experience but denies that any correspondence notion of truth plays a role in acquiring knowledge from observation. Thus he is still subject to the criticism that this leads to scientific knowledge as mere collective opinion.

During revolutions between periods of normal science there is a need

to evaluate competing theories. Kuhn weakens his doctrine of meaning such that some terms in different theories are common but not all of them. This can lead to very difficult problems distinguishing theories (and meanings), solutions of which repudiate many of his own views. From historical studies, Shapere (1977) has gathered evidence that there are rational considerations used to evaluate tentative theories as fruitful for further research (see Section 3.3.4). These depend on an agreed-upon body of information - which Shapere calls a "domain" - outside of Kuhn's normal science and disciplinary matrices. There is scepticism over Kuhn's claim that science oscillates between normal and revolutionary science and that these are "rational reconstructions" on the complex processes involved in the evolution of scientific knowledge.

Although the influence on the Weltanschauungen analyses is declining under the weight of these criticisms, they have had a profound impact on the philosophy of science: in the downfall of logical empiricism, the shift to a dynamic view of science with an emphasis on historical analysis. But, in Section 3.3.4, it will be seen that the focus of the new ideas is quite different: the evaluation of reasoning patterns used in selecting hypotheses or theories worthy of further development, or in deciding on crucial problem areas; and the detailed analysis of conceptual devices actually employed in science.

3.3.3 Structural dynamics of theories

A structural, as opposed to sentential or statement, conception of theories was proposed by Sneed (1971) in a case study of particle mechanics. This is based on a model theoretic formulation which stems from Tarski's work on model theory (1954), and is essentially a semantic conception of theories. Sneed's approach was generalised and refined by Stegmüller (1976) and is known as the "Sneed-Stegmüller synthesis" (Mattessich, 1978). This formalisation has many attractive features - for instance, it can cope with the dynamics of theory change, as well as being a concise symbolism. In the Sneed-Stegmüller synthesis, a theory T consists of a fairly permanent core K, and a set of intended applications I:

$$T = \langle K, I \rangle \qquad \ldots\ldots\ldots (3.1)$$

where K consists of a set of models M, a set of possible models M_p, a set of partial possible models M_{pp}, a set of constraints C and a set of variables v. I is defined by $I\varepsilon/\hat{A}(K)$, where /A is the "application operation":

$$T = < M, M_p, M_{pp}, C, v, I > \quad \ldots\ldots\ (3.2)$$

As the theory is developed the core is extended to include additional laws L, additional constraints C_a, and additional variables v_a:

$$T_e = < M, M_p, M_{pp}, C, v, L, C_a, v_a, I > \quad \ldots\ldots\ (3.3)$$

The notion of a "hard core" of a theory was introduced by Lakatos (1970) and the series of theories T, T_{e1}, T_{e2}, ... is equivalent to a "research programme" in Lakatos' sense (Lakatos' philosophy is considered in the next section). The gradual extension of the theory is very much like the processes occurring during a period of Kuhn's "normal science" and, indeed, Kuhn has supported this version (1975). During a scientific revolution the entire core would be replaced.

The set I of intended applications is not a strict definition but an open set. Typically, it might consist of a set I_o of paradigmatic applications (somewhat like Kuhn's exemplars) and a growing set of applications which the scientist discovers. If this set is growing, i.e. if $I_t \subset I_{t+1}$, then the theory is progressing (much like Lakatos' progressive research programmes whose empirical content is increased). Another notion of progress is that of core refinements. Stegmüller uses the model to cope with Kuhn's ideas of revolutionary science, irrefutability of a theory, and theory dislodgement (without falsification).

Whilst the Sneed-Stegmüller approach offers many possibilities for future development, at present there are some substantial criticisms which can be raised against it. Firstly, there is no account of the way in which competing extensions of the core are evaluated or the conditions under which a core would have to be replaced. Using Shapere's approach (see next section) this could be achieved with certain reasoning-patterns based on a body of accepted knowledge (domain), but this is outside the Sneed-Stegmüller formalisation of a theory (the same criticism as applied to Kuhn's disciplinary matrix in Section 3.3.2). If this is omitted, then there is no guarantee that proliferation of core extensions would ever converge on knowledge and representation of phenomena. Secondly, it is not clear how well different scientific theories will map into this structure, and its application to a variety of case studies is essential in order to validate it. Thirdly, it can be questioned whether scientists actually proceed by looking for new applications for a theory. No doubt this does occur at some times when it is hoped that a theory will have a wide generality, but usually the intended application is well-known

beforehand. Perhaps what is meant here is the widening of the empirical support for the theory over the intended application, a point clearly made by Lakatos (see Section 3.3.4).

With this new formulation, the philosophy of science really becomes a "theory of theories"; some writers have even developed the idea of a theory of science as a general automata theory (Zeigler, 1976; Suppe, 1977). Suppe (1977) suggests that a semantic conception of scientific theories (such as the Sneed-Stegmüller model), together with erotetic logic ("question and answer" logic), may answer the tricky problems about scientific explanation.

3.3.4 Realism and rationality in the philosophy of science

There are, according to Shapere (1974), "patterns of reasoning in the construction or discovery (as well as the ultimate acceptance or rejection) of scientific hypotheses and theories, and that a great deal of illumination of the scientific enterprise can be attained by examining them". The investigations of the current philosophy of science pay close attention to actual scientific practice with the aim of developing a systematic philosophical understanding of the justification of knowledge claims. In this work there is "a strong commitment to both a methodological realism and an epistemological realism" (Suppe, 1977) and this virtually precludes the sociological view of knowledge (e.g. of Kuhn and Feyerabend). The outline below is based on Suppe (1977) and Shapere (1977) but with additional material on Popper and Harré.

Lakatos (1970) conceives of science as a sequence of ever-improving theories T_1, T_2, T_3, ..., which form a "research programme", accompanied by a series of "problem-shifts". If theory T_{i+1} has excess empirical content over T_i, then the problem shift is "theoretically progressive" and if this is corroborated empirically it is empirically progressive (n.b. the criticism of the Sneed-Stegmüller approach in Section 3.3.3), otherwise the problem-shift is "degenerating". The "heuristic power" of a research programme is divided into a positive heuristic which suggests which paths to follow for a progressive problem shift, and a negative heuristic which suggests what paths of research to avoid. Lakatos assumes that the negative heuristic is a "hard core" of the research programme which cannot be modified, but does not go into the considerations for deciding on this. Once a research programme is degenerating it is irrational to proceed with it further. Lakatos does not develop what constitutes a good positive heuristic or procedures for the comparative evaluation of

positive heuristics; in short, he does not give a full account of the rational processes involved. Suppe (1977) offers a very detailed criticism of Lakatos and his key points are summarised below:

1. In a shift from T_i to T_{i+1} there may be a partial semantical re-interpretation of terms in T_i and, since the hard core must remain, this severely limits the reasonable means for modifying theories.

2. Too much emphasis on theory development by testing experimentally and modifying in response (underemphasises theoretical development).

3. Misses out development of new concepts for dealing with a class of phenomena.

4. Lakatos totally ignores "the extent to which what is reasonable is conditioned by the subject matter of the science".

(Lakatos' approach is based on Popper's noninductive methodology as in "Conjectures and Refutations", 1962,where he regards the growth of theories in the following tetradic schema:

$$P_1 \rightarrow TT \rightarrow EE \rightarrow P_2$$

where P_1 is the first problem, TT is a tentative theory, EE is a process of error elimination (as in testing the theory), and P_2 is the redefined problem. Popper concentrates mainly on the EE stage and has little to say about how TT is arrived at, or the shift from stage to stage (i.e. the dynamics).)

An interesting realist approach to theory development is given by Harré (1970) in his theory of models. He regards theories as essentially concerned with the mechanisms of nature, and only derivatively with the patterns of phenomena; a theory as a "statement-picture complex". The chief means of picturing mechanisms in nature is by the use of real or imaginary models and Harré argues that much of the theoretical activity of scientists is spent in this pursuit. He distinguishes between the "subject" of a model - what is modelled - and the "source" of a model - what it is modelled on. This leads to a general categorisation of models: "homomorphs", for which the subject is also the source, and "paramorphs", in which subject and source are different (Harré develops this further into a taxonomy of models). Paramorphs are used in theory development to postulate a hypothetical mechanism for a subject; the mechanism may be from another subject (homomorphic model) or may be an imaginative creation. "Thus, at the heart of a theory are various modelling relations

which are types of analogy" (Harré, 1970). The view of a theory as a statement-picture complex leads to four kinds of hypothesis which are logically and epistemologically distinct:

1. Existential hypotheses - these generate experimental or theoretical (e.g. categorical) research.
2. Descriptions of the model or hypothetical mechanism - empirical pursuit of answers to corresponding questions cannot be undertaken until questions as to the existence of the entities are settled.
3. Causal hypotheses - the power of a hypothetical mechanism to produce the phenomena is queried (conditional statements).
4. Modal transforms - equivalence of different modes of descriptions raises complex issues in epistemology (bi-conditional statements).

The traditional "statement" conception of theories does not yield this variety of hypotheses, and this is a very important aspect of Harré's epistemology. He also considers the relationship between theory and an observation language, but this will not be considered here. Some general criticisms of Harré are not of his approach, but rather what he omits in his account of scientific development; for instance: the way in which competing paramorphs are evaluated as candidates for hypothetical mechanism, the further development of the theory and refinement of the paramorphic model (ultimately homomorphic), or the influence of background knowledge on the reasoning processes. (Harré's conceptual framework is used in Section 3.4 in the outline of some problems in the philosophy of systems science).

Toulmin's philosophy of science is a very rich and metaphorical one. It is based on very detailed studies in the history of science and an evolutionary model of science derived from Darwin's theory of evolution (Toulmin, 1972, "Human Understanding, Vol. I"). He regards the function of science as to build up systems of explanatory and representational techniques with which to reason about phenomena. Theories are introduced in one fell swoop and the incorporation of concepts in prior use requires a "language shift". Furthermore, theories (or models, etc.) are not true or false, but to be judged according to whether they are fruitful in the applications a scientist intends.

The "gene pool" of Toulmin's model is the set of fundamental aims of

science, and the "species" are the separate scientific disciplines. New theories (or other devices) are produced by a process of reproduction and mutation, and the survival of successful theories depends upon reasons (rational considerations) and causes (sociological and other factors). In Toulmin's view it is the progression

$$\text{Conceptual variants (explanatory ideals)} \rightarrow \text{reasoned comparison} \rightarrow \text{selection}$$

which provides science with its objectivity. This is not a correspondence theory of truth (hypothesis $\leftrightarrow$ facts) but rather criticism in the light of experience. The four major criticisms that Suppe (1977) raises against Toulmin are:

1. The model is only a satisfactory model of conceptual change in "compact" (as opposed to diffuse) disciplines, and there is little on what good reasoning actually is.

2. There is little reference to the fitness of theories (but perhaps could be convergence to truth, or fruitful ways of representing phenomena).

3. It is not clear how a subject proceeds rationally or converges on representational techniques that both yield knowledge and are explanatory.

4. To be epistemically reliable and successful (an efficient, robust, adaptive system) the production of conceptual variants will have to be conditioned by the present state of the discipline (analogously to the recent developments in the theory of gene recombinations).

Underlying Shapere's work are three themes (e.g. 1977) expressed as postulates: "I. Scientific development and innovation are often appropriately describable as rational ... II. The rationality involved in specific cases is often generalizable as principles applicable in many other cases ... III. These principles can in some sense be systematized. Shapere indicates these three postulates in case studies over a range of stages of scientific development (his approach is based upon a detailed examination of actual scientific practice). Central to Shapere's philosophy is the concept of a "scientific domain". This consists of items of information (including possible facts, theories, etc.) associated together as a body. By its very nature the domain generates problems: "domain problems" are concerned with a clarification of the domain itself;

"theoretical problems" are those whose solution requires a deeper account of the domain in terms of theories. The domain favours certain types of solution to these problems. Shapere (1977) poses six major questions concerning domains:

1. What considerations lead scientists to regard a certain body of information as a unified subject (as a domain)?
2. How is the description of items in the domain achieved and modified in sophisticated stages of development?
3. What sorts of inadequacies are found in domains and what grounds are there for considering some inadequacies as problems for further research?
4. What considerations lead to the generation of specific lines of research, and what are the reasons for considering some lines of research to be more promising than others?
5. What are the reasons for expecting solutions of certain sorts to be sought for these problems?
6. What are the reasons for accepting a certain solution of a scientific problem regarding a domain as adequate? (p. 523, 1977)

So far, Shapere has mainly considered the following aspects: the grouping of observational facts into domains, rationales for introducing new hypotheses which radically contradict established theory; the role of background information; the maintenance of objectivity (with theory-laden facts); and the role of conceptual devices in science. At a primitive stage in the development of a domain nonproblematic observations and facts are used to establish the appropriateness of theories and their observational interpretations. As the domain progresses to deeper levels of description more theory is involved in observation and this leads to a re-interpretation of the domain (e.g. distinction between direct and non-direct observation). It is this link back to the early stages of the domain through the good reasoning patterns employed which guarantees the objectivity of science. In studying the history of science, Shapere has discovered that there are very many good-reasoning patterns, and that these are affected by the content of the domain in a feedback fashion (c.f. the stability of Toulmin's evolutionary model). Background information (outside the domain) plays a crucial role in the interpretation of observations and objectivity of scientific knowledge; however, there are constraints on how this should be used (for instance, the background

information should be from some field that coheres with successful concepts and theories of other domains). Shapere's analysis of science is very detailed (the account here is very summarised) and offers a great potential for future development. However, Suppe (1977) points to the areas which require most development - Shapere's views on facts, knowledge, and the justification of knowledge claims underlying his entire approach to objectivity and rationality in science.

3.4 Systems Philosophy and the Philosophy of Systems Science

3.4.1 Introduction

This section deals with some aspects of philosophies associated with systems science (for simplicity, "systems science" here refers generally to all aspects of systems - inspired science, both theoretical and applied, such as: systems engineering; systems research; systems methodology; systems approaches to - biology, - medicine, - ecology, - etc.; systems oriented operations research, and so on, and even the abstract towers of general systems theory). A clear distinction must be made between systems philosophies as views of the world, man and society (e.g. Bertalanffy's holistic philosophy, Section 3.4.2; Laszlo's systems philosophy, Section 3.4.3) and philosophical analysis of the various practices of systems science (e.g. Checkland's or Mattessich's methodologies, Section 3.4.4). The review is not intended to be comprehensive but to pick out some of the key points with philosophical significance. Consequently, many of the concepts and analytical devices presented in earlier sections of this chapter will be widely used. Four good sourcebooks for classic references on systems, yet with very different approaches and opinions are von Bertalanffy (1968), Emery (1969), Berlinski (1976), and Mattessich (1978). This critical appraisal fits in with the thesis in the following way: the thesis is a general methodological study of the problem of model validation in science, but with particular reference to models in biology and medicine that are systems oriented. These models are rich in theoretical systems concepts (e.g. feedback, dynamic equilibrium, self-stabilisation, etc.) and are not simply models based on biological data and concepts. Intuitively this suggests that validation may be problematic, since there are at least two distinct levels on which it can be approached - the first is the representational validity of the model in a specific case, and the second relates to the appropriateness

of the systems concepts to biological systems in general. The latter is typical of problems that occur in philosophical analyses of systems science, and which this section attempts to clarify in such a way that the theory of model validity (developed in the next chapter) will be capable of providing techniques for their solution.

3.4.2 The two sides of Bertalanffy

The widespread development and application of systems-oriented sciences and techniques has occurred mainly since the 1960's, although a few scientists were pursuing original work along these lines much earlier (from the 1920's to the 1950's). One of these was the biologist Ludwig von Bertalanffy who, in the 1920's and 1930's, attempted to develop a quantitative theory of metabolism and growth. In doing so he proposed that a biological entity should be conceived organismically as an open system and that new laws, not based on existing laws of physics and chemistry, would have to be found for such systems (1932). An open system is one which interacts freely with its environment (in physical systems this involves exchanges of matter and energy) and Bertalanffy suggested that this provided a general model for systems of any type (physical, social, cognitive, etc.). Furthermore, the logical and mathematical properties of such an abstracted system could be determined and this would form the basis of a "General System Theory". In this aspect of his work, Bertalanffy embraced a complete philosophical view of science and the world (a "systems philosophy") that was dependent upon, but not necessary to, his empirically-based work on theoretical biology. This view was that of a holistic conception which covered and unified the whole of science.

The two sides of Bertalanffy's work - theoretical biology on the one hand, and general system theory (GST) on the other - are reflected in his writings (contrast 1950 or 1968 with 1964) and ultimately lead to many contradictions (some are discussed below). This dualism has not been properly understood by several critics of systems science (e.g. Berlinski, 1976, or Lilienfeld, 1978) and consequently Bertalanffy has received substantial criticisms, notably of his general system theory (which is undoubtedly weak) and the many ambiguities present in his work. These criticisms are fuelled further by Bertalanffy's insistence that he is the founding father of general system theory and the open system concept (both constant themes) and that there is a fundamental link between GST and his work on theoretical biology. These two aspects will be

considered separately and in more detail below (his views on theoretical biology are taken from 1964, and on GST and systems philosophy from various chapters of 1968 - the introduction, and Chapter 2, "The Meaning of General System Theory", originally written in 1956).

At the turn of the century, biologists were still having problems with vitalisms - it was not possible to explain the dynamic and energetic properties of living systems using the current laws of mechanics and thermodynamics and it was tempting to introduce some kind of additional (unobservable, non-physical) force or vital spirit which, scientifically, was very unsatisfactory. By considering a biological organism as an open system across whose boundaries energy and matter may pass, Bertalanffy showed that the closed system laws of thermodynamics no longer applied and that, rather than regressing to a state of unordered equidistribution of energy, such a system would possess characteristics very similar to living systems (1932). The two most important characteristics were that of dynamic equilibrium (as opposed to a minimum energy steady state), for which Bertalanffy coined the term "Fleissgleichgewicht", and that of equifinality whereby two similar systems attain the same end states despite different initial conditions (as in morphogenesis). As well as allowing a quantitative treatment of biological phenomena such as metabolism and growth, this provided a basis for posing teleological questions in a non-vitalistic framework. His claim of originality in this matter is not completely founded since other biologists were tackling the problem at the same time (e.g. Cannon's concept of homeostasis (1929) has proved to be exceedingly fruitful, or Woodger's organismic biology (1930)).

Bertalanffy's 1964 paper is essentially a review in which he considers a variety of types of model used in quantitative metabolism (most of which he has developed). These are open systems, feedback and homeostasis, allometry and the surface rule, and theory of animal growth. The models are fairly simple and usually well-backed up by empirical data. In this paper, he elucidates his philosophy of modelling in biology which is very straightforward and empirically-based and in contrast to his GST. Although there is a close relation of facts to theory and a new model may reinterpret old facts, Bertalanffy nevertheless insists that "the decision whether or not a model is suitable exclusively rests with the facts of observation and experiment" (1964). In assessing some objections that can be raised against modelling, he argues that gross

simplification is necessary given the complexity of biological systems and that ultimately this does lead to better explanations. However, he requires that models should contain few parameters and that all these should be checked experimentally. This latter requirement is one that many physiological modellers or neuro-cyberneticians could not possibly accept, and yet they would regard themselves as far more empirically-oriented than Bertalanffy's GST. It can be seen, therefore, that this side of Bertalanffy cannot be simply dismissed à la Berlinski and that he has made a very significant contribution to the development of theoretical biology.

The same is not true, however, about Bertalanffy's General System Theory and holistic philosophy. GST is based on a holistic philosophy that the world can be understood only by considering it as a series of "wholes" and that these wholes - systems - will exhibit certain standard types of behaviour. Furthermore, this understanding will not be available from a reductionist approach to science. Bertalanffy introduced the concept of GST in his early work on the theory of open systems in biology and gradually developed it into a logical and mathematical formulation (1950). He limited himself in this regard to systems describable by differential equations with which he is able to express precisely the concepts of dynamic equilibrium, equifinality, and even system. The latter seemingly emerges from the consideration that the off-diagonal elements of the state-transition matrix are non-zero. This theory was simply a theory of differential equations and not of systems in the real world. Other workers, such as Mesarović (1964), extended the definition of system and GST to highly abstract set-theoretic terms which have even less to do with reality. In the preface of his 1971 (orig. 1968) book he includes "dynamical system theory, cybernetics, automata theory, systems analysis by set, net, graph theory and others" under the protective wing of GST. To some extent, Bertalanffy did base his system properties on studies of particular scientific fields, but largely on the process of abstraction. This was carried to the extreme by Ashby (1958) who set out to reveal these properties by starting with the "set of all conceivable systems".

In 1954, Bertalanffy set up the Society for General Systems Research (SGSR) with Boulding, Rapoport and others. The aims of this society reveal the underlying holistic systems philosophy:

(1) To investigate the isomorphy of concepts, laws and models in various fields.

(2) To encourage the development of theoretical models in areas which lack them.

(3) To minimise duplication of theoretical work in different disciplines.

(4) To promote the unity of science.

The aims to reduce duplication of work and to promote the unity of science are very fine, but the fundamental flaw is that these will be achieved as a consequence of discovering the general isomorphy of scientific fields. The expected isomorphisms have not been found (except insofar as between the level of theory associated with the use of control theory, cybernetics or other mathematical techniques in different fields) and the subsummation of science into systems science has not occurred. In fact, systems science augments rather than replaces more conventional science.

The way in which Bertalanffy envisages that models are transferred between different fields emerges from his analysis of the levels of description in science (1945, contained in 1968, pp. 84-85). Firstly, there are analogies which are based on the superficial similarity of phenomena in different fields; secondly, there are homologies which involve a transfer of models from one field to another when the "respective laws are formally identical"; and, thirdly, the level of explanation which consists of the specific conditions and laws of a particular field. It is logical homologies which give Bertalanffy the key point in his argument. The formal identity of laws does not relate to traditional laws (if so there would be no need for the homology in the first place) and the argument goes as follows: "If an object is a system, it must have general system characteristics, irrespective of what the system is otherwise". Therefore, having developed a model of one system in one field, this model has complete generality to any other field. But this argument overlooks the central concern of science, i.e. how do we know that the object of our attention is system-A with system-A type properties? Furthermore, this account makes the definition of a system critical for scientific progress and yet the definitions of systems are notoriously ambiguous or general ("A system can be defined as a set of elements standing in interrelations", Bertalanffy, 1968, p. 55).

Other criticisms of GST (in-house criticisms by systems workers) include:

(1) Limit to generalisation and analogy possible in science (Simon, 1969)

(2) GST yields no hypotheses for empirical testing; is dogmatic; contains no self-criticism or pertinent testing/evaluation procedures (Mattessich, 1978).

Thus the systems philosophy of GST leads to an imposition of a view of the world on the world itself, and reduces the process of science to one of simply finding systems in the world which have certain defined properties and not the struggle to find out what the world actually is or how it is so. Bertalanffy uses Kuhn's "paradigm" concept (1962, see Section 3.3.3) as strong evidential support of GST's "new philosophy of nature" (1968, p. xix), yet if he had pursued Kuhn's theory to its logical conclusions he would have found that revolutions occur which overthrow the old paradigm and instate a new paradigm.

3.4.3 Laszlo's systems philosophy

Laszlo's "Introduction to Systems Philosophy. Toward a new paradigm of contemporary thought" (1972) is an example of a systems view of the world and is completely faithful to Bertalanffy's GST. The book purports to offer a new approach to philosophical inquiry based upon systems theory and a "synthetic" approach, yet it does not deal seriously with questions that usually figure prominently in such an inquiry. For instance, the questions of epistemology (the basis of the evolution and justification of knowledge) do not receive consideration until Chapter 11. The structure of the book is as follows:

Firstly, Laszlo puts forward some "primary presuppositions":

"1. The world exists;
and
2. The world is, at least in some respects, intelligibly ordered (open to rational inquiry)." (p. 8)

and some "secondary presuppositions":

"(i) The world is intelligibly ordered in special domains;
or
(ii) The world is intelligibly ordered as a whole."

The next step is to argue that these wholes or special domains are systems whose properties are given by systems theory (or GST). On this basis, Laszlo develops a symbolic theory, "Outline of a general theory of systems" which forms a major chunk of the book. This theory is composed essentially of "second-order models" or "models of models" and thus can be rather general although simple. It has the general form

> "THEORY: $R = f(\alpha, \beta, \gamma, \delta)$ where $\alpha, \beta, \gamma, \delta$ are independent variables having the joint function R ("natural system")" (p. 35)

The variables $\alpha, \beta, \gamma, \delta$ are "but a handful of systems properties" (p. 35) yet nevertheless capable of describing the range of phenomena in the "terrestrial microhierarchy". They are defined thus:

α: property of ordered wholeness (or the "systemic state property")

β: self stabilisation (or "system-cybernetics I")

γ: self adaptation and organisation (or "system-cybernetics II")

δ: dual-functional-structural adaptation (or "holon property")

Using this formulation Laszlo demonstrates the theory for physical, biological, social,and cognitive systems and the mind. In each case the form of the theory is the same as above except that "natural" is replaced by "biological", etc. Finally, the book concludes with a systems view of the philosophical problems of ontology, the mind, epistemology, and ethics. (For instance, epistemology is considered under cognition and based upon a second-order model of a cognitive system.)

It is difficult to know where to begin criticising Laszlo since his style is seemingly erudite yet rather vague, and he often uses terms in a different way from their conventional usage. Therefore, criticism will be directed at two specific issues: the metaphysical assumptions his approach requires, and his views on scientific theories and their validation (in Chapter 11). His primary presuppositions are commitments to a metaphysical realism and rationality. Although he does not adduce evidence that these are fundamental to scientific activity, this is a reasonable philosophical position to take (c.f. Shapere, 1977, see also Section 3.3.4). In other words, the assumption is that there is an object of scientific inquiry and that knowledge of this object is possible. However, this objectivity and rationality are not inherent in the "perspectivist" view of knowledge and science that he presents in Chapter 2. The secondary presuppositions, (i) and (ii), concerning the division and ordering of the

world are completely untenable. Laszlo's associates (i) with the development of special theories and (ii) with general theories. Acceptance of (i) requires the existence of strict boundaries to the application of theories as a metaphysical presupposition. Scientists pragmatically expect their theories to be applicable only in the specific field in which they are working, but do not eliminate the possibility of extending its application as a metaphysical necessity (this would eliminate any analogical exchange of models). It is obvious that Laszlo favours the second secondary presupposition in evidence of the validity of a general theory of systems, and reasons "that the wide empirical applicability of systems concepts argues for the justification of assuming general order as a cogent hypothesis" (p. 11). However, the validity of a general theory is determined only by empirical investigation and not metaphysical necessity. Both of these secondary presuppositions are completely unnecessary and unacceptable as a basis for scientific activity and, since the second one (ii) forms a key step in Laszlo's system philosophy, there is a fundamental flaw in its development.

Laszlo's views on scientific theories and their validation are contained in a section on "Scientific Cognition" in Chapter 11 on "Cognition: Framework for an Epistemology". This is rather strange since he is proposing a scientific approach to philosophy, and a consideration of scientific epistemology should form a natural introduction. He leads up to scientific cognition (and epistemology) by considering cognitive systems in terms of a systems or "gestalt" psychology based on his systems theories of biology and the mind (Chapters 5 and 7). Science forms a "multi-personal natural-cognitive system" in which each scientist can share the same conceptual constructs and experience the same phenomena. Observation of the world depends not only on the state of the world but on the state of the observer in an inter-active sense giving "theory-laden" observations (Hanson, 1958, see Section 3.3.2). At this stage, Laszlo switches from a purely psychological analysis of science to a philosophical analysis, which is an odd mish-mash of positivist and Weltanschauungen views.

Although, in Laszlo's view, theories affect the observation, he states that "confirmation is had when the observation bears out the prediction flowing out of the construct system" (p. 210). (Note that it is not clear whether or not the "construct system" is the theory.) This leads to the standard positivist account of the way in which indirect constructs are tested by using correspondence rules, in which the construct system is

structured in a hypothetico-deductive way. (The criticisms of this view are contained in Sections 3.2.2, 3.2.3, and 3.3.1). He does not analyse epistemological implications of a correspondence theory of truth in which the theory corresponds not with the facts but the theory-seen facts - notoriously difficult yet important ground in the philosophy of science.

Next, Laszlo analyses science as a "systemic control process" and for this purpose uses Kuhn's socio-historical paradigm concept (1962). Kuhn's paradigmatic conception of science was reviewed in Section 3.3.2, but briefly it consists of periods of normal science (under the guidance of a paradigm) interspersed with periods of crisis in which there is a proliferation of new theories and eventual instatement of a new paradigm. Laszlo follows Kuhn's analysis to the letter without remarking that it is based on a historical analysis of the reductionist sciences (the deficiencies of which Laszlo attacks at the beginning of the book), or that it rests on doctrines of meaning, theory incommensurability and confirmation that are totally incompatible with his own views in the previous section. However, Laszlo proposes a cybernetic model of Kuhn's theory that is significant.

Normal science has "system-cybernetics I", that is it is self-stabilising through negative feedback eliminating the error between theory and data, whereas crisis or revolutionary science has "system-cybernetics II" which is self-organising through positive feedback. There are times when science does show such characteristics, and their analysis using these models may prove illuminating (particularly if merged with the Sneed-Stegmüller model-theoretic formulation, Section 3.3.3). The criticisms of Kuhn's theory were presented in detail in Section 3.3.2 and apply equally to Laszlo; for instance, the way in which competing theories are assessed during a revolution (Laszlo - one theory is "confirmed and accepted", p. 218) - implies a neutral observation language and evaluation procedure which contradicts his views on the theory permeation of observations.

Although there are many more criticisms which can be made of Laszlo's work, the inadequacy of his epistemology in satisfying his own presuppositions for empirical inquiry (realist rationalism) undermines his entire approach.

3.4.4 Systems methodologies

Systems methodologies emerged out of the attempt to apply the concepts of general system theory (GST) together with the computer-based techniques

of systems analysis to scientific, technological, and business problems, and consist of sets of methods and rules which, ideally, allow such problems to be well-posed and solved. The main emphasis has been on methodologies applicable to "designed" as opposed to natural systems. A formative work in this area was Hall's "engineering system methodology" (1962) for the systematic design process. Checkland (1972) calls problems related to such systems (technological and business) "real-world problems" which seems to entail a severe attenuation of reality; however, his "Soft Systems Methodology" has received a great deal of attention and has been used widely. This will be briefly outlined below, together with some of the major criticisms. Systems methodologies are almost entirely instrumental, that is they are to be used to achieve some practical end, such as the design or improvement of a system. Consequently, theories capable of dealing with normativistic concepts (such as M'Pherson's multi-objective decision theory, 1979) are associated with such methodologies.

Checkland's "soft-systems" methodology is based on his experience in "action-research" (practical industrial- or business-based research) at Lancaster University and received full articulation in his 1972 paper. It is intended to provide a methodology for solving real-world problems, and consists of nine stages (1972, p. 98):

1. Problem situation
2. Analysis (of what exists at present in the problem situation; a "rich picture")
3. Root definition of relevant systems
4. Conceptual modelling
5. Comparison (between 2 and 4)
6. Definition (of a range of possible changes)
7. Selection (of a desired, agreed-to-be-feasible change)
8. Design (of the agreed change)
9. Implementation of the agreed change.

Stages 3, 4, 8 and 9 involve "further systems thinking" (explained below). Central to Checkland's methodology are stages 3 (root definition) and 4 (conceptual modelling). The root definition consists in an expression of "the basic nature of the system or systems thought to be relevant to the problem situation" (p. 100) and is essentially "a condensed representation of the system(s) in its most fundamental form". There may be a number of root definitions corresponding to the perceptions of different actors in the real situation. Conceptual modelling is a development of the

root definition to model the systems relevant to the problem. It is done by considering the system as a "human activity system" and by allowing only the minimum of activities such that the root definition is satisfied (usually straightforward "events" in a business process). Since the model will form the basis for a decision about the best course of action (to solve the problem) it is obviously very important that the model is valid. Checkland proposes that the conceptual model can be checked against his own "simple general model of a purposeful human activity system" (p. 109) or Beer's organismic model (1972), but since the emphasis is on "usability rather than sophistication" this will not always be necessary. This leads to fairly simple descriptive models of the system.

A severe criticism of Checkland's methodology is that it does not generate models based on the way the system works (i.e. a theoretical model), but simply a descriptive model based on the root definition. Thus there is no guarantee that the implementation of the agreed change (stage 9) would produce the results expected using the model. This criticism has been developed constructively by Molloy and Best (1980) who argue that the methodology can be augmented and construed as a "theory building methodology". This involves a replacement of the statement or "sentential" type of conceptual model (based on the root definition) by a dual "sentential-iconic"model, that is a "statement-picture complex" in Harré's sense (1970, see Section 3.3.4). The picture or iconic model provides a hypothetical underlying mechanism for the system, i.e. a theory. They propose that such a model could be based on the organismic model of Beer (1972) which provides a model of a viable system. Initially, the use of the model would be paramorphic (Harré, 1970), but with development it would become homeomorphic and offer a full theoretical explanation of the system. A further point should be made, that Molloy and Best omit, that the model should be validated in its present role as a model of the specific system, and this can be done to the viable system model as well. The introduction of an explanatory model may lead to a reperception of the original problem and, heuristically, the validity (or value) of the model may have to be regarded in this light. These technical aspects can be understood simply by the maxim that somewhere in the methodology there should be the requirement of improving the understanding of how the system works.

Models of "viable" systems, such as Beer's organismic model (1972) or M'Pherson's"proto-system" (1980), are based on a specific type of system, biological and technological, respectively. The claims for

generality of these models should be assessed by validating them in other areas than their source and not by the GST metaphysical principle ("this is a system, therefore ..."). Furthermore, the model may not be a valid representation of its source material (Beer's model is now erroneous by current neurophysiological theory).

An example of a methodology based on the systems analysis school is that of Mattessich (1978) whose book purports to be an "epistemology of the applied and social sciences". It is essentially a thorough development of decision theory (going through logic, deductive logic, multi-valued logic, inductive logic and probabilistic induction, Von Neumann/ Morganstern decision theory, to probabilistic decision theories) sandwiched between relatively short sections on systems philosophy and epistemology. Mattessich is concerned with applied sciences, or the instrumental use of reason. He is thus concerned with "epistemic utility" as opposed to "cognitive epistemology" or, in other words, truth only if it is useful, and so decision theory (being normative) plays an essential role. In Chapter 7, "Philosophy and the Systems Approach", he shows how there is a strong normativistic element in systems methodology and the philosophies of the Weltanschauungen schools (Hanson, Kuhn, Feyerabend). It is difficult to see how he justifies his emphasis on logic, confirmation theory, etc. in earlier chapters when such philosophers completely reject such an orientation in the philosophy of science. In fact, Mattessich offers very little by the way of methodology for applied and social sciences and his positions on epistemology, theory development and validation are not clear.

3.4.5 Some philosophical problems of systems science - and some possible solutions (with reference to model validation)

This section has considered some of the philosophical issues associated with systems science, ranging from the grand claims of GST to the practically-oriented systems methodologies. Despite the differences between different aspects, there are several common threads or claims: at some stage in scientific activity it is necessary to look at a more global level (synthesis); the theories or methodologies have a wider generality than is usually the case; systems science nearly always entails working across conventional disciplinary boundaries. These raise questions which are both philosophically important and of relevance to the practice of systems science. In particular, these questions are associated with the generality of systems theories/ methodologies, and the analogical transfer of models from one discipline to another, and raise important issues in model validation. The outline

of the philosophy of science in the previous section provides some very powerful philosophical tools with which to pose and tackle these problems, and many will be incorporated in the theory of model validity developed in the next chapter.

The over-optimistic claims made by people such as Bertalanffy and Forrester on the potential of systems science, and its failure to go anywhere near meeting them has led to some stinging criticisms of systems science. Berlinski (1976), in an informed yet negativistic criticism, attacks mainly technical problems associated with systems science; e.g. the definition of a system; the application of Forrester's "system dynamics" to world modelling (Meadows et al., 1972); some cybernetic models in biology, etc.. Lilienfeld's critique (1978) is an ideological one associating the rise of systems theory with the growth of an authoritarian bureaucratic scientific elite and applies equally to the elitification of all areas of intellectual activity. Although the target of his critism is GST (his own interpretation of it), it should not be dismissed lightly and systems scientists should remember that their subject, with all its jargon and complex computerised techniques, makes scientific knowledge even further removed from the average person and a more restricted commodity.

The generality of systems science leads to some difficult problems in model validation. Firstly, general theories such as GST are assumed to apply to any system, no matter what area it is drawn from. If the theory could be validated empirically for some specific systems, could its general validity be inferred? The answer to this is affirmative only if a metaphysical hypothesis is accepted akin to an induction hypothesis (see Section 3.2.3), namely that there is a regularity in the world that holds in all areas. Such a hypothesis is even worse than an induction hypothesis since it asks for the acceptance not just of a regularity, but a regularity of a certain kind (i.e. systems), and not as an empirical hypothesis, but as a metaphysical necessity. At best, a general system theory could be validated (theoretically and empirically) in a finite number of application areas, and, probably, the degree of validity would vary greatly due to theoretical difficulties, measurement problems, etc. In these areas, and using domain knowledge (of each area), a postulate of "appropriate chunking" will apply such that Bertalanffy's "ontological-must" argument holds, i.e. "this chunk of reality (in this area) is a system and therefore must have certain system characteristics". This takes the problem

into a complex middle ground (between general theories and models of specific systems or areas) where some concepts, models and theories are believed to apply to a number of different systems (areas) and these are far removed from GST.

The second type of problem for model validation in systems science occurs in this middle ground where concepts and models are freely transferred between different areas (this is characteristic of, but not limited to, systems science). What epistemological and scientific basis do such models have? What criteria are there for judging the results of the analogical transfer of models between areas? The following considerations may help determine the epistemological basis and heuristic potential of this aspect of systems science:

(i) An analogical model is a paramorph (Harré, 1972) which provides a hypothetical mechanism to explain phenomena in the new area (as a "statement-picture complex", Section 3.3.4); it gives structure and hence understanding to an area (Hanson, 1958, Section 3.3.2).

(ii) Empirical validation may be possible, but a model will certainly introduce new terms and variables ("language-shift" - Toulmin, Section 3.3.4) and its data requirements for extensive empirical validation may not be met until later. An important aspect is to use patterns in the data to reveal structural properties of the system.

(iii) According to Lakatos' theory of research programmes (1970, Section 3.3.4), a research programme is theoretically progressive if the empirical content of a domain is increased, and empirically progressive if this is also empirically validated (as in (ii)).

(iv) The introduction of an analogical model may lead to a redefinition of an old problem in the area or to the generation of new problems which are considered more important or fundamental (Popper's tetradic schema, 1962, Section 3.3.4; Lakatos' problem-shifts, 1970, Section 3.3.4).

(v) In terms of Shapere's "domains" (1977, Section 3.3.4) - does the model solve a domain problem (e.g. in reorganising knowledge) or a theoretical problem (which requires a theory to explain the domain)?

(vi) The use of the model may not satisfy (i) to (v) but may nevertheless be instrumental in meeting some other objective (such as improved health care). However, the satisfaction of such objectives usually requires, and is certainly not impaired by, an increase of knowledge and explanation of phenomena in the area in question brought about by the use of the model.

These are some of the aspects that will be included in the theory of model validity (Chapter 4) with relation to the role of the model in the overall process of model development or scientific activity rather than its representational validity of a specific system. In Chapter 4, this is referred to as "heuristic" validity. The relative emphasis on these two distinct features is determined mainly by the specific modelling objectives in the particular validation study and the stage of domain development. (In Chapter 5, a range of validation methodologies is devised using the theory of model validity. The ε-methodology (Section 5.6) is intended as a heuristic methodology for the validation of analogical and innovative models and develops many of the above considerations.)

Similar types of consideration may be made concerning the validity of the application of systems methodologies. Primarily, these can be assessed pragmatically, i.e. by checking whether a methodology achieves its utilitarian objectives (e.g. design or optimisation of some system). The range of application to different systems or types of problem can be determined from a critical philosophical and theoretical examination of the methodology or by testing it out in practice. Quite often a methodology will embody some theoretical concepts or rest upon the development of a model (as in the Checkland methodology) and in these cases the validity of the theory or theory-building procedure will be of paramount importance.

As systems science progresses, there will be a growing body of systems concepts, models, theories and methodologies, and at the same time a more critical delimiting of their ranges of valid application (theoretically, empirically, pragmatically and heuristically). If this results in a substantive common body or core, then the claims of the visionaries such as Bertalanffy will have been borne out. This core can be regarded as the research aim of systems science which distinguishes it from most other sciences, but it must be strived for in a legitimate scientific manner and emerge as a property of the world. It can never be a metaphysical assumption that systems of this kind do exist and systems scientists must be aware of the possibility that the critical and rational development of their individual researches may lead to an irrevocable and fundamental differentiation of systems theory. In any case, the evidence is now that systems science will only augment conventional science and never replace it.

3.5 Philosophy of Measurement

3.5.1 Introduction

The theory of model validity developed in Chapter 4 (and, indeed, a philosophy of science) calls for detailed considerations of the nature of data available from the system (object, phenomena, etc.) which is modelled. By "data" are meant the records of observations, experiments and measurements of the system. These considerations range from an examination of the possible inaccuracies in measurements, through a delimiting of the extent to which the system is practically and theoretically observable (or measurable), to a philosophical analysis of the foundations of data. In this section, the philosophy of measurement will be outlined with reference to the theory of measurement. The theory of measurement is concerned with an analysis of the logical foundations of measurement and stems from the positivistic analyses of Mach on the concept of mass (1868) and temperature (1896) and Helmholtz's analysis of counting and measuring (1887). Mach intended to show that all scientific theories could be reduced to, or deduced from, the basic elements of sensory experience which were pure facts. However, all measurement contains some theoretical content and is not a simple connection to the facts of reality. The theory of measurement can be regarded as a theory of the elementary level of theory in measurement. Evidence for the relation of theory and measurement is contained in the simultaneous growth of understanding, theories and models of an area and the development of measurement science in that area (e.g. the use of isotopic tracer techniques in biology and the concepts and theories of compartmental analysis).

The philosophy, or theory, of measurement has received little attention in contemporary philosophy of science. There are a number of reasons for this, of which the following two are perhaps the most important: firstly, Hanson's analysis of the theory-ladenness of observations (1956) and the Weltanschauungen philosophies of Feyerabend and Kuhn led to subjectivist epistemologies where the measurements were considered determined by the theory (the criticisms of this view are given in Section 3.3.2); and, secondly, the focus of attention has shifted from the individual theory or model to the overall dynamic process of scientific development (e.g. Lakatos, 1970, Section 3.3.4), in which considerations of particular details of measurement have played only a small role. Despite this, measurement and observation play an essential role in scientific development as the means for finding out about reality. It is likely that a

historical-realist approach such as Shapere's (1977), which focusses on the development of a scientific domain, will provide a much more satisfactory epistemology which will clarify the nature and role of measurement.

The rest of this section is structured thus: firstly, the historical development of measurement theory is outlined; secondly, some recent trends and problem areas are discussed; and, finally, the implications and relevance for the theory of model validity are assessed. For a comprehensive review of measurement theory consult Leaning (1977).

3.5.2 History of measurement theory

Modern measurement theory stems from the work of the nineteenth century mathematicians and physicists on the foundations of mathematics and the elementary concepts in science (Mach (1868, 1896) - mass and temperature; Helmholtz (1887) - counting and measuring). Hölder (1901) formalised the axioms for the measurement of empirical quantities (attributes or properties) that were additive. The British physicist Campbell (1920, 1928) analysed the fundamental nature of physical measurement and gave three rules for measurement:

(i) Measured attributes are capable of being ordered.
(ii) Measured attributes are additive.
(iii) A copy can always be found.

In essence, this forms the basis of a theory of measurement, although nowadays it is recognised that these rules are necessary (not not sufficient) conditions for "extensive" measurement, and that there are many other types of measurement. Many of these were discovered by social scientists for dealing with attributes that are non-extensive. The classic work by Von Neumann and Morganstern (1944) was a theory of utility and chance in economics which was based on the axioms of preference and decision, and stimulated the development of both the theory and practice of measurement in the social sciences (e.g. S. S. Stevens, C. H. Coombs).

Tarski's concept of a relational system (1954) allowed a new axiomatic formulation of measurement theory (Scott and Suppes, 1958). In this approach measurement is regarded as a homomorphic mapping between an empirical relational system and a symbolic relational system, such that the symbols assigned in measurement represent relations between the empirical attributes of objects or events. Such a theory is known as a "representational theory of measurement", and put the analysis of measurement on a formal basis which allowed the methods of modern logic to extend

greatly the range and scope of measurement theory.

Ellis (1966) produced the first book devoted solely to a philosophical analysis of measurement since Campbell (1928). The first presentation of a unified axiomatic theory of measurement was by Pfanzagl (1968). A comprehensive theory of measurement applying both to physical and social sciences was given by Krantz, Luce, Suppes and Tversky (1971) which uses all the devices of modern logic, set theory and mathematics, although still leaves many questions unanswered (see Section 5.3). Measurement had been treated as a numerical assignment, but Finkelstein (1975) showed that it could be generalised to representation by symbols.

At the heart of measurement theory is a definition of measurement (usually expressed formally) as "the objective assignment of symbols to attributes of objects or events in such a way as to describe the relations between them". Measurement theory then proceeds technically by showing what basic empirical properties must be discernible (in respect of an attribute) for that attribute to be measurable on a numerical or any other type of scale. However, the definition also allows the clear articulation of philosophical questions on the nature of measurement. For instance, there has to be a priori identification of the type of object or events appropriate for the attribute; if the attribute is a universal concept, what is its ontological status? At the primitive stages of measurement, the objects are items of everyday experience and the ontology of the attribute is linked to a set of particular empirical operations. As measurement develops the attribute becomes conceptualised and theoretical understanding develops. At a sophisticated level of measurement, the attribute is a variable in a theory (independent of an empirical operation), and the identification of objects is determined by the theory. To deal with these questions properly the theory of measurement needs augmenting with some more philosophical apparatus. However, although it provides a static view of measurement, it nevertheless provides a very good basis for understanding measurement across the whole scientific spectrum.

3.5.3 Some recent trends in measurement theory

Although all measurement depends on theory, it is usual to distinguish between direct and indirect measurement. Direct measurement consists of measurement of an attribute involving that attribute alone. In indirect measurement, however, the value of an attribute may be determined by measuring other related or component attributes and combining the results together in a numerical law or mathematical model. In practice, nearly all

measurement is indirect (or "conjoint" in the psychological literature). A well-known property of laws in physics is that they are dimensionally invariant monomial functions and that attributes (variables) are related multiplicatively. The reason for this has long been of interest in measurement theory (e.g. Bridgman, 1922; Causey, 1969), yet no satisfactory explanation in qualitative empirical terms had been offered. Recently Luce (1978) has shown that dimensionally invariant laws correspond to meaningful qualitative relations on the empirical relational system. The latter are simply the set of automorphisms of the original relational structure. Much work needs to be done in extending the theory of measurement to include indirect measurement based on models. A clear prerequisite for using models in this way is that they have been validated and is therefore highly relevant to this thesis.

Another problem area of measurement theory is that of uncertainty. The primary source of uncertainty in measurement arises from the inability to discriminate precisely the fundamental empirical relations. Only a few attempts have been made to develop a theory of measurement for uncertain relations: the algebraic deterministic theory of semiorder (Luce, 1956); the qualitative probabilistic approach (Domotor, 1969); and a numerical probabilistic theory based on a probabilistic homomorphism (Leaning, 1977). This still remains an area for further research.

3.5.4 Implications for the theory of model validity

The most important implication for the theory of model validity is that the theory of measurement provides a good explanation of the nature of measurement which is a central aspect of the theory of model validity. For instance, it explains the different scale types, the dependency on theory (particularly indirect measurement), the differences in measurement between different sciences, etc. An understanding of this latter point provides a major reason for the differences in validation methodologies in different sciences. In physics, most measurement is indirect, relying on valid general laws, theories, and models. In biology, there is a growing trend towards using models for indirect measurement to cope with the complexity and relative impenetrability of biological systems. In the social sciences, by contrast, measurement is still largely a correlational activity, where theoretical concepts and empirically measurable attributes often do not correspond. One common and interesting feature, however, is the emphasis on the multidimensional character of measurement, and this suggests that theories and models are applied as a whole and not

built up from knowledge of individual empirical attributes.

More practically, the theory of measurement can aid in coping with measurement uncertainty and the meaningful use of measurement data in empirical model validation. For instance, if a variable is measured ordinally, it cannot be used to validate cardinal properties of the model. Similarly, there is a variety of statistics, appropriate to different scale types, which can be used to deal with measurement uncertainty. Individual biological systems within a class (e.g. human cardiovascular systems) show considerable variation when measured and measurement-theoretic considerations can help extract quantitative features for the data that do not show such variation and can be used to validate general models of the class rather than an individual.

3.6 Conclusions

In Chapter 2, most of the scientific literature reviewed treated validity as synonymous with empirical correspondence, which is also the epistemological basis of logical empiricism (Section 3.2). However, logical empiricism has been severely challenged in the recent philosophy of science, in particular for its views on theory or model confirmation (validation) and the basis of scientific knowledge (Section 3.3.1). The extreme views of the Weltanschauungen philosophers (Feyerabend, Kuhn, Section 3.3.2) led to a concept of validity that was equated with coherence with existing theory at most times and non-existent at times of dramatic scientific change. Unfortunately, this approach does not ultimately require that scientific theories or models express information about an objective reality and is therefore defective.

The historical realist school of the philosophy of science (Shapere, Suppe , Section 3.3.4) is based on detailed studies of the actual dynamic process of science (unlike logical empiricism which is essentially a static reconstruction of science). The epistemology of historical realism is dependent on the assumption that an external reality exists and that there are critical methods for obtaining objective knowledge about such a reality. These methods include those based on the empirical and theoretical concepts of validity and, additionally, the use of "good reasoning patterns" which are associated with the recognition of the potential or fruitfulness of a theory, model, research programme, etc. (i.e. "heuristic" potential). The relative importance of empirical, theoretical, and heuristic validity

concepts depends largely on the content and stage of development of the scientific domain associated with a theory or model. There is much evidence, however, that many of the decisions on retaining, modifying, or rejecting a model are made using heuristic considerations.

An important omission of the philosophy of science (including historical realism) is the consideration of pragmatic validity, which is associated with the practical use of theories and models. Although pragmatic validity should not affect representational (theoretical, empirical) or heuristic validity it may have a significant effect on the content and direction of an area of scientific research. The concept of pragmatic validity emerged in the review of Chapter 2, and is operative, for example, in the validation of methodologies for "real-world" problem solving (e.g. systems methodology, Section 3.4.4).

In the next chapter, a theory of model validity is developed in which the four concepts of model validity (empirical correspondence, theoretical coherence, heuristic potential, and pragmatic value) are expressed as sets of validity criteria. In this theory, the relationship between the modelling objectives, the nature of available data, and the validity criteria is carefully analysed and explained. The modelling objectives and available data are taken as indicators of the content and stage of development of the scientific domain associated with a model, although the role of domain knowledge in model validation is also considered separately. The review of the philosophy of measurement (Section 3.5) provides a basis for the theory of data in Chapter 4.

In Chapter 5, the theory of model validity is used to devise a range of validation methodologies. One of these (the ε-methodology, Section 5.6) is heuristically-based and is suitable for dealing with the problems of validation that arise in connection with the analogical transfer of models between different domains that occurs in systems science (Section 3.4). In conclusion, the review of the philosophy of science presented in this chapter has been very useful in explicating the various concepts of model validity listed in these conclusions and in explaining the role of model validation in the ongoing process of scientific development.

CHAPTER 4

A THEORY OF MODEL VALIDITY

4.1 Introduction

4.1.1 Introduction

In this chapter a theory of model validity is presented which forms the central core of the work reported in this thesis. The function of the theory is to clarify and explain the various concepts of model validity that have emerged from the previous two chapters. It does this by classifying the diverse aspects of model validity, and relating them to a set of possible modelling objectives and a theory of data. As such it can be regarded as a form of extended definition of model validity. However, it is a theory because it not only describes and classifies the various aspects of model validity but it explains their relative importance under certain conditions (modelling objectives and data availability) and how the operational use of them in model validation can lead to a satisfaction of modelling objectives. There are two distinct applications of the theory:

(i) In the critical analysis and assessment of model validation in the development of particular research areas (the province of the history and philosophy of science);

(ii) To develop validation methodologies in areas where there is a lack of adequate methodologies, or existing ones are problematic.

The former has played the larger part in the development of the theory (based on scientific and philosophical reviews; Chapters 2 and 3), whereas the latter is more important to the main aim of the thesis, namely the development and application of validation methodologies with emphasis on mathematical models in biology and medicine. The use of the theory in generating methodologies is considered in Chapter 5, and the application of appropriate methodologies for biological models is illustrated in Chapters 6, 7 and 8, and for models in the social sciences in Chapter 9.

4.1.2 Preliminary definitions, assumptions and other considerations

Before outlining the nature, scope and application of the theory (Sections 4.1.3 and 4.1.4), some preliminary definitions, assumptions and other considerations are necessary. Although the definitions of a valid model and model validity are intentionally a little idealistic and general and their inadequacies highlight the need for a theory of model validity, nevertheless in essence they underpin the entire approach. In Section 4.7, some of the philosophical implications of these definitions and assumptions and of the theory itself will be raised. Firstly, an assumption must be made about the primary aim of science.

A. The primary aim of science is the rational acquisition and evolution of knowledge and understanding of phenomena which themselves have an objective existence.

This assumption leads to a realist epistemology that the world exists and that there are legitimate scientific ways of finding out about it, and on which the concepts of model validity and techniques of model validation must be based to avoid the dangers of subjectivist epistemologies. It is not trivial (see Chapter 3).

In science there is a great diversity of knowledge representation devices ranging from descriptive accounts, reports of observations or measurements, diagrams and physical analogues, to the precise abstractions of symbolic and mathematical models and theories. This thesis concentrates on the validity of mathematical (or theoretical) models. A model is defined thus:

D1. A model is a theoretical device which embodies both description and explanation of phenomena.

This definition is based on the use of models in actual scientific practice and differs from the notions of a model on a theory (or formal language) or a model as an exemplar, although it may also satisfy them. A model, therefore, is intended to satisfy the primary aim of science (A). The distinction between models and theories, and the use of models for other (pragmatic) objectives are discussed further below. It should be noted at this stage, however, that the view taken in this thesis is that the successful use of models for pragmatic purposes (such as system optimisation) depends on their scientific nature (i.e. satisfying D1).

The adjective "valid" is derived from the Latin, validus-valere, meaning to be strong. The conventional meanings of valid are strong, sound, legitimate, or efficacious with respect to a set of rules or

purposes. Typically it is used in legal terminology (valid according to legal rules and procedures) or in logic (a valid derivation from a set of axioms). The use intended here is that designating legitimate or efficacious with respect to a specific purpose. A valid model can be defined as:

D2. A valid model is one which satisfies the purposes (or objectives) for which it is intended.

Adopting the definition of a model (D1) with the primary aim (A), a general definition of a valid model for scientific purposes is:

D2a. A valid model is one which represents the phenomena of interest.

By "represent" is meant both a description of the phenomena and an explanation of them. This involves a correspondence of the model with the phenomena themselves and with the other factors that lead to these phenomena (i.e. a mechanism or explanation). Model validity can be simply defined as:

D3. Model validity is the extent to which the model satisfies the purposes for which it is intended; or:

D3a. Model validity is the extent of the representation of the phenomena by the model.

D2a and D3a lead to the notion of an isomorphism or partial isomorphism (homomorphism) between the model and the phenomena. Model validation is then the process of discovering the extent of the homomorphism with the phenomena of interest. This would involve comparing the model with the facts (observations and measurements) concerning the phenomena. D2a can be rephrased as:

D2b. A valid model (of a set of phenomena) is one which corresponds to the facts about the phenomena,

and D3a as

D3b. Model validity is the extent of the correspondence between the model and the facts about the phenomena.

It is at this stage that many problems arise. In general, these are concerned with the philosophical difficulties associated with the correspondence concept of truth (Chapter 3, Section 3.2). Some of these are discussed in Section 4.7, but for the rest of this subsection some realistic scientific considerations will be made which temper the above definitions. The main problem arises from the fact that "facts" are very rarely neutral records of phenomena.

In science knowledge about phenomena is contained in many forms ranging from data records to more or less validated models and theories. So-called "facts" themselves contain theoretical concepts (possibly even structured by the model with which they are being compared). Of course, the scientific basis of models must be in experience, but this is an evolutionary process in which theories and models develop from simple concepts rooted in everyday observation to sophisticated abstractions with which to regard and shape phenomena. The various areas of scientific research form distinct evolving bodies of knowledge, concepts, etc. which may be called "domains" (Shapere, 1974; see also Section 3.3.4).

D4. A domain is a more or less structured body of knowledge related to a certain research area and which may contain data (putative facts), hypotheses, models and theories, etc.

A domain is not necessarily internally consistent and may contain problems (contradictions, inadequate models, incomplete data banks, and so on) which drive the research ahead. As the domain evolves there will be changes in models, theories, conceptual meanings (semantics), important problems, and maybe even in what are considered facts. If the development of the domain is progressive then the major models and theories will be considered highly valid, and a new model will be assessed primarily against these, whereas at an early stage in the domain's history models will be validated mainly against data. Thus the "stage of development" of a domain will affect substantially what is meant by a valid model. The two meanings of model validity distinguished so far are coherency with other theories or models, and correspondence with the facts. Thus the definitions D2b and D3b are insufficient since they entail only correspondence with facts, and the less precise definitions D2a and D3a must be adopted instead. (Philosophies or theories based on D2b and D3b will inevitably lead to an incomplete analysis of science as evidenced by the failure of the positivist programme; see Chapter 3, Section 3.3.1).

The notion of a dynamic domain suggests that models may also play a role in the evolution of knowledge which is not simply checked against other models or theories, or data. This role is that of the model in helping discover new phenomena, gain further understanding, or to redefine the problems of interest, and may be summed up as a recognition of the "heuristic potential" of the model. There is strong evidence that the decision to accept models in science is often made on this basis (see Chapter 3, Section 3.3.4) and yet there are few philosophies of

science or scientific theories of model validity which are explicitly aware of this fact. The next step in this analysis of the concept of model validity is to note that models are frequently used to satisfy some practical or utilitarian purpose, such as the improvement of a system. To the users of models in this way, a model will be considered valid (in a pragmatic sense) if it does indeed achieve the practical end required of it, regardless of whether it satisfies D2a and D3a in representing the phenomena. In other words, if it satisfies the first general definitions of a valid model (D2) and model validity (D3) in terms of the modelling objectives.

The single concept of model validity as correspondence with the facts has therefore been replaced by a multidimensional concept which includes empirical or factual validity, theoretical validity, practical value, and potential for scientific development or understanding. These different aspects are united only at the level of validity as a satisfaction of modelling objectives. It might be argued that this simply equates validity with acceptance, and that validity should relate to empirical and perhaps theoretical aspects only. To some extent the first part of the criticism is correct, and the adjective valid is used because its root meaning is strong, legitimate, sound, acceptable, etc. However, to call a "valid model" an "acceptable model" blurs an important distinction. The various concepts of validity above are the types of scientific reason for accepting a model, but in practice the acceptance of a model by a scientific community is a sociological event, and may be modulated by psychological, social, political, or economic causes. In other words, scientific reasons of validity plus non-scientific causes equals acceptance. Scientific honesty tries to keep the causes as small as possible, a point not fully understood by many political and social analyses of science (which cannot explain therefore why science works). The second part of the criticism is incorrect and leads to incomplete theories of model validity as well as defective validation procedures (see Chapter 2 for many illustrations).

The theory of model validity developed in Section 4.3 extends the above analysis considerably and shows how the different aspects of validity are interrelated and how different aspects become important under different modelling objectives, data theory, and stage of development of the domain. This subsection concludes with an examination of the differences between models and theories.

Usually, models have a greater specificity than theories but more generality than data records. Like theories, and unlike pure data, they may embody explanation (understanding) as well as description (Harré's "statement-picture complex", 1970, see Section 3.3.4). Sometimes a model is considered to be "on a theory" when a theory is reduced by constraints, initial conditions, etc., so that the model satisfies the theory. This assigns a very unimportant role to models and many writers concluded that models were unnecessary (e.g. Braithwaite, 1953). However, models play essential roles in scientific development. For instance, in biology a mathematical model may be based on a descriptive theory with additional empirical data, and will be theoretically (or formally) richer than the theory itself. Although the distinction between models and theories is useful, no strict dividing line can be drawn which does not eliminate most of the interesting properties of models. The theory of model validity may also apply to theories, but this is not elaborated further.

So far, in this chapter, the use of the word "system" has been avoided. It is tempting to talk about a "model of a system" and this implicitly involves a presupposition that the existence of the system is self-evident, but "system" is itself a theoretical concept. It is the modelling of a group of objects/events/phenomena which imparts the structure and properties of system to them. If the model turns out to be valid, then it is justified to consider the grouping a system (or, more correctly, a type of system with the type-properties this entails). This consideration ties up a loose-end in systems science which is generally considered to be model-based.

Another interesting implication for model validity in systems science arises from the concept of domains. The essence of systems science is interdisciplinary, taking models and concepts from one area and applying them to another. This involves a combination of the systems science domain (mathematical modelling, etc.) with a host domain, or domains (e.g. human biology and pharmacology). Since the content and stages of development are very different it is highly likely that the research and modelling objectives, and operative validity concepts will mismatch. This mismatch is often reflected in a partial language barrier, and offers a challenge for the application of the theory of model validity (Chapter 5, Section 5.6).

4.1.3 Nature and scope of the theory

The theory of model validity elucidates the various aspects of model validity and their relation to modelling objectives, etc. It can be regarded as an analytical framework, but a more precise notion is that of a metatheory (introduced in a similar context by Tarski in his famous paper on the concept of truth, 1931). The terms of the metatheory provide a metalanguage with which to talk about the validity of models, and are structured (related) according to the metatheory. The subject of a metatheory is theories (or models) and it is therefore distinguished from theories (or models) whose subjects are empirical phenomena. The concept of the theory of model validity as a metatheory has two immediate implications. Firstly, in order to talk about the validity of models it is necessary also to talk about the models themselves and the phenomena they are intended to represent. This entails that the theory should contain, or be part of, a theory of models and should also contain a section on the nature of data (a theory of data). Secondly, model validity at the level of the model is a semantic concept which, at the level of the (meta)theory, becomes a syntactical property (i.e. obeys certain rules). This suggests that the various aspects of model validity may be expressed as a set of criteria.

The theory is a verbal scientific theory with occasional uses of symbolism for brevity and precision. The structure of the theory is classificatory. Some use of set theory is made.

There are three distinct origins of the theory. The first is the work on the validation of a group of biological models (reported in Chapters 6, 7 and 8) which has formed the major topic of research and has provided many ideas (and problems) on the nature of model validity. The second is an extensive review of the literature of model validity and validation in a wide range of sciences (Chapter 2). Most of the work reviewed concentrated on methodologies for model validation, and tended to be oriented to a particular class of models (e.g. simulation models, see Chapter 2, Section 2.3, or Fishman and Kiviat, 1968). Some exceptions, which have been formative in the development of the theory, are those by Kaplan (1964, Section 2.4.2), who related model validity to model purpose and questioned the validity of data; Hermann (1967, Section 2.4.3) whose seminal paper related a range of modelling objectives to a set of validity criteria; Karplus (1977, Section 2.4.7) who distinguished between deductive and inductive validity; and Gruhl

(1979, Section 2.8.3) who suggested that, in model validation, the overall methodology of modelling should be examined in order to expose the critical points at which uncertainties, or corruptions, are introduced into a model. These contributions were more oriented to the nature of model validity rather than to validation methodologies and this suggests that successful validation methodologies will be based on a good understanding of what is meant by a valid model. Also, in this second area, work on the validity of biological models was reviewed and, in particular, the papers by Berman (1963, Section 2.5.2) and Carson and Finkelstein (1977, Section 2.5.5) were most helpful. The latter suggested that the validity of biological models was problematic because of the lack of subsystem data and/or theory. A related area is that of system identification and, whilst little attention had been paid to model validity by the beginning of the project, an increasing emphasis has been given in the recent literature (e.g. Mehra, 1980; Carson, Cobelli and Finkelstein, 1980; both Section 2.6.4). In this area, the influences range from the social and behavioural sciences (Kaplan, Hermann), to the biological sciences (Berman, Carson and Finkelstein), to energy system modelling (Gruhl), to physical systems (Karplus, Mehra).

The third area is that of work on the philosophy of science (Chapter 3). This has been an important element in the development of the theory. The philosophical status of the theory (which is not intended to be a philosophical theory) lies in the historical realist school of the philosophy of science as espoused by Toulmin (1953, 1972, Section 3.3.4), Lakatos (1970, Section 3.3.4), Harré (1970, Section 3.3.4) and Suppe (1977, Section 3.3.4), although, naturally, the focus on the specific issue of model validity is rather different. Detailed considerations of the philosophical significance of the theory are left until Section 4.7.

The scope of the theory is model validity in any scientific or application area. The theory is intended to be as comprehensive as possible, although additional types of modelling objectives or validity concepts may be required for areas that have not been directly considered. This should not affect the general form of the theory, however. The wide scope of the theory arises from its developmental base in many different areas, as described above, and not as an arbitrary generalisation. This grand claim is justified in part by the fact that the theory is relatively simple.

The theory offers an explanation of model validity which is an acceptable objective in its own right. However, the main application of the theory is for models which cross disciplinary, or domain, boundaries and for which the concept of model validity and legitimate techniques for validation are problematic. This characterises the biological models in the case studies (Chapters 6, 7 and 8) and models in systems science in general, and is the central concern of this thesis.

4.1.4 On the application, testing, and development of the theory

The theory of model validity has two types of application:

(i) Critical or historical analysis of modelling methodologies (including model validation). It provides a very powerful tool for critical analysis.

(ii) In the suggestion of appropriate validation methodologies.

The second application is slightly normativistic: a consideration of the modelling objectives, stage of development (of the domain), and data theory determine what aspects of validity should be important. These can then be used to devise an appropriate programme or methodology for model validation. This use is exemplified in Chapter 5.

The theory is not a metaphysical theory which is self-evidentially true, but one which can be used, tested, and changed. The testing may occur through use in devising validation methodologies for problematic areas, or through historical studies of actual scientific practice. Since the subject of the theory is models (or theories) it cannot strictly be applied to theories of theories (metatheories), but a fruitful way of testing it might be by applying the theory to itself.

Undoubtedly, testing the theory will lead to inadequacies, and it is expected that the solution to these will entail an extension of the set of modelling objectives or validity criteria, or the role of domain knowledge, but not a change of the general form of the theory. If this ultimately proves to be necessary, then the value of the theory will be in encouraging more research effort on the concepts of model validity and the development of efficient and appropriate methodologies for model validation.

4.1.5 Summary of chapter

Section 4.2 briefly outlines the theory of model validity, which is presented in full in Section 4.3. The role of model validation in the

overall modelling process is discussed in Section 4.4. The way in which the theory can be used to devise ("generate") validation methodologies is described in Section 4.5, as a prelude to Chapter 5. Section 4.6 is concerned with the ways in which the theory may be tested, and Section 4.7 deals with the philosophical significance of the theory.

4.2 Outline of the Theory of Model Validity

The theory of model validity presented in the next section has four parts. These are briefly outlined below, together with an introduction to some of the symbolism used in the theory.

(i) Analysis of modelling objectives (Section 4.3.1)

This is concerned with the articulation, analysis, and classification of the various types of modelling objective. In Section 4.1 a definition of model validity was finally adopted that related validity to the role a model is intended to play in an ongoing research programme. In the theory of model validity the various roles are represented as explicit modelling objectives. The set of modelling objectives, O_i, is denoted by $\mathcal{O} = \{O_i\}$. The classification of $\mathcal{O}$ is related to the different ways in which models contribute to the critical evolution of knowledge in a specific scientific domain. (The same is true of the validity criteria (iii).)

(ii) Theory of data (Section 4.3.2)

This part of the theory deals with the means by which empirical phenomena may be represented, ranging from the verbal accounts of observations to the mapping into a formal abstract space (measurement). The different forms of representation are known as data types, D_j, and denoted by the set $\mathcal{D} = \{D_j\}$. The relationship between the available data types and the stage of development of a domain is considered. In general, as a domain develops, evolving theories and models, the repertoire of data types expands, eventually including representation by measurement. The emphasis is on the logico-empirical foundations of data types rather than on specific techniques. The nature of experimentation and the problem of uncertainty are also investigated.

(iii) Validity criteria (Section 4.3.3)

The validity criteria are a set of rules expressed in natural language. They provide either a basis of comparison, or a basis for assessment and cover the range of possible modelling roles. Validation

consists in the application of the appropriate criteria for the model. As far as possible the model should satisfy each applicable criterion; however, this is rarely an exact yes or no decision. The set of validity criteria, V_k, is denoted by $\mathcal{V} = \{V_k\}$.

(iv) Synthesis of the theory (Section 4.3.4)

The final and most important part of the theory is concerned with the overall relationship between modelling objectives ($\mathcal{O}$), data theory (or types) ($\mathcal{D}$), and validity criteria ($\mathcal{V}$). The operative, or relevant, validity criteria for a model are determined by the model's objectives and the available data types. In general, therefore, the theory consists of the map $\mathcal{O} \times \mathcal{D} \rightarrow \mathcal{V}$

Although the meaning of model validity changes with the stage of development (for instance, theoretical sophistication) of a domain (area of research), the theory of model validity itself does not change. It is an assumption of the theory that, for questions of model validity, the nature of a domain is adequately characterised by the modelling objectives and data types at a particular stage of development. The independence, or generality, of the theory is assured by including a comprehensive range of $\mathcal{O}$ and $\mathcal{D}$ to describe domains in any stage of development. In applying the theory to a specific model (or class of models) considerations about the domain enter at the level of $\mathcal{D}$ and $\mathcal{O}$. The appropriate validity criteria, V_m, for the model are then determined by the theory without further reference to the domain. The criteria, V_m, may then be used as the basis of a validation methodology. (This use of the theory is described fully in Section 4.5).

The use of systems models in biology and medicine raises some interesting questions about model validity associated with domains. This type of modelling involves the interaction of at least two domains: the domain of systems modelling, and the domain of human physiology, for instance. The differing nature of the domains will result, most likely, in different modelling objectives and data types. This leads to difficulties at every level, from problem perception, to model formulation and data acquisition, to model validation, and calls for a clear articulation of the modelling objectives, as well as the theoretical foundations and data type requirements of the model. This problem and some solutions are dealt with further in Section 4.4. (Incidentally, the problem is even worse for systems models in the social sciences; see Chapter 9.)

4.3 Theory of Model Validity

4.3.1 Analysis of modelling objectives

4.3.1.1 The nature of modelling objectives

The term "modelling objectives" refers to the purposes a model may serve, the roles a model may play, or the ends a model is intended to achieve in a scientific research programme or practical application. These objectives are related to a set of problems such as inadequate scientific theories, experimental design, or to problems perceived in a socio-technical framework ("real-world" problems). When modelling, objectives are not always explicit; for instance, new areas of research may be too weakly structured to pose problems well and modelling objectives, therefore, will be vague and probably implicit (n.b. this disorder is an essential aspect of early domain development, some implications of which for model validity are discussed in Section 4.4). In established, well-developed domains, modelling objectives will often be implicitly known and yet precise - part of the legitimate research techniques. In this analysis of modelling objectives, a distinction is made between objectives of a general kind and objectives of a specific kind. The former (Section 4.3.1.2) refer to the wider, often longer-term, objectives that do not specify the application of the model, whereas the latter (Section 4.3.1.3) refer to objectives that relate the model to a class of systems or phenomena of intended application. (It might seem that the modelling objectives should include a reference to model validation or, more critically, to the invalidation of the model. However, for formal reasons, this can lead to tautologies or paradoxes stemming from the definition of model validity (D2, Section 4.1.2). To avoid this, a distinction should be made between modelling and validation objectives. Validation objectives depend upon the modelling objectives in a way determined by the theory of model validity (see also Section 4.7)).

4.3.1.2 General modelling objectives

There are two classes of general modelling objectives: scientific and utilitarian.

4.3.1.2.1 General scientific modelling objectives

This class of general objective is associated with the use of a model for the wide aim of the evolution of scientific knowledge and understanding (assumption A in Section 4.1.2). Quite simply, the most general scientific modelling objective is that a model should contribute

to the growth of scientific knowledge and understanding (in practice this is usually related to the scientific domain on which the model is based). As a representation device, it is convenient to distinguish three general scientific modelling objectives:

(a) Description. Model is required as a compact representation of empirical data.

(b) Prediction. On the basis of an existing data set, the model should be capable of making predictions in accordance with new data.

(c) Explanation. The model should provide understanding. In this role, the model may also satisfy description (a) and prediction (b), but usually it has an additional element, namely the concept of an underlying mechanism. (Frequently, an associated objective is the use of a model as a framework for hypothesis testing or theory construction.)

A fourth objective which may be based on (a), (b) or (c) is the use of models for experimental design, i.e. the optimisation of empirical knowledge gained through data acquisition. In general, the type of general scientific objectives that are important, or operative, for a particular model will depend upon the nature, content, and stage of development of the scientific domain of which it is a part. Usually, as a domain develops there is a trend from verbal description to mathematical explanation. (This concept of a dynamically-evolving domain which becomes more theoretical is very important for the epistemological base on which this theory rests, see Section 4.7.) The class of general scientific objective is denoted by θ_s .

4.3.1.2.2 General utilitarian or pragmatic modelling objectives

The class of general utilitarian objectives is associated with the use of the model in a practical application. The following list covers the more important classes of utilitarian objective, but it is probably incomplete:

(a) To improve (or optimise) an existing system, which may range from technical to social to political.

(b) The use of models for design, to meet the requirement of new needs (genuine or synthetic). Once a candidate design has been provisionally accepted, objective (a) may be used.

(c) As an aid for decision makers. By providing predictions of likely future courses of events (e.g. economic forecasts) under different circumstances, the model can help decision makers choose a "best" course of action, or policy.

(d) For "real-world" problem solving. To help define a practical problem and suggest means for its resolution.

(e) As an educational tool. A model may either be intended as a knowledge representation device (surrogate teacher), or form part of a learning system for cognitive development.

A major characteristic of utilitarian objectives is that they frequently conflict. This is a consequence of the differing viewpoints, or frames of reference, of the different users of the model and sometimes the contradictory objectives of an individual user. There are various techniques for dealing with situations like this which can be regarded as normativistic theories of rational choice; however, the amount of preference information (to construct multiattribute, multiobjective utility functions in a piecewise fashion) is very large, making their practical application very difficult. Perhaps the best achievable objective under these circumstances is an awareness and understanding of the various aims that different people have.

Models which are used for utilitarian objectives may be assessed purely pragmatically according to whether the desired practical ends are achieved. However, many general utilitarian objectives also imply general scientific objectives, and the view adopted here is that the value of models used for utilitarian purposes must rest on the scientific validity of the models used or, in other words, on the precise fact that models embody a representation and understanding of the system it is desired to change. The class of general utilitarian modelling objectives is denoted by Θ_u.

4.3.1.3 Specific modelling objectives

The specific modelling objectives specify, or delimit, the intended range of application of a model. This is the class of systems, or phenomena, together with constraints to which the model is intended to apply. To specify the intended range of application (denoted by $\mathcal{R}_I$) some domain knowledge is required, and this can be regarded as a background frame of reference for the use of the model. The two classes of specific objectives are determined by the two classes of general objective, scientific and utilitarian.

4.3.1.3.1 Specific scientific modelling objectives

These specify the class of systems or phenomena that the model is intended to represent, together with a set of constraints concerning time scales, resolution, boundaries, etc. The intended range of application becomes the intended range of representation (still denoted by $\mathcal{R}_I$). It is important to recognise the bimodality of $\mathcal{R}_I$:

(a) Structural or physical modality

This consists of a description of the structure, geometrical properties and topology of $\mathcal{R}_I$

(b) Functional modality

This describes $\mathcal{R}_I$ in terms of its behaviour, functioning, functional properties, dynamics, etc.

The functional modality description is essentially theoretical and depends entirely on the domain knowledge. (Often a model is used to try to solve a problem, difficulty, or inadequacy associated with the domain). The expression of the structural modality is in a more neutral observation language of geometry, structure, and natural language. One consequence of the bimodality of $\mathcal{R}_I$ is that different data types are required when applying empirical validity criteria (Section 4.3.3.4).

This analysis of the intended range of application allows a clear interpretation of the concept of system. $\mathcal{R}_I$ may be regarded as a system if it has certain global structural and functional properties. Structural properties are generally unequivocal, whereas functional properties depend much more on explanatory theories. Thus, if a system S is largely defined structurally (or physically) its ontological status is given. However, if S is defined in the functional modality (e.g. a control system), the ontology of S depends on the validity of the functional theories. If a model M assumes that $\mathcal{R}_I$ is a system S (a functional system) and if M satisfies the appropriate validity criteria, then it may be meaningful to consider $\mathcal{R}_I$ as S. Furthermore, models used in this way (e.g. control system models in biology) provide a legitimate epistemological base for scientific inquiry at the system level (hence the term "model-based systems science").

In established domains, the intended range of application $\mathcal{R}_I$ will act as a fixed reference frame for the model (e.g. in the modelling of physical instruments $\mathcal{R}_I$ is determined structurally by euclidean geometry, and functionally by highly validated classical physical theory),

whereas in rapidly developing domains the formulation and validation of a model may lead to the acceptance of a new frame of reference, constituting a problem shift or change in the specific objectives and possibly a theoretical advance. For this reason it can be seen that it is very important to include information on domain development when expressing the relationship between modelling objectives and appropriate validity criteria (Section 4.3.4).

For model M, the set of specific scientific objectives is denoted by $O_s(M)$. These cannot be classified in the same way as general objectives $\mathcal{O}_s$, since $O_s(M) \rightarrow \mathcal{R}_I$ and, therefore, this would involve a general classification of all empirical phenomena.

In practice, modellers hope to achieve their specific modelling objectives $O_s(M)$ by demonstrating that the model is an accurate representation of $\mathcal{R}_I$, and therefore satisfying at the same time their general objectives $\mathcal{O}_s(M)$. However, if the model does not satisfy $O_s(M)$ it may nevertheless lead heuristically to the satisfaction of $\mathcal{O}_s(M)$. It is possible that, if this heuristic gain is sufficient, the model will be accepted as valid by redefining $O_s(M)$ and $\mathcal{R}_I$.

4.3.1.3.2 Specific utilitarian modelling objectives

In using a model to help in a practical situation, the specific utilitarian modelling objectives are identified with the "real goals" or objectives of the interested parties. These objectives will be of the general kind described in Section 4.3.1.2.2 related to a system of interest; for instance, (a) may be applied to the thyroid assay service in a health care system; or (e) to computer-based education for medical science. Usually, but not always, the specific utilitarian objectives will require that the model represents a certain subsystem or process in the system of interest, i.e. entail specific scientific objectives (these were discussed in Section 4.3.1.3.1). In addition, specific utilitarian objectives always require the specification of the system of interest outside the range of application or scope of the model. Typically, this involves a definition of the "real goals", and the possible courses of action available for change. The course of action is chosen which maximises the satisfaction of the "real goals".

For technical problems (e.g. engineering design) this method works very well, but for more complicated situations (e.g. at a social or economic level) it becomes problematic. Firstly, it may be very difficult to satisfy scientific modelling objectives, and, secondly, the

specification of the system of interest may be incorrect. The latter may follow from a misperception of the "real goals", or from the invalidity of the implicit social theory concerning the means of control, change, and power distribution in the system of interest. Furthermore, utilitarian objectives have a greater variability of change than scientific objectives and, therefore, the assessment of a model's (pragmatic) validity may vary considerably.

These considerations suggest that when models are used primarily for utilitarian objectives the specific scientific objectives (and $\mathcal{R}_T$) as well as the theory implicit in the description of the system of interest should be made very clear. For model M, the set of specific utilitarian objectives is denoted by $O_u(M)$.

4.3.1.4 The relationship between modelling objectives and model type

Within the definition (D1) of a model given in Section 4.1.2 there is a very wide variety of model types, ranging from simple models of observations expressed in natural language to sophisticated mathematical models. The choice of model type is largely determined by the content and stage of development of the particular scientific domain, and this will be reflected in the general (θ_S) and specific (O_s) objectives for the model and the available data types (described in the next section, 4.3.2). To some extent, however, the model type is merely conventional, depending upon the traditions of a research area. A simple classification of model types is given below:

(a) Linguistic model types
  (i) Models based on natural language using a conceptually enriched vocabulary
  (ii) Semiformal models employing symbolic representations (e.g. intuitive set and graph theory, programmatic codes)

(b) Logico-mathematical model types
  (i) Formal logical models
  (ii) Mathematical models, which may be static or dynamic, statistical (black-box) or functional (based on theory, either stochastic or deterministic).

It would be possible in principle to cross-categorise this classification with a range of modelling objectives, but this would be very involved and outside the scope of thesis, and is not necessary for the

theory of model validity. However, an important aspect for model validity is the degree of specification of the model with respect to the modelling objectives, an issue considered in Section 4.3.3 ("Validity criteria"):

(a) Underspecified model type. The model is insufficiently detailed or sophisticated for the modelling objectives (e.g. as a representation of $\mathcal{R}_I$) and available data types.

(b) Overspecified model type. The model is too complex or sophisticated for the modelling objectives and available data types (and stage of domain development).

Occasionally, a model type may be underspecified with respect to the modelling objectives and overspecified with respect to data types (or vice versa), a situation which arises in the use of control system models in biology and medicine.

4.3.2 Theory of data

4.3.2.1 Introduction

Data may be defined as records containing information about empirical phenomena (objects or events). Data vary considerably in the kind and amount of empirical information they contain. These may be classified according to the different empirical representation techniques involved. The theory of data presented here provides a theoretical basis for a simple classification of these different "data types" and introduces various concepts related to data types, such as available and required data types. The theory is not concerned with details of specific data acquisition techniques, but rather with the logico-empirical foundations of data. It has two roles in the theory of model validity: firstly, it is involved in the articulation of the general relationship between modelling objectives and appropriate validity criteria (Section 4.3.4), and, secondly, to aid in structuring the validation of a specific model (or class of models). In the theory of model validity much emphasis is placed on the effect of the scientific domain (specific area of research associated with the model) on model validity and validation, and a theoretical analysis of the nature of data can help in identifying the stage of development of a particular domain.

The theory of data is concerned with the elementary level of theory

that forms the basis of data and demonstrates that all data involve theoretical or linguistic concepts. This implies that data are not unique representations of reality but may change as theories and concepts change. A purely static analysis would suggest that theories determine data; however, a historical analysis of the dynamic development of scientific domains indicates that theory-rich data still retain their realism, or objectivity. At an early stage of domain development, data are based on facts of everyday experience expressed in natural language ("observational data types", Section 4.3.2.2). As the domain develops abstract theories and models data become more theoretically based ("symbolic data types", Section 4.3.2.3).

4.3.2.2 Observational data types

Observational data types are descriptions of empirical phenomena expressed in natural language. As a scientific domain develops the observation language becomes enriched with new concepts and vocabulary (a semantic and syntactic extension). The observation language, or universe of discourse U, forms part of a frame of reference $R = \langle U, L \rangle$ where L is the set of possible statements about the world.

4.3.2.3 Symbolic data types

Symbolic data types are mappings of empirical phenomena into an abstract symbolic space which has certain properties. The elementary empirical operations or relations on the empirical phenomena are represented in the symbolic space, so that symbolic data types are isomorphic (or homomorphic) mappings of aspects of empirical phenomena. Symbolic data types range from intuitively-based representations in a finite space (e.g. graphs of a political ideology, Farbey, Mitchell, Webb, 1979; pattern recognition, Finkelstein, 1975) to formal systems of measurement. In measurement, basic empirical relations of binary ordinality, ternary concatenation, or quaternary ordinality allow mappings into the full ordinal or cardinal number systems. This requires that the basic empirical relations hold universally and is therefore associated with the concept of a universal attribute. In science, (in formal problems) universal attributes are the free variables (parameters and variables) of a mathematical model or theory. Thus in measurement there is a direct link to theory.

Although theory underlies measurement, it is usual to refer to the explicit use of theory as "indirect" measurement. Indirect measurement consists of measuring one or more empirical attributes and performing some physical or mathematical operation on them in order to determine the

value of other attributes. The nature of the operation is dictated by a model or theory. For this purpose, it is necessary that the model or theory should have been previously validated. This raises interesting problems when a model is used for indirect measurement (e.g. parameter estimation) at the same time as it is being validated (this is considered further in Chapter 5).

Symbolic data types are reported alongside observational data types which provide a frame of reference. For instance, the observational data types may identify characteristics of a set of objects or experimental situation whose attributes or results are compactly expressed by symbolic data types.

The logical foundations of measurement (and some forms of finite symbolic data types) are the subject of the theory of measurement. A comprehensive account of the theory of measurement may be found in Krantz et al. (1971) or Leaning (1977).

4.3.2.4 Available and required data types

The specific scientific objectives of a model determine its intended range of application $\mathcal{R}_I$ (Section 4.3.1.3.1). The "available" data types (D_A) for $\mathcal{R}_I$ are the currently available data types from both the structural and functional modalities of $\mathcal{R}_I$. The data types may be observational or symbolic as discussed above, and may also include details of their spatio-temporal limits and resolutions.

In applying empirical validity criteria to a model M of $\mathcal{R}_I$, certain data types are needed to validate its structure and others to validate its behaviour (e.g. in the case of a dynamic mathematical model, the latter will be the values of some or all of the model parameters and the time responses of some or all of the model variables). These will be known as the "required data types" D_M for model M of $\mathcal{R}_I$.

A comparison of the available, D_A, and required, D_M, data types can reveal much concerning the nature of the model M, and the opportunities or difficulties in model validation. This is considered further in Section 4.3.3.4 ("Empirical validity criteria").

4.3.2.5 Data uncertainty

The term "data uncertainty" refers to the general doubt concerning the validity or accuracy of available data, or to the problems associated with data acquisition. Several aspects of data uncertainty are considered below.

4.3.2.5.1 Uncertainty of basic empirical relations in symbolic data types

The basis of symbolic data types (e.g. measurement) is the discrimination of primitive empirical relations (equivalence, order, etc.). In practice there is always a least discernible difference and thus will be manifested as a possible (unknown) error, or uncertainty, on the results of measurement. If the source of the least discernible difference is stochastic then the results of measurements will exhibit statistical error fluctuations, and the statistical theory of errors may be an appropriate tool for their analysis and interpretation.

The uncertainty of the basic empirical relations causes problems in the theory of measurement concerned with demonstrating the logical possibility of meaningful measurement under these conditions. Obviously such measurement is meaningful in practice to scientists, but there is not yet a widely accepted theory of uncertain measurement. (Attempts to develop such a theory include the deterministic algebraic theory of Luce, 1956, and the probabilistic theories of Domotor, 1969, and Leaning, 1977.)

4.3.2.5.2 Uncertainty arising from theoretical inadequacy

Both observational and symbolic data types are theoretically dependent. If the theoretical concepts are new, or have not been validated, then observational and symbolic data will be uncertain (in measurement this is referred to as uncertainty in the concept of an attribute). Symbolic data types, however, depend much more heavily on theories and models. In indirect measurement data uncertainty arises from the uncertainty of the structure or parameters of a mathematical model. This uncertainty can be analysed by "transmitting" expected structural and parametric uncertainties (expressed statistically as probability distributions, or second order statistics) through the model (see, for example, Clifford, 1973).

4.3.2.5.3 Uncertainty arising from data transmission

"Data transmission" refers here to the transfer of data between the site and time of observation or measurement and the final data record. Uncertainty can occur in a great many ways from errors of recording and perception, to problems associated with the frequency of changes in phenomena that are being recorded (for which signal theory is very helpful in analysing).

4.3.2.5.4 Uncertainty in the representation of $\mathcal{R}_I$ (or system)

The available data types D_A from $\mathcal{R}_I$ (intended range of application or system) may not fully describe $\mathcal{R}_I$ and this can lead to uncertainty as to whether D_A provide complete, or at least sufficient, empirical information about $\mathcal{R}_I$. Attempts to resolve this uncertainty are made usually by increasing the spatio-temporal resolution of the data, but it may also be achieved by employing a different theoretical model, thereby affecting the required data types for the functional modality (see also Section 4.3.2.4).

4.3.2.5.5 Data uncertainty associated with experiment

If an experiment on $\mathcal{R}_I$ precisely matches the normally prevailing conditions on $\mathcal{R}_I$ then the data uncertainty is due simply to the sources discussed above. However, experiments are usually performed in a carefully controlled environment, such as a laboratory,and for a relatively small range of experimental conditions. Under these circumstances uncertainty may exist in assuming the generality of the data to non-experimental conditions as well. For instance, do results of in vitro tests on the properties of biological tissues hold for tissues in vivo?, or can functional data based on animal experiments (such as experiments on the neural control of the cardiovascular system in dogs) be used in models of humans?

4.3.2.6 Meaningful operations on data types

A "meaningful operation" on a data type may be defined as one which does not assume that the data type contains more empirical information than it does. In the case of observational data types this requires the logical use of the observation sentences in a deductive sense. The inductive generalisation of an observation sentence may be a valid scientific inference, but it is not a meaningful operation in terms of an observational data type.

A more precise definition of a meaningful operation can be given for symbolic data types: a meaningful operation on a symbolic data type is an admissible transformation which preserves the isomorphism (or homomorphism) of the symbol assignment. The concept of a meaningful operation, or admissible transformation, is very important in model validity. This is because it determines the applicable operations on different symbolic data types (e.g. applicable arithmetic for different measurement types: nominal, ordinal, interval, etc., Leaning, 1977, p. 15) and therefore the correct interpretation in comparing a model with data. For instance,

when data contain stochastic uncertainty, only certain types of statistical measures are meaningful operations for different symbolic data types, as shown in Table 4.1.

Symbolic Data Type	Location	Statistical Measures		
		Dispersion	Association	Significance
Symbolic Data Code	Mode	Information	Transmitted Information	Chi-squared Test
Ordinal Measure-ment	Median	Percentiles	Rank-order Correlation	Sign Test
Internal Measure-ment	Mean	Std. Deviation Av. Deviation	Product-mean Correlation	t-Test F-Test
Ratio Measure-ment	Geometric & Harmonic Mean	% variation		

Table 4.1: Statistical Measures which are Meaningful Operations on various Symbolic Data Types

4.3.3 Validity criteria

4.3.3.1 Introduction

The introductory analysis of this chapter (Section 4.1) showed that there are several distinct meanings to the terms "a valid model" and "model validity", ranging from empirical correspondence to heuristic potential. To each meaning or concept there is a set of tests, means, standards, or rules for judging whether a model is valid with respect to that meaning. For most meanings a model cannot be shown conclusively to be valid (a consequence of the general nature of a model as opposed to the singular nature of empirical data) and the tests will be concerned with judging the extent of the validity of a model. These tests are known as "validity criteria" and are denoted by $V_i \in \mathcal{V}$ (the set of validity criteria). In this section, the set of validity criteria $\mathcal{V}$ is classified and for each class the nature of appropriate tests or rules is explained.

A primary distinction can be made between "internal" validity criteria ($\mathcal{V}_{INT}$) and "external" validity criteria ($\mathcal{V}_{EXT}$). The former

consist of tests on the model without reference to anything outside the model, whereas the latter require reference to aspects external to the model (such as data). $\mathcal{V}_{INT}$ are divided into tests on consistency (Section 4.3.3.2.1) and algorithmic validity (Section 4.3.3.2.2) and, in general, must be completely satisfied before any other aspects are considered. They are therefore prerequisite criteria ("necessary" in philosophical terms). $\mathcal{V}_{EXT}$, on the other hand, depend upon something else and are therefore vulnerable, or contingent. They are divided into representational criteria (Section 4.3.3.3), pragmatic criteria (Section 4.3.3.4), and heuristic criteria (Section 4.3.3.5). The application of the validity criteria to a class of models is determined by considerations of modelling objectives ($\mathcal{O}$), available data types ($\mathcal{D}$), and also the stage of development of the scientific domain. The relationship between $\mathcal{V}$, $\mathcal{O}$ and $\mathcal{D}$ is explained and illustrated in Section 4.3.4 (Table 4.2 in this section provides a compact summary of the validity criteria and associated validity concepts).

4.3.3.2 Internal (or necessary) validity criteria

4.3.3.2.1 Consistency validity criterion, V_{CON}

This criterion requires that the model should contain or entail no logical contradictions. In mathematical models it can be checked by examining for algebraic loops. In formal models (i.e. deductive systems) there are various techniques for proving consistency, although if the model is complex (i.e. a high-order deductive system) it may be possible to prove that consistency is undecidable (Gödel, 1931; see also Section 4.7). For linguistic and semiformal model types (Section 4.3.1.4) it may be difficult to determine consistency completely. The same is true, incidentally, of verifying computer programs with many multi-conditional branching points.

4.3.3.2.2 Algorithmic validity criteria, V_{ALG}

V_{ALG} are variety of tests for checking that the algorithm for solution (analytical) or simulation of the model are correct and lead to accurate solutions. Algorithms for numerical approximation may be checked for stability and assymptotic convergence (e.g. Euler or Runge-Kutta methods for integrating differential equations). The rounding-off errors (associated with the word length for storage of variables in a computer) also should be tested. Simulation models which contain stochastic elements (e.g. pseudo-random binary number generators) should be tested in respect of their statistical properties. (This is referred to as model "verification"

by some authors, see Section 2.3, Fishman and Kiviat, 1968, or Mihram, 1976.)

4.3.3.3 Representational validity criteria, V_{REP}

V_{REP} are divided into empirical (V_{EMP}) and theoretical (V_{THEOR}) criteria and are concerned with testing the extent of the representation of the intended range of application $\mathcal{R}_I$ by model M. They are therefore closely linked to the specific scientific objectives $O_s(M)$ of the model.

4.3.3.3.1 Empirical validity criteria, V_{EMP}

V_{EMP} require that the model should correspond to data available D_A. The most stringent requirement is that the correspondence should hold at all levels from elementary submodels to the overall model structure and behaviour; however, in practice, it may not be necessary to test every aspect of the model empirically. The resolution at which V_{EMP} are applied may be called the "level of validation" (discussed further in Section 4.3.3.3.3).

Some initial tests in V_{EMP} compare the required data types of the model, D_M, with those available, D_A, (the data types include the observational/symbolic distinction, Section 4.3.2, as well as details of spatial and temporal resolution). If $D_M \subseteq D_A$ then the model may receive full empirical validation. This is usually the case for simple models or hypotheses, and the techniques of statistical comparison (hypothesis testing, significance testing, etc.) may play a major role. If $D_M \supset D_A$, the data requirements of the model exceed those currently available, and empirical validation will require an extension of available data (although this is not always possible). In Section 4.4 further consideration is given to the mismatch $D_M \supset D_A$, with some examples.

(V_{EMP} is equivalent to the correspondence notion of truth. Tarski (1930) attempted to formalise the concept of truth as a criterion for deductive systems, but instead he proved that(in sufficiently rich systems) no precise criterion could be formulated which does not lead to an inconsistency, or paradox. He suggested that truth should form a primitive concept in a theory of truth.)

When comparing a model with data, the subtle qualitative characteristics of the data and model should be tested, as well as the comparison of numerical values. Consideration should also be given, of course, to the likely uncertainty on the data, and this can lead to probabilistic assessments of empirical validity (see Chapter 5).

V_{EMP} should be applied as fully as possible over $\mathcal{R}_I$; however, a model is a symbolic generalisation whereas data are records of singular events, and so a model can never be logically proven on the basis of data. The empirical validity can be increasingly confirmed by making ever more severe tests. The emphasis should be on critical tests which aim to invalidate the model, and thereby delimit its valid range of application (denoted by $\mathcal{R}_V$). A decision may be made to accept a model if its $\mathcal{R}_V$ covers areas of $\mathcal{R}_I$ that competing models do not.

4.3.3.3.2 Theoretical validity criteria, V_{THEOR}

The representational validity of a model may also be tested by comparing it with currently accepted theories and models that apply to $\mathcal{R}_I$. In applying V_{THEOR} it is assumed that the theories or models used as a basis have already received validation and established their representational validity (or, at least, they should be the best available representations of $\mathcal{R}_I$). The requirement of V_{THEOR} is that the model should "cohere" with the accepted models or theories. If model M is based on theory T with language L_T then a formal expression of theoretical validity is that M is a valid derivation in L_T from T, in symbols $\vdash_{L_T}: T \rightarrow M$ (or if C_{L_T} denotes the consequence class, $M \in C_{L_T}(T)$). This is in accordance with the concept of a model "on" a theory, and describes correctly the situation where a model is derived from a theory (for instance, by applying the initial conditions and boundary constraints of $\mathcal{R}_I$). If, however, the model has not been derived from the theory, it is not generally possible to determine a decision procedure to decide the validity of M (Gödel, 1931).

In practice, few models are simply derivations from theory, and many involve new hypotheses, assumptions, concepts, etc. which imply that the model is formally richer than the theory. Theoretical validity criteria can still play a large part in model validation, particularly in re-examining assumptions, structure, elementary submodels, etc., but if the theoretical advance is large they will not be appropriate, and other criteria will be used (empirical and heuristic). Strict adherence to the coherence concept of validity would lead to very conservative evolution of models and science.

4.3.3.3.3 The recursive nature of representational validity criteria

Both V_{EMP} and V_{THEOR} can be applied at any level of a model, from elementary submodels to the overall model. Applied at the lowest level in the model they test directly the validity of the basic relationships,

hypotheses, and mechanisms and form a strong (a priori) deductive base for inferring the validity of the overall model. When applied to the overall model (a posteriori) V_{EMP} and V_{THEOR} determine the validity of the aggregate model (e.g. system level) but may also be used for inferring the validity of submodels within the model. Thus the representational validity of the model rests on the validity of its submodels, which in turn can be partly determined from the validity criteria applied to the overall model. In other words, representational validity is a recursive concept.

4.3.3.3.4 Primacy of empirical validity criteria

Ultimately, models represent reality and not other theories or models and therefore the basis of representational validity must be in empirical validity. However, in the long process of evolution of a scientific domain, models and theories develop from an empirical base, having frequent contacts with reality (empirical validity criteria) and become more theoretical yet still embodying knowledge about phenomena. Thus the application of theoretical validity criteria is justified since they rest on an application of empirical validity criteria over a long period of time.

Representational criteria are not the only reasons for accepting a model, and other types of validity criteria may be more important, depending upon the situation.

4.3.3.4 Pragmatic validity criteria, V_{PRAG}

V_{PRAG} are tests of a model in satisfying general and specific utilitarian objectives (Sections 4.3.1.2.1 and 4.3.1.3.2). Utilitarian objectives are fairly straightforward aims, such as the improvement or design of a system. In principle, the application of V_{PRAG} should be simple, involving the definition of measures of effectiveness in the system of interest and then determining whether the objective (e.g. improvement) has been achieved. A large number of utilitarian models may satisfy this scheme, but for others it is problematic. This is because once a model has been used in some practical situation, it modifies that situation in such a way that there is no longer information on how the situation would have developed if the model had not been used. In this case, or where collectively accepted measures of effectiveness are lacking, other tests will be involved in V_{PRAG}. For instance, a model may be critically assessed in terms of the potential benefit it offers, or the understanding that it gives to people involved in the practical situation.

Models used for utilitarian purposes invariably entail specific scientific objectives (i.e. an $\mathcal{R}_I$), and they are therefore subject to the representational criteria V_{EMP} and V_{THEOR}. In particular, such models are frequently required for prediction, and so the predictive validity of models should be tested. (An important point to make here is that predictive validity depends on the knowledge that a model represents the important mechanisms responsible for change, and is not simply a statistical analysis of variances of a regression line.)

4.3.3.5 Heuristic validity criteria, V_{HEUR}

V_{HEUR} are concerned with the assessment of the potential of the model for scientific understanding and discovery, i.e. its role as a heuristic device. They are closely related to the overall aim of science, i.e. as an evolutionary knowledge acquisition process concerning empirical phenomena, and also general modelling objectives (Section 4.3.1.2.1). To some extent V_{HEUR} require that a model satisfy the representational criteria V_{EMP} and V_{THEOR}, but they are more concerned with whether a model will be fruitful or promising for future developments. The type of judgement involved in V_{HEUR} is that exercised by scientists in their day-to-day activity as well as in more important decisions on model validity. Current trends in the philosophy of science are to emphasize the role of good reasoning patterns and heuristics in science as well as the notion of dynamically-evolving scientific domains (see Chapter 3, Section 3.3, or Lakatos, 1970, or Shapere, 1977). Three heuristic validity criteria are outlined below; however, research in this area is at an early stage and there are probably many more:

(i) Expansion of empirical content. If a series of models M_1, M_2, ... in the development of a scientific domain have intended ranges of application $\mathcal{R}_I(M_1)$, $\mathcal{R}_I(M_2)$, then the development is "theoretically progressive" if $\mathcal{R}_I(M_1) \subset \mathcal{R}_I(M_2)$, and "empirically progressive" if $\mathcal{R}_V(M_1) \subset \mathcal{R}_V(M_2)$. If not, then the development is "degenerative" and should be stopped. (Based on Lakatos' theory of research programmes, 1970, see Section 3.3.4).

(ii) Problem shifts and the resolution of anomalies. If a model resolves an outstanding anomaly or problem in the scientific domain then it may be regarded as having heuristic validity, even if it has not yet fully satisfied V_{EMP} or V_{THEOR} (based originally on Einstein, 1905). Associated with a progressive series M_1, M_2 ... there will be a series of problem shifts to P_1, P_2 ... and these may indicate better and more fruitful

directions for research.

(iii) Better "understanding". In some way a model may provide a better understanding of $\mathcal{R}_I$ than is previously available. This may be because it embodies a picture or hypothetical mechanism underlying the phenomena (Harré's "statement-picture complex", 1972; see Section 3.3.4), which will eventually receive empirical validation. A long unresolved problem in the philosophy of science is the nature of scientific explanation, and a clear solution would help in defining what is meant by a better explanation or understanding.

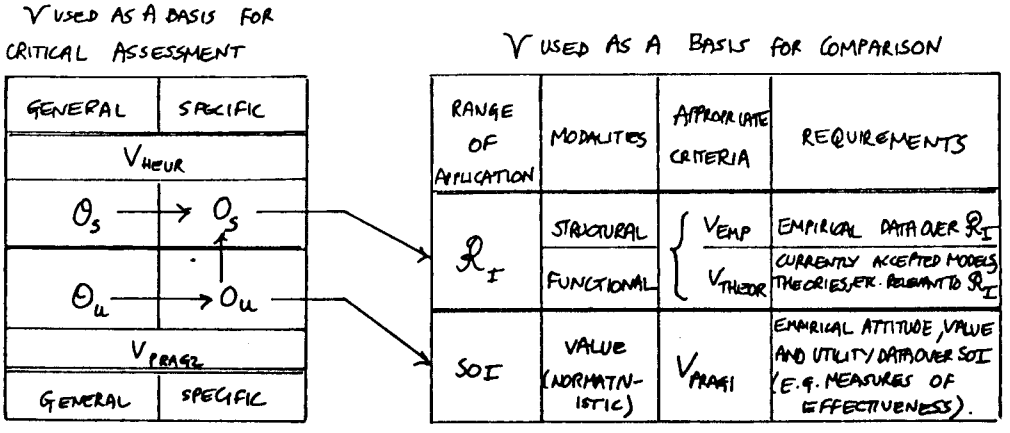
4.3.4 The relationship between validity criteria, modelling objectives and data types

The previous three sections have given an analysis of the range and types of modelling objectives (Θ), a theory of data types ($\mathfrak{D}$), and a classification and explanation of the various validity criteria ($\mathcal{V}$). In each, much reference was made to the other two, and also to what is called the "stage of development of the scientific domain" which refers to the nature (e.g. elementary, advanced, etc.) and content of the specific body of knowledge (including data, theories, models, etc.) associated with the field of research. In this section an attempt is made to explain the relationship between Θ, $\mathfrak{D}$, $\mathcal{V}$ and the stage of domain development on a fairly general level.

Table 4.2 summarises the various validity criteria, their terms of reference and associated validity concepts. In general, the internal criteria, V_{CON} and V_{ALG}, must always be satisfied before any use of a model, in validation or otherwise, and will not be considered further (for statistical models, the application of V_{ALG} may be non-trivial, however). The next step is to delimit the applicability of the various external validity criteria (V_{EMP}, V_{THEOR}, V_{PRAG}, V_{HEUR}) to the types of modelling objective (Section 4.3.1) purely in formal terms (i.e. based on the definitions). Firstly, a distinction can be made between the use of validity criteria as a basis for comparison of a model with data or other models or theories, and the use of validity criteria as a basis for critical or rational assessment. In the former it is the relationship between a model and its specific range of application that is important, whereas in the latter a wider, longer-term view is taken in relation to the specific and general modelling objectives. Into the former category the scientific representational criteria, V_{EMP} and V_{THEOR}, and pragmatic validity criteria, V_{PRAG1}, may be placed (V_{PRAG1} is the set of pragmatic

VALIDITY CRITERIA	ABBREVIATION V	TERMS OF REFERENCE	ASSOCIATED VALIDITY CONCEPTS	
CONSISTENCY	V_{CON}	INTERNAL (NECESSARY)	LOGICAL (AND STATISTICAL) PRE-REQUISITES	
ALGORITHMIC	V_{ALG}			
EMPIRICAL	V_{EMP}	EXTERNAL (CONTINGENT)	REPRESENTATIONAL VALIDITY	EMPIRICAL CORRESPONDENCE
THEORETICAL	V_{THEOR}			THEORETICAL COHERENCE
PRAGMATIC	V_{PRAG}		UTILITARIAN VALUE	
HEURISTIC	V_{HEUR}		SCIENTIFIC POTENTIAL	

TABLE 4.2. VALIDITY CRITERIA, THEIR TERMS OF REFERENCE, AND ASSOCIATED VALIDITY CONCEPTS.



KEY:

Θ_S GENERAL SCIENTIFIC OBJECTIVES

Θ_U GENERAL UTILITARIAN OBJECTIVES

O_S SPECIFIC SCIENTIFIC OBJECTIVES

O_U SPECIFIC UTILITARIAN OBJECTIVES

$\mathcal{R}_I$ INTENDED RANGE OF APPLICATION

SOI WIDER SYSTEM OF INTEREST

FIGURE 4.1. THE GENERAL RELATIONSHIP BETWEEN MODELLING OBJECTIVES, VALIDITY CRITERIA, AND EMPIRICAL DATA.

criteria associated with measures of effectiveness of a model directly related to the system of interest, SOI). The second category contains the scientific heuristic validity criteria, V_{HEUR}, and the more general pragmatic criteria, V_{PRAG2}.

Figure 4.1 illustrates the general relationship between modelling objectives, validity criteria, and empirical data based on the above argument. The arrows on the diagram are to be read as "depends on", "leads to", or "contains"; thus $O_u \rightarrow O_s$ means "the specific utilitarian modelling objectives may contain specific scientific objectives". V_{PRAG2} refers to the set of general, critical pragmatic criteria at the level of overall assessment. Empirical data enter into the comparison of a model over its intended range of application ($\mathcal{R}_I$) in V_{EMP}, and also in the assessment of the effectiveness of a model in the SOI, i.e. in V_{PRAG1}. Figure 4.1 may therefore be regarded as a general form of the map $\mathcal{O} \times \mathcal{D} \rightarrow \mathcal{V}$ and forms the central core of the theory of model validity.

The arrows on Figure 4.1 indicate that the satisfaction of V_{EMP}, V_{THEOR}, and/or V_{PRAG1} may lead to a satisfaction of the modelling objectives, both specific and general. V_{HEUR} and V_{PRAG2} can be interpreted as criteria that forecast that a model will eventually satisfy V_{EMP}, V_{THEOR}, or V_{PRAG1}.

In Figure 4.1, there are several points at which the domain knowledge plays a significant role. The specific modelling objectives (O_s and O_u) are very largely determined and constrained by the domain (e.g. by its content and typical research directions). The articulation of the range of application of the model, $\mathcal{R}_I$ or SOI, requires a great deal of background data and theory (unless the model is completely innovative). The available data (of $\mathcal{R}_I$ and SOI) and currently accepted theories or models required for comparison are closely inter-related and also related to the stage of development of the domain.

The general relationship represented in Figure 4.1 is effectively a general theory underlying the validity of a model. However, in practice, certain validity criteria may appear to be more important, and indeed some will be. The relative importance of validity criteria will depend upon $\mathcal{O}$, $\mathcal{D}$ and the stage of domain development. Furthermore, they may be embedded in the process of model formulation (e.g. from problem perception, to conceptual modelling, to the complete model) so that model validity is a consequence of the modelling methodology. Under these

circumstances the explicit stage of "model validation" in the overall methodology may be simply an empirically refining or tuning process. In Section 4.4 the relationship between the theory of model validity and the overall modelling process is examined in more detail and some specific modelling situations are considered.

The theory can also be used as an analytical framework which can provide a precise analysis of model validity under particular modelling objectives, data types and domain development. For instance, it can structure a validation programme of a complex model (see the first case study on the validation of a dynamic mathematical model of the human cardiovascular system, Chapter 6). Alternatively, it can be used to devise appropriate validation methodologies for classes of models in specific research areas (see Section 4.5, and Chapter 5).

To conclude this section, an important point should be reiterated. The primary aim of models (and science, of course) is to represent and explain real phenomena. Therefore the empirical validity criteria V_{EMP} have a primacy over all other criteria. However, this does not mean that at any one time V_{EMP} will be the most important. It means that the application of V_{EMP} over an extended period of time, from the earliest stages of a scientific domain on, through a developing series of models and theories and data types, forms the epistemological basis of science.

4.4 Implications of the Theory of Model Validity on the Overall Modelling Process

In this section an analysis is made of the conventional "model" (or paradigm) of the overall modelling process using the framework of the theory of model validity (Section 4.4.1). In addition, the problems for model validity and validation that are involved in the use of models in multidisciplinary ("multi-domain") research are considered (Section 4.4.2). The conclusions of Sections 4.4.1 and 4.4.2 suggest that the theory of model validity should form part of research into a wider theory of models, or modelling (Section 4.4.3).

4.4.1 Analysis of the conventional model of the overall modelling process

The conventional model of the overall modelling process, or methodology, has three distinct phases. Firstly, there is the statement of the problem to be considered and the consequent modelling objectives. Secondly,

KEY:

→ ≡ METHODOLOGICAL STEP

⟹ ≡ INFORMATION TRANSFER

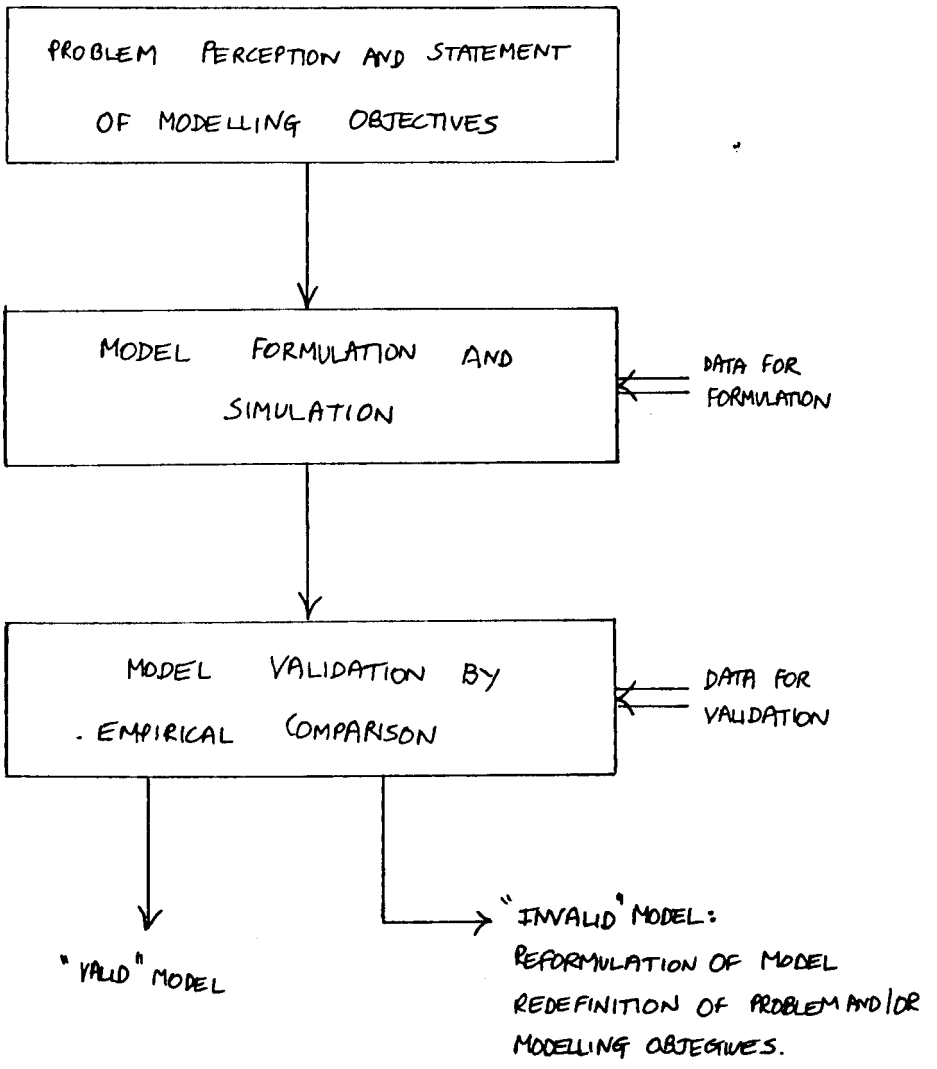


FIGURE 4.2. THE CONVENTIONAL MODEL OF THE MODELLING PROCESS.

the model is formulated and simulated, and, thirdly, the model is validated by comparison with empirical data. Although the second phase is often subdivided, the model represents the essential features of many accounts of the modelling process (e.g. Mihram, 1976; M'Pherson, 1980) and is often quoted as exemplifying "The Scientific Method". Figure 4.2 illustrates the conventional model. If a model passed the empirical comparison test, then it is valid; if not, then the model may have to be modified, or reformulated, or even the problem may have to be redefined.

The main feature of interest here is that model validation is regarded as a distinct phase which follows model formulation and consists of empirical comparisons (i.e. V_{EMP}) which are decisive in determining the validity of the model. For simple models a few iterations of the formulation-validation loop may produce an acceptable model, but for large complex models the conventional model implies that it should be completely reformulated each iteration, an obviously impossible task. In practice the formulation stage of modelling involves many decisions associated with the choice of structure and form of the model, the checking of elementary submodels, assumptions, etc., which implies that validity criteria are being applied implicitly the whole time. In addition, empirical criteria are applied to the individual submodels long before the overall model is complete. The conventional model, therefore, is a coarse picture of the modelling process which wrenches apart the subtle interplay between formulation techniques and validity criteria thereby reducing both to mere mechanical procedures.

Figure 4.3 shows an extended model of the modelling process based on the theory of model validity. The implicit validity criteria in the model formulation process allow a complex iteration between the different stages of formulation that are listed, which depends on how well the specific modelling problem is going (n.b. consequently the order in which they are listed is not important). Included in Figure 4.3 are the external sources of information which affect model formulation and validation. These sources are the background domain and new data from experimental research. It is apparent from Figure 4.3 that model validation is actually "distributed" through the entire process of modelling and therefore the separation of the final empirical validation is largely arbitrary. However, one interpretation is that by the time final empirical validation has started there is already considerable confidence in the validity of the model, and the final stage is simply a calibration,

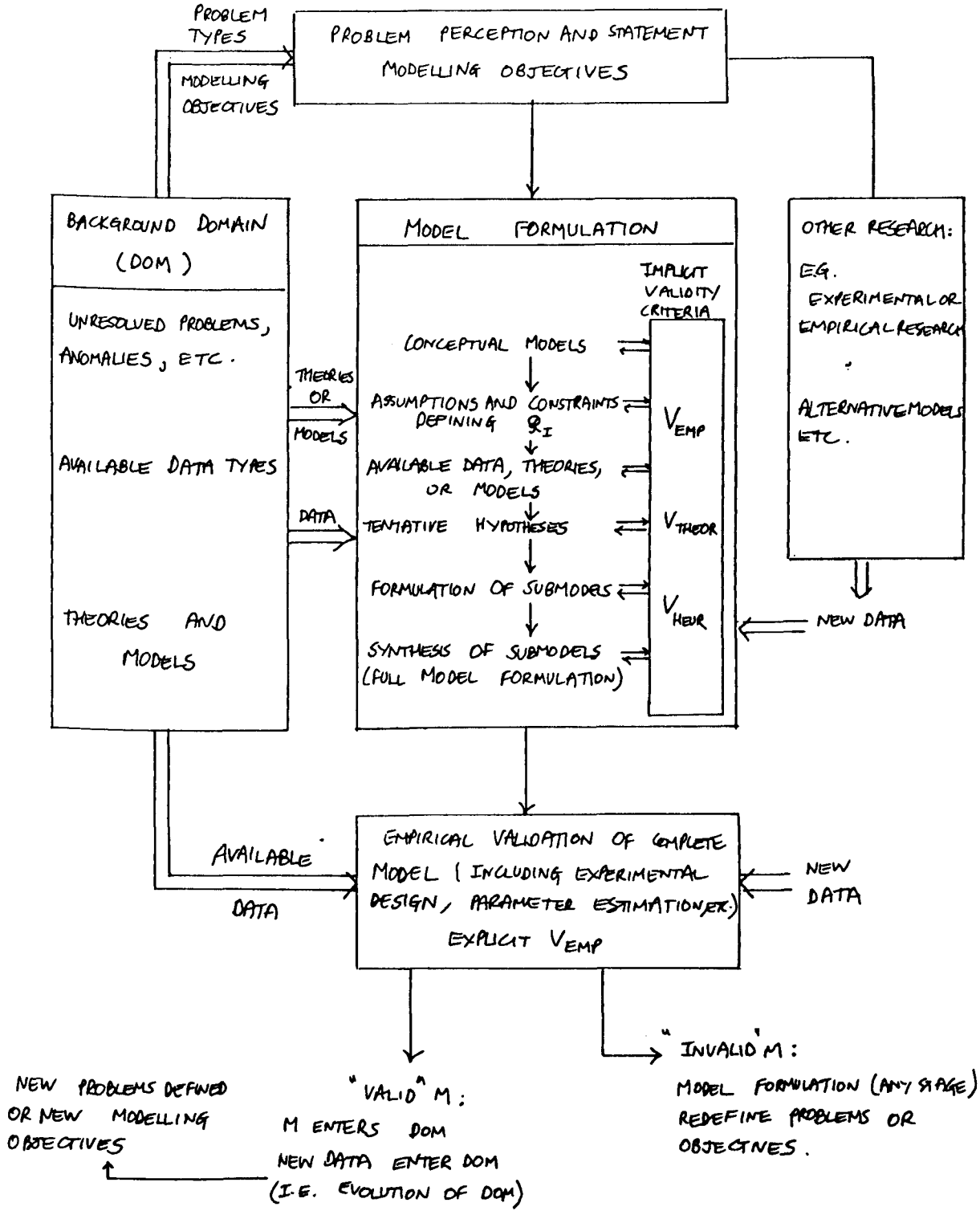


FIGURE 4.3. AN EXTENDED MODEL OF THE MODELLING PROCESS BASED ON THE THEORY OF MODEL VALIDITY.

tuning, or parameter estimation stage. In other words, the validity is established automatically because of the implicit validity criteria in the formulation stage.

This interpretation is in accordance with the situation in the areas of modelling on which the conventional model (Figure 4.2) has been based. These areas are typically technical application areas (such as control engineering) in which models are derived from well-established physical theories and concepts and data are widely available from the intended range of application. The fact that the important validity criteria exist only at an implicit level in the conventional model means that it is misleading when applied to areas (domains) in which theories or models are in a formative stage and there are limited available data. Under these circumstances it is essential that the validity criteria involved during formulation are made explicit. Nevertheless, the need for an extensive validation assessment (independent if possible) of all aspects of model validity once the model is complete is very important.

4.4.2 Multi-domain research

Many models are the result of research in two or more scientific domains, and since the stages of development (theoretical sophistication, available data types, and content) may be very different, model validation can be problematic. This will be illustrated in this section with reference to systems models in biology. The two domains involved are those of biology (DOM_{BIO}) and of systems modelling (DOM_{SYS}), (DOM_{SYS} includes systems theory, cybernetics, control theory, etc. and successful applications).

The intended range of application $\mathcal{R}_I$ of a biological model is defined on the biological domain, whereas the model type (for the specific objectives) is largely determined by DOM_{SYS}. The information on the structural modality of $\mathcal{R}_I$ in DOM_{BIO} (e.g. anatomy) will far exceed the capability of the model, but the functional theories (understanding) of DOM_{BIO} are rarely good enough to describe the functional modality of $\mathcal{R}_I$. Therefore the model provides a theoretical innovation which is usually based on systems control theories. The situation can be summed up by saying that, for the structural modality, the available data types (D_A) exceed those required by the model (D_M), in symbols $D_A \supset D_M$, whereas for the functional modality it is the reverse, $D_M \supset D_A$.

If the main use of the model is in the representation of structure, then it may be fully validated empirically. On the other hand, if the model is intended to explain behaviour it will be very difficult to validate empirically (in full) since there is an insufficient range of data types ($D_M \supset D_A$). (For instance, it is very difficult to obtain continuous measurements of many biological variables (e.g. plasma sodium concentrations) at one site, let alone with high spatial resolution.) Initially, in such applications, the role of the model will be largely as an "analogical construct" in supplying potential increased understanding to DOM_{BIO}, and it will be assessed mainly by heuristic criteria, V_{HEUR}. Eventually, as new techniques of measurement develop, the model can be validated using empirical criteria, V_{EMP}, and the model may be accepted as an adequate explanation of $\mathcal{R}_I$. At the same time, a new domain is established ("biological modelling") which overlaps with DOM_{BIO} and DOM_{SYS} and has its individual body of knowledge as well (it can be regarded as a bridge between DOM_{BIO} and DOM_{SYS}). In general, however, it is probably advisable to limit the mismatch between D_M and D_A in order to allow more complete empirical validation and thereby avoid straining the credulity of biologists too far.

Similar considerations can be applied to questions of validity of systems models in the social sciences (e.g. the Bowers-Mitchell-Webb model of bicommunal conflict, 1979) and this topic is raised in Chapter 9.

4.4.3 The theory of model validity as part of a theory of models

The previous two sections have demonstrated that the concepts of model validity play an intimate role in the ongoing process of modelling and scientific development, and therefore that the theory of model validity should form part of a theory of models or modelling. Some details of such a theory have been outlined, but a full development requires extensive research and is beyond the scope of this thesis.

4.5 The Theory of Model Validity as a Generator of Validation Methodologies

Although the relationship between modelling objectives, validity criteria and empirical data is expressed only at the general level in the theory of model validity (see Figure 4.1), the theory may be applied to a particular modelling situation in order to structure the programme of validation. If the programme applies to a class of models (e.g. the class

of control system models in physiology) it may be regarded as a validation methodology. In Section 4.5.1 it is shown how the theory of model validity may be used to generate validation methodologies, and some illustrations are given in Section 4.5.2.

4.5.1 An outline of the method

The method of generating a validation methodology for a class $\mathcal{M}$ of models begins by considering the modelling objectives of $\mathcal{M}$, and the associated problems and heuristic requirements (such as explanations of a certain type). The next step is to consider the intended range of application of the model ($\mathcal{R}_I$ or SOI), and the information available (in the form of theories, models, data, etc. in the scientific domain). For scientific objectives, if $\mathcal{R}_I$ is well-covered by accepted theories and available data types then the first criteria to apply are V_{THEOR} and if these are satisfied, or unresolved, to pass on to V_{EMP}. On the other hand, if there is scant information on $\mathcal{R}_I$, then V_{HEUR} should be used and new data types should be devised to apply V_{EMP}. Specific utilitarian objectives determine a wider system of interest SOI, as well as a scientific range of application. If $\mathcal{M}$ is to be used for its predictive validity over $\mathcal{R}_I$ this should be checked by V_{EMP} (although V_{THEOR} and V_{HEUR} may be very important in demonstrating the validity of an "underlying mechanism" responsible for the predictions). If data are available from the SOI (e.g. utility functions, measures of effectiveness, etc.) then by using V_{PRAG1}, it is possible to determine if $O_u(\mathcal{M})$ are satisfied.

4.5.2 Illustrations

(a) New research area with very little theoretical development

Typical modelling objectives might be the construction and validation of simple models or hypotheses. Methodology: V_{HEUR} in defining and choice of models, V_{EMP} in validating models (requiring acquisition of empirical data, introduction of new data types). An example is testing of hypotheses in biology.

(b) Established research area with good theoretical base and widely available data

The intended range of application $\mathcal{R}_I$ can be clearly defined. Methodology: V_{THEOR} (and perhaps V_{EMP}), implicit in model derivation, V_{EMP} used for determining parameter values, final calibration, etc. Probably no need to apply V_{HEUR}. If used for utilitarian objectives,

predictive accuracy can be determined by V_{EMP}, and finally the utilitarian effectiveness by V_{PRAG1}. An example is the modelling of technical production processes, in which the final validation stage often consists in a joint parameter estimation-system optimisation algorithm, thereby satisfying V_{EMP} (calibration type) and V_{PRAG1} at the same time.

(c) Analogical construct models

An analogical construct model, or paramorph, is one which has been developed, or whose type comes from a different research area (domain), and is used because it provides structure and potential explanation to the new domain. Methodology: the modelling objectives are primarily for insight and scientific development and initially V_{HEUR} are appropriate. Eventually, however, the model will have to satisfy the representational criteria if it is to become more than a mere analogy. If the model is to be used for utilitarian objectives then it may initially be assessed in terms of the critical basis it provides for satisfying them, i.e. V_{PRAG2}. An example is the use of organismic models in the "real-world" problem solving techniques of "soft-systems" methodologies (Checkland, 1972).

4.6 Testing the Theory

There are three distinct ways in which the theory can be tested: by critical assessment, in historical studies of actual scientific practice, and by testing it out in a practical application. These mirror the three bases from which the theory was developed (reviews of the scientific literature of model validation and the philosophy of science, and work on the validation of mathematical models in biology and medicine). In some ways these tests are rather like applying the theory to itself with critical assessment corresponding to V_{HEUR} and V_{THEOR}, historical studies with V_{EMP}, and testing in a practical application with V_{PRAG}.

4.6.1 Critical assessment

This involves the logical and theoretical analysis of the concepts of validity in the theory and their inter-relations, as well as an examination of the heuristic potential (or explanatory power) of the theory.

4.6.2 Historical studies of actual scientific practice

The use of models in a contemporary or historical developing scientific domain can be studied and it can be determined whether the operative concepts of model validity in such circumstances can be mapped into the theory of model validity.

4.6.3 Testing by practical application

The theory may be tested by examining its ability to structure the validation of models in problematic areas. The case studies (Chapters 6, 7 and 8) are concerned with the validation of three mathematical models in biology.

4.7 Philosophical Significance of the Theory

Although the theory of model validity is intended for scientific application, it contains many philosophical terms and is in some ways a philosophical theory. In this section, some of the philosophical issues in the theory are discussed.

4.7.1 Epistemological basis

The epistemological basis of the theory rests in the empirical criteria which have a primacy over other criteria (see Section 4.3.3.3.4), and whose applications are "spread-out" over the dynamic development of the domain associated with a model, rather than a simple static correspondence of model to facts. It is therefore a realist epistemology (Suppe, 1977).

4.7.2 Comparison with philosophies of science

The four external validity criteria in the theory are:

(i) Empirical correspondence, V_{EMP}

(ii) Theoretical coherence, V_{THEOR}

(iii) Pragmatic value, V_{PRAG}

(iv) Heuristic potential, V_{HEUR}

The theory offers an explanation of the relation between these criteria and between other factors (such as modelling objectives) whilst

stressing the long-term primacy of V_{EMP}. The theory can be compared with different philosophies of science by mapping their views on theory confirmation or acceptance into the set of external validity criteria. The results of this are shown in Table 4.3. The philosophy of science closest

Philosophy of Science	Validity Criteria	Comments
Positivism, Logical Empiricism Early Popper(1935)	V_{EMP} only	In essence, the correspondence theory of truth. Based on the principle of verification (Section 3.2)
Scientific Literature of Model Validation	V_{EMP}, V_{PRAG}	Influence of positivist philosophy on scientific thought. Pragmatic meaning based on utilitarian applications of models. (Chapter 2)
Weltanschauungen Analyses (Kuhn, 1962; Feyerabend, 1975)	V_{THEOR}, (V_{PRAG})	Coherence with theories rather than facts. "Theory-laden facts". Acceptance of model fits in with aims of paradigm (V_{PRAG}). (Section 3.3.2)
Later Popper (1962, 1979)	V_{EMP}, V_{HEUR}	Critical empirical testing provides objective epistemology. Heuristic associated with problem shifts. (Section 3.3.4)
Lakatos' Theory of Research Programmes (1970)	V_{HEUR}, V_{EMP}	Progressive problem shifts (positive heuristic) and empirical testing. (Section 3.3.4)
Historical Realism (e.g. Shapere, 1977)	V_{HEUR}, V_{THEOR}, V_{EMP}	Acceptance of models based on "good reasoning patterns" (heuristics) as well as theoretical coherency and empirical tests. Type of criteria is related to the stage of domain development. (Section 3.3.4)

Table 4.3: Different Philosophies of Science Compared on the Basis of Validity Criteria Implicit in their Theories of Model Acceptance

to the theory of model validity is that of historical realism, although this omits V_{PRAG}. Kuhn's concept of a paradigm (1962) is intuitively appealing to many scientists, yet the implied validity criteria are simply those of theoretical coherency (coherency with the paradigm). This leads to a subjectivist epistemology in which there are no rational means for choosing between competing theories. It is very important to keep the empirical element (i.e. V_{EMP}) in the philosophy of science in order to maintain the objectivity or realism of science.

One aspect of the theory of model validity, which is absent in most philosophies of science, is a consideration of pragmatic validity criteria. Although pragmatic validity does not affect representational validity over $\mathcal{R}_I$ (or the objectivity of scientific knowledge) it can have a large influence on the direction of research in developing models for practical utilitarian purposes. Eventually, this pragmatic "pull" will be manifested in the content of the scientific domain. Thus the concept of pragmatic validity appears to be very important if the effect of socio-technical factors on the development of science is to be understood.

4.7.3 Paradoxes and tautologies in the definition of a valid model

In Section 4.12 a valid model was defined as:

"D2. A valid model is one which satisfies the purposes (or objectives) for which it is intended", or a model is valid iff it satisfies its objectives.

Now suppose that the objectives of model M include the admirable objective O_1: to show that M is valid. Then, from the definition, it follows that M is valid iff M is valid, i.e. a tautology. However, if the objectives include the braver objective O_2: to show that M is invalid, it then follows that M is valid iff it is invalid, i.e. a paradox. This exposes a weakness in the definition of D2. The paradox can be resolved by dividing the objectives into two kinds: "modelling objectives", concerned with the general and specific objectives and intended range of application, and "validation objectives" which may express a desire to validate M. D2 is rephrased as:

D2': A model is valid iff it satisfies its modelling objectives.

4.8 Conclusions

In this chapter, a theory of model validity has been presented which explicates the various concepts of model validity as a set of validity criteria and explains how they are related to modelling objectives, the nature of available data, and background scientific knowledge. The four types of external validity criteria are: empirical correspondence, theoretical coherence, heuristic potential, and pragmatic value. In the theory, modelling objectives are classified into general or specific, and scientific or utilitarian types, and there is also a theory of data. The

underlying motivation for the theory of model validity is the thesis that a good understanding of the nature of model validity is required in order to validate models and to develop validation methodologies. This understanding is further enhanced if the theory is regarded as part of an overall theory of modelling.

As well as providing insight into the nature of model validity, the theory may also be used to devise appropriate methodologies for the validation of models in particular research areas. In the next chapter, a range of four different methodologies suitable for a wide range of modelling problems will be presented, together with an extensive methodology for the systematic comparison of a model with empirical data. The application of the conceptual framework of the theory and some of these methodologies to three biological models is made in Chapters 6, 7 and 8, and to general aspects of the validity and validation of models in the social sciences in Chapter 9.

CHAPTER 5

METHODOLOGIES AND TECHNIQUES FOR MODEL VALIDATION

5.1 Introduction

The theory of model validity developed in Chapter 4 defines the various concepts of model validity and relates them generally to modelling objectives and the nature of available empirical data types (Section 4.3.4). Together with other aspects, the modelling objectives and available data types characterise the stage of development of a particular research area, or scientific domain, and this affects the concepts of validity that are considered important or are operative, as well as the approach towards model validation. In addition to providing a general analytical framework for the critical appraisal or historical analysis of modelling methodologies, the theory of model validity may also be used to devise or generate appropriate validation methodologies for specific research areas (domains). It does this by considering the modelling objectives, available data types, and other indications of domain development (such as theoretical sophistication) and determining the applicable validity criteria and their relative importance (Section 4.5). A validation methodology can be constructed which is the systematic application of these validity criteria to a model. The philosophical basis of the theory is that of historical realism (Section 4.7) which means that models are regarded as representing knowledge and understanding of objects and phenomena that have an objective existence. It follows, therefore, that empirical validity criteria have a primacy over all other criteria (i.e. they are the epistemological base). However, this is on a long-term basis, and at a particular stage in the development of a domain, other criteria may be more important (Section 4.7.1).

In this chapter, a variety of validation methodologies applicable to a range of very different modelling situations are described. Their basis is the theory of model validity, and they make use of the concepts and symbolism of the theory (for an outline, see Table 4.2 and Figure 4.1). Each methodology is presented in an outline flow diagram form, where each block corresponds to a different stage in validation (e.g. the application of a different type of validity criteria) and the arrows indicate the methodological steps. A short discussion on the methodological stages and suggestions for appropriate techniques accompany each diagram.

Two important points must be made on the interpretation of the methodologies. Firstly, each stage is formally distinct with respect to the theory of model validity, but, in practice, stages may be reversed or combined depending on the situation. Secondly,these methodologies do not necessarily form a last step in the overall modelling methodology (Figure 4.2, Section 4.4), but may be intimately embedded in the overall process of model formulation (Figure 4.3). Consequently, a failure to satisfy criteria at a certain stage in the validation methodology may result in a step back to an earlier stage in the model formulation process. These methodological "reverse steps" are highly important to the recursive nature of modelling, and are recognised in these methodologies, but for clarity they are not shown in the diagrams since they may go back to any earlier stages.

The first methodology presented (Section 5.2) is for the comparison of model and data based on the use of features and includes statistical comparisons ("α-methodology"). This methodology is not applied directly but may form the basis, or part, of the full validation methodologies. (Since it proposes a wide variety of techniques, this section is also much longer than the presentations of the other methodologies which are compact.) The second methodology (Section 5.3) is an empirical validation methodology ("β-methodology") which relies on a fairly advanced stage of domain development. The third methodology is based on theoretical as well as empirical criteria (Section 5.4) and is intended for areas where there are reasonable theories and data but the model introduces theoretical development ("γ-methodology"). This methodology is suited to the validation of mathematical models in biology. The fourth methodology (Section 5.5) is for models used for utilitarian objectives ("δ-methodology") and the fifth methodology is a proposal for the validation of models that are innovative ("ε-methodology",Section 5.6).

5.2 A Methodology for Model-Data Comparisons in Empirical Validation (α-Methodology)

The α-methodology is a general methodology for model-data comparisons (i.e. application of empirical criteria) which is intended to extract as much critical empirical information from the data as possible. The four stages of the methodology are shown in Figure 5.1. The entry point should always be the first stage (qualitative comparisons). An emphasis is placed on the use of "features" in empirical validation since they can be

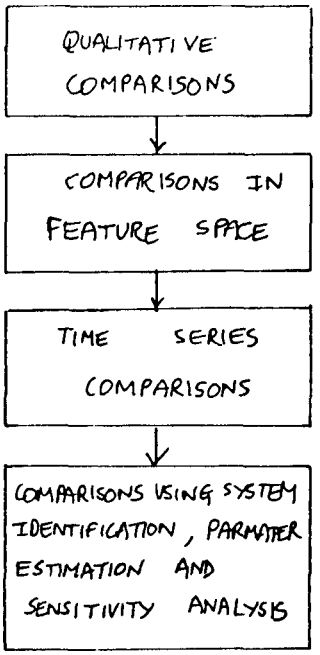


FIGURE 5.1. THE α- METHODOLOGY FOR MODEL-DATA COMPARISONS IN EMPIRICAL VALIDATION.

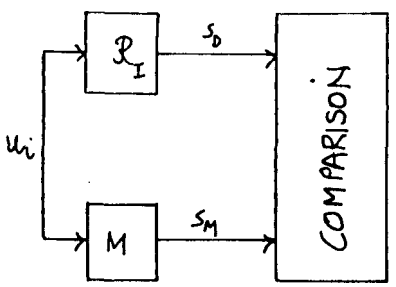


FIGURE 5.2a. COMPARISON OF MODEL WITH QUALITATIVE DATA

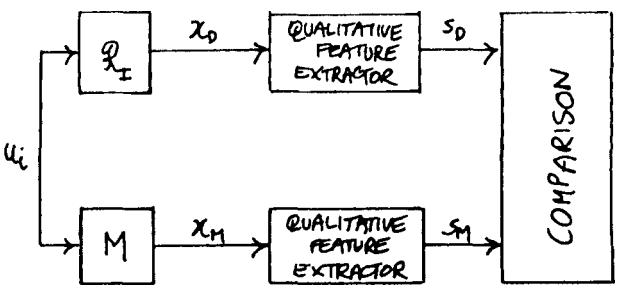


FIGURE 5.2b COMPARISON OF MODEL WITH QUALITATIVE FEATURES EXTRACTED FROM NUMERICAL DATA

selected in order to represent the most important features of the data available (from the intended range of application). The first stage of the methodology applies to any type of model, linguistic or logico-mathematical (as defined in Section 4.3.1.4),whereas the subsequent stages apply to mathematical models. In all stages the emphasis should be on critical comparisons in order to delimit the valid range of application $\mathcal{R}_V$ of the model.

5.2.1 Qualitative comparisons

The intended range of application ($\mathcal{R}_I$) of the model is determined by the specific scientific modelling objectives $O_s(M)$ and has two modalities: physical or structural, and functional (Section 4.1.3.1). Comparisons may be made with data from either of these modalities, and henceforth in the α-methodology the distinction will not be made. Qualitative comparisons may be made with two different data types: firstly, comparisons may be made with observational data types which are linguistic reports of observations and therefore qualitative in nature, and, secondly, comparisons may be made with qualitative features extracted from symbolic or numerical data types. A feature is defined as "an important characteristic or attribute of the data, pertaining to $\mathcal{R}_I$ ". For example, qualitative data or features may range from "the diaphragm has concentric corrugations" to "blood concentrations of carbon dioxide exhibit oscillations shortly after the onset of exercise" or "the system suddenly loses stability".

Qualitative comparisons are extremely important in the empirical validation of all models, and the only means of empirical validation for linguistic-type models. A discussion of the theoretical implications of qualitative comparisons for mathematical models is given at the end of this section. The two kinds of qualitative comparison are shown in Figures 5.2a and 5.2b, where u_i represent the conditions of observation or measurement and are assumed to be applied in the same way to model M; s_D and x_D are the qualitative and numerical data from $\mathcal{R}_I$ (or system); s_M and x_M are qualitative and quantitative aspects of M (form or behaviour). (Sometimes u_i may be regarded as an input and s_D, s_M, x_D, x_M as outputs, but this interpretation does not always apply.) The criterion for qualitative validity is simply:

M is <u>qualitatively valid</u> iff $s_D \sim s_M$ for $\forall u_i \in \mathcal{R}_I$ ..... (5.1)

In other words, if the model reproduces all qualitative aspects or features for all conditions within the intended range of application, it is qualitatively valid.

The qualitative information may be expressed in a formal or informal code which may be manually or automatically extracted from the numerical data. (This can be regarded as an irreversible mapping of a quantitative data type into a relationally-simpler symbolic date type. The mapping should be a "meaningful operation" in terms of the empirical relational system defining the quantitative data type; see Section 4.3.2.6.) Manual codes range from forms of implicit subjective pattern recognition to shorthand pictogrammatic notations (an example is shown in Figure 5.3). An example of an automatic formal code is the computerised sampling and symbolic encoding of a time-varying, or spatially-varying, waveform (for an example and application to validation, see Leaning, 1979). In validating a model of the human respiratory control system, Bali et al. (1976) used feature space pattern recognition (based on a linear classifier) to classify human and model responses to changes of carbon dioxide in inspired air (discussed further in Chapter 8).

Codes such as that shown in Figure 5.3 can be very useful in comparing complex dynamic waveforms such as occur in biological variables (e.g. in the validation of a model of the human cardiovascular system, Chapter 6).

Although qualitative comparisons are only the first stage in the comparison of a mathematical model and data, they are an extremely important prerequisite. This is because there is a close relationship between the occurrence of qualitative features and the structural and functional form of a model. Such a relationship was demonstrated in a study of low-order linear systems (Leaning et al., 1979; see also Batchelor, 1978, on the classification of pole maps in the complex plane). For most nonlinear dynamic systems the qualitative theory of differential equations is the only way of analytically characterising behaviour (Lefschetz, 1963). The sudden qualitative changes in a dynamical system can be associated with the loss of structural stability; René Thom's catastrophe theory arose from the application of the algebraic theory of topology to the problems of qualitative dynamics (Thom, 1975). Nonlinear systems in particular exhibit a wide range of qualitative behaviours, and these should be reproduced in the model before other tests are applied (e.g. prior to parameter estimation, Section 5.2.4, or Mehra, 1980). Once a model has demonstrated the correct qualitative features, the sensitivity

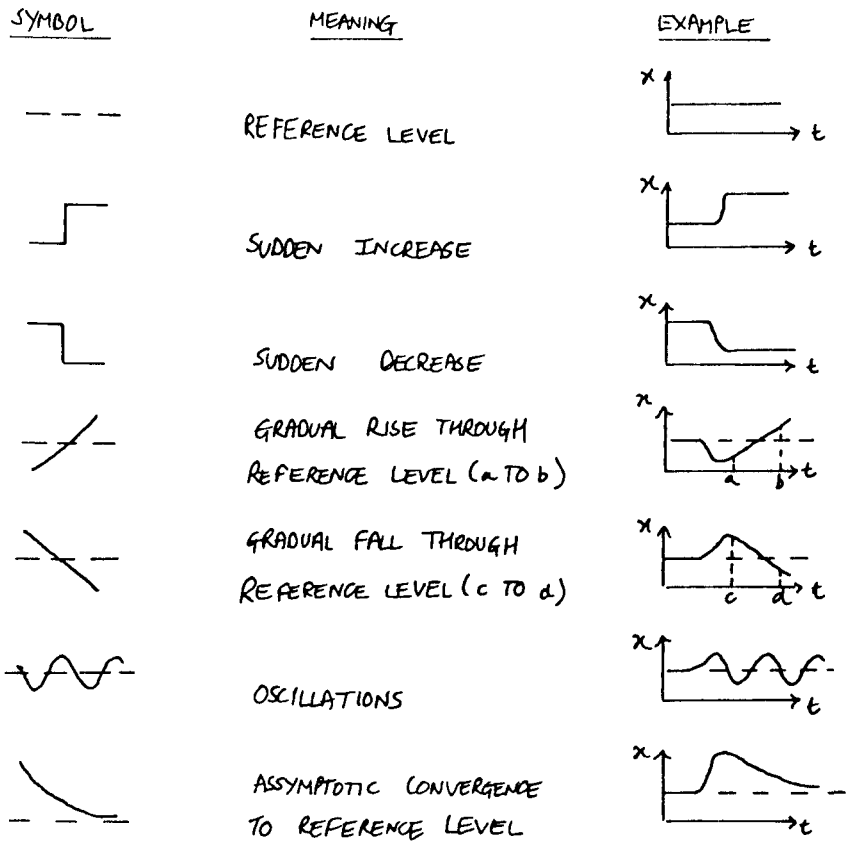


FIGURE 5.3. A SHORTHAND ACTOGRAMMATIC NOTATION FOR QUALITATIVE DYNAMIC FEATURES.

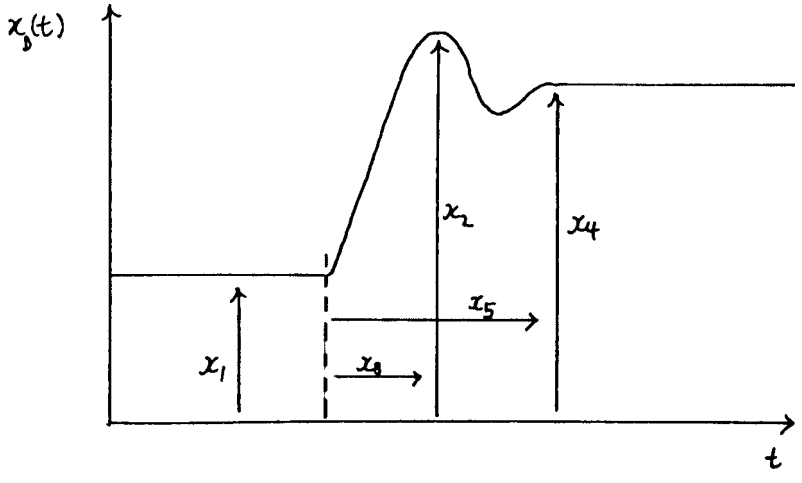


FIGURE 5.4. AN EXAMPLE OF FEATURES OF DYNAMIC RESPONSE DATA

to small variations in parameters should be examined (e.g. by analytic methods, sensitivity analysis, or Monte-Carlo simulation, Section 5.2.4). If the occurrence of the feature is highly sensitive to small parameter changes then it is likely that the model structure is invalid. To sum up: the matching of qualitative features in model and data is important because it indicates that the underlying mechanisms of the model are those of the range of application.

5.2.2 Quantitative comparisons in feature space

Quantitative features may be used for comparison if they are related to important aspects of $\mathcal{R}_I$ (i.e. quantification of qualitative features) or when direct comparisons of model and data are not possible. An example of the latter occurs in a general model of a class of biological systems (e.g. human cardiovascular systems) where there is no single standard system from which to use data for comparison (to some extent the extraction of features can be regarded as a normalisation procedure). Figure 5.4 is an example of a set of features $x = \{x_1, x_2, x_3, x_4, x_5\}$ taken from dynamic transient response data. The features x_1 to x_5 are a compact representation of the dynamic response and also convey the information that it is a step response with overshoot. They can be normalised to the pre-transient value thus:

$$\left.\begin{aligned} y_1 &= \frac{x_2 - x_1}{x_1} \\ y_2 &= x_3 \\ y_3 &= \frac{x_3 - x_1}{x_1} \\ y_4 &= \frac{x_4 - x_1}{x_1} \\ y_5 &= x_5 \end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots \quad (5.2)$$

The features define a feature space which may be used in two ways for validation: direct comparison of feature vectors of model and data, and indirect comparison using a feature space classifier.

5.2.2.1 Direct comparison of feature vectors

Let x_D be the vector of data features and x_M be the vector of model features. The validity criterion may be threshold or statistical:

(i) M is valid iff $|x_M - x_D| < \varepsilon$, where ε is a threshold, or error margin, related to the uncertainty of the data.

(ii) M is valid iff $P[x_M = \zeta_D] > 95\%$, which can be determined statistically (e.g. by assuming x_D are normally distributed) and where ζ_D is the random variable corresponding to x_D.

The distance between x_M and x_D in feature space can be used pseudometrically by which to order a set of competing models (Reggiani and Marchetti, 1975; Argentesi, 1978). A "figure of merit", or measure of model adequacy, F, may be defined in terms of the error between the data features $x_D(k)$ and the model features, $x_M(k)$, $(k = 1, N)$:

$$F = \frac{1}{1 + \frac{1}{N} \sum_{k=1}^{N} w_k \delta_k} \quad , F \in [0, 1]$$

where

$$\delta_k = \left| \frac{x_M(k) - x_D(k)}{x_D(k)} \right|$$

and w_k are weighting factors $(w_k \in [0, 1])$. δ_k is the fractional error of the k^{th} model feature. If there is no error, F = 1, whereas an average error of $\pm 50\%$ between model and data gives F = 0.67. A figure of merit such as this can be very useful as an objective measure for deciding among competing models (e.g. with different parameter values). The precise definition is not critical, but it is important that the measure should be calibrated in terms of the average error.

5.2.2.2 Indirect comparison using a feature space classifier

The mapping of the data into an m-dimensional feature space $X = R_e^m$ may produce a clustering of the data for different experimental conditions $u_i (i = 1, n)$ into distinct classes. Essentially this is a qualitative property of the data ($\mathcal{R}_I$) and it may be used as a further test of the empirical validity of the model. A classifier ρ is constructed and tested using the data set which classifies the feature vector $x_D(u_i) \in X$ into the class s_i which denotes the experimental conditions u_i. The model response $y_M(t)$ is mapped into feature space and classified using ρ as illustrated in Figure 5.5. The form of the classifier shown in Figure 5.5 is deterministic, $\rho = X \to S$, where S is a set of symbols $\{s_i\}$,

$$\rho(x_D(u_i)) = s_i \quad , i = 1, n$$

$$\text{and} \quad \rho(x_D(u_i)) \neq \rho(x_D(u_j)) \quad i \neq j \qquad \ldots\ldots\ldots\ldots \quad (5.3)$$

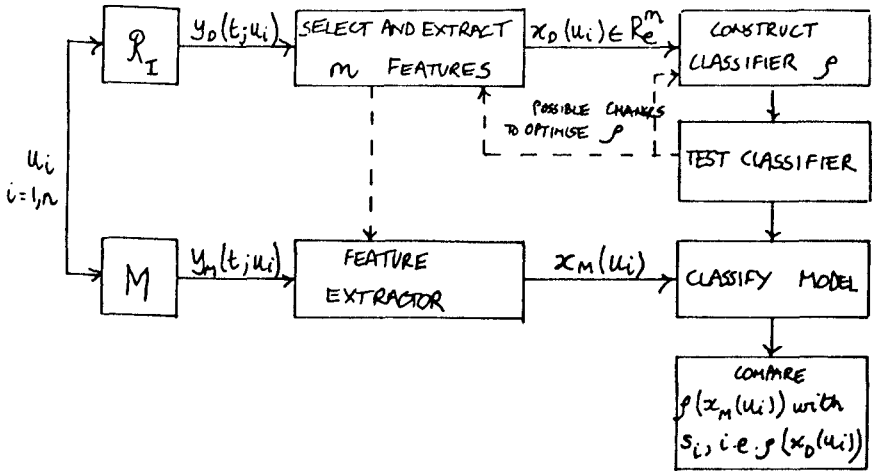


FIGURE 5.5. A METHOD OF MODEL-DATA COMPARISONS BASED ON A FEATURE SPACE CLASSIFIER.

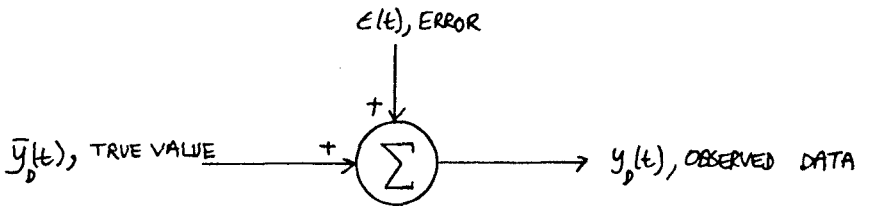


FIGURE 5.6. ADDITIVE ERROR MODEL OF DATA UNCERTAINTY.

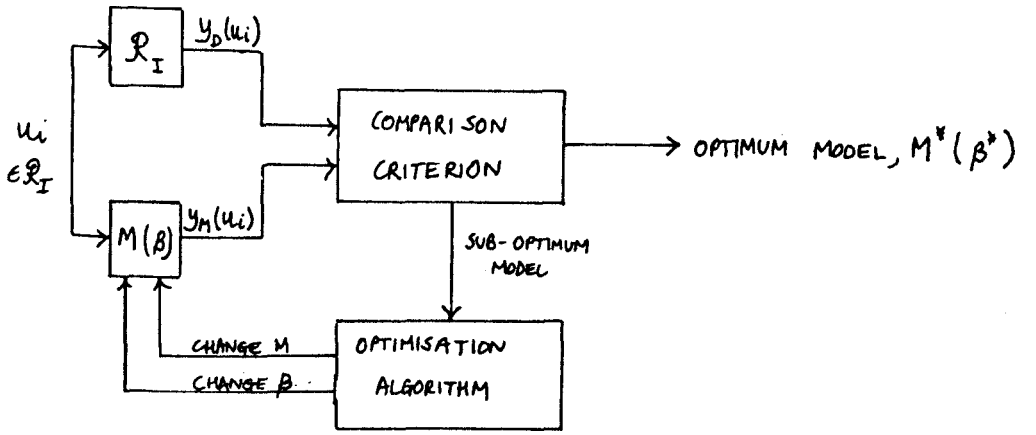


FIGURE 5.7. GENERAL SCHEMA FOR SYSTEM IDENTIFICATION AND PARAMETER ESTIMATION.

The data response for each u_i is mapped into a distinct s_i, and this should hold also for the model. The criterion for validity is therefore:

$$\text{M is valid iff} \quad \rho(x_M(u_i)) = s_i, \quad i = 1, n,$$

$$\text{i.e. iff} \quad \rho(x_M(u_i)) = \rho(x_D(u_i)), \quad i = 1, n \qquad \ldots (5.4)$$

Typical classifiers suitable for this purpose are the linear discriminant classifier (see, for example, Duda and Hart, 1973) and the cellular classifier. A probabilistic classifier may also be used (see Section 5.2.4).

5.2.3 Time series comparisons

The data are continuous $(y_D(t), t\varepsilon[0, T])$ or discrete $(y_D(k), k = 1, N)$, $t = k\Delta T$ over a finite observation time T ($N = T/\Delta T$). The model response is $y_M(t)$ or $y_M(k)$ and is fixed (i.e. the model parameters or structure are not variable). Most empirical model-data comparisons are made at this level and may be simple comparisons based on a maximum allowable error (Section 5.2.4.1) or involve the use of a test stiatistic (Section 5.2.4.2).

5.2.3.1 Maximum allowable error

A maximum allowable error $\varepsilon(t)$ or $(\varepsilon(k))$ is defined which is related to the uncertainty of the data or the desired resolution of $\mathcal{R}_I$ (e.g. model predictions to within $\pm$ 10% of the data). Thus the empirical validity criterion is:

$$\left.\begin{aligned} \text{M is valid iff} \quad & |y_D(t) - y_M(t)| < \varepsilon(t), \quad t\varepsilon[0, T] \\ \text{or iff} \quad & |y_D(k) - y_M(k)| < \varepsilon(k), \quad k = 1, N \end{aligned}\right\} \quad \ldots\ (5.5)$$

The interval $[0, T]$ is known as the "validation interval". Although this criterion appears trivial, it is probably the most widely used empirical validity criterion in the whole of science. Ideally, $[0, T]$ should cover the whole of $\mathcal{R}_I$, but it is very unlikely that data will be available for such a period. (Complete data may be available for many technical models, but for some models whose purpose is for prediction (e.g. econometric models) the validation interval is by definition less than $\mathcal{R}_I$). The criterion can be extended to include u_i:

$$\text{M is valid iff} \quad |y_D(t;u_i) - y_M(t;u_i)| < \varepsilon(t), \quad t\varepsilon[0,T], \quad \forall u_i \in \mathcal{R}_I \qquad \ldots\ (5.6)$$

i.e. for a range of observation conditions, experimental tests, inputs, etc. extending over $\mathcal{R}_I$. Models of biological systems can generally

satisfy the criterion for $t\varepsilon[0, T]$, but data are very rarely available for a wide range of regions of $\mathcal{R}_I$ defined by u_i. More correctly, a biological model (and other models) may satisfy only "local" validity criteria, i.e.

$$|y_D(t;u_i) - y_M(t;u_i)| < \varepsilon(t),\ t\varepsilon[0,T],\ u_i = \bar{u}_i + \delta_i\ \varepsilon\ \mathcal{R}_I,\ i = 1, n \quad \ldots\ldots \quad (5.7)$$

In other words, the criterion is satisfied for a finite number of small regions of $\mathcal{R}_I$. For linear systems, the generalisation of local validity (5.7) to general validity (5.6) is valid if the points $\bar{u}_i$, $i = 1, n$, range across $\mathcal{R}_I$. However, for nonlinear systems, the generalisation cannot be made, unless the $\bar{u}_i$ are associated with the set of qualitative modes of nonlinear behaviour. In this situation there is prior confidence that the structure, or mechanism, of the model is valid. Hence the emphasis on qualitative comparisons in the first stage of the α-methodology is very important. (Similar considerations can be made regarding the generalisability of validity over $[0, T]$ to the interval $[0, \infty]$).

A figure of merit, F, similar to that used with quantitative features (Section 5.2.2.1) may be defined for the comparison of model and data time series. This may be helpful as an objective criterion to select between competing models:

$$F = \frac{1}{1 + \frac{1}{T}\int_0^T w(t)|x_M(t) - x_D(t)|dt}$$

5.2.3.2 Statistically-based comparisons

In this type of comparison the uncertainty of the data is explicitly used in the form of statistics (e.g. mean values, variances, covariances, etc.) or probabilities (e.g. probability distributions, probability density functions, etc.). A wide survey of the variety of statistical techniques for data analysis is beyond the scope of this thesis, and in this section only a short overview will be given. The view taken here is that statistical techniques used in model validation should be as simple as possible, and that the range of models for which advanced statistical techniques are appropriate is really quite small (typically the design or interpretation of experiments which involve a large number of external influencing factors which must be controlled-out, as in psychology or agricultural research).

Although the theoretical basis of statistics is the theory of probability (see, for example, Cox and Hinkley, 1974) which provides a completely general model of data uncertainty, 99% of statistical tests on data are based on an additive error model of uncertainty such as Figure 5.6. The observed value $y_D(t) = \bar{y}_D(t) + \varepsilon(t)$, where $\bar{y}_D(t)$ is a supposed "true" value, and $\varepsilon(t)$ is an error term which is usually assumed to be symmetric (and, frequently, normally distributed). Some points related to this model are discussed at the end of this section.

In statistics, the question "is it likely that model M is valid by comparing model response and data values?" is replaced by the question "is it likely that model M could generate the data values observed?". It is expressed as the null-hypothesis", H_o: the difference between model and observed values is largely accounted for by data error. The testing of H_o is a significance test on a defined test statistic s which is a function of the observations y_D and model values y_M, $s = s(y_D, y_M)$. The random variable associated with s is $S = S(Y_D, y_M)$, where Y_D is the random variable associated with the observed (or realised) data y_D.

For a given set of observations $s = s_D$ is calculated, and the "level of significance" p_D is given by

$$p_D = Pr[S \geq s_D; H_o] \qquad \ldots\ldots\ldots (5.8)$$

Included in H_o is an assumption about the error distribution of the data (e.g. normal or student-t distributions) which allows equation (5.8) to be evaluated. If H_o is to be accepted then it should be unlikely that $S \geq s_D$, and conventionally a small value of 0.01 or 0.05 is assigned to p_D, corresponding to 99% and 95% levels of significance, respectively.

A typical test statistic s is the chi-squared statistic, χ^2. χ^2 goodness of fit statistics can be formed by grouping the data in some way, finding observed and expected (model) frequencies in the cells so formed and taking the test statistic:

$$\chi^2 = \sum \frac{(\text{observed freq.} - \text{expected freq.})^2}{\text{expected freq.}} \qquad \ldots\ldots (5.9)$$

Charts of χ^2 against degrees of freedom (in the fit of model to data) are available on which are plotted lines of level of significance based on the normal distribution. For a set of data and model responses, χ^2 is evaluated and the point located on the χ^2 chart. If $p_D \leq 0.01$ (or some

small number), then the null-hypothesis is accepted.

(For accounts of alternative test statistics and their theoretical basis consult Cox and Hinkley, 1974; or for a standard work on the foundations of statistical inference consult Kendal and Stuart, Vol. 2, 1973).

Equation (5.8) can be used as a criterion of validity if it is remembered that H_o is effectively composed of the model M in question, and an assumed statistical model M_{STAT} (e.g. a normal probability distribution):

$$M \text{ is valid iff } p_D = Pr[S \geq s_D;\ M,\ M_{STAT}] \leq \gamma \quad \ldots\ldots\ (5.10)$$

where γ is a small number (0.01, 0.05, etc.), and iff M_{STAT} is a valid model of the data uncertainty.

To this criterion, the constraints of the validation interval $[0, T]$ and the available test conditions u_i may be added:

$$M \text{ is valid iff } p_D = Pr[S \geq s_D(u_i);\ M(u_i),\ M_{STAT}] \leq \gamma,\ t \in [0,T],$$
$$u_i = \bar{u}_i + \delta_i \in \mathcal{R}_I,\ i = 1, n \quad \ldots\ldots\ (5.11)$$

When the criterion is expressed in this form its similarity to the maximum allowable error criterion (equation (5.7) becomes evident, as do the problems associated with the generalisation of validity beyond $[0, T]$ or u_i, i = 1, n (see Section 5.2.4.1).

For models which are based on a fit, or regression, to the data, the variance-covariance matrices of the residuals $(y_M(k) - y_D(k))$ and the expected error on the coefficients can provide extremely useful information for model validation:

(i) Whiteness of residuals. If the model is valid, then it can be considered to generate the observations, and any residuals, therefore, should be uncorrelated (assuming data errors are uncorrelated). If the variance-covariance matrices of residuals (on variables or coefficients) are found to contain significant non-diagonal terms, then the residuals are highly correlated, non-white, and it is likely that the model could not generate the data (i.e. is invalid).

(ii) If the model satisfies (i), the variance-covariance matrix of the estimated coefficients may be used to predict the expected error between model and data beyond $[0, T]$ (typically it grows steadily). Thus some kind of measure of validity beyond $[0, T]$ is possible. However, the only really acceptable test is the comparison with new data when they become available.

In the methods of "cross-validation" (see Section 2.8.2, or Mosier, 1951, McCarthy, 1976), the data are split into two halves, one for fitting the model (e.g. regression equations), and the other for validating the model (using a test statistic and a level of significance).

Although the use of statistical techniques for comparing model responses with data is very powerful and can be very informative, statistical models and techniques cannot be used successfully on their own. For instance, if a turning point occurs in the data just after T, this will not occur in the model, or be revealed by empirical validation over $[0, T]$. However, a priori theoretical considerations of the behaviour of $\mathcal{R}_T$ may suggest that certain mechanisms should be included in the model, and these will improve the validity of the model beyond $[0, T]$ in a way that no data-based method can. For this reason, simple statistical models (such as ARMA models, Box and Jenkins, 1970) often work better than more sophisticated models (Makridakis and Hibon, 1979).

The additive error model often assumed in statistics (Figure 5.6) requires that the data uncertainty be symmetric and with zero mean. This assumption can be checked against the data, as well as any other models in M_{STAT}. In fact, the best approach is to list all the possible sources of uncertainty in the data and their likely effects (for a framework, see Section 4.3.2.5) and to consider these when comparing the data and model responses.

5.2.4 Comparisons using system identification, parameter estimation, and sensitivity analysis techniques

In these types of comparison, the parameters (and possibly structure) of the model are systematically varied when comparing model with data. System identification and parameter estimation techniques seek to optimise the model by minimising a loss function between the model and data responses (Section 5.2.4.1), whereas the sensitivity analysis techniques reported here are concerned with the extent to which global properties of the model are affected by slight changes in its parameters or structure (Section 5.2.4.2).

5.2.4.1 System identification and parameter estimation techniques

5.2.4.1.1 General form

The general form, or scheme, for system identification and parameter estimation is shown in Figure 5.7, (a standard reference on the subject is Eykhoff, 1974). β denotes an m-dimensional vector of parameters and

M symbolically denotes the model structure. The optimum, or final values are denoted by an asterisk. Three forms for the comparison criterion are:

(i) Least square error (LSE):

$$\left.\begin{aligned} &\min_{M}\ \min_{\beta}\ \{L(y_D,\ y_M(\beta))\} \\ \text{e.g.}\quad L &= \int_0^T |y_D(t) - y_M(t;\beta)|\,dt \\ L &= \sum_{k=1}^{N} (y_D(k) - y_M(k,\ \beta))^2 \end{aligned}\right\} \qquad \ldots\ldots\ (5.12)$$

(ii) Maximum likelihood estimate (MLE)

If $Pr[\varepsilon]$ is the probability distribution of the error on the data, then $L = Pr[y_D - y_M(\beta)]$ is known as the "likelihood" function.

The comparison criterion is:

$$\max_{M}\ \max_{\beta}\ \{Pr[y_D - y_M(\beta)]\} \qquad \ldots\ldots\ (5.13)$$

(In practice, the log likelihood function is used, and the probability distribution is often assumed to be a multivariate normal distribution.)

(iii) Bayesian estimator

This estimator makes use of the a priori expected probability distribution of the parameters and uses Bayes' Theorem.

The optimisation algorithm should search the model parameter space $\mathcal{MP}$, where $\mathcal{M} = \{M_1, \ldots, M_n\}$ finite set of candidate model structures and $\mathcal{P} \subseteq R_c^m$, the allowable parameter space. In practice, the search of $\mathcal{P}$ is conducted for each candidate model using an algorithm such as the hill-climbing or simplex methods (which must satisfy convergence and robustness conditions), and then the optimum values of loss function (or MLE) are compared for the set of candidate models. This allows the selection of an overall optimum model $M^*(\beta^*)$ (which may also be the minimal model that satisfies the criterion).

5.2.4.1.2 Definitions of identifiability and implications for validation

In Section 2.6 the literature of system identification was reviewed for its implications for model validity and validation, and the various concepts of identifiability were examined. Identifiability is concerned with the following problem: "given a class of models $\mathcal{M}$, constraints on parameters $G(\beta) \geq 0$, a set of tests u_i (i = 1, n), and a vector of system

responses y_D, will the scheme of Figure 5.7 lead to a unique $M^*(\beta^*)$?". There are two stages to solution. The first (associated with a priori, or theoretical, identifiability) assumes that y_D are available and completely accurate. The second (associated with a posteriori, practical, or empirical, identifiability) deals with the situation where data contain uncertainty and may not be completely available (e.g. sampled data). Ideally, a model should be shown to be a priori identifiable (a matter of algebraic manipulation), and then used in a practical situation where the variance-covariance matrix of the parameter estimates may be used to determine the loss of identifiability because of data uncertainty (see Section 2.6.4). The following discussion is an analysis of the concept of identifiability in terms relevant to model validity and validation and should not be interpreted as a criticism of system identification whose problems of central concern are different.

The concept of identifiability is an attempt to specify a necessary condition under which the technical optimisation problem has a unique solution, i.e. a true or global maximum is locatable (i.e. the problem as defined by Zadeh, 1962; or Bellman and Åström, 1970). It might seem to follow that the unique model $M^*(\beta^*)$ would then be a correct model of the system or, which is essentially the same, would yield valid predictions about the system which can be used for the purposes of system control. However, this interpretation depends on the validity of the constraints required to articulate the problem completely ($\mathcal{M}$, $G(\beta) \geq 0$, u_i, y_D, - see above). It is obvious that without the constraint of $\mathcal{M}$ there would be no unique solution, since there is an infinite number of conceivable models.

If the formulation of the model has been based on well-validated theory and accurate data then the confidence in the constraints is high (e.g. the modelling and identification of technological systems). Identifiability can be checked, and the model used with much confidence that it is a valid model (theoretically and empirically). But then if the model (i.e. constraints) is valid a priori, identifiability is automatically guaranteed. Final empirical validation tests would be made on the model using the earlier stages of the α-methodology and an analysis of the variance-covariance matrix (or probability distributions) of the parameter estimates.

On the other hand, for models whose formulation does not rest wholly on validated theory (such as complex biological models), then the

confidence in the constraints is low, and the results of an identifiability analysis difficult to interpret. However, parameter estimation results on the model would be used as part of the α-methodology in order to determine the validity of the model. In terms of model validity and validation, therefore, identifiability is not a fundamental criterion, although model validation and identifiability analysis are closely interlinked, particularly in the area of biological modelling (e.g. metabolic modelling). (It is significant that writers on system identification are recently referring much more to questions of model validation, e.g. Mehra, 1980; Carson, Cobelli and Finkelstein, 1980.) An example of identifiability analysis is in the identification of "minimal models" for clinical application in medicine.

5.2.4.1.3 The use of parameter estimation in model validation

The results of model parameter estimation are very important in model validation. If a model is valid, then it should be capable of generating the data under a range of conditions in $\mathcal{R}_I$. The discrepancy between the model and the observed data should therefore simply be due to data uncertainty and will be reflected in the variance-covariance matrices of the residuals and parameter estimates. The var-covar matrix of residuals should match that of the data errors, and if the errors are uncorrelated it should be a diagonal matrix. If the covariances are large, then the model is having difficulty adjusting to the data and is probably invalid structurally. Even if the data errors are correlated, the var-covar matrix of parameter estimates should be largely uncorrelated, and it provides a very effective test of the validity of the model.

An alternative method is to examine the points, or regions, β_i^* in parameter space $\mathcal{P}$ which are the results of parameter estimation in a series of tests u_i, i = 1, n (e.g. β_i^* may be considered to be a hyper-ellipsoid whose diameters are functions of the variances of the components of β_i). A simple criterion for model validity is that of parametric consistency:

$$\text{Model M is valid iff } \bigcap_{i=1}^{n} \beta_i^* \neq 0 \ , \ u_i \in \mathcal{R}_I, \ i = 1, n \quad \ldots\ldots \ (5.14)$$

In other words, there is a single set of parameter values which satisfies an estimation criterion (e.g. minimum loss function) for all test conditions. This was suggested in connection with biological models by Berman (1963) and in practice is frequently used implicitly (e.g. Bali,

1976; Pullen, 1976). This principle can also be used in conjunction with a classification method of model validation (see Sections 5.2.2.2 or 5.2.4.1.4). (An additional criterion is that the parameter estimates should be "feasible".)

5.2.4.1.4 The use of a probabilistic classifier in joint model estimation-validation

In this section, an outline is given of a method for model parameter and structure estimation which is based on a probabilistic classifier and includes a validity criterion. The probabilistic classifier is denoted by $\rho : X \times S \rightarrow [0, 1]$, where X is feature space and S is a finite symbol space. The classifier is constructed from the data set (an ensemble of data, $y_D(\zeta)$ from $\mathcal{R}_I$). ρ is a conditional probability measure, i.e. $\rho(s_k\ x)$ = probability that s_k is the correct symbol to be assigned to location x in feature space. The objective of the classifier is to produce a clear classification of feature space according to the set of input tests u_i = 1, n, (i.e. $\rho(s_j | x_D(u_i))$ should be maximal over S when $s_j = s_i$ (corresponding to u_i)).

The model response is then applied to the classifier (see Figure 5.8) for each test u_i. The parameters β of the current model structure M_α are varied to maximise the probability of correct classification. The next step is the validity criterion:

Model $M_\alpha(\beta)$ is valid iff $\rho(s_i | x_{M_\alpha}(\beta, u_i))$ is maximal for u_i, i = 1, n.

i.e. iff $\rho(s_i | x_{M_\alpha}(\beta, u_i)) \geq \rho(s_j | x_{M_\alpha}(\beta, u_i))$, $\forall s_i, s_j \in S$ .....(5.15)

or, in words, $M_\alpha(\beta)$ is valid if the probability of correct classification for each model response is greater than the probability of misclassification for all u_i (this is the probabilistic equivalent to parametric consistency, Section 5.2.4.1.3).

If the criterion is satisfied M_α is accepted, and the next model structure $M_{\alpha+1}$ is tested. Eventually, a set of competing model structures is obtained which satisfy the validity criterion. There is a variety of possible measures for comparing these models, but the simplest and best is the product of the correct classification probabilities for each test, which is the overall probability of correct classification for the series of tests, P_u. The model whose P_u is maximum is the one which should be selected:

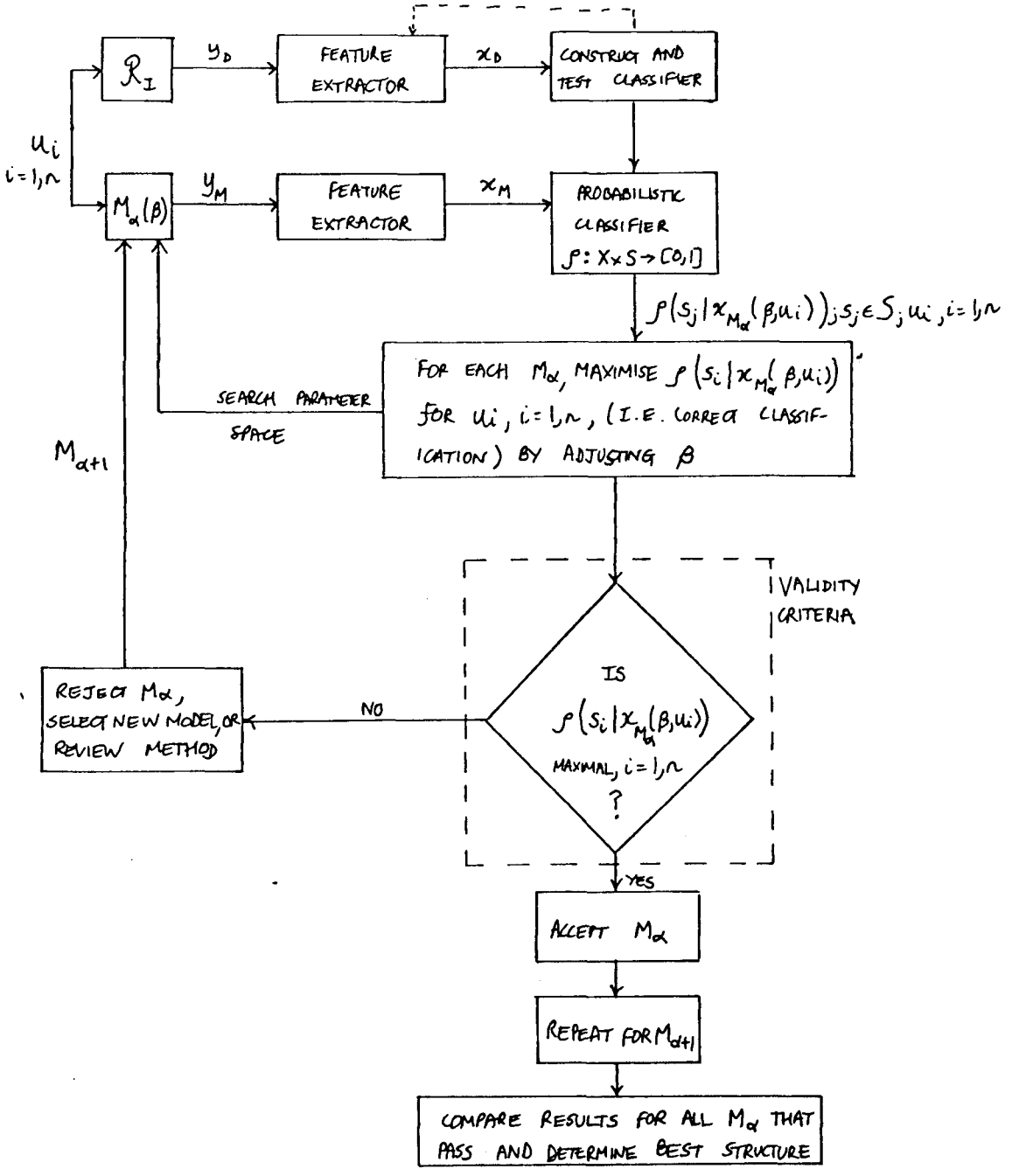


FIGURE 5.8. THE USE OF A PROBABILISTIC CLASIFIER IN JOINT MODEL IDENTIFICATION-PARAMETER ESTIMATION - VALIDATION.

$$P_u(k) = \prod_{i=1}^{n} \rho\left(s_i \mid x_{M_k}(\beta, u_i)\right) \quad \text{for successful models } M_k, M_\ell \ldots$$

the best model M_ℓ is given by:

$$P_u(\ell) \geq P_u(k) \qquad \forall M_k \qquad \ldots\ldots\ldots (5.16)$$

5.2.4.2 Sensitivity analysis

The techniques included in this section investigate the effect on overall model behaviour of small changes in model structure, parameters, or inputs (including initial conditions). In empirical validation these techniques may be used for a variety of purposes:

(i) To examine the dependency of qualitative features of the model (stage 1 of the α-methodology) on various factors. For instance, if a qualitative feature of the model response which has been used as an indicator of model validity disappears with slight variation of parameters, this severely attenuates the valid range of application $\mathcal{R}_V$ of the model and implies that the structure is invalid.

(ii) To determine likely ranges of uncertainty for model parameters that cannot be directly measured.

(iii) To trace through uncertainties in initial conditions, inputs, structure, and model parameters on to the overall response of the model. The range of model responses can then be retested to see if they satisfy empirical validation criteria (e.g. feature or time series comparisons).

(iv) To determine a parameter sensitivity matrix to generate new search directions in parameter space for model parameter estimation.

(v) To determine optimal variables and times for measurement points in order to estimate a certain parameter (i.e. experimental design).

In Section 5.2.4.2.1 a series of techniques is outlined in which one parameter at a time is varied, whereas in Section 5.2.4.2.2 the method of Monte-Carlo simulation, where the entire parameter set is varied, is discussed.

5.2.4.2.1 Sensitivity coefficients and equations

Consider the dynamic model, M;

$$\dot{x} = f(x, \beta; t, u), \quad x \in R_e^n, \ \beta \in R_e^m \qquad \ldots\ldots\ldots (5.17)$$

where u is an input vector. The dependency of state variable x_j on

parameter β_r is represented by the "sensitivity coefficient" C_r^j:

$$C_r^j = \frac{dx_j}{d\beta_r} = C_r^j(x, \beta; t, u) \quad \ldots\ldots\ldots (5.18)$$

For M, there are n x m sensitivity coefficients, all time-varying. The dynamic responses of the sensitivity coefficients can be determined analytically or by simulation from the "sensitivity equations". Differentiating equation (5.17) with respect to β_r, and the j^{th} state variable, x_j, gives:

$$\frac{d}{d\beta_r}\left(\frac{dx_j}{dt}\right) = \frac{d}{d\beta_r}\left(f_j(x, \beta)\right)$$

therefore

$$\dot{C}_r^j = \frac{df_j(x,\beta)}{d\beta_r} + \sum_{i=1}^{n} C_r^i \frac{\partial f_j(x,\beta)}{\partial x_i} \quad \ldots\ldots\ldots (5.19)$$

In total, there are n x m sensitivity equations whose solution requires the prior solution of equation (5.17) for x(t). (For references on sensitivity theory, the classic work is by Tomović, 1963, which is also quite readable; a more general introductory book is Frank, 1978.) In practice the sensitivity equations can be solved analytically for linear models, and by numerical integration for some small non-linear models. However, for complex nonlinear models the determination of the dynamic sensitivity coefficients is intractable by this method. An alternative, yet simple, method is to obtain the approximate sensitivity coefficients by perturbing each parameter β_r by a small amount $\Delta\beta_r$:

$$C_r^j \approx \frac{\Delta x_j(t)}{\Delta\beta_r}, \text{ where } \Delta x_j(t) = x_j(\beta_r + \Delta\beta_r, t) - x_j(\beta_r, t) \quad \ldots (5.20)$$

Sensitivity analysis varies the parameters sequentially out along the axes of parameter space from their nominal position. It therefore leaves many directions uncovered. In Monte-Carlo simulations all the parameters are varied randomly in an attempt to cover more regions of the parameter space surrounding the nominal value.

5.2.4.2.2 Monte-Carlo simulation

In Monte-Carlo simulation, the model is simulated a large number of times (e.g. 50 - 100) with a different set of model parameters (or initial conditions, inputs, etc.) each time. The values of the parameters

are given by:

$$\beta_r = \beta_{r_o} + \Delta\beta_r(k) \qquad r = 1, m$$

where $\Delta\beta_r(k)$ is the random value of a variation for the k^{th} simulation drawn from a probability distribution. On most computers there are facilities for interval or gaussian random number generation which can be used for this purpose. Usually, variations are uncorrelated (i.e. $E\{\Delta\beta_r(k).\Delta\beta_r(\ell)\} = E\{\Delta\beta_r(k).\Delta\beta_s(k)\} = 0$), but sometimes they may be correlated, (e.g. in assessing the effect of correlated noise disturbances on the inputs). Finally, the results of the simulations can be presented as histograms, statistics (e.g. means and var-covar matrices), or probability distributions for the relevant model variables. These may be compared with the data using statistical techniques if appropriate.

(There are other techniques for representing the various forms of uncertainty in a model and their effect on overall response; for an excellent reference, see Schweppe, 1973.)

5.3 An Empirical Validation Methodology (β - Methodology)

The β-methodology is intended for the empirical validation of mathematical models with respect to their specific scientific objectives (i.e. intended range of application, $\mathcal{R}_I$). Prerequisites for the use of the β-methodology are that the models are based on well-established theories, and that plenty of data are available from $\mathcal{R}_I$ (i.e. $D_A \supset D_M$, where D_A are the available data types, and D_M are the data types required by the model). In other words, when the β-methodology is entered the models have established their theoretical validity (usually automatically in the model formulation process). These prerequisites describe the stage of development of domains in which the β-methodology is appropriate. Examples include many models in physics and chemistry, and technical modelling of engineering and industrial systems. A flow diagram for the β-methodology is shown in Figure 5.9.

5.3.1 Preliminary considerations

The first stage of the β-methodology (and also the other methodologies which follow) is an analysis of the modelling objectives and available data types. The specific scientific modelling objectives determine the intended range of application $\mathcal{R}_I$, and the desired spatio-temporal

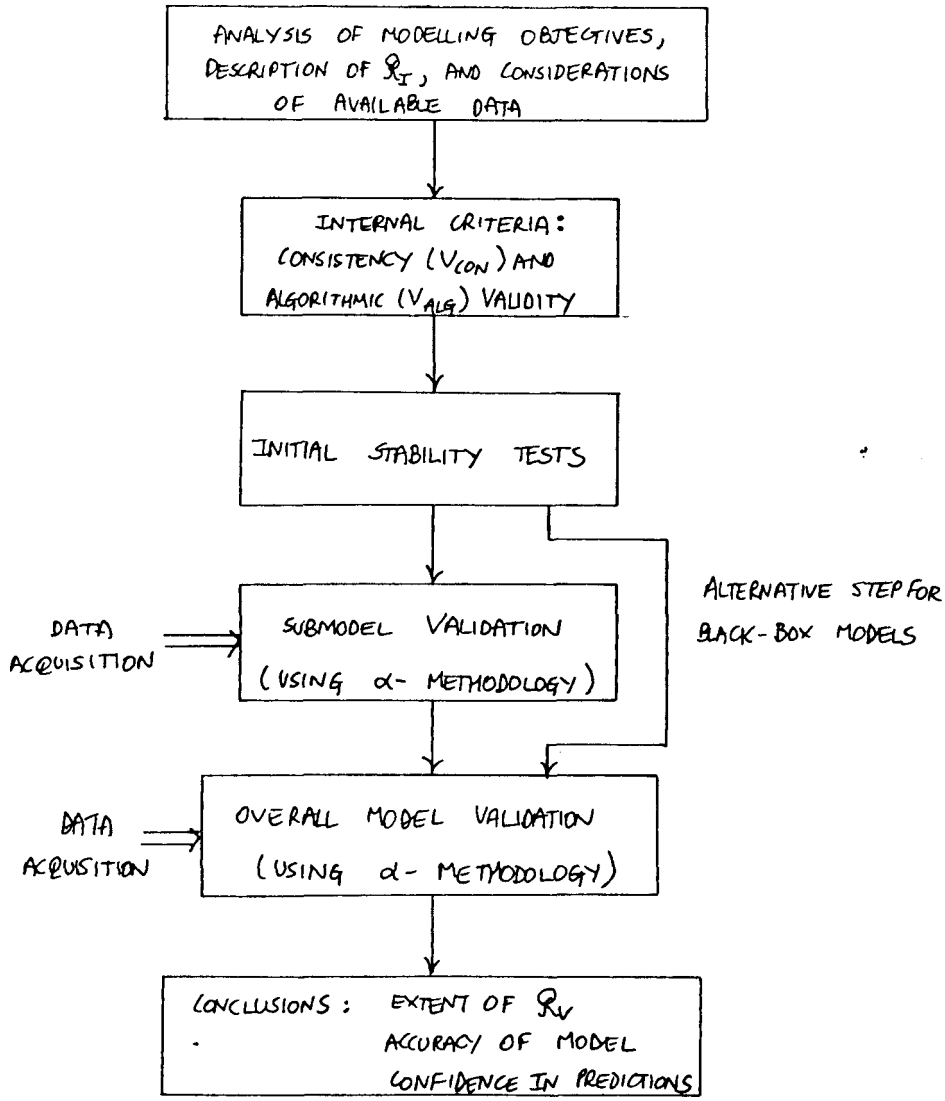


FIGURE 5.9. THE β-METHODOLOGY FOR EMPIRICAL MODEL VALIDATION.

extent and resolution over both physical (structural) and functional modalities should be clearly described. Considerations of data in the β-methodology are concerned with problems of uncertainty. The aim of the β-methodology is to determine critically the empirical range of valid application $\mathcal{R}_V$ of the model.

5.3.2 Necessary conditions

The next two stages are concerned with testing necessary conditions that must be satisfied before the validation proper can start. The first is the application of V_{CON} (consistency) and V_{ALG} (algorithmic) internal validity criteria to the model. These are self-evident (refer to Section 4.3.3.2). The second is a check that the model is not unstable, or prone to instabilities (other qualitative aspects are considered later).

5.3.3 Submodel validation

If the model is an explanatory model (or to increase confidence in a model for predictive purposes), the submodels must be validated empirically. This involves validating the elementary relationships and submodels using the α-methodology.

5.3.4 Overall model validation

The overall model structure and behaviour is validated using the full resources of the α-methodology. ("Black-box" type models for prediction, which are often intended for utilitarian objectives, may step from the third to fifth stages, omitting submodel validation.) This stage will often culminate in a parameter estimation or sensitivity analysis study which will allow the accuracy of the model to be described in statistical terms.

5.3.5 Concluding the β-methodology

The conclusions of the β-methodology should be to delimit $\mathcal{R}_V$ and to assess the accuracy of a model. Occasionally, it may be possible to quantify confidence in model predictions. If $\mathcal{R}_V \supseteq \mathcal{R}_I$ then the model may be considered valid, and be used for other purposes. However, if $\mathcal{R}_V \subset \mathcal{R}_I$ then the model is incompletely valid, and may have to be reformulated (if $\mathcal{R}_V$ is very much smaller than $\mathcal{R}_I$, or there are some significant anomalies, then it may be necessary to challenge the theoretical basis of the model).

5.4 A Theoretical-Empirical Validation Methodology (γ-methodology)

The γ-methodology is a validation methodology based on the explicit application of both theoretical (V_{THEOR}) and empirical (V_{EMP}) criteria. It is suited for models in domains which have reasonable theoretical sophistication and available types, but where the model contains theoretical development and makes requirements for new data types. For example, mathematical models in biology are often based on mainly descriptive theories which they extend into mathematical form, and call for new data types (such as continuous dynamic data, or measurement of new variables or parameters). The γ-methodology may be used alongside the model formulation and development process, or be applied after the completion of the model in which role it is a sort of critical disassembling and reassembling of the model. A flow diagram of the methodology is shown in Figure 5.10.

5.4.1 Preliminary considerations and necessary conditions

The preliminary considerations of the γ-methodology are like that of the β-methodology, except that the intended range of application will not be so well defined, and the available data types (D_A) may be problematic. It is quite likely that the data requirements of the model (D_M) for extensive empirical validation will exceed those available (i.e. $D_M \supset D_A$). For this reason, the theoretical tests are very important. As in the β-methodology, the necessary conditions of V_{CON}, V_{ALG}, and initial stability must be satisfied.

5.4.2 Joint application of theoretical and empirical criteria

The theoretical and empirical criteria are applied together, starting with the elementary submodels, or assumptions, and gradually aggregating the submodels into intermediate submodels and finally the complete model. In this way, the areas of confidence (validity) and uncertainty in the model can be clearly identified. As the submodels are aggregated, the level of validation increases and new properties characteristic of that level must be compared between the model and data. The α-methodology is used for the empirical tests, but because of the limitations of available data, it may only be possible to use the first stage (qualitative comparisons), if any at all.

As the level of validation rises, if the model satisfies V_{THEOR} and/or V_{EMP}, it may be possible to infer indirectly the validity of submodels which did not receive adequate direct validation. This exemplifies

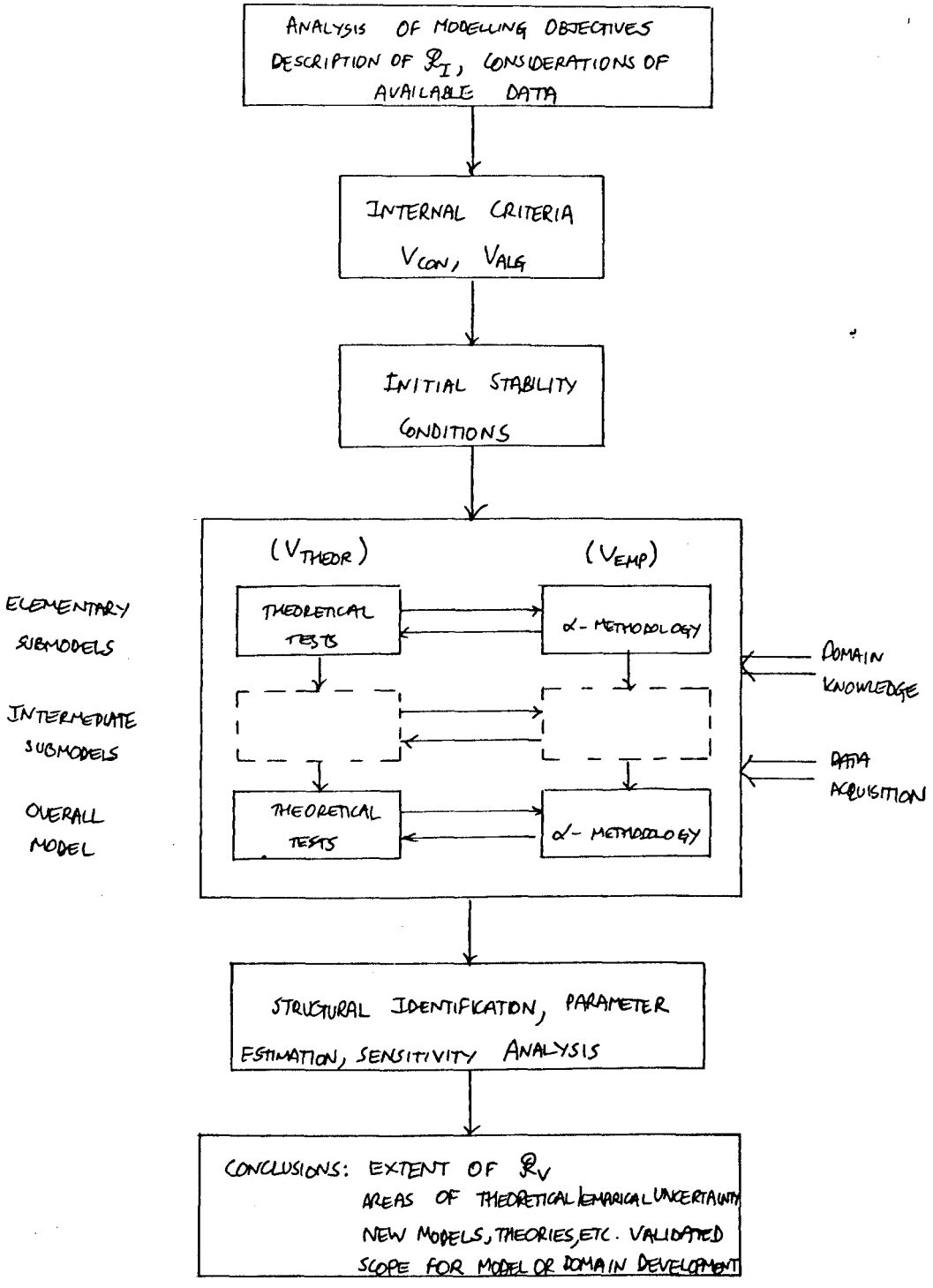


FIGURE 5.10. THE γ- METHODOLOGY FOR COMBINED THEORETICAL - EMPIRICAL VALIDATION.

the use of a model as a test-bed for hypothesis testing. The types of theoretical test suitable include: coherency with theories (when appropriate), critical assessments (e.g. of the kind: "is the structure a sufficiently detailed frame of reference for the functional modality?"), parallels with models or theories in other domains, etc. (If necessary, heuristic criteria may be involved.)

The final level of validation is that of the complete model. Special attention should be paid here to properties that are only emergent at this level, such as stability, control, organisation, and other features. These properties may be associated with the hypothesis that $\mathcal{R}_I$ is a system, or more precisely, a certain type of system (e.g. a control system, or a self-organising system). If they can be validated empirically and theoretically, then it is meaningful to consider $\mathcal{R}_I$ as a system. These considerations answer some difficulties associated with questions of system ontology and epistemology as well as underlying the model-based nature of systems science (see also Section 3.4.5).

5.4.3 Structural identification, parameter estimation, and sensitivity analysis

The final stage of the α-methodology is separated out in the γ-methodology and is used to try to resolve structural and parametric uncertainties in the model. The model must be fairly simple if the techniques of identification and parameter estimation are to work well. The effect of various uncertainties in the model on the overall behaviour can be assessed using the techniques of sensitivity analysis (Section 5.2.4.2).

5.4.4 Concluding the γ-methodology

Firstly, an attempt should be made to delimit $\mathcal{R}_V$, and compare this with $\mathcal{R}_I$. Secondly, the areas of uncertainty (both empirical and theoretical) in the model should be identified. The new submodels which have been adequately validated (e.g. theoretically, and up to "time series comparisons" in the α-methodology) may be put forward for wider acceptance. The model and results of the γ-validation study may suggest directions for future development of the model, new models based on different modelling objectives (hence $\mathcal{R}_I$), or other domain research (e.g. empirical research). These latter aspects may have to be assessed using a heuristic validation methodology (e.g. the ε-methodology, Section 5.6).

5.5 A Validation Methodology for Utilitarian Objectives (δ-Methodology)

The δ-methodology is a general methodology for the validation (and assessment) of models that are intended for primarily utilitarian objectives. Typical application areas are the use of models for prediction to aid decisions (e.g. economic policy decisions), as an educational tool, for improving health-care systems, and in soft-systems methodology. The models may be linguistic or mathematical types. A flow diagram of the methodology is shown in Figure 5.11.

5.5.1 Preliminary considerations and necessary conditions

The first stage of the methodology analyses the modelling objectives. The specific utilitarian objectives determine a system of interest (SOI) in which the model, or results of the model, are to be used. Usually, they also entail scientific objectives, i.e. the representation of an intended range of application $\mathcal{R}_I$ (which may be in SOI). This stage also considers the available data types and model requirements for data in $\mathcal{R}_I$ and SOI.

All models should satisfy the consistency criterion. The algorithmic criteria and initial stability conditions apply only to mathematical models.

5.5.2 Representational validation over $\mathcal{R}_I$

If the model has scientific objectives, then it can be validated over its $\mathcal{R}_I$. The δ-methodology may make use of the β- or γ-methodologies for this purpose, depending on the stage of development of the domain relating to $\mathcal{R}_I$. (If the model is innovative, then the heuristic ε-methodology will be more appropriate, Section 5.6). Typical tests in this stage might be on the predictive validity of the model.

5.5.3 A priori pragmatic validation

The next stage of the δ-methodology is concerned with the critical assessment of the model prior to its actual use in modifying SOI. (In the theory of model validity these were referred to as the V_{PRAG2} criteria; Figure 4.1, Section 4.3.4). Typical considerations may include:

(i) How well does the model act as a device for clarifying the perceptions of the real actors in SOI? (i.e. its function as a rational construction for debate in Checkland's sense)

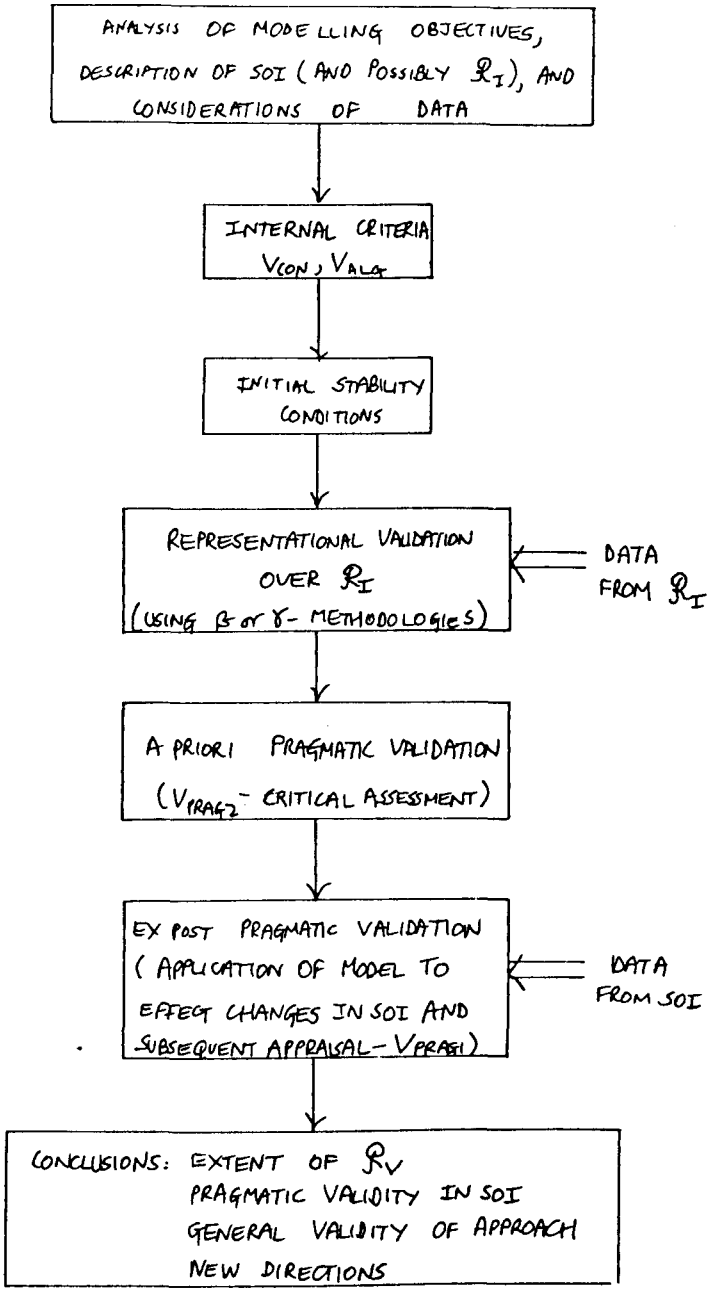


FIGURE 5.11. THE δ-METHODOLOGY FOR MODEL VALIDATION FOR UTILITARIAN OBJECTIVES.

(ii) What range of validity do the additional normativistic models used with the model have? (e.g. decision models, theory of rational choice, uncertainty models of future events).

(iii) What are the social theories implicit in the model or overall methodology? If the approach is based on a misconception of how change can be achieved, any further use of the model will be useless.

It is quite likely that the outcome of the a priori pragmatic validation stage will be negative and that a reverse methodological step will be taken to redefine the problem and modelling objectives. This loop may occur many times before the results of the model are used in practice in SOI.

5.5.4 Ex post pragmatic validation

This stage consists in tests of the model's efficacy after it has been applied to SOI. It may take time before the effects of the decisions, actions, or changes based on the model in SOI are significant, and, therefore, ex post pragmatic validation is a longer-term assessment of the model. (The criteria underlying these tests are V_{PRAG1}, Section 4.3.4.) The changes in SOI must be measured (e.g. utility functions) and then compared with the utilisation objectives. If the utilitarian objectives are satisfied, then it must be assessed whether the relevant changes in SOI are a consequence of employing the model. If so, then the model may be considered to be pragmatically valid. (For complex problems, the techniques of multiobjective-multiattribute utility or value theory may be required. There may also be measurement problems.)

5.5.5 Concluding the δ-methodology

If the model is representational, then the extent of $\mathcal{R}_V$ should be delimited, and the implications that this has on the pragmatic validity of the model in SOI should be made. The results of the application of the model (i.e. how well it actually performed) must be assessed pragmatically in relation to the modelling objectives. It may be possible to draw some conclusions on the general validity of the model (and associated approach) for dealing with problems of a certain type in SOI (based on the a priori as well as ex post pragmatic validation). Finally, some suggestions for new research directions (new problem types, methodologies, models, objectives, etc.) may be made.

5.6 A Validation Methodology for Innovative Models (ε-Methodology)

This methodology is intended to provide some kind of framework for the validation of highly innovative models. "Validation" in this context includes a wide range of possible tests from judgmental decisions to empirical comparisons. The domains in which such models might arise are those in which there is a lack of theory (in general, or of the type involved in the model) and where data types may be limited or problematic. Quite often the models are borrowed from more developed domains on the basis of some isomorphic features between different types of phenomena - "analogical construct models" (Leaning and Webb, 1980), or "paramorphs" (Harré, 1970). The novelty in such models means that they are not amenable to the β or γ validation methodologies (Sections 5.3 and 5.4), yet there is a great need for an approach to assessing these models on a different level (in terms of scientific potential or heuristics) and to show how they might eventually be validated in a γ-methodology, for instance. Consequently, the ε-methodology may also be described as a methodology for "heuristic validation".

Typical examples of this type of model are: the state space (or system dynamics) models used in world modelling (Meadows et al., 1972), and in the modelling of bicommunal conflict (Mitchell, Bowers, Webb; 1978, 1979, 1980). The ε-methodology which is shown in Figure 5.12 is not as prescriptive as the methodologies described previously, rather it indicates the range of possibilities for heuristic validation.

5.6.1 Preliminary considerations and necessary conditions

The analysis of modelling objectives should consider the general as well as specific objectives, e.g. the type of scientific development associated with the use of the model in this domain (theory development, problem shifts, empirical research, etc.). Since the domain may be in a formative or developing stage, the determination of the intended range of application $\mathcal{D}_I$ from the specific objectives may be fuzzy. The preliminary considerations should also take account of the nature of available data types and assess whether they are applicable to the model.

The necessary conditions are the internal consistency and algorithmic criteria, and the initial stability condition. The latter should not be interpreted as advocating a conservative or regularative theory, but simply the requirement that a model should not be wildly unstable at all times.

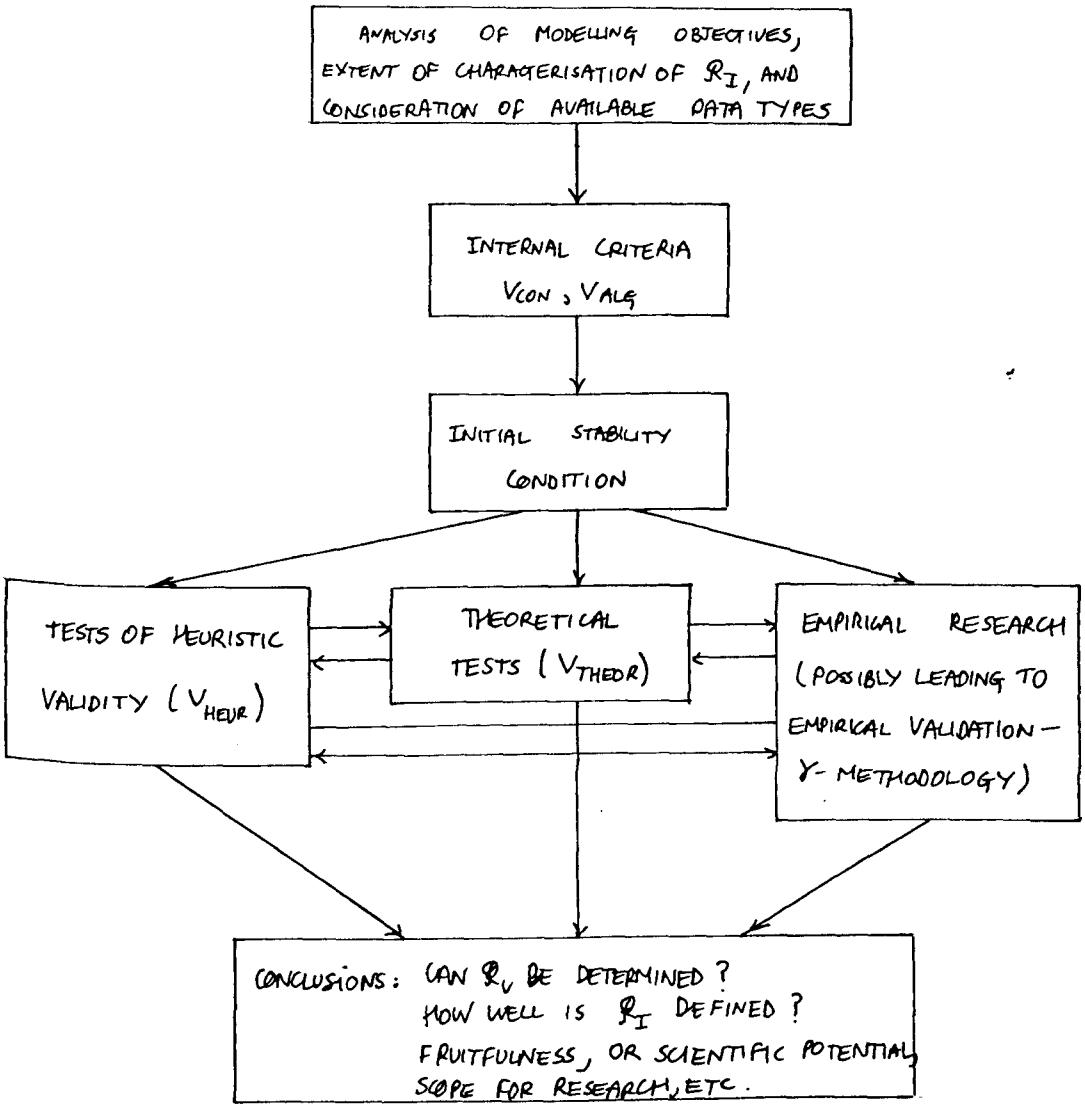


FIGURE 5.12. THE ϵ- METHODOLOGY FOR THE VALIDATION OF INNOVATIVE MODELS.

5.6.2 Tests of heuristic validity

The nature of heuristic validity criteria and some examples were discussed in Section 4.3.3.5. In general, heuristic validity is related to the scope or potential of a model for scientific understanding and discovery. Some tests of heuristic validity are:

(i) Does the model resolve a previous problem (e.g. a theoretical anomaly, or conflicting theories)?

(ii) Does the model suggest new, possibly more fruitful, modelling objectives?

(iii) Does the model extend $\mathfrak{L}_I$ (compared with earlier models)? Is this new $\mathfrak{L}_I$ more amenable to empirical representation and validation?

(iv) Does the model convey a satisfactory understanding of $\mathfrak{L}_I$?

(v) What potential or scope does the model offer for future research?

This type of judgement on a model is essentially concerned with "good reasoning patterns" in science and there may be very many acceptable criteria.

5.6.3 Theoretical tests

Coherency tests of theoretical validity may be made if appropriate theories are available.

5.6.4 Empirical research

If the model is not comparable with available data types in a satisfactory manner, a programme of empirical research may be started which can have several aims:

(i) The acquisition of more data of a previously existing type.

(ii) The definition of new data types, and possibly empirical attributes based on the model, and the realisation in practical observational or measurement techniques.

(iii) The invention of new empirical representation devices.

The empirical research programme is simply a means of finding out more empirical information about the phenomena of interest, which is structured by the model (or a series of developing models). If an

adequate amount of data is obtained, it may then be possible to enter the α-methodology (Section 5.2), although this is envisaged as happening at a much later stage.

5.6.5 Concluding the ε-methodology

The ε-methodology is such an integral part of the overall development process of innovative models that, unlike previous methodologies, it is never really concluded. However, from time to time summaries may be produced containing the following types of point:

(i) Is it possible to determine the empirical range of validity $\mathcal{R}_V$ for the current model M?

(ii) The best characterisation of $\mathcal{R}_I$ at the present time.

(iii) Is $\mathcal{R}_I$ defined better with M?

(iv) The theoretical developments (and new understanding) embodied by M.

(v) The future directions for research which offer the most scope or potential for scientific advance in this area.

5.7 Conclusions

Four methodologies, or programmes, for model validation suitable for different types of models and research areas have been considered in this chapter: the β-methodology (Section 5.3) which is empirically-based and intended for models formulated using well-validated theory and data (e.g. in physical modelling); the γ-methodology (Section 5.4) in which theoretical and empirical validity criteria are jointly applied and is suitable for models which contain a limited amount of theoretical and empirical innovation (e.g. in biological modelling); the δ-methodology (Section 5.5) which is a pragmatic approach to the validation of models that have primarily utilitarian objectives (e.g. models involved in technological system design and optimisation, or in soft-systems methodology); and the ε-methodology (Section 5.6) which is a mainly heuristic programme for the validation of models that are highly innovative (e.g. dynamic models of political conflict, or analogical construct models in general). In addition, an extensive methodology (the α-methodology, Section 5.2) for the systematic comparison of a model with empirical data was described that also includes feature space comparisons.

The methodologies presented in this chapter are based on the conceptual framework of the theory of model validity developed in Chapter 4, and their wide range and detail may be taken, therefore, as evidence of the general applicability of the theory. However, the ultimate test of the theory, or these methodologies, is in the practical application to the validation of a particular model. In the next chapter, the theory of model validity is used as a conceptual framework to structure the validation of a mathematical model of the human cardiovascular system, in which extensive use is made of the α-methodology. The three subsequent chapters (7, 8 and 9) illustrate the use of the theory and other methodologies developed in this chapter in validating two other biological models and in considering some aspects of model validity and validation in the social sciences.

CHAPTER 6

FIRST CASE STUDY - VALIDATION OF A MATHEMATICAL MODEL OF THE HUMAN CARDIOVASCULAR SYSTEM.

6.1 Introduction.

In this first case study, the validity of a complex dynamic mathematical model of the human cardiovascular system (CVS) will be examined. The model was developed by Pullen (1976) in the Department of Systems Science with the specific objectives of studying short term haemodynamics and predicting the effects of rapid cardiovascular active drugs. In this examination extensive use will be made of the theory of model validity (Chapter 4) and many of the validation techniques described in Chapter 5. The structure of the Chapter is as follows:

Firstly, the paradigm of cardiovascular modelling is demonstrated in an historical review of earlier models (Section 6.2) followed by an outline of the Pullen model (Section 6.3). A programme, or methodology for the validation is then developed based on considerations of the modelling objectives and the nature of human CVS data (Section 6.4). The bulk of the Chapter is the validation programme (Section 6.5). Finally, the results of the study are summarised by way of the questtions, "does the model satisfy its objectives?" and, more specifically, "what is the range of validity of the model?" (Section 6.6).

This Chapter is intended to be a full, detailed validation study of a complex biological model, not merely highlights from such a study. It is therefore rather long. The case studies in the following chapters will be much shorter, but it is intended that the structure of this chapter could act as a template for detailed studies of these models and other biological models if desired. For the benefit of non-physiologists the following brief introduction to cardiovascular physiology sets the scene.

6.1.1 Introductory cardiovascular physiology.

The cardiovascular system comprises the heart and blood vessels (and also some parts of the peripheral and central nervous system). It is a general transport system in which blood circulates through the

body distributing to the tissues oxygen from the lungs, certain products of metabolism and substances absorbed from the digestive tract, CO_2 to the lungs and other waste products to the kidneys and liver and also acts as a communication channel for hormonal control. It also plays a role in regulating body temperature as the amount of heat lost from the surface of the body is related to the blood flow through the skin.

It is essential for survival that certain organs (such as the brain) receive oxygen and nutrients at a steady rate; at the same time other tissues' metabolic requirements change widely as the body constantly changes its relationship with the environment. This entails a system that is both tightly controlled and highly flexible, or adaptable.

The cardiovascular system is richly endowed with nerves that emanate from the medulla oblongata in the brain stem at the top of the spinal cord. Changes in the activity of this region cause rapid changes in the properties of the blood vessels and the heart, and neural mechanisms therefore play an important role in the short term control of the CVS. At certain key sites in the blood vessels there are neural sensors sensitive to blood pressure, the outputs of which modify the activity of the medulla oblongata, which in turn modifies CVS parameters in a negative feedback control system with a loop delay of 1-10 seconds. This system acts to maintain arterial pressure and hence blood flows at an acceptable level. Changes of concentrations of O_2 and CO_2, particularly in the brain, also effect the medulla, and the CVS responds quickly to counteract disturbances in these variables.

There are chemical messengers (hormones) in the blood which affect many cardiovascular parameters (such as heart rate, resistance to blood flow, the loss of water and ions through the kidneys, etc.). These hormones have slower dynamics (30secs. - several hours) than the neural mechanisms and play an essential role in the medium term control of the CVS. The levels of hormones are sensitive to many factors such as blood pressure, blood volume, the concentration of ions in the blood and urine, local metabolic requirements, and activity of the medulla for example. The medium term control is therefore complex with many interacting loops. The neural control tends to act as a fast controller of arterial pressure, whereas the medium term hormonal control essentially adapts the CVS to cope with environmental changes.

In the human there are also long term control or adaptive effects that occur during prolonged disturbance to the CVS. These may take place over a period of days to years and typically the properties of heart and blood vessel muscles and the pressure sensors are affected, often permanently.

Mathematical models have been used in cardiovascular research for many years, both for studying detailed aspects of the individual components (such as the heart or arterial tree) but also for understanding the overall organisation and control of the cardiovascular system. The Pullen model is of the latter type and is concerned with the short term neural control and the effects of drugs on the cardiovascular system.

6.2. Modelling the Human Cardiovascular System.

6.2.1. Introduction.

This section provides a brief historical review of some mathematical models used in cardiovascular research. The common feature of these models is their mathematical approach to the cardiovascular system as a controlled system. The role of this review is two-fold: firstly, to serve as an introduction to the description of the Pullen model (Section 6.3); and, secondly, to demonstrate that a paradigm, research programme, or scientific domain, has developed in this area. (The latter provides an escape clause in that it allows one to say that model M has been "fruitful" in contributing to the evolution of research programme R when, perhaps, it fails reasonable tests of theoretical and empirical (representational) validity. In other words, that the model has heuristic validity or potential.) For some detailed reviews of such models see Beneken (1972), Talbot and Gessner (1973), or Pullen (1970 and 1976).

6.2.2. Historical background

William Harvey in the 17th century inferred by quantitative argument and physiological demonstration that the blood circulates from the heart to the arteries is collected by the veins and returns to the heart. Harvey's "proof" is set out in his book "De Motu Cordis et Sanguinis in Animalibus" (1628) and constitutes a lucid validation programme which is analysed in appendix I. It is said that this marks

the beginning of modern medical science, (e.g. Keele, 1976) and it is certainly of great historical importance to this work, and in particular this case study.

Two properties of the heart are necessary in order to understand the early models. Firstly, a property that has been well known for many years, that heart muscle has a natural rhythmicity and, given a supply of oxygenated blood, will spontaneously and regularly beat, pumping blood. Secondly, the famous "Law of the Heart" first formulated by Starling (1866 - 1927). This law states that, within physiological limits, the external stroke work done by the heart is proportional to end-diastolic ventricular volume. The direct consequence of this law is that the heart will automatically balance cardiac output with venous return. These properties may be empirically validated using a "heart-lung preparation" (see, e.g., Lippold and Winton, p.225) which demonstrates that the heart and circulatory system constitute a self-innervating auto-regulative system. If this system is in a constant environment it will be stable.

6.2.3. Models of the auto-controlled cardiovascular system.

These models are based on the observations of Section 6.2.2, that the heart and circulation form a stable system independent of any other form of control (e.g. neural or hormonal). They therefore describe an instantaneous steady state behaviour of the cardiovascular system. Although the auto-controlled cardiovascular system can attain a steady-state equilibrium in a constant environment, the lack of neural or hormonal control in the classical reflex or servo loops causes it to be usually known as the "uncontrolled" cardiovascular system.

In 1955, Guyton proposed a graphical method for the determination of cardiac output. In this method, curves of venous return (VR) and cardiac output (CO) are drawn against a single independent variable, right atrial pressure, (cardiac function curves). The operating point is determined from the intersection of the VR and CO curves, thereby satisfying Starling's law. Guyton's model describes the steady state conditions that the system might achieve given limited environmental disturbances.

The next development of interest is the use of "compartmental" models in cardiovascular modelling. Grodin's model (1959) is based on Starling's law of the heart in which cardiac activity is represented by a linear relationship between end-diastolic ventricular volume and stroke work. The model consists of twenty-three simultaneous equations which must be solved in order to determine the equilibrium values.

Dick and Rideout (1965) developed a compartmental model which has four segments representing the major divisions of the arterial tree, and in which the pumping of the heart is represented by a time-varying compliance of the left ventricle compartment. Beneken (1965) simulated a similar 8-segment model of the uncontrolled circulation on an analogue computer, which included time-varying compliances of both ventricular segments. The atria are lumped with the preceding venous segments.

6.2.4. Models of the neural controlled cardiovascular system.

Whilst the cardiovascular system is stable in an unchanging environment, its stability from moment to moment in a changing environment depends upon a rapid neural control of the heart and blood vessels by the central nervous system. (Historically, the role of the vagus on the heart rate was discovered as early as 1806 by Cyon and Ludwig).

Beneken and De Wit (1967) produced a nineteen segment model of the cardiovascular system. In this model, all four heart compartments have time-varying compliances, and the neural control of heart rate, myocardial contractility, peripheral resistance and venous tone is included. The models of the baroreceptors (which send pressure information to the CNS) are based on Katona's empirical studies on dogs (1965, 1967). The model's behaviour agrees reasonably with standard circulatory responses over short-time periods. Hyndman (1970) developed a "bang-bang" model of neural control in a study of cardiac arrhythmias.

Beneken and Rideout (1968) demonstrated that the transport of a neutral substance in the bloodstream could be modelled by coupling a "slave" model to the blood flow model in a technique known as "multiple modelling". Pullen's model (1976) is based largely on Beneken and De Wit's model together with Hyndman's models of neural control, and the multiple modelling technique (to model drug transport). The method of modelling of local drug effects was introduced by Pullen.

6.2.5. Models of the ultrastable cardiovascular system.

The models of Section 6.2.4. represent, with varying degrees of validity, the major aspects of short-term cardiovascular dynamics with a limited number of environmental changes (those affecting haemodynamic variables). In life, the human cardiovascular system enjoys a stable relationship with a changing environment that entails many other effects which cause changes in the cardiovascular system. These include fast-acting chemical effects, medium-term hormonal changes and longer term fluidic effects and disease processes, for example. Under these conditions the cardiovascular system can be considered to be a self-adaptive control system or, in Ashby's terminology (1956), it is ultrastable.

Guyton's models with Coleman(1967) and Coleman and Granger (1972) are concerned with the overall regulation of the cardiovascular system at this level. The component submodels are usually simple empirical models, and include Guyton's 1955 model of the heart. A recent model which has a similar structure and range of validity is Uttamsingh's model of the human renal - artificial kidney machine system (1981), and a validation study is made of this model in the next chapter. A problem with this type of model (containing many assumptions and empirical models) is that it frequently produces valid overall behaviour whilst containing anomalous invalid behaviours of its submodels.

6.2.6. Other aspects of cardiovascular dynamics.

For an illustration of mathematical models concerned with detailed aspects of the cardiovascular system (rather than as a controlled system) see Bergel (1972). Ohley et al. (1980) describe a validation study of an arterial system model which is based on a finite difference solution to the Navier - Stokes equation, (with experimental data from dog experiments).

6.2.7. Some recent work in the Department of Systems Science.

Rajkumar (1978) investigated the possibility of using model reduction techniques to reduce the Pullen model to a more manageable size. He found that although such techniques are not generally appropriate to such a complex model, it was possible to obtain some improvements in simulation efficiency by increasing the integration step size, applying dynamic reduction techniques to specific areas (such as the

baroreceptor submodels), and by a limited aggregation of compartments. At present Al-Dahan (1979) is continuing the work in three directions: firstly, further validation tests, including detailed parameter sensitivity studies, which have been completed (see Section 6.5.7.1.2.4). secondly, using the model for drug studies (contingent upon the validity of the model); and, thirdly, investigating the feasibility of developing a set of smaller models for specific clinical purposes. These models will have a longer time scale than the Pullen model and therefore include other control mechanisms.

6.3. An Outline of the Pullen Model.

6.3.1. Introduction.

Pullen (1976) describes the central objective of his work as "to produce a pulsatile mathematical model and computer simulation of the controlled cardiovascular system of a normal, resting, conscious, average human suitable for the study of short-term haemodynamics.

The aim was to make a model sufficiently detailed and comprehensive for the study of short-term pharmacokinetics (i.e. drug effects with the major dynamics complete within 2 or 3 minutes) and to use the model to study the overall effects of an injected drug assuming a number of simultaneous actions of the circulating drug at specific sites." (p. 19).

The model is based on the circulatory fluid mechanics model of Beneken and De Wit (1967), with the baroreceptor and neural control models of Katona (1967) and Hyndman (1970). The pharmacodynamic modelling is based on the multiple modelling technique of Beneken and Rideout (1970) for the distribution of a substance in the bloodstream, with Pullen's own method for integrating local drug effects.

For the full detail of the model, its assumptions etc. consult Pullen (1976, Chapters 2,3 and 4), or for an outline that provides slightly more information than here, refer to Al-Dahan et al. (1979). The model has three distinct submodels: a submodel of circulatory fluid mechanics and the heart, a submodel of neural control, and a submodel for pharmacodynamics. These are described separately below. (A full listing of the mathematical model is given in appendix II).

6.3.2. Submodel of circulatory fluid mechanics and the heart (or uncontrolled haemodynamics).

Blood flow through the complex network of vessels in the circulatory system is modelled by the 19 - segment model of Beneken and De Wit (1967) (Fig. 6.1). Each segment is an elastic reservoir with lumped hydrodynamic parameters representing the distributed properties of the appropriate collection of blood vessels (as can be seen from Fig. 6.1, the segments do have significance as physiological partitions). The pumping action of the 4 heart chambers is achieved using actively contracting elastic reservoirs with time varying elastances, and the timing of events within each cardiac cycle (e.g. the duration of ventricular systole) is related to heart period by linear approximations derived by Beneken and De Wit (1967).

General equations characterising a typical segment may be derived by considering two typical segments connected together as shown in Fig 6.2. The static pressure-volume curve of a typical lumped parameter segment is approximated as being linear in the normal operating range. In passive elements, where the compliance (C_1) can be considered to be constant, transmural pressure (P_1) and volume (V_1) are related by the equation

$$P_1 = (V_1 - V_{ul})/ C_1 , \quad V_1 \geqslant V_{ul} \qquad \text{-----(6.1)}$$

where V_{ul} is the unstressed volume. The flow (F_{12}) through the viscous resistance (R_{12}) between the two segments is

$$F_{12} = (P_1 - P_2)/ R_{12} \qquad \text{-----(6.2)}$$

and, from continuity,

$$\frac{dV_1}{dt} = F_{01} - F_{12} \qquad \text{-----(6.3)}$$

In the systemic arteries, inertial effects, wall viscosity, and geometrical and elastic taper are taken into account. Typical equations for an arterial segment are

$$P_1 - P_2 = R_{12} F_{12} + L_{12} \frac{dF_{12}}{dt} \qquad \text{-----(6.4)}$$

$$\frac{dV_2}{dt} = F_{12} - F_{23} , \quad V_2 \geqslant 0 \qquad \text{-----(6.5)}$$

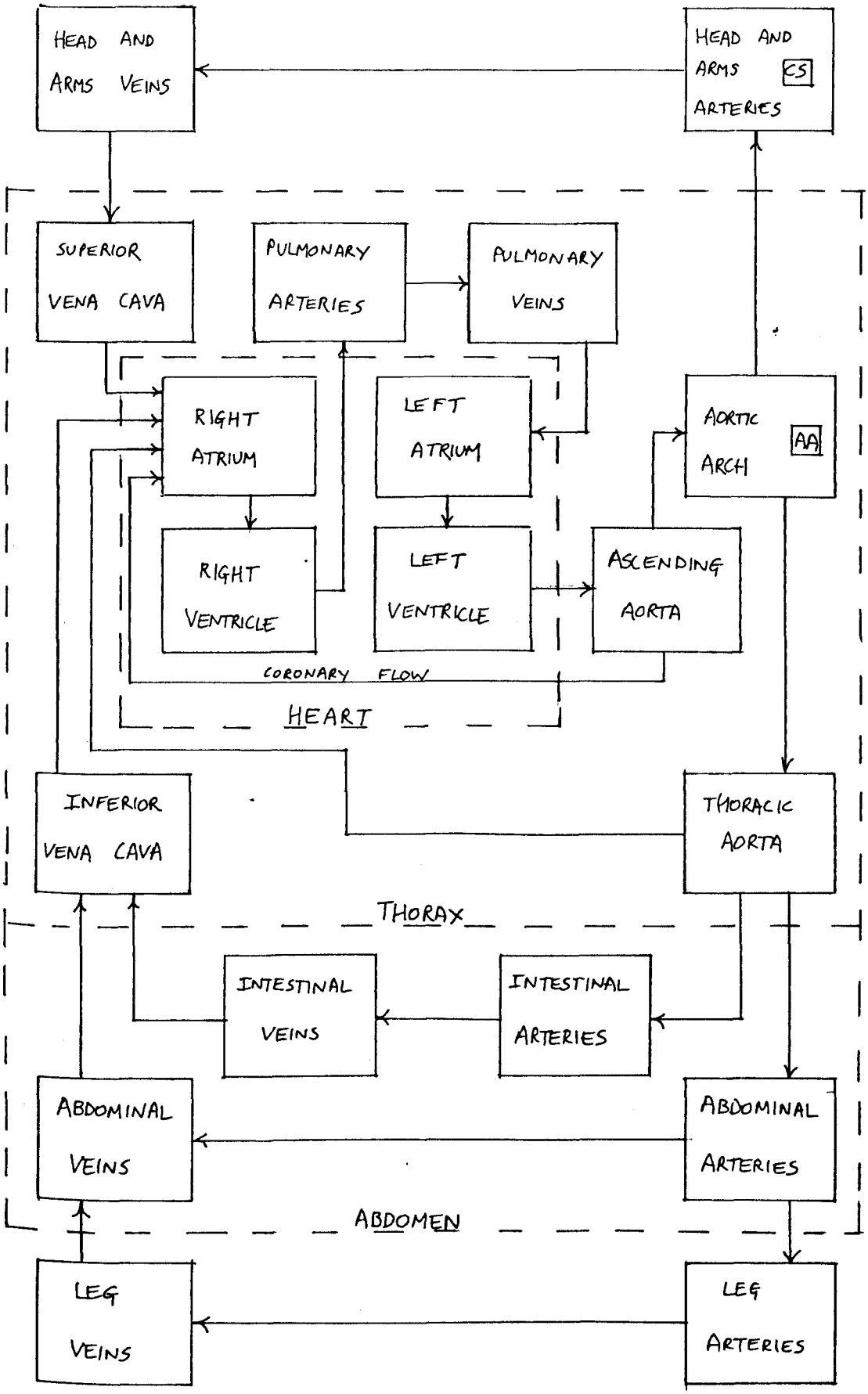


FIGURE 6.1. COMPARTMENTS AND STRUCTURE OF THE PULLEN SUBMODEL OF CIRCULATORY FLUID MECHANICS AND THE HEART (BASED ON BENEKEN & DEWIT, 1967). (AA,CS = AORTIC ARCH AND CAROTID SINUS BARORECEPTORS).

$$P_2 = \frac{1}{C_2}(V_2 - V_{u2}) + R_2' \frac{dV_2}{dt} \qquad \text{-----(6.6)}$$

where L_{12} represents the inertia of the blood, and R_2' is equivalent to wall viscosity.

Equations for the outflow through the ventricular valves also include terms for the inertial and viscous properties of the blood, and a non-linear term representing the pressure drop across the valve.

The resistances of the vascular beds (arterioles, capillaries and venules) are represented by lumped arterio-venous resistances. In modelling the veins non-linearities arise due to the collapsible nature of the veins, and the prescence of venous valves.

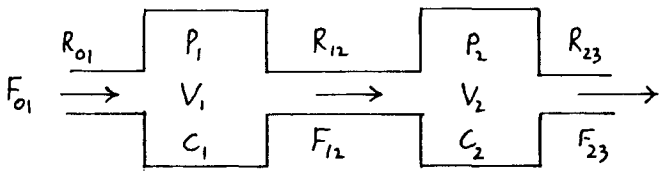
There are also facilities in the model for including the effects of respiration (pressure effects) and orthostasis (postural changes), and for calculating mean arterial pressure, stroke volume, cardiac output and estimated total systemic resistance on a beat by beat basis.

6.3.3. Submodel for neural control.

This submodel consists of models of baroreceptors (which monitor blood pressure and send information to the CNS) and models of the CNS (central nervous system) control which, acting on the baroreceptor outputs, effect changes in certain parameters or variables in the submodel of uncontrolled haemodynamics (Section 6.3.2.)

6.3.3.1 Baroreceptors.

The main baroreceptors for cardiovascular control are located in the walls of the aortic arch and at the division of the carotid artery. Their locations in the model are in the aortic arch and head and arms segments (Fig. 6.1). The models of both baroreceptors are identical and based on Katona's empirical models (1967). A block diagram of an individual baroreceptor model is shown in Fig. 6.3. The baroreceptor output (B) is a linear combination of a dynamic mean pressure estimate (S_B), and a weighted dynamic average (S_C) of the positive rate of change of pressure (S_A), together with a threshold pressure below which firing of the baroreceptor does not occur. The effective input to the CNS is assumed to be a static linear combination of the outputs of the aortic arch and carotid sinus baroreceptors (Fig. 6.4)



F = FLOW RATE ($ML\ SEC^{-1}$)

P = PRESSURE (MM HG)

R = RESISTANCE ($MM\ HG\ SEC\ ML^{-1}$)

V = VOLUME (ML)

C = COMPLIANCE ($ML\ MM\ HG^{-1}$)

FIGURE 6.2. TWO TYPICAL SEGMENTS.

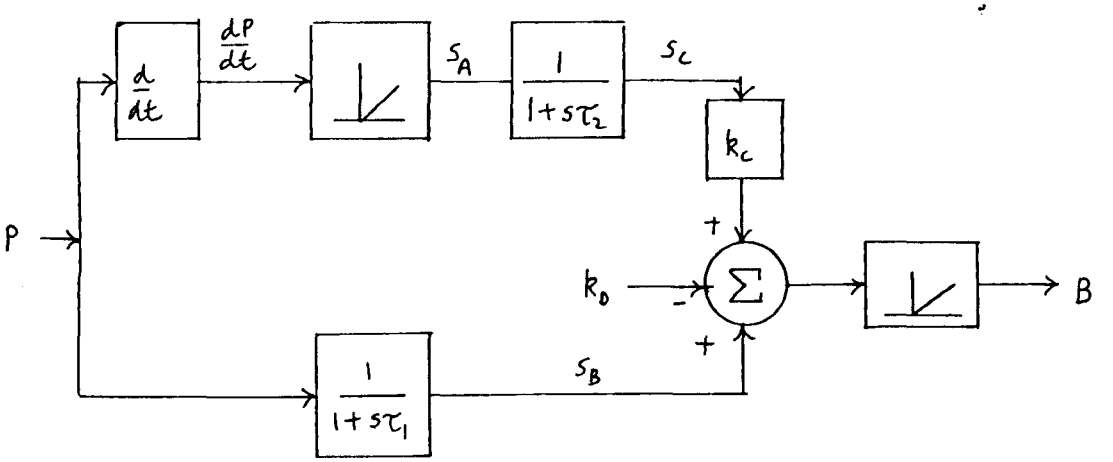


FIGURE 6.3. BLOCK DIAGRAM OF AN INDIVIDUAL BARORECEPTOR MODEL.

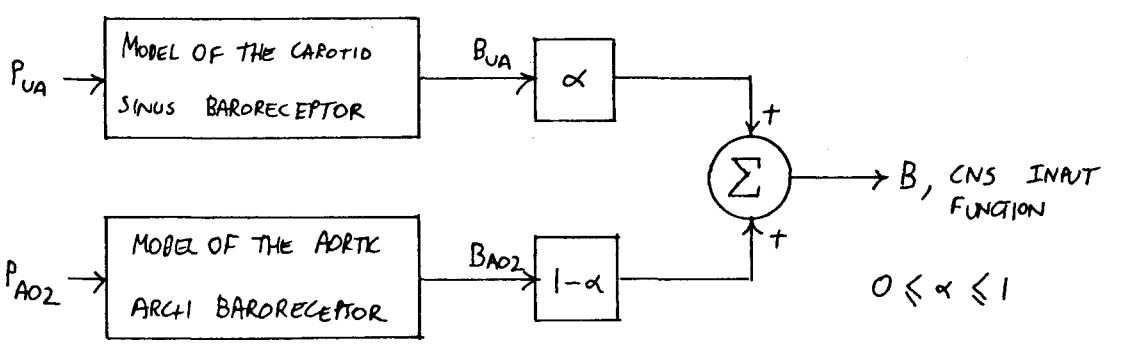


FIGURE 6.4. LINEAR COMBINATION OF OUTPUTS OF AORTIC ARCH AND CAROTID SINUS BARORECEPTORS TO FORM CNS INPUT FUNCTION (B).

6.3.3.2. CNS control of heart rate.

A dynamic version of Katona's two region model (1967) for CNS control of heart rate is adopted. The two regions are for blood pressure greater than and less than normal. The output of the controller sets the heart period for the next cardiac cycle, and is constrained so that heart rate f_H , $30 \leqslant f_H \leqslant 200$ b.p.m.

6.3.3.3. CNS control of myocardial contractility, peripheral resistance, and venous tone.

The separate CNS controllers of myocardility contractility (the force of contraction of the heart), peripheral resistance, and venous tone are all based on Hyndman's "bang-bang" models of CNS control (1970). In each, a dimensionless multiplier modifies the appropriate parameters, (the maximum change is $\pm$ 40%). For peripheral resistance, the parameters are all arterio-venous resistances (with the exception of head and arms, coronary, and pulmonary resistances); and for venous tone, the parameters are the compliances and unstressed volumes of the venous segments.

6.3.4. Submodel for pharmacodynamics.

The "multiple modelling" technique of Beneken and Rideout (1970) is used to represent the transport of a single chemical substance in the bloodstream. A slave 19-segment model is coupled to the main 19-segment blood circulation model so that, for each segment, transport flow is proportional to concentration in the transport model multiplied by blood volume flow. Typical equations for segment 1 shown in Fig 6.2 are

$$\text{Concentration} = m_1/V_1 = w_1 \qquad \text{----(6.7)}$$

$$\text{Mass inflow} = w_{01} F_{01}, \quad w_{01} = w_0, \ F_{01} > 0; \quad w_{01} = w_1, \ F_{01} \leqslant 0 \qquad \text{----(6.8)}$$

$$\text{Mass outflow} = w_{12} F_{12}, \quad w_{12} = w_1, \ F_{12} > 0; \quad w_{12} = w_2, \ F_{12} \leqslant 0 \qquad \text{----(6.9)}$$

The rate of change of mass in segment 1 is given by

$$\frac{dm_1}{dt} = w_{01} F_{01} - w_{12} F_{12} \qquad \text{----(6.10)}$$

Drug injection (intravenous) is simulated by instantaneously increasing the drug mass in the appropriate venous segment at the time of the injection. Drug disposal (absorption, breakdown, decay) is modelled as a first order linear mass decay process in all compartments (time constant $\simeq$ 30sec.).

Local effects of the drug are modelled algebraically using a method developed by Pullen (1976, pp. 78 - 83). The parameters in each segment, which are affected by the drug, are multiplied by a dimensionless variable which is a linear function of the drug concentration in that segment. The parameters, and the drug concentrations upon which they are dependent, are shown in table 6.1.

Region	Parameter	Local Effect	Segment
Heart	Heart rate	↑ = tachycardia ↓ = bradycardia	right atrium
	Systolic elastances	↑ = +ve inotropy ↓ = -ve inotropy	all 4 heart chambers
Systemic circulation	Arterio-venous resistances	↑ = vasoconstriction ↓ = vasodilatation	thoracic, intestinal, abdominal & leg arterial segments
	Venous unstressed volumes	↑ = venodilatation ↓ = venoconstriction	all venous segments
	Venous compliances	↑ = venoconstriction ↓ = venodilatation	

Table 6.1 Effects of Drugs on Model Parameters

For each local effect, it is necessary to specify the direction of drug-induced change (as shown in the table).

6.3.5. Interactions in the Complete Model.

The overall structure of Pullen's model, and the various interactions are illustrated in Fig. 6.5. The complete mathematical model consists of the following number of equations :

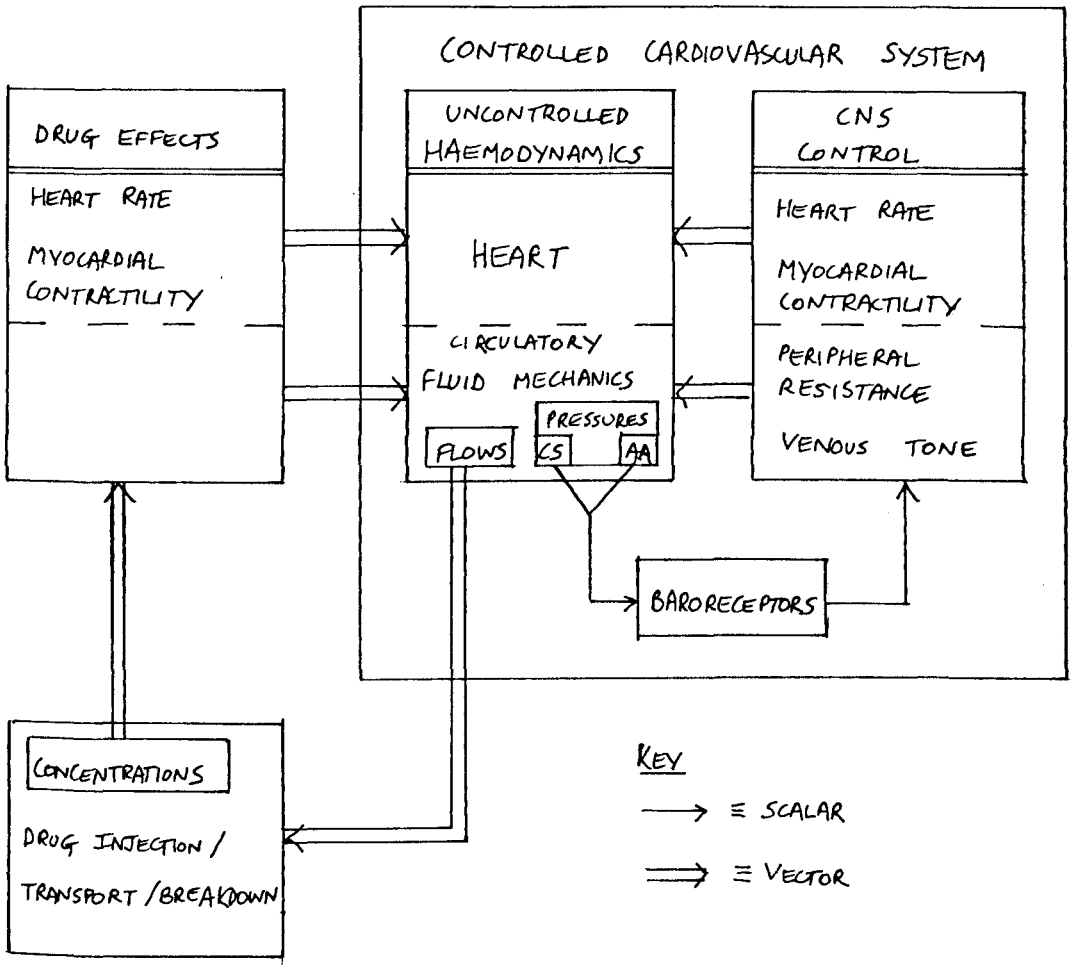


FIGURE 6.5. OVERALL STRUCTURE AND INTERACTIONS IN THE PULLEN MODEL.

Submodel	No. of 1st order differential equations	No of algebraic equations
Circulatory fluid mechanics and heart	31	83
Neural control	11	23
Pharmacodynamics	19	53
Controlled CVS	42	106
Controlled CVS + pharmacodynamics	61	159

The complete model has 178 parameters. The simulation model developed by Pullen (and also used in the validation studies of Section 6.5) to simulate the mathematical model is written in FORTRAN IV. Integration is tackled using Euler's method with a fixed step length of 0.5 msec. Simulation times for the model run on a CDC 7600 are:

Real time (sec)	Model Order	Simulation time (sec)
100	42	39
100	61	56

Rajkumar (1978) showed that the simulation time for the 42nd order model reduced to 15 seconds without appreciable loss of accuracy.

6.4. Development of a Programme of Validation.

6.4.1. Introduction.

The theory of model validity described in Chapter 4 does not offer a unique prescriptive methodology for model validation. Instead, it provides a general analytical framework which can be used to structure the validation of a specific model. The structure will depend on the objectives or purposes of the model, considerations of available data, types (or system data) and the consequent appropriate validity criteria; in other words, the validation methodology is problem-dependent. These three aspects will be considered in this section with reference to the Pullen model. It turns out that the model is logically decomposable in a manner which allows the validation programme to be structured in an elegant hierarchical fashion. An outline of this hierarchical structure is given in Section 6.4.5., together with some specific techniques for validation taken from Chapter 5.

6.4.2. Modelling objectives.

General objectives of Pullen's model include: an increased understanding of cardiovascular physiology; the development of mathematical modelling methodologies in systems science in general, and in biology in particular; and, hopefully, some lessons for systems science and cybernetics on the nature of control in a biological system. These are all obvious and deserve little comment; consideration of whether the Pullen model satisfies any of these objectives will not form part of the validation programme per se, but will be given at the end of this chapter.

The specific objectives are more important in validation because they determine the intended range of application of the model and to a large extent the form of the validity criteria. In Section 6.3.1., the specific objectives of Pullen's model were given as the development of a pulsatile mathematical model of the short-term (2-3 mins) haemodynamics in a normal, resting, conscious human and the prediction of the effects of fast-acting cardiovascular drugs. (Note that the modelling objectives, general and specific, are scientific, not utilitarian or pragmatic). The intended range of application is represented bimodally in a physical (or structural) modality and a functional modality, (for a full treatment see Section 4.3.1.3). The physical modality consists of a description of the elements or parts of the range of application and their geometrical and topological relations, and is not related to behaviour, (i.e. in biological terms, anatomy). The functional modality describes the range of application in terms of its behaviour, functioning, functional properties, dynamics etc., and is represented theoretically by the appropriate laws, theories, models, etc. and empirically by measurements of functional properties. It is important to realise that the structural modality is simply an expression of "what there is," whereas the functional modality consists of available data types plus theories (which are closely related, of course).

For the Pullen model of the human cardiovascular system, the extent of two modalities of the intended range of application are shown in table 6.2.

Modality	Description or Representation
Physical/ structural	heart, blood vessels(and certain anatomical groupings thereof), blood, circulating drugs, baroreceptors (haemodynamic/neural transducers), afferent nerves, CNS (medulla oblongata), efferent nerves (vagus and sympathetic), nerve endings (neural/haemodynamic transducers). general geometric properties and topological features.
Functional Available data types (empirical properties)	static properties, i.e. physical properties such as elastances, inertias, drug effects, etc. dynamic attributes, i.e. volumes, pressures, flow rates, drug concentrations, CNS input and output activities, etc.
Theoretical	Various models, theories, hypotheses of short-term cardiovascular behaviour, e.g. arterial and venous blood flow dynamics, theories of heart action (mechanisms, timings, Starling's "law of the heart", etc.), short-term cardiovascular control theories (based on pressure receptors and neural loops), hormonal control theory, receptor models of drug action, etc.

Table 6.2. Structural and Functional Modalities of the Intended Range of Application of the Pullen Model.

To satisfy its objectives, and therefore be valid, the model should represent these two modalities over a period of two to three minutes for a normal, resting, conscious, young male human.

Most mathematical models in biology involve gross aggregation and idealisation of structure because of the myriad complexity of living organisms, and this is true of Pullen's model. However, the main concern of Pullen's model (and other models) is with the explanation and representation of behaviour at a certain level. Thus model validity is defined largely with respect to the functional modality. Since the structural modality acts as a frame of reference for the functional modality this implies that the structure should be an adequate approximation for the aspects of behaviour of interest in the range of application. When comparing the model's behaviour with normal

human cardiovascular data, the model response should lie within the normal human ranges, or data (metrical) uncertainty. Considerations about data types are made in Section 6.4.3.

6.4.3. The nature of data available from the human cardiovascular system.

Even within this limited domain of cardiovascular physiology, the subject of measurement and observation is a very large one. In this section some fairly general remarks will be made with the emphasis on what aspects are measurable and some data uncertainty problems. As is often the case with biological modelling, short time scales make the acquisition of dynamic data harder rather than easier.

6.4.3.1. Some historical aspects.

By the end of the seventeenth century detailed and accurate observations had been made of the physical structure of the heart and of the networks of arteries and veins (including the capillary link between arterioles and venules). The functional description perhaps began with Harvey (1628) who estimated the daily blood flow out of the heart. Stephen Hales (1677 - 1761) was an experimentalist who made extensive measurements of pressure in man and animals. The physical structure of the mesenteric vascular bed of the dog was meticulously recorded by Mall (1888) who measured not only the diameter of each type of vessel, but the number, cross-sectional area, length and total volume. Unfortunately such a study has not been repeated, and for a human models of the vascular bed are based on percentage distribution of cardiac output to various regions and organs and the dimensions of the major arteries and veins (e.g. Beneken, 1965). Neural and drug aspects were not studied in detail until this century.

6.4.3.2. Presnt day measurement techniques.

For a detailed study consult Cobbold (1974); Hawker (1979) gives a good general review. Heart rate and rhythm are recorded accurately using an ECG. Blood pressure is measured routinely using a pressure cuff around the upper arm and listening for the disappearance of systolic and diastolic pulses with a stethoscope, but can be continuously monitored using catheters and manometers or electronic pressure transducers (see e.g. Gabe, 1972). Central venous and pulmonary artery pressures may be recorded with catheters inserted during surgery.

The number of sites for pressure measurement are limited, and a continuous vector of measurements (corresponding to each segment in the model) is therefore not possible. Blood flow rates can be measured using e-m flowmeters, usually in a peripheral vessel, or using the technique of plethysmography (e.g. on a finger, or arm). Myocardial function can be determined by chest X-ray. Neural activities are not usually measured in human cardiovascular studies, but may be in animal experiments. Drug concentrations may be determined from blood samples at finite intervals from limited sites.

6.4.3.3. Measurement of variables in the Pullen model

"Variables" are the dynamic attributes in the functional modality. In the full model there are 61 state variables, and a complete continuous record of all these would pose great practical problems even if they were all measurable. The following variables are usually available on a continuous, or "beat by beat" basis: heart rate, arterial pressures (typically in aorta or arm), venous pressures, stroke volume, cardiac output, left ventricular ejection time, cardiac ejection fraction, and total peripheral resistance. Variables in the neural control submodel are not measurable because it is an empirical input-output model and they do not have direct physical referents.

The concentration of injected drug could be determined from blood samples, but the rapid dynamics entail that measurements with the spatial and temporal resolution of the model could not be made.

6.4.3.4. Measurement of parameters in the Pullen model

Very few model parameters can be measured, and most are not even standard parameters available in the literature as normal values. Pullen (1976) bases the parameter values largely on those used by Beneken and De Wit (1967) and Hyndman (1970). Beneken (1965) elaborates in detail how he determined the parameter values in his model, but no uncertainties or normal ranges are given. Al Dahan (see Section 6.5.4.3) has attempted to determine such ranges using sensitivity analysis. The parameters for the neural receptors and controllers are based on empirical models fitted to data from dog experiments (Katona 1967). Similar experiments cannot ethically be performed on humans and so there is a problem of comparative physiology.

The parameters describing the local effects of drugs are also non-standard and have to be approximated from descriptive accounts which usually express the local effects in ordinal terms, (see e.g. Burgen

and Mitchell, 1978).

6.4.3.5. Physiological Experiments

The response to certain physiological experiments such as rthostasis (postural change), blood volume changes (haemorrhage), and the Valsalva manoeuvre are well known, and quantitative dynamic responses of the variables in Section 6.4.3.3. are widely available. (Consult any physiological text). In order to use a physiological experiment to validate the model it is necessary that it does not entail changes outside the intended range of application which in turn affect the system, (e.g. chemical effects, shock, emotional changes). In other words, permissible experiments and the range of application are "closed".

6.4.3.6. Pharmacodynamic experiments

As with the determination of local drug effect parameters, data on the overall effects of a drug injection (of a short-term cardiovascular agent) are usually available in qualitative form only, (typically ordinal). The drugs simulated in the Pullen model are frequently naturally occurring hormones (such as noradrenaline, the neurotransmitter of the sympathetic nervous system) or have simple cardiac or vascular effects (e.g. isoprenaline) via or receptors. Thus their overall effects can be "thought-out" in a sort of Gedanken experiment. Presumably this is how many writers obtain results for fast-acting drugs and explains the differing accounts given (particularly where local drug effect and neural reflex are non-cooperatively interacting). Frequently, when experimental drug data are available they are not continuous, and the first measurements are often taken after the initial effects have occurred (i.e. after 1 - 2 mins.). (See the comments on availability of drug plasma concentration data in section 6.4.3.3.).

6.4.3.7. Data uncertainty

This is perhaps the most important consideration of this section. There are four sources of uncertainty :

6.4.3.7.1. Incomplete specification

System complexity, theoretical unobservability and practical constraints imply that only a small subset of variables can be continuously measured and are available as data. (If the model variables

agree with this subset, this does not in general imply that the unobserved variables will also match).

6.4.3.7.2 Parametric uncertainty

As discussed in 6.4.3.3, most parameters represent aggregated properties and are not easily measurable. The values used in the model are therefore very uncertain.

6.4.3.7.3 Definition of a normal system

Although the model is intended to represent a "normal" human, in fact there is no such thing, and the only information available is of statistically average values. The normal population shows a considerable range of behaviour, and ranges of normal arterial pressure, heart rate, cardiac output, etc. are available from a few sources, (e.g. Mountcastle, 1974; Guyton, 1971). Unfortunately, virtually no information is available on the statistical interdependency of these variables (i.e. the non-diagonal elements of the variance-covariance matrix). The implications for empirical validation are that each variable in the model which has a data referent should independently be in its normal range. This is not the same as the model representing a normal human. A trick which is used in the validation study (section 6.5) is to extract features from the data which are subject independent as far as possible, (e.g. in the Valsalva manoeuvre, section 6.5.4.3).

6.4.3.7.4 Metrical uncertainty

This is the measurement uncertainty introduced into each measurement as a consequence of the instrument accuracy, noise effects, or the disturbance of the patient in the process of measurement. A typical metrical uncertainty range for arterial pressure is $\pm$5mm Hg, but given the considerations of sections 6.4.3.7.1 - 3, considerations of metrical uncertainty are not, in general, important in the validation of the Pullen model.

6.4.4. Validity Criteria

6.4.4.1. Consistency criterion

This requires that the mathematical model contains or entails no contradictions. It is a necessary precondition which is usually satisfied in model formulation, (see section 6.5.2).

6.4.4.2. Algorithmic/Simulation criteria

These criteria require that the simulation model is an accurate implementation of the mathematical model. In particular the integration

algorithm of the Pullen model will be assessed, (section 6.5.3).

6.4.4.3. Theoretical criteria

Theoretical criteria require that the model and its submodels should be consistent with physiological theory over the range of application. By "theory", is meant both knowledge of the physical structure of the cardiovascular system and functional explanation of its behaviour. In some respects the model involves great simplification (e.g. the aggregation of the distributed characteristics of the circulatory network), and in others it involves an elaboration of physiological theory into mathematical form (e.g. the neural control and drug submodels).

6.4.4.3. Empirical

In applying these criteria, the model is compared with empirical data. The various empirical criteria applicable to the three different aspects of the model are shown in table 6.4. The criteria are based on the objective requirement of the model to represent a normal human (section 6.4.2) and limitations imposed by the data (section 6.4.3). As the model involves both simplifying assumptions and theoretical innovation, it is highly desirable that each individual submodel should be separately validated empirically. However, as discussed in section 6.4.3, and summarised in table 6.3, suitable data are not always available. Most data used in validation are from experiments on the controlled circulation (section 6.4.3.5).

6.4.5. Structure of the programme of validation.

Firstly, the prerequisite criteria of consistency and algorithmic/simulation/validity should be satisfied. Secondly, the theoretical and empirical validity criteria are applied. These are applied sequentially, starting with the elementary submodels and gradually building up to the whole model so that the validity (or lack of it) is built up deductively. The structure is illustrated in Fig. 6.6, in which it can be seen that the validation is divided into a number of levels, and that the whole structure forms a hierarchy. As shown in table 6.3, not every submodel is validated empirically, and those which are in this study are enclosed in boxes.

The programme of validation proceeds from level 1 to level 4, applying theoretical and, where possible, empirical criteria (table 6.4)

Submodel	Appropriate Data or Experimental Tests for Direct Empirical Validation
Heart	E.C.G. responses; Cardiac Function Curves; etc. (not directly applicable to model, however)
Circulatory Fluid Mechanics	Standard Data on Vessel Volumes, Lengths, etc. (no direct measurement of physical properties).
Neural Control	No suitable data
Drug Distribution	"
Drug Disposal	Limited Pharmacokinetic Data
Local Drug Effects	Data on Specific Action of Drug on Receptor Site (e.g. α - Receptor) not Appropriate.
Uncontrolled Haemodynamics (Heart and Circulatory Fluid Mechanics)	Events and Variables within a Single Cardiac Cycle: Pressure and Flow Waveforms at Various Sites; Cardiac Ejection Dynamics.
Controlled Haemodynamics (Heart, Circulatory Fluid Mechanics, and Neural Control)	Pressure and Flow Waveforms at Various Sites, and Cardiac Variables (Heart Rate, Stroke Volume, etc.) for Human at Rest and During Physiological Experiments that Affect CNS Pressures, Flows, and Haemodynamic Parameters.
Controlled Haemodynamics with Pharmacodynamics (Full Pullen Model)	Circulatory (Pressures and Flows) and Cardiac Changes in Response to Drug Injections, (Usually Qualitative Data).

Table 6.3 Availability of Data and Experimental Tests Appropriate for the Direct Empirical Validation of the Various Submodels.

(VALIDATION PROGRAMME PROCEEDS FROM LEVEL 1 TO LEVEL 4. BOTH THEORETICAL AND EMPIRICAL CRITERIA ARE CONSIDERED. BOXES INDICATE EMPIRICAL VALIDATION. MOVING UP HIERARCHY ≡ DEDUCTIVE VALIDITY. INFERRING DOWN HIERARCHY ≡ INDUCTIVE VALIDITY).

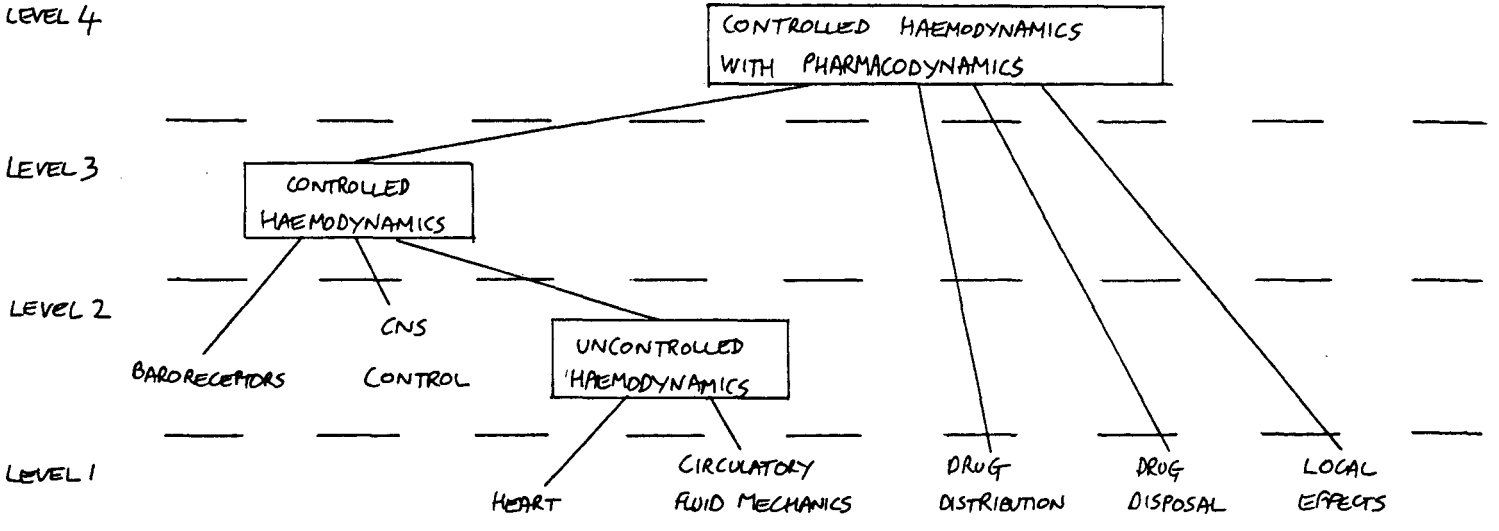


FIGURE 6.6. HIERARCHY (OR TREE) OF VALIDATION FOR PULLEN MODEL.

Aspect of Model	Nature of Data	Applicable Empirical Validity Criteria
Haemodynamic Responses	Quantitative, Dynamic Responses of Heart Rate, Cardiac Output, Arterial Pressure, etc. (see section 6.4.3.3)	Qualitative Similarity; Model Variables in Normal Range; Reproduction of Quantitative Features (in Some Tests); Statistical Tests
Neural Control	No Data	No Direct Empirical Validation
Pharmacodynamic Responses	Descriptive Accounts; Ordinal Responses of Haemodynamic Variables (see section 6.4.3.6)	Qualitative Similarity; Model Variables Change in Same Direction as Data.

Table 6.4. Applicable Empirical Validity Criteria.

to the submodels on each level. Very importantly, if a submodel cannot be empirically validated, inferences can often be made about its empirical validity at a higher level to which it is attached and at which empirical criteria may be applied. Many physiological tests on the controlled circulation (level 3) are used to investigate the integrity of neural control in the human in terms of haemodynamic responses. In validation such tests may also be used effectively to examine the integrity (i.e. validity) of neural control submodels in the Pullen model. Often the integrity of neural control is linked to the existence of key features in haemodynamic response and thus the feature space techniques of model validation proposed in Chapter 5 will be applicable. The details of the full programme of validation (for the application of theoretical and empirical criteria) are shown in table 6.5.

6.5. Results of the Programme of Validation

6.5.1. Introduction

The validation study is presented here in the programme form developed in section 6.4. Firstly, the necessary consistency and algorithmic/simulation validity criteria are checked. Then the empirical and theoretical criteria are applied starting with the elementary submodels and gradually moving up to the overall model with pharmacodynamic responses (i.e. up the hierarchy in fig. 6.6). In this way, the validation tests of the overall model can be interpreted with the help of the results of the empirical and theoretical tests on the submodels. The general conclusions of the programme are given in section 6.6.

6.5.2. Consistency criteria

As remarked in section 6.4.4, consistency is usually satisfied during the process of model formulation. In fact, Pullen had to eliminate some algebraic loops (a consistency problem) when simulating the mathematical model, (Pullen, 1976 pp. 99 and 212).

6.5.3. Algorithmic/Simulation criteria

The model is deterministic and so after verifying the translation into FORTRAN IV (see section 6.3.5) the only remaining checks are for numerical accuracy.

6.5.3.1 Accuracy of integration

The step length, h, of the Euler integrator satisfied $h < 2\tau_{min}$ for convergence (where τ_{min} is the smallest time constant) for flows

Level	Submodel	Validity Criteria	Comments
1	1. Heart 2. Circulatory Fluid Mechanics 3. Drug Distribution 4. Drug Disposal 5. Local Drug Effects	Theoretical " " " "	Examine Assumptions, and Consistency with Physiological Theory; Suggest Experimental Tests for Empirical Validation
2	1. Uncontrolled Haemodynamics	Theoretical Empirical : (i) Qualitative Similarity. (ii) Variables in Normal Ranges	Integration of 1.1 and 1.2 System Effectively Uncontrolled During 1 Cardiac Cycle; Quantitative Data, but Qualitative Similarity Important
	2. Baroreceptors 3. CNS Control	Theoretical "	As for Level 1
3	1. Controlled Haemodynamics	Theoretical	Integration of 2.1, 2.2, and 2.3; Suggest other Experimental Tests
		Empirical: (i) Qualitative Similarity (ii) Variables in Normal Ranges (iii) Feature Comparisons	Data Available for Main Variables under Equilibrium conditions, and Range of Standard Circulatory Tests
4.	1. Controlled Haemodynamics with Pharmacodynamics (Full Pullen Model)	Theoretical	Overall Integration of Model
		Empirical: (i) General Shape of Responses (ii) Main Variables Change in Same Directions as Data	Data Mainly Available from Descriptive Accounts of Effects of Fast-Acting Cardiovascular Drugs.

(N.B. at some levels, criteria are only theoretical owing to lack of appropriate data. This does not imply that the submodel fundamentally cannot be validated empirically. The considerations under "theoretical criteria" in section 6.5 often suggest new critical experimental tests for empirical validation).

Table 6.5 Programme of Validation for the Pullen Model - Systematic Application of Representational (Theoretical and Empirical) Validity Criteria.

1000ml sec^{-1} and h = 0.5 msec. This results in rather a long simulation time, and Rajkumar (1978) has shown that the integrator remains stable if h is increased. The accuracy (with h = 0.5 msec) assessed against results using the Runge-Kutta-Merson method of integration (with accuracy set to 0.01%) is within ± 3%.

6.5.4. Empirical and Theoretical Criteria

These are applied to each submodel in turn at each level in fig. 6.6, from level 1 to level 4, thereby building deductively an idea of the strengths and the limitations of the model. Where data are not available for empirical validation, only theoretical aspects are considered but possible experiments and critical tests for empirical validation are often suggested. In following through the rest of this section, one should bear in mind the modelling objectives, viz. to produce a pulsatile mathematical model of the human cardiovascular system for the study of short-term haemodynamics and predicting the effects of fast-acting cardiovascular drugs.

6.5.5. Level 1 validation

6.5.5.1 Heart submodel - theoretical criteria

The heart is simply characterised as four elastic compartments with time-varying elastances and non-linear modelling of the heart valves (section 6.3.2). The four compartments correspond exactly to the four heart chambers, but this is the extent of the representation of the physical modality. No account is given of the physical interrelation of the chambers, or their geometric shapes. However, for the purposes of representing a normal human these aspects can be considered irrelevant to the model, as they are invariant within the range of application.

The time-varying elastance model was introduced by Dick and Rideout (1965) and represents the active contraction of the heart. This is essentially an input-output model which relates chamber pressures to contained blood volumes during systole. However, there is a close relationship to the force of contraction of the heart muscle, and the elastance functions used by Dick and Rideout (1965), Beneken and De Wit (1967) as well as Pullen are strikingly similar to force of contraction curves (e.g. Guyton, 1971, p. 150). Similarly, the linear approximation for timing of events within each cycle can be linked to basic mechanisms of conduction of the action potential from the S-A node through the Bundle of His to the Purkinje fibres in the ventricular muscle. The empirical validity of the linear approximation could be checked against

an ensemble of ECG data, although this was not available in the present study.

A consequence of the elastance heart submodel is that, if venous return is increased, the end-diastolic volume increases and therefore diastolic pressure rises in proportion. This leads to increased cardiac output which adjusts itself to balance venous return. This is in accordance with Starling's "Law of the Heart" and the theory of cardiac autoregulation (see section 6.2.2, 6.2.3).

The physiological significance (validity) of the heart submodel gives the model parameters an empirical meaning. Indeed, Pullen (1976, p. 120) suggests that left ventricular systolic elastance may be used as a myocardial contractility measure, i.e. an index of heart performance. Of course, before it is used the validity of the submodel should be examined in further detail both theoretically and empirically, (importance checks must also be made on parameter sensitivity, see section 6.5.4.3).

A further aspect to consider is that of the pulmonary and aortic valve ejection dynamics. Pullen (1976, pp. 36-38) uses Beneken and De Wit's (1967) equation:

$$P_{LV} - P_{AO1} = \underbrace{R_{LVAO1} F_{LVAO1}}_{\text{Viscous resistance}} + \underbrace{L_{LV} \frac{dF_{LVAO1}}{dt}}_{\text{Inertia}} + \underbrace{\frac{\rho}{2}\left[\frac{F_{LVAO1}}{A_{AO1}}\right]^2}_{\text{Bernoulli Term}}, \quad F_{LVAO1} \geq 0 \qquad \ldots (6.11)$$

for both valves and, in addition, calculates an "effective area" A_{AO1} of the aortic valve in order to account for curvature of the aortic arch and pulmonary arteries (Pullen, 1976, appendix 5). This calculation is based on a large number of assumptions including laminar flow, and that the blood vessel does not change shape during ejection. Both of these are questionable, in particular the latter since it is well known that the aortic arch is a highly elastic vessel which expands considerably during ejection. Pullen's assumptions lead to a small effective area and consequently, since it is a squared term, a very high estimate of the Bernoulli term (42mm Hg at maximum flow; Pullen, p. 38). The implication of this is that the valves will appear stenosed ventricular pressures will rise very high, and the inertia term will be

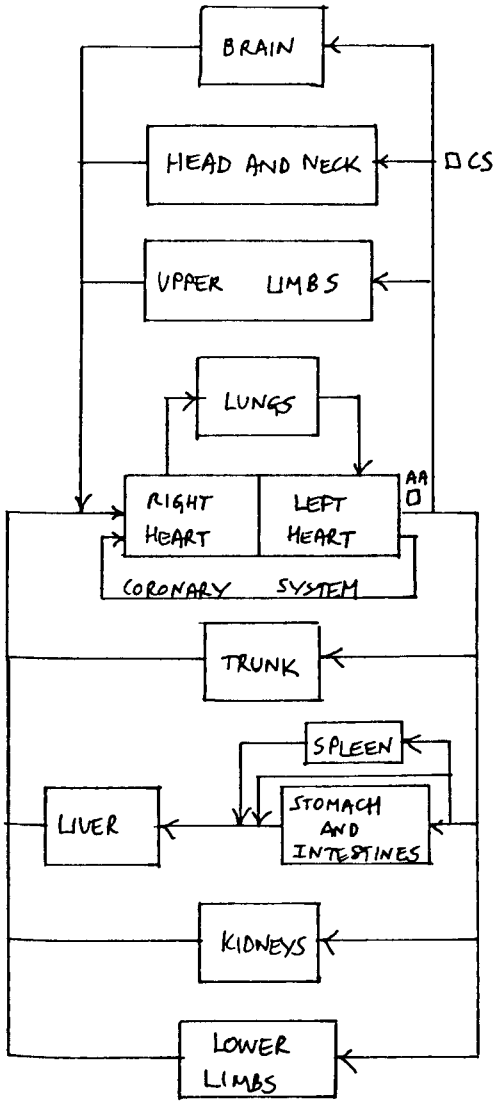
swamped (eliminating the well-known protodiastole when $F_{LVAO1} > 0$, but $P_{LV} < P_{AO1}$). This will be considered further in the empirical validation of level 2 (section 6.5.4.2.).

6.5.5.2. Circulation submodel - theoretical criteria

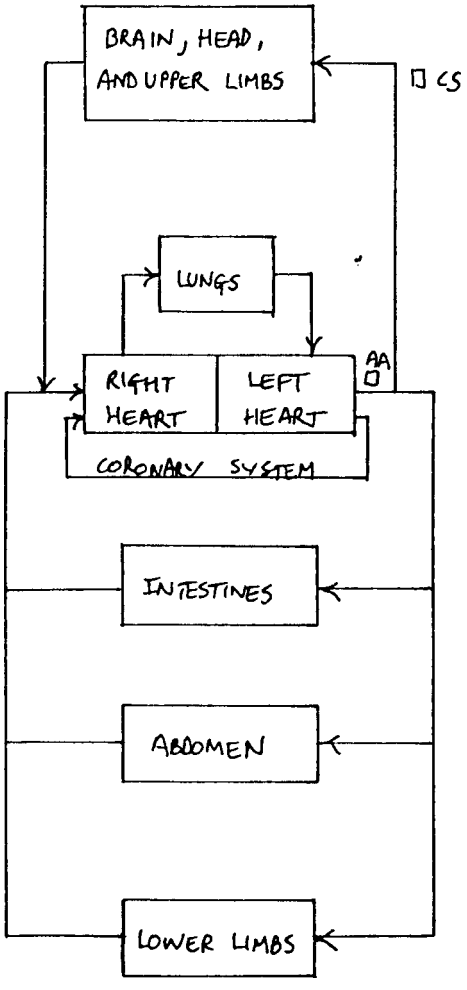
The 15 segments of the circulation submodel (19 including the heart) correspond to the major arterial and venous divisions within the systemic and pulmonary networks. The resolution of the model allows the changing elastic and hydrodynamic characteristics of the arterial tree to be lumped as an approximation to the full Navier-Stokes partial differential equations. The systemic resistance is mainly located in the model between arterial and venous segments, and it is well known that the maximum resistance to blood flow occurs in the capillaries. In the model, however, blood is either in an arterial or venous segment, and the blood contained in the capillaries in the human must be regarded as split between two such segments.

The structure of the circulation submodel is a simplification of the physical modality, but does this impair the validity of the model over the functional modality of the intended range of application? Consider the scheme of the human circulation (itself simplified, taken from Lippold and Winton, 1979, p.213) and the structure of Pullen's circulation submodel (fig. 6.7). It can be seen that the aggregation of the model is mainly in the head, arms and brain and the splanchnic circulations. There are three distinct aspects relevant to functional validity:

(i) As an approximation to the uncontrolled dynamic system. The full haemodynamic system is a stable infinite dimensional system. (By "full dynamic system" is meant the best (most complete) theoretical model of the system. It is not the system itself, nor is it necessarily soluble). The theoretical validity of the approximate finite dimensional model is ideally determined by comparing the pressure and flow variables with the corresponding spatially averaged variables in the full dynamic system. Since the full dynamic system of Navier-Stokes equations is impossible to solve, other methods have to be used. In particular, the convergence of the solution can be examined in models with more or less segments. Rajkumar's 1978 8-segment model of the controlled circulation produces results close to those of the Pullen model and is stable



SCHEME OF HUMAN CIRCULATION (FROM LIPPOLD AND WINTON, 1979, p. 213).



STRUCTURE OF SUBMODEL OF CIRCULATORY FLUID MECHANICS.

FIGURE 6.7. COMPARISON OF STRUCTURE OF SUBMODEL OF CIRCULATORY FLUID MECHANICS AND SCHEME OF HUMAN CIRCULATION.

(Thomas, 1980). It is significant that in the steady-state a simple single division of the systemic circulation (one arterial and one venous segment) produces realistic stable behaviour for aortic and venous pressures and flows. Therefore it is unlikely that the present aggregation of the model produces great inaccuracies vis-a-vis the uncontrolled dynamic system.

(ii) As a reference frame for cardiovascular control.

The resolution of the circulation submodel must be sufficiently detailed to provide information for cardiovascular control within the range of application. The two major sites for neural pressure receptor fibres are located in the aortic arch and carotid arteries (fig. 6.7), and both these areas are represented in the model. These sites send information to the CNS and are the sensors in the fast reflex control of both heart and circulation. (Incidentally, other sites for baroreceptors are available in the model such as the left atrium or venous segments. Evidence suggests that these receptors act as volume sensors and play a role in the longer-term regulation of blood volume).

Two outstanding omissions from the model are separate segments for the kidneys and the brain. The kidneys filter a tremendous volume of blood, removing waste products and selectively excreting ions and water in order to maintain an electrolytic balance and control arterial pressure. The primary mechanism is through the local release of a hormone, renin, and the renin - angiotensia II - aldosterone system which produces vasoconstriction and which has a time scale of a few minutes to days, and so is outside the range of application. The brain is a more important omission since it does play a role in the sensing of changes which result in rapid cardiovascular control. In particular, the activity of the medulla is sensitive to concentrations of O_2 and CO_2 and pH level. Changes in these produce rapid reflex changes in the heart and circulation, however such changes are not entailed within this range of application (limited to direct haemodynamic effects). Thus the submodel is not a representation of <u>all</u> short-term control effects, but within the range of application (haemodynamics) it has sufficient resolution to act as a reference frame for control.

(iii) As a reference frame for pharmacodynamics.

The partition of the circulation should contain the appropriate segments (tissues, organs) on which the injected drug has its direct effects. See section 6.5.5.5.

6.5.5.3. Drug injection and distribution submodel - theoretical criteria

This submodel is described in section 6.3.4. Drug flow is equal to drug mass multiplied by blood volume flow and is essentially a continuity equation (equation 6.10). It assumes that the drug is evenly mixed within each compartment. The validity of this equation depends on both this assumption and the validity of the blood volume flow, i.e. the controlled haemodynamics. However, the presence of the drug modifies the haemodynamic behaviour and therefore its own distribution. The validity also depends on the number of compartments - are there sufficient to approximate the distributed processes of transport (diffusion and flow) in the full dynamic system (refer to section 6.5.5.2) ? A way to investigate this is by a model-based experiment increasing and decreasing the number of compartments and examining convergence of the model.

The empirical validation of the full model with pharmacodynamics takes place at level 4 (because of the complex interaction of many processes) and the drug submodel does not receive empirical validation until then. An experiment can be devised to rectify this unsatisfactory situation which consists of injecting a dye into the bloodstream free from cardiovascular effects. Level 3 of the validation hierarchy (section 6.4, fig. 6.6) is split into two levels (fig. 6.8). If the dye has the same disposal characteristics as the drugs of interest (fast-acting) this also provides a test of the disposal/absorption submodel. Unfortunately, no data on such an experiment were found in the present study.

Other assumptions of the submodel are that drug volume $\ll$ blood volume, and that a single drug is present. The former is easily satisfied, and the problem of multiple drugs can be solved by adding multiple slave models (in this case the possibility of drug interaction exists, however).

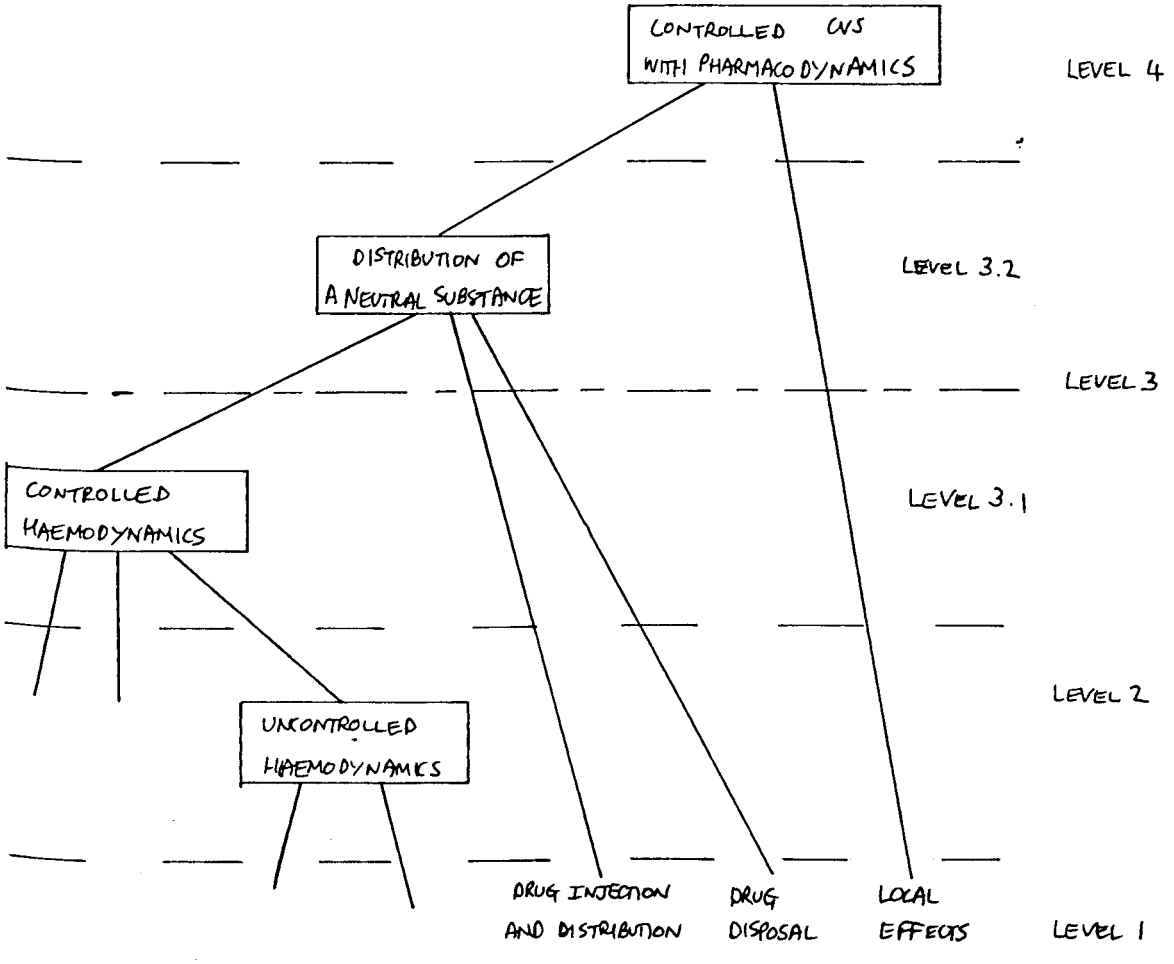


FIGURE 6.8. ALTERNATIVE VALIDATION HIERARCHY FOR EMPIRICAL VALIDATION OF DRUG DISTRIBUTION AND DISPOSAL SUBMODELS USING NEUTRAL SUBSTANCE INJECTED INTO BLOODSTREAM (SEE ALSO FIG. 6.6).

6.5.5.4. Drug disposal submodel - theoretical criteria

The breakdown and absorption of the drug is modelled as a first order linear mass decay process in each segment. This is probably an adequate model for the drugs of interest, although the time constants or half lives are not available as standard data.

6.5.5.5. Local effects submodel - theoretical criteria

The local effect submodel is an algebraic model devised by Pullen (1976, pp. 75-84). It represents the effect of the drug on the hydrodynamic parameters of each segment rather than the underlying processes of drug action (e.g. on α or β receptors, neurotransmission, muscular changes). For example, the resistance R_{12} between two segments (section 6.3, fig. 6.2) is modified in the following fashion:

$$R'_{12} = R_{12} \cdot \sigma_N \cdot \sigma_D \qquad (6.12)$$

where R'_{12} is the normal resistance, σ_N is the neural control multiplier, and σ_D is another dimensionless multiplier representing the effect of the drug which has the linear form:

$$\sigma_D = 1 + K\,w \quad \text{drug increases R (vasoconstriction)} \qquad (6.13)$$

$$\sigma_D = 1/(1 + Kw) \quad \text{drug decreases R (vasodilatation)}$$

where w is the concentration of the drug and K determines the sensitivity to the drug.

The drug effects submodel is therefore linear, and linearly combined with neural control. Since the neural control is highly nonlinear it is unlikely that the principle of superposition is obeyed. Furthermore, no evidence exists on the linearity of the drug receptors, indeed it is likely that receptors will show nonlinear threshold and saturation effects (see e.g. Burgen and Mitchell, 1978, pp. 3 and 233). The quantitative local effects can be determined in experiments on isolated tissues from animals, but modern physiological experiments on humans are concerned with the overall effects of the drug: " major advances in the understanding of how cardiovascular drugs act has come in recent years from new methods .... and include cardiac catheterization for measurement of pressures, Fick and dye methods for measurement of cardiac output, and the attachment of radiopaque markers to the ventricles for the measurement

of ventricular contraction" (Burgen and Mitchell, 1978, pp. 123-4).

Another aspect pertinent to the validity of the drug effects submodel is the partition of the circulation submodel (see also section 6.5.5.2). The present structure is suitable for cardiovascular drugs which have a generally uniform effect (such as the sympathomimetic drugs) but not for drugs which, although fast-acting, have more selective uptake and distribution properties. This suggests different partitions for different modelling objectives. For instance, fig. 6.9 shows an alternative partition based on Zwart et al. (1972) suitable for the study of the uptake, distribution and effect of an anaesthetic drug, halothane. Thus the present model should not be taken as a general framework for any cardiovascular active drug.

These considerations show that this submodel does not have very much theoretical validity and therefore it is a major source of uncertainty in the overall model. The validity of the drug submodels will have to be based on empirical tests at level 4 (refer to fig. 6.6 and section 6.5.4.4). Unfortunately the complexity of the model is justified by the inclusion of pharmacodynamics in the specific modelling objectives and yet the basic drug process and effect models remain substantially unvalidated at this stage.

6.5.6. Level 2 validation

6.5.6.1. Uncontrolled haemodynamics submodel

6.5.6.1.1. Theoretical criteria

The uncontrolled haemodynamics submodel consists of the heart and circulation submodels (section 6.5.5.1 and 6.5.5.2). The integration of the two submodels is achieved by considering heart valve actions and dynamics and is straightforward, as both submodels are in compartmental form. Ideally, the empirical validation would be done by comparing the model with data over two minutes from a human whose neural control loop is open (as discussed in section 6.2.3 the cardiovascular system without neural control does exhibit autoregulation and is stable under limited conditions). Unfortunately, little such data are available (although some are discussed at level 3, section 6.5.7). An alternative method to empirically validate is to consider a single cardiac cycle.

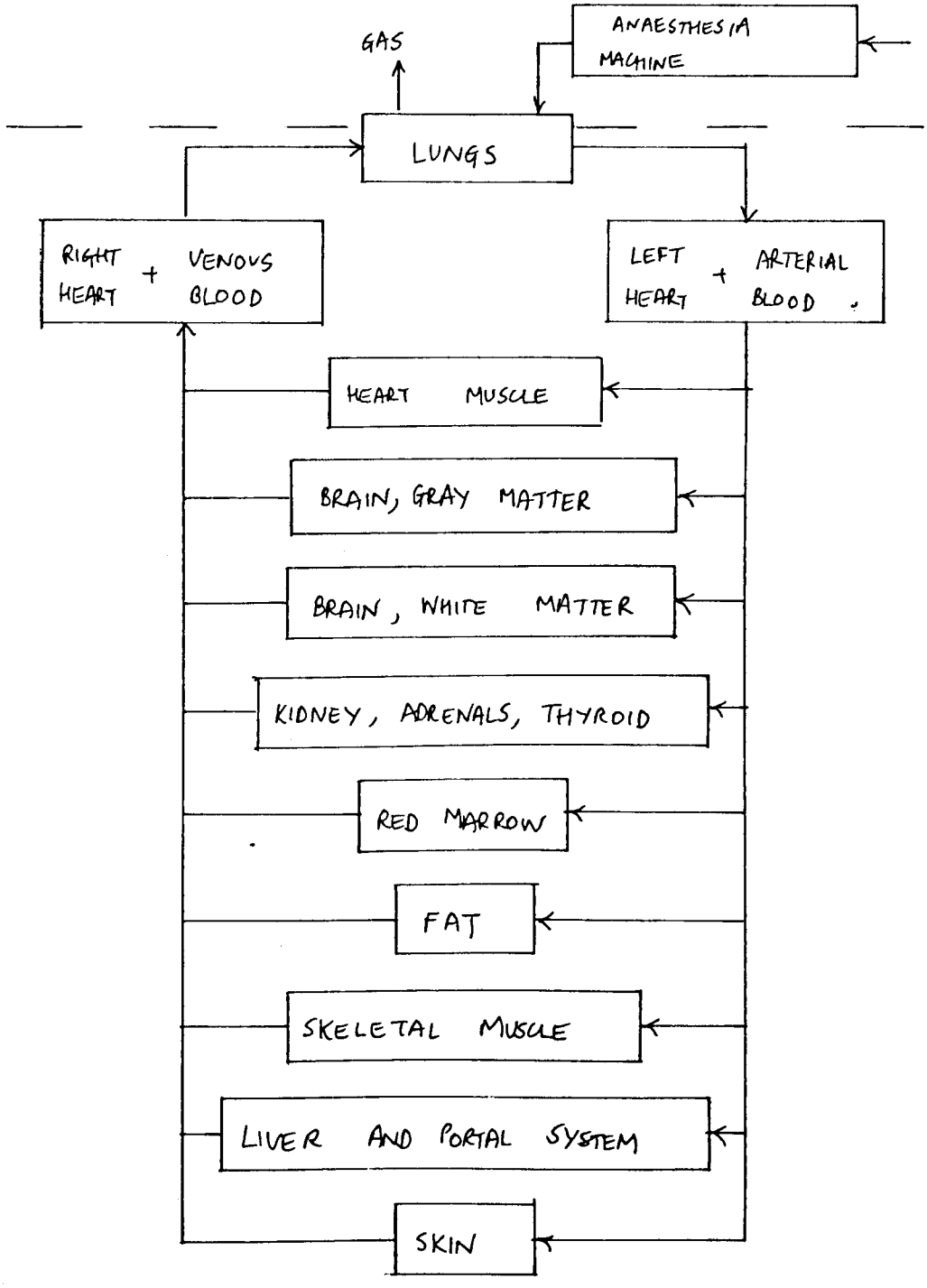


FIGURE 6.9. AN ALTERNATIVE PARTITION OF THE CIRCULATION APPROPRIATE FOR STUDIES OF HALOTHANE, AN ANAESTHETIC DRUG (BASED ON ZWART ET AL., 1972).

6.5.6.1.2. Empirical criteria

Over a single cardiac cycle the cardiovascular system is essentially uncontrolled by the CNS. Haemodynamic data from one cycle may therefore be used to empirically validate the submodel of uncontrolled haemodynamics. Fig 6.10a shows typical aortic, left ventricular and left atrial pressures and aortic blood flow in a typical human (Hawker 1979). The model responses are shown alongside in fig. 6.10b.

There is a close qualitative similarity in most of the variables; in particular, the distinctive "dichrotic notch" in the arterial pressure waveform at the closing of the aortic (S-L) valve and the following dichrotic wave are reproduced in the model (although some fast dynamics are lost). Left atrial pressure remains low in the model ($<$15mm Hg) as it does in the data. During diastole it rises above left ventricular pressure (as the ventricle fills) and shows the small pulse just before systole when the atrium contracts. Aortic pressure in the model is about 10mm Hg higher than the data (but still in the normal range) and the pulse pressures are the same.

A serious discrepancy exists in the left ventricular pressure (P_{LV}) waveform in the model. During systole P_{LV} rises substantially above the aortic pressure and does not show the characteristic flattening as the aortic valve opens (this is well reported in the literature, see also Guyton, 1971 p. 151). A very high P_{LV} arises in the human due to aortic valve stenosis (Walters 1979, p. 62) or increased valve resistance. The error in the model response is therefore probably due to an inaccuracy in the modelling of cardiac ejection dynamics. In section 6.5.5.1. this was considered at some length and it was suggested that the high value of the Bernoulli term leads to a high valvular resistance, and this is confirmed empirically here. Consequences of this are:

(i) Increased total systemic resistance, leading to generalised increase in arterial pressure.

(ii) Increased stroke work of the heart and therefore calculations of heart work based on the model are invalid.

(iii) The "proto-diastole" does not occur. In the human in the latter half of systole, the pressure in the left ventricle drops below that of the aorta, although the S-L valve does not close because

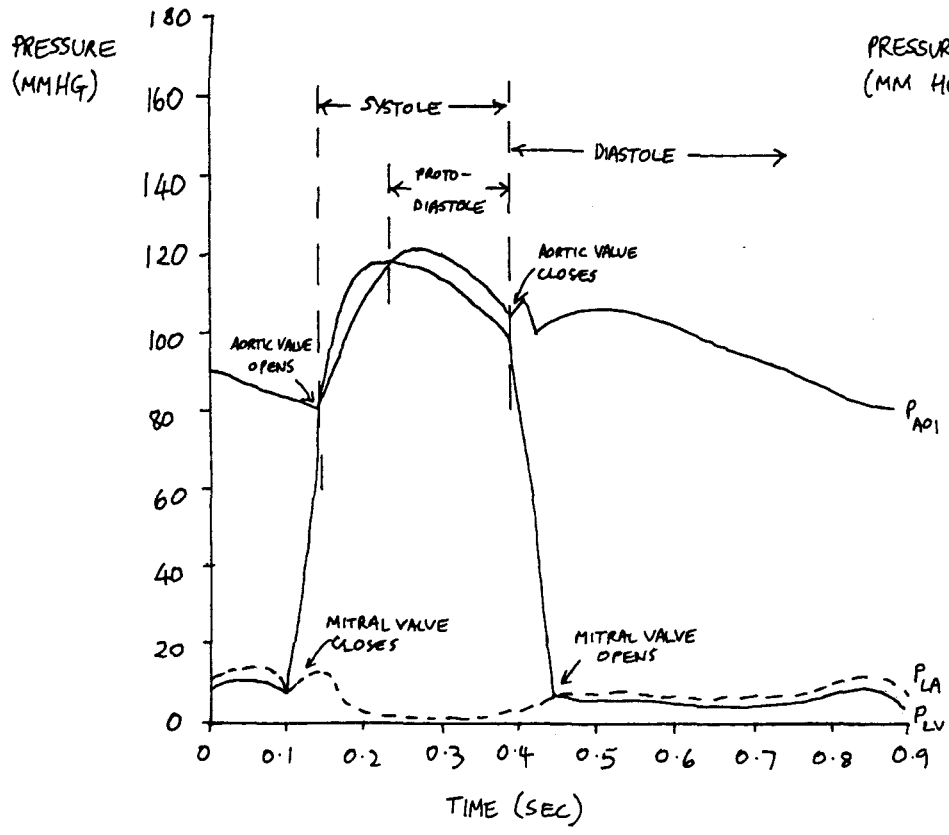


FIGURE 6.10a HUMAN

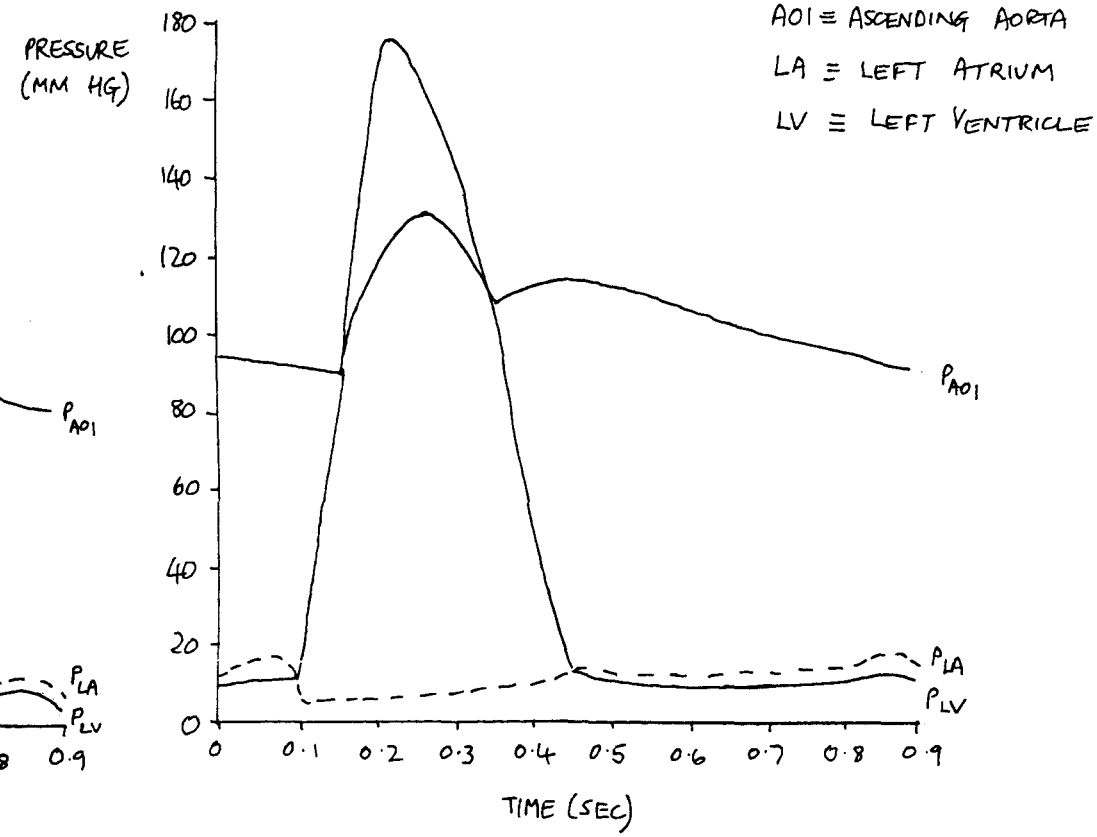


FIGURE 6.10b MODEL

FIGURE 6.10. AORTIC AND LEFT VENTRICULAR PRESSURE WAVEFORMS IN THE MODEL AND HUMAN DURING ONE CARDIAC CYCLE. (DATA FROM HAWKER, 1979, p. 2).

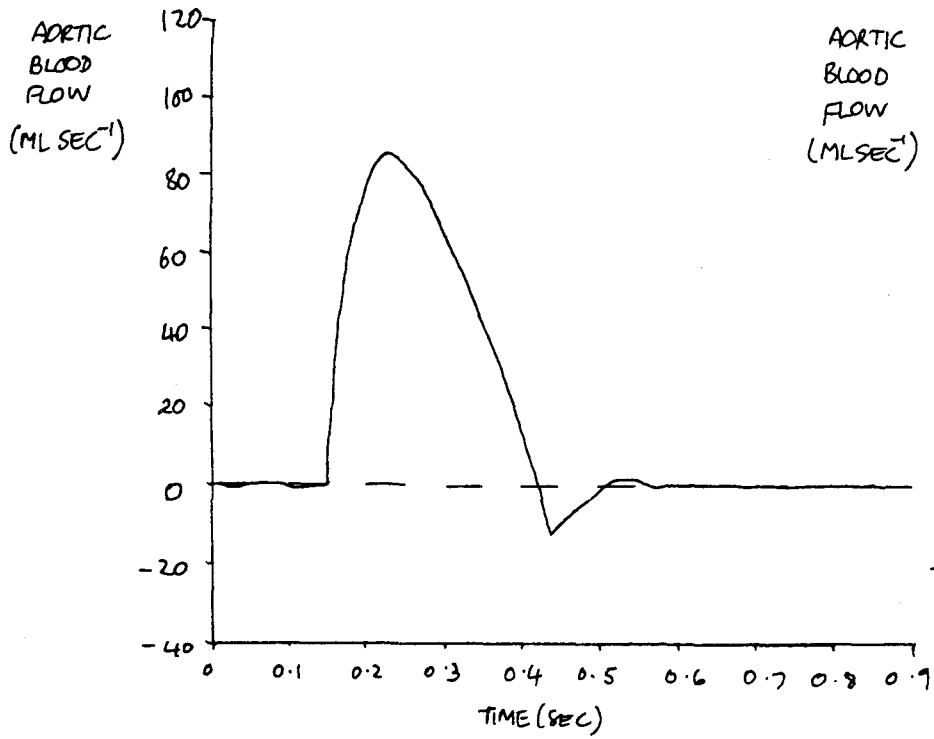


FIGURE 6.10a (CONTD.) HUMAN

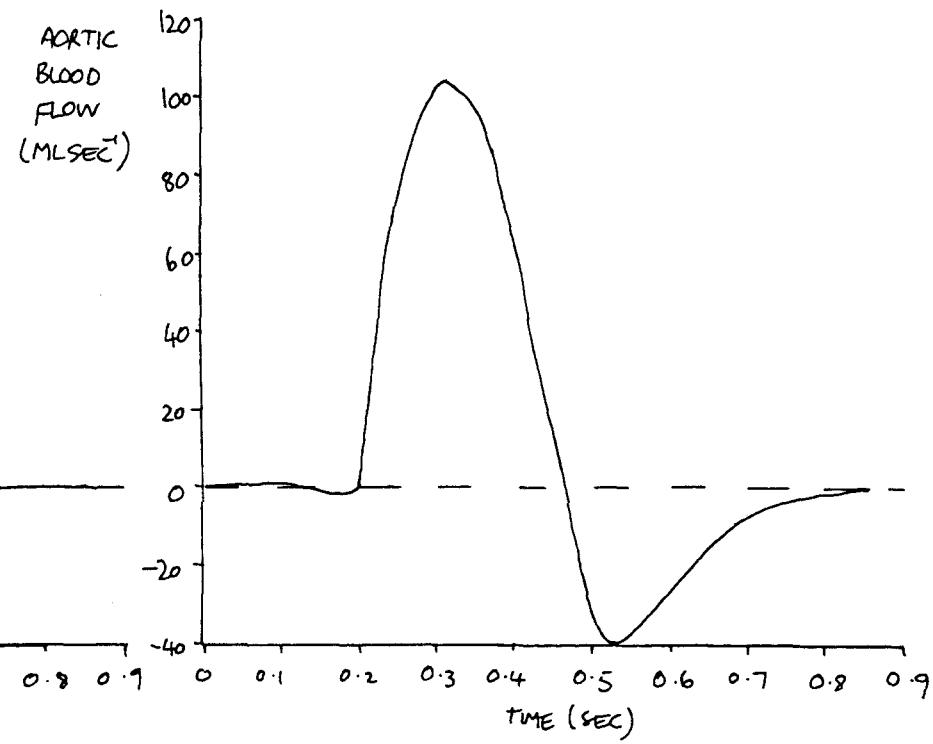


FIGURE 6.10b (CONTD.) MODEL

FIGURE 6.10 (CONTD.) AORTIC BLOOD FLOW WAVEFORMS IN THE MODEL AND HUMAN DURING ONE CARDIAC CYCLE. (DATA FROM HAWKER, 1979, p.2).

the ventricle is still contracted and the blood passes from ventricle to aorta under its own momentum.

The waveforms of aortic blood flow show a close similarity between model and data; the magnitudes are in the same range and the "retrograde flow" also occurs in the model (although rather larger than in the human).

Fig. 6.11 shows the transmission of the pressure pulse through the arterial tree in the human and model. The data are taken from Guyton (1971) and are at the low end of the normal range of pressures. Although the model pressures are generally higher they reproduce the main features of the data:

(i) The further from the heart, the longer until peak (systolic) pressure is reached. A comparison of the average pulse wave velocity can also be made:

	Site	Time until peak	Distance	Estimated wave velocity
Data	Radial	0.05sec	40cm	800 cm sec^{-1}
Model	Leg	0.17sec	90cm	530 cm sec^{-1}

given the uncertainty of the times and distances this is acceptable.

(ii) Systolic pressure increases with distance from the heart due to variation in wave velocity at different stages in the arterial tree. Similarly, diastolic pressure drops slightly, leading to an increased pulse pressure. The extreme increase shown in the dorsalis pedis (a peripheral foot artery in the human) is not reproduced in the model because there is no corresponding compartment. However, in the range of application it is not necessary to have such fine resolution.

As the uncontrolled submodel of haemodynamics stands it has shown to be reasonably valid in empirical tests over a cardiac cycle. However the ventricular pressure and ejection dynamics are uncertain (invalid) as discussed above. The implications of this are that detailed cardiac effects in the model cannot be trusted, and a tendency to arterial pressure to be high.

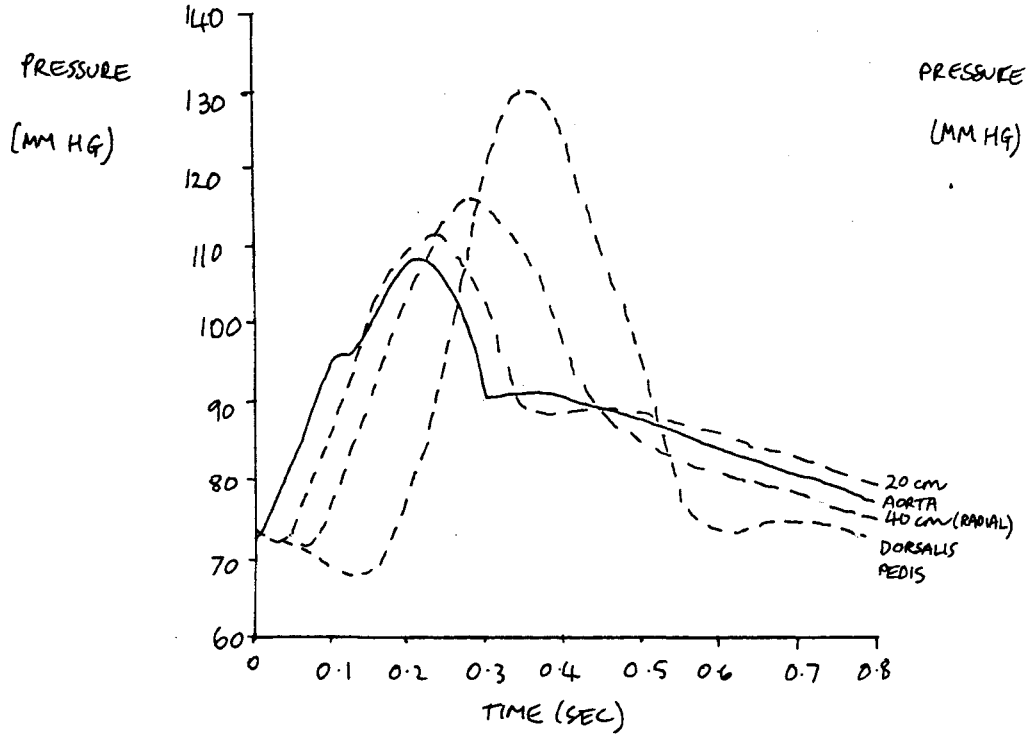


FIGURE 6.11a DATA (FROM GUYTON, 1971, p. 222)

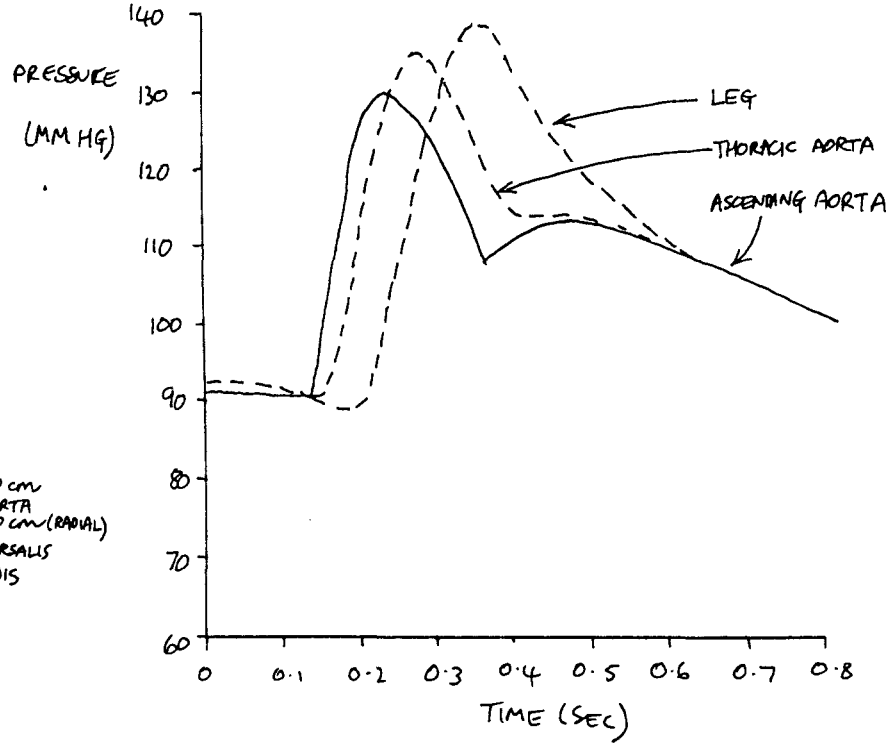


FIGURE 6.11b MODEL.

FIGURE 6.11. PULSE PRESSURE CONTOURS IN DIFFERENT SEGMENTS OF THE ARTERIAL TREE FOR HUMAN AND THE MODEL.

6.5.6.2. Baroreceptors submodels - theoretical criteria

The models of the baroreceptors (section 6.3.3.) are based on Katona's (1967) empirical models derived from dog experiments. These simply represent the pressure level and positive rate of change dependency of the baroreceptor fibres well known in human physiology as well. The parameters (time constants and thresholds) were determined by Pullen. In neurophysiological terms, the outputs of the baroreceptors are a stream of neural impulses travelling down afferent nerves to the brain, however it is a well validated hypothesis that they act as dynamic pressure sensors, and so the representation of the outputs as dynamically averaged pressures in the model is valid. But it is important to stress that the isomorphic nature of the model is lost at this level - i.e. it is a functional black box. A method of determining the parameters of this submodel would be to vary them until the haemodynamic responses of the overall model are closest to the data (see section 6.5.7.1.2).

The adaptation of the baroreceptors - in which their output falls after a prolonged period of hypertension - has a time constant of about 8 days (Guyton, 1971, p.301), and so this mechanism does not need to be included for the range of application.

6.5.6.3. CNS control submodel - theoretical criteria

The CNS control submodel is described in section 6.3.3. In this section, some of the assumptions of this submodel will be investigated.

(i) Linear combination of baroreceptor outputs.

It is assumed that the CNS input function B is a static linear combination of the output of the carotid sinus (B_{UA}) and aortic arch (B_{AO2}) baroreceptors:

$$B = B_{UA} + (1 - \alpha)B_{AO2} \qquad (6.14)$$

where $0 \leqslant \alpha \leqslant 1$

This is the simplest form of combining the inputs, and there is no evidence for it. The basic neural mechanisms in the medulla are unknown, and the only way to validate this is indirectly through the overall model. (see level 3, section 6.5.7) It is a plausible hypothesis

that CNS input B is determined largely by B_{AO2} (aortic arch output) except when B_{UA} (carotid sinus output) falls, indicating a dangerously low flow to the brain. A valuable simulation could be performed to investigate whether the overall model responses are significantly different when different strategies are used.

(ii) A single input function B.

A single CNS input, B, drives the various cardiac and circulatory controllers. Again, this is the simplest representation and no direct evidence can be given for it. Pullen remarks that the inputs for the different controllers could be devised with different linear combinations of B_{UA} and B_{AO2} (Pullen 1976, p. 166). The test for these also would be indirectly through the best fit of haemodynamic responses of the overall model to the data.

(iii) Notion of "centres"

A fundamental implicit assumption of the model is the notion of cardiac and circulatory "centres" in the medulla oblongata and that the function of these centres is for cardiovascular control. A recent trend in neurophysiology is to treat these centres as having a functional rather than a physical modality as the associated changes of neural activity tend to occur in distributed as opposed to local regions of the medulla. The main implication is that the medulla acts as a more distributed controller with a high degree of interaction between different functional centres. However, the understanding of this is insufficiently advanced to construct structurally similar models. In any case, it is likely that the present submodel will be adequate over the range of application.

The value of used in the model (α =0.7)was based by Pullen (1976, p.62) on Dampney's (1971) experiments on dogs. This assigns a greater significance to the carotid sinus baroreceptors as an input for CNS control. Problems of comparative physiology aside, there are other workers (e.g., Glick and Covell, 1968) who consider the aortic arch area to be more important in dogs. Recent experimental studies by Mancia et al.(1978) lend support to this hypothesis in humans, and this aspect is indirectly validated in empirical tests at level 3 (section 6.5.7.1.2).

The form of the sub-models for CNS control (of heart rate, myo-

cardial contractility, and vascular and venous tone) is based on Katona's 2-region model of heart-rate control (1967, derived from dogs) and Hyndman's "bang-bang" cardiovascular controllers (1970, based on human experiments with carotid sinus stimulation). It is the empirical basis of these models which gives them validity as CNS input/output models for Pullen's model, but it is difficult to give them greater physiological validity as many of the internal variables and parameters do not have empirical referents.

6.5.7. Level 3 validation

6.5.7.1. Controlled haemodynamics submodel

6.5.7.1.1. Theoretical criteria

The controlled haemodynamics submodel consists of the heart, circulation and CNS control submodels. It is this level of the hierarchy (see fig.6.6) which receives most empirical validation, as ultimately this submodel is used as a basis for modelling drug effects (Level 4, section 6.5.8). The controlled haemodynamics submodel is based solely on the rapid neural control loop, and many important mechanisms for cardiovascular control are not included. Some of these mechanisms and their time constants are given below:

(i) Chemical effects. Changes in P_{aCO_2} and P_{aO_2} in the brain cause rapid changes in activity in vasomotor and cardiac centres and corresponding reflex changes in the cardiovascular system ($\tau \simeq$ 10sec).

(ii) Hormonal control. Activity in the sympathetic nervous system in cardiovascular control is augmented by the release of catecholamines from the adrenal medulla. Changes in plasma osmality (electrolytic balance) stimulate the release of renin from the JGA in the kidney which controls the cardiovascular system through the renin-angiotensin-aldosterone system ($\tau \simeq$ 5 - 30 mins).

(iii) Kidney control. The kidney selectively excretes water and salts in order to control blood volume and electrolyte concentrations. The former affects venous pressure and hence cardiac output, and salt concentrations have a direct effect on cardiac performance ($\tau \simeq$1-2hr.).

(iv) Thermal homeostasis. Core temperature (under small environmental changes) is controlled by vaso-constriction, -dilation of the skin blood vessels and this affects systemic resistance ($\tau \simeq$10sec. - 1min.)

(v) Baroreceptor adaptation. If blood pressure remains high, the baroreceptor outputs will fall gradually. ($\tau \simeq 5$ days).

It can be seen that hormonal (ii) and kidney (iii) control mechanisms and baroreceptor adaptation (v) have time scales well outside that of the range of application. Chemical (i) and thermal (iv) control have time scales within that of the model, however environmental changes were limited in section 6.4 to those that directly entail haemodynamic changes. Thus, in theoretical terms, the submodel of controlled haemodynamics includes the control mechanisms pertinent over the range of application.

6.5.7.1.2 Empirical criteria.

This section considers four aspects: equilibrium conditions; general circulatory responses; simulation of Mancia's experiment; and model sensitivity analysis.

6.5.7.1.2.1 Equilibrium conditions

These are the conditions that the model achieves when simulating a recumbent, resting human with no environmental disturbances. The values of key model variables are compared with the normal human ranges within which they should lie, in accordance with the criterion of section 6.4.4. (The acquisition of "normal range" data is quite difficult, most references quoting a single normal value, and the ranges have been estimated from a number of sources. The "Handbook of Physiology" (1965) and Mountcastle (1974) proved very useful). Table 6.6 contains model and data values for the equilibrium conditions.

As can be seen from table 6.6 the model variables lie mostly in the middle of the normal ranges, with the exception of arterial pressure which is at the high end of the range. The value of mean arterial pressure (MAP) in the model is very slightly above the end of the normal range although systolic and diastolic pressures in the model are within normal ranges. This is probably due to the standard procedure of calculating MAP one third up from diastolic resulting in a low estimate of the true average value. The high arterial pressure in the model is probably a consequence of the high left ventricular pressure (section 6.5.6.1.2) and, perhaps, uncertainties in the neural control submodel. However, it still satisfies the empirical criterion.

VARIABLE	UNITS	MODEL	NORMAL HUMAN RANGE			NOTES
			UPPER	NORMAL	LOWER	
HEART RATE	B.P.M.	72.9	100 (E)	70 (2)	50 (E)	E - ESTIMATED FROM FH = CO /SV
STROKE VOLUME	ML	69.6	80 (E)	70 (3)	60 (2)	
CARDIAC OUTPUT	ML SEC^{-1}	84.6	134 (1)	100 (1)	50 (1)	RANGE = ±2σ OF YOUNG MALES (1)
MEAN ARTERIAL PRESSURE	MM HG	109.1	109 (1)	95 (2)	80 (1)	RANGE = ±1σ OF YOUNG MALES (20-30 YRS) (1)
SYSTOLIC PRESSURE	MM HG	130.9	135 (1)	120 (2)	109 (1)	
DIASTOLIC PRESSURE	MM HG	90.8	95 (1)	80 (2)	65 (1)	
PULSE PRESSURE	MM HG	40.1	45 (E)	40 (2)	35 (E)	
ESTIMATED SYSTEMIC RESISTANCE	MMHG SEC ML^{-1}	1.29	1.6 (E)	1.2 (E)	0.8 (E)	E - ESTIMATED FROM ETSR = MAP /CO

SOURCES: 1. MOUNTCASTLE (1974).
2. LIPPOLD AND WINTON (1979).
3. WINGATE (1976).

TABLE 6.6. EQUILIBRIUM STATES OF MODEL (CONTROLLED HAEMODYNAMICS) AND NORMAL HUMAN RANGES.

REGION	SOURCE	DATA (ML)	MODEL (ML)
SYSTEMIC ARTERIES	1	1000	585
SYSTEMIC VEINS	1	3200	2780
PULMONARY ARTERIES	1	240	115
PULMONARY VEINS	1	200	540
HEART	1	360	525
(LEFT VENTRICLE	2	140	131)
(RIGHT VENTRICLE	2	140	132)
TOTAL		5000	4545

SOURCES:
1. BERGEL (1972, P.4)
2. HAWKER (1979, P.3)

TABLE 6.7. BLOOD VOLUME DISTRIBUTION IN THE HUMAN AND MODEL.

(A stronger empirical criterion could be used if data were available on the relative interdependence of the data variables, (some are directly related mathematically, others physiologically), i.e. covariance as well as variance statistics. It then would be possible to calculate the probability that the combination of variable values in the model is drawn from a "normal" population. Unfortunately no source for such data has been located).

Further tests may be performed on the distribution of blood volumes and flows in the model. Table 6.7 contains data on the end-diastolic volumes of various regions, and the corresponding model values. The data have a larger total blood volume and more blood in the systemic circulation than the model, but in general the corresponding values are of the right magnitude. Bergel (1972) from whom most of the data are taken admits that the values are "very approximate idealisations". Pullen's values are based on those of Beneken and De Wit (1967), and this shows the very great difficulty of setting-up and validating biological models with respect to even fairly trivial variables. Sensitivity analysis proves very useful in determining how critical the values of such parameters are (see section 6.5.7.1.2.4).

6.5.7.1.2.2 General circulatory responses

These are responses to standard physiological tests on the cardiovascular system which entail directly only haemodynamic changes (i.e. within the range of application). The model is validated against four different tests: (i) postural changes; (ii) blood volume changes; (iii) heart pacing; and, (iv) a Valsalva manoeuvre. Pullen (1976) also used these tests to validate the model, but here they will be considered in more detail, presenting data where possible and insisting on qualitative similarity and feature space comparisons as empirical validity criteria.

(i) Postural changes.

A passive tilt to an upright position (orthostasis) causes large hydrostatic pressure differences which lead to pooling of the blood in the veins. This suddenly reduces venous return and hence arterial pressure. In the normal human there is a reflex through the baroreceptors leading to increased heart rate and systemic resistance. When the system regains its new equilibrium state, arterial pressure remains little changed, cardiac output is decreased, heart rate and systemic resistance are increased

(Lippold and Winton, 1979, p.222), and stroke volume is decreased (Mountcastle, 1974).

In the model, orthostasis is simulated by including a series of hydrostatic pressure generators (Pullen, 1976, p.56) between successive segments. The model response to an upright tilt is shown in fig.6.12. Firstly, consider the qualitative similarity with the descriptive account above. Mean arterial pressure, stroke volume and cardiac output drop suddenly, accompanied by a reflex increase in heart rate. The mean arterial pressure increases due to the increased systemic resistance (after about 20 seconds) although it does not return to the original level. Cardiac output shows an oscillation after the initial drop, but there is no reference to this effect in the data. Finally cardiac output remains lower than the initial value. Stroke volume does not show the oscillations, and so the oscillations of cardiac output are probably due to the overshoot of the heart rate, (possibly the dynamics of the heart rate controller are too rapid). The estimated total systemic resistance (ETSR) shows a very sharp decrease followed by a smooth vasoconstrictive rise. The decrease does not occur in the human, and is an artifact in the model caused by the way in which ETSR is calculated. In general the model has an overall qualitative similarity with the descriptive data, (although some of the rapid effects are suspicious), satisfying the first empirical criterion.

Secondly, some quantitative features of the overall net changes in key variables are compared in Table 6.8. The reduction of stroke volume in the model (20%) is correctly in the normal range (10 - 50%). Arterial pressure does not regain its initial value (as is usually reported), and is significantly lower (-7%). Consequently heart rate is higher (reflex) and increases more than the data. This elevated heart rate has the effect that cardiac output is decreased less in response to reduced stroke volume. The primary cause for these discrepancies is evidently in the change of systemic resistance (+4% as opposed to +20% in the data) which is insufficient to bring the arterial pressure back to normal.

These considerations have certain implications for the validity of the controlled haemodynamics.

1. The response of the heart and circulation submodels to the postural change is to produce a vein pooling and reduced stroke volume as in the data. This gives greater confidence in the validity of these submodels.

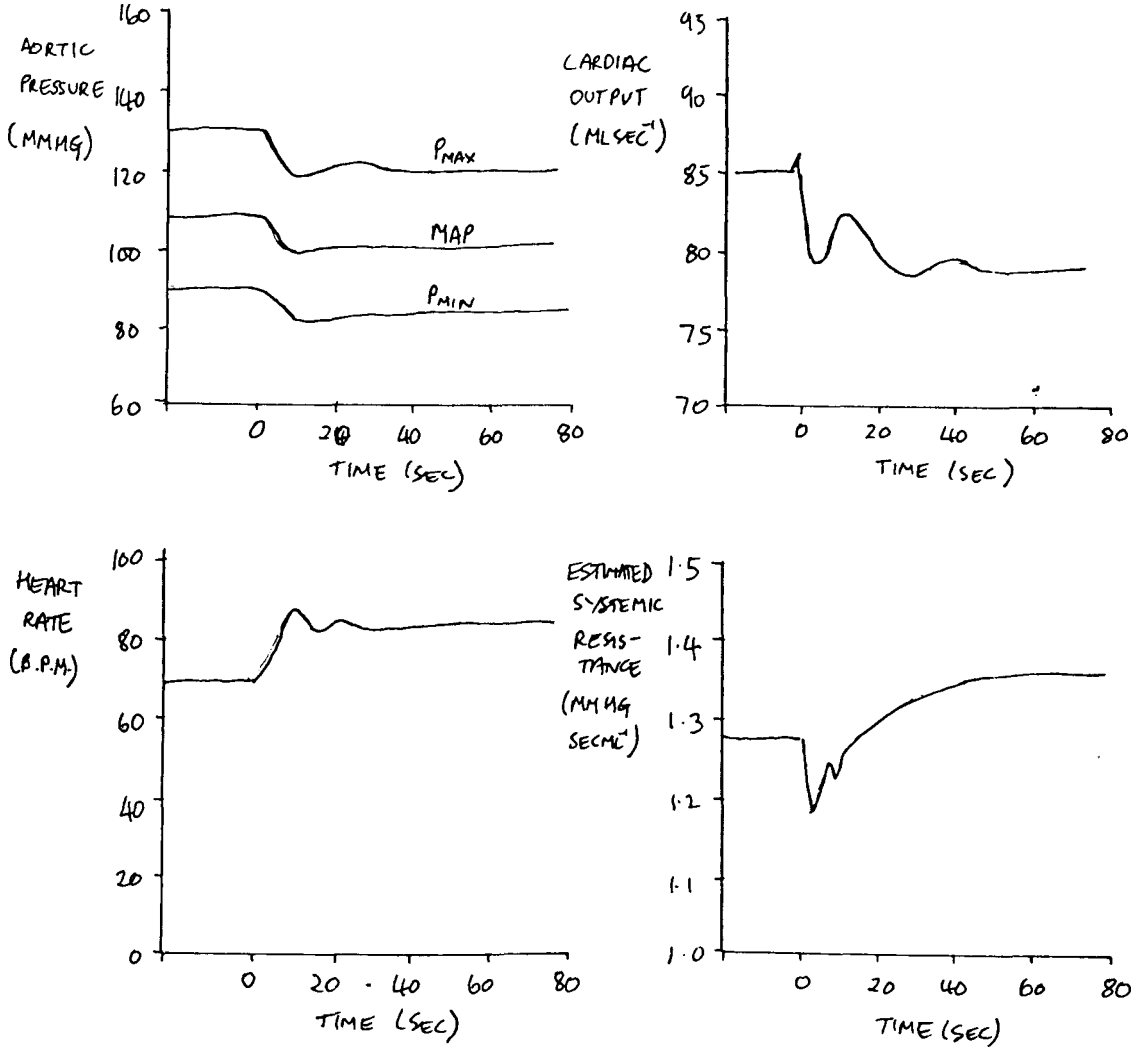


FIGURE 6.12. MODEL RESPONSE TO AN UPRIGHT TILT.

VARIABLE	DATA	MODEL
MEAN ARTERIAL PRESSURE	LITTLE CHANGE	↓ 7%
HEART RATE	↑5-10 BPM	↑ 14 BPM
STROKE VOLUME	↓ 10-50 %	↓ 20%
CARDIAC OUTPUT	↓ 20%	↓ 6%
SYSTEMIC RESISTANCE	↑ 20%	↑ 4%

SOURCE:

MOUNTCASTLE (1974)

TABLE 6.8. CHANGES IN MODEL AND HUMAN AFTER A CHANGE TO AN UPRIGHT FROM A RECUMBENT POSITION.

2. The heart rate reflex of the model to decreasing pressure is rapid and effective, as in the data.

3. The reflex vaso- and veno-constriction in the model is too small in response to decreasing arterial pressure. This reduces the validity of the submodel of neural control of the circulation.

(ii) Blood volume changes

Pullen simulated the effect of a sudden blood loss (e.g. following a haemorrhage) and hypervolaemia (1976, pp 114-115) by instantaneously changing the volume of the head and arms veins segment. The model responses are shown in fig. 6.13. It can be seen that removal of blood produces large, slow effects, whereas increase of blood volume of the same amount ($\simeq 10\%$) produces rapid but smaller changes. This indicates the assymmetry in the neural control, and the fact that baroreceptors are sensitive to positive rates of change of pressure. In both simulations, stroke volume changes significantly, demonstrating that the venous return has changed and that the heart and circulation submodels are behaving correctly.

A descriptive account, with some quantitative details, of the effect of a haemorrhage is given by Lippold and Winton (1979, p.224). The main points are summed up below:

1. Blood volume is reduced, veins are less well filled and output of heart and arterial pressure drop.
2. But, changes are small owing to prompt compensating mechanisms.
3. In an anaesthetised dog, when 10% of the blood volume is removed (from the gut), arterial pressure does not fall, but is maintained by an increase in perpheral resistance.
4. But, if the quantity of blood lost is sufficient to reduce cardiac output to 30 to 50 percent below normal, the arterial pressure falls.

In the model, cardiac output drops by 23% and so a significant drop in arterial pressure is not expected. However, mean arterial pressure drops by 22%. This error can be traced to the inadequate increase of systemic resistance. A similar failure to produce sufficient increase of systemic resistance with falling pressure was found in the previous section.

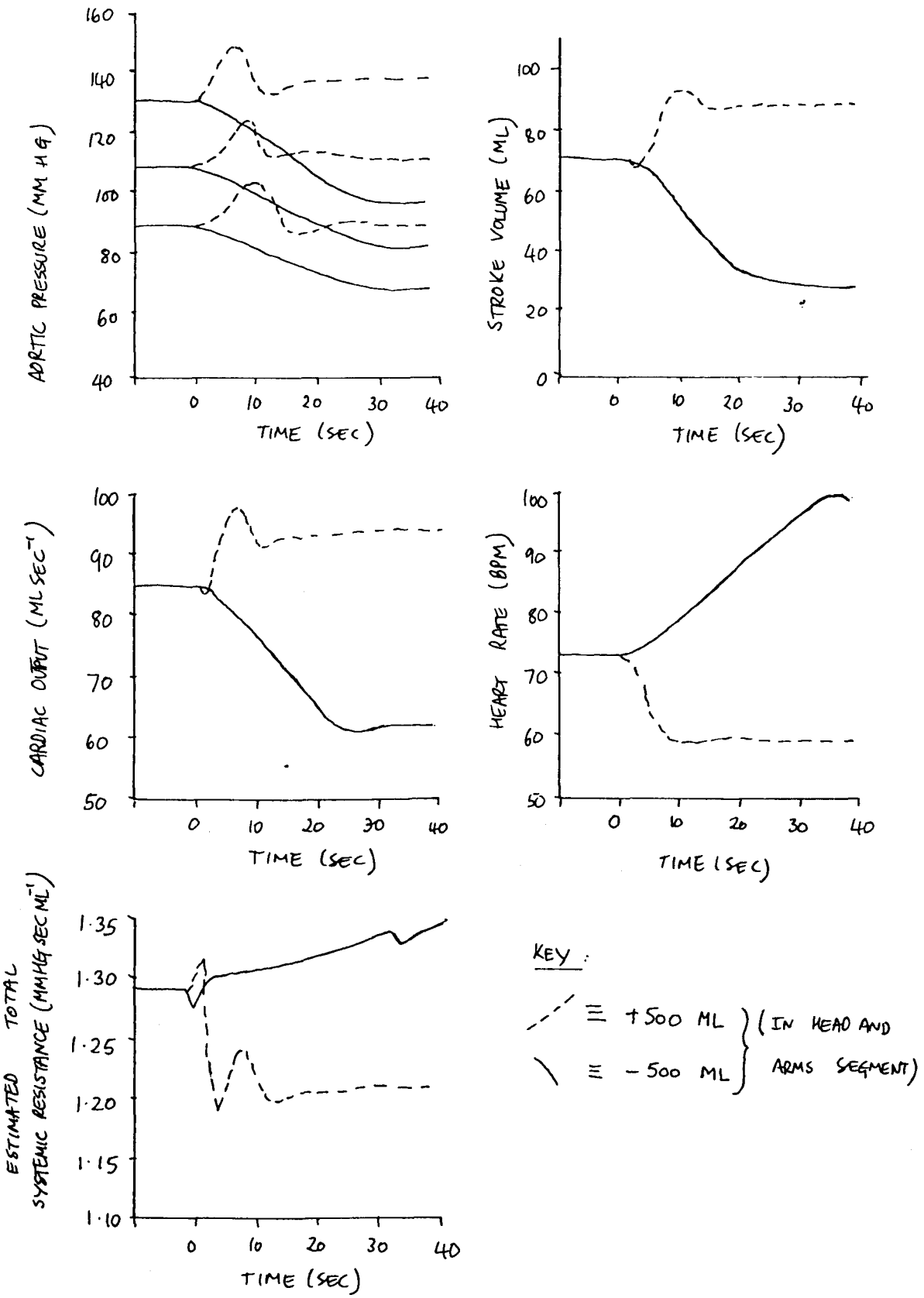


FIGURE 6.13. MODEL RESPONSE TO BLOOD VOLUME CHANGES.

Notice that the heart rate reflex is large (+43%) due to the continued low pressure.

However, in the case of blood volume increase mean arterial pressure rises by only 7.5%, and is rapidly controlled, the systemic resistance drops rapidly and significantly, and the heart rate does not change so much.

These considerations indicate that the range of validity of the model is limited to environmental changes that produce positive pressure changes. For negative pressure changes the model is invalid.

(iii) Cardiac pacing.

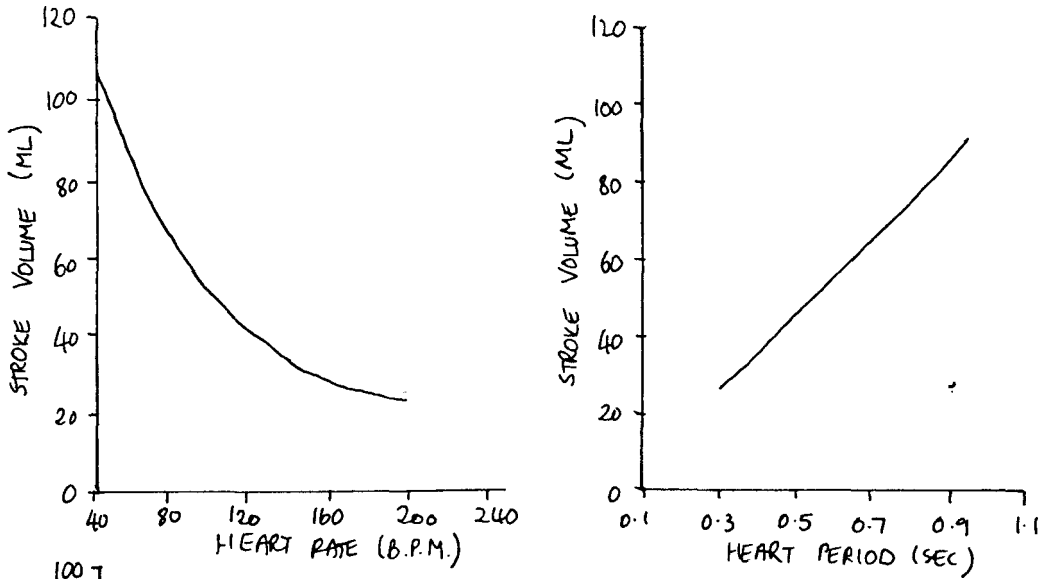
Pullen (1976, pp. 116-117) simulated the experiments of Noble et al. (1966) investigating the effect of changing the heart rate on the cardiovascular function of the conscious dog by pacing the heart using an implanted right atrial electrode. They found that as heart rate increased, the stroke volume fell and the cardiac output either increased or changed very little. A close linear relationship was established between stroke volume and heart rate.

Simulated experiments were performed on these using values of heart period ranging from 0.3sec. to 1.4sec. The results are shown in fig. 6.14. Pullen's model did not reproduce a linear relationship between stroke volume and heart rate, but the relationship between stroke volume and heart period was strikingly linear. The model also showed little variation of cardiac output in accordance with Noble's results.

(iv) Valsalva manoeuvre.

This experiment, originally devised by Valsalva (1666-1723) to test the patency of the eustachian tubes, consists of an attempt to expirate against a resistance (e.g. by closing the glottis, or against a column of mercury). "The effects of this manoeuvre on arterial blood pressure and heart rate are now used as a test of autonomic cardiovascular control mechanisms and of ventricular function" (Hawker, 1979, p.75). For this reason the Valsalva manoeuvre is also a good test of the validity of the controlled haemodynamics.

During the manoeuvre, intra-thoracic and intra-abdominal pressures rise significantly. This increase is transmitted onto the blood



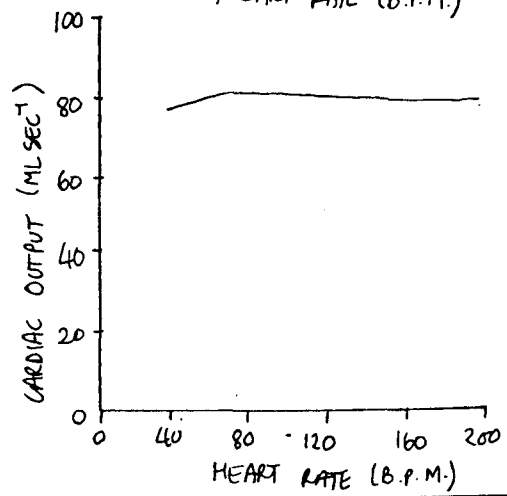
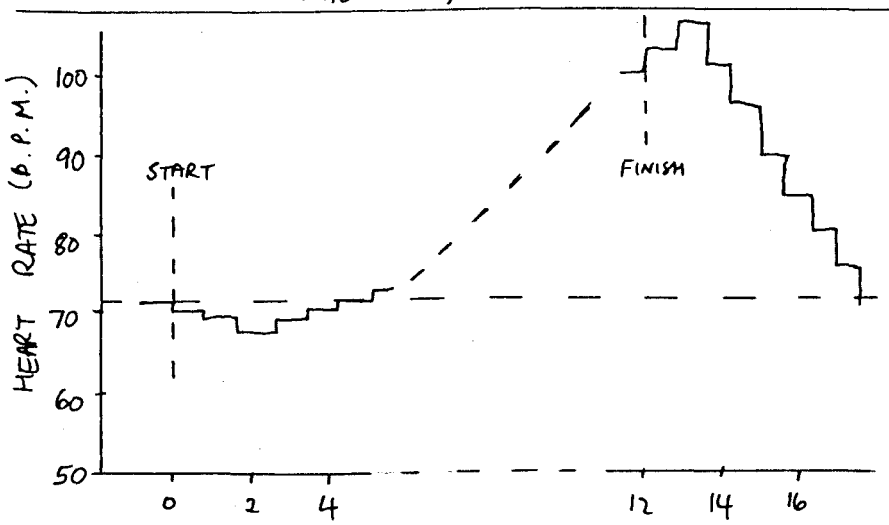
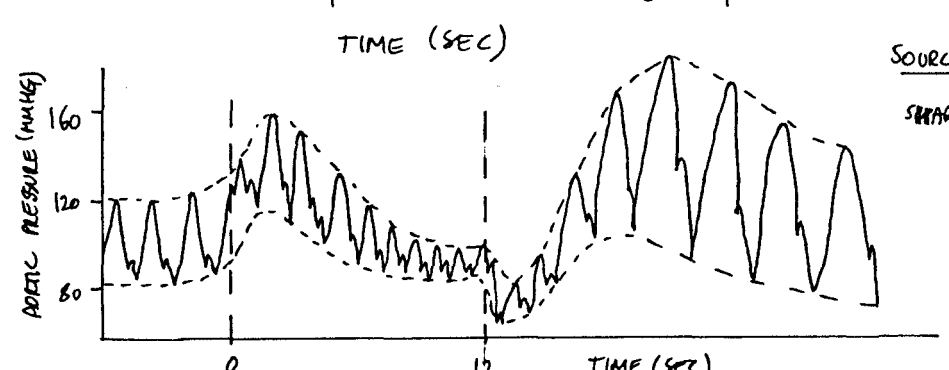


FIGURE 6.14. EFFECT ON THE MODEL OF PACING THE HEART.



SOURCE:
BENEKEN & DEWIT (1967)



SOURCE:
SHARPEY-SCHAFER (1965)

FIGURE 6.15. EFFECTS OF VALSALVA MANOEUVRE ON HEART RATE AND ARTERIAL PRESSURE IN MAN.

contained in the heart and vessels in these regions and venous return is reduced from the periphery. Cardiac output falls and aortic pressure drops. There is a prompt control by the CNS with tachycardia and vasoconstriction, and aortic pressure subsequently rises.

After the manoeuvre, which can only be held for a few seconds, blood pressure drops suddenly (transmitted) and the heart rate increases for a few beats (dynamic delay). Venous return increases as blood floods back into the chest cavity. Cardiac output rises sharply, and produces a large rise in arterial pressure (the vasculature is still constricted), much greater than before the manoeuvre. There is a reflex bradycardia to compensate and eventually the cardiovascular system returns to equilibrium ($\simeq$30sec). This "overshoot" is a characteristic feature of a normal response to a Valsalva manoeuvre (see, e.g. Sharpey-Schafer 1965). Fig. 6.15 shows the quantitative dynamic effects of a Valsalva manoeuvre on the heart rate and blood pressure of a normal human during which intra-thoracic pressure rises by about 40mmHg (taken from Beneken and De Wit, 1967, p.37 and Sharpey-Schafer, 1965).

In the model, the manoeuvre is simulated by setting intra-thoracic and intra-abdominal pressures to +40mmHg for a period of 12 seconds (after Beneken and De Wit, 1967). The model responses of arterial pressure, stroke volume, cardiac output, heart rate and estimated total systemic resistance are shown in fig. 6.16. Firstly, the qualitative similarity of the data (descriptive account and fig. 6.15) and the model will be examined, then a quantitative analysis of features of the heart rate response will be made to investigate indirectly the validity of the submodel of heart rate control.

The model response shows a large drop in stroke volume and cardiac output, indicating that the heart and circulation submodels are responding correctly to the raised pressures. Aortic pressure rises suddenly by about +40mmHg (transmitted directly), but shows a subsequent decline and reduction of pulse pressure (as in the data, fig. 6.15). Heart rate shows an initial decline (due to the sudden rise in pressure) but then there is a substantial tachycardia in response to the falling arterial pressure, in agreement with the data obtained by Beneken and De Wit (1967) shown in fig. 6.15.

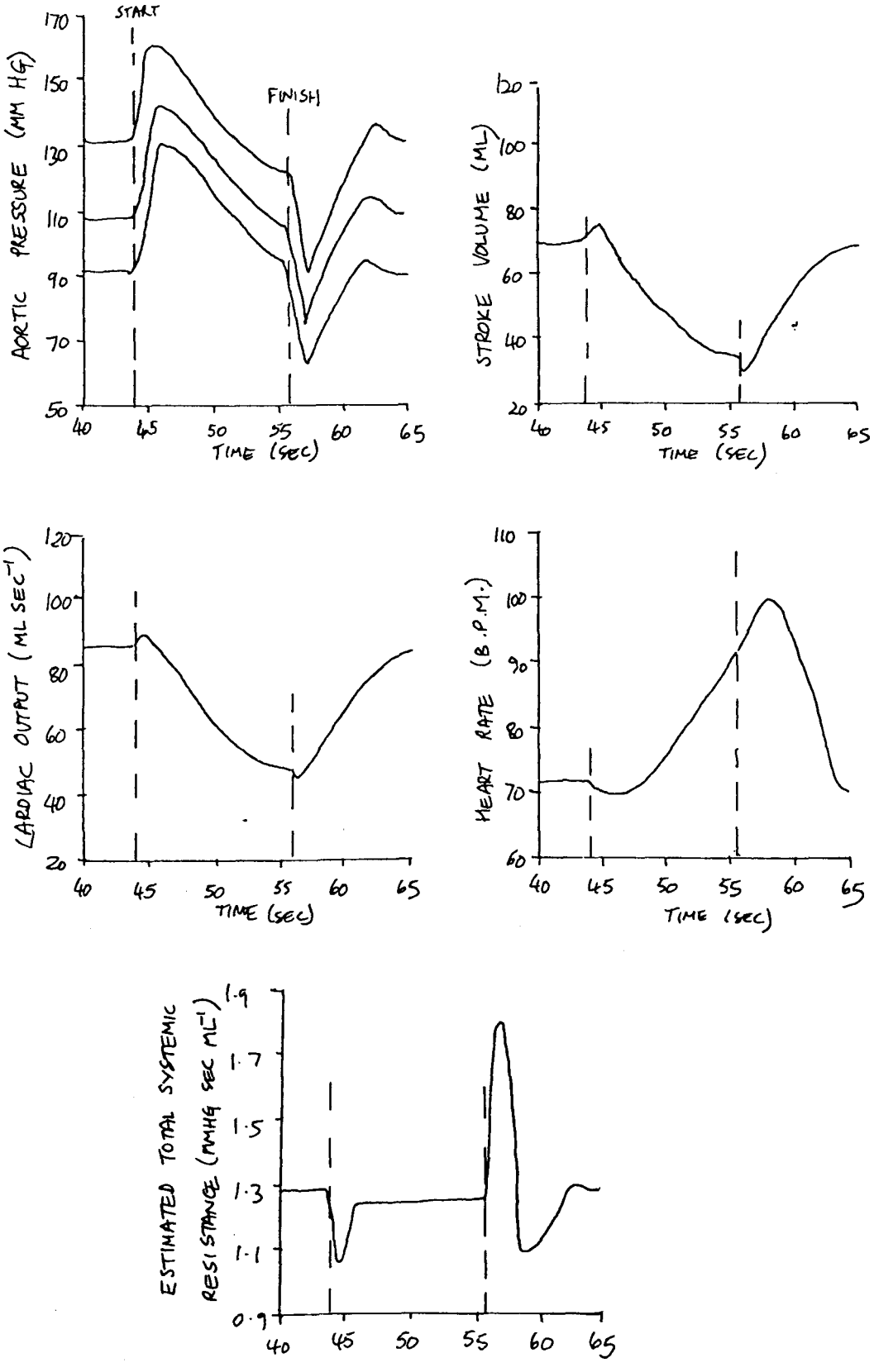


FIGURE 6.16. MODEL RESPONSE TO A VALSALVA MANOEUVRE (INTRATHORACIC AND INTRA-ABDOMINAL PRESSURES SET TO +40 MMHG FOR 12 SEC).

A shorthand symbolic notation (described in chapter 5) will now be used to compare the qualitative aspects of the mean arterial pressure (MAP) in the model and data (fig. 6.17).

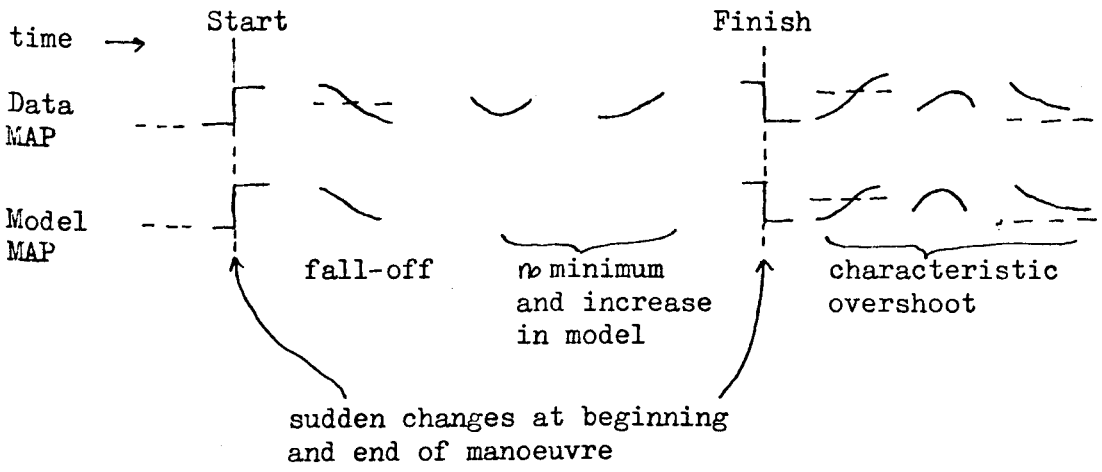


Fig. 6.17. Symbolic Representation of Qualitative Features of Model and Data Mean Arterial Pressure Responses.

The model reproduces the major features except the turning point and increase of MAP during the manoeuvre. This suggests that the reflex vasoconstriction is not sufficient in the model to prevent further decrease of arterial pressure, although there is an overshoot after the end of the manoeuvre. However, as shown in fig. 6.18, the overshoot in the human is much greater than that in the model, and arterial pressure exceeds the maximum value during the manoeuvre.

After the manoeuvre, the heart rate in the model returns quickly to normal, and does not exhibit the large bradycardia in the data (fig. 6.15). This is because of the lack of pressure overshoot due to inadequacy of the vasoconstrictive reflex, and should not be taken as an indication of the invalidity of the submodel of heart rate control. In fig. 6.16 it can be seen that systemic resistance does not rise during the manoeuvre (n.b. the "spikes" in the response are artifacts of the method of calculation); in fact, it shows a slight decrease. This implies an inaccuracy in the submodel of neural control of vascular and venous tone in the response to falling pressure, as inferred in the analysis of the model response to postural changes (see (i) above). (In tabetic subjects there is a similar lack of overshoot owing to the absence of reflex vasoconstriction, (Sharpey-Schafer, 1965, p.1878) and this is why the Valsalva manoeuvre is a critical test of the model).

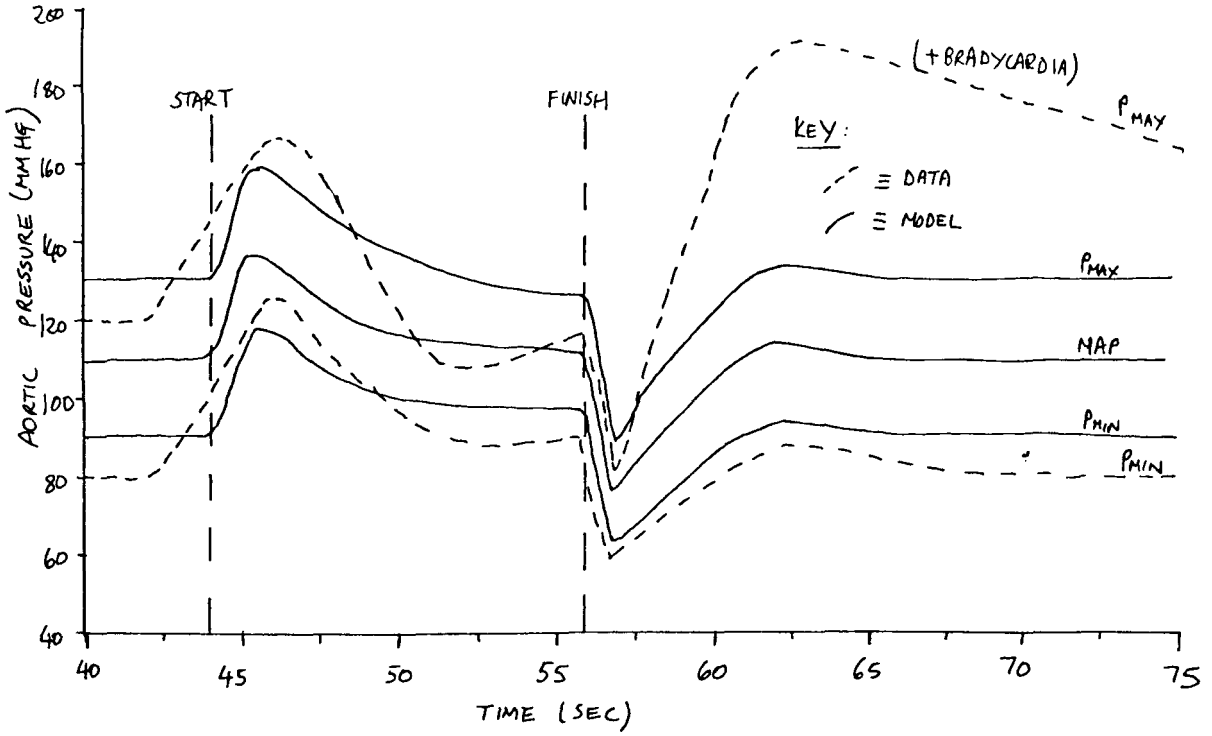
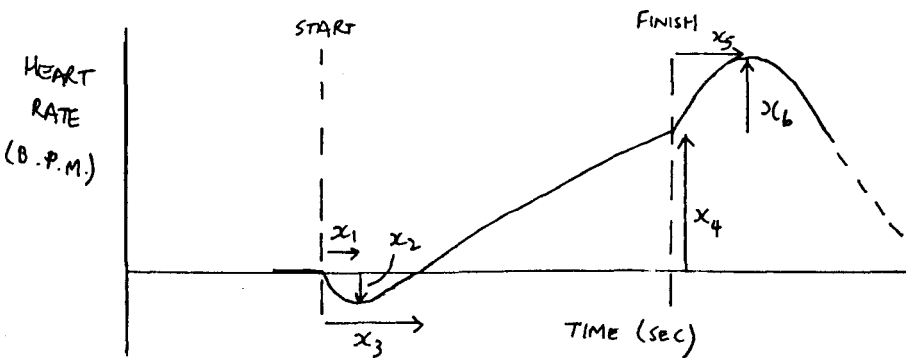


FIGURE 6.18. SYSTOLIC (P_{MAX}), MEAN (MAP), AND DIASTOLIC (P_{MIN}) ARTERIAL PRESSURE WAVEFORMS FOR MODEL AND HUMAN DURING A VALSALVA MANOEUVRE. (DATA FROM SHARPEY-SCHAFER, 1965, p. 1878).



$x_1 \equiv$ NO. OF BEATS TO MINIMUM HEART RATE (HR).

$x_2 \equiv$ INITIAL HR - MINIMUM HR.

$x_3 \equiv$ NO. OF BEATS UNTIL HR REGAINS INITIAL VALUE.

$x_4 \equiv$ NET INCREASE OF HR OVER MANOEUVRE.

$x_5 \equiv$ NO. OF BEATS AFTER FINISH UNTIL MAXIMUM HR.

$x_6 \equiv$ INCREASE OF HR AFTER FINISH.

FIGURE 6.19. FEATURES OF HEART RATE RESPONSE.

In the Valsalva manoeuvre, the carotid sinus and aortic arch baroreceptors are very differently stressed, and so the model response changes significantly with variations of α, the relative contribution of the carotid sinus and aortic arch areas to the CNS input function, (see section 6.3.3, fig. 6.4). Over the manoeuvre, the heart rate response of the model (fig. 6.16) is qualitatively similar to the data (fig. 6.15) and the error introduced by the lack of reflex vasoconstriction does not occur until several seconds after the end of the manoeuvre. A comparison of the heart rate response in the model and data is therefore suitable for assessing the validity of the value of α in the model.

Pullen (1976) used the value $\alpha = 0.7$ in the model giving the responses shown in fig. 6.16 This corresponds to a greater contribution from the carotid sinus baroreceptors, and is based on the work of Dampney et al. (1971) on dogs. However, as pointed out in the theoretical considerations of section 6.5.6.3 the aortic arch area may play a more important role in heart rate control. Beneken and De Wit (1967) ran their model with a range of values of α and concluded that $\alpha = 0.5$ produced a heart rate response of greatest similarity to the data by visual inspection.

Their experiment is repeated here using the Pullen model, which has more detailed baroreceptor and neural control dynamics and using a systematic method of feature comparison. The basis of this method was described in chapter 5. Its advantage, over loss-functional comparisons, is that important aspects of the data (e.g. undershoot, overshoot etc.) can be stressed and comparisons can still be made when absolute values data and model are not sufficiently close (i.e. it constitutes an effective normalisation procedure).

The features extracted from the heart rate response for comparison are denoted by x_i, $i = 1, n$, and are depicted in fig. 6.19. Note that no features are defined beyond a few seconds after the manoeuvre. This is because the arterial pressure does not show the massive overshoot, and so the heart rate reflex is no longer credible in the model. The features of the data and model responses for a range of α from 0.1 to 0.9 are shown in table 6.9. In order to compare each model the fractional differences f_i for each feature ($i = 1,n$) are determined:

SOURCE		x_1	x_2	x_3	x_4	x_5	x_6
DATA		3	1.9	4	29.2	2	8.2
MODEL	α						
M_1	0.9	5	5.4	7	17.5	5	16.2
M_2	0.7	3	3.7	6	17.7	5	12.3
M_3	0.5	1	3.8	6	17.3	5	9.3
M_4	0.3	2	1.2	4	20.1	4	6.1
M_5	0.1	2	0.3	2	23.7	3	2.3

(DATA: BENEKEN AND DEWIT (1967), SEE FIG. 6.15).

TABLE 6.9. FEATURES OF MODEL AND DATA HEART RATE RESPONSES WITH A RANGE OF VALUES OF α (RELATIVE CONTRIBUTION OF CAROTID SINUS AND AORTIC ARCH BARORECEPTORS TO CNS INPUT FUNCTION).

MODEL M_i	α	f_1	RANK	f_2	RANK	f_3	RANK	f_4	RANK	f_5	RANK	f_6	RANK	F	F′ (WITHOUT f_3)
M_1	0.9	0.67	4	1.84	5	0.75	5	-0.40	4	1.5	3	0.98	5	0.494	0.519
M_2	0.7	0.0	1	0.95	3	0.5	2	-0.39	3	1.5	3	0.5	3	0.610	0.682
M_3	0.5	-0.67	4	1.0	4	0.5	2	-0.41	5	1.5	3	0.13	1	0.588	0.648
M_4	0.3	-0.33	2	-0.37	1	0.0	1	-0.31	2	1.0	2	-0.26	2	0.726	0.797
M_5	0.1	-0.33	2	-0.84	2	-0.5	2	-0.19	1	0.5	1	-0.72	4	0.661	0.660

TABLE 6.10. FRACTIONAL DIFFERENCES (f_i) BETWEEN MODEL AND DATA FEATURES DURING A VALSALVA MANOEUVRE.

$$f_i = \frac{x_{i\ \text{model}} - x_{i\ \text{data}}}{x_{i\ \text{data}}}$$

For each feature, the models may be ranked by closeness to the data. The results are shown in Table 6.10. It can be seen that no model has the smallest fractional feature difference for all features, although M_4 is always at least second closest to the data. A figure of merit F, for each model is calculated, based on the average fractional difference:

$$F = \frac{1}{1 + \frac{1}{n}\sum_{i=1}^{n} f_i} \qquad \ldots (6.16)$$

where $\frac{\sum f_i}{n}$ = average factional difference.

$F \in [0,1]$ where F= 1 = no model/data difference
F = 0.5 = 100% fractional difference
F = 0 = ∞ fractional difference

The results are shown in table 6.10. The values of F induce the following ordering on the models; in a manner analagous to Reggiani and Marchetti's concept of model adequacy (1975):

$M_4 > M_5 > M_2 > M_3 > M_1$ where $M_1 \equiv \alpha = 0.9$
$M_2 \equiv \alpha = 0.7$
$M_3 \equiv \alpha = 0.5$
$M_4 \equiv \alpha = 0.3$
$M_5 \equiv \alpha = 0.1$

or, $\alpha = 0.3$ produces a model closest to the data features. This corresponds to a 70% contribution from the aortic arch baroreceptors, and invalidates the value $\alpha = 0.7$ used by Pullen. However, the model is simply modified using the new value. The responses of mean arterial pressure, heart rate and estimated total systemic resistance of the modified model are shown in fig. 6.20. The important differences (apart from the improvement of the heart rate response) are that arterial pressure shows an increased overshoot, and there is some vasoconstriction (c.f. fig. 6.16) although it is still insufficient to produce the spectacular overshoot in the data.

The above illustrates the great potential of using the model as a

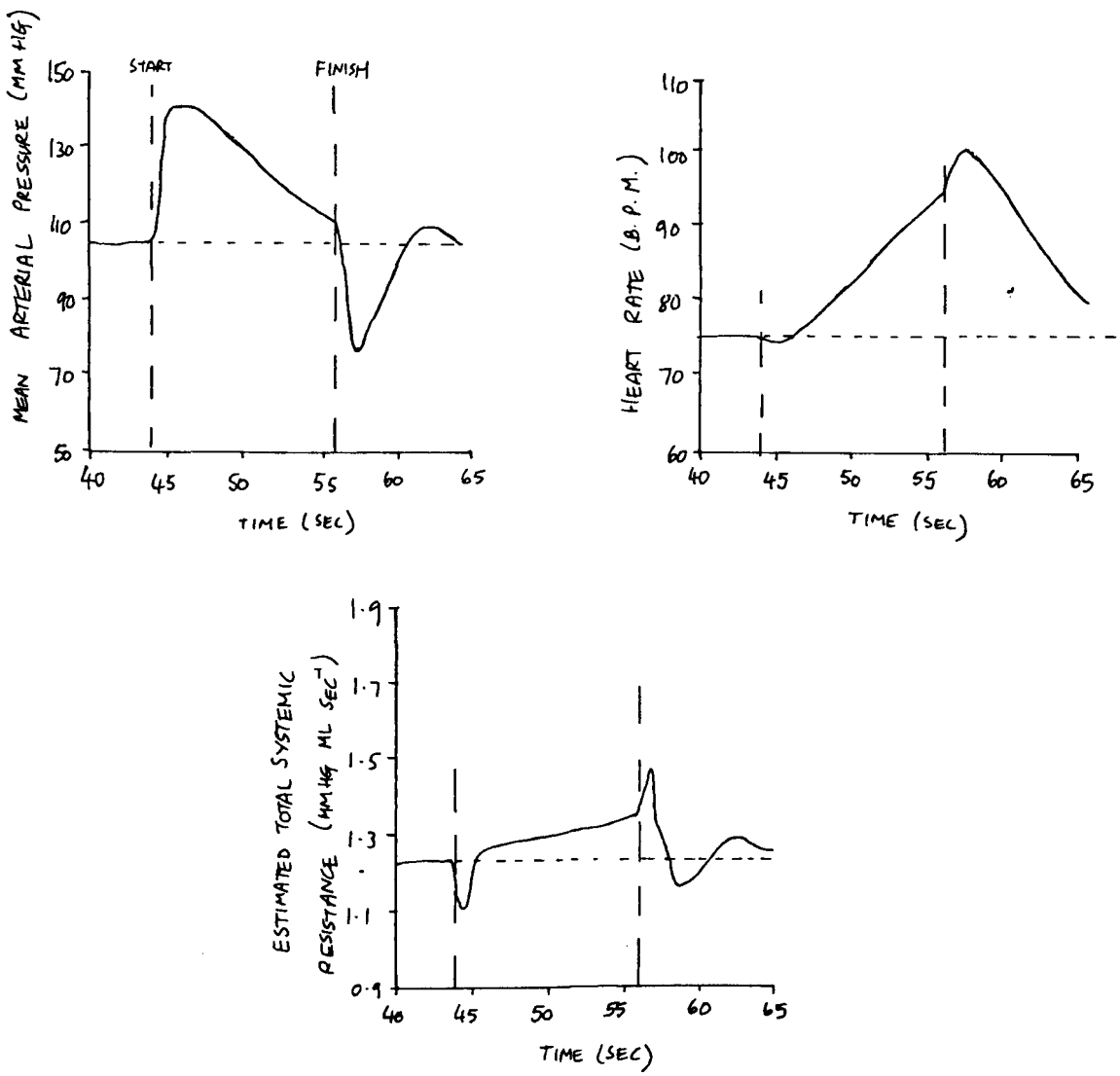


FIGURE 6.20. RESPONSE OF MODIFIED MODEL ($\alpha = 0.3$) TO VALSALVA MANOEUVRE.

test bed for hypothesis testing - providing evidence for the dominance of the aortic arch baroreceptor area in heart rate control - however, great care must be exercised in the choice of features, and the interpretation of results. In particular, the model exhibits defects in the control of systemic resistance, and this manifests clearly in the heart rate response beyond a few seconds after the end of the manoeuvre. For this reason, features may only meaningfully be taken from the response up to the end of the manoeuvre. Nevertheless, this modification of the model receives additional validation in next empirical test.

6.5.7.1.2.3 Simulation of Mancia's Experiment

A physiological experiment by Mancia et al. (1978) was designed to investigate baroreceptor control of heart rate in man. It consisted of two separate experiments: in one, the reflex change of heart rate due to an overall drug induced pressure change was examined; and in the other the heart rate response to direct stimulation of the carotid sinus baroreceptors was achieved using a sealed chamber around the neck of the subject (an experimental innovation was made whereby carotid transmural pressure could be varied both positively and negatively). Their results are shown in fig. 6.21. When both aortic arch and carotid sinus baroreceptors are stimulated in the drug induced pressure changes the slope of the heart period (k_1) is about 3 times that when only the carotid sinus baroreceptors are stimulated (k_2). The conclusion of Mancia et al. was that the aortic arch baroreceptors play a more important role in heart rate control.

Their experiments were simulated on the model. Changes in carotid transmural pressure (ΔCTP) were simply done by adding a term to the input of the baroreceptor submodel. The drug induced pressure changes were simulated using the full model with pharmacodynamics with the appropriate drugs. Mancia et al. used drugs that were vasoactive and free from direct cardiac effect (the α - adrenergic stimulant phenylephrine as a pressor drug, and amyl nitrate as a depressor drug). In the model these are simulated by using local drug effects of vasoconstriction or vasodilatation only. In simulating the drug induced changes, it is not important that the model has not been validated at this level yet (level 4 - next section), since it is only the neural control response to a systemic arterial pressure change that is of interest, not how

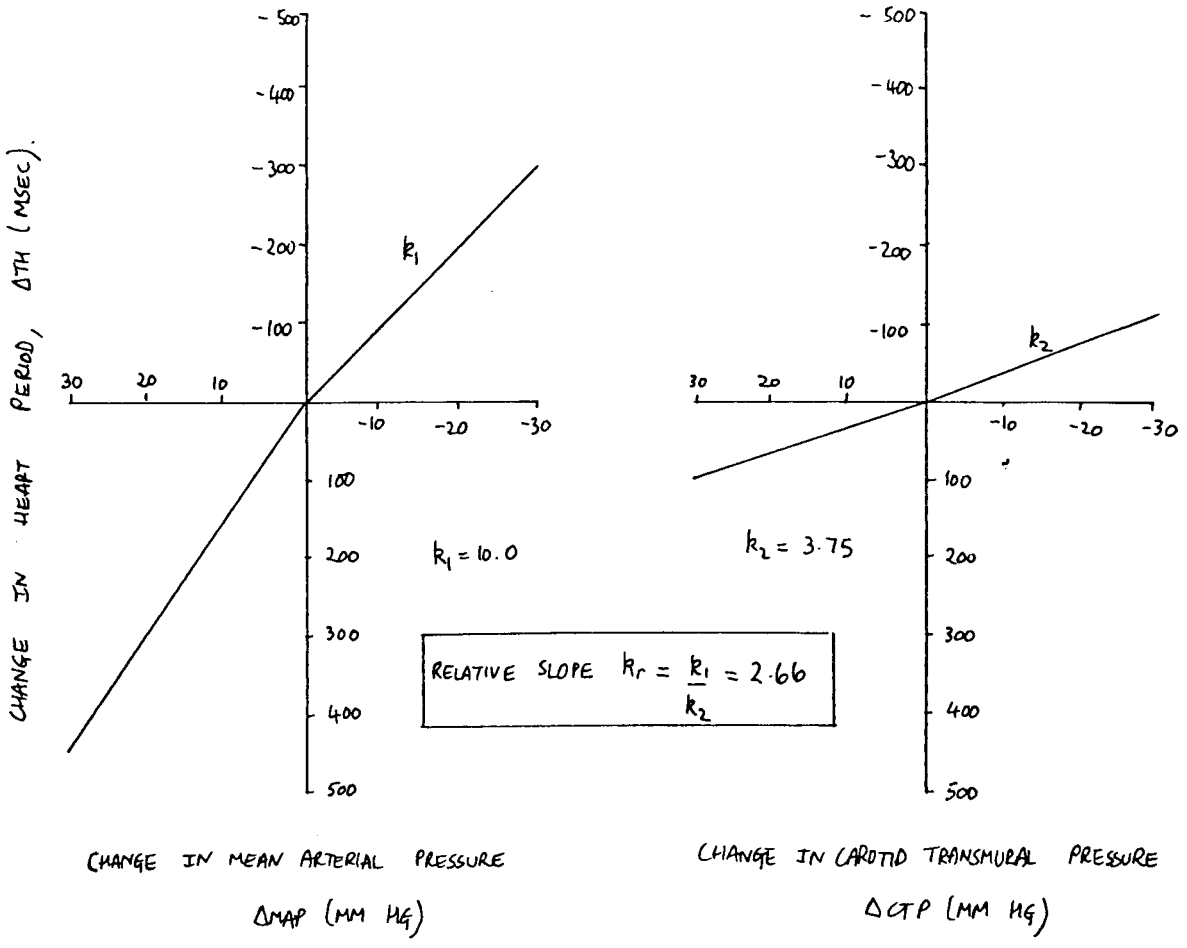


FIGURE 6.21. CHANGE OF HEART PERIOD IN MAN DUE TO DRUG INDUCED CHANGES IN MEAN ARTERIAL PRESSURE (LEFT) AND CHANGES IN CAROTID TRANSMURAL PRESSURE (RIGHT) (FROM MANCIA ET AL., 1978).

α	k_1	k_2	$k_r = k_1/k_2$
0.9	3.6	3.4	1.05
0.7	3.6	2.5	1.44
0.5	4.0	2.0	2.00
0.3	4.0	1.4	2.85
0.1	4.0	0.8	5.00

TABLE 6.11. VARIATION IN RELATIVE SLOPE (k_r) OF MODEL RESPONSE TO CHANGES IN MEAN ARTERIAL PRESSURE (k_1) AND CAROTID TRANSMURAL PRESSURE (k_2) WITH DIFFERENT VALUES OF α.

this change came about. However, for safety, only the heart rate response to falling arterial pressure is used since this has received confirmation in the validation tests of section 6.5.7.1.2.2, (namely (i) postural changes and (iv) Valsalva manoeuvre).

The results of the model simulation for a range of values of α, the relative contribution of carotid sinus and aortic arch baroreceptors are shown in table 6.11. The feature of interest is the relative slope $k_r = k_1/k_2$, since this reflects the relative heart period sensitivity of the two experiments. In fig. 6.21, from the data, $k_r = 2.66$ and from table 6.11 it can be seen that $\alpha = 0.3$ produces $k_r = 2.85$ which is the closest value. This corresponds to a 30% contribution from the carotid sinus and a 70% contribution from the aortic arch areas to the CNS input function in the model. This confirms the conclusions of Mancia et al. (1978) and agrees with the quantitative feature analysis of the Valsalva manoeuvre (section 6.5.7.1.2.2).

As a critical proviso, care should be taken in generalising the validity of this result. It has been established in the model with regard to heart rate control and falling pressure; it has not been demonstrated for the other controllers (myocardial contracility, vasometor tone, and venomotor tone) or for increasing arterial pressure, and inaccuracy in the control of systemic resistance has already been noticed.

6.5.7.1.2.4 Sensitivity Analysis

Al-Dahan (1981) has undertaken an extensive sensitivity analysis of the controlled haemodynamics submodel with respect to many of its parameters. The importance of sensitivity analysis in model validation was discussed in Chapter 5, but briefly it can be used for the following purposes:

1. To check the stability of the model. This has implications for algorithmic/simulation validity, but also for theoretical validity as the cardiovascular system is known to be very stable.
2. To gain confidence in parameter values. Many parameters cannot be measured directly, and sensitivity analysis provides a means for estimating the uncertainty of parameters determined indirectly using the model.

3. To determine a subset of critical parameters which produce large sensitivity coefficients and on which the model's behaviour is highly dependent.

The results of the analysis which were obtained by systematic parameter variation, are shown in table 6.12. There are a very large number of possibilities with such an experiment; for instance, for 20 parameters (chosen on a priori grounds) examining 6 variables over 3 different tests with 4 parameter changes (e.g. ± 15%, ±30%) each time, produces 1440 results. Table 6.12 therefore only shows the most significant results, and for only 2 tests (equilibrium conditions and the Valsalva manoeuvre). In table 6.12 the results for the 10 most sensitive parameters are given in terms of their sensitivity coefficients and relative sensitivities on mean arterial pressure. If x is the normal value of mean arterial pressure and Δx is its charge due to a variation Δp_i in parameter P_i then the <u>sensitivity coefficient</u> $\mathcal{S}_i$ for this parameter is defined by

$$\mathcal{S}_i = \frac{\Delta x}{\Delta p_i} \qquad \ldots (6.17)$$

and the <u>relative sensitivity</u> γ_i by

$$\gamma_i = \frac{\Delta x / x}{\Delta p_i / p_i} \qquad \ldots (6.18)$$

The relative sensitivity represents the degree to which a fractional change in a parameter is reflected in the fractional change of the variable, and allows the relative importance of different parameters to be assessed. In table 6.12 the parameters have been ordered by their relative sensitivities in the Valsalva Manoeuvre.

The sensitivity coefficients have been determined by changes of ± 15% and ± 30% in each parameter, and as can be seen in table 6.12 have produced no instabilities($\gamma_i \gg 1$). Physiologically, the cardiovascular system remains stable despite natural variations in its properties, and therefore the similar behaviour of the submodel of controlled haemodynamics confers validity on this "system property" of the model. Within this set of important parameters there is yet a further critical subset - the neural parameters, k_{18} (gain) and k_{16} (relative baroreceptor contribution, α)

PARAMETER	NORMAL VALUE (p)	EQUILIBRIUM CONDITIONS		VALSALVA MANOEUVRE		DEFINITION OF PARAMETER
		SENSITIVITY COEFFICIENT $\mathcal{S}$	RELATIVE SENSITIVITY γ	SENSITIVITY COEFFICIENT $\mathcal{S}$	RELATIVE SENSITIVITY γ	
k_{18}	1.0	50.01	0.458	55.83	0.815	NEURAL GAIN THRESHOLD
k_{16}	0.7	0.952	0.0061	-23.78	-0.243	$\alpha \equiv$ RELATIVE CONTRIBUTION OF BARORECEPTORS TO CNS INPUT
V_{UPV}	460.0	-0.0294	-0.124	-0.021	-0.144	UNSTRESSED VOLUME OF PULMONARY VEINS
P_{THN}	-4.0	-1.17	0.043	1.04	-0.061	NORMAL THORACIC PRESSURE
a_{RVS}	0.3	30.58	0.084	10.00	0.044	SYSTOLIC ELASTANCE OF RIGHT VENTRICLE
a_{LVS}	1.5	4.44	-0.061	-1.67	-0.036	SYSTOLIC ELASTANCE OF LEFT VENTRICLE
a_{RAD}	0.05	65.52	0.030	-46.72	-0.034	DIASTOLIC ELASTANCE OF RIGHT ATRIUM
a_{RAS}	0.15	26.94	0.037	8.91	0.019	SYSTOLIC ELASTANCE OF RIGHT ATRIUM
V_{URA}	30.0	-0.036	-0.099	-0.0045	-0.019	UNSTRESSED RIGHT ATRIAL VOLUME
R_{RARV}	0.003	-167.4	-0.046	388.2	0.017	RIGHT ATRIO-VENTRICULAR VALVULAR RESISTANCE

SENSITIVITY COEFFICIENT $\mathcal{S} = \Delta x / \Delta p$

RELATIVE SENSITIVITY $\gamma = (\Delta x / x) / (\Delta p / p)$

FOR EQUILIBRIUM CONDITIONS, $x \equiv$ MEAN ARTERIAL PRESSURE IN STEADY STATE (= 109.2 MMHG)

FOR VALSALVA MANOEUVRE, $x \equiv$ MEAN ARTERIAL PRESSURE AT T = 52 SEC (NEAR END OF MANOEUVRE, = 68.5 MM HG)

TABLE 6.12. RESULTS OF SENSITIVITY ANALYSIS ON CONTROLLED HAEMODYNAMICS SUBMODEL.

PARAMETER	CONFIDENCE	PARAMETER	CONFIDENCE
k_{18}	±12%	a_{LVS}	±277%
k_{16}	±41%	a_{RAD}	±294%
V_{UPV}	±69%	a_{RAS}	±526%
P_{THN}	±163%	V_{URA}	±526%
a_{RVS}	±227%	R_{RARV}	±588%

TABLE 6.12. CONFIDENCE RANGES OF PARAMETERS (BASED ON INVERSE SENSITIVITY FOR AN ALLOWABLE ± 10% VARIATION OF MEAN ARTERIAL PRESSURE).

and the unstressed pulmonary venous volume, V_{UPV} - whose relative sensitivities are about 3 - 10 times that of the following group. This group basically comprises heart function parameters, ventricular elastances followed by atrial elastances.

The following interpretation can be given to these results: the controlled haemodynamics submodel is very tightly controlled by the neural control and is largely affected by pulmonary venous properties (determining filling of the left atrium) as well as cardiac function properties. This interpretation is consistent with physiological understanding of overall short-term cardiovascular functioning and therefore provides evidence of the overall validity of the submodel.

The sensitivity analysis results may be used inversely to determine confidence ranges in parameter values for an allowable variation of mean arterial pressure. Table 6.13 contains results (based on the Valsalva manoeuvre) for a range of ± 10% variation on mean arterial pressure calculated from:

$$100 \frac{\Delta p_i}{p_i} = \pm \frac{10.0}{\gamma_i} \% \qquad \ldots (6.19)$$

With the exception of the first three parameters (neural gain, α, and unstressed pulmonary venous volume) the confidence interval is very large. This means that the model response is invariant to large changes in these parameters, and since many of their values cannot be determined by direct measurement (see section 4.3) they are highly uncertain.

An interesting aspect of the application of sensitivity analysis to model-based experimental design can be illustrated with table 6.12. The values of relative sensitivity for k_{16} (relative baroreceptor contribution, α) for the equilibrium conditions and Valsalva manoeuvre are $k_{16} = 0.0061$ and $k_{16} = 0.243$ respectively. This indicates that the results of a Valsalva manoeuvre are much more sensitive to the value of k_{16}, and therefore that it is a good test to determine the validity of the value used for k_{16} in the neural control submodel; indeed, in section 6.5.7.1.2.2 it was used for precisely this purpose.

6.5.8 Level 4 validation.

6.5.8.1. The complete Pullen Model - controlled haemodynamics with pharmacodynamics.

6.5.8.1.1 Theoretical criteria.

There are several major sources of uncertainty within the model. These are: the structural and parametric uncertainty of the neural controller, both theoretically (level 2) and empirically (level 3, particularly in relation to systemic resistance control), the theoretical uncertainties of the drug submodels, especially the disposal and local effects submodels (level 1); and the inability to validate empirically any of the drug submodels prior to this level. Furthermore, it is known that overall drug effects on the cardiovascular system are complex, often interacting with neural control in a non-cooperative manner. Therefore there are many a priori constraints limiting the theoretical validity at level 4 Nevertheless it may be possible that the overall response of the complete model is in agreement with available empirical data.

6.5.8.1.2 Empirical criteria

This section is subdivided into four subsections: the first is a resume of Pullen's results; the second concerns problems of data; the third is a detailed validity analysis of the model's pharmacodynamic responses; the fourth assesses the empirical validity of the distribution and disposal submodels.

6.5.8.1.2.1 Resume of Pullen's results

Pullen (1976) simulated the effects of three sympathomimetic drugs as intravenous injections into the head and arms or leg veins with the appropriate local drug effects "switched on" and values for them, (Pullen, 1976, pp 145 - 151) and gives graphs of the model responses of arterial pressure, stroke volume, cardiac output, heart rate and estimated total systemic resistance for each drug injection. The model is validated by comparing the qualitative directions of change in the model with those reported in the literature (table 6.14). It can be seen that in the case of noradrenaline the inclusion of venoconstriction apparently produces the reduction of heart rate observed in the human, "The results obtained with noradrenaline suggested that venoconstriction may play an

important part in producing the bradycardia commonly observed in practice" (Pullen, 1976, p.150).

Examining table 6.14, there is no other conclusion than that the model is 100% valid in qualitative empirical terms, and, since the drugs have a wide range of cardiac and vasoactive effects, that this is a general validity. However, there are other checks which must be performed prior to the interpretation of this table. For instance, the model responses should be checked for unidirectionality and stability, and the nature of the data carefully considered.

6.5.8.1.2.2 Problems of data

The problems of pharmacodynamic data were discussed generally in section 6.4.3. Typical data available in the literature are in the form of descriptive accounts based on physiological understanding of the local drug effects as well as the overall (experimental) responses. Thus the data are highly "theory-laden" (see reference to Hanson in Chapter 3). As an example, consider the following:
"The action of noradrenaline is rather different [from adrenaline]. The subject pales but feels no palpitation (i.e. pulse pressure constant). Both systolic and diastolic pressures are raised but, since the cardiac output is decreased, noradrenaline must constrict the peripheral vessels strongly. It is interesting to note that the heart usually beats more slowly (bradycardia); and this is due to the large rise in arterial pressure and strong stimulation of the baroreceptors in the aortic arch and carotid sinuses. Reflex inhibition (of the vagus) swamps the rather weak excitatory effect of noradrenaline on the pacemaker [in the heart]" (Lippold and Winton, 1979, p.233). Although descriptive, there is a great deal of information in this account. Unfortunately no ideas of the time scale or quantitative effects are conveyed.

Hawker (1979) uses a shorthand symbolic notation to illustrate the effects of drugs which is extremely compact, and this is used below. Quantitative data are hard to obtain and often more than one drug has been administered (e.g. in cardiac surgery), although a set of dynamic responses for phenylephrine has been found (fig. 6.2.3) which resolves a critical problem concerning the model (see next section).

DRUG	HUMAN OR MODEL	MEAN ARTERIAL PRESSURE	CARDIAC OUTPUT	HEART RATE	TOTAL PERIPHERAL RESISTANCE
METHOX-AMINE	HUMAN	+	0/−	−	+
	MODEL	+	−	−	+
ISOPREN-ALINE	HUMAN	−	+	+	−
	MODEL	−	+	+	−
NORADREN-ALINE	HUMAN	+	0/−	−	+
	MODEL 1	+	−	+	+
	MODEL 2	+	−	−	+

MODEL 1 — WITHOUT VENOCONSTRICTION

MODEL 2 — WITH VENOCONSTRICTION

TABLE 6.14. GENERAL FEATURES OF THE RESPONSES OF THE MODEL AND HUMAN TO INTRAVENOUS INJECTIONS OF METHOXAMINE, ISOPRENALINE, AND NORADRENALINE (+ ≡ INCREASE, 0 ≡ NO CHANGE, − ≡ DECREASE, BASED ON PULLEN, 1976, p. 152).

6.5.8.1.2.3. A detailed validity analysis of the pharmacodynamic responses.

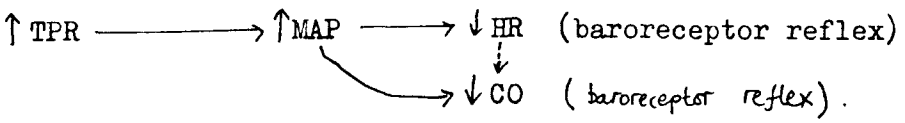
The analysis is divided into a simple case,(i). vasoactive drugs with no cardiac effect - and a complex case,(ii). drugs with vasoactive and cardiac effect. The latter case is divided into cooperative and non-cooperative interaction.

(i) Simple case - vasoactive drug with no cardiac effect

These drugs constrict or dilate the blood vessels producing an increase or decrease of peripheral resistance.

Vasoconstrictive

The α- adrenergic stimulants such as methoxamine or phenylephrine produce a generalised vasoconstriction. The effects are shown below (TPR - total peripheral resistance, MAP - mean arterial pressure, HR - heart rate, CO - cardiac output):



The model response for these variables is shown in fig. 6.22. The rapid and substantial rise of TPR (43%) is accompanied by an increase of arterial pressure (9%) and a reduction of cardiac output (23%) which is partly the neural reflex. Heart rate shows a strange behaviour - it drops by two bpm for 40sec and then rises above normal (+3bpm) until it gradually returns to normal. This is the response classified qualitatively as a decrease in table 6.14, and clearly it is not. The heart rate response should simply be a control reflex to increased arterial pressure, i.e. a unidirectional decrease. Fig. 6.23 shows the effect of differing levels of phenylephrine (intravenously injected) in the human on arterial pressure and heart rate (taken from Mancia et al., 1978). The heart rate reflex is a unidirectional decrease whose maximum charge occurs at about 20 - 30 sec after injection.

The bidirectional model heart rate response throws doubt on the validity of the neural control submodel for heart rate for raised arterial pressure. In the empirical validation at level 3, the heart rate response was satisfactory, but this was for decreasing arterial pressure (postural change and Valsalva manoeuvre).

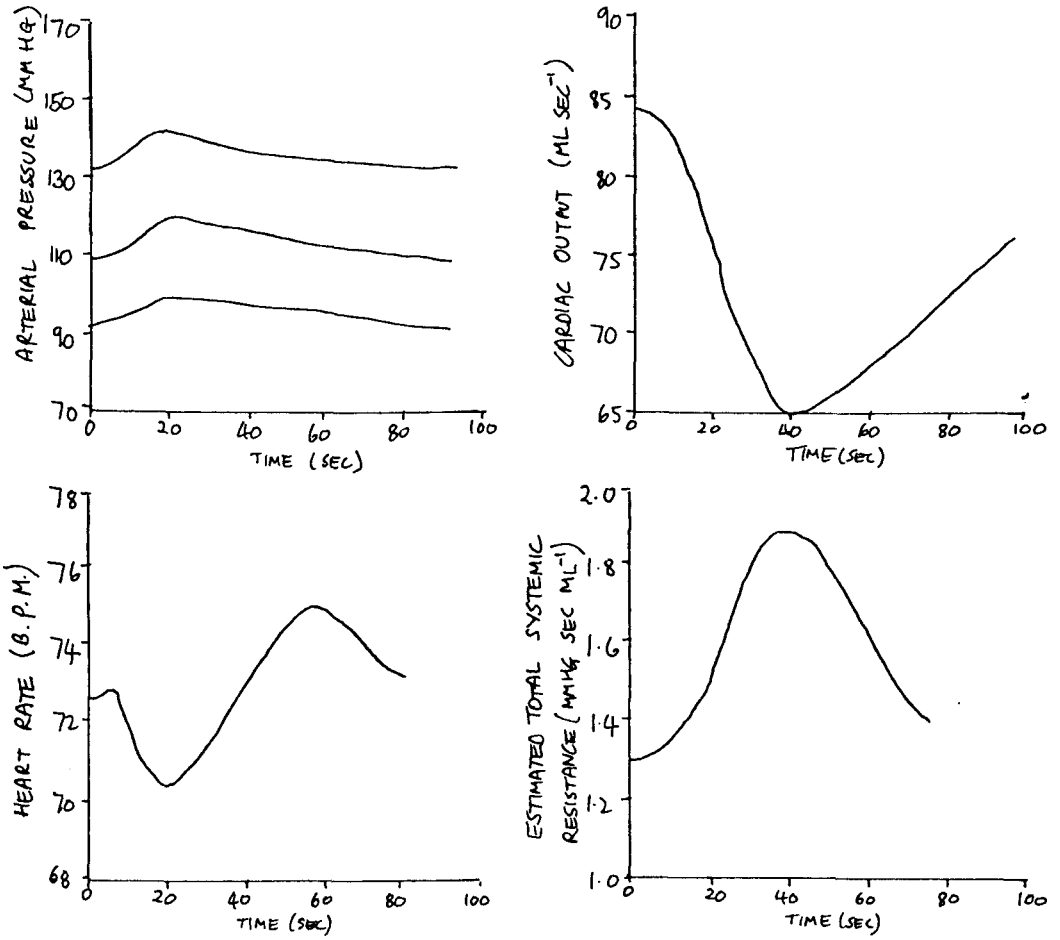


FIGURE 6.22. INJECTION OF PHENYLEPHRINE INTO THE HEAD AND ARMS VEINS SEGMENT AT T=0SEC.

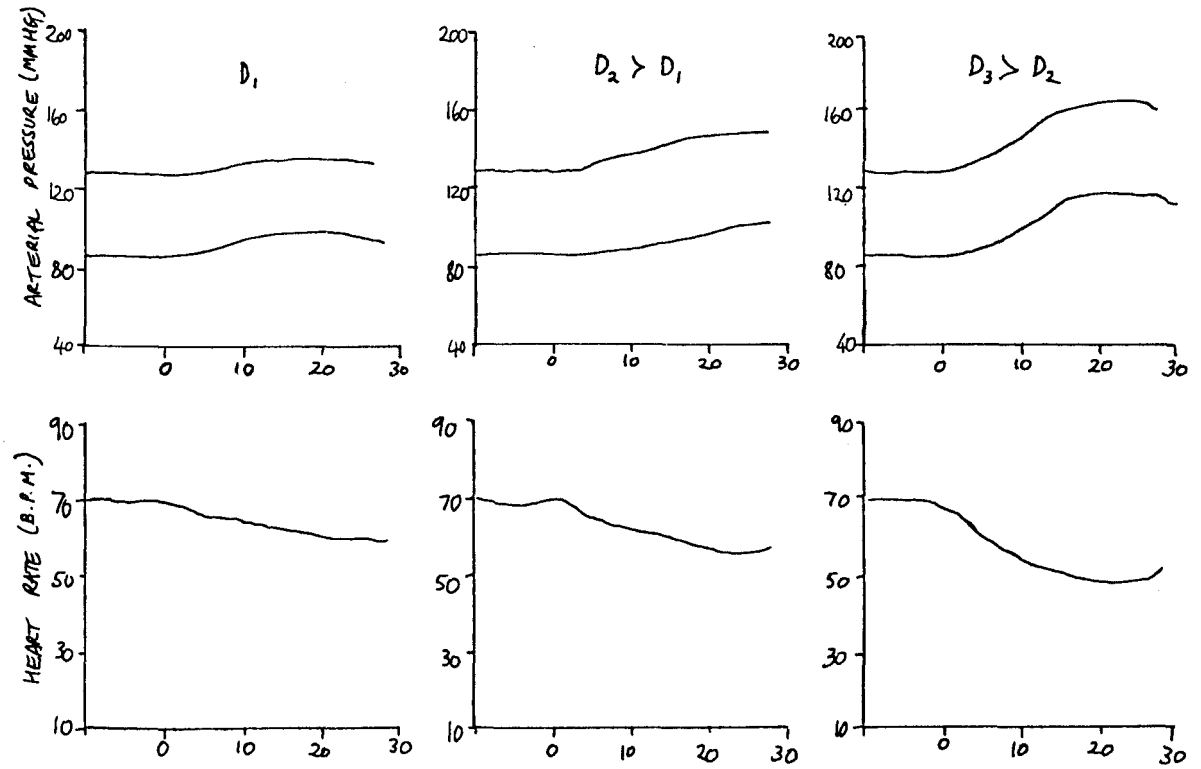


FIGURE 6.23. EFFECT OF INJECTIONS (OF INCREASING DOSES, D_1, D_2, D_3) OF PHENYLEPHRINE ON ARTERIAL PRESSURE AND HEART RATE IN MAN (MANCIA ETAL., 1978, p.238).

Vasodilatative

Given the satisfactory conclusions regarding the validity of the neural controller for heart rate ($\downarrow$ AP), (section 6.5.7.1.2.2), the model response to a vasodilatative drug should be acceptable. A drug such as amyl nitrate acts as a depressor on the sympathetic nervous system leading to a reduction of vascular tone:

vasodilatation $\longrightarrow \downarrow$TPR $\longrightarrow \downarrow$MAP $\longrightarrow \uparrow$HR (reflex)

$\longrightarrow \uparrow$CO (")

The response of the model is obtained by switching vasodilatation "on" in the model and is illustrated in fig. 6.24. The results agree closely with reported accounts (e.g. Burgen and Mitchell, 1978, pp.126-127): systemic resistance drops greatly (41%) and is reflected in the fall of arterial pressure (22%), which also exhibits a slight increase of pulse pressure (5%) ("arterial pulsation becomes marked", p.126). Cardiac output rises by 20%. The heart rate response to the lowered pressure is unidirectional and there is a significant tachycardia (14%) in accordance with the data.

Thus the model produces a valid response in the case of dilatative vasoactive drugs, when the arterial pressure falls and there is a reflex tachycardia. This also supports the conclusions of the level 3 validation. At that level it was found that the reflex vasoconstriction under this condition is insufficient, however in the case of a drug the strong direct effect of the drug on the vasculature can be considered to override the weaker neural reflex.

(ii) Complex case

Drugs that affect both the heart and blood vessels.

Cooperative interaction

Isoprenaline has its actions almost entirely on the β -receptors. There is therefore marked vasodilatation (in muscle) and strong positive inotropic and chronotropic actions on the heart (Burgen and Mitchell, 1978 p. 124):

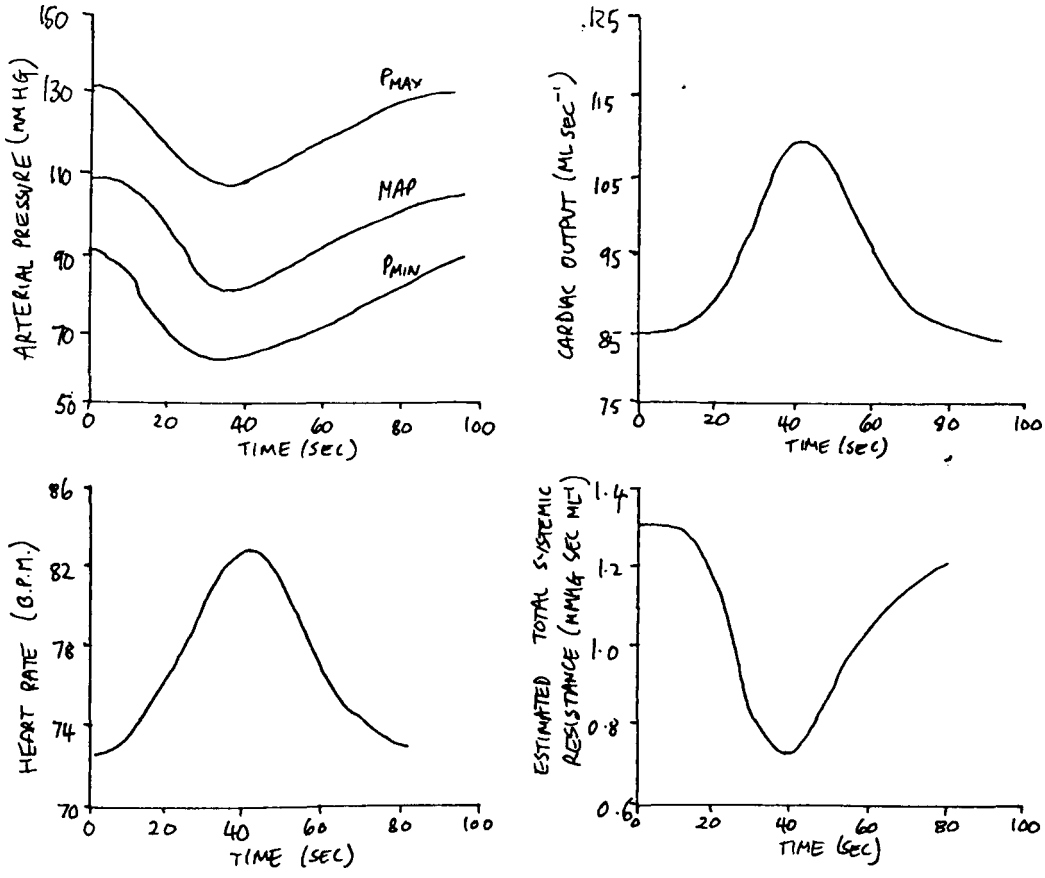


FIGURE 6.24. INJECTION OF A VASODILATATIVE DRUG (E.G. AMYL NITRATE) INTO THE HEAD AND ARMS VEINS SEGMENT OF THE MODEL AT T = 0 SEC.

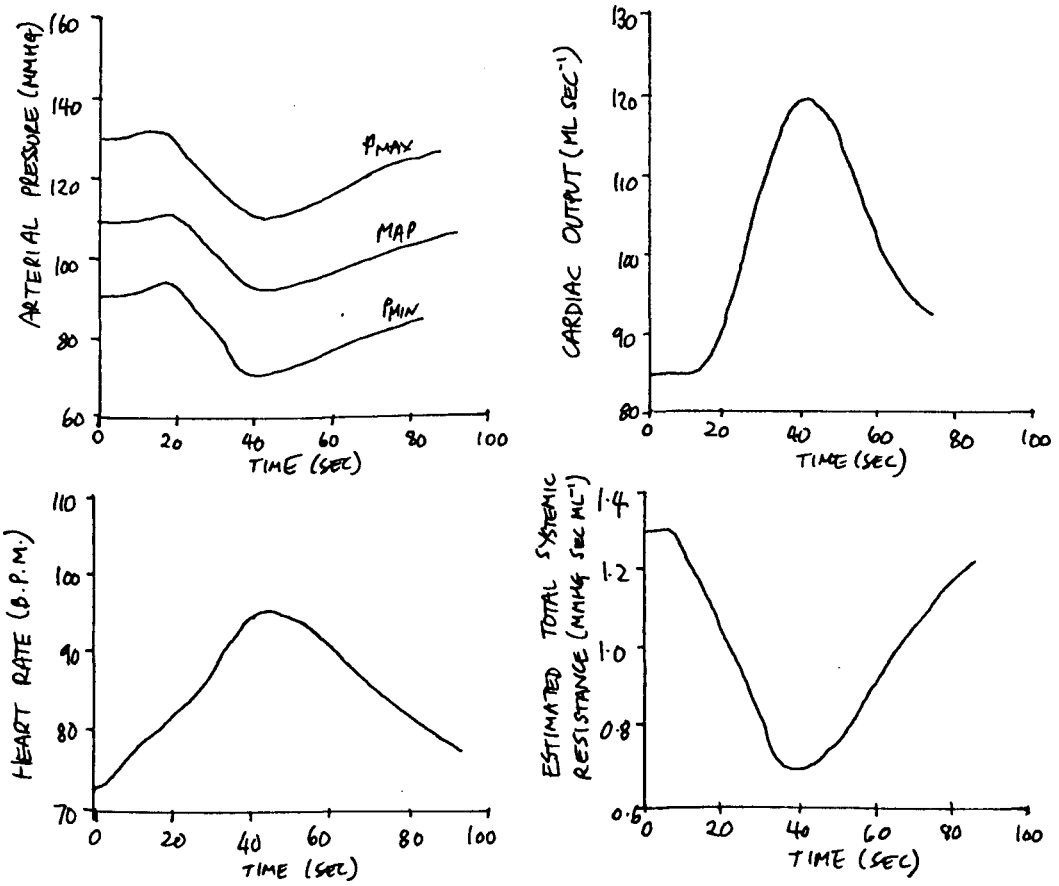
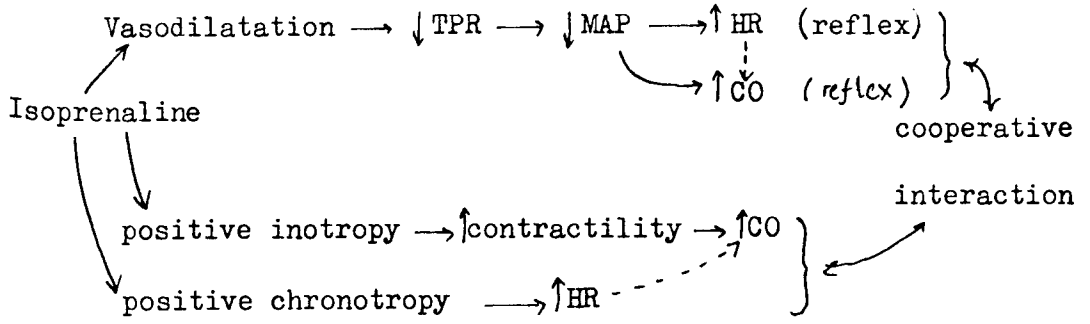


FIGURE 6.25. INJECTION OF ISOPRENALINE INTO THE HEAD AND ARMS VEINS SEGMENT OF THE MODEL AT T = 0 SEC.

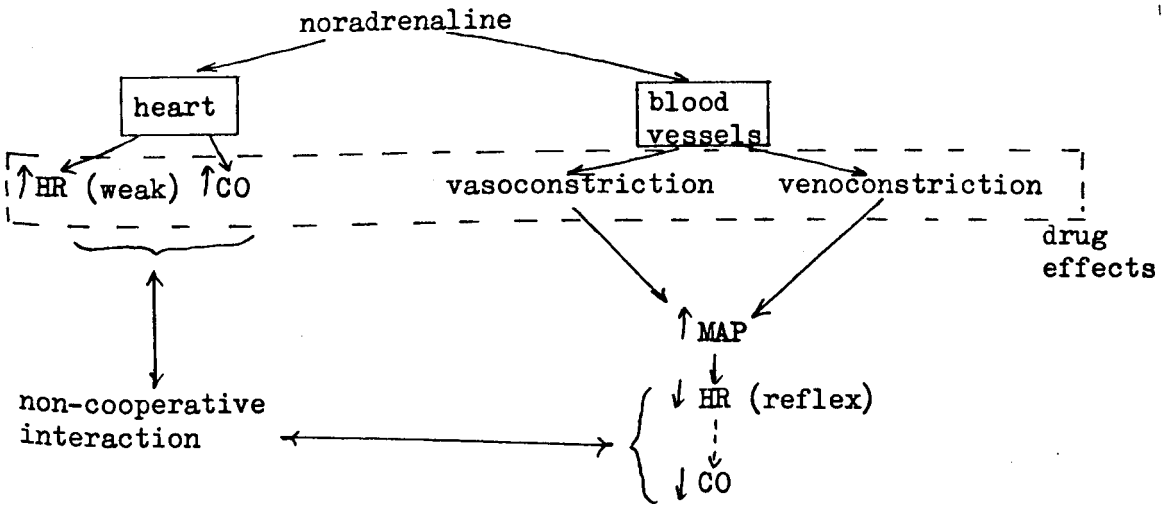


The action of this drug will be described as "cooperative interaction", since the reflex changes on the heart as a consequence of the drug effect on peripheral resistance are in the same direction (and augmenting) the direct cardiac effects of the drug. In the model, isoprenaline is simulated by switching "on" vasodilatation, tachycardia, and positive inotropy (Pullen, 1976, p. 148). The results are shown in fig. 6.25.

The model exhibits the reduction of systemic resistance (drug induced) and the consequent drop in arterial pressure. Heart rate and cardiac output increase significantly indicating the cooperative reflex and drug effect. The model response for this drug is therefore satisfactory.

Non-Cooperative interaction - the enigma of noradrenaline.

"Noradrenaline is a catecholamine acting predominantly on the sympathetic α-receptors and has in most species a relatively weak action. It produces a considerable increase of blood pressure and slowing of the heart with cutaneous vasoconstriction", (Burgen and Mitchell 1978, p.124). The effects are summed up in the following diagram.



The reflex slowing of the heart tends to counteract the direct effect of noradrenaline in increasing stroke volume, so that overall the cardiac output may not be much changed. Since the reflex neural changes on the heart as a consequence of the effect of the drug on the blood vessels are in the opposite direction to the direct cardiac effects of the drug, the action of this drug will be known as "non-cooperative interaction".

The action of noradrenaline is simulated firstly in the model by switching "on" vasoconstriction, tachycardia, and positive inotropy and setting the σ - values to $\sigma_1 = 400$, $\sigma_2 = 50$, $\sigma_3 = 50$ ($\sigma_1 \equiv$ vasoconstriction, $\sigma_2 \equiv$ tachycardia, $\sigma_3 \equiv$ positive inotropy), (see section 6.3.4, and Pullen, 1976, p. 149). The results are shown in fig. 6.26. The model reproduces the vasoconstriction and the arterial pressure rises as in the human. Cardiac output drops despite the positive inotropic drug effect. Heart rate rises greatly (17%), whereas in the human it falls (see section 6.5.8.1.2.2, and Burgen and Mitchell, 1978 p. 124). The rise in heart rate is presumably due to the dominance of the drug's positive chronotropic effect in the model.

Pullen reports that, "it is found that adjustments of the constants σ_1, σ_2, and σ_3 cannot reproduce the required bradycardia response", (p. 150), yet, as table 6.14 shows, the response of the model for methoxamine is precisely the desired one for noradrenaline. This has the σ - values, $\sigma_1 = 400$, $\sigma_2 = \sigma_3 = 0$, i.e. a mainly vasoactive drug. However, despite this, Pullen then includes venoconstriction (denoted by $\sigma_4 = 10$) as an effect of the drug. The modified results of the model are shown in fig. 6.27.

The systemic resistance rises to a greater value than before, because of the venoconstriction, and arterial pressure rises rapidly as expected. Heart rate is high for a couple of beats (initial drug effect) and then drops significantly as a reflex to the high arterial pressure as in the human. Notice, however, that in this simulation, Pullen has reduced the drug-induced tachycardia parameter σ_2 from 50 to 10. After 20 seconds the heart rate increases sharply and exceeds the normal value. (This effect is not reported in the available data concerning noradrenaline, but it does indicate the possibility that although sampled measurements indicate a unidirectional change the real system may have more complex behaviour, particularly if early measurements

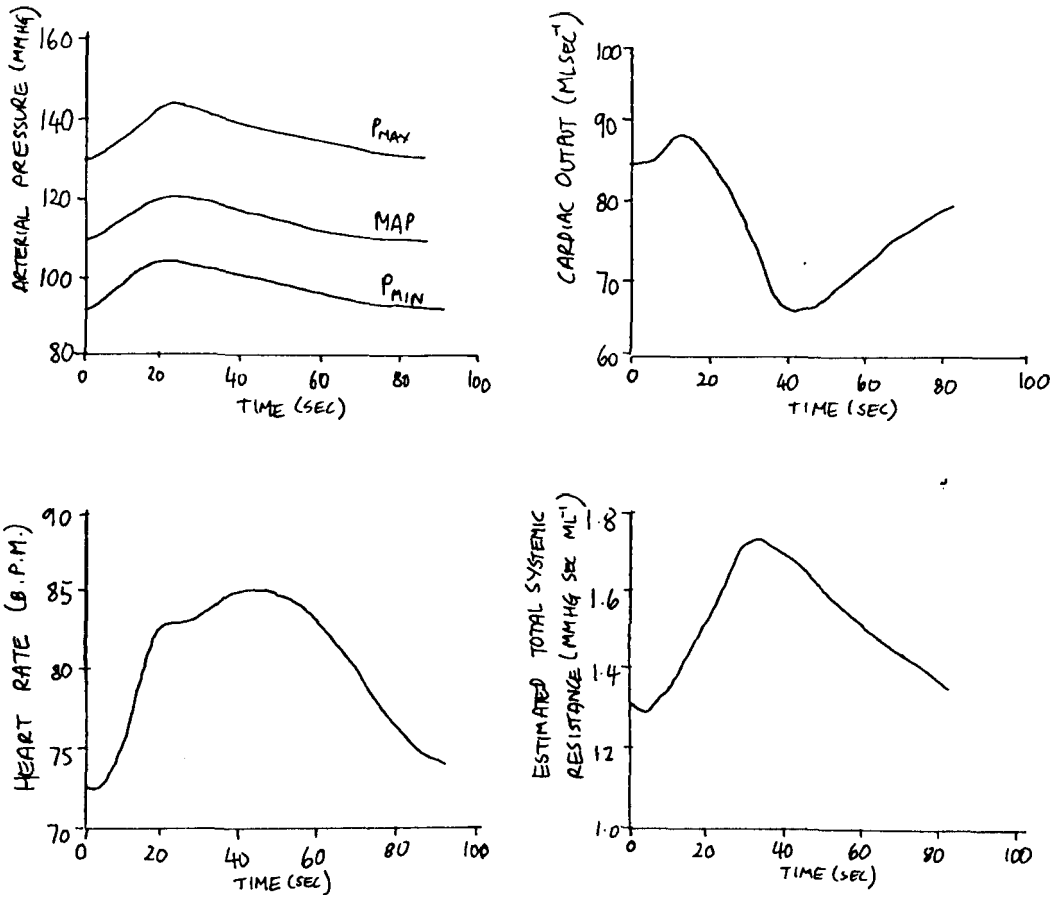


FIGURE 6.26. INJECTION OF NORADRENALINE INTO HEAD AND ARMS VEINS SEGMENT OF THE MODEL AT T= 0 SEC ($\sigma_1 = 400, \sigma_2 = 50, \sigma_3 = 50, \sigma_4 = 0$).

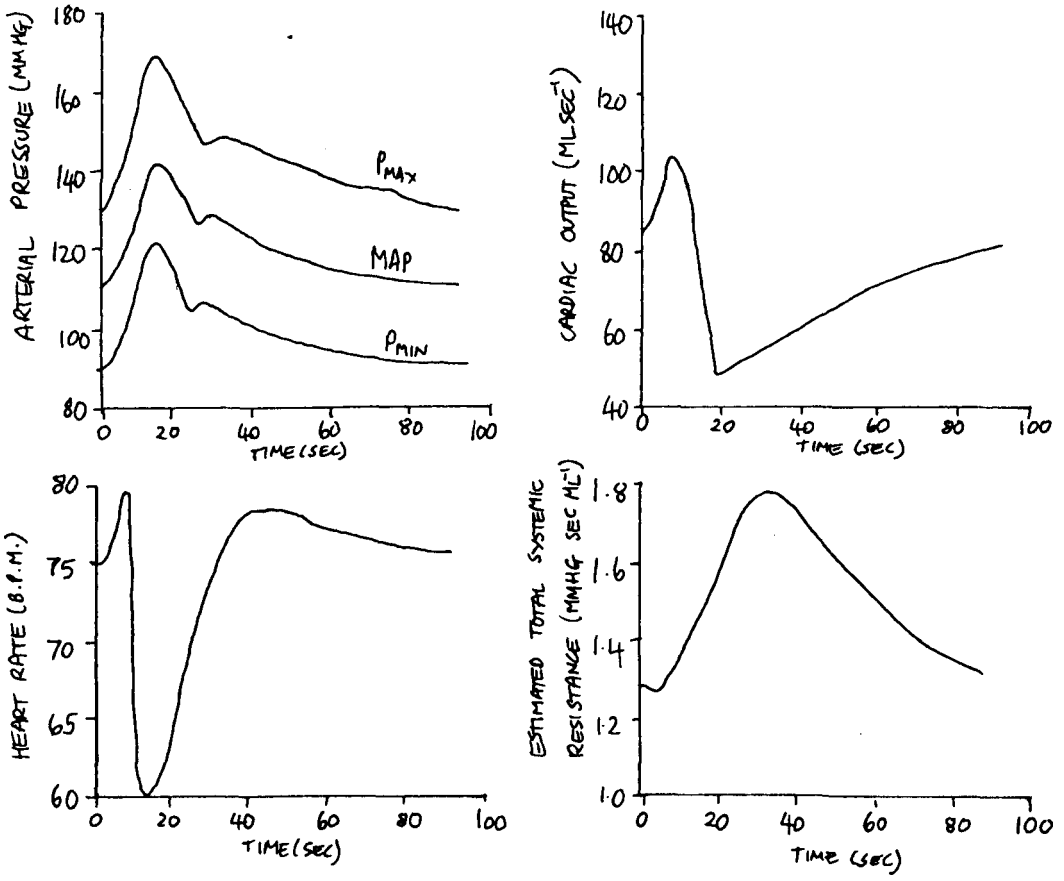


FIGURE 6.27. EFFECT OF NORADRENALINE IN THE MODEL WITH VENOCONSTRICTION ($\sigma_1 = 400, \sigma_2 = 10, \sigma_3 = 50, \sigma_4 = 10$).

($t < 20$sec) are not available. This must either be due to the drug effect or a defect in the neural control submodel. Since the tachycardic effect has been reduced (i.e. σ_2), it is the control which must be at fault, as was found with methoxamine, a purely vasoconstrictive drug. The venoconstriction included by Pullen is a well-known effect of noradrenaline. It is known because noradrenaline occurs naturally in the human and has a role as a hormonal controller of the cardiovascular system, (e.g. Hawker, 1979, p. 62).

The results of this subsection are summarised in table 6.15. In this table the qualitative responses are represented by small diagrams in order to convey most information. It is easy from table 6.15 to delimit the range of validity: for a drug-induced fall in arterial pressure, the model is qualitatively valid in its major variables; for a drug-induced rise in arterial pressure, the model has a qualitatively invalid heart rate response, and this is reflected elsewhere in the model.

6.5.8.1.2.4. Empirical validation of the distribution and disposal submodels

The problems of empirical validation of the drug distribution and disposal submodels were discussed at level 1. (section 6.5.5). Some inferences can be made on the validity of these two submodels from the overall pharmacodynamic responses. Fig. 6.22 shows the dynamics of the model response to an injection of phenylephrine, with the peak in arterial pressure occurring at $\simeq 20$ seconds. A similar dynamic is exhibited in the data of Mancia et al. (1978) in fig. 6.23, indicating that the drug is transported to the significant sites and is disposed in the correct times. To gain greater confidence in the empirical validity of the distribution and disposal effects, a series of tests using neutral dyes should be performed if such data become available.

DRUG		MEAN ARTERIAL PRESSURE	CARDIAC OUTPUT	HEART RATE	TOTAL PERIPHERAL RESISTANCE	COMMENTS	
METHOXAMINE OR PHENYLEPHRINE	DATA						SIMPLE CASE
	MODEL					VASOCONSTRICTION, UNACCEPTABLE HR RESPONSE	
AMYL NITRATE	DATA						
	MODEL					VASODILATATION ACCEPTABLE RESPONSE	
ISOPRENALINE	DATA						COOPERATIVE INTERACTION
	MODEL					VASODILATATION, CHRONOTROPY, AND INOTROPY. ACCEPTABLE RESPONSE	
NORADRENALINE	DATA						COMPLEX CASE
	MODEL 1					VASOCONSTRICTION, NO VENOCONS., LARGE DRUG TACHYCARDIA, UNACCEPTABLE HR RESPONSE	NON-COOPERATIVE INTERACTION
	MODEL 2					VASOCONSTRICTION, VENOCONSTRICTION, REDUCED DRUG TACHYCARDIA, UNACCEPTABLE HR RESPONSE	
	MODEL 3					VASOCONSTRICTION, NO VENOCONSTRICTION NO DRUG TACHYCARDIA (AS PHENYL-EPHRINE) UNACCEPTABLE HR RESPONSE	

TABLE 6.15. AN ALTERNATIVE TABLE FOR THE QUALITATIVE VALIDATION OF THE PULLEN MODEL WITH PHARMACODYNAMICS.

6.6 Conclusions

6.6.1. Summary and conclusions of the results of the programme of validation.

In the validation study of section 6.5, it was clear that the validity of the model does not completely cover the intended range of application. As far as the physical modality of the range of application is concerned, the model is a very approximate aggregation. However, it was demonstrated that the level of approximation was more than detailed enough to reproduce effects pertinent to the functional modality.

In the functional modality, the aspects of neural control and drug effects have the least theoretical validity, since they are not based on previously validated physiological theory or models (to some extent the model is concerned with the development of such theory). The empirical tests on the model established that the basic haemodynamic behaviour of the heart and circulatory fluid mechanics (excluding neural and drug effects) is in adequate agreement with available data. (Exceptions are the ventricular pressures which rise too high in the model. This has been traced to an inaccuracy in the modelling of the valve ejection dynamics and produces little change in the rest of the model. In any case, the fault can easily be rectified).

The response of the model to increased blood volume, and hence arterial pressure, was very similar to reported accounts indicating that under these circumstances the neural control models were acting correctly together. However, when the arterial pressure was lowered the response of the model was unsatisfactory - although the heart rate reflex was correct, it was found that the slower vasoconstrictive reflex was insufficient and the model could not counteract the changes. The error was most noticeable after a Valsalva manoeuvre when the model's behaviour was strikingly similar to that of a human with impaired neural function. During the Valsalva manoeuvre, the heart rate response was unaffected by this error, and was used to test an hypothesis concerning relative contributions of the two baroreceptor sites to heart rate control. Contrary to the assumption of the model, it was found that with a greater contribution from the aortic arch site the heart rate response was closest to the data.

By contrast, in the empirical tests of the pharmacodynamic responses, drugs that tended to reduce arterial pressure produced qualitatively acceptable results, whereas drugs that induced increased arterial pressure gave incorrect model responses of heart rate.

The apparent contradiction of the unacceptable model response to haemodynamically-reduced arterial pressure and acceptable model response to drug-reduced arterial pressure has a simple explanation: the former results from an inadequate vasoconstrictive reflex (in the CNS control submodel of peripheral resistance) whereas, in the latter, any vasoconstrictive reflex (even if invalid) is masked by the stronger vasodilatative effect of the drug. However, the discrepancy between the results for increased arterial pressure is more difficult to explain. In the case of increased arterial pressure due to increased blood volume, all CNS cardiac and vascular control submodels are operative and, together, produce a response valid qualitatively and quantitatively for important features. With drug-induced arterial pressure increases the CNS vascular control submodel is always over-ridden by the direct vasoconstrictive effect of the drug (and sometimes on other CNS control submodels) and under these conditions the model heart rate response is qualitatively incorrect. A possible interpretation is that when all CNS control submodels are operative any defect in the heart rate control is not manifested. The empirically valid range of application of the model is shown, for arterial pressure changes, in table 6.16.

It can be seen from table 6.16 that there is a severe attenuation of the intended range of application of the model in functional terms. This suggests that the model contains defects in the CNS control submodels. Until these are rectified the complexity of the structure of the haemodynamics model cannot be justified, and the interpretation of the pharmacodynamic responses of the model remains problematic. Furthermore, the only validation of the dynamics of drug distribution and disposal is at a qualitative theoretical level, and there is therefore substantial uncertainty concerning this aspect of the model.

The conclusion of the programme of validation must be that the Pullen model fails to satisfy adequately its specific modelling objectives by failing appropriate tests of theoretical and empirical representational validity over the intended range of application. There is no doubt,

Aspect of Model	Disturbance (and Cause)	Empirical Validity	Comments
Controlled Haemodynamics	↑AP (Increased Blood Volume)	Qualitative Quantitative (Features of Key Variables)	Qualitative Changes Correct Quantitative Features (% Changes) of Key Variables in Normal Ranges
	Equilibrium Conditions	Qualitative Quantitative	Important Qualitative Events Reproduced Within Cardiac Cycle. Key Variables in Normal Ranges.
	↓AP (Haemorrhage or Upright Tilt)	Qualitative Quantitative Error	Qualitative Changes Correct. Quantitative Discrepancy with Normal Data (by Feature Comparison). (Insufficient control in Model Traced to Inadequate Vasoconstriction Invalidity of CNS Resistance Control Submodel)
Controlled Haemodynamics with Pharmaco-dynamics	↑AP (Drug Effects Include Vaso-constriction)	Qualitative Error	Qualitative Discrepancy on Heart Rate Response (Invalidity of CNS Heart Rate Control Submodel)
	↓AP (Drug Effects Include Vaso-dilatation	Qualitative	Key Model Variables Exhibit Same Qualitative Changes as Data

(n.b. "Key variables" include mean arterial pressure, pulse pressure, cardiac output, stroke volume, heart rate, and total peripheral resistance).

Table 6.16. Empirically Valid Range of Application ($\mathcal{R}_v$) of the Pullen Model for Arterial Pressure Disturbances

however, that the model contributes to the understanding of the short-term behaviour of the cardiovascular system and also that it has heuristic validity in other respects,(for instance as a lesson or warning of the appropriateness of complex mathematical models in biology), thus satisfying some of its more general objectives. However, care should be exercised in the heuristic use of the model for hypothesis testing or theory development until the above problems have been solved.

One of the great difficulties in formulating or validating mathematical models in biology is the lack of comprehensive quantitative dynamic data. This is certainly true of the human cardiovascular system yet the results of the programme of validation imply that the model can still be improved on the basis of the presently available data, particularly by modifying the CNS control submodels. If this is achieved then the next step in validation would require an extension of the available data (for instance, continuous records of pressure, drug concentrations, etc. from a number of sites). Perhaps a better programme is to develop simpler models whose data type requirements are not so far removed from those available. This is supported by the theoretical considerations of the validity of the model in representing the structural modality of the human cardiovascular system, which suggest that a fairly simple structure is an adequate frame of reference for modelling at the level of control and overall dynamic behaviour. An attractive feature of simpler models is that they can satisfy, more directly, utilitarian modelling objectives, such as the improvement of diagnosis and therapy in a health-care system.

6.6.2. General conclusions.

The programme of model validation, which is based on the theory of model validity presented in Chapter 4, has provided an extensive and very critical analysis of the validity of the Pullen model. The method of applying both theoretical and empirical criteria starting at the level of elementary submodels and gradually building up to the overall model clearly exposes the areas of validity and uncertainty in the model and allows a precise delimitation of the valid range of application. The critical conclusions of the programme of validation are nevertheless constructive in that they determine what areas of the model require modification (and occasionally explain how), and also they suggest new

specific objectives or directions for research which may be more fruitful. In the next chapter, a validation study is presented of a mathematical model of the human renal - artificial kidney machine system. This model is ultimately intended for practical use in a health - care system and, in satisfying most of its specific modelling objectives, it proves to be more valid than the Pullen model.

CHAPTER 7

SECOND CASE STUDY - THE VALIDATION OF A MATHEMATICAL MODEL OF THE HUMAN RENAL-ARTIFICIAL KIDNEY SYSTEM

7.1 Introduction

The subject of the second validation study is a mathematical model of the human renal-artificial kidney system developed by Uttamsingh (1977, 1981) in the Department of Systems Science. The model is intended primarily to help in the health-care of patients with kidney failure undergoing haemodialysis (on an artificial kidney machine) by providing predictions of the clinical state of a patient during and after periods of dialysis. It is a dynamic representation of the function of the kidneys in man, in both normal and disease states, and their role in the excretion of waste products and the overall bodily control of fluid and electrolyte levels, as well as the effect of haemodialysis.

This case study contrasts with the preceding one in that the model under consideration has utilitarian objectives (i.e. clinical application, or the improvement of a health-care system) as opposed to solely scientific objectives (the understanding of fast cardiovascular dynamics). However, the requirement of the utilitarian objectives is for empirical (predictive) validity. This suggests that an appropriate validation methodology is the δ-methodology (for utilitarian modelling objectives), described in Chapter 5, with an emphasis on the testing of the representational validity of the model. The framework of the programme of validation for the cardiovascular model in Chapter 6, which is effectively a γ-methodology (a theoretical/empirical validation methodology, Chapter 5), is therefore also an appropriate one for the renal model. If the model fails some tests of representational validity, but still retains its predictive validity, it will still satisfy its utilitarian objectives and be considered valid for this purpose. (For models with utilitarian objectives, both an intended range of application $\mathcal{R}_I$ as well as the wider system of interest SOI are specified. However, the representational validity criteria over $\mathcal{R}_I$ may be relaxed if the model obviously meets its utilitarian objectives over SOI.)

The validation study in this chapter is not complete,but is intended to illustrate the nature and problems of validation of the renal model as

an example of a model with largely utilitarian objectives (i.e. clinical application). In particular, attention is given to the degree of representational validity which is adequate for the successful application of the model. A full validation programme would be structured along the lines of the validation of the cardiovascular model in Chapter 6 with an additional emphasis on the application or use of the model. In the following section a very brief outline of renal physiology from a systems point of view is given which provides a simple, yet necessary, frame of reference for non-physiologists.

7.1.1 Brief outline of renal physiology

The kidneys may be regarded as a pair of multivariable sensors and controllers which act in unison in controlling a large number of bodily systems by a process of selective excretion of water and other substances from the body (in the urine) and by the secretion of chemical controllers (hormones) into the bloodstream. They are located in the back of the abdomen and each supplied with a large artery. The blood flow rate through the kidneys is fairly constant and very large (about one-quarter of the blood circulation) which means that the entire blood volume can be treated by the kidneys in a short time. Three kinds of kidney function may be distinguished: (a) pure excretion; (b) control by selective excretion; and (c) hormonal control secretion. These are considered separately below.

The kidneys are usually identified with their function of pure excretion, (a), in the removal of waste products from the blood. Urea, an end product of metabolism, and creatinine (associated with muscular activity) are the main waste products excreted via the kidneys, but other substances (possibly toxic) are also removed. A greater proportion of the wastes than fluid is excreted by the kidneys, and so the concentration of urea, say, in the urine is about 60 times that in the blood.

The selective excretion of water, salts, and phosphates by the kidneys results in the control of a number of bodily systems: body fluid, blood pressure, electrolytic balance, and blood acidity (pH). The ability to do this is linked to the structure of the kidneys, which consist of millions of tiny tubules (nephrons) across whose walls water, salts and other substances may pass from the blood and be wholly or partially reabsorbed back into the kidney blood which then returns to the general circulation. The reabsorption is also controlled by hormones, such as ADH which

increases the reabsorption of water from the nephrons if the fluid level in the body falls and thereby decreases fluid loss in the urine. The control of the electrolytic balance (the proportion of sodium and potassium salts) by the kidneys is essential for the correct functioning of cells, the basic units fundamental to life.

The kidneys are also sensitive to various variables and secrete hormones into the blood stream in order to control them. The major hormonal control system is the renin-angiotensin—aldosterone system which monitors and controls blood pressure, sodium reabsorption, and potassium excretion. The kidneys also secrete erythropoietin, when the demand for oxygen increases, which has the effect of stimulating red blood cell development (the carriers of oxygen in the blood).

The failure of the kidneys is complex, but in general it results in a diminished ability to excrete waste products and effectively control numerous bodily systems. Fortunately, the kidneys have a large reserve capacity and are capable of adapting to partial failure (one kidney is perfectly adequate for life). If kidney function is nearly completely impaired the levels of toxic substances will rise and the control of fluidic, electrolytic, and other systems will become unstable, finally resulting in death unless action is taken. This may involve the filtering of blood external to the body in an artificial kidney machine (haemodialysis) to remove wastes and rebalance the blood, or, optimally, the transplantation of a functioning kidney.

Renal physiology involves the study of many interacting levels of control and its full treatment is a complicated subject, made slightly easier, however, by viewing it from a control or systems approach. (For an introductory account, consult Lippold and Winton, 1979, pp. 130-148; or, for a detailed explanation, Part 5 of Guyton, 1971.)

7.2 Background and Outline of the Uttamsingh Renal Model

7.2.1 Introduction and modelling objectives

The general objective of the Uttamsingh model is utilitarian - the improvement of a health care system. Specifically, the model is intended to aid in the treatment of patients with renal failure who are undergoing dialysis on an artificial kidney machine. The specific utilitarian objective is mapped into a set of specific scientific objectives - the

representation of the human/artificial kidney machine system and the prediction of the clinical state of a patient on or off dialysis. The predictions of the model, if correct, then allow the design of optimal or improved therapies for patients undergoing dialysis and therefore satisfy the specific utilitarian objective. A typical aim for the design of dialysis therapies would be the maximisation of periods between dialyses and the minimisation of time spent on the dialysis machine, as well as the maintenance of the patient's feeling of "well-being".

In order to accomplish the specific scientific objectives, the model includes the major factors that determine the overall clinical state of a patient with renal failure who may be undergoing dialysis: fluid levels and arterial pressure; concentrations of electrolytes, hormones,and metabolites in the blood; thermoregulation; kidney function; and the effect of haemodialysis. Some earlier models (see Section 7.2.2) were based solely on the prediction of levels of metabolites (e.g. urea and creatinine) and electrolytes in the blood, and were simple uncontrolled dynamic compartmental models. However, the Uttamsingh model is a representation of the overall role of the kidneys in both the excretion of substances and the control of, and interaction with, various subsystems. Thus the validity of the model predictions depends not simply on the comparison of the model predictions with overall response data,but also on the degree to which the model is a correct representation of the dynamics and control mechanisms of the human renal and associated subsystems. In this sense, the model can meet wider requirements than its specific utilitarian objectives. For instance, the model may reveal unwanted interactions between subsystems under certain conditions, and preventive action can be taken to avoid these in the treatment of patients (e.g. the onset of nausea, feelings of discomfort, or vomiting). More importantly, in this form, the model is essentially an explanatory model and may therefore be used for general scientific objectives. For instance, the model may be employed as a "test bed" for hypothesis testing.

Uttamsingh (1981) summarises the three objectives of his model as follows:

"1. Prediction of the state of the patient during and after dialysis;

2. Examination of the structure of the model and data generated in order to search for the causes of the unexplained phenomena that are occasionally observed during dialysis;

3. Utilization of the model to test various hypotheses concerning the renal-body fluid system."

(The implications of the modelling objectives for model validity and validation are discussed further in Section 7.3.1.)

The next section (Section 7.2.2) describes some previous models of the human renal system which have a range of different objectives. The subsequent section (Section 7.2.3) outlines the Uttamsingh model. This is presented at a purely verbal, or conceptual, level and the full details of the mathematical model may be found in Uttamsingh (Chapter 4, 1980). (An account of the model in an early stage of development and some preliminary validation tests are contained in Uttamsingh (1977). Contrasting this with the 1981 thesis provides an interesting study in the dynamics of model development and indicates the intimate role that validation plays in the ongoing process of model development.) A full listing of the mathematical model is provided in Appendix III.

7.2.2 Previous models

Earlier models divide very clearly into two types: explanatory models (for scientific objectives) and predictive models for clinical application. In general, the former are more complex than the latter. Explanatory models may focus at different levels from detailed kidney function to overall control processes. The model of Guyton et al. (1972) on the overall regulation of the circulation includes an empirical representation of kidney function as one of many interacting control mechanisms. A similar model directed at overall kidney function is that of Cameron (1977). The model of Cage et al. (1977), on the other hand, is concerned with the action within the nephrons of the kidney's medulla.

There are a number of simpler models which have been developed specifically for use in clinical application to dialysis. For instance, the model by Walker et al. (1975) is a simple representation of the dynamics of urea and creatinine which can be used for prediction of an individual patient both on and off dialysis. The model is simple enough to implement on a programmable calculator. The model by Lott et al. (1977) is intended for the investigation of different generalized strategies for dialysis.

In most of these models, the techniques of compartmental analysis are used. A compartment or "pool" corresponds to the level of a particular

substance at various sites in the body (e.g. intracellular and extracellular compartments for urea, electrolytes, etc.). The shift of a substance between compartments (e.g. by diffusion, chemical change, decay, etc.) is represented by linear or nonlinear functions. The modelling of many of the renal subsystems is made difficult by the fact that there is a complex interaction between these functions and the levels (or concentrations) of the substances themselves (such as electrolytes or hormones). Consequently, the models of urea and creatinine, which are physiologically inert, are the simplest, and have therefore been used more in clinical applications.

7.2.3 The Uttamsingh renal model

Although the Uttamsingh model is intended primarily for clinical application, its form is more closely related to the explanatory models above (such as Guyton's or Cameron's) than to the simpler, clinically-oriented models. This is a consequence of the intention (objective) to be able to predict the overall state of a patient during and after dialysis, and not simply the level of a few variables (such as plasma urea concentration) during dialysis. In formulating the model, a set of variables was chosen which indicates the clinical state of a human in relation to the functioning of the renal system. This set includes: arterial pressure; body temperature; concentrations of electrolytes, hormones and wastes in the blood; and urine flow rate and composition. Other variables are also involved which interrelate this set. The choice of variables indicates the subsystems that will have to be included in the final model.

The model may be considered to be composed of eight submodels: thermoregulatory model; cardiovascular model; kidney function model; fluid and electrolytic balance model; urea and creatinine dynamics model; hormonal dynamics model; kidney failure model; and a model of the effect of dialysis on an artificial kidney machine. These are considered separately below. The overall integration of the model and the relationship between the various submodels are discussed in Section 7.2.3.9. (A full listing of the mathematical model is given in Appendix III.)

7.2.3.1 Thermoregulatory system submodel

The thermoregulatory system submodel consists of a model of the passive thermal system and a model of the neural regulation of temperature (Figure 7.1). The passive system is represented as two compartments, one for the core and one for the skin (Figure 7.1(b)). Two first order differential equations for skin and core temperatures describe the exchange of heat between the core and skin, and include terms for the metabolic generation of heat in the core, the conduction of heat in the blood flow from the core to the skin (as well as by thermal diffusion), heat loss to the environment, and insensible heat loss. The model was derived from a 7-compartment model of thermoregulation by Hardy and Stolwijk (1966).

The submodel of the neural regulation of temperature is based on a set-point theory for the control of deep body (core) temperature (Hardy and Stolwijk, 1966). In the body there are temperature sensors located in the core and skin which send information to the hypothalamus, which, in turn, modifies the activity of the vasomotor centre in the medulla. This controls the peripheral resistance of arterioles in the skin, and hence the heat loss from core to skin through the skin blood flow, thereby counteracting any changes in core or skin temperature (Figure 7.1(a)). In the model there is a piecewise linear approximation for the relationship between skin total peripheral resistance and core temperature. If the core temperature is in the normal range, the resistance is further modulated by the skin temperature. (In the human, if the core temperature exceeds, or is less than, critical levels for life, shivering or sweating of the skin may also occur, but these mechanisms are not included in the model.)

7.2.3.2 Cardiovascular system submodel

The cardiovascular submodel is based largely on the circulation model of Guyton, Coleman and Granger (1972, see also Section 6.2.5). In this submodel heart function is represented graphically using Guyton's technique (1955) which equates cardiac output with venous return. Curves can be drawn relating both cardiac output and venous return to right atrial pressure (filling pressure) of the heart. The cardiac output curve is known as a "cardiac function curve", and depends upon various factors (e.g. neural output, concentration of potassium ions, etc.) which alter its shape and the effectiveness of the heart as a pump. The curve of venous return against right atrial pressure is known as a "systemic

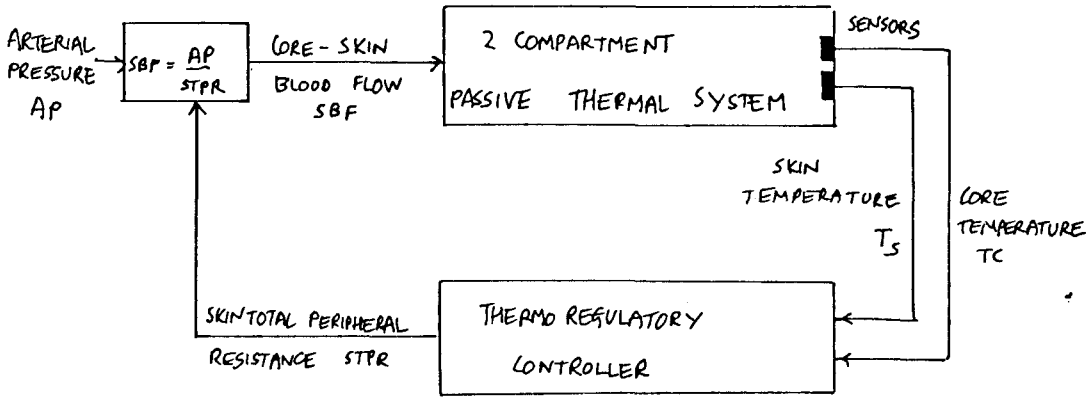


FIGURE 7.1.(a). OVERALL FEEDBACK STRUCTURE.

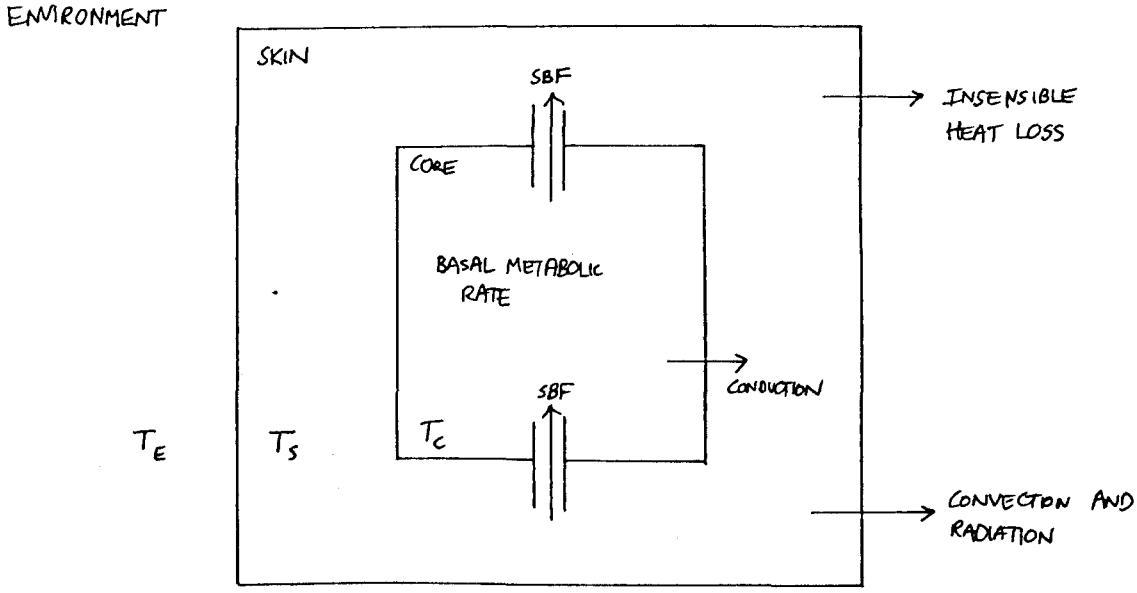


FIGURE 7.1(b). TWO COMPARTMENT PASSIVE THERMAL SYSTEM.

FIGURE 7.1. THERMOREGULATORY SYSTEM SUBMODEL.

function curve" and is modified by the mean systemic pressure of the circulation. The mean systemic pressure is the pressure that the blood would be at throughout the body if blood flow was zero, and is related monotonically to the volume of blood. The main aspects of the cardiovascular system submodel are depicted in Figure 7.2.

The blood volume is, in turn, a linear function of the extracellular fluid volume. (In Figure 7.2 the extracellular fluid volume is shown as the time integral of the net fluid input rate, but in the full model, fluid is distributed between intracellular and extracellular compartments in order to balance osmotic pressures; see Section 7.2.3.4.) The systemic and cardiac function curves allow the determination of cardiac output given the mean systemic pressure and the level of cardiac hypoeffectiveness. Arterial pressure is then equal to the product of cardiac output and total peripheral resistance. Total peripheral resistance is controlled by thermoregulation and the presence of a vasoconstricting hormone, angiotensin II (associated with the renal control of blood pressure). The modelling of the effect of angiotensin II on peripheral resistance is based on the experimental results of Deheneffe et al. (1976). In the model of Guyton et al. (1967, 1972) the loop of the cardiovascular model is closed by deriving an empirical relationship between arterial pressure and urine flow rate; however, in the Uttamsingh model, an explicit model of the kidney function is provided which indirectly closes the loop (see next section).

7.2.3.3 Kidney function submodel

Although the kidney consists of millions of separate nephrons, the processes that occur in each are very similar and, for the purposes of the model, all kidney function is assumed to take place in one very large nephron. The kidney function submodel is an algebraic representation of the instantaneous flows of fluid and substances in various sections of the nephron. The model is based on currently accepted theories concerning the formation of the glomerular filtrate and the subsequent processing in the major sections of the nephron (e.g. Guyton, 1971, pp. 397-). The sections of the nephron which are modelled are Bowman's Capsule (in which the preliminary filtering of the blood takes place across the glomerular membrane to form the glomerular filtrate), the proximal tubules (in which a large proportion of fluid and wanted solutes are reabsorbed back into the blood), the loop of Henle (in which the osmolality of the urine, and hence body electrolytic balance, is carefully

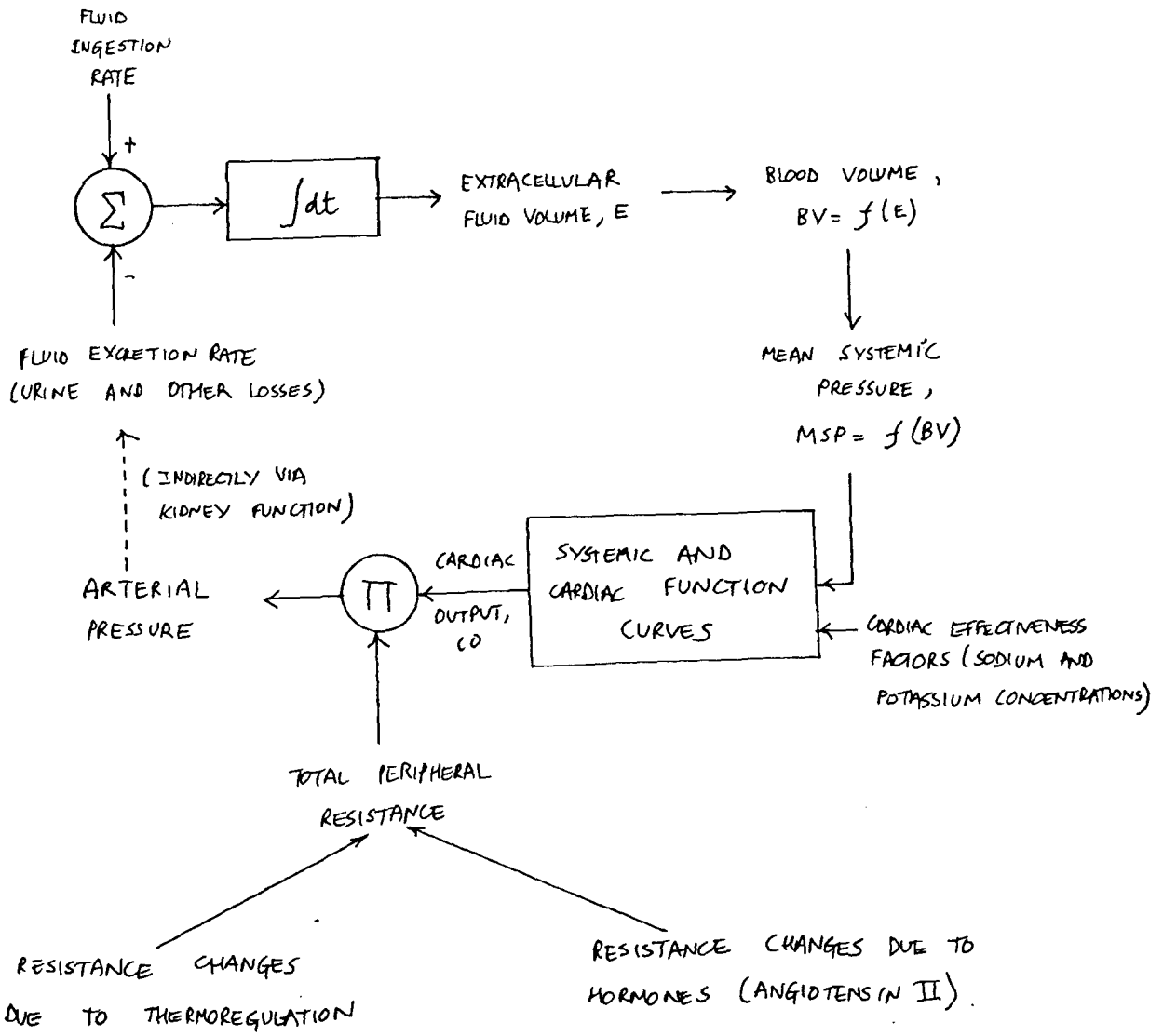


FIGURE 7.2. SUBMODEL OF THE CARDIOVASCULAR SYSTEM (BASED ON GUYTON ET AL., 1967, 1972).

controlled), and the distal tubules (in which fluid and sodium are further reabsorbed and potassium secreted under the controlling influence of hormones) (Figure 7.3).

In each section, there are equations for the reabsorption of fluid and sodium (the potassium in the urine is assumed to be secreted in the distal tubules). The excretion of the waste products, urea and creatinine is not modelled in the kidney function submodel since they are not reabsorbed (see Section 7.2.3.5). In the distal tubules, the reabsorption of fluid is dependent on the level of ADH (anti-diuretic hormone), and the reabsorption of sodium and secretion of potassium is a function of the level of aldosterone (see Section 7.2.3.6).

Although the actions of the various sections of the nephron are well-known in qualitative terms, the understanding of quantitative effects is much more uncertain; many experiments, for instance, can be performed only on animals. In the Uttamsingh model, the relationship between glomerular filtration rate (GFR) and arterial pressure is based on the results of Shipley and Study (1951), the absorption characteristics of the loop of Henle are derived from the animal experiments of Landwehr et al. (1968), and the effect of ADH on fluid reabsorption is taken from the model of Dehaven and Shapiro (1970). Other relations were derived by careful interpretation of detailed physiological texts (e.g. Guyton, 1971) and by adjustment until the overall model response was most realistic (e.g. the choice of a minimum value for glomerulo-tubular balance in the proximal tubule, Uttamsingh, 1981).

7.2.3.4 Fluid and electrolytic balance submodel

The fluid and electrolytes (sodium and potassium) in the model are divided into two compartments which represent the intracellular and extracellular "pools" of these substances. Balance equations are written for the fluid volumes and electrolyte masses in each compartment. The inputs and outputs to this submodel are into or out of the extracellular compartment only and consist of: ingestion rates of fluid, sodium and potassium through the gut and the excretion rates of these through the kidneys in urine (in man there are also some losses in the faeces and from the surface of the skin). The basic structure of the submodel is shown in Figure 7.4. It can be seen that there is an additional loss of fluid and electrolytes through the artificial kidney machine when the patient is undergoing dialysis (the exchange of solutes with the kidney machine is

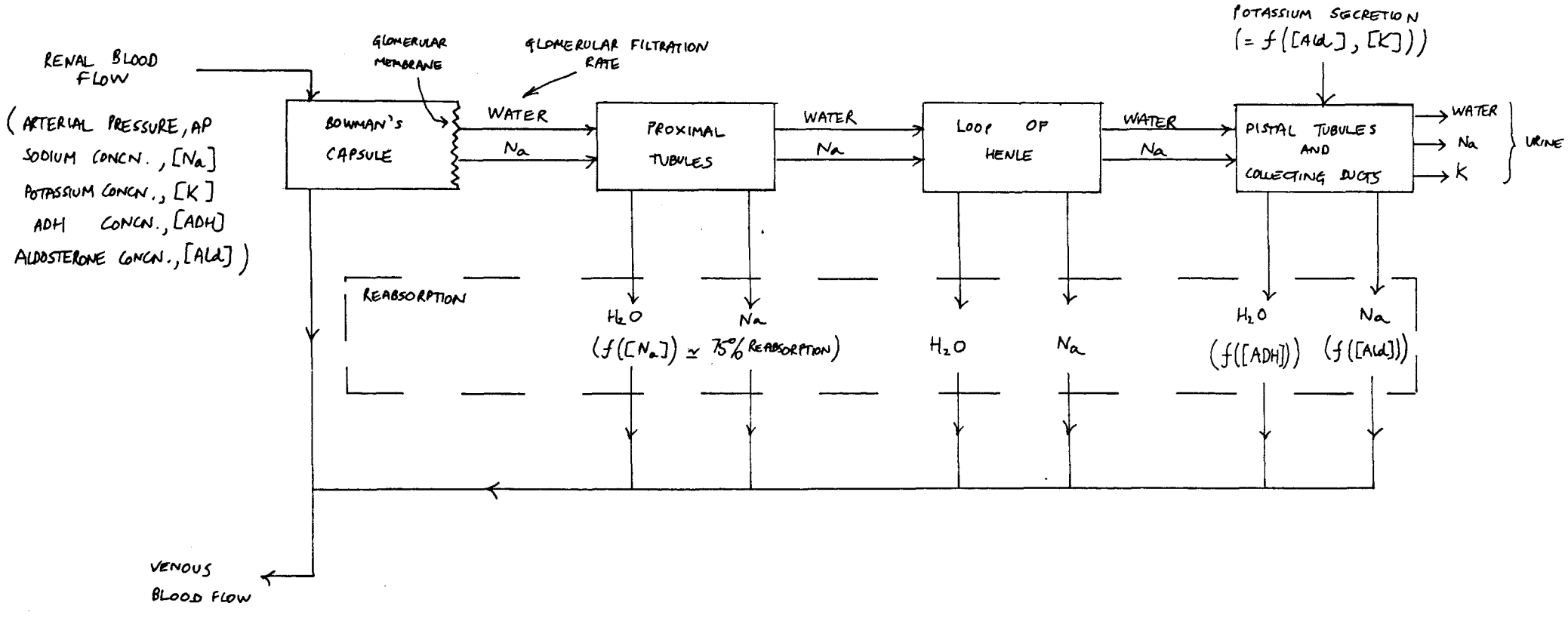


FIGURE 7.3. STRUCTURE OF THE KIDNEY FUNCTION SUBMODEL (AS A SINGLE NEPHRON).

based on a process of diffusion, and there may be a net flow into the patient if the concentration of solute in the dialysing fluid is greater than that of the blood).

In the model it is assumed that the active transport and passive diffusion of sodium and potassium ions across the cell membrane balance exactly and therefore only fluid is transferred between the intracellular and extracellular compartments in the submodel. Consequently, the masses of electrolytes in the intracellular compartment remain constant. In the human, water is transferred rapidly across the cell membrane in order to maintain osmotic equilibrium. This is modelled by an exchange of water between the two compartments which ensures that the intracellular and extracellular osmolalities are equal (the exchange is performed at each integration step of the model, ≈ 1 min). The osmolality depends on the concentrations of electrolytes in a fluid and hence the fluid volume as well as electrolyte masses. Thus the fluid and electrolyte systems are intimately connected, as summarised in Figure 7.5. (It is the interaction between the positive and negative feedback loops in Figure 7.5 which allows the conditions in the cell to exist and be stable in order that vital cellular processes may occur.)

The extracellular fluid volume is composed of the plasma volume of the blood and the volume of the interstitial fluid. It is assumed that electrolytes are evenly distributed throughout both volumes such that their osmolalities are equal. In the cardiovascular submodel, the blood volume is expressed as a function of the extracellular fluid volume (see Figure 7.2).

7.2.3.5 Urea and creatinine dynamics submodel

Urea and creatinine are the major end products of metabolism within the cell. In the model, the production of urea and creatinine is represented by constant rate inputs to the intracellular compartment (Figure 7.6). Balance equations are written for intracellular and extracellular urea and creatinine. The exchanges across the cell membrane, and losses by excretion (and through the kidney machine) are diffusion processes, and therefore concentration dependent. Thus the intracellular and extracellular fluid volumes are time-varying inputs to the otherwise linear urea and creatinine dynamics. The parameters of the submodel and the generation rates for urea and creatinine are based on Cooney (1976) and Frost and Kerr (1977).

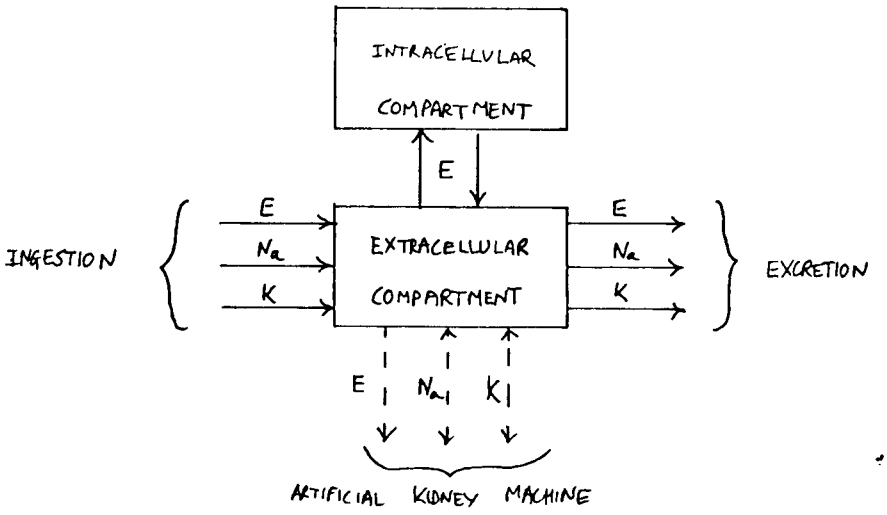


FIGURE 7.4. BASIC STRUCTURE OF THE FLUID AND ELECTROLYTE BALANCE SUBMODEL.

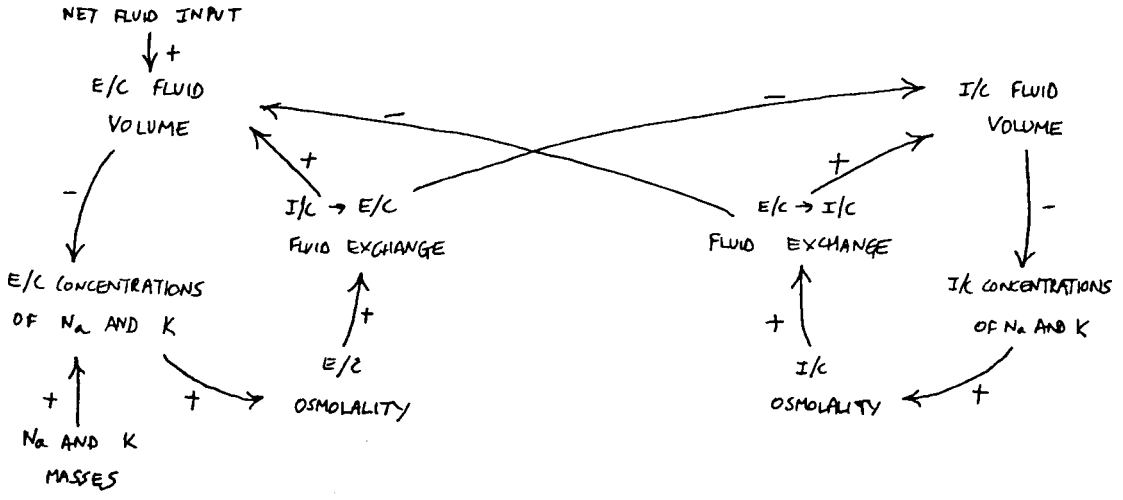


FIGURE 7.5. ARROW DIAGRAM SHOWING HOW OSMOREGULATION IS REPRESENTED IN THE RENAL MODEL.

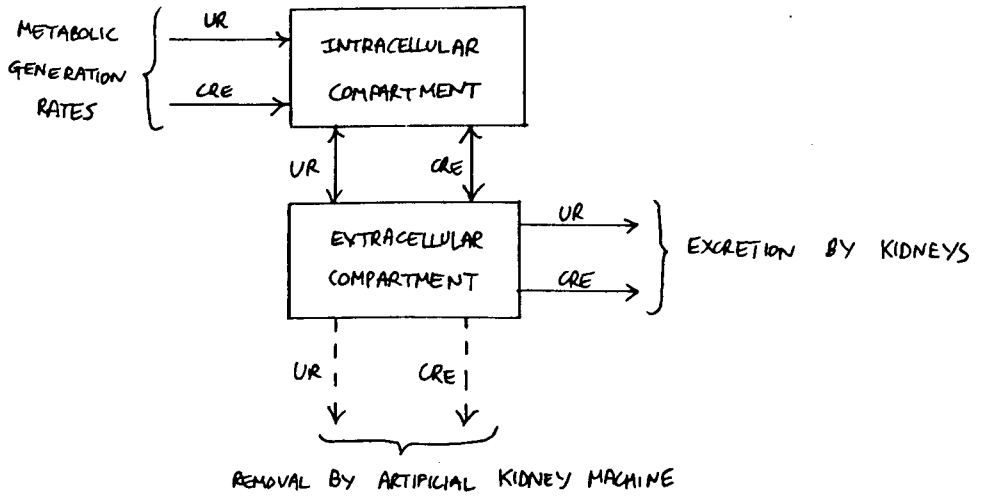


FIGURE 7.6. SUBMODEL OF UREA AND CREATININE DYNAMICS.

7.2.3.6 Hormonal dynamics submodel

The two hormonal systems having a major effect on the human renal system are the ADH (antidiuretic hormone) and the renin-angiotensin-aldosterone systems. Both of these systems are included in the model, and each hormone is simply represented as a first order dynamic system (or a one-compartment model). The submodel is complicated by the fact that the secretion, or creation, rates and clearance rates of the hormones are often complex nonlinear functions. The ADH model is shown in Figure 7.7, and the renin-angiotensin-aldosterone model in Figure 7.8.

The secretion of ADH occurs in the hypothalamus-pituitary system and it is well-known to be affected by both plasma osmolality and by increased blood volume (ADH, in turn, modifies the rate of water loss from the kidneys and hence controls these variables by negative feedback). The functional relationship in the ADH submodel between ADH secretion and excess fluid volume is based on Reeve and Kulhanek (1967), and the relationship with plasma osmolality is based on Dehaven and Shapiro (1970). The additive combination of the two effects (see Figure 7.7) is based on Johnson et al. (1970). The decay, or clearance, rate of ADH is a function of the concentration of ADH in plasma (Bigelow et al., 1973).

The secretion rate of renin in the model (Figure 7.8) is assumed to be a function of the sodium flow rate in the distal tubules of the kidney (see Section 7.2.3.3). This embodies different theories concerning the mechanism for renin release from the macula densa cells of the juxtaglomerular apparatus (Uttamsingh, 1981, or Section 7.4.1). The equation for renin secretion is derived from Thurau (1971). Renin is an enzyme which acts on its substrate releasing angiotensin I, which is rapidly converted to angiotensin II. The rate of formation of angiotensin I (and hence angiotensin II) is given by the Michaelis-Menten equation in terms of the concentrations of renin and its substrate. The parameters for the renin-angiotensin kinetics were derived by Haas and Goldblatt (1967).

Aldosterone is secreted from the adrenal cortex into the plasma (the adrenal glands are located on top of the kidneys). The factors affecting aldosterone secretion rate in the model are the concentration of angiotensin II and the plasma concentration of potassium. The relationships in the model were extrapolated from the results of Blaine et al. (1972; dependency on ang. II) and Seif (1974; effect of plasma potassium concentration). The renin-angiotensin-system controls blood pressure via the vasoconstrictive effect of angiotensin II (Section 7.2.3.2), and plasma

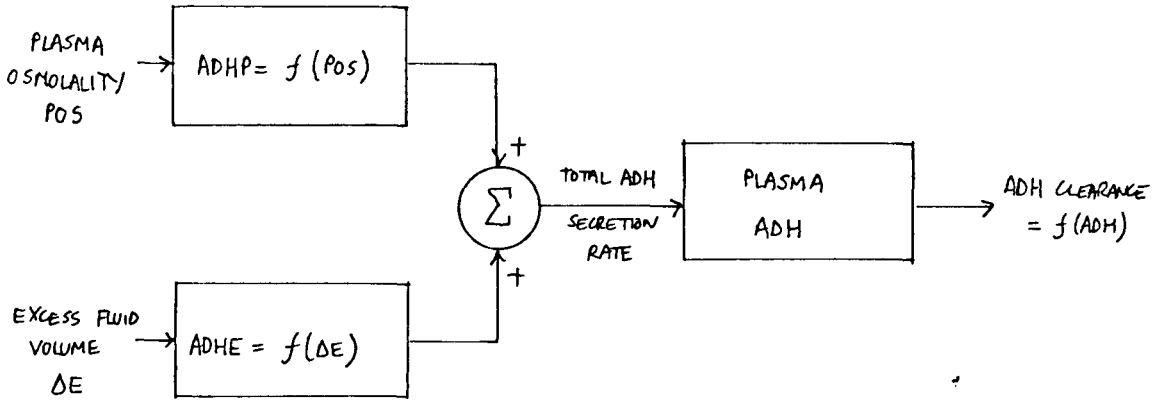


FIGURE 7.7. ADH DYNAMICS SUBMODEL.

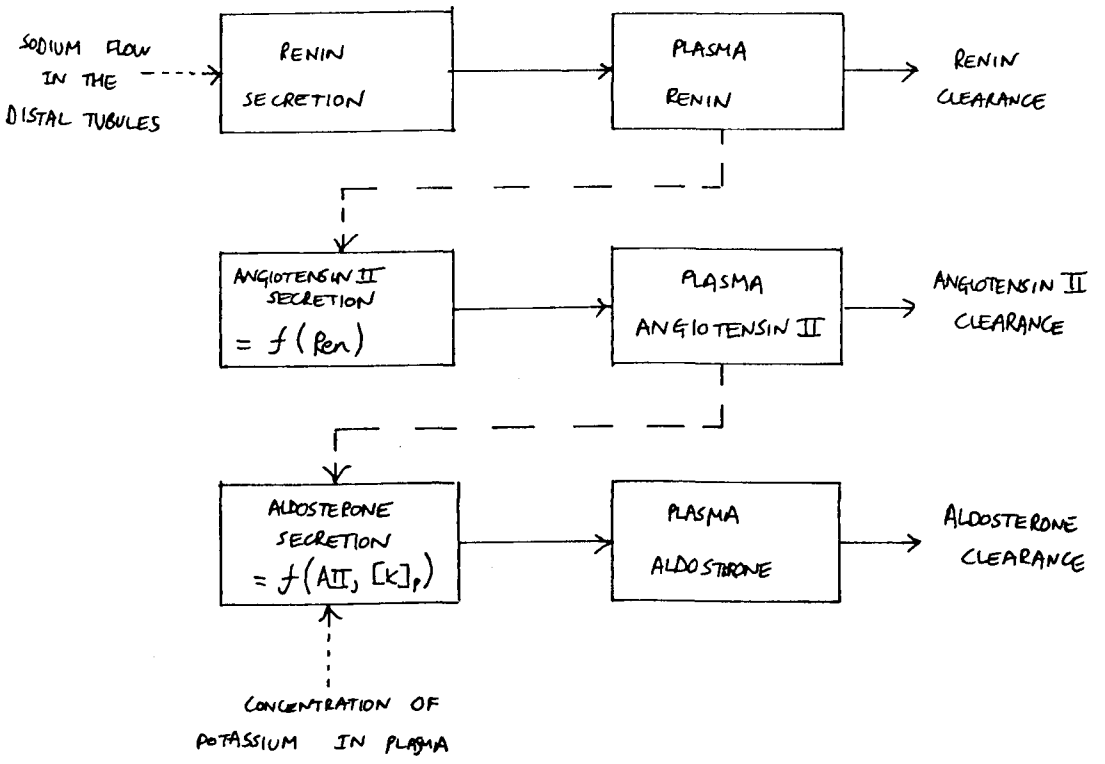


FIGURE 7.8. SUBMODEL OF THE RENIN - ANGIOTENSIN II - ALDOSTERONE SYSTEM.

sodium and potassium levels by the effect of aldosterone on kidney function (Section 7.2.3.3).

The clearance rates for renin, angiotensin II, and aldosterone were determined from the values of steady-state concentrations of the hormones, assuming linear mass removal dynamics. Although the hormonal system dynamics submodel is based on empirical results and other models where possible, there is still a large amount of uncertainty. This is partly because of the extrapolation of some relationships from animal experiments, and partly because of the lack of full understanding of many of the basic mechanisms involved. (The same is true concerning the modelling of the effects of the hormones on the renal system).

7.2.3.7 Kidney failure submodel

The complete or partial failure of the kidney is represented by four dimensionless parameters which multiply the appropriate terms in the model equations describing the excretion or secretion of substances by the kidney. Each parameter lies in the interval $[0, 1]$ where the value unity indicates that kidney function is normal, and the value zero indicates that there is no remaining kidney function. The four parameters are defined as follows:

FACT1 = Remaining fractional ability to excrete water and sodium
FACT2 = Remaining fractional ability to secrete renin
FACT3 = Remaining fractional ability to excrete potassium
FACT4 = Remaining fractional ability to excrete urea and creatinine.

The adequacy of this technique for representing renal failure is discussed in Section 7.4.1. In Section 7.4.3, these parameters are varied in an estimation procedure which fits the model to dialysis data from an individual patient.

7.2.3.8 Artificial kidney machine submodel

During haemodialysis, arterial blood from the patient is pumped through the artificial kidney machine and returned to a vein. In the machine, blood is separated from the dialysis fluid by a semi-permeable membrane across which diffusion of solutes and leakage of water occur. The rate of transfer of each solute depends upon the blood flow rate and the concentration gradient across the membrane, and is modelled by a first order diffusion equation. (The dialysis fluid contains no urea and creatinine and so these are removed rapidly from the blood. The

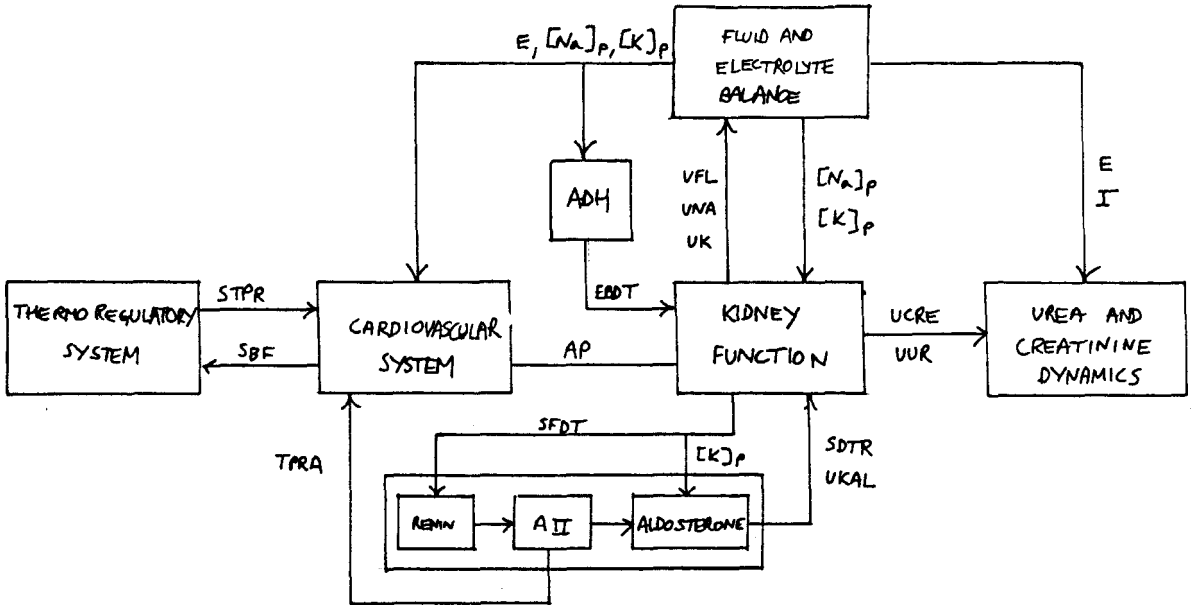
concentrations of electrolytes in the dialysis fluid are carefully controlled to regulate the electrolytic balance of the patient.) The terms representing the water loss and transfer of solutes in the dialysis machine are subtracted from the balance equations for water, sodium, potassium, urea, and creatinine in the extracellular compartment.

7.2.3.9 Overall structure of the model

The interactions between the separate submodels to form the complete model are illustrated in Figure 7.9. In the model, as in the human, there are many levels of behaviour ranging from the compartmental dynamics of various substances, to control systems (thermal, cardiovascular, electrolytic balance), to the overall adaptive control system formed by combining the submodels. The kidney failure submodel (not shown in Figure 7.9) has the effect of modifying the kidney function submodel. The artificial kidney machine submodel (also not shown) interacts with both fluid and electrolytic balance, and urea and creatinine submodels.

7.2.3.10 Model simulation

The overall model consists of 16 first-order differential equations and approximately 50 algebraic equations. Many equations are highly nonlinear (often approximated by piecewise linear fits) and discontinuous, and therefore the model is not amenable to analytic solution or treatment. The simulation model is written in FORTRAN IV and the differential equations are solved using numerical integration (Uttamsingh, 1981). An integration step of 1 minute results in a stable solution using a first-order Euler integrator. The model is used to represent an individual patient by using the clinical data as initial conditions for a simulation run. Table 7.1 shows the information required to specify the initial conditions and inputs for a patient undergoing dialysis. A typical simulation run is of a patient on dialysis for 4 or 6 hours and the following 24 or 48 hours, with results printed every 30 minutes or 1 hour. The original program written by Uttamsingh (1981) is in an interactive form which allows easy use by clinicians unfamiliar with computing.



KEY:

- AP ≡ ARTERIAL PRESSURE
- E ≡ EXTRACELLULAR FLUID VOLUME
- EBDT ≡ FLUID FRACTION REABSORBED IN THE DISTAL TUBULES
- I ≡ INTRACELLULAR FLUID VOLUME
- $[K]_p$ ≡ PLASMA POTASSIUM CONCENTRATION
- $[Na]_p$ ≡ PLASMA SODIUM CONCENTRATION
- SBF ≡ CORE – SKIN BLOOD FLOW
- SDTR ≡ SODIUM REABSORBED IN THE DISTAL TUBULES
- SFDT ≡ SODIUM FLOW IN THE DISTAL TUBULES
- STPR ≡ SKIN TOTAL PERIPHERAL RESISTANCE
- TPRA ≡ SKIN RESISTANCE DUE TO ANGIOTENSIN II (AII)
- UCRE ≡ URINE CREATININE FLOW RATE
- UFL ≡ URINE FLUID FLOW RATE
- UK ≡ URINE POTASSIUM FLOW RATE
- UKAL ≡ URINE POTASSIUM FLOW RATE
- UNA ≡ URINE SODIUM FLOW RATE
- UUR ≡ URINE UREA FLOW RATE

FIGURE 7.9. INTERACTIONS AND OVERALL STRUCTURE OF THE RENAL MODEL (KIDNEY FAILURE AND ARTIFICIAL KIDNEY MACHINE SUBMODELS ARE NOT SHOWN).

State of Patient	Ability to excrete sodium and water Ability to secrete renin Ability to excrete potassium Ability to excrete urea and creatinine Effectiveness of heart Pre-dialysis weight
Average Daily Ingestion Rates	Average daily fluid intake Average daily sodium intake Average daily potassium intake
Initial Values of Model variables	Core temperature Skin temperature Extracellular fluid volume Sodium concentration in plasma Potassium concentration in plasma Intracellular fluid volume Sodium concentration in intracellular fluid Potassium concentration in intracellular fluid Plasma renin concentration Plasma angiotensin II concentration Plasma aldosterone concentration Arterial pressure Plasma ADH concentration Urea concentration in extracellular fluid Creatinine concentration in extracellular fluid
Dialysis data	Sodium concentration in dialysate Potassium concentration in dialysate Length of time on dialysis Ultrafiltration pressure Average blood flow rate through kidney machine

Table 7.1: Information Required to Simulate Model

7.3 Development of a Programme of Model Validation

The programme of validation for the Pullen cardiovascular model (Chapter 6) was structured using the theory of model validity (Chapter 4) and lengthy considerations of the modelling objectives, available data types, and the appropriate validity criteria. In developing a programme of validation for the Uttamsingh renal model the same approach (i.e. from "first principles") may be adopted or, by contrast, a standard validation methodology (e.g. as proposed in Chapter 5) suitable for this type of model may be employed. The latter approach is problematic in that the model has primarily utilitarian objectives yet it is essentially a representational or explanatory model, whereas the former approach is very time-consuming. In this section, a programme of validation will be developed for the renal model which is based on standard methodologies but which are selected and modified by brief considerations of model objectives, data, and appropriate validity criteria.

In Chapters 4 and 5 the importance of the scientific "domain" (body of data, theories, models, etc. pertinent to the area) was stressed in affecting the model type as well as operative validity criteria and acceptable validation methodologies. The domain of renal physiology (a subdomain of human physiology) is a complex one, much of which is expressed in quantitative dynamic terms, and this implies that validation will consist largely in the comparison of model and data time series. However, there are many areas of uncertainty in the renal physiology domain, such as the precise nature of nephron function, the control effects and dynamics of hormones or, possibly, other mechanisms or control systems that have not yet been discovered. Some of these may be understood only qualitatively, or may have been investigated only in animals. The model, therefore, is a theoretical advance and may have to be judged partly by general heuristic considerations.

7.3.1 Modelling objectives

The objectives of the Uttamsingh renal model were presented in Section 7.2.1. The primary general objective of the model is utilitarian - the improvement of a health care system. The specific utilitarian objective is as an aid in the management of patients with impaired renal function undergoing periods of dialysis on an artificial kidney machine, and this defines the "system of interest" (SOI) in which the model is to be used. The model is required to provide predictions concerning the

clinical states of such patients which may be used to improve or optimise their therapies. Thus the specific utilitarian objectives entail specific scientific objectives and a corresponding "intended range of application" ($\mathcal{R}_I$). $\mathcal{R}_I$ is therefore the clinical state of the human renal system with a range of possible failure modes and both during and between periods of dialysis. $\mathcal{R}_I$ includes events within the renal system that take place over a time scale of 4 hours (i.e. over a dialysis) to 1 week. The time resolution of $\mathcal{R}_I$ is approximately 30 minutes, but some important effects which occur in a shorter time are also included. The model is intended to be general and also patient-specific. (The "clinical state", in $\mathcal{R}_I$, is the substate of the overall state of the patient pertinent to the renal system, given current physiological understanding. It includes the states of the thermal, cardiovascular, fluidic, electrolyte, waste, and hormonal systems as well as the renal and artificial kidney systems.)

It is clear that, in order to provide predictions of such a wide number of variables, the model must be a representation of the individual subsystems and the way they interact, and should include parameters, initial conditions, etc. which allow predictions to be made of an individual patient. In other words, the model should be isomorphic with $\mathcal{R}_I$. Consequently, the model embodies an explanation of $\mathcal{R}_I$ and may therefore satisfy general scientific objectives (e.g. insight, hypothesis testing, etc.) as well as general utilitarian objectives, a point made by Uttamsingh (1981; see his Section 3.3). Although this makes the model scientifically more interesting than, say, a statistical forecasting model, unfortunately it complicates the programme of validation. This is discussed further in Section 7.3.3.

7.3.2 Available data types

The full validation of the model requires three sources of data: (i) data from the renal system for representational validation; (ii) data concerning the future clinical state for predictive validation; and (iii) data from the SOI on the effectiveness of the model in use. Data source (ii) is, in fact, a subset of data source (i), but is separated because the predictive empirical validity of the model is directly related to its pragmatic validity (i.e. for utilitarian objectives). Since the model has not been used in practice, no data are available from source (iii) (pragmatic validation must therefore be on a critical level, see Section 7.3.3). In the rest of this section, some remarks

will be made on the availability, uncertainty, and other aspects of data types from the first two sources.

The major variables of the model are easily measurable in the human and are usually routinely measured in patients with renal failure (e.g. before and after dialysis). They include: arterial pressure; skin and core temperatures; plasma concentrations of sodium potassium, urea, creatinine, and hormones; estimates of extracellular and intracellular fluid volumes; and total body weight. The average (as opposed to instantaneous)urine flow rate and composition can be determined. It is these variables which are important to the clinician and against which the model must be compared in order to test its predictive validity. They may be used jointly to check the representational validity of the model.

Ideally, empirical validation (for representation) should also compare the "internal" variables of the model, such as the details of flows of fluid and electrolytes in various parts of the kidney nephrons, the secretion rates of various hormones, intracellular concentrations, etc. However, the behaviour of many such variables cannot be determined in vivo in humans, although it is frequently known in qualitative terms (i.e. an observational data type). Most of the quantitative data for these variables is available from experiments on animals and care must be taken in applying them to humans. (Where model variables are not available as quantitative data, the validity of their responses may sometimes be tested by using the concept of identifiability and the techniques of parameter estimation.)

Many of the parameters in the model cannot be directly measured but may be determined indirectly using models or by extrapolation from animal data. In Section 7.4 one of the empirical validation tests is concerned with the use of the model itself for the indirect measurement of the kidney failure parameters and the urea generation rate.

A simulation of the model requires other information concerning the average daily ingestion rates of fluid, sodium, and potassium, and details of the dialysis therapy. Such data are available. In predicting the behaviour of a system such as the renal system, it is important that the assumed ingestion rates are correct since they have a significant effect on the model response. Patients with renal failure (particularly those in hospital) are usually on strictly controlled diets which ensure the validity of such data. Other information, such as the occurrence of vomiting during dialysis or the removal of bedding (affecting thermoregulation),

is rarely recorded, yet is very important when interpreting the model response against data.

In comparing the model with data, the uncertainty of the data must be taken into account. The measurement uncertainties of some of the major variables are shown in Table 7.2. In practice, the total uncertainty will be more than that shown because of effects on the patient which are not modelled yet produce changes in $\mathcal{R}_I$. A further source of uncertainty arises from the time resolution of the data. Measurements are usually made just before and just after dialysis, and so the data only represent trends in the behaviour of $\mathcal{R}_I$.

7.3.3 Appropriate validity criteria and programme of validation

If the model is to be used simply as a predictive device to improve the SOI (patient management during dialysis therapy), then the programme of validation should be a δ-methodology for utilitarian modelling objectives, as described in Chapter 5. The various stages of the methodology are shown in Figure 7.10. The symbols V_{ALG}, V_{CON}, V_{EMP}, V_{PRAG1}, and V_{PRAG2} denote the types of validity criteria and refer to algorithmic, consistency, empirical, direct pragmatic, and general pragmatic validity criteria, respectively. The first stage, which is a necessary precondition, is concerned with initial tests of consistency, algorithmic validity, and stability (for more details consult Chapter 5).

The second stage involves the comparison of the model responses with data from $\mathcal{R}_I$ in order to test its predictive validity and therefore involves empirical validity criteria. This is concerned only with the validity of predictions of variables of interest to the clinician (which include arterial pressure, skin and core temperatures, plasma concentrations (in particular of urea, sodium, and potassium), urine flow rate, etc.). The comparison of model and data may be based on features such as the change in a variable (e.g. arterial pressure) over dialysis rather than its absolute value. Sensitivity analysis may be used to determine the confidence in the model predictions which should lie within data uncertainty intervals. The predictive validity can be established in this way since data are widely available concerning the states of patients before dialysis, just after dialysis, and at some later time (typically the start of the next dialysis). In other words, data are available over the time span of $\mathcal{R}_I$ (at this overall level).

Variable	Estimated Percentage Uncertainty
Time	± 0.3%
Arterial Pressure	± 5%
Plasma Sodium	± 1%
Plasma Potassium	± 5%
Plasma Urea	± 2%
Plasma Creatinine	± 2%

Table 7.2: Measurement Uncertainties of the Major Variables

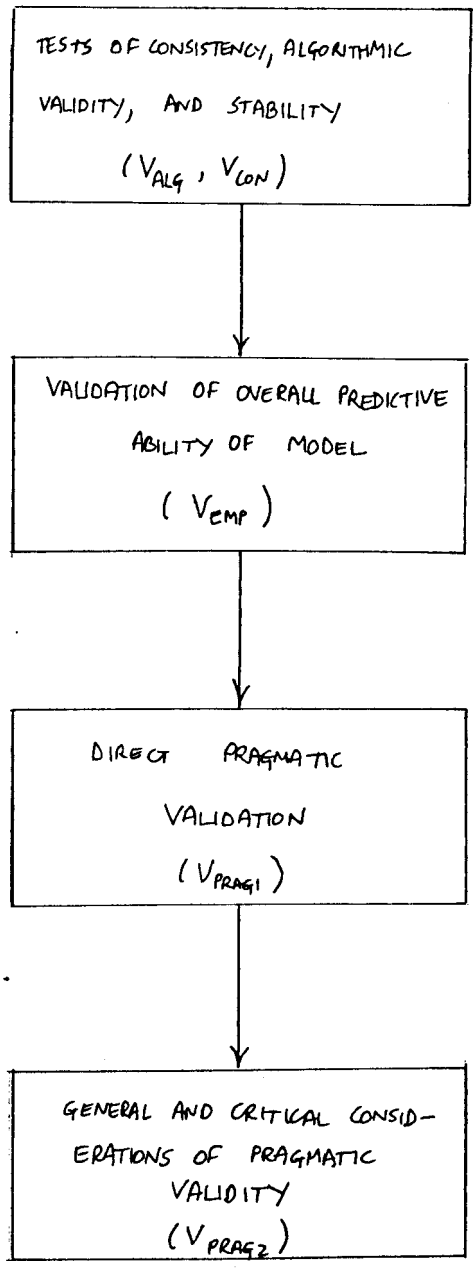


FIGURE 7.10. A PROGRAMME OF VALIDATION BASED ON THE δ-METHODOLOGY FOR UTILITARIAN MODELLING OBJECTIVES.

The third stage is concerned with the validity of the model in terms of its use in SOI, i.e. its direct pragmatic or utilitarian validity. Since the model has not yet been used in practice, no data are available on the effect of SOI of using the model. However, a partial test can be made of the pragmatic validity of the model as a predictive instrument in renal health-care by asking questions such as: "Can the predictions of the model lead to an improvement of dialysis therapy?". It would be necessary to examine the accuracy of model predictions, and also the ability of the model to predict situations which are either clinically satisfactory (such as stability) or clinically bad (such as rapidly falling arterial pressure, or significant temperature changes) in advance of the clinician.

The final stage is a more general, critical assessment of the model and approach in tackling the problem of renal dialysis therapy. In this stage, for instance, the time, effort, and resources expended in developing the model might be compared with the benefits likely to follow from its use, and also with alternative methods for improving dialysis therapy.

As far as the specific utilitarian objectives of the model are concerned, the most important stage of the above programme is the third (direct pragmatic validation). If it passes the preliminary tests of the kind indicated, then it should be implemented in a practical clinical situation as a predictive aid for dialysis therapy (i.e. a "clinical trial") and data acquired from SOI (the wider renal health-care system, as opposed to the renal system itself or $\mathcal{R}_I$) in order to test its pragmatic validity fully. However, before this stage is entered, the second stage should be examined further.

The second stage is aimed at determining the representational validity of the model at an overall level by comparison with empirical data. For simple models, or black-box models, this is satisfactory, but for complex models, such as the Uttamsingh renal model, which are based explicitly on an understanding of the system and its subsystems it is inadequate. The testing of the representational validity of the renal model should include explicit validation of the submodels as well as the overall model and involve tests of theoretical coherency and empirical correspondence. In this way, the validity of the model mechanisms that generate the predictions, and not simply the predictions, can be determined. Furthermore, this allows an assessment of whether the complexity of the model is necessary in order to achieve the type and accuracy of

predictions required by the specific utilitarian modelling objectives. The alternative method for assessing representational validity is depicted in Figure 7.11. V_{THEOR} denote theoretical validity criteria. In Figure 7.11 the importance of involving domain knowledge (i.e. up-to-date renal physiology) and new data for validation is stressed.

This method has the additional advantage of being precisely the type required for the validation of models with scientific objectives and therefore it is suitable for the other intended objectives of the renal model such as hypothesis testing and discovering new patterns of behaviour (see Section 7.3.1). In fact, it has the same structure as that used in the validation of the Pullen cardiovascular model (Chapter 6), and also the γ-methodology (Chapter 5). In applying the theoretical and empirical criteria, at each level, both qualitative and quantitative aspects should be considered. For empirical comparisons, the α-methodology based on features (Chapter 5) and culminating in a sensitivity analysis or a parameter estimation procedure is appropriate. The disassembly of the model into elementary submodels is obviously guided by the separate physiological functions that the model represents (Section 7.2), and the whole or partial reassembly forms a tree structure (Figure 7.12). In going from level 1 to level 5 the validity of the overall model is built up deductively and areas of confidence or uncertainty can be identified. At the same time, inferences may be made "down" the tree about the validity of individual submodels.

The boxes in Figure 7.12 indicate the submodels which are empirically validated in this study (Section 7.4) and also Uttamsingh's thesis (1981). In formulating the model there was implicit validation of levels 1 and 2. However, a full validation of the model would require detailed theoretical and empirical tests of these levels. In testing the model at the higher levels, standard physiological experiments (such as water, saline or hormonal loading) are powerful tests of the model since they are designed to test the effectiveness of various bodily systems and the causes of incorrect responses are often well-known. This form of model validation may be described as "pathological". It must be ensured, however, that the significant changes caused by these tests lie within $\mathcal{R}_I$.

For the general scientific objectives of the renal model, the third and fourth stages of Figure 7.10 are inappropriate. Figure 7.13 outlines a general methodology for the validation of the renal model. After the second stage the methodology splits according to whether scientific or

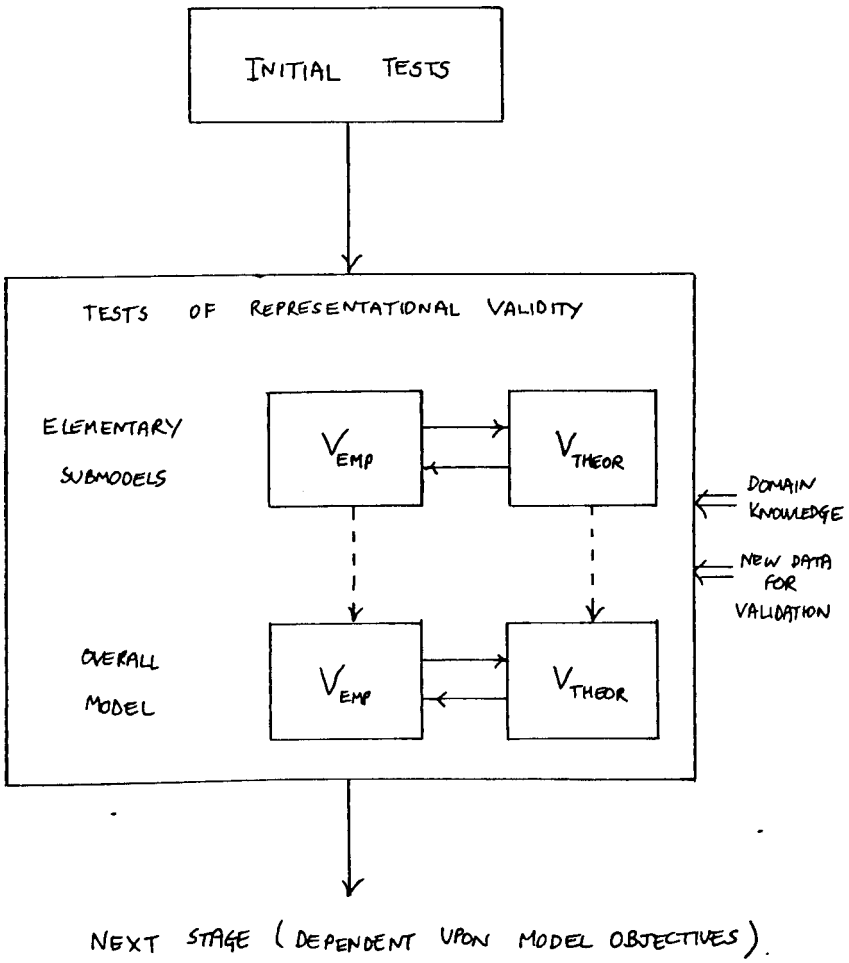


FIGURE 7.11. AN ALTERNATIVE METHODOLOGY FOR ASSESSING THE REPRESENTATIONAL VALIDITY OF THE MODEL.

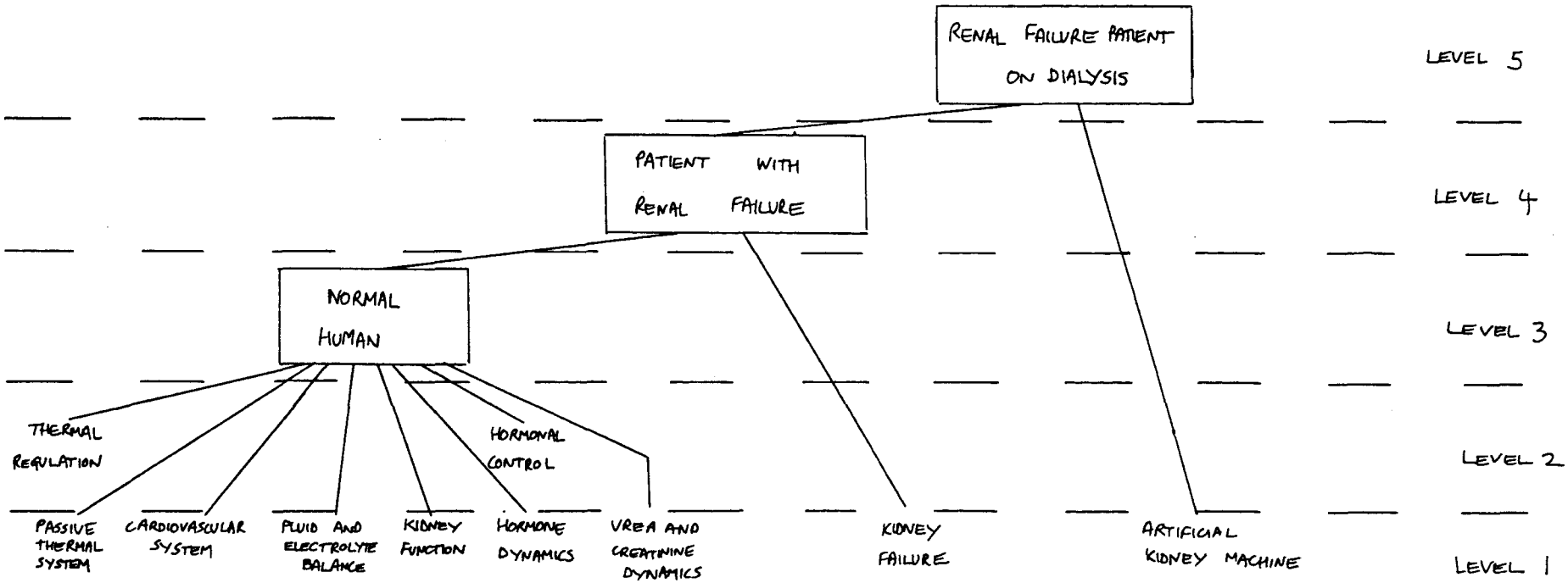


FIGURE 7.12. HIERARCHY OR TREE OF VALIDATION FOR RENAL MODEL.

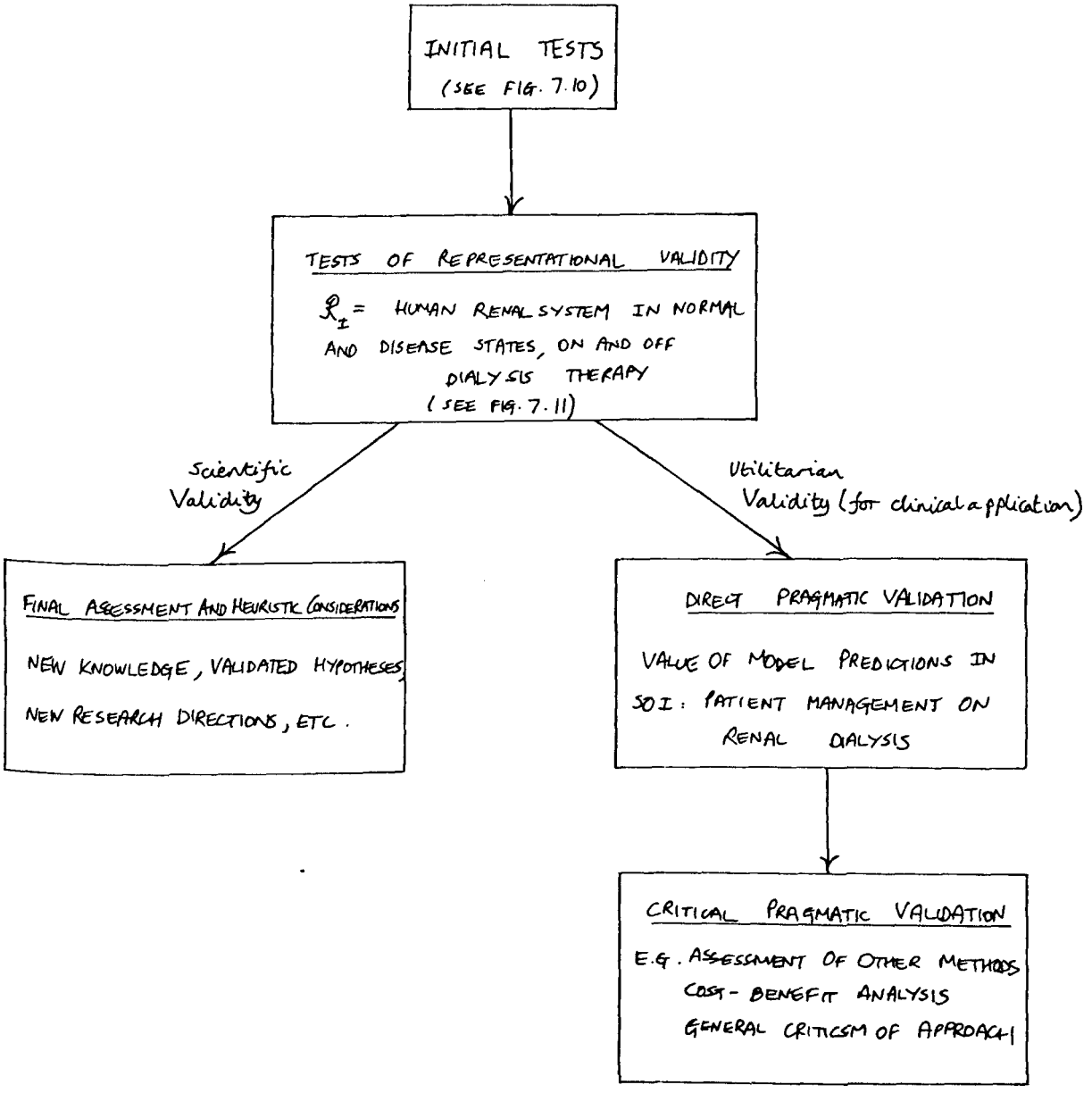


FIGURE 7.13. AN OUTLINE OF A GENERAL VALIDATION METHODOLOGY FOR THE RENAL MODEL.

utilitarian validity is being examined. Both of these paths may be followed, but it is clear that conflicting decisions on the validity of the model may be made (see Section 7.5). In the next section, some results of the validation study are presented. These are sufficient, however, to make decisions on the validity of the model for scientific and utilitarian objectives, and also to generalise on the suitability of this type of model for clinical application.

7.4 Results of Validation

In this section, some of the results of applying the programme, or methodology, of validation developed in Section 7.3 to the Uttamsingh renal model will be presented. The emphasis will be on empirical tests performed on levels 3, 4, and 5 of the tree of validation shown in Figure 7.12 or, in other words, tests performed on the more or less complete model. In Section 7.4.1 some brief considerations on the theoretical validity of a few submodels (levels 1 and 2) are made (in a full validation study empirical criteria should also be applied to these levels). Section 7.4.2 outlines a series of tests performed on level 3 (normal model), level 4 (model with renal failure), and level 5 (failure model and dialysis therapy). These are aimed at showing that the overall responses of the model are realistic and also to make some inferences about the validity of individual submodels. The predictive validity of the model for patients undergoing periods of dialysis is assessed in Section 7.4.3, together with the possibility of tuning the model to individual patients by using parameter estimation.

This section, therefore, deals with the assessment of the representational validity of the model (the second stage of the methodology in Figure 7.13). The final stages of the methodology associated with pragmatic validity and general scientific validity (or heuristic validity) are considered in Section 7.5.

7.4.1 Some theoretical tests on submodels

These tests are aimed at determining the theoretical validity of the submodels. Theoretical validity criteria require that a model should cohere with physiological theory over $\mathcal{R}_I$. This simply means that each submodel should, if possible, conform to, or be based on, the currently accepted understanding (or explanation) of the subsystem it is intended

to represent. If the submodel is coherent with well-established theory then its theoretical validity establishes confidence in its deductive contribution to the representational validity of the overall model. An example of this is the cardiovascular submodel which is based on the graphical techniques and mathematical models of Guyton (1955, 1967, 1971, 1972) and which have been thoroughly tested empirically and theoretically. Another example is the submodel of the effects of haemodialysis (Renkin, 1956).

If there is no appropriate theory, or existing theories are contradictory or inadequate (e.g. descriptive theories when quantitative dynamic theories are required), the submodel must be validated in a different way. Ideally, this is by empirical validation. However, in the case of renal modelling, the appropriate experiments and measurements are often difficult to make (hence the theoretical inadequacy). Under these circumstances, it may be possible to determine indirectly the representational validity of a submodel by inference from the empirical validity of the overall model. Theoretical validation tests on submodels are very important in delimiting areas of uncertainty in a model and selecting those submodels for special attention when the model is validated as a whole. In the renal model, the main areas of uncertainty are the hormonal dynamics and control, and the representation of renal failure. Some aspects of these are considered below.

7.4.1.1 Renin release in the hormonal dynamics submodel

This is an application of theoretical validity criteria at level 1 in the hierarchy of validation (see Figure 7.12). The nature and role of renin in the renal system were discussed in the model description (Section 7.2.3.6). There is much uncertainty concerning the precise nature of the mechanisms affecting the release of renin and a variety of seemingly conflicting theories of renin release. This creates some difficulty in deciding against which theory to test the submodel. Renin is released into the blood from the granular cells (macula densa cells) of the juxtaglomerular apparatus which lies between, and in contact with, the afferent arteriole and the beginning of the distal tubule of each nephron. It is the siting of the juxtaglomerular apparatus, together with the experimental results on the long-term control of arterial pressure by renin, which has led to the various theories of renin release. The three main theories are summarised below.

(i) The intrarenal vascular receptor theory (Tobias et al., 1959)

This theory maintains that a reduction of the stretch of the arteriolar wall (due to reduced arterial pressure) leads to an increased rate of renin release from the granular cells. This causes an increase in the level of plasma aldosterone which promotes sodium reabsorption in the distal tubules. The reduced rate of sodium excretion results in an osmotically-induced increase of extracellular fluid volume, and blood pressure consequently rises.

(ii) The macula densa sodium load theory (Vander and Miller, 1964)

This states that a decrease in the total sodium ("sodium load") arriving at the macula densa cells through the tubular fluid results in an increase in the rate of renin release. Aldosterone acts to increase the reabsorption of sodium and control the plasma concentration of sodium, and hence extracellular fluid volume and arterial pressure.

(iii) The macula densa intraluminal sodium concentration theory (Thurau, 1972)

In this theory, the rate of renin release is proportional to the concentration of sodium in the tubular fluid entering the distal tubules. The increased level of angiotensin II in the plasma causes constriction of the afferent arteriole thereby reducing glomerular filtration rate and the rate of sodium excretion.

In the model (Uttamsingh, 1981), the rate of release of renin, RS, is given by:

$$RS = 0.0163 - 0.0093 \times SFDT \qquad \ldots\ldots\ldots (7.1)$$

where SFDT is the total sodium flow entering the distal tubules. The model therefore explicitly satisfies theory (ii) (the macula densa sodium load theory). However, since in the model SFDT depends on the glomerular filtration rate (GFR), which in turn is a function of the arterial pressure (AP), RS is indirectly related to arterial pressure and hence satisfies the intrarenal vascular receptor theory (i). Theory (iii) is concerned with the control of GFR by renin, mediated by the vasoconstrictive action of angiotensin II. In the model there is an experimentally derived relationship between GFR and AP in which the controlling role of the renin-angiotensin II system is implicit.

Thus the representation of renin secretion in the hormonal submodel is coherent with two of the proposed theories, (i) and (ii), and the model itself resolves the apparent contradiction with theory (iii). However, a

caveat must be added. The experiments of Lee (1969) offer invalidating evidence for the sodium load theory (ii), although this may be the result of other uncontrolled factors affecting renin release. In any case, the model is an excellent vehicle for investigating the complex interaction between multi-control loops and for examining different mechanisms for renin release.

7.4.1.2 ADH secretion and removal in the hormonal dynamics submodel

Although it is well-known that the ADH secretion rate is a function of both plasma osmolality and excess fluid volume, the precise way in which the two signals for ADH release combine is uncertain. When the signals are cooperatively interacting (i.e. increased plasma osmolality and negative excess fluid volume, or vice versa) the form of the combination of the signals makes little difference to the net rate of secretion of ADH. However, when the signals are non-cooperatively interacting (i.e. increased plasma osmolality and positive excess fluid volume, and the converse) the form of the combination is critical to the net secretion rate of ADH, and hence the control of extracellular fluid volume and plasma osmolality. Thus the lack of a complete theory of ADH secretion contributes to the uncertainty of the overall model, particularly in the case of non-cooperatively interacting stimuli.

Nevertheless, this test of theoretical validity suggests a critical empirical test on the overall model that may be used to determine the correct form for the combination of the signals. When a hypertonic saline load is given to a normal patient, both plasma osmolality and excess fluid volume increase, giving rise to conflicting signals for ADH secretion. If the model is simulated for such a test (which lies in $\mathcal{R}_I$) and compared with human response data for urine flow rate, any discrepancy between the model and data will very likely be caused by an incorrect combination of the signals for ADH secretion. In this way, a number of competing combination functions can be assessed and the best chosen. The results of the test are presented in Section 7.4.2.1.2. (This test is an application of empirical validity criteria to level 3 of Figure 7.12).

In addition to the uncertainty associated with the secretion of ADH, the theory concerning the removal rate of ADH at low plasma concentrations is inadequate. This is a consequence of the practical difficulties of measuring ADH dynamics at low concentrations. In the empirical validation at level 3 (Section 7.4.2.1), some tests are performed by which the ADH removal function can be indirectly validated.

7.4.1.3 Kidney failure submodel

The kidney failure submodel was described in Section 7.2.3.7. The modelling of kidney failure by four dimensionless multipliers, or parameters, which represent the remaining fractional ability of the kidneys to perform certain functions is original and therefore cannot be checked for theoretical validity directly. Instead, the submodel must be assessed by reference to the physiological theory of renal failure. As they stand, the parameters are inadequate since they allow the following physiologically impossible situation:

$$FACT1 = 0, \quad FACT3 \neq 0, \text{ and } FACT4 \neq 0 \qquad \ldots\ldots\ldots (7.2)$$

This would mean that the kidney may excrete potassium, urea and creatinine, but not water or sodium. However, solutes can only be excreted in solution and this therefore represents a non-physiological situation. In practice, a clinician would never specify such a combination of parameter values for a renal failure patient. However, the situation may arise in the automatic estimation of the parameters using the model (see Section 7.4.3). There may be other constraints on, or relations between, the FACT parameters which are related to particularly pathological states of the kidneys, and it may be necessary to include these explicitly in the kidney failure submodel.

In renal failure, patients show a considerable degree of adaptation in an attempt to overcome the incorrect functioning of the kidney. Thus there may be functional changes in the behaviour of the remaining nephrons which cannot be modelled using the multiplier technique. Uttamsingh (1981) argues that the submodel is a satisfactory representation for patients with only a small fraction of kidney function remaining. However, this leaves problematic the modelling of patients who have partial renal failure or, for instance, patients who have suffered acute renal failure and are beginning to regain kidney function after treatment.

Patients with renal failure frequently adapt their basal metabolic rate (in some patients BMR increases, whereas in others it decreases). In the model, the parameter representing the generation rate of urea (GUREA) may be adjusted in order to account for this change. However, GUREA is not a parameter that is usually known to the clinician. In the empirical validation at level 5 (Figure 7.12) it is shown how the value of GUREA may be determined by parameter estimation.

7.4.2 Empirical validation of levels 3, 4 and 5

In this section, empirical validity criteria are applied to levels 3, 4 and 5 of the validation hierarchy (Figure 7.12) which refer to the model of a normal human (Section 7.4.2.1), model of a patient with renal failure (Section 7.4.2.2), and the model of a failure patient undergoing dialysis (Section 7.4.2.3), respectively. The response of the model in terms of the overall clinical variables (urine flow rate; arterial pressure; plasma electrolyte, hormonal and solute concentrations, temperature, etc.; see Section 7.2.1) should correspond qualitatively and quantitatively to the data from humans (see Section 7.3.3). If possible, the error between model and data should lie within the uncertainty interval of the data (Table 7.2). However, in the case of renal failure, the FACT (failure) parameters are uncertain and a good match between model and data requires adjustment of these parameters. The use of parameter estimation techniques and the predictive validity of the model are assessed in Section 7.4.3.

7.4.2.1 Empirical validation at level 3 - model of a normal human

Three empirical tests of the model are made in this section: the effect of a water load (Section 7.4.2.1.1), the effect of a hypertonic saline load (Section 7.4.2.1.2), and the effect of an aldosterone load (Section 7.4.2.1.3). The results are taken from Uttamsingh (1981) who includes an additional test - the comparison of the renal model with simulations from the model of Guyton et al. (1972) on the effect of saline loading after the reduction of renal mass. The tests, for which data on the average, healthy human responses are available, are critical tests in that they affect all the subsystems represented by the model, and in particular they are effective tests of the function of the hormonal control systems (e.g. refer to Hawker, 1978, p. 23) and therefore of the more uncertain parts of the model. Some general comments on the results of the level 3 validation are made in Section 7.4.2.1.4.

7.4.2.1.1 The effect of a water load

The effect of an ingestion of 1 litre water is simulated in the model by increasing the volume of the extracellular fluid by 1 litre after a pure delay of 15 minutes. The response of the model and human data for the urine flow rate is shown in Figure 7.14. With the exception of the oscillatory burst in the data, the model and data match closely in qualitative shape, quantitative levels and, in particular, the duration of the

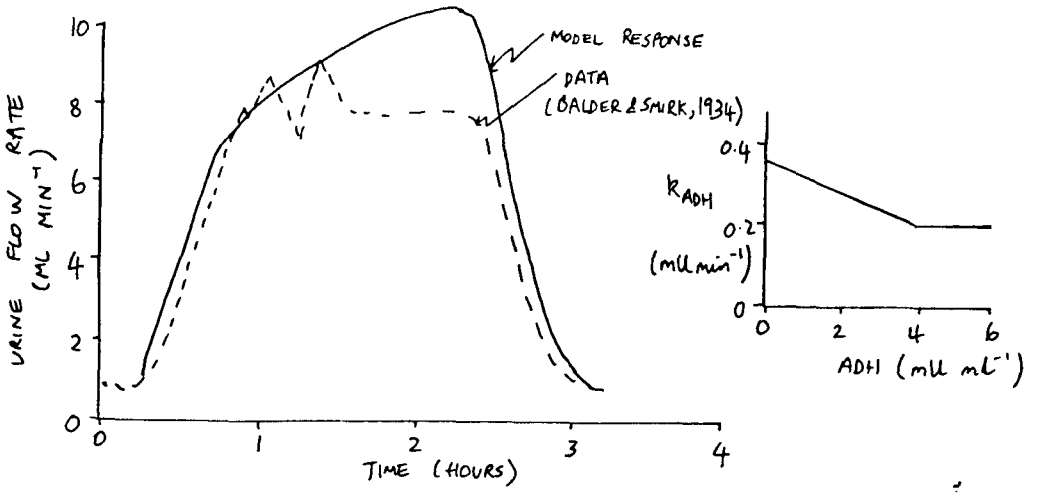


FIGURE 7.14. MODEL AND HUMAN URINE FLOW RESPONSE AFTER INTAKE OF 1 LITRE WATER.

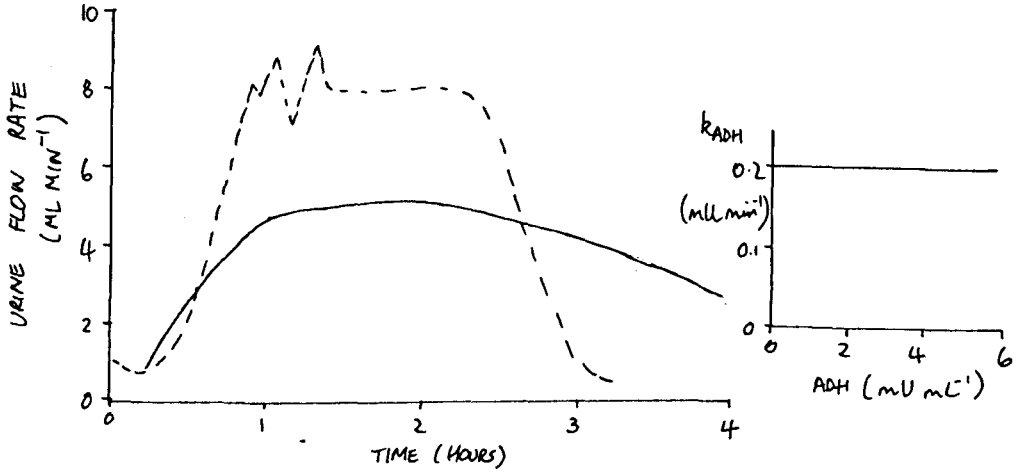


FIGURE 7.15. MODEL AND HUMAN URINE FLOW RESPONSE AFTER INTAKE OF 1 LITRE OF WATER, WITH CONSTANT ADH CLEARANCE RATE (k_{ADH})

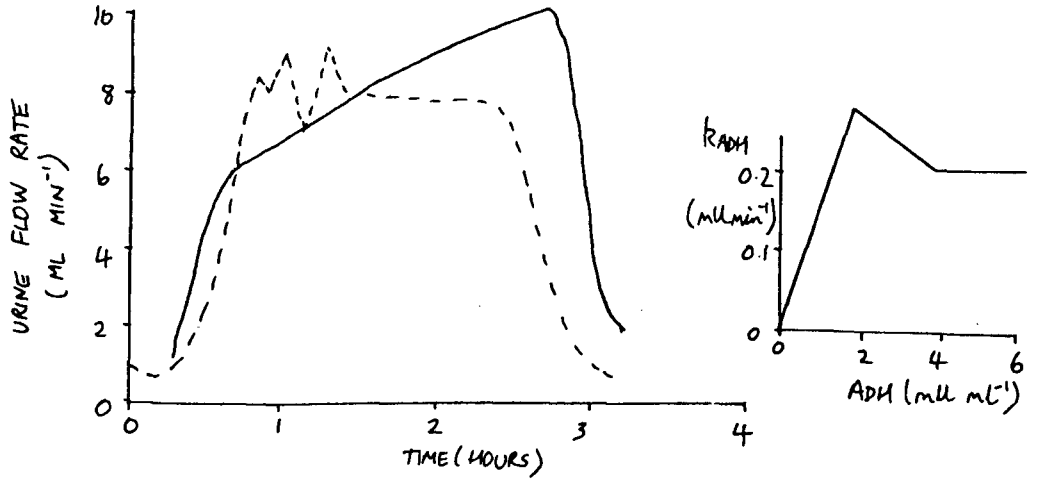


FIGURE 7.16. MODEL AND HUMAN URINE FLOW RESPONSE AFTER INTAKE OF 1 LITRE OF WATER, WITH AN ALTERNATIVE FUNCTION FOR k_{ADH}.

period of elevated urine flow. After recent experiments, the oscillatory effect in the data, which are taken from Balder and Smirk (1934), was revealed to be a peculiarity in the response of a particular patient (Uttamsingh, 1981). The close correspondence between the model and data urine flow rate responses suggests that the submodels of kidney function and ADH (which plays an essential role in controlling urine flow rate and body fluid level) are valid. The inset to Figure 7.14 depicts the form of the nonlinear clearance factor for ADH, k_{ADH}. In the hormonal submodel, the differential equation for ADH is:

$$\frac{d}{dt}(ADH) = ADHS - k_{ADH} \times ADH \quad \ldots\ldots\ldots (7.3)$$

where ADH = plasma concentration of ADH, and ADHS = net secretion rate of ADH into plasma. At low levels of ADH, k_{ADH} increases and therefore ADH is removed more rapidly from the plasma. This promotes diuresis and therefore increased urine flow in order to return to extracellular fluid volume to normal. (This is effectively a nonlinear control system for body fluid control.)

Figures 7.15 and 7.16 demonstrate the response of the model with two different functions for k_{ADH} (the difference is in the values of k_{ADH} at low ADH concentrations). In Figure 7.15, k_{ADH} is constant and results in a urine flow rate response which is much less than the human and of much longer duration, and can therefore be rejected. The k_{ADH} function in Figure 7.16 increases as ADH falls below 4 mU ml^{-1} but drops to zero below 2 mU ml^{-1}. The resulting model response is rather peaky, and not as close to the data as Figure 7.14. However, given the variation of the normal human population, it cannot be rejected so easily as Figure 7.15.

These tests provide evidence, through the overall model response, of the validity of the representation of ADH clearance in the hormonal dynamics submodel, which was identified in the theoretical validation at level 1 as an area of uncertainty (Section 7.4.1.2), as well as the empirical validity of the clinical response of the model. Another source of uncertainty in the ADH dynamics is the form of the combination of signals for the secretion of ADH. In the water load test, excess fluid volume rises and plasma osmolality falls resulting in cooperative signals for ADH secretion, which is consequently not very sensitive to the way in which the signals are combined, and hence this has little effect on the conclusions made here about ADH clearance. In the next section, the model urine flow rate response is examined following a hypertonic saline load,

when the signals for ADH secretion are non-cooperatively interacting.

7.4.2.1.2 The effect of a hypertonic saline load

A hypertonic saline load consists of a salt solution in which the concentration of salt is greater than that of plasma. Uttamsingh (1981) simulated the effect of a hypertonic saline load on a patient who had previously been deprived of water for 16 hours. This was based on the experiments of Dean and McCance (1949). The response of urine flow rate in one human and the model is illustrated in Figure 7.17. The qualitative shape, quantitative levels, and dynamic aspects match very closely. Since the model urine flow response to a hypertonic saline load is very sensitive to the form of combination of signals for ADH secretion, a number of alternative functions were tested and the results are presented in Figures 7.18 - 7.20. In Figure 7.20 the net secretion of ADH (ADHS) is simply equal to the average of that due to excess fluid volume (ADHSV) and that due to the hyperosmolality of plasma (ADHSP). With this function, the urine flow remains at a low level, and therefore must be rejected.

The model responses in Figures 7.18 and 7.19, as well as the final model (Figure 7.17), are the result of a combination function for ADHSV and ADHSP which is a parameterised function of the excess fluid volume, ECV:

$$ADHS = f(ADHSV, ADHSP, ECV, a, b, c) \quad \ldots\ldots \quad (7.4)$$

where the parameters a, b, c are weighting factors. The parameters were adjusted until the model and data responses were as close as possible. The resulting function was incorporated in the model (Figure 7.17; see Uttamsingh, 1981).

In addition, Uttamsingh (1981) simulated the model to determine the minimum value for glomerular tubular balance (i.e. the minimum fraction of glomerular filtrate which is reabsorbed in the proximal tubules) using results from a saline load test (Uttamsingh, 1981). Thus the various tests presented here allow inferences to be made on the validity of the submodels, such as the hormonal and kidney function submodels, which were previously uncertain because of the inadequacy of theory and experimental data available. However, the possibility must be admitted that other forms of the submodels might produce the same overall model response or, in other words, that the identified model is nonunique. For the scientific objectives of the model, the most important aspect is that some candidate submodels can be rejected, leaving one or a number to be advanced for further research.

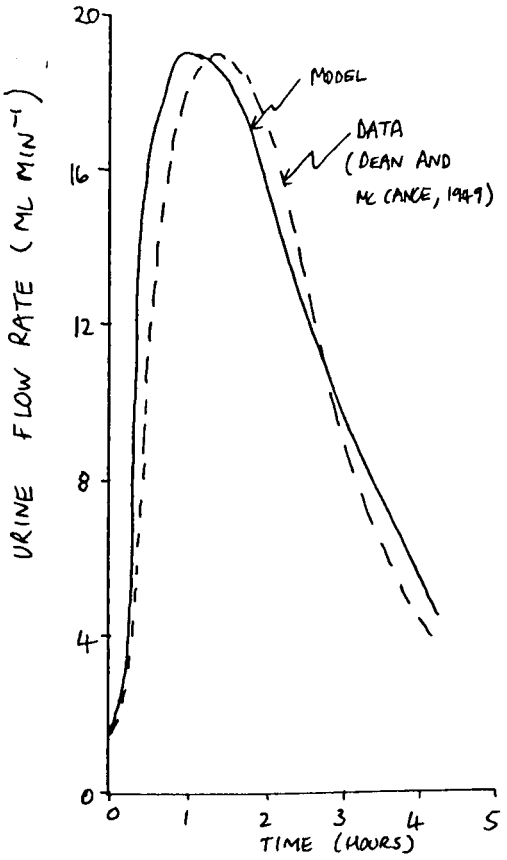


FIGURE 7.17. MODEL AND HUMAN RESPONSES TO HYPERTONIC SALINE LOAD.

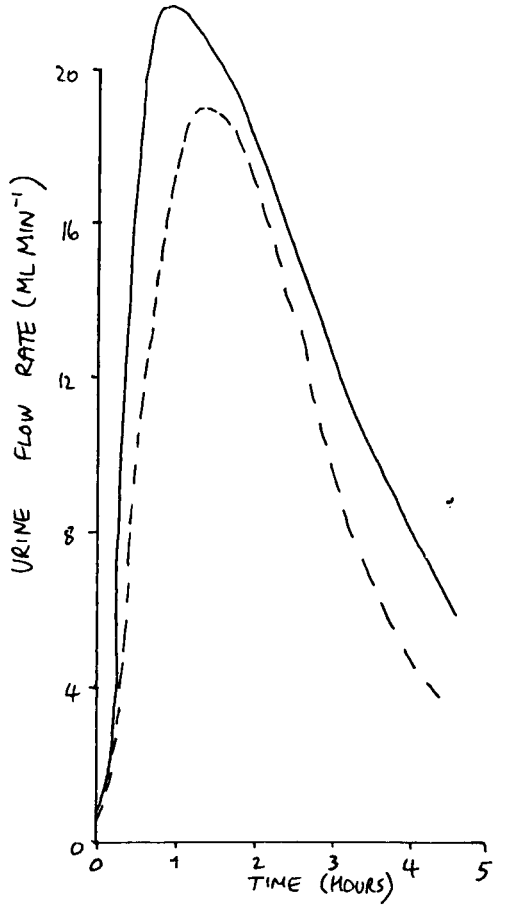


FIGURE 7.18. MODEL AND HUMAN RESPONSES TO HYPERTONIC SALINE LOAD, WITH FUNCTION (a).

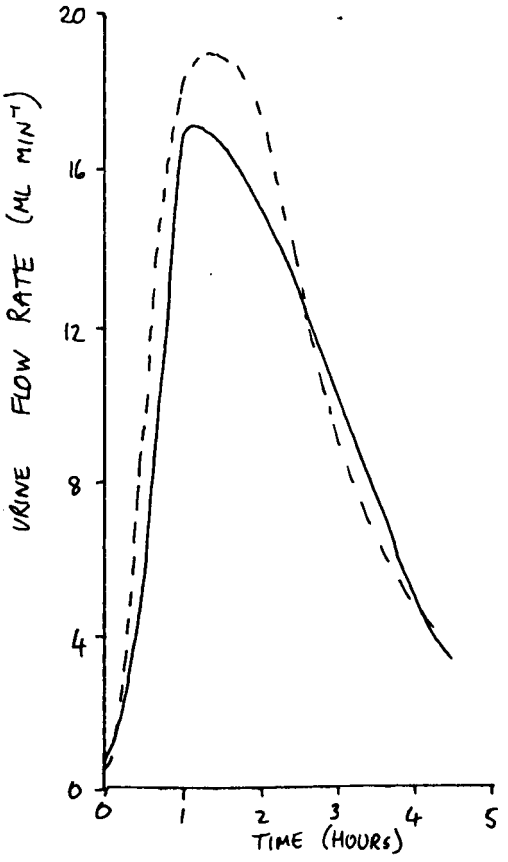


FIGURE 7.19. MODEL AND HUMAN RESPONSES TO HYPERTONIC SALINE LOAD, WITH FUNCTION (b).

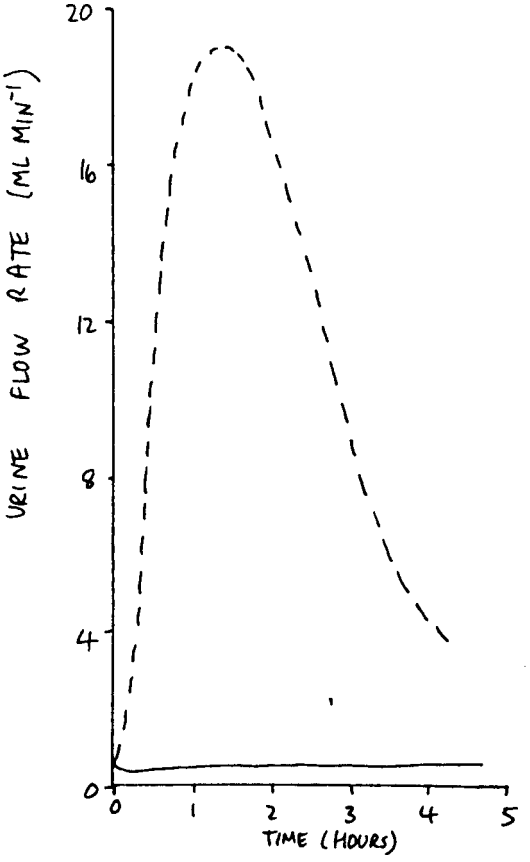


FIGURE 7.20. MODEL AND HUMAN RESPONSES TO HYPERTONIC SALINE LOAD, WITH FUNCTION (c).

For the utilitarian objectives, the empirical validity of the clinical response of the model is most important and this is confirmed in these tests for urine flow.

7.4.2.1.3 The effect of an aldosterone load

The effect of daily intramuscular injections of deoxycorticosterone acetate (a mineralocorticoid whose effects are very similar to aldosterone) was simulated in the model by multiplying the secretion rate of aldosterone by a factor of four (Uttamsingh, 1981). The results of the simulation and data, averaged from Davis and Howell (1953), are shown in Figure 7.21 for extracellular fluid volume, arterial pressure, aldosterone level, and sodium excretion rate. Aldosterone promotes the reabsorption of sodium in the distal tubules. Both model and data show the sudden drop in urine sodium excretion rate shortly after the beginning of the experiment. This indicates that the effect of aldosterone on kidney function is correctly modelled. The model also shows the gradual rise in sodium excretion rate in order to balance sodium ingestion rate as do the data.

The reduced excretion of sodium leads to increased plasma sodium concentration and osmolality. This results in an expansion of the extracellular fluid volume to counteract the hyperosmolality. In the model extracellular fluid volume rises, closely matching the data. This suggests that the control mechanisms for fluid balance and osmolality (i.e. fluid shift osmosis, and the ADH submodel) are operating correctly. Blood pressure rises (as a consequence of the increased extracellular fluid volume) in both model and data, providing confirmation of the validity of the cardiovascular submodel.

7.4.2.1.4 Comments on the results of level 3 validation

The model of a normal human with no kidney failure has exhibited acceptable responses in the above tests, and it has also been possible to infer the validity of some of the submodels (for these tests). However, the model-data comparisons have been made largely on the basis of urine flow rate responses and, with the exception of the aldosterone tests, the empirical validity of the responses of the other clinical variables remains untested. Equally, there are many other experiments which may be performed which still lie in $\mathcal{R}_I$ and against which the model may be validated (e.g. different daily ingestion rates of water and sodium, repetitions of water or saline load tests using different volume intakes, injections of other hormones, etc.). In order to determine fully the

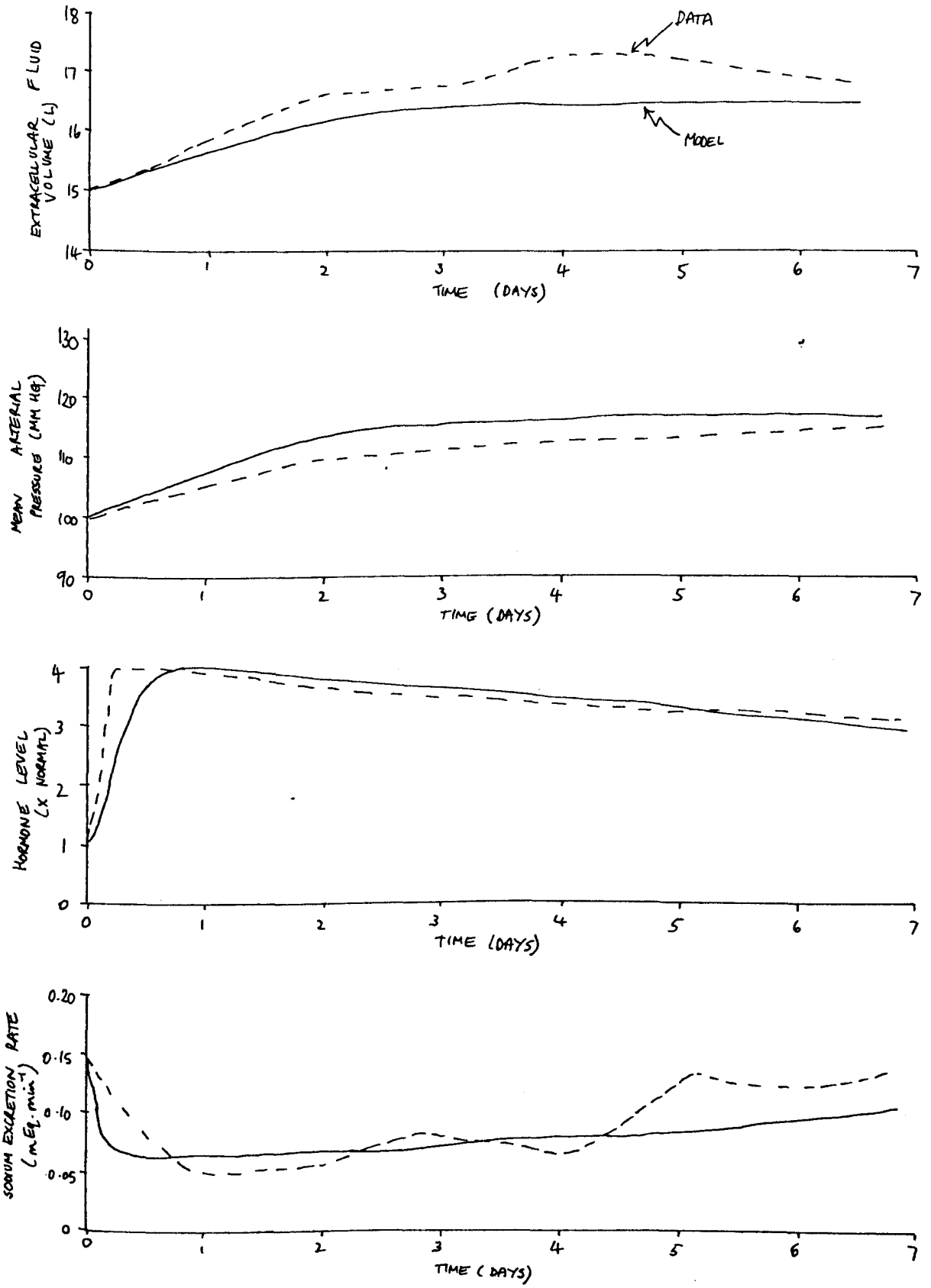


FIGURE 7.21. MODEL AND HUMAN RESPONSE TO AN ALDOSTERONE LOAD

empirical validity of the level 3 model these tests should be performed and the responses of arterial pressure, plasma-sodium, -potassium, -urea, -creatinine, -ADH, -renin, -angiotensin, and -aldosterone concentrations, fluid volume, urine flow rate and composition, and temperature should be measured, if possible, and compared with the model. When these tests are concluded it will be possible to delimit clearly the empirically valid range of application ($\mathcal{R}_V$), and also to assess more precisely the uniqueness and representational validity of the uncertain submodels (e.g. hormone dynamics and control submodels).

7.4.2.2 Empirical validation at level 4 - model of a patient with renal failure

The next level in the validation hierarchy (Figure 7.12) is concerned with the model applied to a patient with kidney failure, but not undergoing dialysis. The model is used to simulate a patient, who has acute renal failure, in the period between two successive dialyses for which data are available (i.e. at the end of the first dialysis and the beginning of the second). In order to run the model the parameter values, initial conditions, ingestion rates, etc. (with the exception of details of dialysis therapy) contained in Table 7.1 must be specified. Table 7.3 shows the model and human responses, for several clinical variables, at the beginning and end of a 17-hour period. The parameters representing the renal failure of this patient were obtained in the parameter estimation study to be described in Section 7.4.3. With the exception of plasma sodium concentration, the variables in the model change in the same direction as the data. The discrepancy in plasma sodium is small (+ 4%) although greater than the estimated measurement uncertainty ($\pm$ 1%). However, the concentrations of sodium in the model and human are both still below normal values.

As expected, both plasma urea and creatinine increase over this period of dialysis. The change in plasma urea in the model is 13% greater than the data, which is acceptable. On the other hand, plasma creatinine in the model shows a substantial increase compared with the data. This indicates that the generation rate of creatinine in the model may be too high. The value of this parameter is difficult to specify and changes with the degree of muscular activity.

The increase of extracellular fluid volume in the model is a result of the inadequate excretion of water, and leads to an increase of arterial

MODEL KIDNEY FAILURE PARAMETERS: FACT1 = 0.1, FACT2 = 0.5, FACT3 = 0.1, FACT4 = 0.0, GUREA = 0.050

TIME (HOURS)	PLASMA UREA ($g.L^{-1}$)		PLASMA CREATININE ($g.L^{-1}$)		PLASMA SODIUM ($mEq.L^{-1}$)		PLASMA POTASSIUM ($mEq.L^{-1}$)		ARTERIAL PRESSURE ($mmHg$)		EXTRACELLULAR FLUID VOLUME (L)	
	M	D	M	D	M	D	M	D	M	D	M	D
0	2.33	2.33	0.066	0.066	131.0	131.0	5.3	5.3	103.3	103.3	19.2	–
17	3.54	3.12	0.087	0.070	132.8	129.0	6.3	5.9	112.9	106.7	19.9	–

TABLE 7.3. MODEL AND HUMAN RESPONSES IN ACUTE RENAL FAILURE, OFF-DIALYSIS (PATIENT DG, UTTAMSINGH, 1981).

DETAILS OF DIALYSIS THERAPY:

CONCENTRATION OF SODIUM IN DIALYSATE = 135.0 $mEq.L^{-1}$

CONCENTRATION OF POTASSIUM IN DIALYSATE = 2.3 $mEq.L^{-1}$

TIME ON DIALYSIS = 4 HR

POST COIL PRESSURE = 90.0 MMHG.

FILTRATION RATE = 1.11 $mL.min^{-1}$

BLOOD FLOW RATE THROUGH MACHINE = 150 $mL.min^{-1}$

TIME (HOURS)	PLASMA UREA PUR ($g.L^{-1}$)		PLASMA CREATININE PCRE ($g.L^{-1}$)		PLASMA SODIUM PNA ($mEq.L^{-1}$)		PLASMA POTASSIUM PK ($mEq.L^{-1}$)		ARTERIAL PRESSURE AP ($mm.Hg$)		URINE FLOW RATE UFL ($mL.min^{-1}$)	
	M	D	M	D	M	D	M	D	M	D	M	D
0 (PRE)	2.75	2.75	0.075	0.075	128.0	128.0	6.0	6.0	97.0	97.0	0.06	0.03
4 (POST)	1.73	1.67	0.051	0.052	133.1	131.0	4.2	4.9	100.6	93.3	0.06	0.03

FACT1 = 0.05, FACT2 = 1.0, FACT3 = 0.05, FACT4 = 0.1, GUREA = 0.040

TABLE 7.4. MODEL AND HUMAN RESPONSES ACROSS A 4 HOUR PERIOD OF DIALYSIS (PATIENT DG, UTTAMSINGH, 1981).

pressure which is confirmed by the data. (Arterial pressure is affected by many factors ranging from short-term disturbances and control effects, to external changes, to longer-term adaptive changes. Consequently, only the trends in arterial pressure data, over a period of a few hours, are relevant to the empirical validation of the renal model, which does not not include the short-term or very long-term effects.)

In general, therefore, the model satisfies empirical validity criteria at level 4 in qualitative and approximate quantitative terms for the major clinical variables. An important anomaly is the response of plasma sodium concentration which is qualitatively incorrect and this should be investigated further. Over a long simulation such as this, many external factors affect the renal system. For simplicity of the model many of these are not modelled, and therefore the match between model and data may deteriorate. Nevertheless, the model has empirical validity in the prediction of general trends for a patient with renal failure in the periods between dialyses. If a more accurate prediction of the state of the patient is required, the model plasma urea concentrations may be used as a more reliable indicator.

In the next section the model is used to represent a patient undergoing dialysis therapy (level 5 in Figure 7.12). Under these conditions, where external factors are controlled, the patient is at rest, the dialysis machine has a large effect, and the time scale is shorter, it will be seen that the model can track the human response much more closely.

7.4.2.3 Empirical validation at level 5 - model of a renal failure patient during dialysis

Level 5 is the final level in the validation hierarchy (Figure 7.12) and is concerned with the validation of the model applied to a patient with renal failure who is undergoing dialysis. This level of validation is of primary importance for the specific utilitarian objectives, which require that the model should be capable of predicting the effect of dialysis on a patient. In a simple model, the validation programme might start directly at this level. However, the Uttamsingh renal model is complex and a prior validation of some of its simpler modes, or submodels, is essential. Thus the results of the previous two sections (validation at levels 3 and 4), whilst indicating the need for more comprehensive validation and possibly model modification, are extremely important. Although there were some significant differences between the model and the data (especially at level 4) these were largely associated with longer-

term controls or external factors that are not included in the model, and which are minimal over the relatively short and controlled conditions of dialysis. It is therefore legitimate to proceed to this next level of validation, although the inadequacies should be kept in mind.

In this section, the simulation of one dialysis is compared with available data in order to test that the model response is realistic and does not produce significant discrepancies. A fuller analysis of the validity at level 5 is presented in the next section (Section 7.4.3). The model is set up to simulate a patient who has acute renal failure and greatly reduced kidney function (the full details of this patient, D.G., may be found in Uttamsingh, 1981). The details of the dialysis therapy, values of the kidney failure and urea generation parameters, and the responses of the model and human clinical variables over a 4-hour dialysis are presented in Table 7.4.

The failure parameters for this model simulation were obtained by an iterative parameter estimation procedure described in the next section (for dialysis D2). The model and human response are very close, and the responses for urea and creatinine are excellent. The apparently large difference in arterial pressure is nevertheless within the normal range of variation, and arterial pressure basically shows little change in both the model and human. A more significant discrepancy is the plasma sodium response which rises more in the model than the human (the inadequacy of the plasma sodium response was also noted in Section 7.4.2.2). At first glance, therefore, the response of the model in tests at level 5 is not only realistic but appears to have predictive validity, or accuracy, as well.

In order to establish the generality of the empirical validity of the model in representing dialysis, Uttamsingh (1981) compared the model response with data from 8 dialyses on 6 patients exhibiting a wide range of clinical symptoms (e.g. acute/chronic renal failure). On the whole, the model responses were acceptable, although not as good as Table 7.4 because of the uncertainty in the initial estimates of the failure parameters. In the next section, the simulations of three dialysis on a single patient are analysed in depth. In particular, an attempt is made to quantify the empirical predictive validity of the model at level 5 and to determine the values of the kidney failure parameters using parameter estimation techniques. The results of parameter estimation also have implications for the representational validity of the model and its submodels.

7.4.3 Empirical validation at level 5 - parameter estimation and the quantification of predictive accuracy

In the case of the simulation of a single dialysis presented in the previous section, there was a close correspondence between the values of the clinical variables in the model and human at the end of the dialysis period. The model response was the result of one of the parameter estimation procedures described below, which are used to estimate the values of the renal failure parameters for an individual patient by minimising the error between model and data over the period of dialysis. The motivation for the use of parameter estimation was the lack of precision with which the renal failure parameters could be specified by the clinician, and the consequent mismatches that occurred (particularly in plasma urea concentrations) between the model and the data in the 8 dialyses simulated by Uttamsingh (1981) using the clinician-specified values.

The aims of this section are: (i) to report the results of two parameter estimation procedures applied to a single patient during three periods of dialysis; (ii) to assess the potential and acceptability of parameter estimation using the renal model; and (iii) to quantify the accuracy of model predictions over the course of dialysis and also for several hours after the end of dialysis. The third aim is related to the utilitarian objective of the model and is concerned with testing whether the model predictions are sufficiently accurate to aid in the design of dialysis therapies.

Details of the patient state and dialysis therapy for the three dialyses are given in Section 7.4.3.1. The two parameter estimation procedures are outlined in Section 7.4.3.2 and the results presented in Section 7.4.3.3. In Section 7.4.3.4 an attempt is made to quantify the predictive validity of the model. Some remarks on the theoretical and practical identifiability of the model are made in Section 7.4.3.5. Finally, in Section 7.4.3.6, some preliminary conclusions are drawn (the final conclusions, including the assessments of scientific and pragmatic validity of the model, are presented in Section 7.5). The work presented below was carried out in conjunction with Uttamsingh.

7.4.3.1 Details of patient state and dialysis therapy

The three dialyses were performed on patient DG during his period of acute renal failure. During this time the patient had only a very small urine flow rate, although he was still secreting renin. Table 7.5

gives details of the ingestion rate data, dialysis therapy, and the clinician's estimate of the failure parameters for the three dialyses. Figures 7.22 - 7.24 show the responses of arterial pressure (AP), plasma sodium concentration (PNA), plasma potassium concentration (PIC), plasma urea concentration (PUR), and plasma creatinine concentration (PCRE) over the dialyses. From the values at the start of dialysis it can be seen that the patient has very high plasma levels of urea, and low plasma levels of sodium. Because of the very high urea levels, the patient was dialysed daily. The dialyses D1, D2, and D3 are three subsequent (but not consecutive) ones taken from a period of three weeks of daily dialysis. Towards the end of this period the patient regained some kidney function, as is indicated in the failure parameters (Table 7.5). In Section 7.4.3.4 (Tables 7.10 and 7.11) the model is run on after the end of each dialysis until the start of the next dialysis and compared with the pre-dialysis data (i.e. on the days immediately following D1, D2, and D3). (For more details of the patient state and model simulations consult Uttamsingh (1981, Chapter 6) in which all the values required in Table 7.1 are given.)

The very high levels of urea in the patient are due to a high basal metabolic rate. The value of the urea generation rate parameter (GUREA) is not known to the clinician and so, in the initial model runs, the normal value was used (0.015 g min^{-1}; see Table 7.5). This resulted in a discrepancy for plasma urea responses which suggested that this patient would be an interesting subject for parameter estimation, particularly to determine his heightened urea generation rate (i.e. indirect measurement).

7.4.3.2 Parameter estimation procedures

Two procedures were used for parameter estimation: an automatic least-squares estimator, and an interactive, systematic manual procedure. These are outlined below.

7.4.3.2.1 Least-squares parameter estimation

Least-squares parameter estimation was performed using a modified version of the "GIDENT" software package for the estimation of parameters in nonlinear dynamic models with multiple outputs devised by Roberts (1977). GIDENT is written in FORTRAN IV for use on a CDC7600, and employs standard subroutines from the NAG Library. In GIDENT the minimisation of the

	PARAMETER	DIALYSIS			COMMENTS
		D1	D2	D3	
CLINICIAN'S ESTIMATE OF RENAL FAILURE PARAMETERS	FACT1	0.0	0.0	0.37	(THESE PARAMETERS HAVE BEEN DEFINED PREVIOUSLY).
	FACT2	0.87	0.87	0.87	
	FACT3	0.0	0.0	0.62	
	FACT4	0.0	0.0	0.12	
	GUREA	0.015	0.015	0.015	UREA GENERATION RATE ($g.min^{-1}$)
INGESTION RATES	FLUDDAY	0.6	0.75	2.7	WATER INTAKE/DAY (L/day)
	SODDAY	190.0	50.0	120.0	SODIUM INTAKE/DAY ($mEq.day^{-1}$)
	POTDAY	40.0	50.0	60.0	POTASSIUM INTAKE/DAY ($mEq.day^{-1}$)
DIALYSIS THERAPY	SODDIA	135.0	135.0	139.0	CONCENCENTRATION OF SODIUM IN DIALYSATE ($mEq.L^{-1}$)
	POTDIA	2.3	2.3	2.3	CONCENTRATION OF POTASSIUM IN DIALYSATE ($mEq.L^{-1}$)
	PCPR	30.0	90.0	100.0	POST COIL PRESSURE ($mm\ Hg$)
	QB	150.0	200.0	150.0	BLOOD FLOW RATE THROUGH MACHINE ($ml.min^{-1}$)
	T	4.0	4.0	4.0	TIME ON DIALYSIS ($hr.$)

(IN D3 THE MACHINE WAS PRIMED WITH 0.4L BLOOD).

TABLE 7.5. ESTIMATES AND INITIAL VALUES OF RENAL FAILURE PARAMETERS, INGESTION RATE DATA, AND DETAILS OF DIALYSIS THERAPY FOR PATIENT DG FOR THREE DIALYSES.

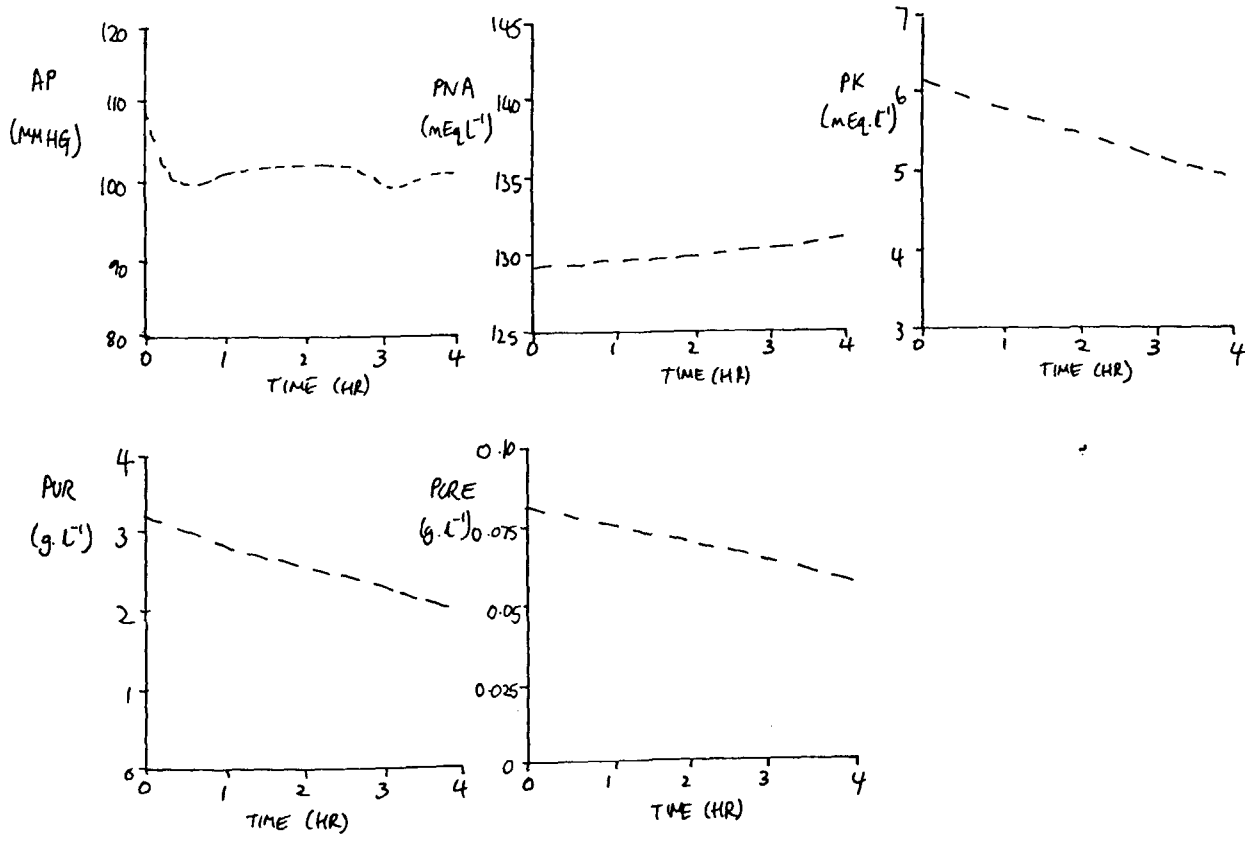


FIGURE 7.22. CLINICAL RESPONSE DATA FOR PATIENT DG OVER DIALYSIS D1

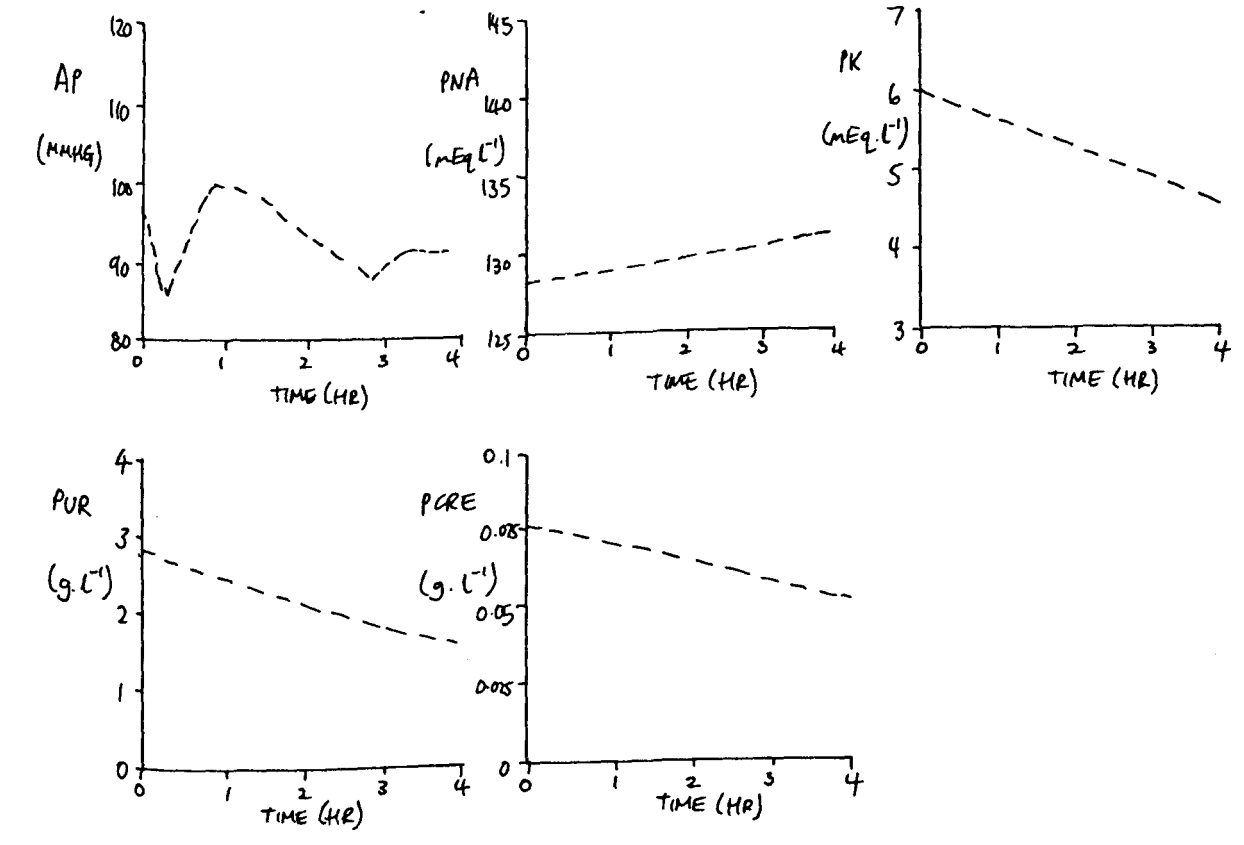


FIGURE 7.23. CLINICAL RESPONSE DATA FOR PATIENT DG, OVER DIALYSIS D2

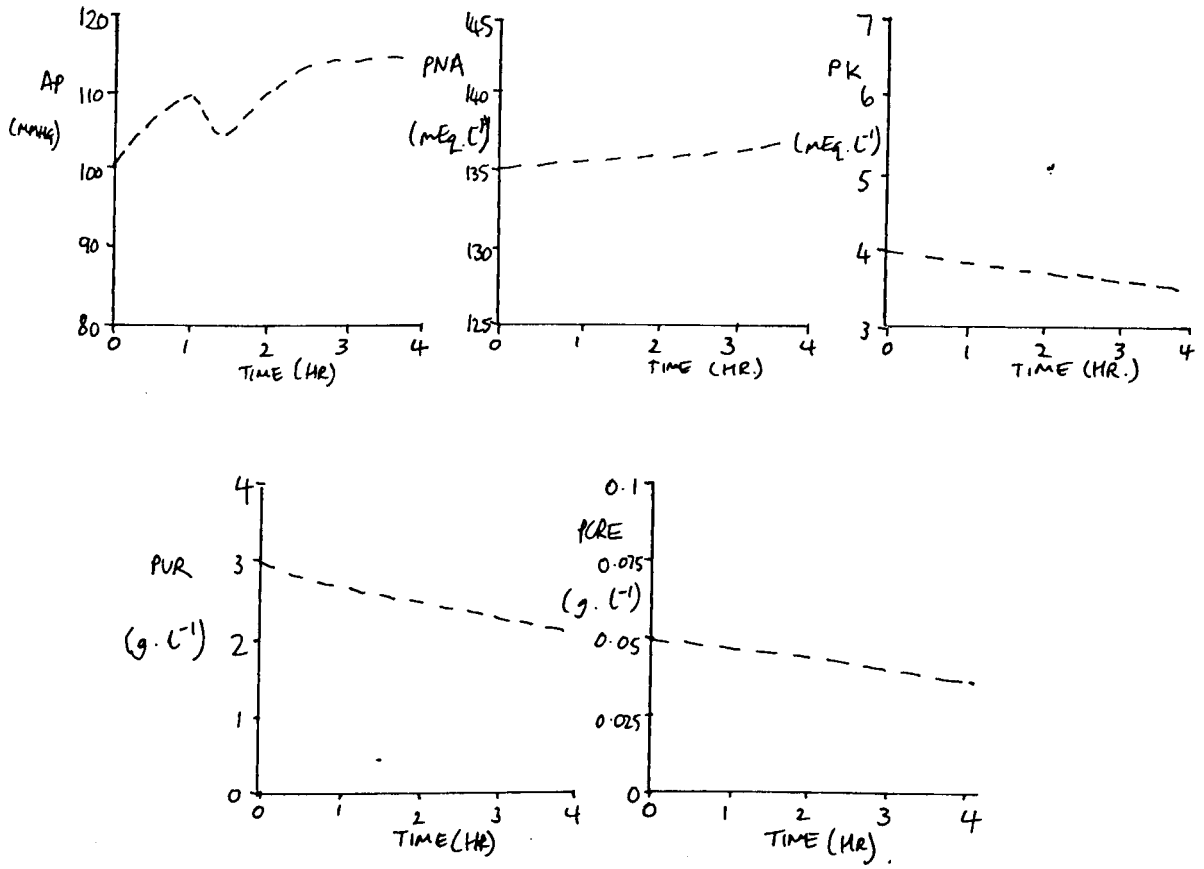


FIGURE 7.24 CLINICAL RESPONSE DATA FOR PATIENT DG OVER DIALYSIS D3.

weighted sum of squared residuals is conducted using a Simplex optimisation algorithm to search parameter space. The integrator was replaced with a first-order Euler integrator which proved to be more stable for the renal model, as well as faster in computer time. The renal model was inserted as a subroutine which was called by GIDENT for integration over the period of dialysis for a specified set of parameter values at each iteration.

The outputs and parameters used in the optimisation are shown in Table 7.6. The model variables PUR and PCRE were multiplied by appropriate factors so that their individual contributions to the weighted sum of squared errors were of the same order of magnitude as that of arterial pressure. The Simplex optimiser works best with parameters normalised to unity, and so the model parameters were redefined for the estimator, as indicated in Table 7.6. For simplicity, only three parameters (FACT1, FACT4, and GUREA) were estimated. (In an earlier sensitivity analysis the model response was found to be more sensitive to these parameters than to FACT2 and FACT3.) Furthermore, FACT4 and GUREA are directly related to the dynamics of plasma urea in the model which it was hoped to improve. The choice of outputs was dictated in a similar manner. Arterial pressure was included as an output as data were available for this variable every 30 minutes during dialysis. Plasma urea and creatinine responses in the patient during the dialysis were obtained by linear interpolation between the start and finish values (see Figures 7.22 - 7.24).

The error terms for each variable were evaluated by comparing the model outputs with the data (based on Figures 7.22 - 7.24) every 30 minutes for the 4 hours on dialysis (i.e. 9 sample points). The squared errors were multiplied by time-dependent weighting factors (see Table 7.6) and then added to form the total sum of squared errors. The error residuals on the arterial pressure responses were much larger than those on plasma urea and creatinine, and the weighting factors were appropriately reduced. Because of the uncertainty of the linear interpolation of the urea and creatinine data, the intermediate weighting factors for the errors on these variables were set to half the initial and final values (see Table 7.6).

The initial parameter values for each dialysis were taken as normal (i.e. FACT1 = 1.0, FACT4 = 1.0, and GUREA = 0.015). (In GIDENT, there are also facilities for sensitivity analysis after the location of the least

		DEFINITION	TYPICAL VALUES	TIME - DEPENDENT WEIGHTINGS	
OUTPUTS	Y_1	AP	100	$(0.2, 0.1, \ldots, 0.1, 0.2)$	
	Y_2	50 PUR	150	$(1.0, 0.5, \ldots, 0.5, 1.0)$	
	Y_3	1000 PCRE	70	$(1.0, 0.5, \ldots, 0.5, 1.0)$	
PARAMETERS	P_1	1 - FACT1	0.9	0.0	INITIAL VALUES (NORMAL PATIENT).
	P_2	1 - FACT4	0.9	0.0	
	P_3	10 (1 - UREA)	0.85	0.85	

TABLE 7.6. OUTPUTS AND PARAMETERS USED IN THE LEAST SQUARES ESTIMATOR.

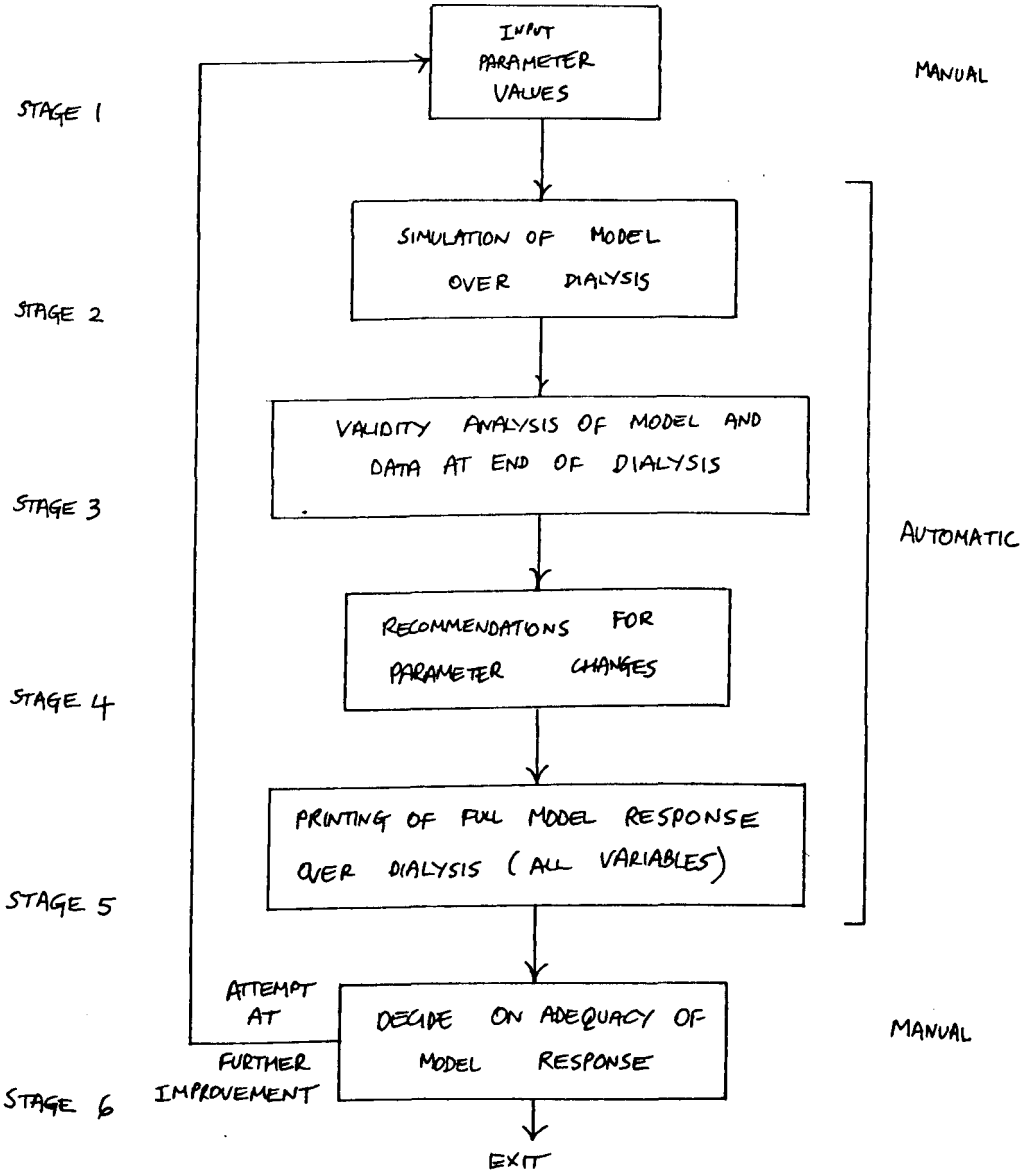


FIGURE 6.25. SYSTEMATIC PROCEDURE FOR MANUAL PARAMETER ESTIMATION

squared error has been found. The results of the sensitivity analysis are then used to evaluate the variances of the parameter estimates using a student t distribution. This was not used, however, since there were systematic errors in the residuals of the renal model, whereas the residuals must be random with zero mean for the analysis to be valid. For the full technical details and user specifications of GIDENT refer to the manual; Roberts, 1977.)

In setting up GIDENT for use with the renal model, some simple considerations of identifiability were made. These are reviewed in Section 7.4.3.5 but, briefly, they indicated that plasma creatinine should be included as an output in order that FACT4 and GUREA are uniquely identifiable. Some comments on the usability of GIDENT with the renal model are made in the presentation of results (Section 7.4.3.3).

7.4.3.2.2 Interactive systematic manual parameter estimation

Because of the difficulties associated with using an automatic parameter estimation algorithm with a complex model, an alternative systematic manual estimation procedure was devised based on an interactive version of the model and run on a PRIME 550. In this procedure, all five failure parameters (FACT1 to FACT4 and GUREA) are set by the user. Figure 7.25 depicts the six stages of the procedure. Of these stages, only the first and the last are manual, the rest are automatically performed by the computer and the results displayed at the terminal. Each stage will be outlined, beginning with stage 2.

In stage 2, the model is simulated over the 4-hour dialysis period using the parameters specified in stage 1. The values of the main clinical variables (AP, PNA, PK, PUR, and PCRE) in the model, at the end of dialysis are compared with the data in stage 3. If the error is less than the uncertainty interval of the data, this information is displayed. In stage 4, the errors on the model variables are analysed and compared with a set of rules for determining the directions of parameter changes. These rules were elaborated in advance using sensitivity analysis. The recommendations for parameter changes are then displayed. To aid in the choice of a new parameter set, the full model response at 30-minute intervals over the dialysis is printed in stage 5.

On the basis of the computer-supplied information, from the above stages, and other physiological data or knowledge, the user decides whether the model response is adequate or in need of further improvement

(stage 6). In the latter case, a new set of parameter values may be input (stage 1). The advantages of this method are that the user may take into account any additional information that is available and pertinent, and also that the process of determining the directions for parameter changes is done automatically and can highlight contradictory discrepancies between the model and the data. An example of the use of the procedure for dialysis D1 is shown in Figure 7.26 (for simplicity the output of the full model response has been omitted). The first parameter change (i.e. to (2)) follows the recommendations exactly; however, the second change (to (3)) does not increase FACT4 as required. This is because FACT1 is set to zero, implying zero urine flow rate, and therefore no urea or creatinine excretion (i.e. FACT4 = 0.0). The urine flow rate in the human is 0.15 ml min^{-1} which corresponds closely to simulation (2). Thus the decision may be made to accept the parameters of (2) even though the fit of (3) is slightly better.

7.4.3.3 Presentation of results

For each of the three dialyses, four sets of results will be presented: (i) the model response using normal parameter values (denoted by β_N); (ii) the model response using the parameter values given by the clinician (β_C); (iii) the model response using optimal values from the least-squares estimator (β^*_{LSE}); and (iv) the model response using final values from the systematic manual procedure (β^*_{SYS}). In each case the responses will be compared with the post-dialysis data.

The least square estimate was very difficult to set up and proved very sensitive to the choice of time-dependent weightings. Because the best values for the parameters were often at the physiological constraints it was difficult to halt the estimator. The main difficulty, however, was the location by the algorithm of a very high FACT4 and GUREA mode, with FACT1 = 0. Theoretically, the model is uniquely identifiable if both PUR and PCRE are included as outputs (see Section 7.4.3.5), but in practice the PCRE response is not as sensitive to FACT4 as is the PUR response, and this can make the model practically unidentifiable. The high GUREA mode is unfeasible physiologically, since no urea may be excreted if FACT1 = 0. The problem was avoided by increasing the weightings of PCRE compared to PUR, but this reduced the accuracy of the fit to the urea data, of course.

The problem with the systematic manual estimation procedure was

(1) FACT1 = 1.000
FACT2 = 1.000
FACT3 = 1.000
FACT4 = 1.000
GUREA = 0.015

Per-dialysis results: data and simulation

	Pre	Post	
	Data	Data	Model
PNA	128.000	131.000	133.861
PK	6.000	4.900	3.963
PUR	2.754	1.670	0.840
PCRE	0.075	0.052	0.045
AP	97.000	93.300	97.852
CTEMP	36.000	0.000	36.545
STEMP	34.000	0.000	34.842

Validity analysis of model response

AP lies in uncertainty interval of data

Advice on parameter changes

Try decreasing FACT1 ← PNA
Try decreasing FACT3 ← PK
Try decreasing FACT4 ← PUR
Try increasing GUREA ← PUR
Try decreasing FACT4 ← PCRE

(Output of full response, UFL = 1.25 at t = 4 hr)

(2) FACT1 = 0.100
FACT2 = 0.500
FACT3 = 0.100
FACT4 = 0.100
GUREA = 0.060

(Continued)

Figure 7.26: Example of Systematic Manual Parameter Estimation applied to Dialysis D1

Per-dialysis results: data and simulation

	Pre	Post	
	Data	Data	Model
PNA	128.000	131.000	133.190
PK	6.000	4.900	4.173
PUR	2.754	1.670	1.813
PCRE	0.075	0.052	0.048
AP	97.000	93.300	98.512
CTEMP	36.000	0.000	36.543
STEMP	34.000	0.000	34.844

Validity analysis of model response

PCRE lies in uncertainty interval of data

Advice on parameter changes

Try decreasing FACT1 ← PNA
Try decreasing FACT3 ← PK
Try increasing FACT4 ← PUR
Try decreasing GUREA ← PUR
Try decreasing FACT1 ← AP
Try decreasing FACT2 ← AP

(Output of full response. Variable of interest, UFL = 0.13 at t = 4 hr)

(3) FACT1 = 0.000
FACT2 = 0.250
FACT3 = 0.000
FACT4 = 0.000
GUREA = 0.050

Per-dialysis results: data and simulation

	Pre	Post	
	Data	Data	Model
PNA	128.000	131.000	133.113
PK	6.000	4.900	4.213
PUR	2.754	1.670	1.664
PCRE	0.075	0.052	0.048
AP	97.000	93.300	97.368
CTEMP	36.000	0.000	36.549
STEMP	34.000	0.000	34.843

(Continued)

Figure 7.26 (cont.)

Validity analysis of model response

PUR lies in uncertainty interval of data

PCRE lies in uncertainty interval of data

AP lies in uncertainty interval of data

Advice on parameter changes

Try decreasing FACT1 ← PNA
Try decreasing FACT3 ← PK

(Output of full response, UFL = 0.0 at t = 4 hr)

Figure 7.26 (continued)

that it did not always yield the same results, although mostly they were quite close. This depended on what factors were considered to be more important at the particular time. In part it reflects the lack of sensitivity of the model response to some parameters (especially FACT2 and FACT3), and hence a genuine uncertainty. This suggests that more information may be needed in order to determine precisely all the parameters. However, for predictive purposes, a rough estimate of the less sensitive parameters will be satisfactory. The results presented below for the systematic procedure are those which were consistently obtained a number of times.

Tables 7.7, 7.8, and 7.9 show the responses across dialyses D1, D2, and D3 of the patient and the model (with the four sets of parameter values), including the qualitative changes. Firstly, the estimated parameter values will be discussed and, secondly, the various responses will be analysed. A quantitative assessment of the accuracy of the model with the different parameter values is made in Section 7.4.3.4.

In general, the estimated parameters have a high value of GUREA indicating the high metabolic rate of the patient (the urea generation rate is about 3 to 4 times normal at approximately 0.05 g min^{-1}). In the LSE estimation of D3, however, GUREA remains at the normal level. This is because FACT1 and FACT4 are estimated as zero, whereas the patient has had some real recovery of renal function and therefore excretion of urea. Since no urea is being excreted in the model, GUREA remains low to compensate. The LSE parameters (β^*_{LSE}) show no improvement of kidney function over the dialyses, demonstrating that insufficient information is contained in the outputs AP, PUR, and PCRE. On the other hand, the systematic estimates (β^*_{SYS}) show a slight improvement in kidney function over the three dialyses (FACT1 goes from 0.1 to 0.05 to 0.15), although the improvement is not as great as estimated by the clinician.

In all three dialyses, there is close tracking of PUR by the model with β^*_{LSE} or β^*_{SYS}. This is good evidence of the validity of the urea dynamics submodel, and the modelling of the removal of urea through the artificial kidney machine. Usually, the model response for PUR (with β^*_{LSE} or β^*_{SYS}) is more accurate than the response with clinician parameters (β_C). The PCRE response is significantly improved by estimation compared with the normal response, but shows little change between the responses with β_C and those with β^*_{LSE} or β^*_{SYS}. The large systematic error in PCRE in D1 does not occur in D2 and D3 (and therefore the generation

INITIAL VALUES IN HUMAN AND MODEL (T = 0 HR)					
PUR ($g\ \ell^{-1}$)	PCRE ($g\ \ell^{-1}$)	PNA ($mEq\ \ell^{-1}$)	PK ($mEq\ \ell^{-1}$)	AP (mm Hg)	UFL ($m\ell\ min^{-1}$)
3.17	0.078	129.0	6.2	110.0	0.15

Source	FAILURE PARAMETERS				
	FACT1	FACT2	FACT3	FACT4	GUREA
β_N	1.0	1.0	1.0	1.0	0.015
β_C	0.0	0.87	0.0	0.0	0.015
β^*_{LSE}	0.0	0.87	0.0	0.0	0.055
β^*_{SYS}	0.1	0.5	0.1	0.0	0.050

POST-DIALYSIS VALUES (T = 4 HR)						
Source	PUR	PCRE	PNA	PK	AP	UFL
DATA	↓ 2.33	↓ 0.066	↑ 131.0	↓ 5.3	↓ 103.3	0.15
β_N	↓ 1.04	↓ 0.051	↑ 134.1	↓ 4.1	↑ 112.0	1.25
β_C	↓ 2.12	↓ 0.055	↑ 133.3	↓ 4.4	↑ 110.8	0.0
β^*_{LSE}	↓ 2.28	↓ 0.055	↑ 133.3	↓ 4.4	↑ 110.8	0.0
β^*_{SYS}	↓ 2.27	↓ 0.055	↑ 133.4	↓ 4.4	↑ 112.6	0.13

Table 7.7: Patient and Model Variables before and after Dialysis D1 for Different Parameter Values

INITIAL VALUES IN PATIENT AND MODEL (T = 0 HR)					
PUR ($g\ \ell^{-1}$)	PCRE ($g\ \ell^{-1}$)	PNA ($mEq\ \ell^{-1}$)	PK ($mEq\ \ell^{-1}$)	AP (mm Hg)	UFL ($m\ell\ min^{-1}$)
2.75	0.075	128.0	6.0	97.0	0.03

Source	FAILURE PARAMETERS				
	FACT1	FACT2	FACT3	FACT4	GUREA
β_N	1.0	1.0	1.0	1.0	0.015
β_C	0.0	0.87	0.0	0.0	0.015
β^*_{LSE}	0.0	0.87	0.0	0.0	0.056
β^*_{SYS}	0.05	1.0	0.05	0.1	0.040

POST-DIALYSIS VALUES (T = 4 HR)						
Source	PUR	PCRE	PNA	PK	AP	UFL
DATA	↓ 1.67	↓ 0.052	↑ 131.0	↓ 4.9	↓ 93.3	0.03
β_N	↓ 0.84	↓ 0.048	↑ 133.9	↓ 4.0	↑ 97.9	1.18
β_C	↓ 1.76	↓ 0.051	↑ 133.1	↓ 4.2	↑ 97.4	0.0
β^*_{LSE}	↓ 1.93	↓ 0.051	↑ 133.1	↓ 4.2	↑ 97.4	0.0
β^*_{SYS}	↓ 1.73	↓ 0.050	↑ 133.1	↓ 4.2	↑ 100.6	0.06

Table 7.8: Values of Patient and Model Variables before and after Dialysis D2 for Different Parameter Values

INITIAL VALUES IN PATIENT AND MODEL (T = 0 HR)					
PUR ($g\ \ell^{-1}$)	PCRE ($g\ \ell^{-1}$)	PNA ($mEq\ \ell^{-1}$)	PK ($mEq\ \ell^{-1}$)	AP (mm Hg)	UFL ($m\ell\ min^{-1}$)
3.0	0.05	135.0	4.1	101.0	0.71

Source	FAILURE PARAMETERS				
	FACT1	FACT2	FACT3	FACT4	GUREA
β_N	1.0	1.0	1.0	1.0	0.015
β_C	0.37	0.87	0.62	0.12	0.015
β^*_{LSE}	0.0	0.87	0.0	0.0	0.015
β^*_{SYS}	0.15	0.5	0.1	0.05	0.050

POST-DIALYSIS VALUES (T = 4 HR)						
Source	PUR	PCRE	PNA	PK	AP	UFL
DATA	↓ 2.27	↓ 0.040	↑ 137.0	↓ 3.5	↑ 115.0	0.71
β_N	↓ 0.90	↓ 0.034	↑ 140.0	↓ 3.5	↑ 115.0	5.90
β_C	↓ 1.75	↓ 0.036	↑ 138.4	↓ 3.5	↑ 119.2	2.16
β^*_{LSE}	↓ 1.92	↓ 0.036	↑ 137.7	↓ 3.4	↑ 118.9	0.0
β^*_{SYS}	↓ 2.0	↓ 0.036	↑ 137.9	↓ 3.4	↑ 119.2	0.8

Table 7.9: Values of Patient and Model Variables before and after Dialysis D3 for Different Parameter Values

rate of creatinine may be assumed to be normal). As PCRE levels are low, it is quite likely that measurement uncertainty on this variable is higher than the $\pm$ 2% indicated in Table 7.2.

Plasma sodium (PNA) and potassium (PK) concentrations in the model during D1 and D2 show greater changes across dialysis than in the data, but the error is reduced in the model with β_C, β^*_{LSE}, or β^*_{SYS} compared with the model with β_N. In D3, the model tracks PNA and PK closely with both sets of estimated parameters (β^*_{LSE} and β^*_{SYS}). By this time, the electrolyte concentrations in the patient had returned to normal levels, whereas in D1 and D2 PNA was low and PK was high. This may point to an over-compensatory action of the model in restoring electrolyte balance.

In the responses of arterial pressure (AP) across the dialyses, the data show a slight fall ($\simeq$ 5 mm Hg) in D1 and D2, whereas the model exhibits little change. Given the variability of arterial pressure in the human, this is an adequate fit. Furthermore, a small improvement is made in reducing the error in both β^*_{LSE} and β^*_{SYS}. In D3, both the patient and model show a significant rise in arterial pressure (of 15 - 20 mm Hg) over dialysis, which is due to the priming of the artificial kidney machine with 0.4 ℓ of blood (see Table 7.5).

The data for the urine flow rate (UFL) response (ml min^{-1}) were obtained from daily averages of the patient's urinary output. Although this is a very rough estimate of the short-term urine flow rate which varies over the day, it is better than no estimate at all. From Tables 7.7 to 7.9 it can be seen that in the model with β_N, UFL is at least one order of magnitude too high. In D1 and D2, both β_C and β^*_{LSE} produce zero UFL,andin D3 β^*_{LSE} still gives zero UFL, whereas β_C results in an over-estimate of UFL. The responses of UFL for the model with β^*_{SYS} are very close to the averaged data, and this is a consequence of the comparison of this variable in the estimation procedure (Section 7.4.3.2.2). This must be taken as additional confirming evidence of the estimated values of β^*_{SYS}, in particular the value of FACT1 which is directly related to UFL. Equally, it may be taken as evidence for the ability of the renal model to represent an individual patient on dialysis or, in other words, for the empirical predictive validity of the model at level 5 (in Figure 7.12).

The preliminary analysis suggests, therefore, that the model is a reasonably accurate representation of a patient undergoing dialysis. Since the kidney function of the patient changes over the three dialyses,

the results imply a wider generality for the representational validity of the model to other patients with renal failure. The accuracy of model responses (especially PUR) was clearly improved using parameter estimation, and this improvement is assessed quantitatively in the next section.

7.4.3.4 Quantification of predictive accuracy

Although the results of Section 7.4.3.3 are initially promising, it is not easy to draw definite conclusions from the individual results presented in Tables 7.7, 7.8, and 7.9. There is need, therefore, for some kind of overall objective measure of the predictive accuracy of the model. This will provide a summary with which to assess whether any improvements can be had from parameter estimation, and also whether the model has adequate predictive validity (at level 5) to satisfy its specific utilitarian objectives (as a predictive tool in the health-care of renal dialysis patients). To achieve this, the following "figure of merit", F, is defined:

$$F = \frac{1}{1 + \frac{1}{n} \sum_{i=1}^{n} \delta_i} \qquad \ldots\ldots (7.5)$$

where

$$\delta_i = \left| \frac{x_i(M) - x_i(D)}{x_i(D)} \right| \qquad \ldots\ldots\ldots (7.6)$$

where x_i is the variable which is compared between the data (D) and model (M). δ_i is the fractional error. If there is no difference between the model and data, F = 1; if there is an average 100% fractional difference, F = 0.5. The relationship between F and the average error is shown in Table 7.12 (the actual definition used for F is not critical, but it is important to calibrate it in terms of an average error, or uncertainty; see Section 5.2.2.1).

The figures of merit for the model with the four different sets of parameters (β_N, β_C, β^*_{LSE}, and β^*_{SYS}) were evaluated for the post-dialysis results for D1, D2, and D3. The model simulations for D1 and D3 were run on to the next day and compared with data that were available at the start of the next dialysis (see Tables 7.10 and 7.11). Two figures of merit were used. The first was based on the four plasma biochemical variables (PUR, PCRE, PNA, and PK), and the second on PUR and PCRE alone

PATIENT AND MODEL VALUES 17 HRS AFTER END OF D1						
Source	PUR ($g\ \ell^{-1}$)	PCRE ($g\ \ell^{-1}$)	PNA ($mEq\ \ell^{-1}$)	PK ($mEq\ \ell^{-1}$)	AP (mm Hg)	UFL ($m\ell\ min^{-1}$)
DATA	↑ 3.12	↑ 0.070	↓ 128.0	↑ 5.9	↑ 106.7	0.05
β_N	↓ 0.16	↑ 0.063	↑ 160.0	↑ 4.7	↓ 75.8	0.55
β_C	↑ 2.52	↑ 0.076	↑ 134.7	↑ 5.7	↑ 121.3	0.0
β^*_{LSE}	↑ 3.61	↑ 0.076	↑ 134.7	↑ 5.7	↑ 121.3	0.0
β^*_{SYS}	↑ 3.48	↑ 0.076	↑ 135.2	↑ 5.4	↑ 122.2	0.13

(For values of β_N, β_C, β^*_{LSE}, and β^*_{SYS}, see Table 7.6)

Table 7.10: Values of Patient and Model Variables 17 hours after end of Dialysis D1 for Different Parameter Values

PATIENT AND MODEL VALUES 24 HRS AFTER END OF D3						
Source	PUR ($g\ \ell^{-1}$)	PCRE ($g\ \ell^{-1}$)	PNA ($mEq\ \ell^{-1}$)	PK ($mEq\ \ell^{-1}$)	AP (mm Hg)	UFL ($m\ell\ min^{-1}$)
DATA	↑ 2.52	↓ 0.038	↑ 142.0	↑ 3.6	↓ 100.0	1.14
β_N	↓ 0.11	↑ 0.045	↑ 141.3	↑ 4.8	↓ 111.6	1.8
β_C	↓ 1.42	↑ 0.065	↑ 142.0	↑ 5.0	↑ 120.1	1.63
β^*_{LSE}	↑ 2.38	↑ 0.063	↓ 129.6	↑ 6.6	↑ 143.7	0.0
β^*_{SYS}	↑ 3.06	↑ 0.062	↓ 132.3	↑ 4.8	↑ 139.6	0.19

(For values of β_N, β_C, β^*_{LSE}, and β^*_{SYS}, see Table 7.8)

Table 7.11: Values of Patient and Model Variables 24 hours after end of Dialysis D3 for Different Parameter Values

(i.e. concerned with the accuracy of the urea and creatinine dynamics). Arterial pressure was not included because of its greater variability. The data for urine flow are daily averages and so this was also excluded.

The results are shown in Tables 7.13 and 7.14. It can be seen that there is a more or less general improvement in the F values associated with the estimated parameters (β^*_{LSE} and β^*_{SYS}). The F values for the longer-term prediction in D1 and D3 are less than those post-dialysis, i.e. the uncertainty of the predictions increases with time, as expected. For the four variables together, the F values for β^*_{LSE} and β^*_{SYS} post-dialysis are greater than or equal to 0.914, corresponding to an average error of less than 10%. The predictions of urea and creatinine alone (Table 7.14) over dialysis for β^*_{LSE} and β^*_{SYS} are not quite so accurate, falling to F = 0.887 (an average error of 13%) which is still very satisfactory.

The longer-term predictions of the model are clearly not so accurate, particularly for the model with β_N and β_C (in D3, Table 7.14, the 24-hour predictions of urea and creatinine have an average error > 50%). For D1, however, the 17-hour predictions of urea and creatinine with the estimated parameters, β^*_{LSE} and β^*_{SYS}, have a maximum average error of 12% which is very good. The deterioration in longer-term predictive accuracy in D3 is probably due to the uncertainty of the degree to which kidney function has been restored, whereas in D1 the rate of excretion of fluid is still very low. When the model is used to represent a patient with some renal function, more accurate data on urine flow and composition will lead to better parameter estimates, and hence improved predictive accuracy.

The improvement of predictive accuracy of the model using parameter estimation compared with using the clinician specified parameters is marginal. However, the results indicate that parameter estimation may be used for this marginal improvement, to confirm or improve the clinician specified parameters, to estimate parameters when they are not known in advance, and, finally, to establish the representational validity of the model.

7.4.3.5 Notes on model identifiability - theoretical and practical

As well as its implications for representational model validity, parameter estimation was used to try to improve the tracking of plasma urea concentration (PUR). This is a very important variable in the human since, although non-toxic, it is a primary waste product and directly

F	AVERAGE ERROR (%)
0.952	± 5
0.909	± 10
0.833	± 20
0.769	± 30
0.714	± 40
0.667	± 50

Table 7.12: Relationship between Figure of Merit (F) and Average Error

SOURCE	D1		D2	D3	
	POST-DIALYSIS	17 HOURS	POST-DIALYSIS	POST-DIALYSIS	24 HOURS
β_N	0.795	0.727	0.837	0.838	0.730
β_C	0.900	0.916	0.932	0.922	0.722
β^*_{LSE}	0.914	0.850	0.923	0.932	0.701
β^*_{SYS}	0.914	0.921	0.950	0.940	0.762

Table 7.13: Figures of Merit for the three Dialyses, based on Errors on PUR, PCRE, PNA and PK

SOURCE	D1		D2	D3	
	POST-DIALYSIS	17 HOURS	POST-DIALYSIS	POST-DIALYSIS	24 HOURS
β_N	0.719	0.656	0.777	0.726	0.637
β_C	0.886	0.878	0.939	0.859	0.635
β^*_{LSE}	0.914	0.891	0.920	0.887	0.737
β^*_{SYS}	0.912	0.909	0.973	0.901	0.703

Table 7.14: Figures of Merit for the Three Dialyses, based on Errors on PUR and PCRE

related to other highly toxic nitrogenous substances that must be kept at a low concentration in the blood. In patients with renal failure who have a high basal rate of metabolism (such as DG above) the need therefore arises to perform frequent dialysis to prevent these substances, and hence urea, from rising too high. For this reason, PUR was used as an output and the parameters FACT4 (remaining fractional ability to excrete urea and creatinine) and GUREA (generation rate of urea) were included. The other variables were arterial pressure (AP) and plasma creatinine (PCRE), and the other parameter was FACT1 (remaining fractional ability to excrete water and sodium).

Before undertaking the parameter estimation, some elementary considerations of identifiability were made to see if the estimation problem had been well posed. It was assumed, firstly, that the parameter FACT1 would affect all variables, in particular AP, and therefore that it was uniquely identifiable theoretically. To assess the identifiability of FACT4 and GUREA, the urea and creatinine dynamics submodel was decoupled from the model and some simplifying assumptions were made. Figure 7.27 shows a diagram of the full compartmental structure of the urea and creatinine dynamics submodel. It is coupled to the rest of the model through the values of intracellular (ICFV) and extracellular (ECFV) fluid volumes. It is assumed, for the purposes of the analysis, that the changes in these volumes are much less than those that take place in the values of urea and creatinine masses. Furthermore, it is assumed that the dynamics of urea and creatinine exchange between intracellular and extracellular compartments are more rapid than the urine and dialysis dynamics, so that the intracellular and extracellular compartments may be lumped together (Figure 7.28).

In Figure 7.28, x_M and y_M represent the concentrations in the total fluid volume (V) of urea and creatinine, respectively. It is assumed that the values of total fluid volume, V, and the normal parameters k_1, k_2, k_3, and k_4 are known. The dynamic responses for urea and creatinine responses are given by:

$$x_M(t) = \frac{g_u}{a} + \left(x_o - \frac{g_u}{a}\right)e^{-at} \qquad \ldots\ldots\ (7.7)$$

$$y_M(t) = \frac{g_c}{b} + \left(y_o - \frac{g_c}{b}\right)e^{-bt} \qquad \ldots\ldots\ (7.8)$$

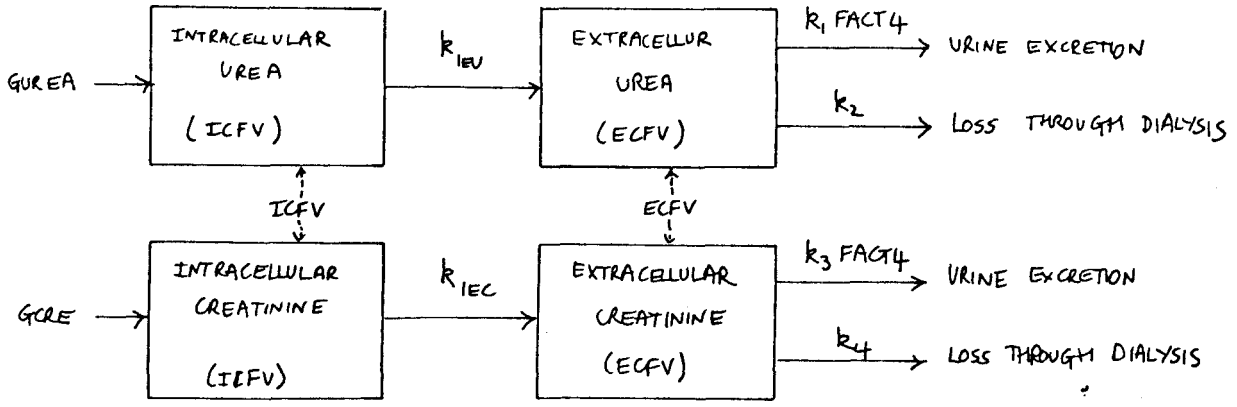


FIGURE 7.27. FULL UREA AND CREATININE DYNAMICS SUBMODEL.

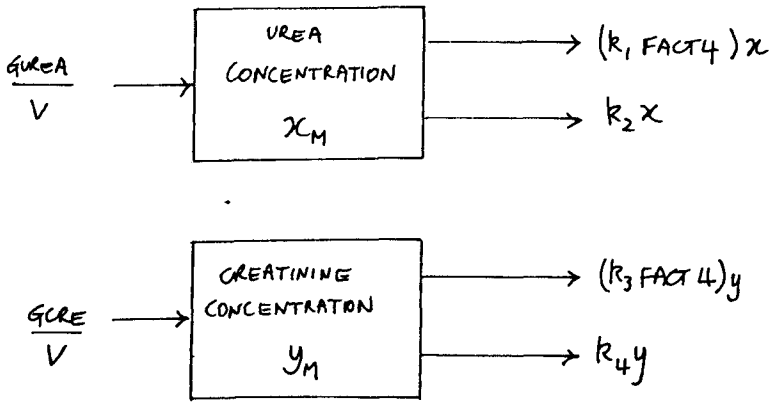


FIGURE 7.28. SIMPLIFIED UREA AND CREATININE SUBMODEL FOR IDENTIFIABILITY ANALYSIS.

where

$$g_u = \frac{GUREA}{V} \quad \ldots\ldots (7.9)$$

$$g_c = \frac{GCRE}{V} \quad \ldots\ldots (7.10)$$

$$a = k_1 FACT4 + k_2 \quad \ldots\ldots (7.11)$$

$$b = k_3 FACT4 + k_4 \quad \ldots\ldots (7.12)$$

and where x_o and y_o are the initial concentrations of urea and creatinine, respectively. Measurements of the concentrations of urea and creatinine in the patient are represented by x_D and y_D and given by:

$$x_D(t) = x_2 + (x_1 - x_2)e^{-\alpha t} \quad \ldots\ldots (7.13)$$

$$y_D(t) = y_2 + (y_1 - y_2)e^{-\beta t} \quad \ldots\ldots (7.14)$$

where x_1, y_1 are the initial values and x_2, y_2 are the steady-state values. (This assumes, of course, that the human response is shown to be first order).

If only data for plasma urea (x) are used, compare equation (7.7) with equation (7.13):

$$x_D(t) = x_2 + (x_1 - x_2)e^{-\alpha t} \quad \ldots\ldots (7.7)$$

$$x_M(t) = \frac{g_u}{a} + \left(x_o - \frac{g_u}{a}\right)e^{-at} \quad \ldots\ldots (7.13)$$

by direct substitution:

$$a = \alpha \quad \ldots\ldots (7.14)$$

$$x_o = x_1$$

and $$\frac{g_u}{a} = x_2 \quad \ldots\ldots (7.15)$$

If the dynamic information is available (i.e. α), then

$$FACT4 = \frac{\alpha - k_2}{k_1} \quad \ldots\ldots (7.16), \text{ from (7.11) and (7.14)}$$

and $$GUREA = \alpha x_2 V \quad \ldots\ldots (7.17), \text{ from (7.9) and (7.15).}$$

However, the data for plasma urea are available only before and after dialysis, and in Section 7.4.3.1 the response was linearly interpolated between these values. Thus, if α is unknown, the value of a cannot be determined and hence any values of g_u and a which satisfy (7.15) are

acceptable. With no dynamic information, therefore, FACT4 and GUREA are parameter unidentifiable.

In the model, only GUREA is affected by the high basal metabolic rate and not GCRE, and it can be assumed that GCRE is known. If the data for creatinine (7.14) are used and compared with (7.8):

$$y_D(t) = y_2 + (y_1 - y_2)e^{-\beta t} \qquad \ldots\ldots (7.14)$$

$$y_M(t) = \frac{g_c}{b} + \left(y_o - \frac{g_c}{b}\right)e^{-bt} \qquad \ldots\ldots (7.8)$$

then, from the steady-state valyes y_1 and y_2:

$$y_o = y_1$$

and $$\frac{g_c}{b} = y_2$$

Since g_c is known, $b = \frac{g_c}{y_2}$ can be determined and hence

$$\text{FACT4} = \frac{b - k_4}{k_3} \qquad \ldots\ldots (7.18), \text{ from } (7.12)$$

and $$\text{GUREA} = (k_1\text{FACT4} + k_2)x_2V \qquad \ldots\ldots (7.19), \text{ from } (7.17)$$

Thus, if steady-state data for both urea and creatinine are used, the model is parameter identifiable and, in theory, the values of FACT4 and GUREA may be uniquely determined. (In Section 7.4.3.3 the results of the parameter estimation procedures supported the assumption that GCRE is normal.)

However, in practice, the concentrations of creatinine are much less than urea and therefore the dynamics of creatinine are not so sensitive to FACT4. This can lead to uncertainty in the estimation of FACT4 based on the creatinine response. In addition, the data for creatinine are more prone to measurement error because of their low values. Any theoretical or practical problems of identifiability could be easily solved if data were available on the urine flow. It can be shown that GUREA is uniquely identifiable from periodic measurements of plasma urea (e.g. across dialysis) and measurements of the mass of urea in urine from micturitions (as long as the intervals between micturitions are timed).

7.4.3.6 Summary

In this section (i.e. Section 7.4.3) the use of parameter estimation at level 5 of the validation hierarchy (Figure 7.12) has been investigated. Despite and size and complexity of the renal model it was possible to use a least square error parameter estimation procedure. However, the difficulties encountered in setting it up and of including physiological constraints and other data mitigate against the further use of such a procedure on the full model (and, indeed, on similar large biological models). By contrast, the systematic manual procedure is easy to implement, and any other information that is available can be easily incorporated. In the three dialyses studied, the data on average daily urine flow were used as an additional variable, but other variables might be available for different patients. Unfortunately, the procedure would be very difficult to automate fully.

Both parameter estimation procedures produce consistent parameter values over the first two dialyses. The fact that the model can cope with different patient conditions and dialysis therapies without parameter change is evidence for the broad structural validity of the model. The parameter values estimated for the third dialysis differed from the values for the first two. This was a consequence of the real changes in kidney function that occurred in the patient. The changes in the kidney failure parameters (FACT1 - FACT4) were tracked well by the systematic manual procedure, but not by the least squares estimates.

In all three dialyses, both estimation procedures led to an increase of the accuracy of the model response over dialysis. For the four biochemical variables (plasma concentration of urea, creatinine, sodium and potassium) the average model error was less than 10% compared with the data. The longer-term predictions of the model (up to the beginning of the next daily dialysis) for the first dialysis were very good, particularly the urea and creatinine responses (average error less than 12%). However, for the third dialysis, the longer-term predictions were substantially inaccurate (average error > 50%). This must be due to the uncertainty of the parameter estimates, and suggests that measures of urine flow rate and composition, even if averaged over a period of hours, must be used in comparing the model and patient when there is a partial regain of kidney function. In the concluding section of this chapter (Section 7.5) a final assessment is made of the predictive validity of the model for use in improving the renal health-care system.

Finally, the results of this section have implications for the representational validity of the model at level 5. The model responses for urea and creatinine were very close to the patient data over dialysis (especially urea which was consistently accurate). The responses of plasma sodium and potassium concentration in the model were satisfactory, although not so good at low sodium and high potassium levels. This may indicate an inadequacy in the electrolyte balance submodel. The urine flow rate in the model was very sensitive to FACT1 but was easily matched to the data, providing indirect validation of the submodels of kidney function and kidney failure. Although there were differences between the arterial pressure in the model, the main trends in the data were captured, suggesting that the cardiovascular submodel lacks short-term controls. The behaviour of the thermoregulatory submodel was not included in the parameter estimation and therefore no inferences can be made about its validity. Similarly, because of kidney failure, the hormonal control loops (i.e. the ADH and renin-angiotensin II - aldosterone systems) are open and therefore inoperative.

7.5 Conclusions

The conclusions to this chapter are divided into two. Firstly, some conclusions on the validity of the renal model, in relation to its objectives, are made in Section 7.5.1. Secondly, some more general conclusions on the programme of validation, the validation results, and their implications for the validation of this type of biological model are made in Section 7.5.2.

7.5.1 Conclusions on the validation of the Uttamsingh renal model

In Section 7.4 the results of the application of theoretical and empirical validity criteria to a variety of levels in the validation hierarchy were presented. The assessment of whether the representational validity of the model is adequate for the modelling objectives has been left until now. Since the modelling objectives clearly divide into two - scientific objectives and utilitarian objectives (see Section 7.3.1) - these are considered separately below.

7.5.1.1 Pragmatic validity of the renal model

The specific utilitarian objectives of the model (Section 7.3.1) require that it should be used as a predictive tool in the system of interest (SOI) which is the care of patients with renal failure undergoing dialysis therapy. The intended range of application ($\mathcal{R}_I$) of the model is therefore the human renal system in a variety of failure modes both on and off dialysis. The model should predict accurately the values of the major clinical variables over dialysis and in the periods between dialyses. The ultimate test of the pragmatic validity of a model is the ex post assessment of the effects on the SOI of using the model. However, since the renal model has not yet been used in a clinical application, the evaluation must be in terms of the potential benefit of the predictive validity of the model over $\mathcal{R}_I$.

In Section 7.4.3.6 it was shown that, with the aid of parameter estimation, the model could generate accurate predictions of plasma urea concentration and other plasma biochemical variables over dialysis (to within 10%) and the main trends in arterial pressure. The longer-term predictions up to the next dialysis were not so accurate, although the predictions for plasma urea were within 10% of the data. Therefore, in principle, the model could be used to predict the outcomes of various dialysis therapies and thereby improve them. The predictions of plasma urea may be used to time optimally the next dialysis. Thus the model appears to have satisfied preliminary pragmatic validity criteria with regard to its intended clinical application, or specific utilitarian objectives.

However, the model is complex, the parameter estimation is difficult to implement, and practical considerations weigh against using it in its present form. More importantly, however, the empirical validity of the model established in the tests of Section 7.4.3 is undermined by the uncertainties in the model associated with some of the submodels (e.g. the thermoregulatory, hormonal, and electrolyte balance submodels). Ironically, the complexity of the model was justified by the intended clinical application.When the model is used to represent a kidney failure patient on dialysis many of the submodels are wholly or partially inoperative. For instance, the ADH control loop is open because the level of plasma ADH does not affect urine flow (there is very little anyway). It is not known what happens to such a system when it goes open loop. Similarly, the sophistication of the kidney function submodel

(which is isomorphic with the functioning of a normal nephron) is not fully used when modelling a patient with severe renal failure. Under these circumstances the model is operating in a degenerative mode, and it is difficult to justify the extra behavioural resources that are available.

The predictive accuracy of the plasma urea and creatinine responses during and between dialyses is evidence for the validity of the urea and creatinine dynamics submodel and the artificial kidney machine submodel. These two submodels could be used on their own for the accurate prediction of urea and creatinine, and could form part of a suite of models available to the clinician in the dialysis unit and implemented, if possible, on a microcomputer. The overall model could be comprehensively reduced to include only factors relevant to a renal failure patient. The thermoregulatory and hormonal submodels would be the most likely candidates for removal. The final form of the model could then be used for overall trend prediction. It might utilise parameter values that had been estimated using simple models tracking urea and creatinine, fluid, and electrolytes. With the simple models it would also be possible to generate uncertainty bands for the predictions. If such a system proved beneficial to the clinician in the design and control of dialysis therapy, it might be possible then to automate the therapy design procedure.

7.5.1.2 Scientific validity of the renal model

The empirical validation of the model applied to a normal human (level 3), to a patient with renal failure (level 4), and to a renal failure patient on dialysis (level 5) demonstrated that the overall responses of the major clinical variables were close to available data (Sections 7.4.2 and 7.4.3), and therefore established the broad empirical representational validity. In addition, in the simulations of a normal human, the validity of some aspects of the hormonal and kidney function submodels were indirectly tested. The results of the parameter estimation procedures indicated the ability of the model to represent accurately kidney failure patients (in different states of failure) on dialysis. Nevertheless, substantial areas of uncertainty were exposed, particularly concerning the theoretical and empirical bases of the hormonal and thermoregulatory submodels, and there is tremendous scope for further empirical validation and model development. (It must be regarded as fortuitous that the most uncertain submodels are largely inoperative when modelling a renal

failure patient, and therefore that the model can yield accurate predictions of such a patient on dialysis to satisfy its specific utilitarian objective.)

Thus the model partially satisfies its specific scientific objective - the representation of $\mathcal{R}_I$ (the human renal system and associated subsystems) - although there is plenty of room for improvement. General scientific objectives are associated with heuristic validity criteria. It is obvious that the model contributes to the "understanding" of the human renal system, but more precisely the model is useful for the following purposes:

(i) Investigating the complex (and often counter-intuitive) interaction of many processes and control systems centred on the kidneys. (The coordination of different physiological control systems is an area of considerable lack of knowledge.)

(ii) As a test bed for hypothesis testing. This was amply illustrated in the indirect validation of hypotheses for the secretion and clearance of ADH at level 3.

(iii) The structuring of empirical research. The data requirements for empirical validation often suggests critical tests that could resolve uncertainties.

In conclusion, the renal model offers plenty of scope for scientific development, and without too many obvious inadequacies.

7.5.2 General conclusions

The scientific and utilitarian modelling objectives have been shown to lead to different requirements for the renal model and to different conclusions on the validity of the model. The present form of the model is more suited to scientific objectives, whereas for clinical application simpler models, possibly a suite, would be more appropriate. These conflicting objectives make validation a lengthy process and rather difficult (the study reported in Section 7.4 is incomplete).

The programme of validation (based on the γ and δ validation methodologies in Chapter 5) and the analytical framework of the theory of model validity (Chapter 4) have produced a detailed and critical assessment of the renal model with close attention to the modelling objectives. As was the case with the Pullen cardiovascular model (Chapter 6), the validation

study is very long. This appears to be unavoidable in large biological models in which there are always many areas of theoretical uncertainty and lack of empirical data. The limited resources available for validation, and the undoubted need for validation, imply that biological models must be simpler if comprehensive validation is to be performed.

In the next chapter, a mathematical model of the human respiratory control system will be examined. The extent of the empirically validated range of application will be described, together with some of the problems impeding further model validation and development.

CHAPTER 8

THIRD CASE STUDY - THE VALIDATION OF A MATHEMATICAL MODEL OF THE HUMAN RESPIRATORY CONTROL SYSTEM.

8.1. Introduction.

The third and final biological model to be considered as a case study for model validation is a breathing model of the human respiratory system and its control developed in the Department of Systems Science by Bali (1974, 1976). The model is based on that of Grodins (1967) with the main additions of a compartment to represent muscle tissue and the modelling of events within the respiratory cycle in order to investigate possible respiratory control strategies or hypotheses, the primary objective of the model. In this chapter, the model will be outlined, a suitable programme of validation described, and the extent of the representational validity will be delineated. In addition, the heuristic validity of the model both in formulating new control strategies and as an aid to experimental respiration physiology will be assessed.

As well as the basic model developed by Bali, two modifications to the submodel of respiratory control will be considered: the feedforward controller of Saunders (1980), and the timing-effect and combined chemical-neurogenic drive controller of Sarhan et al. (1979, 1980). Strictly, therefore, this case study is concerned with the development of a series of respiratory models which constitute a model-based research programme closely related to developments in experimental physiology. This highlights the view that model validation is not simply a last stage in modelling but a continually repeated step in the ongoing process of model and scientific development.

The structure of the chapter is as follows: in section 8.1.1. a very brief outline of respiratory physiology is given to provide an introduction to some of the concepts and terminology. The model, and previous models on which it is based, will be described in section 8.2 (including the Saunders and Sarhan modifications). Section 8.3. is concerned with the development of a suitable programme of model validation, based on an analysis of the modelling

objectives and available data types, and using the framework of the theory of model validity (Chapter 4). The results of model validation, which determine the present extent of its representational validity, are presented in section 8.4. The conclusions of section 8.5. are two-tiered. Firstly, the results of section 8.4. are summarised, the empirically valid range of application is identified, and the heuristic validity of the model is assessed. Secondly, some general conclusions are made on the value of the approach taken in this case study and the implications on the use of the theory of model validity for the validation of biological models in general.

8.1. 1. A very brief introduction to respiratory physiology.

The respiratory system acts in conjunction with the cardiovascular system as a transport system for the intake and distribution of oxygen (O_2) and for the removal of carbon dioxide (CO_2). The cells of the body require O_2 (at approximately 250 ml./min., at rest) for the oxidation of fuels (carbohydrate, fat, protein) in order to satisfy vital energy needs. CO_2, an end product of oxidation, is toxic in high concentrations and has therefore to be removed from the body.

Blood from the various tissues returns to the heart largely depleted of O_2 and high in CO_2 (venous blood). After passing through the right side of the heart it enters the pulmonary circulation of the lungs. The air passages of the lungs terminate in millions of tiny sacs (alveoli) which are surrounded by blood capillaries. Here, the CO_2 in the blood diffuses into the alveolar air, and O_2 from the alveoli diffuses into the blood, rebalancing the partial pressures of the gases in the blood and alveoli. The oxygen enriched blood returns to the left side of the heart where it is pumped at high pressure into the systemic arteries and thence to the tissues, supplying oxygen and completing the cycle. The exchange of gases from alveoli to blood is known as "external respiration" whereas the exchange from blood to tissues is known as "internal respiration".

The gases are carried in the blood either by chemical combination with haemoglobin or by solution in the plasma. The total

concentrations of O_2 and CO_2 in the blood are related to their individual partial pressures, and also to the concentrations of CO_2 and O_2, respectively (the Haldane effect and Bohr shift). As well as its toxic effect, CO_2 affects the acidity of body fluids (through the carbonic acid - bicarbonate buffer system).

The tissues may be considered as four separate compartments: the brain, the muscles, the organs, and other tissues. The demand for O_2 delivery and CO_2 removal varies greatly between these compartments. For instance, the supply to the brain must be constant (cessation of O_2 supply and CO_2 removal for 4 minutes can result in irreversible brain damage), whereas the demands of muscle tissue vary by orders of magnitude depending upon the level of exercise. These requirements entail a control system that is both tight and yet highly flexible or adaptable, which is capable of eliminating small disturbances as well as adapting to significant changes in its relationship to the environment. The full understanding of this control system requires consideration of both the respiratory and cardiovascular control systems and their coordination. However, for the purposes of this introduction (and, indeed, Bali's model) it will be assumed that the cardiovascular control system maintains the flow of blood as a carrier for O_2 and CO_2 as an independent fluidic system with volumetric and pressure sensors. (For a control-theoretic account of the cardiovascular system, consult Chapter 6).

In man there are sensors sensitive to the partial pressures of oxygen (P_{O_2}) and carbon dioxide (P_{CO_2}) and to the acidity (pH) of body fluids, which are known as "chemoreceptors". In addition there are sensors in the muscles which monitor the degree of muscular activity. All these sensors transmit information to the central nervous system (CNS) through afferent nerves. The signals are processed in a diffuse area of the low brain at the top of the spinal cord (known as the "respiratory centre") which in turn modulates the activity of motor nerve cells which control the rate and depth of breathing via efferent nerves to the muscles of the chest and diaphragm. In this way the intake of O_2 and output of CO_2 are controlled. Fig. 8.1 illustrates the structure of the control system and also some of the terminology.

The control system is more sensitive to increases in P_{CO_2} (known

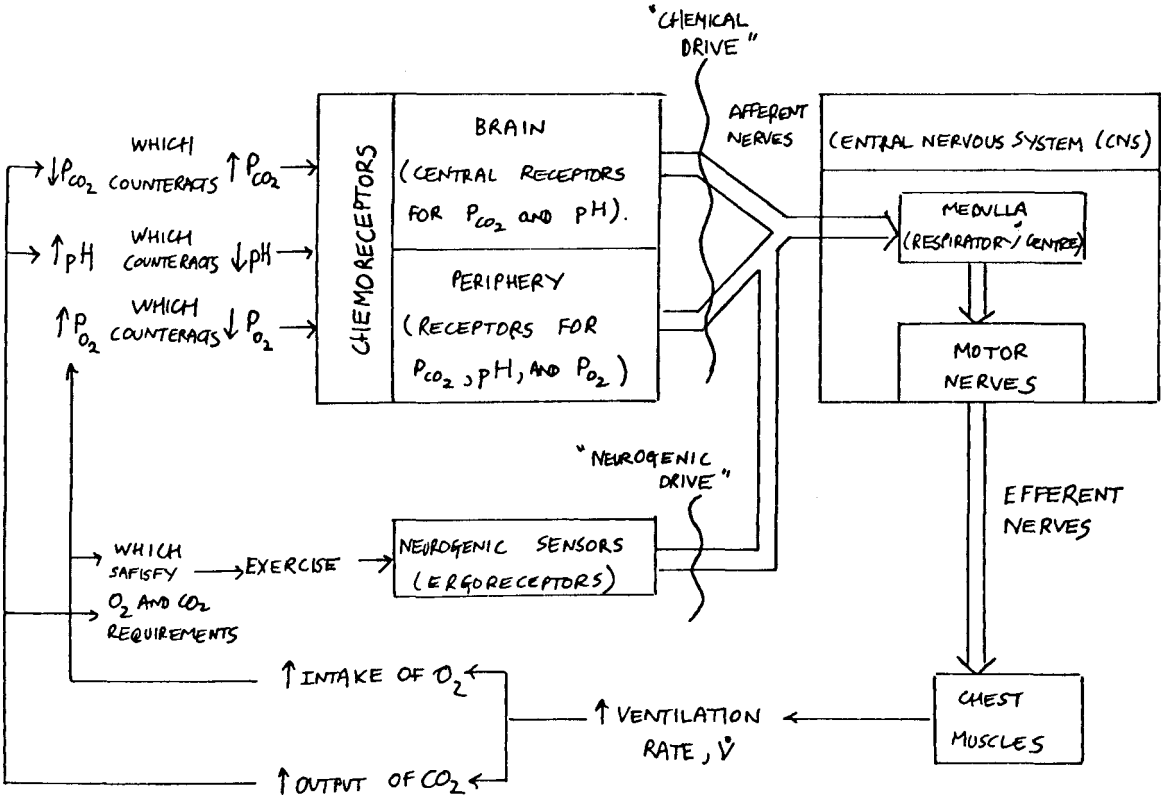


FIGURE 8.1. SIMPLE MODEL OF THE RESPIRATORY CONTROL SYSTEM.

as "hypercapnia") and increases of acidity than to decreases of P_{O_2} ("hypoxia"), however it can be seen from fig. 8.1 that for all these variables the system has negative feedback control. Originally, it was thought that the control of breathing during exercise was effected through this negative feedback as well. However, this model cannot account for many of the phenomena that occur during exercise (e.g. isocapnia, or constancy of P_{CO_2}, implying no chemical drive). The current view is that there are muscular receptors sensitive to the level of exercise, and hence the uptake of O_2 and production of CO_2. From fig. 8.1, it can be seen that the control of breathing for an exercise stimulus is not a negative feedback system - the disturbance, exercise, is not eliminated, rather the system adapts to cope with the gas transport requirements of this new state. The controversy over breathing control in exercise is perhaps partly due to the failure of experimental physiologists to recognise that a quite different control model was requireed and this is reflected in the type of experiments that have been traditionally performed and which do not yield the necessary information, (the situation is now changing, e.g. Kao et al., 1979).

The primary source of uncertainty in the understanding of the respiratory system is how the system is controlled. As well as the problems of control in exercise, there are the problems associated with the interaction of exercise with chemical drives; even within the chemical drive, itself, the interaction between O_2 and CO_2 drives is not understood for all combinations (e.g. hypocapnic hypoxia). This uncertainty can be traced to the fact that the control occurs in the CNS, the understanding of which is still at an early stage. In particular, many of the neuroanatomic pathways and regions of the medulla are not precisely known and the various controllers, such as the respiratory "centre", have a functional rather than physical or structural significance. (The same source of uncertainty was identified in cardiovascular control in Chapter 6, which ultimately led to the invalidity of many of the overall responses of the Pullen cardiovascular model).

The primary aim of the Bali model - the development and testing of respiratory control hypotheses - is therefore closely related to the concerns of experimental respiratory physiology. Furthermore,

from the pioneering work of Haldane on CO_2 control of breathing (1905), through the "multiple factor theory" of Gray (1945), to the graphical approaches of Cunningham and associates (e.g. 1963, 1977), there is a long tradition of quantitative treatment of respiration in physiology. In Bali's model (see section 8.2), most of the equations used in the physio-chemical modelling of the controlled system are based on those developed, tested, and widely used by physiologists. This contrasts strongly with the Beneken (1967) and Pullen (1976) models of the cardiovascular system which introduced new mathematical descriptions of the arterio-venous system which are still largely alien to physiologists. In the terminology of the theory of model validity (Chapter 4), the Bali model is less of an "analogical construct" or "paramorph" than the Pullen model, and there is greater confidence in the validity of the submodels of the controlled system leading to a deductive base for the validity of the overall model.

8.2. Background and Outline of the Bali Respiratory Model.

8.2.1. Introduction and modelling objectives.

The following extract from Saunders, Bali, and Carson (1980) summarises the main modelling objective and the major features of the model:

"The purpose of this work is to design a model which would allow us to test hypotheses on the control of breathing. Since these hypotheses are concerned with breath-to-breath events, cyclic ventilation must be represented and since dead space and shunt have major effects on gas exchange it is wise to include them. If we wish to examine one of the most controversial questions, the control of breathing during exercise, a separate muscle compartment is required. These minimal requirements lead to a model with 14 [non-linear] first order differential equations and nearly 80 main variables excluding dummies used in computation".

In addition it should be added that the model was intended to aid in the devising of control hypotheses as well as testing them (clearly illustrated in Saunders' 1980 paper). In section 8.2.2 some of the previous respiratory models are reviewed and in section

8.2.3. the Bali model is outlined. Attention is mainly given to the submodel of respiratory control which, together with the Saunders and Sarhan modifications, is presented in mathematical terms. The submodel of the controlled system based on physico-chemical modelling is described verbally.

8.2.2. Previous models.

The first explicit mathematical model of the respiratory control system was made by Gray (1945). This algebraic model was concerned with the steady-state responses to CO_2 inhalation, lack of O_2, and metabolic disturbances in the acid-base balance. The first dynamic analysis was made by Grodins, Gray, et al. (1954) which could only accept CO_2 breathing as an input. This original model was extended and refined with the development of computing techniques, finally resulting in the comprehensive model of Grodins, Buell, and Bart (1967) which is something of a classic. It was sufficiently general to accommodate a variety of inputs (CO_2 breathing, hypoxia, and metabolic disturbances in acid-base balance). The following outline of the Grodins model is taken from the 1967 paper:

"It treats the chemical buffering and gas transport systems in reasonable detail, including both Haldane and Bohr effects, and it recognizes the presence of many transport delays [of blood flow] which are themselves dependent variables. It permits convenient exploration of a variety of possible control functions, including the role of CSF [the fluid bathing the brain and spinal cord] hydrogen ion concentrations and of O_2 - CO_2 interaction at the peripheral chemoreceptors".

The Grodins model of the controlled respiratory system (i.e. gas exchange, transport, chemical buffering, control of blood flow) forms the basis of the Bali model described in section 8.2.3. Another class of models has been specifically concerned with the role of CSF in respiratory regulation (e.g. Horgan and Lange, 1962). For a more detailed review of these and other models of the respiratory system consult Bali (1976).

8.2.3. Outline of the Bali respiratory control system model.

Following the original distinction made by Grodins et al. (1967), the model may be considered as two interacting submodels: a submodel of the controlled system (including the dynamics of gas exchange, transport, etc) and a submodel of the controlling system (the chemoreceptors, other receptors, CNS centres, and effector mechanisms). As explained above, the submodel of the controlled system will be described verbally (section 8.2.3.1) whereas the submodel of the controller will be presented mathematically. In section 8.2.3.2 the control submodel used by Bali is outlined. Saunders' feedforward controller for respiratory control in exercise (1980) is described in section 8.2.3.3, and Sarhan's timing-effect controller that incorporates an additive neurogenic exercise drive is outlined in section 8.2.3.4. Notes on the overall structure and computer simulation of the model are made in section 8.2.3.5.

Since the model is largely based on the Grodins model (with the exception of the controller submodel) this will often be referred to, and the major differences indicated. For a full description of the model and its development see Bali (1976); shorter accounts can be found in Sarhan (1979) or Saunders et al. (1980). A full listing of the mathematical equations of the model is given in Appendix IV.

8.2.3.1. Submodel of the controlled system.

The structure of the submodel of the controlled system is shown in fig. 8.2. The differences from the Grodins model are:

(i). A variable volume lung which models events within the respiratory cycle, and leads to oscillatory blood gas concentrations and pH level.

(ii). A dead space which is divided into a fixed anatomical dead space and a partly expansible alveolar dead space.

(iii). A shunt of mixed venous blood past the lung which can represent a number of pathological states (such as emphysema).

(iv). A separate compartment for muscle tissue, which is important in the control of breathing during exercise.

(v). The CSF compartment of the Grodins model is deleted.

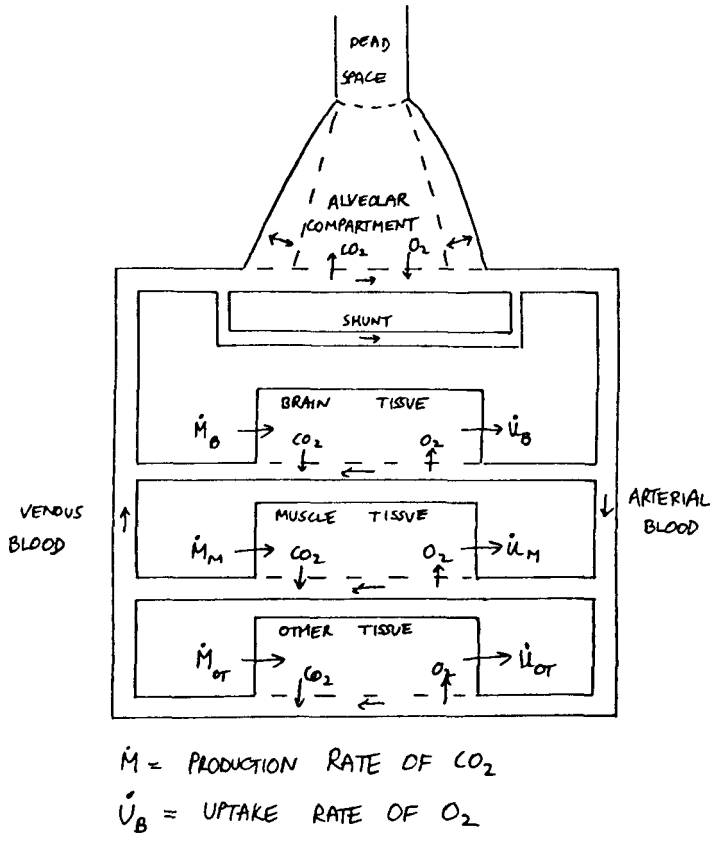


FIGURE 8.2. THE STRUCTURE OF THE SUBMODEL OF THE CONTROLLED SYSTEM.

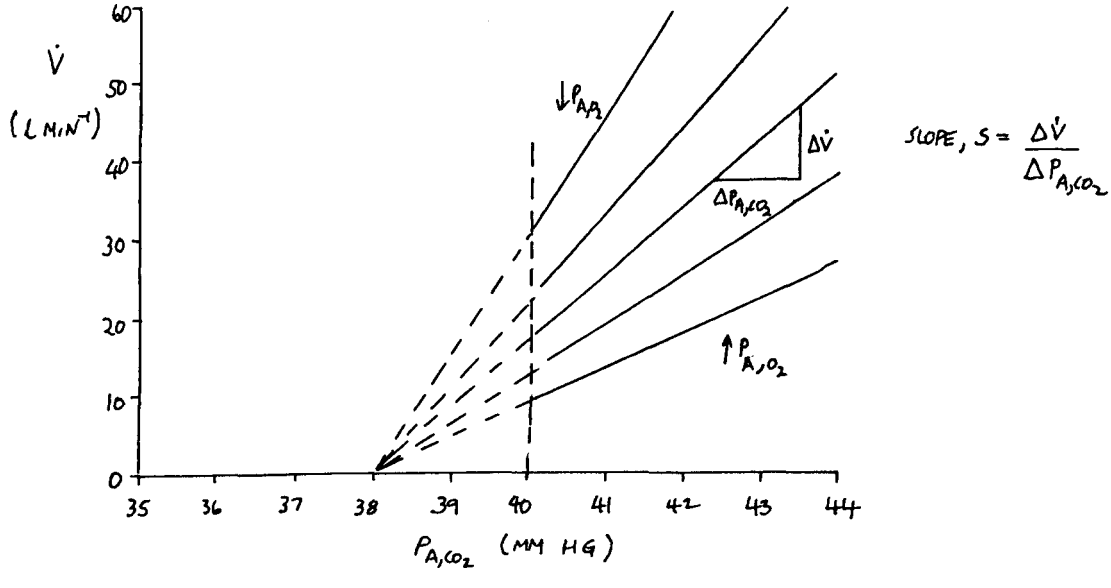


FIGURE 8.3. EFFECT OF ALVEOLAR P_{O_2} (P_{A,O_2}) ON THE RELATION BETWEEN VENTILATION RATE ($\dot{V}$) AND ALVEOLAR P_{CO_2} (P_{A,CO_2}) (TAKEN FROM CUNNINGHAM & LLOYD, 1963).

In the model, the controller is represented simply as a black-box relating changes in model variables to the control of breathing. Thus there is no correspondence to the physical structure of the CNS control system, and only a very rough correspondence to the functional behaviour (only the input - output relation is modelled; neural events are not considered). For this and other reasons, the controller submodel is a large source of uncertainty in the overall model. It is hoped that by comparing overall model responses with different controller submodels with data from the controlled system it will be possible to discriminate between different controllers and thereby test them in order to reduce this uncertainty.

Although the isomorphism of the model with the physical structure of $\mathcal{R}_I$ is not as important as the isomorphism with the functional modality, increased structural accuracy will lead to increased confidence in the functional validity. Most importantly, in studying respiratory control, the structure should correspond at the sites where signals are received for CNS control (i.e. brain tissue P_{CO_2}, arterial blood P_{O_2} and P_{CO_2}, muscle tissue exercise level $\dot{M}_m$). The extension of the Grodins model (1967) by Bali (1976) essentially increased the spatial resolution in order to investigate functional control hypotheses.

The modelling of the controlled system is based on good qualitative physiological theory of physical and chemical processes, such as the dynamics of gas exchange, gas dissociation curves, acid base balance (the Henderson - Hasselbalch equation), etc.. The controller submodels are derived from experimental steady-state results on humans under a limited number of conditions (usually hypercapnia in CO_2 breathing) and uncertainty exists in extrapolating the results beyond these conditions (see also section 8.3.2.).

In the programme of validation, the representational validity of the submodel of the controlled system should be first demonstrated. Then the overall model may be used as a test-bed for inferring the validity of the controller submodels. In doing so, the tests which are performed (the sources of data) should also lie within $\mathcal{R}_I$.

Mass balance equations are written for O_2 and CO_2 in each of the tissue compartments, and the mixed venous blood which depend on the transport delays of the blood. Two sets of equations describe the gas concentrations in the dead space and lung (alveoli) during inspiration and expiration. The partial pressures of O_2 and CO_2 in the alveoli and arterial blood are assumed to equilibrate. The equations for arterial and venous O_2 and CO_2 dissociation curves are based on those of Grodins et al. (1967) which include the Haldane and Bohr effects (the dissociation curves relate the concentration of a gas to its partial pressure, and other factors). The circulatory time delays are expressed as functions of the total and local (i.e. brain, muscle, etc.) blood flows. These flows are in turn controlled by the partial pressures of O_2 and CO_2 according to the empirical equations derived by Grodins et al. (1967). (The control loop: P_{O_2}, P_{CO_2} → chemoreceptors → cardiovascular centres in CNS → blood flow → P_{O_2}, P_{CO_2}, is therefore implicit in the "controlled" system. As Grodins remarked, full understanding of the control of respiration will eventually require explicit models of both cardiovascular and respiratory control, and their interaction).

8.2.3.2. The Bali controller submodel.

The Bali controller is based on a sinusoidal pattern of breathing, and is a combined arterial O_2 and CO_2 controller derived from the data of Cunningham and Lloyd (1963) in the general form suggested by Lloyd et al, (1958). The experimental relationships between ventilation rate ($\dot{V}$) and P_{O_2} and P_{CO_2} are shown in figs. 8.3. and 8.4. The results suggest a multiplicative interaction between the O_2 and CO_2 drives. For low P_{CO_2} (hypocapnia), the results were extrapolated. The equations of the controller are given by:

$$\dot{V} = S_1 (P_{a,CO_2} - 38.0) \quad \text{for } P_{a,CO_2} \geqslant 40.0 \qquad \text{—— (8.1)}$$

where

$$S_1 = 2.2 \left(1.0 + \frac{16.0}{P_{a,O_2} - 30.0} \right) \qquad \text{—— (8.2)}$$

$$\dot{V} = S_2 \left(P_{a,CO_2} - 20.0 \right) + 4.0 \quad \text{for } 20.0 \leqslant P_{a,CO_2} < 40.0 \qquad \text{—— (8.3)}$$

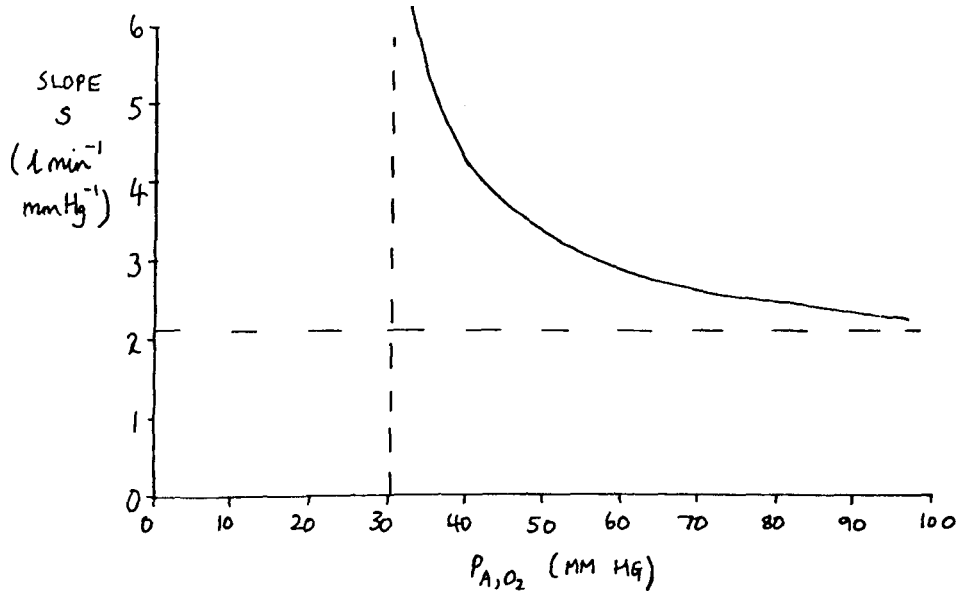


FIGURE 8.4. HYPERBOLIC RELATION BETWEEN SLOPE ($\Delta\dot{V}/\Delta P_{A,CO_2}$) AND ALVEOLAR P_{O_2} (P_{A,O_2})

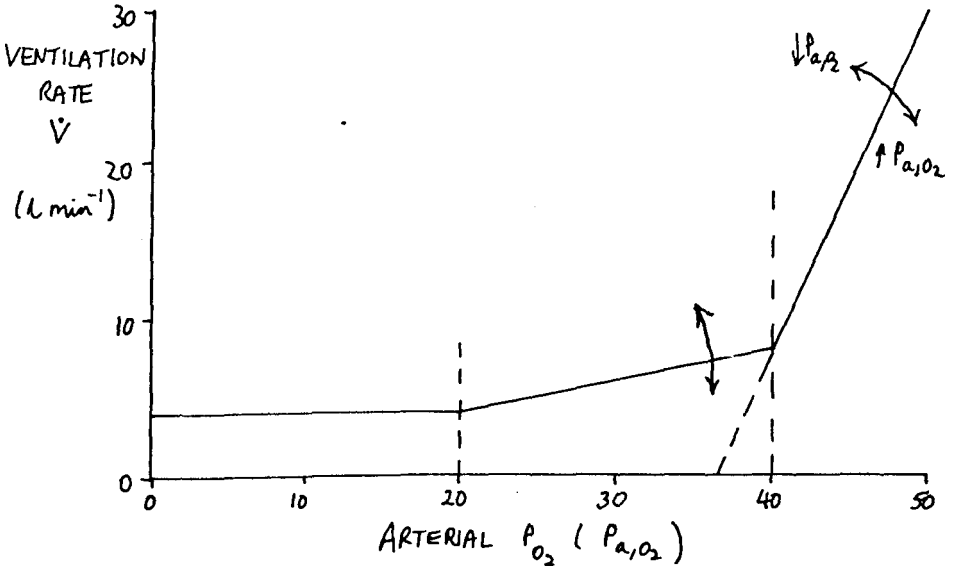


FIGURE 8.5. GENERAL FORM OF THE CONTROLLER USED BY BALI (1976).

where

$$S_2 = 0.1\ (S_1 - 2.0) \qquad \text{——— (8.4)}$$

and

$$\dot{V} = 4.0 \qquad \text{for } 0 \leq P_{a,CO_2} < 20.0 \qquad \text{——— (8.5)}$$

P_{a,O_2} and P_{a,CO_2} refer to the arterial partial pressures of O_2 and CO_2, respectively. The general form of the controller is depicted in fig. 8.5.

8.2.3.3. The Saunders controller submodel (for exercise)

Saunders (1980) postulated a "feedforward" controller for breathing during exercise which he tested using the model with a sinusoidal pattern of breathing. In this controller, the ventilation rate is related to the rate of change of arterial P_{CO_2} :

$$\dot{V} = 6.15 \max \frac{dP_{a,CO_2}}{dt} - 1.0 \qquad \text{——— (8.6)}$$

which applies during exercise. For joint exercise and CO_2 breathing the equation becomes (Saunders, 1980) :

$$\dot{V} = f\ (\text{chemical}) + f \max \frac{dP_{a,CO_2}}{dt} - 7.4 \qquad \text{——— (8.7)}$$

where f (chemical) is the controller described in section 8.2.3.2.

8.2.3.4. The Sarhan controller submodel.

By representing the pattern of breathing as an assymmetric triangular waveform, separate controllers can be provided for inspiration and expiration, thereby allowing timing effects within the respiratory cycle to be investigated. The general form of the timing-effect controller (Sarhan, 1979) is shown in fig. 8.6, and is based on the experimental results of Cunningham and Gardner, (1977). The equations for the controller, which include an oxygen term are given by:

$$\frac{VT}{T_I} = 0.062 \left(1 + \frac{23.2}{P_{a,O_2} - 30.0}\right)\overbrace{\left(P_{a,CO_2} - 35.2\right)}^{\text{"Drive"}} \qquad \text{——— (8.8)}$$

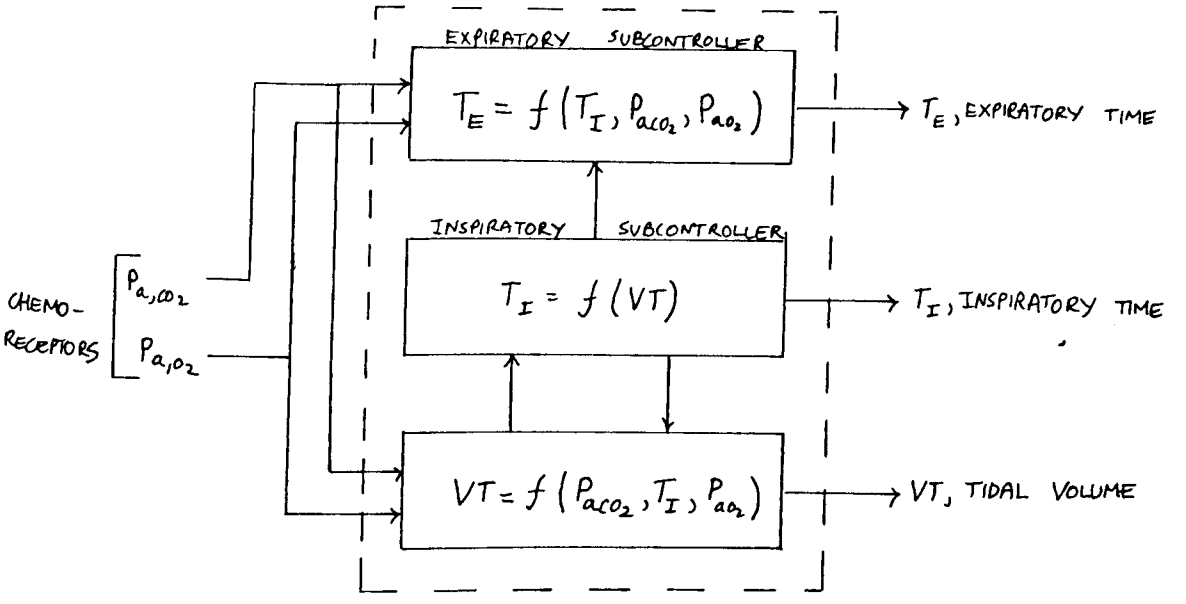


FIGURE 8.6. GENERAL FORM OF THE SARHAN TIMING-EFFECT CONTROLLER SUBMODEL (BASED ON THE DATA OF CUNNINGHAM AND GARDNER, 1977).

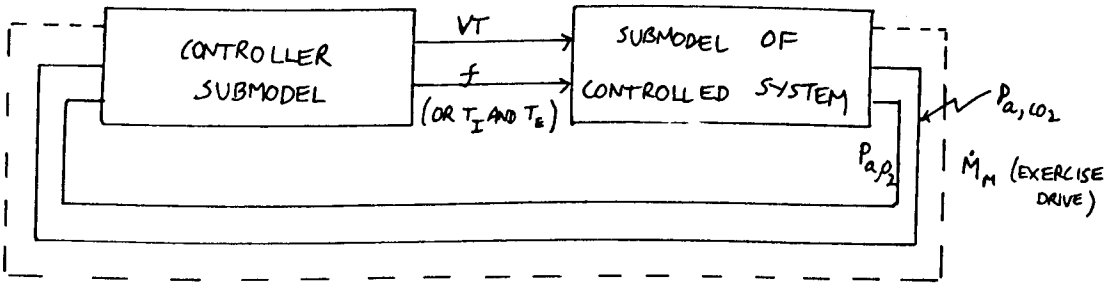


FIGURE 8.7. OVERALL STRUCTURE OF THE MODEL OF THE RESPIRATORY CONTROL SYSTEM.

and

$$T_E = 0{\cdot}64\, T_I + \frac{11{\cdot}1}{((P_{a,CO_2} - 35.2) + 2.73)} \qquad \text{——— (8.9)}$$

The control of breathing during exercise is modelled using a "neurogenic" drive which is related to the muscular activity during exercise expressed as the metabolic production rate of CO_2 ($\dot{M}_M$), under the influence of the work of Kao, et al., 1979). This may be represented by:

$$\frac{VT}{T_I} = f\,(\text{Chemical}) \oplus g\,(\dot{M}_M) \qquad \text{——— (8.10)}$$

where $\oplus$ indicates a combination operation, (the precise form of this is unknown, as is the nature of neurogenic receptors), and f(Chemical) refers to equation (8.8).

8.2.3.5. Overall structure and simulation of the model.

The total model consists of 14 non-linear first order differential equations and approximately 50 algebraic equations. The interaction between the controlling and controlled submodels is relatively simple, as shown in fig. 8.7. The model was simulated in FORTRAN IV on a CDC7600 (at the University of London Computer Centre) requiring 400 seconds of computer time for 30 minutes of real time. The differential equations were solved using a fourth-order Runge-Kutta technique, and the implicit functions (the gas dissociation equations) were solved by the Newton-Raphson technique.

In the next section, a suitable programme of validation for the respiratory model is developed.

8.3. A Programme for Model Validation.

The form of the development of a programme of validation is by now familiar - analysis of modelling objectives (section 8.3.1.) and available data types (section 8.3.2.), leading to appropriate validity criteria and structuring of a programme of validation (section 8.3.3.)

However, at the outset, a suitable methodology can be suggested. This is the γ - methodology (for theoretical/empirical validation, see

Chapter 5) together with the α - methodology for empirical comparisons making use of the feature techniques (to overcome data variability problems; see Chapter 5), and which has been used extensively in the previous two case studies of biological models.

8.3.1. Modelling objectives.

The primary objective of the model is the testing of hypotheses (controller submodels) on the control of breathing in man (section 8.2.1.). The general scientific objective is therefore hypothesis formulation and testing, and the specific scientific objective is a representation of the human respiratory system which is sufficiently accurate to allow inferences on the validity of individual submodels. The specific scientific objective determines the intended range of application ($\mathcal{R}_I$) of the model. The articulation of $\mathcal{R}_I$ requires the use of physiological knowledge and the distinction between the structural (or physical) and functional modalities of $\mathcal{R}_I$. The structural modality of $\mathcal{R}_I$ consists of the lungs, the vascular network (arteries and veins), the various types and locations of tissues, the chemical and neural receptors, and the regions of the brain associated with respiratory control. The description of the functional modality is linked closely to the understanding of the behaviour of $\mathcal{R}_I$, and consists of the dynamic responses of respiratory frequency, tidal volume, concentrations (or partial pressures) of O_2 and CO_2 in various tissues or the blood, blood pH, input-output relations of the controller, etc. The time resolution of $\mathcal{R}_I$ is 0.25sec which allows events to be reproduced within the respiratory cycle ($\simeq$ 2 - 5 sec).

It is clear that because of the immense anatomical complexities, the model can only be a very rough approximation of the structural modality of $\mathcal{R}_I$. However, the structural modality of the model provides a frame of reference for the functional modality of the model (the structural resolution is related to the number of variables). Therefore the criterion of validity for the representation of the structural modality of $\mathcal{R}_I$ is that it should be sufficiently detailed to allow an accurate representation of the functional modality. (The primary objective of the respiratory model is the development of functional models of respiratory control. In models concerned with the understanding of anatomical structure, the relation between structural and functional modalities is reversed).

8.3.2. Available data types.

The main respiratory variables (frequency, tidal volume, alveolar gas composition) may be monitored continuously using a spirometer, which may also measure the total metabolic O_2 demand. The values of alveolar P_{O_2} and P_{CO_2} may be taken as measures of the arterial partial pressures. Samples of blood may also be used to determine arterial partial pressures and pH (this is not usually performed continuously). The compartmental parameters (volumes, metabolic O_2 uptake and CO_2 production) are usually known for a normal human, although their determination for an individual would be complex. Similarly, gas exchange and dissociation parameters are widely known in the normal case. Although the major variables of interest ($\dot{V}$, f, VT, P_{a,CO_2}, P_{a,O_2}) are easily measurable, the experimental test conditions required to investigate the behaviour for various combinations of $\Delta P_{a,CO_2}$, $\Delta P_{a,O_2}$, and Δ exercise are not all easy to set up.

The effects of increases in P_{a,CO_2} (hypercapnia) can be investigated by increasing the concentration of CO_2 in the inspired air, and at the same time the effects of altering P_{a,O_2} by increasing or decreasing the concentration of O_2 in inspired air can be examined (in effect the experiments of Cunningham and Lloyd, 1963). However, if the concentration of O_2 in the inspired air is reduced on its own (i.e. hypoxia), the increased ventilation rate leads to a "washing-out" of the CO_2 (hypocapnia). Under these circumstances it is very difficult to construct graphs of the relationship between $\dot{V}$, f (or T_I and T_E), and P_{a,O_2} for different P_{a,CO_2} levels because of the variability of the induced hypocapnia. A recent advance in experimental respiratory physiology is to use computer-controlled valves for regulating the concentration of CO_2 in the inspired air in order to maintain a constant hypocapnic level by monitoring the alveolar P_{CO_2} in a negative feedback loop. Similar controllers are used to systematically vary P_{O_2}. The results of Young (1970) suggest that the relationship between $\dot{V}$, P_{a,CO_2} and P_{O_2} in hypocapnic hypoxia is not as well behaved as the graph of fig. 8.5 indicates (for $P_{a,CO_2} \lesssim 40$ mmHg).

Any step change in the concentration of O_2 or CO_2 in inspired air will lead to a reflex change in ventilation rate and consequently an

additional change in the input rate resulting in a non-step net input. In the case of CO_2 breathing the increased ventilation rate response leads to an extra increase of the intake of CO_2 - stable positive feedback. The Fenn-Craig technique (1963) allows a true step change of inspired CO_2 to be applied.

In man, the precise measurement of neural activity associated with respiratory control centres in the brain is not possible to achieve. However, in the model the controller is represented as an input - output relation between physical respiratory variables and so this is no problem.

The investigation of respiratory control during exercise is complicated by the fact that the stimuli and receptors for the supposed "neurogenic" drive are not exactly known. In addition, there is a controversy between those who favour a purely chemical control (e.g. Saunders, 1980), and those who accept the validity of a neurogenic drive (e.g. Kao, et al., 1979). It is likely that a full explanation will be based on both respiratory and cardiovascular control systems which poses another challenge to the primarily single-input single-controller single-output conceptual models of experimental physiology.

8.3.3. A suitable programme of validation.

In order to use the model to test the validity of the various controller submodels (section 8.2.3.2-4) - the primary objective - the validity of the submodel of the controlled system (section 8.2.3.1.) should first be established. This suggests that the validation programme should begin by applying theoretical and empirical criteria to the elementary submodels (including assumptions) and gradually work up to the overall structure and behaviour of the model. Ideally, a fair amount of confidence in the representational validity of the submodel of the controlled system should be obtained before considering the overall model with controller. The general form of the programme of validation suitable for the respiratory model is shown in fig. 8.8. Initially, some prerequisite criteria for model stability should be satisfied. Then follows the disassembly/reassembly representational validation. In examining the overall validity of the model, parameter estimation and system identification techniques may be useful. Since the available data show the normal variability of a biological population,

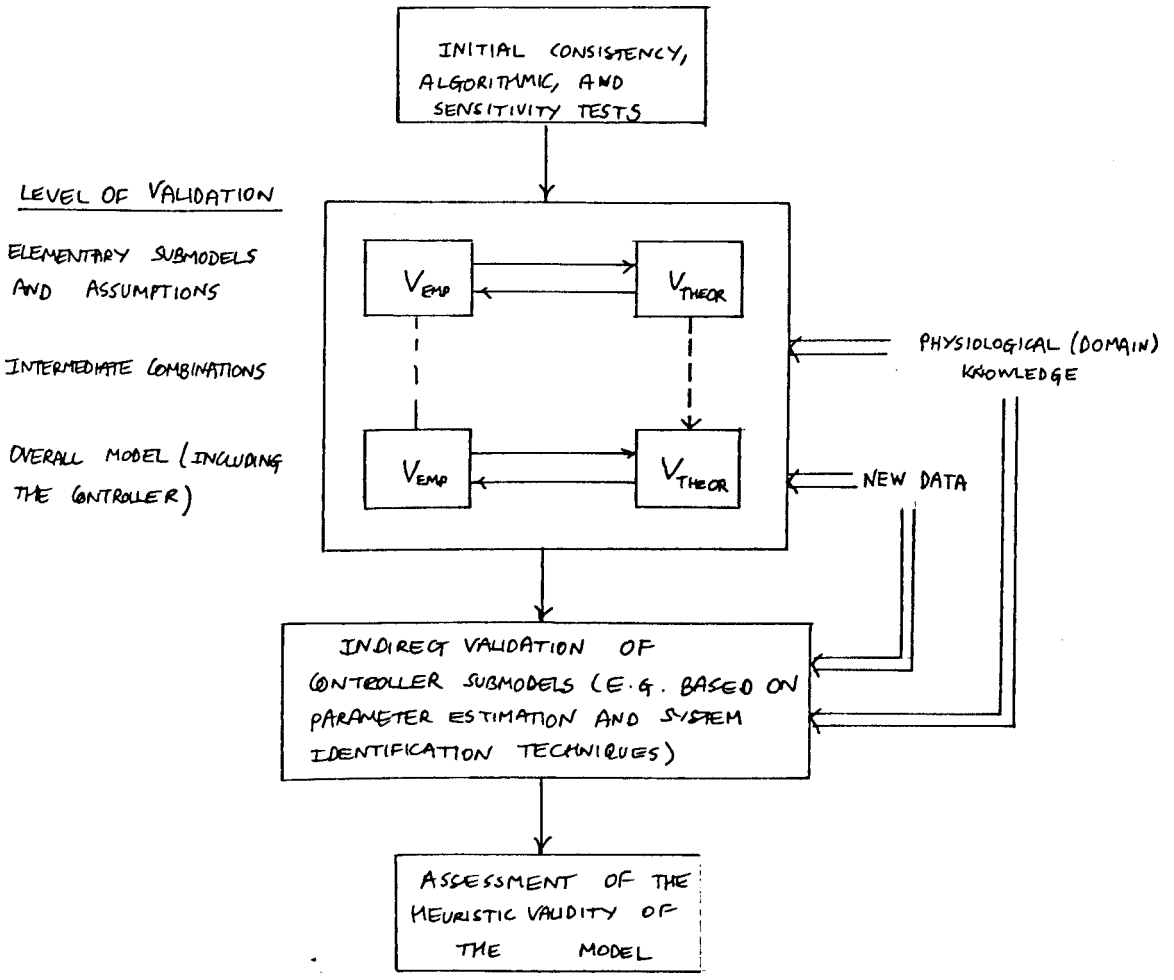


FIGURE 8.8. PROGRAMME OF VALIDATION FOR THE RESPIRATORY MODEL (BASED ON THE Y- METHODOLOGY, SECTION 5.4).

comparisons based on feature space techniques will be very useful (as described in the α-methodology, section 5.2). The final stage of the programme is an assessment of the heuristic validity of the model. This is a critical assessment of the potential of the model as a tool in physiological research, in particular its ability to be used for understanding the respiratory control system, for formulating and testing new controller hypotheses, and for stimulating experimental and theoretical research.

It is clear from fig. 8.8 that the programme of validation proposed for the respiratory model corresponds exactly to the δ-validation methodology (Chapter 5, section 5.4), and which was used in the validation of the cardiovascular and renal models (Chapters 6 and 7). This appears, therefore, to be a fairly general programme for the validation of physiological system models. An important aspect of the programme is the decomposition of the model to form a hierarchy or tree of validation. This makes it very clear which submodels are valid or well-based and which remain unvalidated and uncertain. The concept of the "level" or "depth" of validation is also important. A decomposition of the respiratory model is shown in fig. 8.9.

The empirical validation results presented in the next section (8.4) all concern validity at level 4 in fig. 8.9. The model response is compared with data from a normal human in a range of experimental tests: increased CO_2 inhalation, decreased O_2 breathing, and exercise. As far as current understanding is concerned, these tests do not cause significant changes outside $\mathcal{R}_I$ which then affect the respiratory system. Empirical tests could be performed at level 3 in a variety of ways, e.g.: (a). by examining model variables over one respiratory cycle when the system is effectively uncontrolled; or (b). comparing the model with data from a patient with impaired neural function, but these are not reported.

8.4. Results of Validation.

In this section, some of the results of the application of empirical validity criteria to the respiratory control system model (level 4 of fig. 8.9) in a number of physiological tests will be presented. In addition, in section 8.4.1., some remarks on the theoretical validity (and possible empirical tests) of some of the submodels are made. The results are given for each of the controller submodels

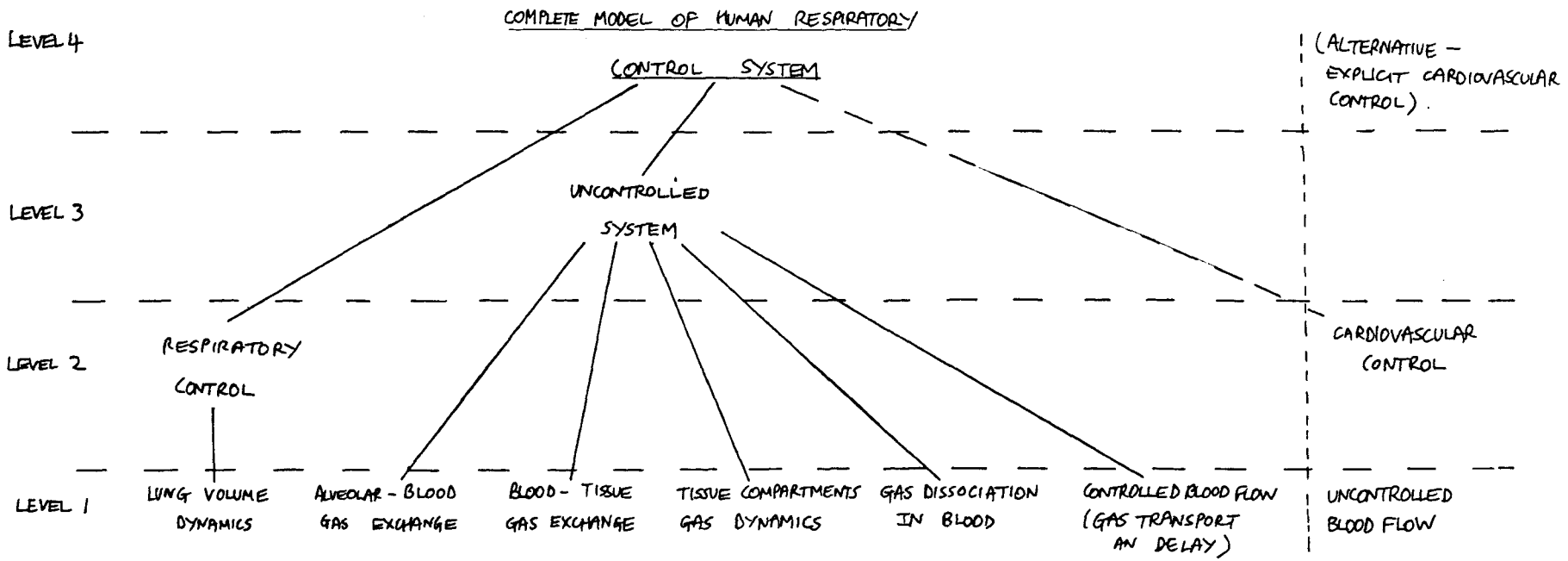


FIGURE 8.9. HIERARCHY OR TREE OF VALIDATION FOR THE BALI RESPIRATORY MODEL.

described in section 8.2.3 in an attempt to determine their valid ranges of application. The threee physiological tests which are examined are: (i). increased CO_2 concentration in inspired air (section 8.4.2): (ii). decreased O_2 concentration in inspired air (section 8.4.3); and (iii). the effect of exercise (section 8.4.4).

8.4.1. Notes on the representational validity of the submodels.

With the exception of respiratory control and blood flow submodels, the submodels are based on physico-chemical principles of modelling such as mass balance, chemical buffering systems, acid-base balance, etc. and many of the parameters of these submodels can be determined physiologically (i.e. as opposed to fitting the parameters of the overall model). This establishes a good degree of confidence in the theoretical and empirical validity of the submodel of the uncontrolled system (level 3 in fig. 8.9). The simulations of the original model (Grodins et al., 1967) and of the gradual extensions by Bali (1976) provide additional confirmation. These extensions were designed so that all the factors or mechanisms that are currently known to play a role in respiratory control would be represented in the model (e.g. sites for chemoreceptors, separate compartments for brain, muscle, and other tissues, cyclic ventilation, and the control of tidal volume, frequency of breathing, or T_I and T_E).

The submodel of the controlled blood flow (and hence dependent time delays for gas transport between compartments) is the empirical model derived by Grodins et al. (1967), (the blood flows are expressed as functions of the arterial P_{CO_2} and P_{O_2}). The empirical basis of the submodel supports its representational validity, however it would be good to know the accuracy of the fitted equations (e.g. the expected variances of the coefficients). An eventual modification of the respiratory model would be to remove the implicit cardiovascular control and to have an explicit cardiovascular control submodel (see fig. 8.9). From a systems point of view, the full understanding of either the respiratory or cardiovascular systems requires an understanding of both and their mutual coordination.

8.4.2. Increased CO_2 concentration in inspired air.

The fraction of CO_2 in the inspired air was abruptly increased from 0 to 3, 5, 6 and 7% to match the data of Reynolds et al. (1972). In the model, ventilation rate increases slowly and arterial P_{CO_2} increases more rapidly. A comparison of the model with the original controller (section 8.2.3.2) and the data is shown in table 8.1. (taken from Saunders et al., 1980) which shows a reasonable agreement except that the half time of the ventilation off-transient in the model does not decrease with increased CO_2 fraction and the off-transient undershoots in alveolar P_{CO_2} are too large for low CO_2 concentrations. The discrepancy during the transient may arise because the controller is based on the steady-state experimental results of Cunningham and Lloyd (1963).

The steady-state response of the model using the Sarhan timing-effect controller (section 8.2.3.4) for the same CO_2 breathing tests is shown in table 8.2 (taken from Sarhan et al., 1979). These results are also in good agreement with the data of Reynolds et al. (1972).

The results of these empirical tests confirm the validity of the overall model and therefore, indirectly, the validity of the two controller submodels. However, the extent of validity (i.e. $\mathcal{R}_V$) is clearly limited to conditions of raised arterial P_{CO_2} (hypercapnia), and also to steady-state changes. Since the respiratory cycle in man is asymmetric and approximately triangular rather than sinusoidal, applying representational validity criteria to the two controllers will result in the elimination of the Bali controller and the further confirmation of the timing-effect controller of Sarhan et al. (1979).

Bali (1976) used parameter estimation and feature space pattern recognition techniques to demonstrate that the controller based on arterial P_{CO_2} and P_{O_2} could not optimally fit the data with a unique set of parameter values for all CO_2 concentrations (i.e. the parametric consistency criterion, section 5.2.4.1.3) whereas a controller deriving inputs of P_{CO_2} from both brain and arterial compartments, and P_{O_2} from the arterial compartment could produce a unique optimal fit. Whilst it is known that respiratory control is effected through pH sensors in the CSF (i.e. brain) as well as peripheral (arterial)

	STEADY-STATE		½-TIME ON-TRANSIENT (SEC)		OVERSHOOT (%)		½-TIME OFF-TRANSIENT (SEC)		UNDERSHOOT (%)	
	M	D	M	D	M	D	M	D	M	D
MINUTE VENTILATION ($L.min^{-1}$)	12-44	11-41	42-106	64-135	0	0	33-42	43-12	0	0
ALVEOLAR P_{O_2} (MM HG)	122-142	126-144	62-51	72-45	0	0	74-120	49-84	0	0
ALVEOLAR P_{CO_2} (MM HG)	42-55	44-56	4-10	7-13	47	24	4-1	6-6	140-50	27-72

(FOR 3% CO_2)

TABLE 8.1. COMPARISON OF MODEL (WITH BALI CONTROLLER SUBMODEL) WITH DATA FOR 3-7 % CO_2 BREATHING (REYNOLDS ET AL, 1972). (EACH ENTRY SHOWS THE RANGE OF RESPONSES FOR 3-7% CO_2 BREATHING).

	PRE-STIMULUS VALUES		STEADY-STATE	
	M	D	M	D
MINUTE VENTILATION ($L\,MIN^{-1}$)	8	6	13-49	11-41
ALVEOLAR P_{O_2} (MM HG)	103	102	119-137	126-144
ALVEOLAR P_{CO_2} (MM HG)	40	43	42-55	44-56
RESPIRATORY FREQUENCY (MIN^{-1})	17	11	18-26	13-26

TABLE 8.2. COMPARISON OF MODEL (WITH SARHAN CONTROLLER SUBMODEL) WITH STEADY-STATE DATA FOR 3-7% CO_2 BREATHING (REYNOLDS ET AL, 1972). (CONTROLLER INCLUDES OXYGEN TERM IN THE INSPIRATORY FLOW EQUATION).

chemoreceptors, the use of automatic parameter estimation techniques etc. are probably too sophisticated at this stage in the model development. One of the difficulties is that of the variability of data corresponding to a normal human. This means that there are differences between the model and the data which cannot be eliminated by simply adjusting controller parameters. Furthermore, the theoretical and practical identifiability of the estimation problem posed with the model should also be examined, (see, for instance, the validation of the renal model in Chapter 7).

8.4.3. Decreased O_2 concentration in inspired air.

Table 8.3 shows the model results (with the Bali controller) for 7, 8, and 9% O_2 breathing compared with the data of Reynolds and Milhorn (1973). The steady-state results show a larger ventilation with lower alveolar P_{CO_2} for the model, but despite this the alveolar P_{O_2} was also lower than the data. The on-transients for ventilation rate were several times slower in the model and there were large overshoots not seen in the data. The corresponding results for the model with the timing-effect controller are given in table 8.4 (taken from Sarhan et al., 1979). Although the steady-state ventilation rate matches the data more closely than in table 8.3, there is a very significant discrepancy in the steady-state value of alveolar P_{O_2}.

Clearly, neither of the sets of model responses satisfies the empirical validity criteria. This is due to an inadequacy in the controller submodels, both of which were based on experiments for increased CO_2 breathing. In hypoxia, the arterial P_{CO_2} also falls (due to a variety of factors, mainly the "washing-out" of CO_2 due to the high ventilation rate) into the hypocapnic region to the left of 40mmHg in fig. 8.5. The extrapolation of the controller equation in this region is uncertain and not based on unequivocal experimental results, (the problems of experimentally investigating hypocapnic hypoxia were discussed in section 8.3.2). The inadequacy of the model response to hypoxia is therefore not too surprising.

8.4.4. The effect of exercise.

Exercise was simulated in the model by simultaneously increasing the O_2 uptake and CO_2 production in the muscle compartment together

	STEADY-STATE		½ TIME ON-TRANSIENT (SEC)		OVERSHOOT (%)		½ TIME OFF-TRANSIENT (SEC)		UNDERSHOOT (%)	
	M	D	M	D	M	D	M	D	M	D
MINUTE VENTILATION ($L\ MIN^{-1}$)	18-12	15-10	174-167	54-33	97-36	0	13-24	21-27	0	0
ALVEOLAR P_{O_2} (MM HG)	32-36	39-45	17-18	15-29	0	0	—	168-63	0	0
ALVEOLAR P_{CO_2} (MM HG)	24-31	31-35	195-200	105-69	0	0	120-70	120-80	0	0

TABLE 8.3. COMPARISON OF MODEL (BALI CONTROLLER) WITH DATA FOR 7-9% O_2 BREATHING (REYNOLDS AND MILHORN, 1973).

	PRE-STIMULUS VALUES		STEADY-STATE	
	M	D	M	D
MINUTE VENTILATION ($L\ MIN^{-1}$)	8	6	17-9	15-10
ALVEOLAR P_{O_2} (MM HG)	103	103-107	30-31	39-45
ALVEOLAR P_{CO_2} (MM HG)	40	42	27-35	31-35
RESPIRATORY FREQUENCY (MIN^{-1})	17	12	12-11	15-13

TABLE 8.4. COMPARISON OF MODEL (SARHAN CONTROLLER) WITH DATA FOR 7-9% O_2 BREATHING (REYNOLDS AND MILHORN, 1973).

with an appropriate increase in cardiac output (Saunders, Bali and Carson, 1980). In addition, a reservoir of 400 ml O_2 was added to the muscle compartment (simulating myoglobin O_2 storage), and the gain of CO_2 in the controller equations was increased to match the increased production of CO_2. Fig. 8.10a shows the results of the model simulation using the Bali controller and a typical human response. The most important features of the human response are the sudden increase in ventilation rate at the start of exercise followed by a gradual rise, and the isocapnia of P_{CO_2}. In the model neither of these occur; there is a delay of about 6 seconds before the ventilation rate responds with a large overshoot, and P_{CO_2} exhibits very large overshoot and undershoot.

The delay in the model response is the time before the increased CO_2 output is sensed as an overall increase of P_{CO_2} at the peripheral chemoreceptors (i.e. the speed of the signal transmitted in the blood). The invalidity of the model response with the Bali controller suggests that there is an alternative control mechanism for the control of breathing during exercise. The controller postulated by Saunders (1980, see section 8.2.3.3) is sensitive to the maximum rate of change of arterial P_{CO_2}, and produces the response shown in fig. 8.10b for 100 W exercise. This compares very well with the human response (fig. 8.10a), demonstrating the sudden rise of ventilation rate at the beginning of exercise and isocapnia in the steady-state. The instantaneous response of the controller which is still chemically driven via arterial P_{CO_2}, arises because of the haemodynamic effect of the suddenly increased cardiac output and blood flow in increasing the oscillatory frequency (and hence max dP_{CO_2}/dt) of the existing arterial P_{CO_2} variations.

Although the Saunders' controller works satisfactorily as a single drive during exercise, it is very sensitive to the parameter values, and there are problems when combining it with the other controllers (it cannot simply be turned off in hypercapnia). Furthermore, the controlled model response during exercise will be very sensitive to the changes in blood flow dynamics, and the correct modelling of cardiovascular control is therefore important. An alternative approach to the control of breathing during exercise is to use an additive combination of chemical and neurogenic drives. Although the precise location and form of receptors is uncertain there is good experimental evidence

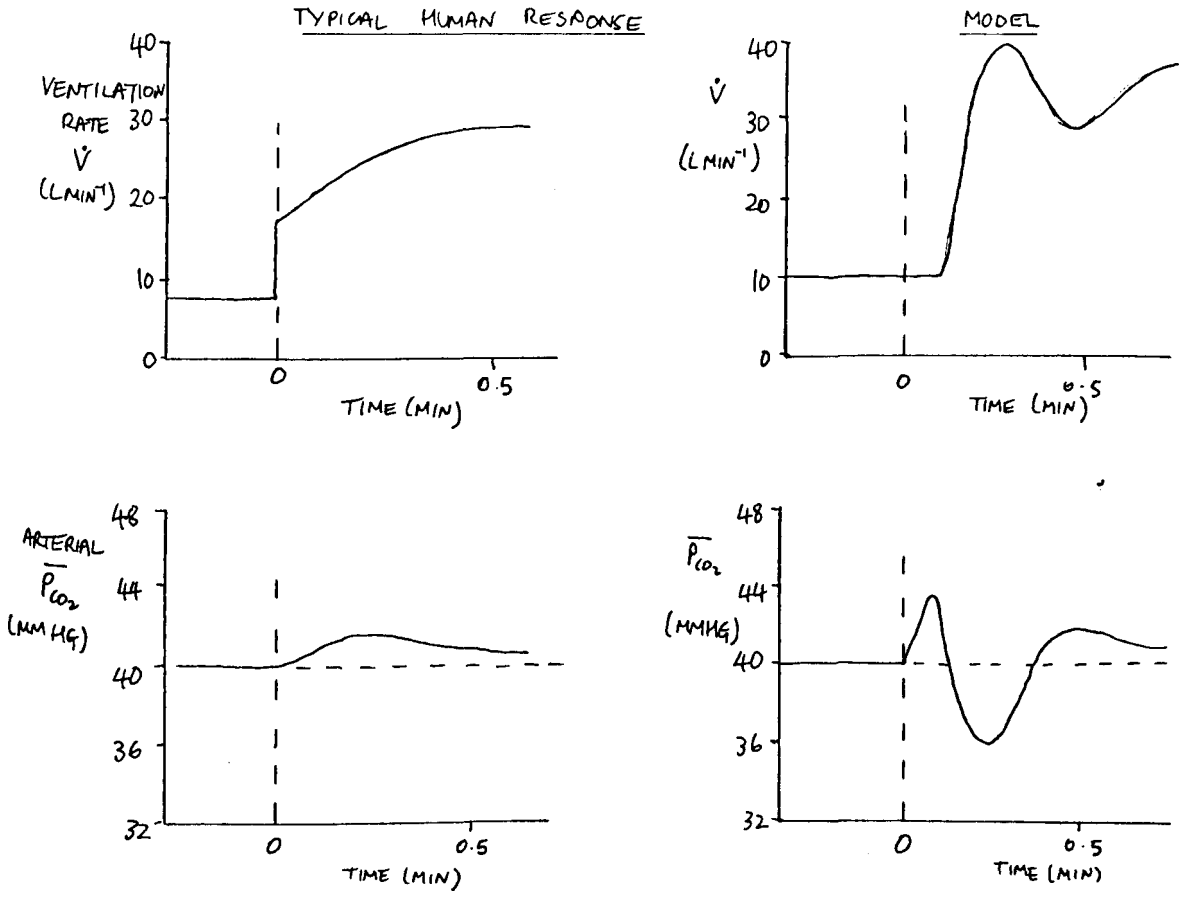


FIGURE 8.10a MODEL (BALI CONTROLLER) AND HUMAN RESPONSE DURING 100W EXERCISE

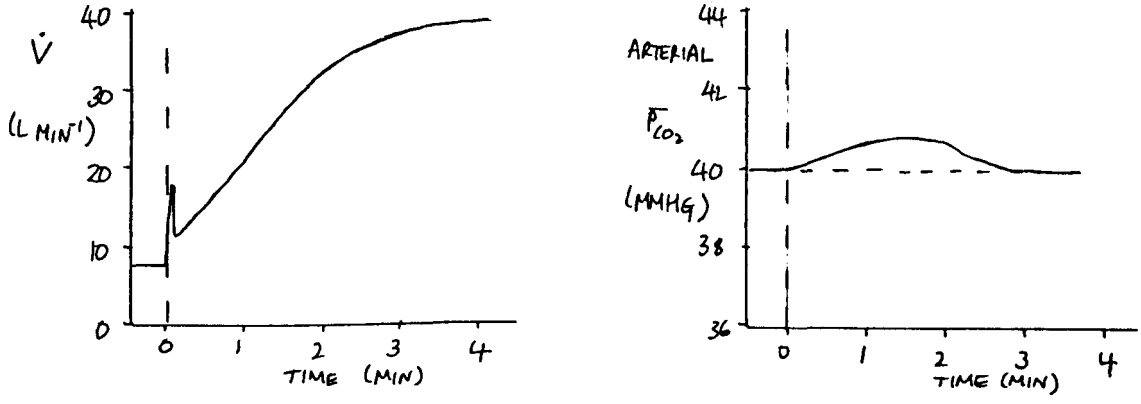


FIGURE 8.10b MODEL RESPONSE FOR 100W EXERCISE (TAKEN FROM SAUNDERS, 1980).

that they exist (Kao et al., 1979). Sarhan et al. (1980, see also section 8.2.3.4) have used a neurogenic drive, which is a function of the CO_2 production in muscle, in combination with the chemical drive and the preliminary results are promising.

8.5. Conclusions.

In section 8.5.1, the results of validation presented in section 8.4. will be summarised and an assessment made of whether the respiratory model satisfies its objective. The general conclusions to the chapter are made in section 8.5.2.

8.5.1. Conclusions on the validity of the model of the respiratory control system.

The submodel of the uncontrolled system (or respiratory "plant") has been shown in theoretical and overall empirical tests to be a sufficiently accurate representation for the purpose of investigating respiratory control hypotheses. However, if possible, additional empirical tests should be performed on this submodel (level 3 in fig. 8.9). One of the difficulties is associated with the implicit cardiovascular control, and this may have to be separated eventually.

In the analysis of the available data (section 8.3.2) it was shown that insufficient data were available for complete formulation of the controller submodels for all possible disturbances, and in the model (section 8.2.3) various approximate extrapolations were made (e.g. for hypocapnic hypoxia). Nevertheless, in section 8.4, data from physiological tests on CO_2 and O_2 breathing, and exercise were used to validate or invalidate the controller submodels indirectly via the overall model responses. The results, which determine the empirically valid range of application ($\mathcal{R}_V$) of the model are illustrated in fig. 8.11 for changes in arterial P_{O_2} and P_{CO_2}. They apply to both the sinusoidal (Bali, 1976) and timing effect (Sarhan et al., 1979) controls, although the latter offer a more accurate representation of events within the respiratory cycle. New experimental data are required in order to formulate controllers for hypocapnia.

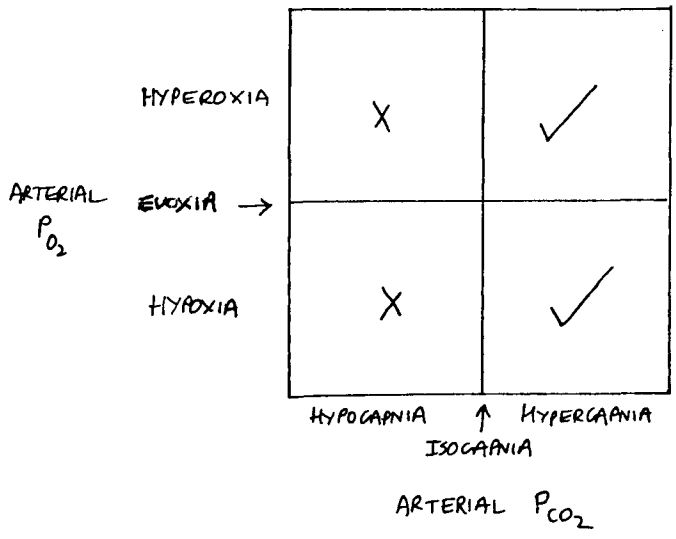


FIGURE 8.11. THE EMPIRICALLY VALID RANGE OF APPLICATION (R_V) OF THE MODEL FOR CHANGES IN P_{O_2} AND P_{CO_2}. (THIS APPLIES TO BOTH SINUSOIDAL (BALI) AND TIMING-EFFECT (SARHAN) CONTROLLERS).

The respiratory controllers based solely on chemical drive produced unacceptable results in the simulation of the effects of exercise. However, the feedforward controller of Saunders (1980) and combined chemical/neurogenic drive control of Sarhan et al. (1980) gave responses that compared favourably with typical human responses, in a preliminary analysis.

Although the empirically valid range of application ($\mathcal{R}_V$) does not cover the whole of the intended range of application ($\mathcal{R}_V$) of the model, it is being increasingly extended with the introduction of better controller submodels. Moreover, the submodel of the uncontrolled system is sufficiently valid to allow the effective and critical indirect validation of the controller submodels. This is evidenced by the invalidation or rejection of some submodels. In this sense, the model satisfies its primary objective - the testing of hypotheses on the control of breathing (section 8.2.1) - and must therefore be considered "valid" for this heuristic purpose. However, the model satisfies heuristic criteria beyond this main objective - there is no doubt that it contributes to the understanding of respiratory control, and offers plenty of scope for investigating the multiple interacting control loops and devising new controller strategies for testing, and also that it can be of help in experimental physiology.

8.5.2. General conclusions.

In this chapter, the validation of the Bali model of the human respiratory control system (and some modifications of it) has been structured using the framework of the theory of model validity (Chapter 4). The programme of validation was very similar to those used in the validation of the Pullen cardiovascular model (Chapter 6) and the validation of the Uttamsingh renal model (Chapter 7) and is essentially the γ-validation methodology described in Chapter 5. As well as providing a systematic basis for the validation of the respiratory model, the programme allows a critical final delimiting of the empirically valid range of application of the model and the major sources of uncertainty in the model (section 8.5.1).

The respiratory model has emerged better than the cardiovascular and renal models in terms of satisfying its objectives. The reasons for this are: (a). the terms of reference, objectives, etc. of the

model are closely related to respiratory physiology; and (b). the model is well-based on quantitative physiological theory and has less theoretical innovation or uncertainty than the other models. The lesson is simple and obvious - models of biological systems which introduce large amounts of new theory or concepts (i.e. analogical models) will be very difficult to validate in anything but heuristic terms.

In the next Chapter, some of the problems associated with the validity and validation of models in the social sciences will be considered.

CHAPTER 9

FOURTH CASE STUDY - SOME ASPECTS OF VALIDITY AND VALIDATION OF MATHEMATICAL MODELS IN THE SOCIAL SCIENCES

9.1 Introduction

The title of this chapter covers an extremely wide and dangerous area - model validity and validation in the epistemological and methodological minefield of the social sciences. Some observations will be made on the nature of theory, laws, data, etc. in these sciences and the implications for modelling (objectives, theory and data requirements, validity, and validation) will be investigated. The aim is to show that the multidimensional concept of model validity propounded in the theory of model validity (Chapter 4) has meaning for models in the social sciences and that this can be operationalised in model validation, a general methodology for which will be outlined. The approach adopted effectively widens the conventional meaning of validation to include all the scientific considerations involved in deciding to pursue or to drop a research programme based on a model (or series of models) and does not equate validation simply with checking against empirical data. This latter view is regarded as a false positivistic reconstruction which does not apply either to the social or physical sciences (see Chapters 3 and 4).

The validation methodology, which is proposed, is not a strict series of tests that must be followed exactly, but rather indicates the range of possible considerations that may be made about models which may be at many different levels of advancement, and is based on the ϵ-methodology described in Chapter 5. It is then used with the conceptual framework provided by the theory of model validity to consider the validity and validation of some illustrative areas of mathematical modelling in the social sciences: econometric modelling; world modelling; dynamic modelling of bicommunal political systems, and the analogical use of models (as in organismic models). These types of models may be regarded as "systems models" since they share the following features: (i) synthesis of analytical understanding in order to cope with complex interactions; (ii) an interest in global as well as local properties of the subject matter, (iii) frequent use of borrowed modelling techniques, concepts, ad models; and (iv) their consequent inter-disciplinary nature. (To some extent, all models organise their source

material into a system - more precisely, a type of system - and the close relationship between the use of models and system epistemology is also considered).

Unfortunately, all of these features make model validity problematic (in addition to all the problems of theory and data in the social sciences) and, unless the model research programme includes at least some validation stages, the models are open to a great deal of criticism (as in world modelling). These problems are mitigated by providing guidelines for model validation which will help speed the choice and application of appropriate empirical tests, theoretical considerations, etc. thereby maximising the scientific acceptability of a model whilst working within fixed modelling resources.

An additional aim of this Chapter is to reduce the divide between "hard" and "soft" systems science. The theory of model validity provides a conceptual framework, or common language, which broadens the understanding of modelling in both areas sufficiently to allow common dialogue. It is then possible to identify specific similarities and differences, rather than merely to state that differences exist, (in fact many of the supposed differences disappear). Eventually, this type of analysis could be deepened with the use of a theory of modelling based on detailed examination of the actual scientific use of models, in a wide variety of areas.

The structure of the Chapter is as follows: in § 9.2 some general features of theory, laws, data, etc. in the social sciences and their implications for modelling are examined. A methodology for model validation is proposed in § 9.3. In § 9.4 the methodology is applied to a number of models. Finally, in § 9.5, some concluding remarks are made.

9.2 Some Aspects of Laws, Theories, Models, and Data in the Social Sciences

The term "social sciences" refers here to the scientific study of social systems - their subsystems, institutions, and functionings - and includes psychology, sociology, economics, and political science. This section attempts to highlight some general, hopefully incontrovertible, features of laws, theories, and data in the social sciences, that have been selected for their relevance to modelling and model

validity. Discussion of specific issues of modelling in the social sciences is left until § 9.4, (modelling frequently entails working across conventional disciplinary boundaries, and general considerations such as presented here would appear, therefore, to be essential to a full understanding of the nature of modelling and for dealing with problems of model validation).

9.2.1 Observations on general features of laws, theory, and data in the social sciences

The observations will be made briefly under the following headings: (i) Laws; (ii) Theory; (iii) Data; (iv) Experiment; (v) Complexity; (vi) Self-reference; and (vii) Systems. In § 9.2.2 the implications for modelling are examined.

(i) Laws

In general there are no laws in the social sciences that have a universally accepted range of application (in contrast to physical sciences).

(ii) Theory

Without doubt, theory is the main form of expression in the social sciences. The majority of theories are expressed in an enriched natural language and only a few are mathematical (exceptions include theories in economics and psychology). There are often multiple, conflicting theories for a particular subject or system, and these may reflect different critical, political, or ideological views taken as a basis for inquiry.

(iii) Data

The problem of data in the social sciences is relating empirical information (and techniques) to concepts, variables, or parameters that occur in a theory or model. Many variables in social theories do not have a directly measurable empirical referent. Even in economics where many of the variables are physical (e.g. money, commodities, population, etc.) there is still a large gap between variables that occur in economic theory and those which are available as economic statistics, (econometrics is the study of mathematical models based on the latter). Much empirical work in the social sciences is concerned with discovering the structure or relationships in multidimensional data using correlational-

type techniques. Data-based methods, however, do not yield the theoretical understanding they promise. Models offer a stepping stone between theory and data. In addition, there are tremendous practical problems of data acquistion in the social sciences, e.g. uncertainty associated with spatio-temporal aggregation, inherent variability, use if data acquired for other purposes, etc.

(iv) Experiment

It is not generally possible to devise an experiment on a social system which will allow many factors to be controlled in order that the relationship between small subsets can be systematically examined. This experimental control is a feature of some parts, but by no means all, of the physical sciences and psychology. The implication is that the full complexity of the phenomena or systems have to be dealt with as a whole. Most theories in the social sciences are therefore general, or global, and this makes the theory - data distance even larger.

(v) Complexity

Social systems are highly complex - they possess a great number of subsystems which interact in a very large number of ways. Attempts to link events in an orderly manner are often confounded by variable dynamic feedbacks which cannot be controlled in an experimental situation. Social systems appear, therefore, to be highly recursive, although they are not unique in this respect (e.g. biological organisms). Despite the real complexity, the study of social systems is made more tractable by the division into arbitrary disciplines (justified, in part, by the high degree of commonality).

(vi) Self-reference

An additional difficulty, in the study of social systems, is that the investigators often form part of the system (or similar system) that is being investigated. Thus the values, beliefs, etc. held as a consequence of belonging to a particular society may affect the understanding of that and other societies. Regardless of problems of objectivity, it is clear that self-reference in the study of social sciences is closed, i.e. social systems affect social theories which in turn may profoundly affect social systems (e.g. Marxism, economic theory, Freudian psychology).

(vii) Systems

The word "system" effectively means "assembly of things" but without the physical restrictions this entails. It is never really used scientifically on its own, but with at least one adjective, as in "social system" or "dynamic system" or "non-linear dynamic social system". The combined term designates a combination of local and global structural and/or functional properties and is essentially a theoretical concept. As soon as the intended range of application ($\mathcal{R}_I$) of a theory or model is described as a "Σ-system", say, an entire body of knowledge, theories, models, etc. is invoked. It is important to realise that, except in trivial instances, the identification of $\mathcal{R}_I$ with a Σ-system has the same epistemological status as the identification of $\mathcal{R}_I$ with a model or theory.

9.2.2 Implications for modelling: objectives, validity, validation

On the traditional view of modelling (i.e. a model derived from a theory, or fitted to data) the problematic nature of theory and data (as compared to the physical sciences) would seem to make modelling in the social sciences a hopeless case. However, this view of modelling completely ignores the extremely varied roles that models may play in the evolution and justification of scientific knowledge claims in both the physical and social sciences. Nevertheless, the nature of theory and data in the social sciences means that model validity cannot be equated simply with theoretical coherence or with empirical correspondence. The objectives, or intended uses, of models in the social sciences are more concerned with the developmental aspects of theory construction or empirical research and may include the following:

(i) The structuring of a fuzzy intended range of application ($\mathcal{R}_I$). This may reveal theoretical and practical inadequacies.

(ii) Stimulating empirical research in a direction which is regarded as "fruitful".

(iii) As a tool for theory development, or testing the consistency of competing theories. (A typical example is the extension of a verbal theory into a mathematical model).

(iv) As a concise means of communicating ideas, or empirical results.

(v) In solving a model, by analytical or simulation techniques, under various conditions, pseudo-experiments can be performed on the model that could not be performed on the real system (e.g. the investigation of hypothetical or "what-would-have-happened-if" worlds, similar to the use of counterfactuals in historical analysis).

(vi) To provide a picture of the system which gives an explanation as a whole (i.e. an analogical construct).

(vii) Practical or utilitarian objectives. Despite the difficulties of theoretical and empirical validation models are used in a practical situation to modify some system of interest (e.g. models for prediction in econometrics, the rich picture - root definitions - conceptual model schema for "real-world problem solving" in soft systems methodology).

The four external validity criteria identified in the theory of model validity in Chapter 4 were empirical correspondence, theoretical coherence, pragmatic value, and heuristic potential. Except in a few cases (e.g. econometric models) the first is inappropriate as single criterion. The relative importance of the four criteria in determining the validity of a model will depend very much on the modelling objectives and the content and stage of development of the particular domain (field of research) associated with the model. In general, however, the most important set of criteria for model validation in the social sciences are heuristic. These are concerned with the evaluation of the potential of a model for discovery, developing scientific understanding, encouraging empirical research, etc. These criteria are related to "good reasoning patterns" in science and many can be explicity stated. In the next section a general methodology for model validation is proposed which is based mainly on heuristic criteria (which are described in detail) but also includes tests based on empirical, theoretical, and pragmatic validity criteria.

9.3 A General Methodology for Model Validation in the Social Sciences

9.3.1 Nature and scope of the methodology

The general methodology for model validation outlined in this section is a flexible programme capable of dealing with the validation of models of many different types in the social sciences. Although the emphasis is on validation tests associated with heruristic potential (as in the ϵ-validation methodology, § 5.6), the methodology contains

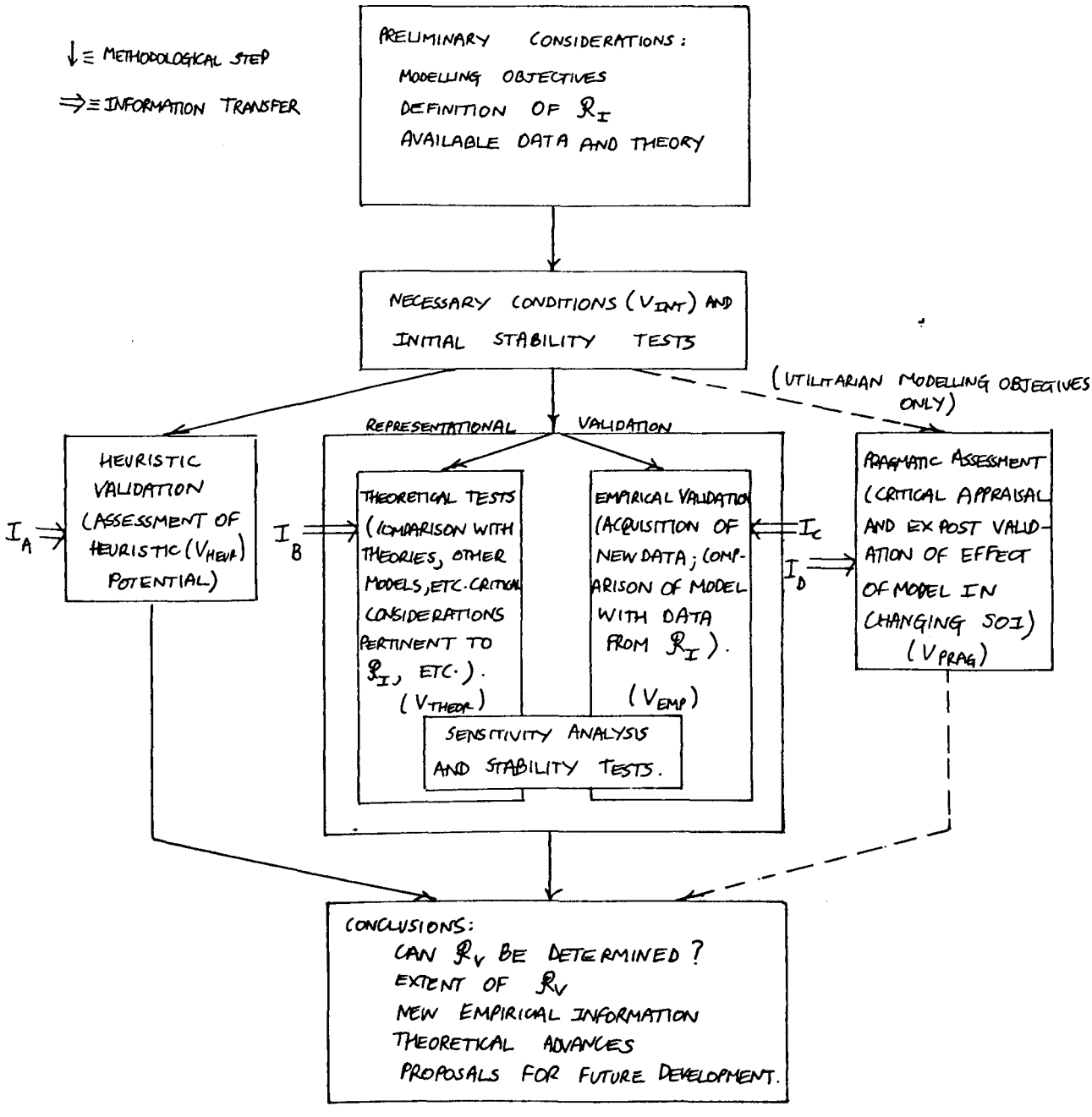
the available repertoire of assessment devices. It is based on the theory of model validity (which elaborates the relationship between modelling objectives, the nature of data (and domain development), and validity criteria, developed in Chapter 4) and on the ϵ-validation methodology for innovative models described in Chapter 5. Many of the stages of the methodology would be performed in the process of model formulation and development and therefore, it is best to regard it as embedded in the overall modelling methodology rather than simply as a final assessment procedure (although it may be very effectively used for this purpose). The necessary conclusions to this view are that the validity or acceptability of a model will depend to a great extent on the methodology used to construct it, and that validation will be often a process of methodological criticism (e.g. the critique of "The Limits to Growth" by Cole et al., 1973; see § 9.4.2). In § 9.4, the validation methodology is used to consider the validity and validation of models in four illustrative areas of model-based research in the social sciences.

9.3.2 Outline of the validation methodology

The validation methodology is depicted in fig. 9.1. Each box in the diagram represents a different validation stage and the systematic logical progression of validation proceeds down the page. However, in practice the results of a particular stage of validation may be used to alter the model (or even more general aspects such as modelling objectives or the intended range of application, $\mathcal{R}_I$) and in principle forward or reverse methodological steps may be made between any stages (this is discussed more fully in Chapter 4, § 4.4.1). Each stage of the methodology is discussed separately below.

9.3.2.1 Preliminary considerations

The first stage of the methodology consists of a detailed analysis of the modelling objectives and the available data and theory. This stage should clarify how well the intended range of application ($\mathcal{R}_I$) is understood. If a model is highly innovative, $\mathcal{R}_I$ may be possibly only well defined with the help of the model and validation tests will have to be more stringent than usual. The appropriate validity criteria and validation tests and their relative importance should be identified at this stage. For instance, a model that is based on a domain (field of research) that is empirically oriented and which is to



I_A = GENERAL INFORMATION FROM SCIENTIFIC DOMAIN (PROBLEMS, RESEARCH AREAS, THEORIES, MODELS, DATA, ETC.).
I_B = ACCEPTED (OR POSSIBLE) THEORIES OR MODELS OF $\mathcal{R}_I$
I_C = EXISTING AND NEW DATA FROM $\mathcal{R}_I$
I_D = DATA FROM SOI ON EFFECT OF MODEL

FIGURE 9.1. A GENERAL METHODOLOGY FOR THE VALIDATION OF MODELS IN THE SOCIAL SCIENCES.

be used for utilitarian objectives (e.g. econometric models, see § 9.4.1) will be mainly concerned with empirical validation (§ 9.3.2.3.1) and pragmatic assessment (§ 9.3.2.5). Differences between approaches emerge clearly in the preliminary considerations (e.g. critical theorists taking the heuristic validation/ theoretical tests route). Other aspects that may be considered at this stage which are important at later stages include: identification of theoretical problems associated with the domain (e.g. apparently inconsistent theories); and areas of $\mathcal{R}_I$ in which available data are uncertain, inadequate, or unavailable.

9.3.2.2 Necessary conditions and initial stability tests

The necessary conditions are the satisfaction of internal validity criteria of consistency and algorithmic/simulation validity (see Chapter 4), which are self-evident and easily checked. In a complex model simulated on a computer it is very important that these tests are made since obvious mistakes can be easily hidden and there is a tendency to trust even the strangest results from a large model. The initial stability tests apply to a mathematical model or computer simulation and require that the model be reasonably stable in preliminary solutions or simulation runs. Stability tests require careful consideration in models of social systems since it is possible that $\mathcal{R}_I$ is not itself fundamentally stable (see § 9.3.2.3.3).

9.3.2.3 Representational validation

Representational validation consists of the application of empirical (§ 9.3.2.3.1) and theoretical (§ 9.3.2.3.2) criteria. Detailed stability and sensitivity tests are also included (§ 9.3.2.3.3), and these may be considered to apply both empirical and theoretical validity criteria (see fig. 9.1). This stage of the methodology is concerned with the extent to which the model represents the intended range of application, $\mathcal{R}_I$. The empirically validated range of application, $\mathcal{R}_V$, of the model is the portion of $\mathcal{R}_I$ which has satisfied the appropriate empirical tests, ($\mathcal{R}_V$ may be described as the "empirical support" of the model).

9.3.2.3.1 Empirical validation

Empirical validation is the comparison of the model (form, function, and behaviour) with empirical data from $\mathcal{R}_I$. Closely associated with these comparisons is a programme of empirical research

aimed at extending the empirical information available from $\mathcal{I}_I$. Quite often this programme will be structured with the aid of the model (e.g. in devising new data types). Empirical validity criteria require that the model and data should match to within data uncertainty over the spatio-temporal resolution of $\mathcal{I}_I$. A suitable methodology for empirical validation, which proceeds from qualitative comparisons to feature analysis to time series comparisons and parameter estimation, is the α-methodology developed in Chapter 5 (§ 5.2). In the long-term it is the occasional touches of a research programme with reality through empirical validation that provides an epistemological basis.

Typical problems of empirical validation of models in the social sciences are associated with the generality of models and the inadequacy of available data. A model which describes a social or political situation in general may have to be augmented by additional hypotheses, initial conditions, constraints, etc. in order to represent a particular instance of the general situation. These additions are necessary in order to validate empirically such a model. There may be insufficient data for validation for both practical and theoretical reasons. As discussed in § 9.2.1, many variables (concepts) used in theories and models in the social sciences cannot be measured, and a large part of the empirical validation of many models will be concerned with the extent to which operational techniques can be developed for the measurement of previously unmeasurable variables in the model.

It is possible to disassemble some models such that each submodel may be validated separately, and the overall validation can be structured hierarchically by the gradual reassembly of the model. This allows a clear identification of the areas of validity and uncertainty in the model. However, many phenomena in social systems are highly recursive and interdependent, and this approach may not always be feasible. In this case, empirical validation can only be performed on the model as a whole and this makes its validity even more problematic.

One of the strongest tests of empirical validity occurs when a prediction of the model subsequently takes place. This is particularly effective when the prediction is of a rare or unusual event.

9.3.2.3.2 Theoretical tests

Theoretical tests consist of checks of model coherency with existing theories or models, and other critical considerations.

One of the difficulties is the existence of multiple, conflicting theories that occur in the social sciences. Occasionally, a model may allow such conflicting theories to be integrated, but in general the model has to be compared with the theory that is considered "best". The enforcement of theoretical validity criteria depends on how well established current theories are, a point which should emerge from the preliminary considerations (§ 9.3.2.1). If a model does not cohere well with existing theories it must be validated in other ways - i.e. by empirical (§ 9.3.2.3.1) or heuristic (§ 9.3.2.4) validation.

9.3.2.3.3 Stability and sensitivity analysis

Sensitivity analysis is concerned with the extent to which global properties of a dynamic model are retained despite small variations in parameter values, initial conditions, structure, etc. A full description of the available techniques and their uses in model validation is given in Chapter 5 (§ 5.2.4.2). Some additional comments will be made here on the use of these techniques in validating models in the social sciences. Frequently, because of data and theory problems, sensitivity analysis is the main focus of model validation (see § 9.4.3).

Most models in the social sciences are highly nonlinear. Consequently, the sensitivity of the model to one parameter (or initial condition, etc.) is often highly dependent on at least one other parameter in a way that fundamentally cannot occur in a linear system. Many sensitivity analysis techniques (e.g. analytic solutions of Tomoric's sensitivity equations, or systematic perturbation methods) vary one parameter at a time from its nominal value and therefore may not always expose the dramatic changes than can occur when two or more parameters change slightly together. One-by-one parameter variation methods steer blind orthogonal courses through the rich diversity of nonlinear phenomena. A similar criticism can be raised against methods of uncertainty transmission through nonlinear systems based on first-order Taylor approximations (and the Jacobian matrix) or perturbation techniques. Ideally, the sensitivity of global properties (such as stability) are determined by topological methods, but this is rarely possible in complex models. An alternative solution is to use Monte Carlo model simulations (see § 5.2.4.2) or analytic

techniques in which a set of parameters may be varied simultaneously.

Another problem is whether the global property, taken as an indicator of the model's validity (theoretical or empirical), actually occurs in the intended range of application ($\mathcal{R}_I$). For instance, if it is assumed that a certain socio-economic structure is stable, this may imply a conservative, or regularative, social theory which may be criticised. Global, or "systems", properties of a model and $\mathcal{R}_I$ should not be accepted as a prior assumption but should be the subject of both theoretical and empirical investigation. If a model has received adequate theoretical and empirical validation over $\mathcal{R}_I$, and contains certain systems properties (e.g. stability, oscillations, catastrophes, etc.), these properties may also be ascribed to $\mathcal{R}_I$. Conversely, a critical test of a model is in matching its global qualitative properties to those of $\mathcal{R}_I$. Since models are the most effective means of studying complex interactions, they therefore also provide an epistemological basis for the investigation of systems in systems science.

9.3.2.4 Heuristic validation

Most social science models have neither a complete theoretical nor an adequate data base against which to be validated. The main assessments of these models are made in heuristic terms, i.e. the extent to which a model is a tool for understanding and discovery. Heuristic considerations form part of the day to day activities of all scientific research and are, in effect "good scientific reasoning patterns". Whilst there are a very large number of good heuristics, it is nevertheless possible to identify a small number of criteria to determine the heuristic potential of a model which are fairly objective. These criteria recognise that a model forms part of a series of models, or an ongoing research programme. The following criteria may be used to test whether a model has heuristic potential:

(i) The model has an increased range of application over its predecessor, i.e. $\mathcal{R}_I(k) \supset \mathcal{R}_I(k-1)$

(ii) The empirical support of the model is greater than its predecessor, i.e. $\mathcal{R}_V(k) \supset \mathcal{R}_V(k+1)$

(iii) The model helps improve the understanding of the phenomena of interest. (This is difficult to specify, probably because

the nature of scientific explanation is not well understood. A helpful way to approach this, however, is to examine the range of questions which can be posed and answered with the model. In some senses, a model imparts an understanding of the system as a whole. Additional considerations may be of the theoretical elegance or conceptual simplicity of the model).

(iv) The model allows the examination of hypothetical, "what-would-have-happened-if-worlds"; a pseudo-experiment; counterfactuals.

(v) An empirical research programme is initiated, sustained, and structured as a consequence of the model. (An example is when a theoretical concept is expressed as a model variable which is then measurable).

(vi) The model allows a clearer definition of $\mathcal{R}_I$, (i.e. it provides a good conceptual model, see (iii)).

(vii) An outstanding problem in the domain (field of research) is solved by the model (e.g. some previously unexplained phenomena, or conflicting theories).

(viii) The model leads to a "better" set of problems and new modelling objectives (i.e. the model may fail tests of representational validity yet suggests new directions for research).

At first sight it seems that a model should satisfy heuristic criteria positively and not negatively (e.g. as a consequence of the critical outcry against a very bad model, see § 9.4.2). However, in the long run, bad models are probably fundamental to the dialectic of science.

9.3.2.5 Pragmatic assessment

The pragmatic validity of a model (with utilitarian objectives) for use in a wider system of interest (SOI) may be assessed in critical terms, or by ex post validation after the model has been used and data are available on its effect (see Chapter 5, § 5.5). Examples of models with utilitarian objectives occur in econometrics (§ 9.4.1) and soft systems methodology (§ 9.4.4). Although it is possible to validate this type of model purely pragmatically, without regard to scientific validity (heuristic and representational), the view taken here is that scientific validity is essential for successful utilitarian application.

9.3.2.6 Concluding the methodology

The validation methodology forms part of the overall modelling methodology of a model. In many areas of the social sciences theories models and data are in a developmental stage (e.g. world modelling, conflict dynamics) and a final, definitive conclusion on the validity of a model may be inappropriate. However, a summary of the results could be produced that includes the following points:

(i) Whether the empirical support, $\mathcal{R}_v$ (the empirically valid range of application), of the model can be determined and, if so, its extent.

(ii) New empirical information available, (including new data representation devices).

(iii) Theoretical advances.

(iv) The appropriateness of modelling and proposals for future development.

9.4 The Validity and Validation of Models in Four Illustrative Areas of Modelling in the Social Sciences

The four areas of modelling in the social sciences that have been selected for illustration are econometrics (§ 9.4.1), world modelling (§ 9.4.2), conflict dynamics (§ 9.4.3), and the use of analogical models in general (§ 9.4.4). The aim is to analyse the different models within the framework of the validation methodology (§ 9.3) and, in some cases, to suggest additional validation tests that should be performed. (Each area will be discussed only in sufficient detail to introduce aspects of model validity and validation and is not intended as a balanced introduction to the subject area).

9.4.1 Econometric modelling

Econometric models were originally devised in order to test economic theories (such as Keynes'). Today, however, they are intended primarily for utilitarian objectives - the control or management of an economy, avoidance of economically unsatisfactory situations (e.g. inflation), etc. and do this by providing predictions of the state of an economic system beyond the present time. An econometric model is so called because the variables that occur in it are all measurable, and consists of a set of simultaneous difference equations (relating

groups of variables) and a large data base. The equations are fitted to the data using statistical techniques, and the process is repeated continually as the data base is enlarged (a good reference on econometrics is Desai, 1976). Econometricians generally assume that expanding the data base will eventually lead to models with predictive validity. The complexity of current econometric models and data bases can only be handled on large computers (e.g. the Treasury, Cambridge, and Wharton models). The range of application ($\mathcal{R}_{\mathcal{I}}$) of an econometric model may be the economic system of a town, region, country, or international trade, but in all cases the model will be a highly aggregated representation of the economic events that occur. The system of interest (SOI) is the economic system ($\mathcal{R}_{\mathcal{I}}$), of course, but in addition it includes the political and social systems that are highly important in any economic decision.

The objectives and empirical basis of econometric models suggests that validation should focus on empirical tests and, to some extent, pragmatic assessment (see fig. 9.1). Econometric models have received a great deal of empirical validation using statistical techniques (e.g. Naylor et al., 1966; Young et al., 1973; Desai, 1976) to the extent that a good fit to existing data is often considered to be the sole criterion of validity (Friedman, 1953). The predictive validity of a model may be assessed by comparison with historical data, and with new data when it becomes available. This approach hinges on the false assumption that a data base can be built which will eventually contain details of the range of all possible economic behaviour.

An alternative approach in empirical validation is to examine the ability of econometric models to reproduce key features or qualitative events ("turning points" , House, 1974) in historical data as well as having good overall statistical fit. This can establish that the relationships, or mechanisms, in the model actually occur in the economic system and are not arbitrary equations which happen to reproduce data (Berlinski, 1976). The econometric model then becomes effectively a tool for theory development. The insight provided into a critical event may be far more valuable than the accurate prediction of a period of economic calm. Although there is an increasing emphasis on these other aspects of validity (theoretical and heuristic) in econometrics (e.g. House, 1974; Desai, 1976),

econometric models have not yet generated new theories which allow complex economic systems to be better understood. (In fact, most econometric models are implicitly based on Keynesian economic theory). (By contrast, meteorological modelling, which has a similar structure - large data bases, simultaneous finite difference models, a complex global data acquisition network - has always paid attention to the use of large models for theory testing and development, and new theories, particularly of global climatological patterns, have emerged; Monin, 1972). The knowledge that a model embodies a good understanding of the economic system also increases the confidence in the empirical validity of predictions beyond the validation interval.

One of the main problems in the theoretical assessment of econometric models is that of aggregation; the extent to which variables may be lumped together in global models or theories (macroeconomics) yet still satisfy small scale theories (microeconomics). Leontief derived conditions (in terms of the marginal rates of substitution of economic variables) under which aggregation is valid, however this can be shown to limit permissive functions to the linear kind only (Green, 1964). Aggregation problems occur in many models in the social sciences (e.g. world models, § 9.4.2).

To sum up: econometric models are data-based and the main thrust of validation is empirical (statistical comparisons). However, theoretical tests are important and this is reflected in the growing emphasis in empirical validation on the reproduction of critical turning points which provide good evidence of the validity of the funcitonal relationships or mechanisms in a model. On the whole, econometric models are too large and institutionalised for heuristic validity to be an important factor for acceptability although, in principle, they do offer tremendous scope for empirical research and theoretical development (two important heuristic criteria).

9.4.2 World modelling - "The Limits to Growth"

In response to the Club of Rome's "Project on the Predicament of Mankind", a team of scientists led by Professor Meadows of MIT developed a world model in an attempt to investigate the likely changes of the world's population, capital, agricultural and natural

resources, and pollution over the next 100-150 years. The model (known as "World 3") was based on that of Forrester ("World 2", 1971). The findings of Meadows and his group were first published, strangely, in a popular account, "The Limits to Growth" (Meadows et al., 1972), in which the dramatic growth and eventual collapse of the world's population before the year 2100 was dogmatically predicted. The full technical report (Meadows et al., 1974), in which the results of a wider range of simulation runs and scenarios were presented, was rather less dogmatic and more cautious but nevertheless predicted the same "behaviour mode" of eventual collapse.

Both World 2 and World 3 were formulated using Forrester's "Systems Dynamics" technique (Forrester, 1968), which is basically a way of constructing first order differential equations that allows multiplicative interactions of variables. World 2, which Forrester developed on the basis of his own common sense insight (Forrester, 1971), consists of five sectors: population, capital (economics), agriculture, natural resources, and pollution. World 2 was not based on currently accepted theories or data. Meadows' World 3 has the same basic structure as World 2 but is much more elaborate containing about three times as many mathematical equations and many of the numerical relationships are estimated for empirical data (Meadows et al., 1972, 1974). Although the complete model is complicated, many variables in the model represent gross aggregations (such as developed and third world economies) and many of the submodels are highly elementary (e.g. the population dynamics).

On the basis of historical data (from 1900-1970) both models may be run forward in order to project current trends - the so-called "standard run" (Meadows et al., 1972, p.124). This run "assumes no major change in the physical, economic, or social relationships which have historically governed the development of the world system" (Meadows et al., 1972). World 2 and World 3 predict a massive growth in the world's population and pollution followed by a severe collapse before 2100. In "The Limits to Growth", the standard run is followed by a series of modifications to World 3 which attempt to produce a stable equilibrium without collapse. The measures which are apparently required to avoid the prospect of doom include a drastic reduction of the birth rate and a cessation of industrial growth.

The two world models have received a great deal of criticism since they were originally published. This was particularly fuelled

by the dogmatic, assertive, pseudo-scientific claims of "The Limits to Growth". The best collection of criticism is the book edited by Cole, Freeman, Jahoda, and Pavitt (1973) in which the model (World 3) is assessed from technical, theoretical, and ideological viewpoints. Berlinski (1976) criticised the mathematical basis of systems dynamics in terms of the theory of dynamical systems. Rather than review the main criticisms of the world models, World 3 (i.e. the model developed by Meadows et al. and reported in "The Limits to Growth") will be assessed in terms of each stage of the validation methodology outlined in § 9.3. The five stages to be considered are: (i) necessary conditions; (ii) empirical validation, (iii) theoretical tests; (iv) sensitivity analysis, and (v) heuristic assessment.

(i) Necessary conditions

The consistency validity criterion requires that a model should contain or entail no contradictions and is usually satisfied very simply, however Julien, Freeman, and Cooper (in Cole et al; 1973, Chapter 6) showed that the capital submodel contained many internal inconsistencies which were partly responsible for the overshoot and collapse response. No errors arose in the simulation of the equations (i.e. algorithmic/simulation criteria) using an Euler system since all the time constraints in World 3 are long. However, if fast dynamics are included an Euler system will probably result in errors.

(ii) Empirical validation

Meadows et al. (1972) use the response of the standard run from 1900-1970 as evidence for the empirical validity of the model. Since there are no constraints on time in the model it should also be capable of predicting backwards in time but, as Cole and Curnow (Cole et al., 1973) demonstrated, the original model shows a catastrophic collapse from an infinite population in 1880. Furthermore, if the model run is started 30 years earlier, say, the date of the eventual collapse is brough forward by 30 years. If the model was started in 1850, the collapse would occur in 1970. These results show that the model is not capable of reproducing periods of historical data in general and its good fit to the 1900-1970 period in the standard run is purely fortuitous.

(iii) Theoretical tests

While the model fails to satisfy empirical validity criteria, its emergence from theoretical tests is even worse. The two submodels which are most strongly attacked are those representing population and capital. The population submodel basically follows Malthus' ideas on population expressed in 1830 and is not based on currently accepted demographic theories; in particular, it omits many important social feedback loops on the control of birth rate. Similarly, the capital submodel does not make full use of empirically validated economic theory and is very rigid. The aggregation of the developed and third world countries in the model eliminates the possibility of investigating many important crisis situations (e.g. the oil crisis). Other areas of the model that are theoretically inadequate are the pollution and available resources sectors. In addition, technological change is treated as an exogenous factor yet this has tremendous impact on the model's behaviour. The theoretical inadequacies can be traced largely to the omission of social and political processes in the model which have a profound impact on the world system.

In theoretical terms, therefore, the model is invalid. This could be justified if the model had passed some fairly stringent empirical tests (which it hasn't) or if it offered tremendous heuristic potential.

(iv) Sensitivity analysis

Meadows et al. (1972) offer additional confirmation of the basic behaviour mode in a series of sensitivity tests in which certain parameters (such as available natural resources) are varied by large amounts. Despite these changes the model still exhibits the overshoot and collapse pattern. It is interesting that this feature is cited as evidence of the validity of the model (Meadows et al., 1972, p.121). Usually, if a model of a dynamic system which has been evolving (but relatively stable) exhibits a rapid instability it is rejected as invalid. Cole and Curnow (in Cole et al., 1973) repeated the sensitivity tests and in addition simulated the model with a number of parameters changed simultaneously by small amounts. In some of these latter tests parameters were also varied slowly with time in order to introduce adaptive social and technological feedbacks into the model. Under these circumstances the

model no longer exhibits the overshoot-collapse behaviour.

In effect, the standard run of World 3 is driven hard into growth and early collapse by the rigid capital and population submodels. If these submodels are made slightly more elastic, or adaptive, by relatively small parameter changes the model will run for centuries without catastrophe. Sensitivity analysis reveals the dogmatic assertions on the impending catastrophe to be without foundation, in terms of the model itself.

There is a lesson here for the sensitivity analysis of nonlinear models. Tests in which separate parameters are individually varied (even by large amounts) at the beginning of a run will not in general reveal the complete range of the model's behaviour. It is therefore very important to perform sensitivity tests in which more than one parameter is varied in each run and, if possible, to investigate the effect of time-varying parameter changes.

(v) Heuristic assessment

Having failed all the other validation tests, the only possible redeeming feature for World 3 is that of heuristic validity. Unfortunately the story is the same. The gross aggregation of the model virtually precludes interesting theoretical development and does not help understanding of the real physical, social, and political issues involved in global dynamics. Although the model requires a great deal of data this is at such an aggregated level which makes it useless for any further research.

As it stands, World 3 fails every stage in the validation methodology of § 9.3. Paradoxically, the very badness of the model has generated a tremendous critical response which has exposed many of the problems inherent in world modelling and indicated legitimate directions for future research. "Thinking about the Future" (Cole et al., 1973) is a far more valuable and scientifically acceptable book than "The Limits to Growth" (Meadows et al., 1972), but it would never have been written if the latter had not been published in the first place.

9.4.3 Modelling bicommural conflict

In this section a model of bicommunal political system is examined. The model, which was developed in the Department of Systems Science, has a triadic structure consisting of submodels of the dominant community, the economic subsystem, and the dominated community

(Not , Mitchell, and Janes, 1974). Recent work has focussed on the dominated community, the community that is assumed to be less powerful with respect to numbers, organisation, and mobilisable resources (Bowers, Mitchell, and Webb, 1978,1979,1980). The model of the dominated community was formulated mathematically using Forrester's technique of Systems Dynamics (Forrester, 1968). The main objectives of this work have been: (i) to observe the dynamics of conflict, (ii) to identify "opt-out" points on the conflict spiral; and (iii) to investigate the potentialities of dynamic systems analysis in a complex area with high recursivity (Bowers, Mitchell, and Webb 1979a). (Detailed description of the model, including its theoretical basis, is given in Bowers et al., 1979a, the results of the initial simulation runs in 1978, and a readable summary of the work in 1979b).

The present stage of development of the model of the dominated community will now be assessed using the validation methodology of § 9.3 (fig. 9.1). As with the world model (§ 9.4.2) the following aspects are considered: (i) necessary conditions, (ii) empirical validation; (iii) theoretical tests; (iv) sensitivity analysis; and (v) heuristic assessment. Each corresponds to a different concept of validity (see Chapter 4).

(i) Necessary conditions

Clearly the model contains no obvious contradictions,however it is very difficult to exhaust all the possibilities of inherent inconsistencies. The simulation of the model's equations uses an Euler system (Bowers et al., 1979, p.15, following Forrester, 1968). Given the nature of conflict (e.g. uncontrolled positive feedbacks) it is likely that rapid dynamics may suddenly occur. In a linear system the shortest time constants can be determined and the Euler step chosen accordingly. However, in nonlinear systems, the dynamics are functions of the state variables (especially using the multiplicative techniques of systems dynamics) and extremely fast dynamics may occur (as in catastrophe). Under these circumstances there is no guarantee that an Euler system will reproduce the full range of nonlinear behaviour and it may lead to cummulative integration errors. Some topological considerations of the mathematical equations should be made to check the nature of likely structural change. (These are minor points, but clear omissions).

(ii) Empirical validation

The model is effectively a development of a number of theories of political conflict into mathematical form. Consequently, many of the model variables were not previously measurable (i.e. on a numerical scale). Empirical validation for such a model does not consist simply in the comparison of the model with data (from a specific conflict situation) but includes the empirical research which attempts to operationalise the measurement of model variables; in effect, the model shapes the empirical investigation. The lengthy report of Mitchell and Webb (1978) demonstrates that empirical references for the model variables can be defined and scales and indices for their measurement constructed (some are admittedly subjective). An interesting outcome of this work was in the measurement and meaning of political ideology determined from paragraph analysis of documents of the Scottish National Party using graph theory (Farbey, Mitchell, and Webb, 1979). This introduces a new data type (i.e. symbolic relational structure) and appears to have general application to the measurement ideology from documents as well as in the semantic analysis of domain-specific documents (Farbey et al., 1980).

The comparison of the model's response with data collected from a conflict situation (using the new measures) has not yet been made. The model is a general model of bicommunal conflict and does not represent a specific conflict situation. In order to do so it will have to be augmented with additional submodels and specific constraints of the situation will have to be applied. An appropriate data base of time series data is also required. When the comparisons are eventually made emphasis should be placed on the reproduction of important qualitative features (such as conflict growth) before statistical comparisons. If the empirical validity is established in a number of cases (e.g. Scottish nationalism, the Basque Community in Spain, and the Flemish community in Belgium) and if the number of additional submodels required in each is not too large, then it will be legitimate to claim that the model has some general empirical validity. It is highly unlikely, of course, that the model will predict the outcomes of real conflict situations very accurately because it is deterministic whereas social and political processes have an obviously stochastic nature.

(iii) Theoretical tests

Whilst World 3 paid almost total disregard to accepted theories (§ 9.4.2), systems dynamics was used in bicommunal conflict modelling as a means of extending verbal political theories into mathematical form. Consequently, much attention was given to the coherence of the model with theories of political conflict. The model expanded the theoretical description of conflict into a number of levels such that theories which had previously been regarded as incompatible or contradictory were resolved and integrated within the framework of the model (Bowers, Mitchell, and Webb, 1980).

(iv) Sensitivity analysis

Most of the validation tests carried out on the model have concerned sensitivity analysis. These have involved changing a parameter value, initial condition, etc. and observing the effect on the stability of the model. The difficulty of interpreting sensitivity analysis results in the context of models in the social sciences has been discussed earlier in this Chapter. The problems range from whether or not the subject of the model, $\mathfrak{P}_I$, is inherently stable to the incredibly wide range of sensitivity tests that can be performed on nonlinear dynamic models.

(v) Heuristic assessment

A number of heuristic validity criteria were described in § 9.3.2.4. For an innovative or highly developmental model these criteria are by far the most important (see Chapters 4 and 5). In the physical sciences models eventually converge to correspondence with empirical data, which then becomes the most important criterion. Given the complexity of social and political systems it seems unlikely that this convergence will take place during the lifetime of an individual model. Throughout the many reports on bicommunal modelling there is an underlying theme that the results must be judged heuristically by the insights they afford into conflict situations. However, there is a tremendous asymmetry between the formal, precise treatments of modelling, the sensitivity tests, and measurement and the rather ill-structured assessments of heuristic validity or potential. Contrary to the implicit assumption, heuristic criteria are not vague but can be given precise expression (§ 9.3.2.4) which allows

a fairly objective assessment of the heuristic validity of a model. The following aspects will be considered: (a) theoretical advance; (b) empirical support, (c) understanding; (d) problem shifts; (e) empirical research; (f) new directions for research.

(a) Theoretical advance

The model allows a clearer definition of $\mathcal{R}_I$ and extends the resolution with which it may be described theoretically; i.e. it is a theoretical elaboration (this includes the mathematical expression of verbal theory) and advance.

(b) Empirical support

Although the empirically valid range of application (or empirical support), $\mathcal{R}_V$, has not yet been determined, substantial progress has been made towards the operational procedures for measuring model variables which will soon make this possible.

(c) Understanding

In vague terms it can be said that the model advances the understanding of political conflict situations. More precisely, it offers an increased potential for explanation by offering an expanded repertoire of questions which may be posed and answered concerning such situations. In jargon, more " counterfactuals" and their consequences can be examined than with previous theories of conflict.

(d) Problem shifts

Prior to the model a perceived problem in the domain of political conflict was the incompatibility of existing theories. In integrating a number of theories, the model effectively solved this problem.

(e) Empirical research

In operationalising the measurement of variables, the model helps to structure empirical investigation of dynamic conflict situations. Particularly successful results were associated with the use of graph theory in the measurement of ideology.

(f) New directions of research

In addition to further validation tests, the model is to be used to investigate what combination of conditions are necessary

to stimulate a mass-movement in a conflict situation. An attempt is to be made to reduce the submodel of the dominated community and to simulate it with the dominant and economic system submodels (Bowers et al., 1979b).

In conclusion, the model emerges positively from the heuristic and theoretical tests in the validation methodology. Although empirical validation has still to include empirical comparisons, results so far are encouraging. On another level, the conclusions appear to offer partial evidence for the applicability of systems dynamics techniques to problems in the social sciences. However, this undervalues considerably the interaction of scientific judgement and technical skills of the individual workers in the research team. Comparing the positive results of this model with the criticisms of the world models (§ 9.4.2), both of which use systems dynamics, it is clear that the use of this technique is purely incidental and that there are other more fundamental theoretical and methodological factors involved.

9.4.4 Analogical models

An analogical model is one whose subject is different from its source. This involves the transfer of a model (or modelling techniques) from one domain in which it has been developed to a new domain. A basis for the use of analogical models is the perception of theoretical and empirical similarities between the two domains. Although analogical modelling is an important feature of all science, it is perhaps most widely used in systems science, particularly when applied to social systems.

When an analogical model is introduced into a new domain it effectively gives a new explanation or theory as a whole. This type of model raises interesting problems of model validity and validation (and epistemology) which were discussed in Chapter 3 (§ 3.4.5). Examples of analogical models are control system models in biology and dynamic mathematical models introduced into the social sciences. At first it is unlikely that theoretical or empirical validity criteria could be applied to an analogical model (e.g. if the model's data type requirements exceeded the available data types in the new domain). Validation must proceed mainly heuristically (as with the conflict model, § 9.4.3) until appropriate measurement systems have

been developed. Gradually, if a model is accepted, it will be gradually changed and integrated into the domain. By the time it satisfies empirical correspondence tests it may have changed significantly from its original form and will no longer be analogical.

In systems science, analogical models are frequently used for utilitarian objectives (i.e. in improving, modifying or designing a system of interest). Checkland's soft systems methodology for "real world problem solving" (Checkland, 1972) suggests the use of organismic models such as that of Beer (1972) or models based on his "human activity system". According to Checkland these models must be assessed according to how they contribute to the solution of the problem and not as representations of an $\mathcal{R}_I$. In other words, he advocates pragmatic validity criteria alone. M'Pherson's "protosystem" (M'Pherson, 1980) is another example of an analogical model for problem solving. Whilst M'Pherson includes empirical as well as pragmatic validity criteria, heuristic and theoretical validity criteria, which are crucially important, are omitted. Checkland's and M'Pherson's methodologies concentrate so much on solving the real-world problem that the reason for the use of a model becomes obscured. Surely it is simply this: a model embodies scientific understanding or explanation, and understanding is a pre-requisite for effecting change.

9.5 Conclusions

After considering some general features of laws, theories, models, and data in the social sciences (§ 9.2.1), it was clear that validity could not be equated simply with empirical correspondence or theoretical coherence (§ 9.2.2). A validation methodology was presented which was based on the theory of model validity (Chapter 4) and included the full range of validity criteria (§ 9.3). This methodology is original in that an emphasis is placed on the importance of heuristic assessment for innovative models.

In § 9.4, the validation methodology was used successfully to examine the validity and validation of models in four illustrative areas of modelling in the social sciences. In particular, the methodology explained the inadequacies of world models (§ 9.4.2) and illustrated the potentialities of nonlinear dynamic mathematical modelling of bicommunal political systems (§ 9.4.3). These results may be taken as evidence that the multidimensional concept of model validity, and the conceptual framework of the theory of model validity, proposed in Chapter 4 has meaning for models in the social sciences. In addition, the validation methodology could be used to structure the validation of such models in the future.

This Chapter concludes the four case studies. The first three case studies (Chapters 6, 7 and 8) dealt with the validity and validation of three mathematical models in biology in considerable detail and demonstrated the applicability of the theory of model validity in that area. In developing the theory it was intended that it should apply to models generally in science (and engineering). The results of this Chapter confirm that the theory is also applicable to models in the social sciences, a particularly problematic area, which is an encouraging sign for its generality. In the next Chapter, some methodological, theoretical, and practical implications of the work reported in the last six Chapters on modelling and validation in systems science and biomedicine will be examined.

CHAPTER 10

METHODOLOGICAL, THEORETICAL AND PRACTICAL IMPLICATIONS FOR MODELLING AND VALIDATION IN SYSTEMS SCIENCE AND BIOMEDICINE

10.1 Introduction

The previous six chapters have presented a great deal of methodological investigation and detailed practical validation case studies. (In Chapter 4, a theory of model validity was developed which explicated validity as a multi-dimensional concept and provided a conceptual framework for the analysis of model validity and validation. This was used in Chapter 5 to devise methodologies suitable for the validation of models with a range of objectives and at different stages of development. Chapters 6, 7 and 8 consisted of the practical validation of three biological models, and, in Chapter 9, some aspects of model validity and validation in the social sciences were considered generally, and with reference to some illustrative areas.) In this chapter, the implications of the results of this work on methodological, theoretical, and practical aspects of modelling and validation in systems science and biomedicine will be examined. The aim is to show how research in both areas can benefit from, or be guided by, these results in particular, and this type of methodological study in general. Firstly, however, a brief summary will be made of the theory of model validity.

In the theory of model validity (Chapter 4), a valid model is defined as one which satisfies its modelling objectives. The latter are classified into scientific objectives (which are essentially concerned with the representation and understanding of phenomena) and utilitarian objectives (which are concerned with the creation or improvement of a system, solving practical problems, etc.). Validity criteria (or tests for model validity) are classified into internal and external criteria. Internal criteria consist of consistency and algorithmic (or simulation) criteria and are necessary prerequisites. External criteria are contingent (dependent upon factors external to the model, e.g. data or theories) and are further classified as follows: representational criteria, which consist of empirical criteria (correspondence with empirical data) and theoretical criteria (coherence with accepted theories or models); heuristic criteria (which examine the potential of a model for improved explanation, new research,

etc.); and pragmatic criteria (which are concerned with the extent to which a model can lead to better design of a system, solution of a practical problem, etc.). Each criterion may be equated with a different concept of validity; the most common notion of validity as simply empirical correspondence is replaced, therefore, by a richer multi-dimensional concept. The relationship between modelling objectives and appropriate validity criteria is articulated in the theory of model validity, together with the requirements for, and constraints of, empirical data (the theory contains a theory of data). Throughout the theory, the importance of the content and stage of development of the domain (field of research associated with a model) in affecting a model and its validity are emphasised.

The structure of the chapter is as follows: in Section 10.2 the possible benefits for systems science on a general methodological level are examined. Some more practical recommendations for modelling and validation in biology and medicine are made in Section 10.3. In effect this section summarises the general conclusions made in Chapters 6, 7 and 8. Finally, in Section 10.4, suggestions are made on the direction of future research programmes into model validity and validation.

10.2 Benefits for Systems Science

The term "systems science" here refers to the loosely associated areas known as "systems approaches to ....", "systems research", "systems methodology" (hard and soft), etc. The main features which characterise systems science include: (i) an attempt to deal with complex phenomena, systems, and problems (both natural and artificial); (ii) the development of a body of systems concepts and theories with wide applicability; (iii) a commitment to synthesis as well as analysis; and (iv) working across conventional disciplinary boundaries. Models (verbal and mathematical) provide a means of integrating knowledge, studying complex relations and investigating global (or system) as well as local properties and have therefore played an exceedingly important role in the development of systems science. Consequently, the results of this methodological study into model validity and validation have direct implications for systems science. These will be examined at two levels: firstly, the implications for validation of systems models in specific application areas; and, secondly, the general methodological implications for systems science as a whole.

Various methodologies for model validation have been devised (Chapter 5) and successfully applied to systems models in biology and the social sciences (the conclusions on the validities of the models vary considerably, however; see Chapters 6, 7, 8 and 9). These methodologies are based on the theory of model validity which offers an increased understanding of model validity. (Some specific recommendations for modelling and validation in biology and medicine are made in Section 10.3). Systems models are invariably based on models or modelling techniques developed in one area, or scientific domain (typically dynamical systems theory, control theory, etc.), which are applied to phenomena or problems in a different area (e.g. biology, social sciences). This may represent a theoretical advance (e.g. extension of theories, introduction of new concepts, mathematical formalisation,etc.), and often the data requirements of the model far exceed the available data in the new area. Consequently model validity cannot be equated simply with empirical correspondence and/or theoretical coherence (e.g. as in the modelling of technological (physical) systems). Another problem for validation is that the conventional approaches, objectives, and criteria of acceptability in the new area may differ considerably from the area in which the model or modelling techniques originate. Some of the validation methodologies were designed specifically to cope with these difficulties (the γ-methodology for models with a limited amount of theoretical and empirical innovation, Section 5.4; the ε-methodology for highly innovative, or analogical, models, Section 5.6; and a general methodology for models in the social sciences; Section 9.3). The main features of these methodologies are:

(i) The inclusion of objective criteria for heuristic assessment.

(ii) Detailed considerations of modelling objectives, and data requirements and availability.

(iii) Theoretical validity tests if possible (it is not wise to ignore existing theoretical achievements.

(iv) Empirical validation conceived as an empirical research programme, extending the available data types (i.e. devising new measurement systems) and making comparisons with the model when possible. (A methodology for model-data comparisons has also been developed which ranges from qualitative tests through feature space comparisons to model parameter estimation procedures - the α-methodology, Section 5.2).

(v) Pragmatic assessment for utilitarian modelling objectives.

The emphasis in the present approach has been in creating validation methodologies which expose and deal with the full extent of the validation problem. It is important to recognise that many stages of validation (i.e. application of validity criteria) are fundamentally embedded in the overall process of model formulation and development, and that validation is not simply a "bottom of the page" empirical calibration procedure. Consequently, if a model emerges badly from validation (i.e. is "invalid") an improvement will probably require a change of the modelling methodology as well as the model.

Models are increasingly complex - a consequence of the availability of powerful computing facilities and the nature of systems science. More complex models require greater time for formulation and validation. Unfortunately, most development has occurred on the formulation side of modelling in systems science rather than validation (this is probably because in technical modelling of physical systems, where many of the modelling techniques originate, models are based on well-established bodies of theories and data, resulting in an automatic validity). The validity of complex systems models in areas such as biology, the social sciences, etc. based on these techniques is therefore problematic (in other words, there is an inadequate realisation of a model's objectives). The use of the validation methodologies developed here will result in the inclusion of critical validation tests at key points in modelling methodologies. Hopefully this will result in an improved distribution of the limited modelling resources so that a model will be more able to realise its objectives.

On a general, or metatheoretical, level the theory of model validity acts as a common conceptual framework for analysing models and validity in all application areas of systems science. This commonality was achieved by broadening the concept of validity as used in different areas, including heuristic and pragmatic validity, and relating validity to modelling objectives, as opposed to a simplistic treatment of validity (e.g. validity as empirical correspondence). The conceptual framework can help in the identification of both similarities and differences between systems models, modelling techniques, and validity in different areas. Thus it satisfies the systemic research aim of unity at a metatheoretical level (i.e. a theory of theories, or models) and can determine the extent to which unity is achieved at the level of theory, model, and methodology (e.g. do the systems concepts such as feedback, stability, etc., have a general applicability?).

It appears that the generality claimed for many systems models or concepts is essentially analogical (i.e. the generality only holds for heuristic validity), and has not been demonstrated for the full range of validity criteria.

The problem of the definition and epistemology of "system" has also been considered (see Sections 3.5.4, 4.4.2, and 9.4.4). Definitions of "system", by itself, are usually either trivial, or formal and over-restrictive (and ontologically dubious). In practice, "system" is used to refer vaguely to a collection (or grouping, assembly, etc.) of phenomena, events, people, or objects which, for some reasons, are interesting as a collection (or grouping, assembly, etc.) and not simply as individuals. However, when used with an adjective, as in "Σ system", a precise set of local and global, structural and functional properties is identified. A Σ system is therefore a theoretical entity or model. If a model M (which is a Σ system) is used to represent an intended range of application, $\mathcal{R}_I$ then the decision as to whether or not $\mathcal{R}_I$ has the properties of a Σ system will depend on the extent to which M has been validated. The formulation and validation of models therefore provides a legitimate epistemological basis for the investigation of systems in systems science. In separating Σ system from $\mathcal{R}_I$ a difficult problem in the philosophy of systems science is solved.

10.3 Recommendations for Modelling and Validation in Biology and Medicine

In this section, the conclusions of the validation studies of the three biological models (Chapters 6, 7 and 8) will be briefly summarised, and some recommendations for modelling and validation in biology and medicine will be made that are based on the developments reported in this thesis.

The objectives of the model of the human cardiovascular system examined in Chapter 6 are scientific - the investigation of short-term haemodynamics and pharmacodynamics in man. The model is highly complex, containing 61 first order differential equations and 159 algebraic equations, and there is considerably insufficient physiological theory and data available to validate it fully. However, it was demonstrated that, on the basis of current physiological understanding and available data, the model had many defects. These were traced to the neural control and drug submodels. This implies that there is scope for model improvement

even though the theoretical/data base of the model is insufficient. Although the high detail of the model requires an extension of available data, the information this would yield is of very limited physiological and medical significance. If the model was simplified, the gap between required and available data would be diminished and, more importantly, greater confidence could be placed on the indirect validation of the submodels of neural control and drug action.

By contrast, the model of the human renal-artificial kidney machine system (Chapter 7) has primarily utilitarian objectives - the improvement of the health care of patients with kidney failure undergoing dialysis therapy - although it is also intended for the study of renal control processes in normal and disease states. The model includes simple representations of the major control systems that are associated with the kidneys (hormonal, fluidic, electrolytic, and thermal). Most of these are based on partial understanding and inadequate data (e.g. from animal experiments) and are therefore a large source of uncertainty in the model. Fortunately, many of these controllers do not operate when a patient is in a state of severe renal failure, and the model yields fairly accurate predictions for the major clinical variables during and between periods of dialysis. These predictions may be used to optimise dialysis therapy (e.g. minimisation of time on dialysis and maximisation of time between dialyses). However, the accuracy of the model in this degenerate mode does not justify the inclusion of the controllers. The extra complexity slightly undermines confidence in the predictions and makes model parameter estimation very difficult and problematic. This suggests that a simpler model should be used for the clinical application, and the full model for investigating normal renal control mechanisms.

The model of the human respiratory system assessed in Chapter 8 is intended for the scientific objective of investigating respiratory control. Many parts of the model (e.g. lung-gas exchange curves, blood gas dissociation, etc.) are based on well—validated physiological theory. In addition, there is a long tradition of quantitative approach to respiratory physiology and many of the model variables are measured in physiological laboratories. Of the three biological models examined, the respiratory model is the one most fully validated. It is also possible to test (and reject) various hypotheses for respiratory control using the model thereby directly satisfying its main objective.

In making general recommendations for modelling and validation in

biology and medicine, five separate topics will be considered: (i) scientific objectives; (ii) clinical application; (iii) models for teaching; (iv) use of features; and (v) suites of models. These recommendations are made directly on the basis of the developments and results reported in this thesis, in particular the theory of model validity, the validation methodologies, and the results of the validation case studies.

(i) Scientific objectives are associated with the use of a model for understanding, explanation, stimulating experimental research, etc. In biology, these objectives may require models to represent a system accurately, on the one hand, or to be a heuristic device, on the other. In the former case, the γ-methodology (Section 5.4) for model validation which systematically applies theoretical and empirical criteria to various disassemblies of a model has proved to be a powerful and critical methodology. In comparing a general model with biological data from a particular system or individual, the methods of feature space comparisons are very useful (see Section 5.2, or (iv), below). Many models are used in theoretical biology to provide a new focus for theoretical or experimental research and may be incommensurable with existing biological theory and data (e.g. models of neural nets). The ε-methodology for model validation (Section 5.6) which provides objective criteria for heuristic assessment may be used in the initial validation of this type of model until new data become available for empirical validation.

The aspects of biological systems which play a major role in control and adaptation are the neural, hormonal, and genetic systems. These are all areas in which understanding is still at an early stage. There is tremendous scope here for the use of models as exploratory devices and to introduce mathematical theory. An example is the way in which different neural control loops interact and are coordinated in the medulla. The overall effects on bodily systems of different types of hypothetical interactions may be investigated using a model. This may suggest critical experiments that would not usually be undertaken within the conventional areas of physiology.

(ii) Models are being increasingly used in medicine for clinical applications (i.e. a utilitarian modelling objective). Major uses include the optimisation of clinical measurement, or assay, systems; the improvement or automation of diagnosis and prognosis; the improvement of therapy; or the improvement of health-care in a wider sense. Models for clinical applications are generally simpler than those for scientific objectives

and are attractive for a number of reasons. Firstly, their simplicity means that their data requirements match those available and the models may be more fully validated empirically (i.e. in time-series comparisons). Secondly, this allows a parameter estimation problem to be well-posed and solved for system identification or indirect measurement. Thirdly, less time is spent on one model (this has the additional advantage that there will be less tendency to retain a model, because of sunk costs, after it has failed validation tests). Fourthly, they may be simulated on microcomputers which are widely available. Finally, their conceptual and mathematical basis is easily communicable.

A suitable validation methodology for clinical models is the β-methodology (Section 5.3) which has an emphasis on empirical tests (in particular, those based on parameter estimation techniques). It is also important to consider the pragmatic validity of these models. (Empirical validation is concerned with the extent to which a model represents a particular system; pragmatic validation investigates whether or not utilitarian objectives are satisfied.) The δ-methodology (Section 5.5) provides a framework for the critical assessment and ex post evaluation of a model. For example, empirical validation will determine if a model can correctly estimate parameters for an individual patient, whereas pragmatic validation will assess whether these parameter values promise, or result in, improvements of diagnosis or therapy.

(iii) Models may also be used for teaching purposes to represent some physiological system under normal,pathological, or therapeutic conditions. The validity of this type of model should be assessed in terms of their didactic effect on the student. Typically, the model is simulated on a computer using an interactive program. The representational validity of this type of model can be relaxed - its behaviour should not appear different from a physiological system, but it does not have to correspond exactly to a particular instance - and, consequently, there is much potential for models in this role.

(iv) In comparing a general model of a type of biological system with empirical data from individual systems, problems arise because of the inherent variability of biological systems and phenomena. Statistical techniques may be used to obtain population averages, but this may filter out important and interesting characteristics (especially dynamic phenomena of variable timing). An alternative approach is to define features of the data obtained from biological systems which capture the important

characteristics and, at the same time, reduce the variability between individual systems (i.e. an effective normalisation procedure). A general method for comparing features of data and model in validation is included in the α-methodology (Section 5.2).

(v) One of the main problems for model validation in biology and medicine occurs when a model has multiple objectives. For example, the renal model (Chapter 7) is required for investigating renal control mechanisms (a scientific objective) and for improving renal dialysis therapy (a utilitarian objective). It is quite likely that a single model will not fully satisfy all its objectives. In the case of the renal model, the two objectives are clearly conflicting. There is no reason why one large model should be the only acceptable solution. Instead, a set or suite of interrelated models may be developed where each model has a different objective. Each model can be clearly validated against its single objective and conflicting requirements can be eliminated. Such a suite of models could be stored in a single computer file together with an operating or executive program and might consist of a small model for parameter estimation, a larger model for examining overall behaviour, and a probabilistic decision model for diagnosis, prognosis, and therapy selection, for example.

10.4 Suggestions for Future Research Programmes

There are four distinct areas in which research on model validity and validation may be undertaken: (i) philosophy; (ii) methodology; (iii) theoretical aspects; and (iv) validation case studies of individual models. It benefits a research programme if more than one area is considered; ideally, all four areas should be involved.

(i) Philosophy. The incorporation of the theory of model validity into a theory of models, or modelling. The epistemology of modelling, its implications for model validity and validation, and for systems concepts and theories in systems science. This work should be based on a detailed examination of actual scientific practice.

(ii) Methodology. The development of critical, effective validation methodologies appropriate to specific application areas (e.g. physical, biological, social, or political modelling) and which cover the full range of validity criteria - internal, and external (empirical, theoretical, pragmatic and heuristic). The results of the current study suggest that this may involve changing overall modelling methodologies. An aim should be

the development of methodologies (joint model formulation/validation) with which modelling objectives may be realised at the expense of minimum use of resources (e.g. money, time, computing, data acquisition, etc.)

(iii) Theoretical aspects. Further work on the problem of identifiability (theoretical and practical aspects). Classification of non-linear dynamic systems and their qualitative properties (extremely important in model validation).

(iv) Validation case studies. Much understanding of the nature of model validity and techniques for model validation can be obtained by selecting a group of illustrative models from a particular research area as case studies for validation. In this thesis, the subjects for the case studies were biological models. The next stage would be to consider in detail other areas such as economic modelling or political modelling in which model validity is more problematic (economics and econometrics would probably be the more fruitful area). The results of these case studies may also be used to test and develop the theory of model validity developed in Chapter 4. Since many models are complex, such case studies may be time-consuming. Research, therefore, should be directed also at methods for devising a minimal, yet critical and effective, set of tests for validation.

CHAPTER 11

CONCLUSIONS

This thesis has investigated in depth the meaning and nature of model validity and the ways in which models may be validated. As models have become increasingly complex and applied to new areas, such as biology, ecology, energy systems, and the social sciences, it is widely recognised that the problems of model validity and validation have become much more difficult and important. In a review of the scientific literature of model validity and validation (Chapter 2) it was found that, whilst numerous concepts of validity and techniques for validation abound, there is no framework or satisfactory explanation of how the various concepts are related to each other and to other factors, or of what tests to use in validating a particular model. A consequence of this is that many validation methodologies are simplistic (e.g. restricted to input-output data comparisons) and some models are not validated at all. Validation methodologies which are entirely data-based may work adequately for models in control engineering (or other technical areas, where the models are based on classical physics) but when used in new areas, such as biology or the social sciences, where there are great theoretical and data problems, they are often totally inappropriate. It was clear at an early stage in the work, therefore, that validity and validation would have to be considered from a much wider perspective. The review of the philosophy of science (Chapter 3) revealed many different ideas on the nature of model validity and at the same time stressed the importance of a sound epistemological base. An interesting finding of this review was that the current trend in the philosophy of science is the study of actual ongoing scientific practice and development, rather than the old positivist concern with the logical properties of completed theories.

An innovative theory of model validity was proposed in Chapter 4. This is a conceptual framework which gives meaning to the different aspects of model validity and explains how they are related to each other and to other factors. As such, it provides a basis for analysing models and their validity and for designing appropriate validation methodologies for all application areas of models. In so doing it is unique and original (an exception is the seminal paper of Hermann, 1967, in which many of the ideas were first suggested although a full conceptual framework was not

actually developed). The basic structure of the theory is triadic, consisting of an analysis of modelling objectives, a theory of data, and a set of validity criteria. The validity criteria (or tests for validity) represent the various concepts of model validity and are closely related to modelling objectives. They are classified as follows:

1. Internal criteria (Prerequisite criteria with no reference outside a model)
   - 1.1 Consistency criterion (A model should contain or entail no contradictions, Section 4.3.3.2.1)
   - 1.2 Algorithmic criteria (Faithful solution or simulation of a model, Section 4.3.3.2.2)

2. External criteria (Reference to external factors, e.g. data or theory)
   - 2.1 Representational criteria (Tests of the extent to which a model represents phenomena or system, Section 4.3.3.3)
     - 2.1.1 Empirical criteria (Correspondence with empirical data, Section 4.3.3.3.1)
     - 2.1.2 Theoretical criteria (Coherence with accepted theories or models, Section 4.3.3.3.2)
   - 2.2 Heuristic criteria (Tests of heuristic, or scientific, potential, Section 4.3.3.5)
   - 2.3 Pragmatic criteria (Assessment of the value of a model for practical use, Section 4.3.3.4)

An important aspect of the theory is the inclusion of a number of objective tests for heuristic validity, which may be used in the validation of highly innovative models, as in systems science and the application of models to new areas. There is also a detailed classification of modelling objectives (Section 4.3.1) in which a primary distinction is made between scientific objectives (for representation, explanation, hypothesis testing, etc.) and utilitarian objectives (for practical applications). The theory of data (Section 4.3.2) helps in considering the nature of available data, and the implications of such data for model validity (particular emphasis is on data uncertainty). The final and most important aspect of the theory is the elucidation of the general relationship between modelling objectives, data, and validity criteria, (Section 4.3.4). This relationship may be used to determine appropriate validity criteria and data requirements for the validation of a model.

Throughout the theory there is explicit consideration of the influence on model validity of the content and stage of development of the scientific domain (field of research) associated with a model.

The first test of the theory of model validity was in using it to develop validation methodologies suitable for different modelling objectives and stages of scientific development (Chapter 5). A wide variety of methodologies was devised, ranging from empirically-based methodologies (e.g. for models in control engineering), through empirical/theoretical methodologies (e.g. for biological models), to heuristic methodologies (e.g. for political models). The latter methodology is substantially new and offers tremendous scope for the validation of innovative models, for instance in systems science. The range and detail of the validation methodologies provided initial evidence for the power and general applicability of the theory of model validity.

The next stage was the application of the conceptual framework of the theory of model validity and appropriate validation methodologies (and techniques) to three mathematical biological models (a model of the human cardiovascular system, Chapter 6; a model of the human renal-artificial kidney machine system, Chapter 7; and a model of the human respiratory control system, Chapter 8). The results were critical and detailed assessments of the models which determined clearly the extent of their representational validity and whether or not they satisfy their modelling objectives. In addition, the areas of inadequacy or uncertainty within the models have been identified and suggestions made for future model development. Biological models deal with complex phenomena, are often based on inadequate theory, and have data requirements which usually exceed those available. Consequently, they raise many problems for model validity and validation, and the successful results of these three case studies are practical evidence for the appropriateness of the validation methodologies used and the value of the conceptual framework (i.e. the theory of model validity). Furthermore, the results themselves have proved important in the development of the models.

Further support for the theory of model validity was obtained in the final case study in which more general aspects of model validity and validation in the social sciences were investigated (Chapter 9). After considering the nature of theory, models, and data in the social sciences (as compared with the physical and life sciences), a general methodology, based on the theory of model validity, for model validation was proposed.

This is a broad validation methodology with particular emphasis on methods for the objective assessment of heuristic validity. The methodology was applied to some illustrative areas of modelling in the social sciences (including world modelling, and models of bicommunal political conflict) and it was found to help in identifying clearly the potentialities and problems of models and validation in the various areas.

Finally, in Chapter 10, the overall implications of the work for methodological, theoretical, and practical aspects of modelling and validation in systems science, and biology and medicine were examined. In both these areas it was shown that the theory of model validity (Chapter 4) leads to an improved understanding of the nature of modelling and model validity in general, and that the validation methodologies (Chapter 5) are suitable for the critical and effective validation of a wide range of types of model. For modelling in biology and medicine detailed recommendations were made, mainly on the basis of the results of the biological case studies, for the types of model appropriate for different modelling objectives (e.g. research vs. clinical application) and for suitable techniques and methodologies for validation. Lastly, it was suggested that there are four distinct areas for future research into model validity and validation - philosophy, methodology, theoretical aspects, and practical validation studies - and that research programmes would be most successful if at least two areas were pursued simultaneously.

The work reported in this thesis contributes to the improved understanding and explanation of the concept of model validity and offers a repertoire of practical validation methodologies. On another level, the work is a broad methodological study of the kind which is urgently required in systems science. More practically, however, much of the thesis has been concerned with the extensive validation of three specific biological models.

A model may be regarded as a system of equations and, to paraphrase the famous words of Paul Dirac, "it is more important to have beauty in one's model than to have it fit experiment".

APPENDIX I

AN HISTORICAL CASE STUDY IN MODEL VALIDATION - WILLIAM HARVEY'S DISCOVERY OF THE CIRCULATION OF THE BLOOD

William Harvey (1597 - 1657) may be regarded as the father of modern physiology and medical science. Although this fails to recognise the great contributions of other early scientists such as Vesalius (1514 - 1564), Servetus (1511 - 1553), Fabricius of Aquapendente (1537 - 1619), and Columbus (1516 - 1559), as well as the Islamic scientists, and overemphasises the role of the individual in historical processes, it is true that Harvey personally made several profound scientific discoveries and advances. His three main works are "Prelectiones Anatomiae Universalis" (1615 - 1628), "De Motu Cordis et Sanguinis in Animalibus" (1628), and "De Generatione Animalium" (1651). The "Prelectiones" were prepared from Harvey's lecture notes and reveal the germination and growth of his thoughts on medicine as a whole, as well as the two great problems - generation of animals, and the movement of the heart and blood. In "De Generatione", he propounds an epigenetic theory of generation with the heart playing the control role.

Harvey is best remembered as the discoverer of the systemic circulation of the blood, which he described in his most celebrated book, "De Motu Cordis". This little book of 72 pages is a masterpiece of beautifully logical scientific writing. Not only was the discovery profound in itself, the scientific methods Harvey used (in particular quantitative argument) were also novel. In this appendix, Harvey's demonstration of the circulation of the blood in "De Motu Cordis" will be outlined, and analysed as an historical validation case study. This relates not only to the main topic of the thesis (model validity and validation) but also to the subject of the main case studies - biological modelling - especially Chapter 6 (the validation of a mathematical model of the human cardiovascular system). (For a more general account of Harvey's life and work, consult Keele, 1978).

AI.1 An Outline of Harvey's Demonstration of the Circulation of the Blood in "De Motu Cordis".

The early chapters of "De Motu Cordis", ("On the Movement of the

Heart and Blood in Animals"), are taken up with a critical review of previous work and remarks on the various phases of the movement of the heart based on Harvey's careful clinical and anatomical observations of cold blooded animals. In Chapter 8, he reports how he found that the amount of blood passing from the veins to the arteries in the heart is more than could be possibly derived from ingested food.

"In consequence, I began privately to consider if it had a movement, as it were, in a circle. This hypothesis I subsequently verified ..." (DMC, p.58). (The references, "DMC", are to the translation by Franklin, 1957).

In order to verify (validate) the hypothesis, Harvey makes three postulates which he subsequently demonstrates:

1. "that the blood is continuously and uninterruptedly transmitted by the beat of the heart from the vena cava into the arteries in such amount that it cannot be supplied by the ingesta".

2. "that the blood is continually, evenly, and uninterruptedly driven by the beat of the arteries into every member and part, entering each in far greater amount than is sufficient for its nutrition or can be supplied to it by the whole mass of blood".

3. "similarly, ... that the veins themselves are constantly returning this blood from each and every member to the region of the heart". (Chapter 9; DMC, p. 61).

In demonstrating the first postulate, Harvey took the great step of introducing quantitative measurement and theory into physiology and medicine:

"In man, then, let us take the amount that is extruded by the individual beats, and that cannot return into the heart because of the barrier set in its way by the valves, as half an ounce, or three drachms, or at least one drachm. In half an hour the heart makes over a thousand beats ... If you multiply the drachms per beat by the number of beats you will see that in half an hour either a thousand times three drachms or times two drachms, or five hundred ounces, or other such proportionate quantity of blood has been passed from the heart into the arteries, that is, in all cases blood in greater amount than can be found in the whole of the body ...

In the above sort of way, by calculating the amount of blood transmitted ... let us convince ourselves that the whole amount of blood mass goes through the heart from the veins to the arteries and similarly makes the pulmonary transit.

Even if this may take more than half an hour or an hour or a day for its accomplishment, it does nevertheless show that the beat of the heart is continuously driving through that organ more blood than the ingested food can supply, or all the veins together at any time contain."

(Chapter 9; DMC, pp. 62 - 63)

In Chapter 11, Harvey enters on the well known series of experiments in which he ligatures the veins and arteries of the arm in order to confirm the second postulate.

"Just as in a tight ligature the arteries above the ligature are distended and pulsate, but not those below it, so - per contra - in a medium tight ligature the veins below the ligature swell up and are resistant, but those above behave quite differently." (DMC, p. 73).

This demonstrates that blood goes outward with its strength from the heart down the arteries, and returns in the veins with less vigour to the heart. Since microscopes were not yet available to observe the capillary beds, Harvey could only speculate that the blood goes from the arteries to the veins through "invisible porosities". (The capillaries were observed by Malpighi, circa 1670).

The third postulate receives confirmation in Chapter 13. Firstly, Harvey describes the anatomy and function of the venous valves, (Fabricius, Harvey's anatomy teacher in Padua, had made detailed anatomical studies of the venous valves, but did not realise their functional importance). The twin flaps of each valve "are so ready to come together and act in unison that they completely prevent any backflow from the root of the vein into any other branches" (DMC, p. 82). In order to provide external proof of the action of the venous valves and additional confirmation of the third postulate, he performs another series of ligature experiments, with a medium-tight ligature.

Figure AI.1 shows Harvey's illustrations of the experiments (the only diagrams in "De Motu Cordis"). In the top illustration (figura 1) swellings occur at intervals (B, C, D, D, E, F) which are produced by the valves. Blood cannot be "milked" downwards past a valve. But if pressure is placed on a valve (H) and the blood milked

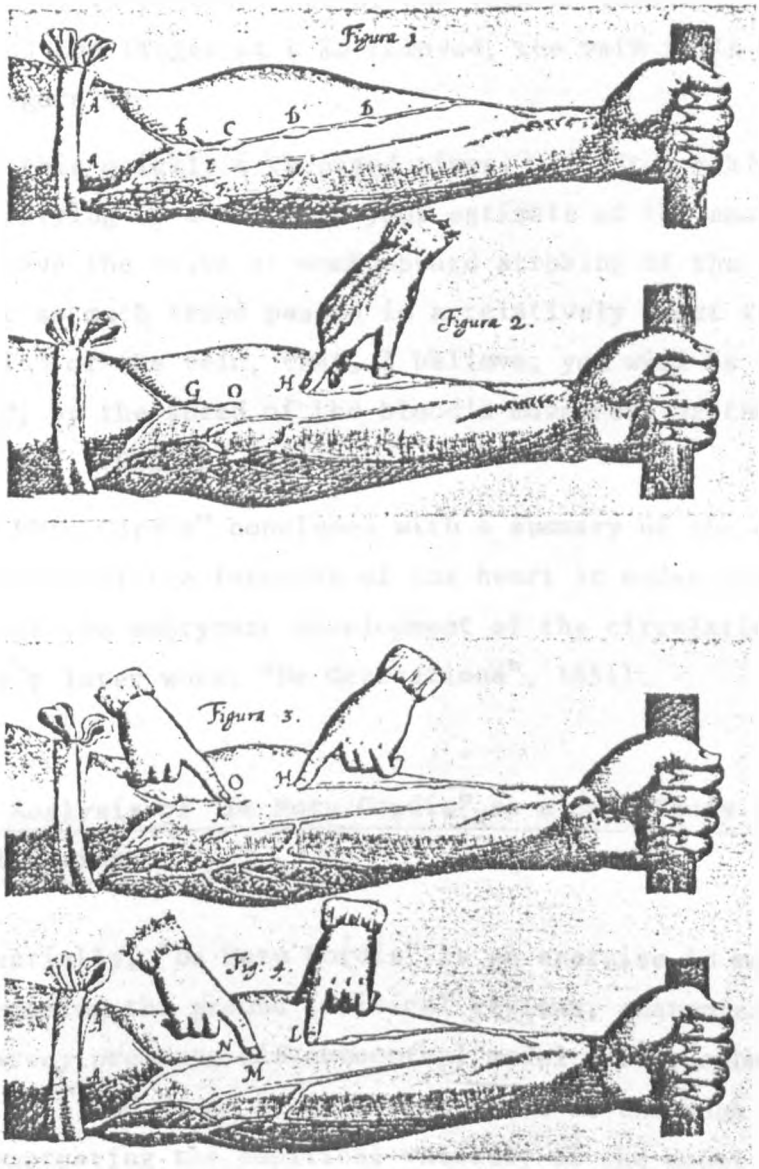


FIGURE AI.1. ILLUSTRATIONS TAKEN FROM HARVEY'S "DE MOTU CORDIS" (1628) SHOWING THE USE OF A LIGATURE TO DEMONSTRATE THE THIRD POSTULATE AND THE CIRCULATION OF THE BLOOD. (THE LETTERING REFERS TO THE DEMONSTRATION OF THE ACTION OF THE VENOUS VALVES).

upwards past the next valve (O) the section of vein (H - O) remains empty (figura 2). Even if pressure is applied above the valve (K) blood will still not pass backwards through valve (O) (figura 3). Finally, if the blood is pressued out of the vein LN into the region above the valve, and the finger at L is removed, the vein fills up quickly from below (figura 4)

"Do this quickly a thousand times. If after making a calculation (by multiplying by a thousand your estimate of the amount which is raised above the valve at each upward stroking of the vein), you will find that so much blood passes in a relatively short time through the one portion of the vein, that, I believe, you will be completely convinced, by the speed of the blood's movement, of the fact that it circulates." (DMC, p. 86).

"De Motu Cordis" concludes with a summary of the argument, a consideration of the function of the heart in wider terms, and an analysis of the embryonic development of the circulation (much extended in Harvey's later work, "De Generatione", 1651).

AI.2 An Analysis of "De Motu Cordis" as a Case Study in Model Validation

Essentially, "De Motu Cordis" is an exercise in model validation. After preparing the ground (critical reviews, anatomical observations, etc.), Harvey presents his conceptual model (or hypothesis) of the circulating blood at an early stage. Most of the book is concerned with demonstrating the empirical validity of the model. Since the passage of blood from arteries to veins could not be observed at the time, Harvey devised three critical validation tests (i.e. testing the three postulates) the results of which could only be satisfied by his model. These tests will now be considered with respect to the different validity criteria: empirical, theoretical, heuristic, and pragmatic (see Chapter 4).

All the tests are explicitly empirical. The first involves a quantitative determination of the blood flow rate from the heart. The very high value is consistent with the circulating blood model, although Harvey gave no upper and lower bounds for blood velocity. The second and third tests rely mainly on qualitative observations of the behaviour of the veins and arteries during ligature experiments. These provided

additional confirmation that the blood flows from arteries to veins. The basic process of transport from arteries to veins could not be directly observed, but only inferred, and so the tests are effectively indirect empirical tests. (The same type of indirect validation is used to validate the neural control (or other control) submodels in the biological models of Chapters 6, 7 and 8).

Harvey put forward a radical new model about blood flow - a discovery - which contradicted and therefore could not be assessed against the conventional (medieval) ebb and flow model of the blood flow, in which arterial and venous flows are separate. However, whilst in functional terms his model was original, in structural or anatomical terms (structure of the heart, venous valves, etc.) his ideas were coherent with those of earlier anatomists, such as Fabricius. At the same time as validating Harvey's model, the empirical tests provided counterevidence against the ebb and flow model.

Although the acceptance of Harvey's model is based on empirical criteria, it also satisfies heuristic criteria. Firstly, it offered an explanation of data that could not be explained previously and resolved the anomalies of the ebb and flow model. Secondly, "De Motu Cordis" is an excellent example of good scientific reasoning (not simply deductive logic).

Harvey's discovery of the circulation was so soundly based that it was soon accepted (within his lifetime). This acceptance did not require an assessment of the pragmatic value of the new model. In fact, an early criticism was that the ebb and flow model was perfectly adequate for medical application. Of course, the discovery and the new techniques were soon found to be of great importance for practical medicine as well as physiology.

APPENDIX II

A MATHEMATICAL MODEL OF THE HUMAN CARDIOVASCULAR SYSTEM

List of Symbols Used

a	elastance (reciprocal compliance)
A	cross-sectional area
b	variable associated with myocardial contractility control
B	baroreceptor output
C	compliance
d	variable associated with venous tone control
f	frequency
F	flow
g	acceleration due to gravity
G	hydrostatic pressure difference
h	integration step length
K	constant
l	length
L	inertance
m	mass
M	injected mass
n	number of g's of acceleration
P	pressure
q	variable associated with peripheral resistance control
R	resistance
t	time
T	period
u	variable associated with heart rate control
v	velocity
V	volume
ω	concentration
x	variable associated with time-varying compliance generation
y	variable associated with respiration
α	constant
μ	kinematic viscosity
ρ	density
σ	variable associated with drug effects
τ	time constant

List of Subscripts

Compartmental subscripts (after Beneken & Dewit, 1967) :

AA	Abdominal arteries
AO1	Ascending aorta
AO2	Aortic arch
AO3	Thoracic aorta
AV	Abdominal veins
CA	Leg arteries
IA	Intestinal arteries
CV	Leg veins
IV	Intestinal veins
IVC	Inferior vena cava
LA	Left atrium
LV	Left ventricle
PA	Pulmonary arteries
PV	Pulmonary veins
RA	Right atrium
RV	Right ventricle
SVC	Superior vena cava
UA	Head and arm arteries
UV	Head and arm veins

Other subscripts :

AO	Aorta
SA	Systemic artieries
SC	Systemic circulation
SV	Systemic veins
ABD	Abdomen
BRONC	Bronchial
CC	Critical closure
COR	Coronary
D	Diastolic
H	Heart
HEAD	Head and arms
IE	Inhalation-exhalation
INT	Intestinal
TV	Thoracic veins
VC	Vena cava
AOT	Aortic arch & thoracic aorta
LPA	Lower part arteries
LPV	Lower part veins
LEG	Legs
LUNG	Lungs
MAX	Maximum
MIN	Minimum
N	Normal
R	Respiratory
TH	Thorax
T	Total
U	Unstressed

Units Employed

Pressure	mmHg
Flow	ml sec^{-1}
Volume	ml
Compliance	ml $mmHg^{-1}$
Elastance	mmHg ml^{-1}
Resistance	mmHg sec ml^{-1}
Inertance	mmHg sec^2 ml^{-1}

(1 mm Hg = 1332 gm cm^{-1} sec^{-2})

The Complete Mathematical Model (Pullen, 1976)

A.II.1 The Circulatory Fluid Mechanics Model

Right atrium:

$$\frac{dV_{RA}}{dt} = F_1 - F_{RARV} \quad , \quad V_{RA} \geq 0 \tag{AII.1}$$

$$P_{RA} = a_{RA} \, (V_{RA} - V_{URA}) \tag{AII.2}$$

$$F_{RARV} = \begin{cases} F_2 & , \quad F_2 > 0 \\ 0 & , \quad F_2 \leq 0 \end{cases} \tag{AII.3}$$

$$F_1 = F_{SVCRA} + F_{IVCRA} + F_{BRONC} + F_{COR} \tag{AII.4}$$

$$F_2 = (P_{RA} - P_{RV}) / R_{RARV} \tag{AII.5}$$

Right ventricle:

$$\frac{dV_{RV}}{dt} = F_{RARV} - F_{RVPA} \quad , \quad V_{RA} \geq 0 \tag{AII.6}$$

$$\frac{dF_{RVPA}}{dt} = \frac{P_{RV} - P_{PA} - R_{RVPA} F_{RVPA} - \left(\frac{\rho}{2A^2_{PA}}\right) F^2_{RVPA}}{L_{RV}} \quad , \quad F_{RVPA} \geq 0 \tag{AII.7}$$

$$P_{RV} = a_{RV} \, (V_{RV} - V_{URV}) \tag{AII.8}$$

Pulmonary arteries:

$$\frac{dV_{PA}}{dt} = F_{RVPA} - F_{PAPV} \quad , \quad V_{PA} \geq 0 \tag{AII.9}$$

$$P_{PA} = (V_{PA} - V_{UPA}) / C_{PA} \tag{AII.10}$$

$$F_{PAPV} = \begin{cases} (P_{PA} - P_{PV}) / R_{LUNG} , & P_{PV} > P_{CC} \\ (P_{PA} - P_{CC}) / R_{LUNG} , & P_{PV} \leq P_{CC} \end{cases} \qquad \text{(AII.11)}$$

Pulmonary veins:

..22

$$\frac{dV_{PV}}{dt} = F_{PAPV} - F_{PVLA} , \quad V_{PV} \geq 0 \qquad \text{(AII.12)}$$

$$P_{PV} = (V_{PV} - V_{UPV}) / C_{PV} \qquad \text{AII.13)}$$

$$C_{PV} = \begin{cases} C_{PVN} , & V_{PV} > V_{UPV} \\ K_6 C_{PVN} , & V_{PV} \leq V_{UPV} \end{cases} \qquad \text{(AII.14)}$$

$$F_3 = \frac{(P_{PV} - P_{LA}) V^2_{PV}}{R_{PVLA} V^2_{UPV}} \qquad \text{(AII.15)}$$

$$F_{PVLA} = \begin{cases} F_3 , & F > 0 \\ K_1 F_3 , & F_3 \leq 0 \end{cases} \qquad \text{(AII.16)}$$

Left atrium:

$$\frac{dV_{LA}}{dt} = F_{PVLA} - F_{LALV} , \quad V_{LA} \geq 0 \qquad \text{(AII.17)}$$

$$P_{LA} = a_{LA} (V_{LA} - V_{ULA}) \qquad \text{(AII.18)}$$

$$F_4 = (P_{LA} - P_{LV}) / R_{LALV} \qquad \text{(AII.19)}$$

$$F_{LALV} = \begin{cases} F_4 , & F_4 > 0 \\ 0 , & F_4 \leq 0 \end{cases} \qquad \text{(AII.20)}$$

Left ventricle:

$$\frac{dV_{LV}}{dt} = F_{LALV} - F_{LVAO1} \quad , \quad V_{LV} \geq 0 \qquad \text{(AII.21)}$$

$$\frac{dF_{LVAO1}}{dt} = \frac{P_{LV} - P_{AO1} - R_{LVAO1}F_{LVAO1} - \left(\frac{\rho}{2A^2_{AO1}}\right)F^2_{LVAO1}}{L_{LV}} \quad , \quad F_{LVAO1} \geq 0 \qquad \text{(AII.22)}$$

$$P_{LV} = a_{LV} \quad (V_{LV} - V_{ULV}) \qquad \text{(AII.23)}$$

Ascending aorta:

$$\frac{dV_{AO1}}{dt} = F_{LVAO1} - F_{AO1AO2} - F_{COR} \quad , \quad V_{AO1} \geq 0 \qquad \text{(AII.24)}$$

$$\frac{dF_{AO1AO2}}{dt} = (P_{AO1} - P_{AO2} - R_{AO2}F_{AO1AO2}) \; / \; L_{AO2} \qquad \text{(AII.25)}$$

$$P_{AO1} = \frac{1}{C_{AO1}} (V_{AO1} - V_{UAO1}) + \frac{K_8}{C_{AO1}} \cdot \frac{dV_{AO1}}{dt} \qquad \text{(AII.26)}$$

$$F_{COR} = (P_{AO1} - P_{RA}) \; / \; R_{COR} \qquad \text{(AII.27)}$$

Aortic arch:

$$\frac{dV_{AO2}}{dt} = F_{LAO1AO2} - F_{AO2UA} - F_{AO2AO3} \quad , \quad V_{AO2} \geq 0 \qquad \text{(AII.28)}$$

$$\frac{dF_{AO2UA}}{dt} = \frac{P_{AO2} + P_{TH} - P_{UA} - R_{UA}F_{AO2UA} - G_{AO2UA}}{L_{UA}} \qquad \text{(AII.29)}$$

$$\frac{dF_{AO2AO3}}{dt} = \frac{P_{AO2} - P_{AO3} - R_{AO3}F_{AO2AO3} + G_{AO2AO3}}{L_{AO3}} \qquad \text{(AII.30)}$$

$$P_{AO2} = \frac{1}{C_{AO2}} (V_{AO2} - V_{UAO2}) + \frac{K_8}{C_{AO2}} \cdot \frac{dV_{AO2}}{dt} \qquad \text{(AII.31)}$$

Head and arms arteries:

$$\frac{dV_{UA}}{dt} = F_{AO2UA} - F_{UAUV} \quad , \quad V_{UA} \geq 0 \qquad \text{(AII.32)}$$

$$P_{UA} = \frac{1}{C_{UA}} \left(V_{UA} - V_{UUA} \right) + \frac{K_8}{C_{UA}} \cdot \frac{dV_{UA}}{dt} \qquad \text{(AII.33)}$$

$$F_{UAUV} = (P_{UA} - P_{UV}) / (R_{HEAD}\, \sigma_{HEAD}) \qquad \text{(AII.34)}$$

Head and arms veins:

$$\frac{dV_{UV}}{dt} = F_{UAUV} - F_{UVSUC} \quad , \quad V_{UV} \geq 0 \qquad \text{(AII.35)}$$

$$P_{UV} = d_3 \ (V_{UV} - V_{UUV}) / C_{UV} \qquad \text{(AII.36)}$$

$$C_{UV} = \begin{cases} C_{UVN} & , \quad V_{UV} > V_{UUV} \\ K_6 C_{UVN} & , \quad V_{UV} \leq V_{UUV} \end{cases} \qquad \text{(AII.37)}$$

$$V_{UUV} = V_{UUVN} / d_4 \qquad \text{(AII.38)}$$

$$F_5 = \frac{(P_{UV} - P_{SVC} - P_{TH} + G_{UVSVC})\ V^2_{UV}}{R_{UV} V^2_{UUVN}} \qquad \text{(AII.39)}$$

$$F_{UVSVC} = \begin{cases} F_5 & , \quad F_5 > 0 \\ K_9 F_5 & , \quad F_5 \leq 0 \end{cases} \qquad \text{(AII.40)}$$

Thoracic aorta:

$$\frac{dV_{AO3}}{dt} = F_{AO2AO3} - F_{BRONC} - F_{AO3IA} - F_{AO3AA} \quad , \quad V_{AO3} \geq 0 \qquad \text{(AII.41)}$$

$$\frac{dF_{AO3IA}}{dt} = \frac{P_{AO3} + P_{TH} - P_{IA} - P_{ABD} - R_{IA} F_{AO3IA} + G_{AO3IA}}{L_{IA}} \qquad \text{(AII.42)}$$

$$\frac{dF_{AO3AA}}{dt} = \frac{P_{AO3} + P_{TH} - P_{AA} - P_{ABD} - R_{AA} F_{AO3AA} + G_{AO3AA}}{L_{AA}} \qquad \text{(AII.43)}$$

$$P_{AO3} = \frac{1}{C_{AO3}} (V_{AO3} - V_{UAO3}) + \frac{K_8}{C_{AO3}} \cdot \frac{dV_{AO3}}{dt} \quad \text{(AII.44)}$$

$$F_{BRONC} = \frac{P_{AO3} - P_{RA} - G_{AO3RA}}{R_{BRONC} \cdot q_4 \cdot \sigma_{BRONC}} \quad \text{(AII.45)}$$

Intestinal arteries:

$$\frac{dV_{IA}}{dt} = F_{AO3IA} - F_{IAIV} \quad , \quad V_{IA} \geq 0 \quad \text{(AII.46)}$$

$$P_{IA} = \frac{1}{C_{IA}} (V_{IA} - V_{UIA}) + \frac{K_8}{C_{IA}} \cdot \frac{dV_{IA}}{dt} \quad \text{(AII.47)}$$

$$F_{IAIV} = (P_{IA} - P_{IV})/(R_{INT}\, \sigma_{INT}\, q_4) \quad \text{(AII.48)}$$

Intestinal veins:

$$\frac{dV_{IV}}{dt} = F_{IAIV} - F_{IVIVC} \quad , \quad V_{IV} \geq 0 \quad \text{(AII.49)}$$

$$P_{IV} = d_3 (V_{IV} - V_{UIV})/C_{IV} \quad \text{(AII.50)}$$

$$C_{IV} = \begin{cases} C_{IVN} & , \quad V_{IV} > V_{UIV} \\ K_6 C_{IVN} & , \quad V_{IV} \leq V_{UIV} \end{cases} \quad \text{(AII.51)}$$

$$V_{UIV} = V_{UIVN}/d_4 \quad \text{(AII.52)}$$

$$F_6 = \frac{(P_{IV} - P_{IVC} + P_{ABD} - P_{TH} - G_{IVCIV}) V^2_{IV}}{R_{IV}\, V^2_{UIVN}} \quad \text{(AII.53)}$$

$$F_{IVIVC} = \begin{cases} F_6 & , \quad F_6 > 0 \\ K_{10} F_6 & , \quad F_6 \leq 0 \end{cases} \quad \text{(AII.54)}$$

Abdominal arteries:

$$\frac{dV_{AA}}{dt} = F_{AO3AA} - F_{AAAV} - F_{AACA} \quad , \quad V_{AA} \geq 0 \quad \text{(AII.55)}$$

$$\frac{dF_{AACA}}{dt} = \left(P_{AA} - P_{CA} + P_{ABD} - R_{CA}F_{AACA} + G_{AACA} \right)/L_{CA} \quad \text{(AII.56)}$$

$$P_{AA} = \frac{1}{C_{AA}} (V_{AA} - V_{UAA}) + \frac{K_8}{C_{AA}} \cdot \frac{dV_{AA}}{dt} \quad \text{(AII.57)}$$

$$F_{AAAV} = (P_{AA} - P_{AV})/(R_{ABD} q_4 \sigma_{ABD}) \quad \text{(AII.58)}$$

Abdominal veins:

$$\frac{dV_{AV}}{dt} = F_{AAAV} + F_{CVAV} - F_{AVIVC} \quad , \quad V_{AV} \geq 0 \quad \text{(AII.59)}$$

$$P_{AV} = d_3 (V_{AV} - V_{UAV})/C_{AV} \quad \text{(AII.60)}$$

$$C_{AV} = \begin{cases} C_{AVN} & , \quad V_{AV} > V_{UAV} \\ K_6 C_{AVN} & , \quad V_{AV} \leq V_{UAV} \end{cases} \quad \text{(AII.61)}$$

$$V_{UAV} = V_{UAVN}/d_4 \quad \text{(AII.62)}$$

$$F_7 = \frac{(P_{AV} - P_{IVC} + P_{ABD} - P_{TH} - G_{IVCAC}) V^2_{AV}}{R_{AV} V^2_{UAVN}} \quad \text{(AII.63)}$$

$$F_{AVIVC} = \begin{cases} F_7 & , \quad F_7 > 0 \\ K_{11} F_7 & , \quad F_7 \leq 0 \end{cases} \quad \text{(AII.64)}$$

Leg arteries:

$$\frac{dV_{CA}}{dt} = F_{AACA} - F_{CACV} \quad , \quad V_{CA} \geq 0 \quad \text{(AII.65)}$$

$$P_{CA} = \frac{1}{C_{CA}} (V_{CA} - V_{UCA}) + \frac{K_8}{C_{CA}} \cdot \frac{dV_{CA}}{dt} \quad \text{(AII.66)}$$

$$F_{CACV} = (P_{CA} - P_{CV})/(R_{LEG} q_4 \sigma_{LEG}) \quad \text{(AII.67)}$$

Leg veins:

$$\frac{dV_{CV}}{dt} = F_{CACV} - F_{CVAV} \quad , \quad V_{CV} \geq 0 \tag{AII.68}$$

$$P_{CV} = d_3(V_{CV} - V_{UCV})/C_{CV} \tag{AII.69}$$

$$C_{CV} = \begin{cases} C_{CVN} & , \quad V_{CV} > V_{UCV} \\ K_6 C_{VN} & , \quad V_{CV} \leq V_{UCV} \end{cases} \tag{AII.70}$$

$$V_{UCV} = V_{UCVN}/d_4 \tag{AII.71}$$

$$F_8 = \frac{(P_{CV} - P_{AV} - P_{ABD} - G_{AVCV})V^2_{CV}}{R_{CV} V^2_{UCVN}} \tag{AII.72}$$

$$F_{CVAV} = \begin{cases} F_8 & , \quad F_8 > 0 \\ K_{12} F_8 & , \quad F_8 \leq 0 \end{cases} \tag{AII.73}$$

Inferior vena cava:

$$\frac{dV_{IVC}}{dt} = F_{AVIVC} + F_{IVIVC} - F_{IVCRA} \, , \, V_{IVC} \geq 0 \tag{AII.74}$$

$$P_{IVC} = (V_{IVC} - V_{UIVC})/C_{IVC} \tag{AII.75}$$

$$C_{IVC} = \begin{cases} C_{IVCN} & , \quad V_{IVC} > V_{UIVC} \\ K_6 C_{IVCN} & , \quad V_{IVC} \leq V_{UIVC} \end{cases} \tag{AII.76}$$

$$F_9 = \frac{(P_{IVC} - P_{RA} - G_{IVCRA}) \; V^2_{IVC}}{R_{IVC} V^2_{UIVC}} \tag{AII.77}$$

$$F_{IVCRA} = \begin{cases} F_9 & , \quad F_9 > 0 \\ K_5 F_9 & , \quad F_9 \leq 0 \end{cases} \tag{AII.78}$$

Superior vena cava:

$$\frac{dV_{SVC}}{dt} = F_{UVSVC} - F_{SVCRA} \quad , \quad V_{SVC} \geq 0 \tag{AII.79}$$

$$P_{SVC} = (V_{SVC} - V_{USVC})/C_{SVC} \tag{AII.80}$$

$$C_{SVC} = \begin{cases} C_{SVCN} & , \quad V_{SVC} > V_{USVC} \\ K_6 C_{SVCN} & , \quad V_{SVC} \leq V_{USVC} \end{cases} \tag{AII.81}$$

$$F_{10} = \frac{(P_{SVC} - P_{RA} + G_{SVCRA})\, V^2_{SVC}}{R_{SVC} V^2_{USVC}} \tag{AII.82}$$

$$F_{SVCRA} = \begin{cases} F_{10} & , \quad F_{10} > 0 \\ K_5 F_{10} & , \quad F_{10} \leq 0 \end{cases} \tag{AII.83}$$

Time-varying compliances of atria and ventricles:

$$\frac{dU_{10}}{dt} = 1.0 \ (U_{10} \text{ set to zero at end of cardiac cycle}) \tag{AII.84}$$

$$T_{AS} = K_{22} + K_{23} T_H \tag{AII.85}$$

$$T_{AV} = T_{AS} - K_{24} \tag{AII.86}$$

$$T_{VS} = K_{25} + K_{26} T_H \tag{AII.87}$$

$$x_1 = \pi/T_{AS} \tag{AII.88}$$

$$x_2 = \pi/T_{VS} \tag{AII.89}$$

$$x_3 = \begin{cases} 0 \ , & U_{10} > T_{AS} \\ \sin (x_1 U_{10}), & U_{10} \leq T_{AS} \end{cases} \tag{AII.90}$$

$$x_4 = \begin{cases} U_{10} - T_{AV} & , \quad U_{10} > T_{AV} \\ 0 & , \quad U_{10} \leq T_{AV} \end{cases} \tag{AII.91}$$

$$x_5 = \begin{cases} 0 & , \ x_4 > T_{VS} \\ \sin (x_2 x_4) & , \ x_4 \leq T_{VS} \end{cases} \tag{AII.92}$$

$$a_{LA} = x_3 \{ b_2 \sigma_{LA} a_{LAS} - a_{LAD} \} + a_{LAD} \quad \text{(AII.93)}$$

$$a_{LV} = x_5 \{ b_2 \sigma_{LV} a_{LVS} - a_{LVD} \} + a_{LVD} \quad \text{(AII.94)}$$

$$a_{RA} = x_3 \{ b_2 \sigma_{RA} a_{RAS} - a_{RAD} \} + a_{RAD} \quad \text{(AII.95)}$$

$$a_{RV} = x_5 \{ b_2 \sigma_{RV} a_{RVS} - a_{RVD} \} + a_{RVD} \quad \text{(AII.96)}$$

Respiration:

$$\frac{dy_2}{dt} = 1.0 \quad (y_2 \text{ set to zero at end of respiratory cycle}) \quad \text{(AII.97)}$$

$$y_1 = \begin{cases} y_2 , & y_2 \leq T_{IE} \\ 0 , & y_2 > T_{IE} \end{cases} \quad \text{(AII.98)}$$

$$P_{TH} = K_1 + (K_2 - K_1) \sin (\pi y_1 / T_{IE}) \quad \text{(AII.99)}$$

$$P_{ABD} = K_3 + (K_4 - K_3) \sin (\pi y_1 / T_{IE}) \quad \text{(AII.100)}$$

Calculation of (MAP), (SV), (CO), (ETSR) :

$$(MAP) = \frac{1}{T_H} \int_{t_1}^{(t_1 + T_H)} P_{AO1} \, dt \quad (t_1 = \text{start of a cardiac cycle}) \quad \text{(AII.101)}$$

$$(SV) = \int_{t_1}^{(t_1 + T_H)} F_{LVAO1} \, dt \quad \text{(AII.102)}$$

$$(CO) = (SV)/T_H \quad \text{(AII.103)}$$

$$ETSR = (MAP)/(CO) \quad \text{(AII.104)}$$

True total systemic resistance: (TTSR)

$$R_A = R_{UA} + R_{HEAD}\,\sigma_{HEAD} + \frac{R_{UV}V^2_{UUVN}}{V^2_{UV}} \quad \frac{R_{SVC}V^2_{USVC}}{V^2_{SVC}} \tag{AII.105}$$

$$R_B = q_4\ \sigma_{BRONC}\ R_{BRONC} \tag{AII.106}$$

$$R_C = R_{CA} + q_4 R_{LEG}\sigma_{LEG} + {}^{R_{CV}V^2}{}_{UCVN} / V^2_{CV} \tag{AII.107}$$

$$R_D = q_4\ R_{ABD}\ \sigma_{ABD} \tag{AII.108}$$

$$R_E = R_C R_D/(R_C+R_D) + R_{AA} + R_{AV}\ V^2_{UAVN}/V^2_{AV} \tag{AII.109}$$

$$R_F = R_{IA} + q_4\ \sigma_{INT} R_{INT} + R_{IV}\ V^2_{UIVN}/V^2_{IV} \tag{AII.110}$$

$$R_G = R_E R_F/(R_E + R_F) + R_{IVC} V^2_{UIVC}/V^2_{IVC} \tag{AII.111}$$

$$R_H = R_{AO3} + R_B R_G/(R_B + R_G) \tag{AII.112}$$

$$R_I = R_{AO2} + R_A R_H/(R_A + R_H) \tag{AII.113}$$

$$(TTSR) = R_{COR}\ R_I/(R_{COR} + R_I) \tag{AII.114}$$

A.II.2 The Neural Control Model

Aortic arch baroreceptors:

$$\frac{dS_3}{dt} = (P_{AO2} - S_3)/\tau_1 \tag{AII.115}$$

$$\frac{dS_4}{dt} = (S_2 - S_4)/\tau_2 \tag{AII.116}$$

$$S_1 = \frac{dP_{AO2}}{dt} \tag{AII.117}$$

$$S_2 = \begin{cases} S_1 , & S_1 > 0 \\ 0 , & S_1 \leqslant 0 \end{cases} \qquad \text{(AII.118)}$$

$$S_5 = K_{13} (S_3 + K_{14} S_4 - K_{15}) \qquad \text{(AII.119)}$$

$$B_{AO2} = \begin{cases} S_5 , & S_5 > 0 \\ 0 , & S_5 \leqslant 0 \end{cases} \qquad \text{(AII.120)}$$

Carotid sinus baroreceptors:

$$\frac{dS_8}{dt} = (P_{UA} - S_8)/\tau_1 \qquad \text{(AII.121)}$$

$$\frac{dS_9}{dt} = (S_7 - S_9)/\tau_2 \qquad \text{(AII.122)}$$

$$S_6 = \frac{dP_{UA}}{dt} \qquad \text{(AII.123)}$$

$$S_7 = \begin{cases} S_6 , & S_6 > 0 \\ 0 , & S_6 \leqslant 0 \end{cases} \qquad \text{(AII.124)}$$

$$S_{10} = K_{13} (S_8 + K_{14} S_9 - K_{15}) \qquad \text{(AII.125)}$$

$$P_{UA} = \begin{cases} S_{10} , & S_{10} > 0 \\ 0 , & S_{10} \leqslant 0 \end{cases} \qquad \text{(AII.126)}$$

C.N.S. input function:

$$B = (1 - K_{16}) B_{AO2} + K_{16} B_{UA} \qquad \text{(AII.127)}$$

C.N.S. control of heart rate:

$$\frac{dU_4}{dt} = (U_1 - U_4)/U_3 \qquad \text{(AII.128)}$$

$$\frac{dU_6}{dt} = (U_5 - U_6)/\tau_3 \qquad \text{(AII.129)}$$

$$\frac{dU_7}{dt} = (U_6 - U_7)/\tau_4 \qquad \text{(AII.130)}$$

$$U_1 = \begin{cases} K_{17}(B - K_{18}) & , \quad B > K_{18} \\ 0 & , \quad B \leqslant K_{18} \end{cases} \qquad \text{(AII.131)}$$

$$U_2 = \frac{dU_1}{dt} \qquad \text{(AII.132)}$$

$$U_3 = \begin{cases} K_{19} & , \quad U_2 > 0 \\ K_{20} & , \quad U_2 \leqslant 0 \end{cases} \qquad \text{(AII.133)}$$

$$U_5 = \begin{cases} K_{18} & , \quad B > K_{18} \\ B & , \quad B \leqslant K_{18} \end{cases} \qquad \text{(AII.134)}$$

$$U_9 = \begin{cases} 2.0 & , \quad U_8 \geqslant 2.0 \\ U_8 & , \quad 0.3 < U_8 < 2.0 \\ 0.3 & , \quad U_8 \leqslant 0.3 \end{cases} \qquad \text{(AII.135)}$$

$$U_8 = K_{21} \cdot \sigma_H (U_4 + U_7) \qquad \text{(AII.136)}$$

$\dot{T}_H$ is set to the value of U_9 at the end of the cardiac cycle.

C.N.S. control of peripheral resistance:

$$\frac{dq_2}{dt} = (q_1 - q_2)/\tau_5 \qquad \text{(AII.137)}$$

$$\frac{dq_3}{dt} = (q_1 - q_3)/\tau_6 \qquad \text{(AII.138)}$$

$$q_4 = K_{29}q_3 + (1 - K_{29})\, q_2 \qquad \text{(AII.139)}$$

$$q_1 = \begin{cases} K_{27} & , \quad B > K_{18} \\ K_{28} & , \quad B \leqslant K_{18} \end{cases} \qquad \text{(AII.140)}$$

C.N.S. control of myocardial contractility:

$$\frac{db_2}{dt} = (b_1 - b_2)/\tau_8 \qquad \text{(AII.141)}$$

$$b_1 = \begin{cases} K_{34}, & B > K_{18} \\ K_{35}, & B \leqslant K_{18} \end{cases} \qquad \text{(AII.142)}$$

C.N.S. control of venous tone:

$$\frac{dd_2}{dt} = (d_1 - d_2)/\tau_7 \qquad \text{(AII.143)}$$

$$d_1 = \begin{cases} K_{30}, & B > K_{18} \\ K_{31}, & B \leqslant K_{18} \end{cases} \qquad \text{(AII.144)}$$

$$d_3 = 1 + K_{32}\,(d_2 - 1) \qquad \text{(AII.145)}$$

$$d_4 = 1 + K_{33}\,(d_2 - 1) \qquad \text{(AII.146)}$$

A.II.3 The Pharmacokinetics Model

Right atrium:

$$\frac{dm_{RA}}{dt} = \omega_{SVCRA} F_{SVCRA} + \omega_{COR} F_{COR} + \omega_{BRONC} F_{BRONC} + \omega_{IVCRA} F_{IVCRA} - \omega_{RARV} F_{RARV} - m_{RA}/\tau_9 \qquad \text{(AII.147)}$$

$$\omega_{RA} = m_{RA}/V_{RA} \qquad \text{(AII.148)}$$

Right ventricle:

$$\frac{m_{RV}}{dt} = \omega_{RARV} F_{RARV} - \omega_{RVPA} F_{RVPA} - m_{RV}/\tau_9 \qquad \text{(AII.149)}$$

$$\omega_{RV} = m_{RV}/V_{RV} \qquad \text{(AII.150)}$$

Pulmonary arteries:

$$\frac{dm_{PA}}{dt} = \omega_{RVPA}F_{RVPA} - \omega_{PAPV}F_{PAPV} - m_{PA}/\tau_9 \quad \text{(AII.151)}$$

$$\omega_{PA} = m_{PA}/V_{PA} \quad \text{(AII.152)}$$

Pulmonary veins:

$$\frac{dm_{PV}}{dt} = \omega_{PAPV}F_{PAPV} - \omega_{PVLA}F_{PVLA} - m_{PV}/\tau_9 \quad \text{(AII.153)}$$

$$\omega_{PV} = m_{PA}/V_{PV} \quad \text{(AII.154)}$$

Left atruim:

$$\frac{dm_{LA}}{dt} = \omega_{PVLA}F_{PVLA} - \omega_{LALV}F_{LALV} - m_{LA}/\tau_9 \quad \text{(AII.155)}$$

$$\omega_{LA} = m_{LA}/V_{LA} \quad \text{(AII.156)}$$

Left ventricle:

$$\frac{dm_{LV}}{dt} = \omega_{LALV}F_{LALV} - \omega_{LVAO1}F_{LVAO1} - m_{LV}/\tau_9 \quad \text{(AII.157)}$$

$$\omega_{LV} = m_{LV}/V_{LV} \quad \text{(AII.158)}$$

Ascending aorta:

$$\frac{dm_{AO1}}{dt} = \omega_{LVAO1}F_{LVAO1} - \omega_{AO1AO2}F_{AO1AO2} - \omega_{COR}F_{COR} - m_{AO1}/\tau_9 \quad \text{(AII.159}$$

$$\omega_{AO1} = m_{AO1}/V_{AO1} \quad \text{(AII.160)}$$

Aortic arch:

$$\frac{dm_{AO2}}{dt} = \omega_{AO1AO2}F_{AO1AO2} - \omega_{AO2UA}F_{AO2UA} - \omega_{AO2AO3}F_{AO2AO3} - m_{AO2}/\tau_9 \quad \text{(AII.161)}$$

$$\omega_{AO2} = m_{AO2}/V_{AO2} \quad \text{(AII.162)}$$

Head and arms arteries:

$$\frac{dm_{UA}}{dt} = \omega_{AO2UA}F_{AO2UA} - \omega_{UAUV}F_{UAUV} - m_{UA}/\tau_9 \quad \text{(AII.163)}$$

$$\omega_{UA} = m_{UA}/V_{UA} \quad \text{(AII.164)}$$

Head and arms veins:

$$\frac{dm_{UV}}{dt} = \omega_{UAUV}F_{UAUV} - \omega_{UVSVC}F_{UVSVC} - m_{UV}/\tau_9 + M\delta(t) \quad \text{(AII.165)}$$

(assuming mass M injected at t = 0)

$$\omega_{UV} = m_{UV}/V_{UV} \quad \text{(AII.166)}$$

Thoracic aorta:

$$\frac{dm_{AO3}}{dt} = \omega_{AO2AO3}F_{AO2AO3} - \omega_{BRONC}F_{BRONC} - \omega_{AO3IA}F_{AO3IA} - \omega_{AO3AA}F_{AO3AA} - m_{AO3}/\tau_9 \quad \text{(AII.167)}$$

$$\omega_{AO3} = m_{AO3}/V_{AO3} \quad \text{(AII.168)}$$

Intestinal arteries:

$$\frac{dm_{IA}}{dt} = \omega_{AO3IA}F_{AO3IA} - \omega_{IAIV}F_{IAIV} - m_{IA}/\tau_9 \quad \text{(AII.169)}$$

$$\omega_{IA} = m_{IA}/V_{IA} \quad \text{(AII.170)}$$

Inferior vena cava:

$$\frac{dm_{IVC}}{dt} = \omega_{AVIVC}F_{AVIVC} + \omega_{IVIVC}F_{IVIVC} - \omega_{IVCRA}F_{IVCRA} - m_{IVC}/\tau_9 \quad \text{(AII.181)}$$

$$\omega_{IVC} = m_{IVC}/V_{IVC} \quad \text{(AII.182)}$$

Superior vena cava:

$$\frac{dm_{SVC}}{dt} = \omega_{UVSVC}F_{UVSVC} - \omega_{SVCRA}F_{SVCRA} - m_{SVC}/\tau_9 \quad \text{(AII.183)}$$

$$\omega_{SVC} = m_{SVC}/V_{SVC} \quad \text{(AII.184)}$$

Concentrations appropriate to directions of flow:

$$\omega_{SVCRA} = \begin{cases} \omega_{SVC}, & F_{SVCRA} > 0 \\ \omega_{RA}, & F_{SVCRA} \leq 0 \end{cases} \quad \text{(AII.185)}$$

$$\omega_{COR} = \begin{cases} \omega_{AO1}, & F_{COR} > 0 \\ \omega_{RA}, & F_{COR} \leq 0 \end{cases} \quad \text{(AII.186)}$$

$$\omega_{BRONC} = \begin{cases} \omega_{AO3}, & F_{BRONC} > 0 \\ \omega_{RA}, & F_{BRONC} \leq 0 \end{cases} \quad \text{(AII.187)}$$

$$\omega_{IVCRA} = \begin{cases} \omega_{IVC}, & F_{IVCRA} > 0 \\ \omega_{RA}, & F_{IVCRA} \leq 0 \end{cases} \quad \text{(AII.188)}$$

$$\omega_{RARV} = \begin{cases} \omega_{RA}, & F_{RARV} > 0 \\ \omega_{RV}, & F_{RARV} \leq 0 \end{cases} \quad \text{(AII.189)}$$

$$\omega_{RVPA} = \begin{cases} \omega_{RV}, & F_{RVPA} > 0 \\ \omega_{PA}, & F_{RVPA} \leq 0 \end{cases} \quad \text{(AII.190)}$$

$$\omega_{PAPV} = \begin{cases} \omega_{PA}, & F_{PAPV} > 0 \\ \omega_{PV}, & F_{PAPV} \leq 0 \end{cases} \quad \text{(AII.191)}$$

Intestinal veins:

$$\frac{dm_{IV}}{dt} = \omega_{IAIV}F_{IAIV} - \omega_{IVIVC}F_{IVIVC} - m_{IV}/\tau_9 \quad \text{(AII.171)}$$

$$\omega_{IV} = m_{IV}/V_{IV} \quad \text{(AII.172)}$$

Abdominal arteries:

$$\frac{dm_{AA}}{dt} = \omega_{AO3AA}F_{AO3AA} - \omega_{AAAV}F_{AAAV} - \omega_{AACA}F_{AACA} - m_{AA}/\tau_9 \quad \text{(AII.173)}$$

$$\omega_{AA} = m_{AA}/V_{AA} \quad \text{(AII.174)}$$

Abdominal veins:

$$\frac{dm_{AV}}{dt} = \omega_{AAAV}F_{AAAV} + \omega_{CVAV}F_{CVAV} - \omega_{AVIVC}F_{AVIVC} - m_{AV}/\tau_9 \quad \text{(AII.175)}$$

$$\omega_{AV} = m_{AV}/V_{AV} \quad \text{(AII.176)}$$

Leg arteries:

$$\frac{dm_{CA}}{dt} = \omega_{AACA}F_{AACA} - \omega_{CACV}F_{CACV} - m_{CA}/\tau_9 \quad \text{(AII.177)}$$

$$\omega_{CA} = m_{CA}/V_{CA} \quad \text{(AII.178)}$$

Leg veins:

$$\frac{dm_{CV}}{dt} = \omega_{CACV}F_{CACV} - \omega_{CVAV}F_{CVAV} - m_{CV}/\tau_9 \quad \text{(AII.179)}$$

$$\omega_{CV} = m_{CV}/V_{CV} \quad \text{(AII.180)}$$

$$\omega_{PVLA} = \begin{cases} \omega_{PV}, & F_{PVLA} > 0 \\ \omega_{LA}, & F_{PVLA} \leqslant 0 \end{cases} \quad \text{(AII.192)}$$

$$\omega_{LALV} = \begin{cases} \omega_{LA}, & F_{LALV} > 0 \\ \omega_{LV}, & F_{LALV} \leqslant 0 \end{cases} \quad \text{(AII.193)}$$

$$\omega_{LVAO1} = \begin{cases} \omega_{LV}, & F_{LVAO1} < 0 \\ \omega_{AO1}, & F_{LVAO1} \geqslant 0 \end{cases} \quad \text{(AII.194)}$$

$$\omega_{AO1AO2} = \begin{cases} \omega_{AO1}, & F_{AO1AO2} < 0 \\ \omega_{AO2}, & F_{AO1AO2} \geqslant 0 \end{cases} \quad \text{(AII.195)}$$

$$\omega_{AO2UA} = \begin{cases} \omega_{AO2}, & F_{AO2UA} > 0 \\ \omega_{UA}, & F_{AO2UA} \leqslant 0 \end{cases} \quad \text{(AII.196)}$$

$$\omega_{AO2AO3} = \begin{cases} \omega_{AO2}, & F_{AO2AO3} > 0 \\ \omega_{AO3}, & F_{AO2AO3} \leqslant 0 \end{cases} \quad \text{(AII.197)}$$

$$\omega_{UAUV} = \begin{cases} \omega_{UA}, & F_{UAUV} > 0 \\ \omega_{UV}, & F_{UAUV} \leqslant 0 \end{cases} \quad \text{(AII.198)}$$

$$\omega_{UVSVC} = \begin{cases} \omega_{UV}, & F_{UVSVC} > 0 \\ \omega_{SVC}, & F_{UVSVC} \leqslant 0 \end{cases} \quad \text{(AII.199)}$$

$$\omega_{AO3IA} = \begin{cases} \omega_{AO3}, & F_{AO3IA} > 0 \\ \omega_{IA}, & F_{AO3IA} \leqslant 0 \end{cases} \quad \text{(AII.200)}$$

$$\omega_{AO3AA} = \begin{cases} \omega_{AO3}, & F_{AO3AA} > 0 \\ \omega_{AA}, & F_{AO3AA} \leqslant 0 \end{cases} \quad \text{(AII.201)}$$

$$\omega_{IAIV} = \begin{cases} \omega_{IA}, & F_{IAIV} > 0 \\ \omega_{IV}, & F_{IAIV} \leqslant 0 \end{cases} \quad \text{(AII.202)}$$

$$\omega_{IVIVC} = \begin{cases} \omega_{IV}, & F_{IVIVC} > 0 \\ \omega_{IVC}, & F_{IVIVC} \leqslant 0 \end{cases} \quad \text{(AII.203)}$$

$$\omega_{AAAV} = \begin{cases} \omega_{AA}, & F_{AAAV} > 0 \\ \omega_{AV}, & F_{AAAV} \leqslant 0 \end{cases} \qquad \text{(AII.204)}$$

$$\omega_{AACA} = \begin{cases} \omega_{AA}, & F_{AACA} > 0 \\ \omega_{CA}, & F_{AACA} \leqslant 0 \end{cases} \qquad \text{(AII.205)}$$

$$\omega_{CVAV} = \begin{cases} \omega_{CV}, & F_{CVAV} > 0 \\ \omega_{AV}, & F_{CVAV} \leqslant 0 \end{cases} \qquad \text{(AII.206)}$$

$$\omega_{AVIVC} = \begin{cases} \omega_{AV}, & F_{AVIVC} > 0 \\ \omega_{IVC}, & F_{AVIVC} \leqslant 0 \end{cases} \qquad \text{(AII.207)}$$

$$\omega_{CACV} = \begin{cases} \omega_{CA}, & F_{CACV} > 0 \\ \omega_{CV}, & F_{CACV} \leqslant 0 \end{cases} \qquad \text{(AII.208)}$$

Effect of drug on heart rate:

$$\sigma_H = \begin{cases} 1 + \sigma_2 \omega_{RA}, & \text{bradycardia} \\ \dfrac{1}{1 + \sigma_2 \omega_{RA}}, & \text{tachycardia} \end{cases} \qquad \text{(AII.209)}$$

Effect of drug on peripheral resistance:

$$\sigma_{BRONC} = \begin{cases} 1 + \sigma_1 \omega_{AO3}, & \text{vasoconstriction} \\ \dfrac{1}{1 + \sigma_1 \omega_{AO3}}, & \text{vasodilatation} \end{cases} \qquad \text{(AII.210)}$$

$$\sigma_{INT} = \begin{cases} 1 + \sigma_1 \omega_{IA}, & \text{vasoconstriction} \\ \dfrac{1}{1 + \sigma_1 \omega_{IA}}, & \text{vasodilatation} \end{cases} \qquad \text{(AII.211)}$$

$$\sigma_{ABD} = \begin{cases} 1 + \sigma_1 \omega_{AA}, & \text{vasoconstriction} \\ \dfrac{1}{1 + \sigma_1 \omega_{AA}}, & \text{vasodilatation} \end{cases} \qquad \text{(AII.212)}$$

$$\sigma_{LEG} = \begin{cases} 1 + \sigma_1 \omega_{CA} & , \text{ vasoconstriction} \\ \dfrac{1}{1 + \sigma_1 \omega_{CA}} & , \text{ vasodilatation} \end{cases} \quad \text{(AII.213)}$$

$$\sigma_{HEAD} = \begin{cases} 1 + \sigma_1 \omega_{UA} & , \text{ vasoconstriction} \\ \dfrac{1}{1 + \sigma_1 \omega_{UA}} & , \text{ vasodilatation} \end{cases} \quad \text{(AII.214)}$$

Effect of drug on myocardial contractility:

$$\sigma_{RA} = \begin{cases} 1 + \sigma_3 \omega_{RA} & , \text{ positive inotropy} \\ \dfrac{1}{1 + \sigma_3 \omega_{RA}} & , \text{ negative inotropy} \end{cases} \quad \text{(AII.215)}$$

$$\sigma_{RV} = \begin{cases} 1 + \sigma_3 \omega_{RV} & , \text{ positive inotropy} \\ \dfrac{1}{1 + \sigma_3 \omega_{RV}} & , \text{ negative inotropy} \end{cases} \quad \text{(AII.216)}$$

$$\sigma_{LA} = \begin{cases} 1 + \sigma_3 \omega_{LA} & , \text{ positive inotropy} \\ \dfrac{1}{1 + \sigma_3 \omega_{LA}} & , \text{ negative inotropy} \end{cases} \quad \text{(AII.217)}$$

$$\sigma_{LV} = \begin{cases} 1 + \sigma_3 \omega_{LV} & , \text{ positive inotropy} \\ \dfrac{1}{1 + \sigma_3 \omega_{LV}} & , \text{ negative inotropy} \end{cases} \quad \text{(AII.218)}$$

APPENDIX III

A MATHEMATICAL MODEL OF THE HUMAN RENAL-ARTIFICIAL KIDNEY MACHINE SYSTEM

List of Symbols Used

a	Cardiac function parameter
A	Angiotensin II plasma concentration; area of dialysis membrane
ADH	ADH plasma concentration
ADHS	Total ADH secretion rate
ADHSP	ADH secretion rate due to plasma osmolality
ADHSV	ADH secretion rate due to excess fluid volume
ALD	Aldosterone plasma concentration
ALS	Total aldosterone secretion rate
ALSA	Aldosterone secretion rate due to angiotensin II concentration
ALSK	Aldosterone secretion rate due to plasma potassium concentration
AP	Arterial pressure
AS	Total angiotensin II secretion rate
AVOS	Average intracellular and extracellular fluid osmolality
b	Cardiac function parameter
BMRC	Core basal metabolic rate
BMRS	Skin basal metabolic rate
BV	Blood volume
c	Specific heat of blood
C_{Bi}, C_{Di}	Concentration of sodium potassium in blood and dialysate
C_c, C_s	Thermal capacities of core and skin
C_I, C_E	Concentration of urea or creatinine in intracellular and extracellular fluid
CE	Cardiac effectiveness
CEK	Cardiac effectiveness due to plasma potassium concentration
CENA	Cardiac effectiveness due to plasma sodium concentration
CO	Cardiac output
DADH	Removal rate of ADH
DAP_o	Patient - specific correction factor for arterial pressure
DTPR	Total peripheral resistance due to plasma angiotensin II concentration
DWV	Excess fluid volume

E	Extracellular fluid volume
E_{in}	Ingestion rate of fluid
EBDT	Fraction of intraluminal fluid reabsorbed in the distal tubules
EBLH	Fraction of intraluminal fluid reabsorbed in the loop of Henle
EDTR	Fluid reabsorbed from the distal tubules
EFDT	Fluid flow into the distal tubules
EFLH	Fluid flow into the loop of Henle
ELHR	Fluid reabsorbed in the loop of Henle
EPTR	Fluid reabsorbed in the proximal tubules
FACT1-4	Kidney failure parameters, defined in A.III.8.1
FLUMIN	Fluid ingested per minute
FNA	Sodium flow through the glomerular membrane
G	Generation rate of urea or creatinine
G'	Generation rate of urea or creatinine in renal failure
GFR	Glomerular filtration rate
GTB	Glomerular tubular balance
I	Intracellular fluid volume
IC	Intracellular osmotic constant factor
IHL	Insensible heat loss
IK	Intracellular potassium concentration
IK'	Intracellular potassium concentration after osmotic balance
INA	Intracellular sodium concentration
INA'	Intracellular sodium concentration after osmotic balance
IOS	Intracellular osmolality
$k_{I,E}$	Cell permeability constant for urea or creatinine
K	Permeability of dialysis membrane for electrolytes or waste products
K_{CS}	Thermal conductance between core and skin
K_r	Rate of clearance of urea or creatinine through urine
K_{SE}	Surface to environmental heat transfer coefficient
MSP	Mean systemic pressure
PC	Plasma (or extracellular) osmotic constant factor
PCP	Pressure across dialysis membrane
PK	Plasma potassium concentration
PK'	Plasma potassium concentration after osmotic balance

PNA	Plasma sodium concentration
PNA'	Plasma sodium concentration after osmotic balance
POS	Plasma (or extracellular) osmolality
POTDIA	Concentration of potassium in the dialysate
POTMIN	Potassium ingested per minute
PV	Plasma volume
Q_B	Blood flowrate through dialysis machine
R	Plasma renin concentration
RAP	Right atrial pressure
RHL	Heat loss in core
RS	Renin secretion rate
RVR	Resistance to venous return
SBF	Core to skin blood flow
SDTR	Sodium reabsorbed from the distal tubules
SFDT	Sodium flow into the distal tubules
SFLH	Sodium flow into the loop of Henle
SLHR	Sodium reabsorbed from the loop of Henle
SODDIA	Concentration of sodium in the dialysate
SODMIN	Sodium ingested per minute
SPTR	Sodium reabsorbed from the proximal tubules
STPR	Skin total peripheral resistance
T_c	Core temperature
T_E	Environmental temperature
T_s	Skin temperature
TEK	Total extracellular potassium mass
TENA	Total extracellular sodium mass
TIK	Total intracellular potassium mass
TINA	Total intracellular sodium mass
TPR	Total peripheral resistance
TPR_{TH}	Total peripheral resistance due to thermoregulation
UFL	Urine fluid flow rate
UK	Urine potassium flow rate
UKAL	Urine potassium flow rate due to aldosterone
UKH	Urine potassium flow due to plasma potassium concentration
ULTRF	Fluid loss rate through dialysis membrane
UNA	Urine sodium flow rate

VR	Venous return
ρ	Blood density

Units Employed

ADH concentration	$mU.\ell^{-1}$
Aldosterone and Angiotensin II Concentrations	$ng.\ell^{-1}$
Osmolality	$mosm.\ell^{-1}$
Pressure	mm Hg
Renin concentration	$GU.\ell^{-1}$ (Goldblatt units)
Sodium and Pottasium masses	m Eq
Temperature	^{o}C
Thermal capacitance	$cals.^{o}C^{-1}$
Thermal conductivity	$cals.\ min^{-1}\ ^{o}C^{-1}$
Time	min.
Urea and creatinine masses	g.

The Complete Mathematical Model

(For a description of model development, consult Uttamsingh, 1981, Chapter 4).

A.III.1 Thermoregulatory System Submodel

A.III.1.1 Passive thermal system submodel

$$C_c \frac{d T_c}{dt} = BMRC - K_{CS} (T_c - T_s) - \rho cSBF (T_c - T_s) - RHL \quad .. \quad (1)$$

$$C_s \frac{d T_s}{dt} = BMRS + K_{CS} (T_c - T_s) + \rho cSBF (T_c - T_s) - K_{SE}(T_s - T_E) - IHL \quad .. \quad (2)$$

A.III.1.2 Submodel of thermal control

$$\left.\begin{array}{ll} STPR = 4484.3 & \text{if } T_c < 35.0 \\ STPR = -2882.8\ T_c + 105382.0 & \text{if } 35.0 \leq T_c < 36.4 \\ STPR = 19.3\ T_s - 209.8 & \text{if } 36.4 \leq T_c < 37.0 \text{ and } T_s \leq 34.1 \\ STPR = 36.9\ T_s - 809.9 & \text{if } 36.5 \quad T_c < 37.0 \text{ and } T_s > 34.1 \\ STPR = -256.2\ T_c + 9927.8 & \text{if } 37.0 \leq T_c < 38.5 \\ STPR = 64.1 & \text{if } 38.5 \leq T_c \end{array}\right\} \quad .. \quad (3)$$

A.III.1.3 Interactions with the cardiovascular system submodel

$$SBF = \frac{AP}{STPR} \quad .. \quad .. \quad .. \quad (4)$$

$$TPR_{TH} = \frac{20.934\ STPR}{20.934 + STPR} \quad .. \quad .. \quad .. \quad (5)$$

A.III.2 Cardiovascular System Submodel

A.III.2.1 Extracellular fluid and blood volumes

$$\frac{dE}{dt} = E_{in} - UFL \qquad .. \quad .. \quad .. \quad (6)$$

$$\left.\begin{aligned} BV &= 0.33E && \text{if } E \leq 21.0 \\ BV &= 0.0156E + 6.6 && \text{if } E > 21.0 \end{aligned}\right\} \qquad .. \quad .. \quad (7)$$

$$MSP = 3.5\,BV - 10.5 \qquad .. \quad .. \quad .. \quad (8)$$

A.III.2.2 Total peripheral resistance

$$\left.\begin{aligned} DTPR &= 0.037A - 1.0 && \text{if } A \leq 27.0 \\ DTPR &= 5.44\ \log_{10}A - 7.8 && \text{if } A > 27.0 \end{aligned}\right\} \qquad .. \quad (9)$$

$$TPR = TPR_{TH} + DTPR \qquad .. \quad .. \quad .. \quad (10)$$

A.III.2.3. Effect of N_a and K concentrations on cardiac effectiveness

$$\left.\begin{aligned} CENA &= 1.0 && \text{if PNA} < 148.0 \\ CENA &= -0.0125\,PNA + 2.85 && \text{if PNA} \geq 148.0 \end{aligned}\right\} \qquad .. \quad (11)$$

$$\left.\begin{aligned} CEK &= 1.0 && \text{if PK} < 6.5 \\ CEK &= -0.065\,PK + 1.43 && \text{if PK} \geq 6.5 \end{aligned}\right\} \qquad .. \quad (12)$$

$$CE = 0.5\ (CENA + CEK) \qquad .. \quad .. \quad .. \quad (13)$$

A.III.2.4 Systemic function curves

$$RVR = 0.07\ TPR \qquad .. \quad .. \quad .. \quad (14)$$

$$VR = \frac{(MSP - RAP)}{RVR} \qquad .. \quad .. \quad .. \quad (15)$$

(equation (15) is solved simultaneously with equation (16); see A.III.2.6. This corresponds to the intersection of systemic function (venous return) and cardiac function (cardiac output) curves - Guyton's 1955 method).

A.III.2.5 Cardiac function curves

$$CO = a\ RAP + b \quad .. \quad .. \quad .. \quad (16)$$

where: (i) for $RAP \leq 2.0$

$a = 3.0,\ b = 5.25$	if $CE > 0.85$
$a = 2.5,\ b = 3.75$	if $0.85 \geq CE > 0.62$
$a = 1.7,\ b = 2.125$	if $0.62 > CE$

(ii) for $2.0 < RAP \leq 4.0$

$a = 0.875,\ b = 9.5$	if $CE > 0.85$
$a = 0.625,\ b = 7.5$	if $0.85 \geq CE \geq 0.62$
$a = 0.375,\ b = 4.75$	if $0.62 > CE$

(iii) for $4.0 < RAP$

$a = 0.0,\ b = 13.0$	if $CE > 0.85$
$a = 0.0,\ b = 8.75$	if $0.85 \geq CE \geq 0.62$
$a = 0.0,\ b = 6.25$	if $0.62 > CE$

A.III.2.6 Determination of right arterial pressure, cardiac output, and arterial pressure

$$RAP = \frac{MSP - b\ RVR}{1 + a\ RVR} \quad .. \quad .. \quad .. \quad (17)$$

$$CO = \frac{a\ MSP + b}{1 + a\ RVR} \quad .. \quad .. \quad .. \quad (18)$$

$$AP = CO.TPR + DAP_o \quad .. \quad .. \quad .. \quad (19)$$

A.III.3 Kidney Function Submodel

A.III.3.1 Submodel of glomerular function

$$\left.\begin{array}{lll} GFR = 0.0 & \text{if} & AP \leq 20.0 \\ GFR = 1.92\,AP - 38.4 & \text{if} & 20.0 < AP \leq 75.0 \\ GFR = -0.00808\,AP^2 + 2.195\,AP - 13.6 & \text{if} & 75.0 < AP \leq 120.0 \\ GFR = 0.035\,AP + 129.2 & \text{if} & 120.0 < AP \end{array}\right\} \quad (20)$$

$$FNA = \frac{GFR \,.\, PNA}{1000.0} \quad .. \quad .. \quad .. \quad (21)$$

A.III.3.2 Proximal tubule segment submodel

(i) Sodium

$$GTB = -0.0357\ PNA + 5.815 \qquad \text{where } 0.75 \leq GTB \leq 1.0 \quad .. \quad (22)$$

$$SPTR = GTB.FNA \quad .. \quad .. \quad .. \quad .. \quad (23)$$

$$SFLH = FNA - SPTP \quad .. \quad .. \quad .. \quad .. \quad (24)$$

(ii) Water

$$EPTR = GTB \,.\, GFR \quad .. \quad .. \quad .. \quad .. \quad (25)$$

$$EFLH = GFR - EPTR \quad .. \quad .. \quad .. \quad .. \quad (26)$$

A.III.3.3 Submodel of the Loop of Henle

(i) Sodium

$$SLHR = 0.8\ SFLH \quad .. \quad .. \quad .. \quad .. \quad (27)$$

$$SFDT = SFLH - SLHR \quad .. \quad .. \quad .. \quad .. \quad (28)$$

(ii) Water

$$EBLH = -0.01\ EFLH + 0.65 \quad .. \quad .. \quad .. \quad (29)$$

$$ELHR = EBLH \,.\, EFLH \quad .. \quad .. \quad .. \quad .. \quad (30)$$

$$EFDT = EFLH - ELHR \quad .. \quad .. \quad .. \quad .. \quad (31)$$

A.III.3.4 Submodel of distal and collecting segments

(i) The action of ADH

$$\left.\begin{array}{ll} EBDT = 0.0 & \text{if } ADH \leq 0.765 \\ EBDT = 0.383\,ADH - 0.293 & \text{if } 0.765 < ADH \leq 3.0 \\ EBDT = -0.0383\,ADH^2 + 0.364\,ADH + 0.109 & \text{if } 3.0 < ADH \leq 5.0 \\ EBDT = 0.0012\,ADH + 0.9653 & \text{if } 5.0 < ADH \end{array}\right\} \quad .. \quad (32)$$

$$EDTR = EBDT \,.\, EFDT \quad .. \quad .. \quad .. \quad .. \quad (33)$$

$$VFL = EFDT - EDTR \quad .. \quad .. \quad .. \quad .. \quad (34)$$

(ii) The action of aldosterone

$$\left.\begin{array}{ll} SDTR = 0.6\,SFDT & \text{if } ALD \leq 0.0 \\ SDTR = SFDT\,(0.003\,ALD + 0.596) & \text{if } 0.0 < ALD \leq 85.0 \\ SDTR = SFDT\,(0.00021\,ALD + 0.833) & \text{if } 85.0 < ALD \leq 800.0 \\ SDTR = SFDT & \text{if } 800.0 < ALD \end{array}\right\} \quad .. \quad (35)$$

$$UNA = SFDT - SDTR \quad .. \quad .. \quad .. \quad .. \quad (36)$$

$$UKH = 0.107\,PK - 0.505 \quad .. \quad .. \quad .. \quad (37)$$

$$\left.\begin{array}{ll} UKAL = 0.00028\,ALD - 0.0062 & \text{if } ALD \leq 85.0 \\ UKAL = 0.00009\,ALD + 0.0224 & \text{if } 85.0 < ALD \end{array}\right\} \quad .. \quad (38)$$

$$UK = UKH + UKAL \quad .. \quad .. \quad .. \quad .. \quad (39)$$

A.III.4 Submodel of Hormonal System

A.III.4.1 Submodel of ADH system

$$POS = 2.11\,PNA \quad .. \quad .. \quad .. \quad .. \quad (40)$$

$$\left.\begin{array}{ll} ADHSP = 0.348\,POS - 103.43 & \text{if } POS \geq 299.5 \\ ADHSP = 0.0285\,POS - 8.04 & \text{if } POS < 299.5 \end{array}\right\} \quad .. \quad (41)$$

$$DWV = E - E_N \quad .. \quad .. \quad .. \quad .. \quad (42)$$

$$\left.\begin{array}{ll} ADHSV = 0.0 & \text{if } DWV \geq 1.8 \\ ADHSV = 0.15 - 0.083\,DWV & \text{if } 1.87 > DWV \geq 1.0 \\ ADHSV = 0.813 - 0.75\,DWV & \text{if } 1.0 > DWV \geq -1.2 \\ ADHSV = 1.7. & \text{if } -1.2 > DWV \end{array}\right\} \quad (43)$$

$$\left.\begin{array}{ll} ADHS = \dfrac{17.0\,DWV \,.\, ADHSV + ADHSP}{17.0 + DWV} & \text{if } POS > 299.6 \text{ and } DWV > 2.0 \\ ADHS = \dfrac{(33.0\,DWV - 32.0)\,ADHSV + ADHSP}{(33.0\,DWV - 32.0) + 1.0} & \text{if } POS > 299.6 \text{ and } 1.0 \leq DWV \leq 2.0 \end{array}\right\} \quad (44)$$

$$ADHS = 0.5\,(ADHSV + ADHSP) \qquad \text{for all other conditions}$$

$$\left.\begin{array}{ll} DADH = 0.206 & \text{if } ADH > 4.0 \\ DADH = 0.374 - 0.042\,ADH & \text{if } ADH \leq 4.0 \end{array}\right\} \quad (45)$$

$$PV = 0.6\,BV \qquad .. \qquad .. \qquad .. \qquad .. \quad (46)$$

$$\frac{d\,AD + 1}{dt} = \frac{ADHS - ADH \,.\, DADH}{PV} \qquad .. \qquad .. \qquad .. \quad (47)$$

A.III.4.2 Submodel of the renin - angiotensin II - aldosterone system

(i) Renin

$$RS = 0.0163 - 0.0093\,SFDT \qquad .. \qquad .. \qquad .. \quad (48)$$

$$\frac{dR}{dt} = \frac{RS - 0.135\,R}{PV} \qquad .. \qquad .. \qquad .. \qquad .. \quad (49)$$

(ii) Angiotensin II

$$AS = 583.3\,R\,.\,PV \quad .. \quad .. \quad .. \quad .. \quad (50)$$

$$\frac{dA}{dt} = \frac{AS - 4.04\,A}{PV} \quad .. \quad .. \quad .. \quad .. \quad (51)$$

(iii) Aldosterone

$$\left.\begin{aligned} ALSA &= A && \text{if } A \leq 18.0 \\ ALSA &= 4.43A - 61.7 && \text{if } 18.0 < A \leq 34.0 \\ ALSA &= 0.78A + 62.5 && \text{if } 34.0 < A \end{aligned}\right\} \quad .. \quad (52)$$

$$ALSK = 21.64\,PK - 55.5 \quad .. \quad .. \quad .. \quad (53)$$

$$ALS = 0.25\,(3.0\,ALSA + ALSK) \quad .. \quad .. \quad .. \quad (54)$$

$$\frac{d\,ALD}{dt} = \frac{ALS - 0.62\,ALD}{PV} \quad .. \quad .. \quad .. \quad (55)$$

A.III.5 Submodel of the Artificial Kidney Machine

A.III.5.1 Ultrafiltration of water

$$\left.\begin{aligned} ULTRF &= 0.0139\,PCP + 0.7 && \text{if } PCP < 100.0 \\ ULTRF &= 0.042\,PCP - 2.1 && \text{if } PCP \geq 100.0 \end{aligned}\right\} \quad .. \quad (56)$$

A.III.5.2 Diffusion of electrolytes and water products

For concentrations, C_{Bi} and C_{Di}, of species i in blood and dialysate, respectively:

where $K_1 = 0.10$ for sodium

$K_2 = 0.05$ for potassium

$K_3 = 0.07$ for urea

$K_4 = 0.06$ for creatinine

$$\frac{d(C_{Bi} \cdot E)}{dt} = Q_B \, (C_{Bi} - C_{Di}) \left[\exp \left(- \frac{K.A}{Q_B} \right) - 1 \right] \quad .. \quad (57)$$

A.III.6 Submodel of Fluid and Electrolyte Balance

A.III.6.1 Sodium and potassium balance

(i) Off-dialysis

$$\frac{d\,(TENA)}{dt} = SODMIN - UNA \quad .. \quad .. \quad .. \quad (58)$$

$$\frac{d\,(TEK)}{dt} = POTMIN - UK \quad .. \quad .. \quad .. \quad (59)$$

(ii) On-dialysis

$$\frac{d\,(TENA)}{dt} = Q_B \, (PNA - SODDIA) \left[\exp \left(- \frac{K_1.A}{Q_B} \right) - 1 \right] + SODMIN - UNA' \quad (60)$$

$$\frac{d\,(TEK)}{dt} = Q_B \, (PK - POTDIA) \left[\exp \left(- \frac{K_2.A}{Q_B} \right) - 1 \right] + POTMIN - UK' \quad (61)$$

(iii) Concentrations of extracellular sodium and potassium

$$PNA = \frac{TENA}{E} \quad .. \quad .. \quad .. \quad .. \quad (62)$$

$$PK = \frac{TEK}{E} \quad .. \quad .. \quad .. \quad .. \quad (63)$$

(iv) Balance and concentrations of intracellular sodium and potassium

$$\frac{d\,(TINA)}{dt} = 0 \quad .. \quad .. \quad .. \quad .. \quad (64)$$

$$\frac{d\,(TINA)}{dt} = 0 \quad .. \quad .. \quad .. \quad .. \quad (64)$$

$$\frac{d\,(TIK)}{dt} = 0 \quad .. \quad .. \quad .. \quad .. \quad (65)$$

$$INA = \frac{TINA}{I} \quad .. \quad .. \quad .. \quad .. \quad (66)$$

$$IK = \frac{TIK}{I} \quad .. \quad .. \quad .. \quad .. \quad (67)$$

A.III.6.2 Fluid balance (osmosis) submodel

(i) Off-dialysis

$$\frac{dE}{dt} = FLUMIN - UFL \quad .. \quad .. \quad .. \quad .. \quad (68)$$

(ii) On-dialysis

$$\frac{dE}{dt} = FLUMIN - UFL - ULTRF \quad .. \quad .. \quad .. \quad (69)$$

(iii) Osmosis

$$POS = PNA + PK + PC \quad .. \quad .. \quad .. \quad .. \quad (70)$$

$$IOS = INA + IK\ I\ IC \quad .. \quad .. \quad .. \quad .. \quad (71)$$

$$AVOS = \frac{POS \,.\, E + IOS \,.\, I}{E + I} \quad .. \quad .. \quad .. \quad (72)$$

$$E' = \frac{POS \,.\, E}{AVOS} \quad .. \quad .. \quad .. \quad .. \quad (73)$$

$$I' = \frac{IOS \,.\, I}{AVOS} \quad .. \quad .. \quad .. \quad .. \quad (74)$$

$$PNA' = \frac{PNA \cdot AVOS}{POS} \quad .. \quad .. \quad .. \quad .. \quad (75)$$

$$PK' = \frac{PK \cdot AVOS}{POS} \quad .. \quad .. \quad .. \quad (76)$$

$$INA' = \frac{INA \cdot AVOS}{IOS} \quad .. \quad .. \quad .. \quad .. \quad (77)$$

$$IK' = \frac{IK \cdot AVOS}{IOS} \quad .. \quad .. \quad .. \quad .. \quad (78)$$

A.III.7 Urea and Creatinine Dynamics Submodel

(Equations are written for the concentrations, C_I and C_E, of a general species. Parameter values for urea and creatinine are given in (iii).).

(i) Off-dialysis (normal)

$$\frac{d(C_I \cdot I)}{dt} = G - k_{I,E}(C_I - C_E) \quad .. \quad .. \quad .. \quad (79)$$

$$\frac{d(C_E \cdot E)}{dt} = k_{I,E}(C_I - C_E) - K_r \cdot C_E \quad .. \quad .. \quad (80)$$

(ii) On-dialysis (with kidney failure)

$$\frac{d(C_I \cdot I)}{dt} = G' - k_{I,E}(C_I - C_E) \quad .. \quad .. \quad .. \quad (81)$$

$$\frac{d(C_E \cdot E)}{dt} = k_{I,E}(C_I - C_E) - K_r' \cdot C_E - Q_B \cdot C_E \cdot \left[\exp\left(-\frac{K.A}{Q_B}\right) - 1\right] \quad (82)$$

(iii) Parameter values

Parameter	Urea	Creatinine	Units
G	0.021	0.00042	$g.min^{-1}$
G'	Depend on Metabolic Changes		$g.min^{-1}$
$k_{I,E}$	0.7	0.4	$\ell.min^{-1}$
k_r	0.14	0.014	$\ell.min^{-1}$
k_r'	Depend on Kidney Failure		$\ell.min^{-1}$

A.III.8 Submodel of Renal Failure

A.III.8.1 Definition of failure parameters

FACT1 ≡ Remaining fractional ability to excrete water and sodium

FACT2 ≡ Remaining fractional ability to secrete renin

FACT3 ≡ Remaining fractional ability to excrete potassium

FACT4 ≡ Remaining fractional ability to excrete urea and creatinine

A.III.8.2 Modified equations

(i) GFR

$$\left.\begin{aligned} GFR &= 0.0 && \text{if } AP \leq 20.0 \\ GFR &= FACT1\ (1.92\,AP - 38.4) && \text{if } 20.0 < AP \leq 75.0 \\ GFR &= FACT1\ (-0.00808\,AP^2 + 2.195\,AP - 13.6) && \text{if } 75.0 < AP \leq 120.0 \\ GFR &= FACT1\ (0.035\,AP + 129.2) && \text{if } 120.0 < AP \end{aligned}\right\} \quad (20a)$$

(ii) EBLH

$$EBLH = -\frac{0.01\ EFLH}{FACT1} + 0.65 \qquad \text{if } FACT1 > 0.0 \qquad .. \ (29a)$$

(iii) UK

$$UK = FACT3\ (UKH + UKAL) \quad .. \quad .. \quad .. \quad (39a)$$

(iv) RS

$$RS = FACT2\ \left(0.163 - \frac{0.0093 \cdot SFDT}{FACT1}\right) \text{ if } FACT1 > 0.0 \quad .. \quad (42a)$$

(v) Urea and creatinine

The general equations for concentration C_E of species in extracellular compartments are modified as follows:

(a) Off-dialysis

$$\frac{d\ (C_E \cdot E)}{dt} = k_{I,E}\ (C_I - C_E) - k_r \cdot C_E \cdot FACT4 \quad .. \quad .. \quad (20a)$$

(b) On-dialysis

$$\frac{d\ (C_E \cdot E)}{dt} = k_{I,E}\ (C_I - C_E) - k_r \cdot C_E \cdot FACT4 - Q_B \cdot C_E \left[\exp\ \left(-\frac{K.A}{Q_B}\right) - 1\right] \quad .. \quad .. \quad (82a)$$

APPENDIX IV

MATHEMATICAL MODEL OF THE HUMAN RESPIRATORY CONTROL SYSTEM

List of Symbols used

Variables

C	Concentration
D	Diffusion rate
f	Frequency of respiration
F	Fractional concentration
$\dot{M}$	Metabolic production rate
P	Partial pressure
pH	Blood acidity
RQ	Respiratory exchange rate (Respiratory quotient)
Q	Quantity
$\dot{Q}$	Blood flow rate
t	Time
$\dot{T}$	Instantaneous transfer rate to/from alveolar compartment
τ	Pure time delay
T_I	Inspiratory time
T_E	Expiratory time
$\dot{U}$	Utilisation rate
V	Volume
$\dot{V}$	Ventilation
VT	Tidal volume

Subscripts

a	Arterial
A	Alveolar
B	Brain tissue
c	Capillary
CO_2	Carbon dioxide
D	Dead space
D ANAT	Anatomical dead space
D ALV	Alveolar dead space
D PHYS	Physiological dead space
E	Expired, Expiratory
FRC	Functional residual capacity
I	Inspire, Inspiratory
J	Gas (O_2, CO_2)
L	Lung
M	Muscle tissue
N	Normal
O_2	Oxygen
OT	Other tissue
s	Shunted fraction
v	Venous
$\bar{v}$	Mixed venous blood

Units Employed

Pressure	mm Hg	Concentration of gas	$\ell.\ell$ blood^{-1}
Volume	ℓ	Metabolic rate	ℓ STPD.min^{-1}
Flow	ℓ.min^{-1}	Respiratory frequency	breaths.min^{-1}

The Complete Mathematical Model

(The structure of the model is shown in Chapter 8, Figure 8.2. For a commentary on the equations and the numerical parameter values, etc. consult Sarhan et al., 1979.)

A.IV.1 Pattern of Breathing

A.IV.1.1 Sinusoidal pattern of breathing

$$V_A = V_{FRC} + \frac{VT}{2}(1 - \cos 2\pi ft) \quad \ldots\ldots (1)$$

$$f = \frac{\dot{V}}{VT} \quad \ldots\ldots (2)$$

$$V_{FRC} = 2.9 - 0.312\ VT \quad \ldots\ldots (3)$$

$$\dot{V}_A = \pi f.VT.\sin 2\pi ft \quad \ldots\ldots (4)$$

$$VT = 0.288\ \dot{V}^{\frac{1}{2}} \quad (\text{if } \dot{V} > 10.47\ \ell.\text{min}^{-1}) \quad \ldots\ldots (5)$$

$$VT = 0.089\ \dot{V} \quad (\text{if } \dot{V} \leq 10.47\ \ell.\text{min}^{-1}) \quad \ldots\ldots (6)$$

$$\dot{V}_{A(\text{actual})} = \pi f.VT.\sin 2\pi ft + \left(D_{O_2} - D_{CO_2}\right).\frac{863}{713} \quad \ldots\ldots (7)$$

$$D_{CO_2} = \dot{Q}.\left(C_{v,CO_2} - C_{a,CO_2}\right) \quad \ldots\ldots (8)$$

$$D_{O_2} = \dot{Q}.\left(C_{a,O_2} - C_{v,O_2}\right) \quad \ldots\ldots (9)$$

(The factor 863/713 in (7) is introduced to account for the difference between S.T.P. and B.T.P.S., Body Temperature and Pressure Standard units.)

A.IV.1.2 Triangular waveform pattern of breathing

Inspiration

$$V_A = V_{FRC} + \frac{VT}{T_I}.\ t \quad \ldots\ldots (10)$$

$$\dot{V}_A = VT/T_I \quad \ldots\ldots (11)$$

Expiration

$$V_A = V_{FRC} + VT - \frac{VT}{T_E} \cdot t \quad \ldots\ldots (12)$$

$$\dot{V}_A = - VT/T_E \quad \ldots\ldots (13)$$

$$\dot{V}_{A(actual)} = - VT/T_E + (D_{O_2} - D_{CO_2}) \cdot \frac{863}{713} \quad \ldots\ldots (14)$$

A.IV.2 Respiratory Control Submodel

A.IV.2.1 Control submodel for sinusoidal pattern of breathing

$$\dot{V} = 4.0 \ \ell.min^{-1} \ (\text{if } P_{a,CO_2} < 20.0 \text{ mm HG}) \quad \ldots\ldots (15)$$

$$\dot{V} = 0.1 \ (SS - 2.0)(P_{a,CO_2} - 20.0) + 4.0 \ \ell.min^{-1} \ (\text{if } 20.0 \leq P_{a,CO_2} \leq 40.0) \quad \ldots\ldots (16)$$

$$\dot{V} = SS(P_{a,CO_2} - 38.0) \ \ell.min^{-1} \ (\text{if } P_{a,CO_2} \geq 40.0) \quad \ldots\ldots (17)$$

$$\text{where } SS = 2.2(1.0 + 16.0/(P_{a,O_2} - 30.0)) \quad \ldots\ldots (18)$$

A.IV.2.2 Control submodel for triangular pattern of breathing

$$\text{Drive} = P_{a,CO_2} - 35.2 \quad \ldots\ldots (19)$$

$$VT/T_I = 0.11 \ (\text{Drive}) \quad \ldots\ldots (20)$$

$$T_I = 1.29 - 0.007 \ VT \ (\text{if } VT < 2.048\ell) \quad \ldots\ldots (21)$$

$$T_I = \frac{0.65}{(VT - 0.88)} + 0.59 \ (\text{if } VT \geq 2.048 \ \ell) \quad \ldots\ldots (22)$$

$$T_E = 0.64 \ T_I + 11.1/(\text{Drive} + 2.73) \quad \ldots\ldots (23)$$

$$VT = \frac{VT}{T_I} \cdot T_I \quad \ldots\ldots (24)$$

$$T_{total} = T_I + T_E \quad \ldots\ldots (25)$$

$$f = \frac{60}{T_{total}} \quad (\text{breaths.min}^{-1}) \qquad \ldots\ldots (26)$$

$$\dot{V} = f.VT \qquad \ldots\ldots (27)$$

A.IV.3 Submodel of the Controlled System

A.IV.3.1 CO_2 store equations

Brain tissue

$$\frac{dC_{B,CO_2}}{dt} = \left(\dot{M}_{B,CO_2} + \dot{Q}_B\left(C_{a,CO_2}(\tau_{L-B}) - C_{Bv,CO_2}\right)\right)/V_B \qquad \ldots\ldots (28)$$

Muscle tissue

$$\frac{dC_{M,CO_2}}{dt} = \left(\dot{M}_{M,CO_2} + \dot{Q}_M\left(C_{a,CO_2}(\tau_{L-M}) - C_{Mv,CO_2}\right)\right)/V_M \qquad \ldots\ldots (29)$$

Other tissues

$$\frac{dC_{OT,CO_2}}{dt} = \left(\dot{M}_{OT,CO_2} + \dot{Q}_{OT}\left(C_{a,CO_2}(\tau_{L-M}) - C_{OTv,CO_2}\right)\right)/V_{OT} \qquad \ldots (30)$$

Mixed venous blood

$$\bar{C}_{v,CO_2} = \left(\dot{Q}_B.C_{Bv,CO_2}(\tau_{B-L}) + \dot{Q}_M.C_{OTv,CO_2}(\tau_{M-L}) + \dot{Q}_M.C_{Mv,CO_2}(\tau_{M-L})\right)/\dot{Q}$$

$$\ldots\ldots (31)$$

A.IV.3.2 O_2 store equations

Brain tissue

$$\frac{dC_{B,O_2}}{dt} = \left(-\dot{U}_{B,O_2} + \dot{Q}_B\left(C_{a,O_2}(\tau_{L-B}) - C_{Bv,O_2}\right)\right)/V_B \qquad \ldots\ldots (32)$$

Muscle tissue

$$\frac{dC_{M,O_2}}{dt} = \left(-\dot{U}_{M,O_2} + \dot{Q}_M\left(C_{a,O_2}(\tau_{L-M}) - C_{Mv,O_2}\right)\right)/V_M \qquad \ldots\ldots (33)$$

Other tissues

$$\frac{dC_{OT,O_2}}{dt} = \left(-\dot{U}_{OT,O_2} + \dot{Q}_{OT}\left(C_{a,O_2}(\tau_{L-M}) - C_{OTv,O_2}\right)\right)/V_{OT} \quad \ldots\ldots (34)$$

Mixed venous blood

$$C_{\bar{v},O_2} = \left(\dot{Q}_B.C_{Bv,O_2}(\tau_{B-L}) + \dot{Q}_M.C_{Mv,O_2}(\tau_{M-L}) + \dot{Q}_{OT}.C_{OTv,O_2}(\tau_{M-L})\right)/\dot{Q} \quad \ldots\ldots (35)$$

A.IV.3.3 The alveolar gas exchange and dead space compartments

$$V_{D\ ANAT} + V_{D\ ALV} = V_D \text{ (i.e. } V_{D\ PHYS} \text{ the physiological dead space)} \quad \ldots\ldots (36)$$

$$V_A + V_{D\ ALV} = VL \quad \ldots\ldots (37)$$

Dead space equations during inspiration

$$\frac{dF_{D,CO_2}}{dt} = \dot{V}_{A(actual)}\left(F_{I,CO_2} - F_{D,CO_2}\right)/V_D \quad \ldots\ldots (38)$$

$$\frac{dF_{D,O_2}}{dt} = \dot{V}_{A(actual)}\left(F_{I,O_2} - F_{D,O_2}\right)/V_D \quad \ldots\ldots (39)$$

Alveolar equations during inspiration

$$\frac{dF_{A,CO_2}}{dt} = \left[\left(F_{D,CO_2} - F_{A,CO_2}\right)\left(\dot{V}_{A(actual)} - 0.2\frac{dVT}{dt}\right) + D_{CO_2}\right]/\left(V_A - V_{D\ ALV}\right) \quad \ldots\ldots (40)$$

$$\frac{dF_{A,O_2}}{dt} = \left[\left(F_{D,O_2} - F_{A,O_2}\right)\left(\dot{V}_{A(actual)} - 0.2\frac{dVT}{dt}\right) - D_{O_2}\right]/\left(V_A - V_{D\ ALV}\right) \quad \ldots\ldots (41)$$

Dead space equations during expiration

$$\frac{dF_{D,CO_2}}{dt} = \left[\left(\dot{V}_{A(actual)} - 0.2\frac{\dot{dVT}}{dt}\right)\left(F_{D,CO_2} - F_{A,CO_2}\right)\right]/V_D \quad \ldots\ldots (42)$$

$$\frac{dF_{D,O_2}}{dt} = \left[\left(\dot{V}_{A(actual)} - 0.2\frac{\dot{dVT}}{dt}\right)\left(F_{D,O_2} - F_{A,O_2}\right)\right]/V_D \quad \ldots\ldots (43)$$

Alveolar equation during expiration

$$\frac{dF_{A,CO_2}}{dt} = D_{CO_2}/(V_A - V_{D\ ALV}) \quad \ldots\ldots (44)$$

$$\frac{dF_{A,O_2}}{dt} = -D_{O_2}/(V_A - V_{D\ ALV}) \quad \ldots\ldots (45)$$

A.IV.3.4 Alveolar arterial equilibrium

$$P_{A,CO_2} = 713.F_{A,CO_2} \quad \ldots\ldots (46)$$

$$P_{A,O_2} = 713.F_{A,O_2} \quad \ldots\ldots (47)$$

$$P_{a,O_2} = P_{A,O_2} \quad \ldots\ldots (48)$$

$$P_{a,CO_2} = P_{A,CO_2} \quad \ldots\ldots (49)$$

A.IV.3.5 Arterial gas dissociation curves

$$C_{a,CO_2} = (BH\ CO_3)_b + \overbrace{0.375\left((Hb) - C_a(HbO_2)\right)}^{\text{"Haldane effect"}} + 0.0006732\ P_{a,CO_2} - 0.62\ \log_{10}\left((C_{a,CO_2} - 0.0006732\ P_{a,CO_2})/0.01\ P_{a,CO_2}\right) \quad \ldots\ldots (50)$$

$$C_{a,O_2} = \frac{0.023}{760} P_{a,O_2} + C_a(HbO_2) \quad \ldots\ldots (51)$$

where $C_a(HbO_2) = HB\ \left(1 - \exp\left(-X.P_{a,O_2}\right)\right)^{2.2}$

A.IV.3.6 Venous blood dissociation curves

The same as for arterial gas dissociation.

A.IV.3.7 Venous tissue equilibrium

In this model, tissue and venous blood are considered to have different dissociation curves, i.e. $C_{T,CO_2} \neq C_{vT,CO_2}$. It is assumed, however, that $P_{T,CO_2} = P_{vT,CO_2}$. In the tissue dissociation curve, the Haldane effect is not included.

A.IV.3.8 The pH equation

$$pH = 7.4 - 0.64 \log_{10}(P_{A,CO_2}/40.0) \quad \ldots\ldots (52)$$

A.IV.3.9 Blood flow

$$\dot{Q} = \dot{Q}_B + \dot{Q}_M + \dot{Q}_{OT} \quad \ldots\ldots (53)$$

$$\dot{Q}_{OT} = 0.16\ (\dot{Q} - \dot{Q}_B) \quad \ldots\ldots (54)$$

A.IV.3.10 Circulatory delay times

τ_{L-B} = delay time from lung to brain tissue

$$= 1.062\dot{Q} + 0.015/\dot{Q}_B \quad \ldots\ldots (55)$$

τ_{L-M} = delay time from lung to muscles (assumed equal to delay time from lung to other tissues)

$$= 1.062/\dot{Q} + 0.735/\dot{Q}_M \quad \ldots\ldots (56)$$

$$\tau_{B-L} = 0.06/\dot{Q}_B + 0.188/\dot{Q} \quad \ldots\ldots (57)$$

$$\tau_{M-L} = 2.94/\dot{Q}_M + 0.188/\dot{Q} \quad \ldots\ldots (58)$$

A.IV.3.11 The shunt effect

The model incorporates a blood shunting effect in order to simulate certain diseases. A small quantity of blood, using between 1% and 5% of the total cardiac output, fails to pass through the pulmonary capillaries but instead is shunted through non-aerated vessels, either in the lungs themselves or in the heart. This blood mixes with the aerated blood in the left heart and slightly reduces the P_{O_2} of the blood before it enters the arterial tree. This is known as venous admixture of blood and its effect is illustrated in Figure 8.2. In some diseases of pulmonary circulation, such as emphysema, the shunted blood amounts to more than 50% of the total cardiac output.

Thus, blood C_{a,O_2} is calculated as follows:

$$C_{a,O_2} = \left((\dot{Q} - \dot{Q}s).C_{c,O_2} + \dot{Q}_s.C_{v,O_2}\right)/\dot{Q} \quad \ldots\ldots (59)$$

where

$\dot{Q}$ = total blood flow ($\ell.min^{-1}$)

$\dot{Q}_s$ = fraction of $\dot{Q}_T$ shunted

and C_{c,O_2} = concentration of capillary oxygen.

A.IV.3.12 Cardiac flow control

$$\dot{Q} = (Q_N + \Delta Q_{O_2} + \Delta Q_{CO_2} - Q_B)/\tau_1 \quad \ldots\ldots (60)$$

$$\Delta Q_{O_2} = 9.6551 - 0.2885\, P_{a,O_2} + 2.9241 \times 10^{-3}(P_{a,O_2})^2$$

$$- 1.0033 \times 10^{-5}(P_{a,O_2})^3 \quad [\text{if } P_{a,O_2} < 104 \text{ mm Hg}] \quad \ldots. (61)$$

$$\Delta Q_{O_2} = 0 \quad [\text{if } P_{a,O_2} \geq 104 \text{ mm Hg}] \quad \ldots\ldots (62)$$

$$\Delta Q_{CO_2} = 0.3(P_{a,CO_2} - 40.0) \quad [\text{if } 40 \leq P_{a,O_2} \leq 60] \quad \ldots\ldots (63)$$

$$\Delta Q_{CO_2} = 0 \quad [\text{if } P_{a,CO_2} < 40, \text{ or if } P_{a,CO_2} > 60 \text{ mm Hg}] \quad \ldots\ldots (64)$$

A.IV.3.13 Cerebral blood flow control

$$Q_B = (Q_{BN} + \Delta Q_{B,O_2} + \Delta Q_{B,CO_2} - Q_B)\tau_2 \quad \ldots\ldots (65)$$

$$\Delta Q_{B,O_2} = 0 \quad [\text{if } P_{a,O_2} \geq 104 \text{ mm Hg}] \quad \ldots\ldots. (66)$$

$$\Delta Q_{B,O_2} = 2.785 - 0.1323\, P_{a,O_2} + 2.6036 \times 10^{-3}(P_{a,O_2})^2 - 2.324 \times 10^{-5}(P_{a,O_2})^3$$

$$+ 7.6559 \times 10^{-8}(P_{a,O_2})^4 \quad [\text{if } P_{a,O_2} < 104 \text{ mm Hg}] \quad \ldots\ldots (67)$$

$$\Delta Q_{B,CO_2} = 2.323 \times 10^{-2} - 3.1073 \times 10^{-2}\, P_{a,CO_2} + 8.0163 \times 10^{-4}(P_{a,CO_2})^2$$

$$[\text{if } P_{a,CO_2} < 38] \quad \ldots\ldots (68)$$

$$\Delta Q_{B,CO_2} = 0 \quad [\text{if } 38 \leq P_{a,CO_2} \leq 44] \quad \ldots\ldots (69)$$

$$\Delta Q_{B,CO_2} = -15.58 + 0.7607\ P_{a,CO_2} - 1.2947 \times 10^{-2}\left(P_{a,CO_2}\right)^2$$

$$+ 9.3918 \times 10^{-5}\left(P_{a,CO_2}\right)^3 - 2.1748 \times 10^{-7}\left(P_{a,CO_2}\right)^4$$

$$\left[\text{if } P_{a,CO_2} > 44\right] \qquad \ldots\ldots (70)$$

REFERENCES

Adams, J.M., Attinger, E.O., and Attinger, F.M., "Cluster analysis of respiratory time series", Biol. Cybern., 28, 183-190, 1978.

Al-Dahan, M., Rajkumar, N., Pullen, H.E., Finkelstein, L., Hill, D.W., and Carson, E.R., "Mathematical Modelling of the Human Cardiovascular System", Research Memorandum DSS/MA-D-NR-HEP-LF-DWH-ERC/188, Dept. Systems Science, The City University, London, 1979.

Al-Dahan, M., et al, "Sensitivity Analysis and Monte Carlo Simulations of a Model of the Human Cardiovascular System", Research Memorandum (to be published), Dept. Systems Science, The City University, London, 1979.

Argentesi, F.,"A methodology for the evaluation of model adequacy", Progress in Cybernetics and Systems Research, III, John Wiley, New York, 184-190, 1978.

Ashby, W.R., "Introduction to Cybernetics", Chapman and Hall, 1956.

Ashby, W.R., "General systems theory as a new discipline", General Systems, 3, 1-6, 1958.

Atkins, G.L., "Multicompartmental Models in Biological Systems", Methuen, 1969.

Baldes, E.J., and Smirk, F.H.,J.Physiol.(London),82,62,1934.

Bali, H.N., and Carson, E.R., "Towards a Breathing Model of the Respiratory System", Research Memorandum DSS/HNB-ERC/65, The City University, London, 1974.

Bali, H.N., Carson, E.R., Hill, W.J., Finkelstein, L., "Validation of models through recognition of patterns in measurement", Proc. VII IMEKO Congress, London (Acta Imeko, 1976, 225-234), 1976.

Bali, H.N., "A Mathematical Model of the Human Respiratory System which Exhibits Breathing and Includes Blood Flow and Ventilation Controls", Ph.D. Thesis, The City University, London, 1976.

Baram, Y., "Separation and validation of linear models: model set compatibility", Proc. Lawrence Symps. on Systems and Decision Sciences, Berkeley, CA, USA, Western Periodicals Co., 256-60, 1977.

Batchelor, B.G., "The classification of pole maps in the complex plane", J. Cybern., 8, 237-51, 1978.

Beer, S., "The Brain of the Firm", Penguin Books, 1972.

Bellman, R., Åström, K.J., "On structural identifiability", Mathematical Biosciences, 7, 329-339, 1970.

Bekey, G.A., "Models and reality: some reflections on the art and science of simulation", Simulation, 29, No. 5, 161-4, 1977.

Beneken, J.E.W., "A mathematical approach to cardiovascular function", Institute of Medical Physics, TNO, Utrecht, Netherlands, Rep. No. 2 - 4 - 5/6, 1965.

Beneken, J.E.W., and De Wit, B., "A physical approach to haemodynamic aspects of the human cardiovascular system", in "Physical Bases of Circulatory Transport", Reeve, E.B., and Guyton, A.C., (ed.), W.B. Saunders, Philadelphia, 1967.

Beneken, J.E.W., and Rideout, V.C., "The use of multiple models in cardiovascular system studies: transport and perturbation", I.E.E.E. Trans. Bio. Med. Eng., 15, 281 - 289, 1968.

Beneken, J.E.W., "Some computer models in cardiovascular research", in "Cardiovascular Fluid Mechanics", Bergel, D.H. (ed.), Academic Press, 173 - 223, 1972.

Bergel, D.H. (ed.), "Cardiovascular Fluid Dynamics", Academic Press, 1972.

Berlinski, D., "On Systems Analysis; An Essay concerning the Limitations of some Mathematical Models in the Social, Political and Biological Sciences", MIT Press, 1976.

Berman, M., "The formulation and testing of models", Ann. N.Y. Acad. Sci., 108, 182-194, 1963.

von Bertalanffy, L., "Theoretische Biologie", Berlin, Borntraeger, 1932.

von Bertalanffy, L., "An outline of general system theory", Brit. J. Phil. Sci, 1, 139-164, 1980.

von Bertalanffy, L., "General System Theory: Formulations, Development, Applications", Penguin Books, 1968.

von Bertalanffy, L., "Basic Concepts in Quantitative Biology of Metabolism", (orig. 1964), 163-196, in von Bertalanffy, 1968.

Bigelow, J.H., De Haven, J.C., and Shipley, M.L., "Systems analysis of the renal function", J. Theor. Biol., 41, 287-322, 1973.

Black, M., "Models and Archetypes", in "Models and Metaphors", Cornell Univ. Press, Ithaca, N.Y., 1962.

Blaine, E.H., Davis, J.O., and Harris, P.D., "A steady state control analysis of the renin - angiotensin - aldosterone system", Circ. Res., 30, 713-730, 1972.

Blalock, H.M., "Measurement in the Social Sciences, (Theories and Strategies", Macmillan, 1974.

Blaug, M., "Economic Theory in Retrospect", 1962.

Bowers, D.A., Mitchell, C.R., and Webb, K., "A Dynamic Model of Bicommunal Political Systems. IV Initial Simulation Runs. V Summary and Conclusions", Research Memorandum DSS/DAB-CRM-KW/163, Dept. Systems Science, The City University, London, 1978.

Bowers, D.A., Mitchell, C.R., and Webb, K., "A Dynamic Model of Bicommunal Political Systems. I Model Description", Research Memorandum DSS/DAB-CRM-KW/160, Dept. Systems Science, The City University, London, March, 1979a.

Bowers, D.A., Mitchell, C.R., and Webb, K., "Systems Dynamics in Situations of Conflict between Two Communities", Research Memorandum DSS/DAB-CRM-KW/193, Dept. Systems Science, The City University, London, August, 1979b.

Box, G., and Jenkins, G.M., "Time-Series Analysis, Forecasting and Control", Holden-Day, 1970.

Braithwaite, R.B., "Scientific Explanation: A study of the function of theory, probability and law in science", CUP, 1953.

Bridgman, P.W., "Dimensional Analysis", Yale University Press, 1922.

Brown, R.F., Godfrey, K.R., "Problems of determinacy in compartmental modelling with application to bilirubin kinetics", Math. Biosci., 40, 205-224, 1978.

Bunge, M., "Philosophy of Physics", and "Method, Model and Matter", both D. Reidel, Dordrecht, Holland, 1973.

Burgen, A.V.S., and Mitchell, J.F. (ed.), "Gaddum's Pharmacology", 4th edn., Oxford University Press, 1978.

Cage, P.E., et al, "A model of the human renal medulla", Comp. and Biomed. Res., 10, 561-584, 1977.

Cameron, W.H., "A model framework for computer simulation of overall renal function", J. Theor. Biol., 66, 551-572, 1977.

Campbell, N.R., "Physics: The Elements", Cambrige University Press, 1920, reprinted as "Foundations of Science", Dover, 1957.

Cambell, N.R., "An Account of the Principles of Measurement and Calculation", Longmans Green, 1928.

Cannon, W.B., "Organisation for physiological homeostasis", Physiol. Res., 9, 397, 1929.

Carnap, R., "Testability and Meaning", Phil. Sci., 3, 420-468; 4, 1-40, 1936, 1937.

Carnap, R., "Logical Foundations of Probability", University of Chicago Press, (orig. 1980), 2nd ed., 1962.

Carson, E.R., Finkelstein, L, "Problems of identification in metabolic systems", in Proc. 3rd IFAC Symp. on Identification and System Parameter Estimation, North Holland, 1973.

Carson, E.R., Finkelstein, L., "General problems of validation of control system models in physiology", IFAC Symposium on Control Mechanisms in Bio- and Eco-Systems, Leipzig, 1977.

Carson, E.R., Cobelli, C., and Finkelstein, L., "The identification of metabolic systems. A review.", Proc. 5th IFAC Symp. on Identification and System Parameter Estimation, Darmstadt, W. Germany, Pergamon Press, Oxford (in press), 1979.

Causey, R.L., "Derived measurement, dimensions and dimensional analysis", Phil. Sci., 36, 252, 1969.

Checkland, P.B., "Towards a systems-based methodology for real-world problem solving", Journal of Systems Engineering, 3, 87-116, 1972.

Chrostowski, J.D., Evensen, D.A., Hasselman, T.K., "Model verification of mixed dynamic systems", Trans. ASME, J. Mech. Des., 100, No. 2, 266-73, 1978.

Churchman, C.W., Ackoff, R.L., Arnoff, E.L., "Introduction to Operations Research", John Wiley, N.Y., 1957.

Cobbold, R.S.C., "Transducers for Biomedical Measurements: Principles and Applications", New York: Wiley, 1974.

Cobelli, C., Salvan, A., "Reliability of parameter estimation in biological compartmental models. A computer analysis of several factors influencing the estimates", Ident. and Syst. Para. Est., Vol. I, IV IFAC Symposium, 401-411, 1976.

Cobelli, C., and DiStefano III, J.J., "Parameter and structural identifiability concepts and ambiguities: a critical review and analysis", Am. J. Physiol., (Regulatory, Integrative, and Comparative Physiology), May 1980.

Cole, H.S.D., Freeman, C., Jahoda, M., and Pavitt, K.L.R., "Thinking about the Future. A Critique of 'The Limits of Growth", Chatto and Windus, 1973.

Cooney, D.O., "Biomedical Engineering and Instrumentation. Vol. 2: Biomedical Engineering Principles", New York: Dekker, 1976.

Cox, D.R., and Hinkley, D.V., "Theoretical Statistics", London: Chapman and Hall, 1974.

Cunningham, D.J.C., and Lloyd, E.B. (eds.), "The Regulation of Human Respiration", Oxford: Blackwell, 1963.

Cunningham, D.J.C., and Gardner, W.N., "A quantitative description of the pattern of breathing during steady-state CO_2 inhalation in man with special emphasis on expiration", J. Physiol., 272, 613-632, 1977.

Dampney, R.A.L., Taylor, M.G., and McLachlan, E.M., "Reflex effects of stimulation of carotid sinus and aortic baroreceptors on hindlimb vascular resistance in dogs", Circulation Research, 29, 119-127, 1971.

Davis, J.O.,and Howell, D.S., Endocrinology,52,245,1953.

Dean, R.F.A., and McCance, R.A.,"The renal responses of infants and adults to the administration of hypertonic solutions of sodium chloride and urea", J.Physiol.,109,81-97,1949.

De Haven, J.C., and Shapiro, N.Z., "Simulation of the renal effects of antidiuretic hormone in man", J. Theor. Biol., 28, 261-286, 1970.

Deheneffe, J., et al, "Response of aldosterone and blood pressure to angiotensin II infusion in anephric man", Circ. Res., 39, no. 2., 183-190, 1976.

Desai, M., "Applied Econometrics", Phillip Allen, 1976.

Dick, D.E., and Rideout, V.C., "Analog simulation of left heart and arterial dynamics", Proc. 18th ACEMB, Philadelphia, 78, 1965.

DiStefano III, J.J., and Cobelli, C., "On parameter and structural identifiability: nonunique observability/ reconstructability for identifiable systems, other ambiguities and new definitions", I.E.E.E. Trans. Autom. Contr. 1980.

Domotor, Z., "Probabilistic Relational Structures and their Applications", Technical Rep. No. 144, Stanford Univ., Inst. Math. Studies in the Social Sciences, 1969.

Einstein, A., "On a Heuristic Viewpoint Concerning the Production and Transformation of Light", (trans., orig. 1905) in Boorse and Motz, eds. "The World of the Atom", New York: Basic Books, 545-557, 1966.

Ellis, B., "Basic Concepts of Measurement", Cambridge University Press, 1966.

Emery, F.E. (ed.), "Systems Thinking", Penguin Books, 1969.

Eykhoff, P., "System Identification :Parameter and State Estimation", John Wiley, New York, 1974.

Farbey, B.A., Mitchell, C.R., and Webb, K., "A Meaning and Measurement of Political Ideology: The Case of Scottish Nationalism", Research Memorandum DSS/BAF-CRM-KW/190, Dept. Systems Science, The City University, London, July, 1979.

Farbey, B.A., Mitchell, C.R., and Webb, K., "A Computer-Based System for the Interactive Analysis of Domain - Specific Documents", Research Memorandum DSS/BAF-CRM-KW/210, Dept. Systems. Science, The City University, London, September 1980.

Fenn, W.O., and Craig, A.B., J. Appl. Physiol., 18, 1023, 1963

Feyerabend, P., "An attempt at a realistic interpretation of experience", Proc. of the Aristotelian Society, New Series, 58, 143-170, 1958.

Feyerabend, P., "Against Method. Outline of an anarchistic theory of knowledge", (orig. 1970), Verso, 1975.

Finkelstein, L., "Representation by symbols as an extension of the concept of measurement", Kybernetes, 4, 215-23, 1975.

Fishman, G.S., Kiviat, P.J., "The statistics of discrete-event simulation", Simulation, 10, 185-195, 1968.

Forrester, J.W., "Principles of Systems", Cambridge, Mass., USA: Wright-Allen Press, 1968.

Forrester, J.W., "World Dynamics", Cambridge, Mass., USA: Wright-Allen Press, 1971.

Frank, P.M., "Introduction to System Sensitivity Theory", Academic Press, 1978.

Friedman, M., "The Methodology of Positive Economics", in "Essays in Positive Economics", Univ. Chicago Press, 1953.

Frost, T.H., and Kerr, D.N.S., "Kinetics of hemodialysis: a theoretical study of the removal of solutes in chronic renal failure compared to normal health", Kidney Intern., 12, 41-50, 1977.

Gabe, I.T., "Pressure measurement in experimental physiology", in "Cardiovascular Fluid Mechanics", D.H. Bergel (ed.), 11 - 49, Academic Press, 1972.

Glick, G., and Covel, J.W., "Relative importance of the carotid and aortic baroreceptors in the reflex control of heart rate", Am. J. Physiol., 214, 955-961, 1968.

Gödel, K., "On Formally Undecidable Propositions", New York: Basic Books, 1962 (orig. 1931).

Gold, H.J., "Mathematical Modelling of Biological Systems. An Introductory Guide book", John Wiley and Sons, 1977.

Golomb, S.W., "Mathematical models: uses and limitations", Astronaut. Aeronaut., 57-59, 1968.

Gray, J.S., "The Multiple Factor Theory of Respiratory Regulation", Army Air Force School of Aviation Medicine Project Report No. 386 (1,2,3) 1945.

Green, H.A.J., "Aggregation in Economic Analysis", Princeton Univ. Press, 1964.

Grodins, F.S., Gray, J.S., Schroeder, K.R., Norins, A.L., and Jones, R.W., "Respiratory responses to CO_2 inhalation. A theoretical study of a non-linear biological regulator", J. Appl. Physiol., 7, 283-308, 1954.

Grodins, F.S., "Integrative cardiovascular physiology: A mathematical synthesis of cardiac and blood vessel hemodynamics", Quart. Rev. Biol., 34, 93-123, 1959.

Grodins, F.S., Buell, J., and Bart, A.J., "Mathematical analysis and digital simulation of the respiratory control system", J. Appl. Physiol., 22, 260-276, 1967.

Gruhl, J., "Model validation", Proc. Int. Conf. on Cybernetics and Society, October, 1979, sponsored by I.E.E.E. Systems, Man and Cybernetics Society, 536-541, 1979.

Guyton, A.C., and Coleman, T.G., "Long-term regulation of the circulation: interrelationships with body fluid volumes", in "Physical Bases of Circulatory Transport", Reeve, E.B., and Guyton, A.C. (eds.), 170-201, W.B. Saunders, Philadelphia, 1967.

Guyton, A.C., "Determination of cardiac output by equating venous return curves with cardiac response curves", Physiol. Rev., 35, 123-129, 1955.

Guyton, A.C., "Textbook of Medical Physiology", W.B. Saunders, Philadelphia, 1971.

Guyton, A.C., Coleman, T.G., and Granger, H.J., "Circulation: overall regulation", Ann. Rev. Physiol., 34, 13-46, 1972.

Haas, E., and Goldblatt, H., "Kinetic constants of the human renin and human angiotensin reaction", Circ. Res., 20, 45-55, 1967.

Haldane, J.S., and Priestley, J.G., "The regulation of the lung ventilation", J. Physiol., 32, 225-226, 1905.

Hall, A.D., "Methodology for Systems Engineering", Van Nostrand, 1962.

"Handbook of Physiology", Hamilton, W.F., and Dow, P. (eds.), American Physiological Society, 1965.

Hanson, N.R., "Patterns of Discovery", Cambridge University Press, 1958.

Hanson, N.R., "Observation and Explanation: A Guide to the Philosophy of Science", George Allen and Unwin, 1971.

Hardy, J.P., and Stolwijk, J.A.J., "Partitional calorimetric studies of man during exposures to thermal transients", J. Appl. Physiol., 1966.

Harré, R., "The Principles of Scientific Thinking", Macmillan, 1970.

Harré, R., "The Philosophies of Science", Oxford University Press, 1972.

Harvey, W., "Excertatio de Motu Cordis et Sanguinis in Animalibus", 1628, (translated by K.J. Franklin, "Movement of the Heart and Blood in Animals", Blackwell Scientific Publications, Oxford, 1957).

Hawker, R.W., "Notebook of Medical Physiology: Endocrinology", Churchill Livingstone, 1978.

Hawker, R.W., "Notebook of Medical Physiology: Cardiopulmonary", Churchill Livingstone, 1979.

Helmholtz, H., "Counting and Measuring", (orig. 1887), trans. by C.L. Bryan, Van Nostrand, 1930.

Hempel, C.G., and Oppenheim, P., "A definition of 'degree of confirmation'", Phil. Sci., 12, 98-115, 1945.

Hempel, C.G., and Oppenheim, P., "Studies in the logic of explanation", Phil. Sci., 15, 135-175, 1948.

Hempel, C.G., "Deductive - Nomological vs. Statistical Explanation" in Feigl and Maxwell (eds.) "Minnesota Studies in the Philosophy of Science", vol. III, 98-169, 1962.

Hermann, C.F., "Validation problems in games and simulations, with special reference to models of international politics", Behavioural Science, 12, 216-231, 1967.

Hertz, H., "The Principles of Mechanics", (orig. 1887), Dover, 1956.

Hevesey, G., "The Absorption and Translocation of lead by plants", Biochem. J., 17, 439, 1923.

Hickey, F.C., and Brownwell, G.L., "Dynamic analysis of iodine metabolism in normal subjects", J. Clin. Endocrin. Metab., 14, 1423, 1954.

Hintikka, J., "On a combined system of inductive logic", in "Studia Logico - Mathematica et Philosophica in Honorem Rolf Nevanlinna", Helsinki, 1965.

Hölder, O., "Die axiome der quantität und die lehre von mass", Math. Phys. Klasse, 53, 1-64, 1901.

House, P.W., "Diogenes revisited - the search for a valid model", Simulation, 23, 117-125, 1974.

Horgan, J.D., and Lange, R.L., "Analog computer simulation of periodic breathing", IRE Trans. Biomed. Electr., BME-9, 221-228, 1962.

Hume, D., "Enquiry concerning Human Understanding", 1748, OUP, 1902.

Hyndman, P.W., "A Digital Simulation of the Human Cardiovascular System and its use in the Study of Sinus Arrythmia", Ph.D. Thesis, Imperial College, University of London, 1970.

Ignall, E.J., Kolesar, P., and Walker, W.E., "Using simulation to develop and validate analytic models: some case studies", Oper. Res. (USA), 26, No. 2, 237-253, 1978.

Jacquez, J.A., "Compartmental Analysis in Biology and Medicine", Elsevier Publ. Co., 1972.

Johnson, J.A., Zehr, J.E., and Moore, W.W., "Effects of separate and concurrent osmotic and volume stimuli on plasma ADH in sheep", Am. J. Physiol, 218, 1273-1280, 1970.

Kahne, S., "Model credibility for large-scale systems", I.E.E.E Trans. Syst. Man & Cybern., SMC-6, No. 8, 586-590, 1976.

Kalman, R.E., "Mathematical description of linear dynamical systems", SIAM J. Control, 1, 152-192, 1963.

Kalmus, H., "Introduction - Control and Regulation as Interactions within Systems", in H. Kalmus (ed.), "Regulation and Control in Living Systems", New York: John Wiley and Sons, 1966.

Kao, F.F., Mei, S.S., and Kalia, M., "Interaction between neurogenic exercise drive and chemical drive", in "Central Nervous Control Mechanisms in Breathing", C. von Euler, and H. Lagercrantz (eds.), Pergamon, 75-89, 1979.

Kaplan, A., "The Conduct of Inquiry. Methodology for Behavioural Science", Chandler Publ. Co., 1964.

Karplus, W.J., "System Identification and Simulation - A Pattern Recognition Approach", Fall Joint Computer Conference, 1972.

Karplus, W.J., "The spectrum of mathematical modelling and systems simulation", Proc. 4th AICA Congress on Simulation of Systems, Delft, Netherlands, 23-28 August, 1976, Amsterdam, Netherlands: North Holland, 5-13, 1977.

Katona, P.G., "Computer Simulation of the Blood Pressure Control of Heart Period", Sc. D. Thesis, MIT, Cambridge, Mass., 1965.

Katona, P.G., Barnet, O., and Jackson, W.D., "Computer simulation of the blood pressure control of heart period", in Kezdi, P., (ed.), "Baroreceptors and Hypertension", 191-199, Pergamon, 1967.

Keele, K.D., "The Life and Works of William Harvey", Endeavour 2, No. 3, 104-107, 1978.

Kendal, M.G., and Stuart, A., "The Advanced Theory of Statistics. Vol. 2 Inference and Relationship", Charles Griffin (1st edn. 1946), 1973.

Keynes, J.M., "The General Theory of Employment, Interest, and Money", orig. 1936, Macmillan, 1967.

Kheir, N.A., "A validation procedure for missle system simulation models", Proc. 1976 Conf. on Modeling and Simulation, Pittsburgh, Pa., 7, pt. 1, 534-539, 1976.

Kheir, N.A., and Holmes, W.M., "On validating simulation models of missile systems", Simulation, 30, 1978.

Krantz, D.H., Luce, R.D., Suppes, P., and Tversky, A., "Foundations of Measurement. Vol. I: Additive and Polynomial Representations", Academic Press, 1971.

Kuhn, T.S., "The Structure of Scientific Revolutions", Univ. Chicago Press, 1970, (2nd edn., orig. 1962).

Kuhn, T.S., "Reflections on My Critics", in I. Lakatos and A. Musgrave (eds.), "Criticsm and the Growth of Knowledge", Cambridge University Press, 231-278, 1970.

Kuhn, T.S., "A Formalism for Scientific Change", 5th Int. Conf. of Logic, Methodology and the Philosophy of Science, 1975.

Kuhn, T.S., "Second Thoughts on Paradigms", in F. Suppe (ed.), "The Structure of Scientific Theories", University of Illinois Press, 459-482, 1977.

Lakatos, I., and Musgrave, A. eds., "Criticism and the Growth of Knowledge", Cambridge University Press, 1970.

Landwehr, D.M., et al, "Effect of reduction in filtration rate on renal tubular sodium and water reabsorption", Am. J. Physiol., 215, 687-695, 1968.

Laszlo, E., "Introduction to Systems Philosophy. Toward a New Paradigm of Contemporary Thought", Gordon and Breach, 1972.

Leal, D.J., Roberts, P.D., "Verification of mathematical dynamic models", Research Memorandum DSS/DJL-PDR/120 (confidential report), Dept. Systems Science, The City University, London, 1977.

Leaning, M.S., "Foundational Studies in Measurement. An Attempt to Formalise Uncertainty in the Theory of Measurement", M.Sc. Thesis, Dept. Systems Science, The City University, London, 1977.

Leaning, M.S., Carson, E.R., and Finkelstein, L., "The Use of Feature Space Methods in Model Validation", Research Memorandum DSS/MSL-ERC-LF/181, Dept. Systems Science, The City University, London, 1979.

Lee, M.R., "Renin and Hypertension. A Modern Synthesis", London: Lloyd-Luke, 1969.

Lefschetz, S., "Differential Equations: Geometric Theory", New York: John Wiley and Sons, 1963.

Lilienfeld, R., "The Rise of Systems Theory. An Ideological Analysis", John Wiley, 1978.

Lindholm, F.A., et al, "Assessing model adequacy .... in integrated circuit simulation", I.E.E.E. J. Solid-State Circuits, vol. SC-6, 213-222, 1971.

Lippold, O.C.J., and Winton, F.R., "Human Physiology", 7th edn., Churchill, Livingstone, 1979.

Lloyd, B.B., Jukes, M.G.M., and Cunningham, D.J.C., Quart. J. Exp. Physiol., 43, 214, 1958.

Lott, R.S., et al, "The Application of Mathematical Models on an Interactive Computer Graphics Display for Renal Patient Management", in "Technical Aspects of Renal Dialysis", T.H. Frost (ed.), Pitman, 1977.

Luce, R.D., "Semiorders and a theory of utility discrimination", Econometrica, 24, 178-191, 1956.

Luce, R.D., "Dimensionally invariant numerical laws correspond to meaningful qualitative relations", Phil. Sci., 45, 1-16, 1978.

Mach, E., "Über die Definition der Masse", Carls Repertorium (Munich), 4, 355-359, 1868.

Mach, E., "The Analysis of Sensations", (orig. 1886), Dover, 1959.

Makridakis, S., and Hibon, M., "Accuracy of forecasting: An empirical investigation", J.R. Statistical Soc. A., 142, Part 2, 97-145, 1979.

Mancia, G., Ferrari, A., Gregorini, L., Ludbrook, J., and Zanchetti, A., "Baroreceptor control of heart rate in man", in "Neural Mechanisms in Cardiac Arrythmias", Schwartz, P.J., Brown, A.M., Malliani, A., and Zanchetti, A., (eds.), Raven Press, 323-333, 1978.

Mattessich, R., "Instrumental Reasoning and Systems Methodology. An Epistemology of the Applied and Social Sciences", Reidel, Dordrecht, 1978.

McCarthy, P.J., "The use of balanced half-sample replication in cross-validation studies", J. Am. Statist. Assoc., 71, 596-604, 1976.

Meadows, D.H., Meadows, D.L., Randers, J., and Behrens III, W.W., "The Limits to Growth. A Report for the Club of Rome's Project on the Predicament of Mankind", New York, Universe Books, 1972, (Pan Books, London, 1974).

Meadows, D.L., Behrens III, W.W., Meadows, D.H., Naill, R.F., Randers, J., and Zahn, E.K.O., "Dynamics of Growth in a Finite World", Wright-Allen Press, 1974.

Mehra, R.K., "Nonlinear system identification: selected survey and recent trends", IFAC V Symposium on Identification and System Parameter Estimation, Preprints, Vol 1, Pergamon Press, 77-83, 1979.

Mesarovic', M.D., "Foundations for a General Systems Theory", in M.D. Mesarovic' (ed.), "Views on General Systems Theory", John Wiley, 1-24, 1964.

Mihram, G.A., "Simulation: Statistical Foundations and Methodology", Academic Press, N.Y., 1972.

Mihram, G.A., "The Modeling Process", I.E.E.E. Trans. Syst. Man & Cybern., SMC-2, 621-629, 1972.

Mihram, G.A., "Some practical aspects of the verification and validation of simulation models", Oper. Res. Quarterly, 23, 17-29, 1972.

Mihram, G.A., "Four questions regarding the credibility of simulation", and "Four further queries concerning simular credibility", Modeling and Simulation, 7, Pt. II, 1225-1231 and 1231-1234, 1976.

Miller, K.S., Murray, F.J., "The mathematical basis for the error analysis of differential analyzers", J. Math. Phys. (Cambridge, Mass), 1953.

Miller, D.R., "Sensitivity analysis and validation of simulation models", J. Theor. Biol., 48, 345-60, 1974.

Mitchell, C.R., and Webb, K., "A Dynamic Model of Bicommunal Political Systems: III Definitions, Scales, Indices", Research Memorandum DSS/CRM-KW/162, Dept. Systems Science, The City University, London, May, 1978.

Molloy, K.J., and Best, D.P., "The Checkland methodology considered as a theory building methodology", in "Progress in Cybernetics and Systems Research" (in print), John Wiley, 1980.

Monin, A.S., "Weather Forecasting as a Problem in Physics", (orig. 1969), trans. P. Superak, MIT Press, 1972.

Mosier, C.J., "I. Problems and designs of cross-validation", in "Symposium: The Need and Means of Cross-Validation", reported in Education and Psychological Measurement, 11, 5-11, 1951.

Mountcastle, V.B. (ed.), "Medical Physiology", 13th edn., Vol. II, C.V. Mosby Co., 1974.

M'Pherson, P.K., "The Design of Multi-Dimensional Value Criteria for Use in Decision, Design and Technology", Research Memorandum DSS/PK/187, Dept. Systems Science, The City University, London, 1979.

M'Pherson, P.K., "Systems and systems science? An attempt to answer some ontological and epistemological questions", in "Progress in Cybernetics and Systems Research" (in print), John Wiley, 1980.

Naylor, T.H., Balintfy, J.L., Burdick, D.S., and Kong Chu, "Computer Simulation Techniques", John Wiley, N.Y., 1966.

Naylor, T.H., and Finger, J.M., "Verification of computer simulation models", Management Sci., XIV, 92-101, 1967.

Naylor, T.H., Burdick, D.S., and Sasser, W.E., "The Design of Computer Simulation Experiments", in book of same title, ed. T.H. Naylor, Duke Univ. Press, 3-35, 1969.

de Neufville, J.I., "Validating policy indicators", Policy Sciences, 10, 171-188, 1978.

von Neumann, J., and Morgenstern, O., "Theory of Games and Economic Behaviour", Princeton University Press, 1944.

Noble, M.I.M., Trenchard, D., and Guz, A., "Effect of changing heart rate on cardiovascular function in the conscious dog", Circulation Research, 19, July, 1966.

Noton, M., Mitchell, C.R., and Janes, F.R., "The systems analysis of conflict", Futures, 114-132, April, 1974.

Nunnally, J.C., "Introduction to Psychological Measurement", McGraw-Hill, 1970.

Ohley, W.J., Kao, C., and Jaron, D., "Validity of an arterial system model: a quantitative evaluation", I.E.E.E. Trans. Biomed. Eng., BME-27, 203-211, 1980.

Panjabi, M., "Validation of mathematical models", (letter), J. Biomech., 12, No. 3, 238, 1979.

Pfanzagl, J., "Theory of Measurement", John Wiley, 1968.

Popper, K.R., "The Logic of Scientific Discovery", (orig. 1935), Hutchinson, 1972.

Popper, K.R., "Conjectures and Refutations", Basic Books, 1962.

Popper, K.R., "Objective Knowledge. An Evolutionary Approach", (orig. 1972), Clarendon Press, 1979.

Pullen, H.E., "Cardiovascular Modelling and Simulation", M.Sc. Thesis, The City University, London, 1970.

Pullen, H.E., "Studies in the Modelling and Simulation of the Human Cardiovascular System with Application to the Effect of Drugs", Ph.D. Thesis, Dept. Systems Science, The City University, London, 1976.

Rajkumar, N., "Reduction of a Mathematical Model of the Human Cardiovascular System", M.Sc. Thesis, Dept. Systems Science, The City University, London, 1978.

Reeve, E.B., and Kulhanek, L., "Regulation of Body Water Content. A Preliminary Analysis", in "Physical Bases of Circulatory Transport", E.B. Reeve and A.C. Guyton (eds.), Chapter 18, 151-177, 1967.

Reggiani, M.G., Marchetti, F.E., "On assessing model adequacy", I.E.E.E. Trans. Syst. Man & Cybern., SMC-5, No. 3, 322-330, 1975.

Reichenbach, H., "The Rise of Scientific Philosophy", University of California Press, 1951.

Renkin, E.M., "The relation between dialysance, membrane area, permeability and blood flow in the artificial kidney", trans. Am. Soc. Artif. Intern. Organs Transpl., 2, 102-105, 1956.

Reynolds, W.J., Milhorn, H.T., and Holloman, G.H., "Transient ventilatory response to graded hypercapnia in man", J. Appl. Physiol., 33, 47-54, 1972.

Reynolds, W.J., and Milhorn, H.T., "Transient ventilatory response to hypoxia with and without controlled alveolar P_{CO_2}", J. Appl. Physiol., 35, 187-196, 1973.

Rice, D.B., Smith, V.L., "Nature, the experimental laboratory, and the credibility of hypotheses", Behav. Sci., 9, 239-246, 1964.

Roberts, P.D., "Parameter Estimation in Non-Linear Dynamic Mathematical Models - User Manual for Subroutines IDENT and PERAS", Research Memorandum No. DSS/PDR/127, Dept. Systems Science, The City University, London, 1977.

Rowland, J.R., and Holmes, W.M., "Simulation validation with sparse random data", Comput. Electr. Eng. (USA), 5, 37-49, 1978.

Sarhan, N., Carson, E.R., and Finkelstein, L., "Mathematical Modelling and Computer Simulation of the Human Respiratory Control System", Research Memorandum No. DSS/NS-ERC-LF/199, The City University, London, 1979.

Sarhan, N., Carson, E.R., and Finkelstein, L., "The Regulation and Control of Respiration: Mathematical Modelling and Computer Simulation", Research Memorandum No. DSS/NS-ERC-LF/214, 1980.

Saunders, K.B., Bali, H.N., and Carson, E.R., "A breathing model of the respiratory system: the controlled system", J. Theor. Biol., 84, 135-161, 1980.

Saunders, K.B., "Oscillations of arterial CO_2 tension in a respiratory model: some implications for the control of breathing in exercise", J. Theor. Biol., 84, 163-179, 1980.

Schaeffer, B.M., "Model validation using simulation and sequential decision analysis", Bull. Oper. Res. Soc. Am. (USA), 23, suppl. 1, p. B/69, 1975.

Schlesinger, S., et al, "Developing standard procedures of simulation, validation and verification", Proc. 1974 Summer Computer Simulation Conference, Society for Computer Simulation, La Jolla, Calif., 927-933, 1974.

Schweppe, F.C., "Uncertain Dynamic Systems", Prentice-Hall, 1973.

Scott, D., and Suppes, R., "Foundational aspects of theories of measurement", J. Symb. Logic, 23, 113-128, 1958.

Seif, F.J., "Systemanalyse der Aldosteronsekretion", Stuttgart: Georg Thieme Verlag, 1974.

Shapere, D., "Galileo: A Philosophical Study", University of Chicago Press, 1974.

Shapere, D., "Scientific theories and their domains", in F. Suppe (ed.) 518-565, 1977.

Sharpey - Schafer, E.P., "Effect of respiratory acts on the circulation", in "Handbook of Physiology", Hamilton, W.F. and Dow, P. (eds.), Section 2, 3, 1875-1886, American Physiological Society, 1965.

Sheppard, C.W., "The theory of the study of tracers within a multi-compartment system using isotopic tracers", J. Appl. Phys., 19, 70, 1948.

Shipley, R.E., and Study, R.S., "Changes in renal blood flow, extraction of inulin, glomerular filtration rate, tissue pressure and urine flow with acute alterations of renal artery blood pressure", Am. J. Physiol., 167, 676-688, 1951.

Siklossy, L., Roach, J., "Model verification and improvement using DISPROVER", Artif. Intell, (Netherlands), 6, No. 1, 41-52, 1975.

Snee, R.D., "Validation of regression models: method and examples", Technometrics, 19, no. 4, 415-428, 1977.

Sneed, J.D., "The Logical Structure of Mathematical Physics", Reidel, Dordrecht, 1971.

Stegmüller, W., "The Structure and Dynamics of Theories", Springer-Verlag, 1976.

Suppe, F. (ed.), "The Structure of Scientific Theories", University of Illinois Press, 1977.

Suppes, P., "Models of Data", in "Studies in the Methodology and Foundations of Science", New York, Humanities Press, 24-35, 1969.

Sussman, H., and Zahler, R., "Catastrophe theory as applied to the social and biological sciences: A critque", Synthese, 37, 117-216, 1978.

Talbot, S.A., and Gessner, U., "The Cardiovascular System and its Controls", Ch. 16 in "Systems Physiology", John Wiley, 1973.

Tarski, A., "Contributions to the theory of models", Indagationes Mathematicae, 16, 572-88, 1954.

Tarski, A., "The Concept of Truth in Formalized Languages", (orig. 1931), in "Logic, Semantics, Metamathematics. Papers from 1923 to 1938.", translated by J.H. Woodger, Oxford: Clarendon Press, 1956.

Tawari, P.V., "A systematic procedure for validating and evaluating dynamic models", Univ. Washington, Seattle, USA, Thesis, 1974.

Thom, R., "Topological models in biology", Topology, 8, 313-335, 1969.

Thom, R., "Structural Stability and Morphogenesis", Benjamin, 1975.

Thomas, G.,"Mathematical Modelling of the Human Cardiovascular System", Dept. Systems Science, The City University, B.Sc. Thesis,1980.

Thurau, K., "Local regulation of blood flow", Circ. Res., Suppl. I, 28, 1971.

Thurau, K., et al, "Activation of renin in the single juxtaglomerular apparatus by sodium chloride in the tubular fluid at the macula dersa", Circ. Res. 31, (Suppl. II), 182-186, 1972.

Tinbergen, J., "An Econometric Approach to Business Cycle Problems", Herman (Paris), 1937.

Tobias, L., et al, "Effect of high perfusion pressures on the granulations of juxtaglomerular cells in an isolated kidney", Kidney. J. Clin. Invest., 38, 605-610, 1959.

Tomovic', R., 1963, "Sensitivity Analysis of Dynamic Systems", New York: McGraw-Hill.

Toulmin, S., "The Philosophy of Science", Hutchinson, 1953.

Toulmin, S., "Human Understanding. Vol. I", Princeton University Press, 1972.

Uttamsingh, R.J., and Carson, E.R., "A Mathematical Model of the Patient-Artificial Kidney System", Research Memorandum No. DSS/RJU-ERC/122, Dept. Systems Science, The City University, London, 1977.

Uttamsingh, R.J.," A Systems Approach to Health Care in Renal Dialysis", Ph.D. Thesis, Dept. Systems Science, The City University, London, (to be published), 1981.

Van Horn, R., "Validation", in T.H. Naylor (ed.), "The Design of Computer Simulation Experiments", Duke University Press, 1969.

Vander, A.J., and Miller, R., "Control of renin secretion in the anesthetized dog", Am. J. Physiol., 207, 537-546, 1964.

Vemuri, V., "Modeling of Complex Systems. An Introduction", Academic Press, 1978.

Walker, W.E., et al, "Application of a programmable pocket calculator to a single compartment mathematical model of solute kinetics", Computer Programs in Biomedicine, 75, 99-104, 1975.

Walters, G., "The Language of Pathology", HM + M Publishers, 1979.

Wigan, M.R., "The fitting, calibration, and validation of simulation models", Simulation, 18, 188-192, 1972.

Wingate, P., "The Penguin Medical Encyclopaedia", Penguin Books, 1976.

Wood, D.O., "Notes on the MIT Experiment in Establishing Model Assessment Laboratory", MIT Energy Lab, Cambridge MA., USA, 1978.

Woodger, J.H., "The 'concept of organism' and the relation between embryology and genetics", Quart. Rev. Biol., 5, 1-3, 1930.

Worsley, B.H., Lax, L.C., "Selection of a numerical technique for analysing experimental data of the decay type, with special reference to the use of tracers in biological systems", Biochem. Biophys. Acta, 59, 1, 1962.

Young, J.M., "Some Factors Affecting Respiration in Man", Ph.D. Thesis, Oxford University, 1970.

Young, P., et al, "Macro-economic modelling - A case study", 3rd IFAC Symposium, paper C-3, 145-165, 1973.

Zadeh, L.A., "From circuit theory to system theory", Proc. IRE, 50, 856-865, 1962.

Zeeman, E.C., "Catastrophe Theory", Sci. Amer., 234, 65, 1976.

Zeigler, B., "Theory of Modeling and Simulation", John Wiley, 1976.

Zilversmit, D.B., et al., "On the calculation of turnover time and turnover rate from experiments involving the use of labelling agents", J. Gen Physiol., 26, 253, 1943.

Zwart, A., Smith, N.T., and Beneken, J.E.W., "Multiple model approach to uptake and distribution of halothane", Comp. Biomed. Res., 1972.

INDEX OF KEY TERMS

Most of the thesis has been written in ordinary language, with the addition of technical vocabularies (such as biology and mathematical modelling) whose meanings are well known. However, in considering the philosophical, methodological, theoretical and practical aspects of model validity and validation, it was necessary to introduce many new concepts, and to give more precise meanings to existing terms. Most of these were defined and explained at length in Chapter 4 ("Theory of Model Validity"), but since they are used throughout the thesis, in particular the case studies, the more important and frequently occurring terms and phrases are described briefly below as a reference index, together with a reference to the section in which they were introduced.

1. Data type

   An empirical representation device (§4.3.2.1) classified into:

   (i) Observational data type. Description of empirical phenomena in ordinary language. (§4.3.2.2)

   (ii) Symbolic data type. Mapping of empirical phenomena into a precise, abstract symbolic space (e.g. pattern classification, measurement). (§4.3.2.3)

   The data types corresponding to a model are known as "required data types" (denoted by D_M), and those that are actually observable or measurable are known as "available data types". (D_A), (§4.3.2.4)

2. Domain

   A more or less structured body of knowledge related to a certain research area and which may contain data (putative facts), hypotheses, theories, models, etc. (§4.1.2)

3. Methodology

   A set of principles, rules, or techniques. The study of methods, techniques, criteria, etc. used in scientific practice.

4. Modality

The intended range of application ($\mathcal{R}_I$, see 7) has two modalities or modes (§4.3.1.3.1):

(i) Structural or physical modality. The structure, geometrical properties, and topology of $\mathcal{R}_I$.

(ii) Functional modality. The behaviour, functioning, functional properties, dynamics, etc. of $\mathcal{R}_I$.

5. Model

A theoretical knowledge representation device which embodies both description and explanation of phenomena. (§4.1.2)

6. Objectives

This refers to "modelling objectives" - the purposes a model may serve, the roles a model may play, or the ends a model is intended to achieve in a scientific research programme or practical application (§4.3.1.1). A detailed classification of modelling objectives is made in §4.3.1 based on the following cross-classification:

(i) "General" or "specific" objectives. General objectives - wide, long-term objectives. Specific objectives - relate model to a class of systems or phenomena of interest.

(ii) "Scientific" or "utilitarian" objectives. Scientific objectives - associated with the evolution and testing of scientific knowledge in a domain (see 2), representation of phenomena, etc. Utilitarian objectives - the model is intended for use in a practical situation.

7. Range of application

(i) Intended range of application, $\mathcal{R}_I$. The class of systems or phenomena that a model is intended to represent, together with a set of constraints concerning time scales, resolutions, boundaries, etc. (§4.3.1.3.1)

(ii) Empirically valid range of application, $\mathcal{R}_V$. The extent of $\mathcal{R}_I$ for which the model has satisfied empirical validity criteria. (§4.3.3.3.1)

8. Stage of development

This refers to the degree of theoretical sophistication (i.e. mathematical as opposed to verbal theories), repertoire of available data types (see 1), the range of scientific and methodological techniques, etc. of a domain (see 2) associated with a model. (§4.1.2)

9. System

The term "system" is used loosely in general parlance to mean a collection of associated objects, events, phenomena, ideas, rules, etc., for which, however, there is no generally accepted, yet substantive, definition. "Σ - system" is used here to denote precisely a generic, theoretical or empirical, system whose structural and functional properties can be clearly defined.

10. System of interest, SOI

The wider system, beyond the range of application (or representation) of the model, in which the model (or its conclusions, predictions, etc.) is to be used for utilitarian objectives, i.e. practical application (§4.3.1.3.2)

11. Valid model

A valid model is one which satisfies the modelling objectives for which it is required (§4.1.2). Model "validity" is the extent to which the modelling objectives are satisfied. The different concepts of validity depend on the validity criteria. (see 13)

12. Validation

The process of determining the validity of a model. The systematic application of appropriate validity criteria (see 13) for a given set of modelling objectives. A "programme" or "methodology" of validation is a series of tests for a model, or class of models, which is based on considerations of modelling objectives, data requirements, and appropriate validity criteria. (§4.3.4, §4.5, and Chapter 5)

13. Validity criteria

Tests, rules, means, bases for comparison, or standards for determining the validity of a model (§4.3.3.1). Each validity criteria explicates a different concept of model validity, and is related to different modelling objectives. The validity criteria are classified as follows:

1. Internal validity criteria

   Do not require reference outside model:

   1.1 Consistency validity criterion. Model should contain or entail no contradictions. (§4.3.3.2.1)

   1.2 Algorithmic (or simulation) validity criteria. Require correct and accurate solution or simulation of model. (§4.3.3.2.2)

2. External validity criteria

   Require reference to domain, theories, models, or data outside model:

   2.1 Representational validity criteria. Concerned with testing the extent of the representation of $\mathcal{R}_I$ (see 7):

      2.1.1 Empirical validity criteria. Require that model should agree with empirical data from $\mathcal{R}_I$ - "empirical correspondence". (§4.3.3.3.1)

      2.1.2 Theoretical validity criteria. Model should be consistent with accepted theories or models appropriate to $\mathcal{R}_I$ - "theoretical coherence". (§4.3.3.3.2)

   2.2 Pragmatic validity criteria. Tests of model in satisfying utilitarian objectives either by critical assessment, or evaluation of effect on SOI (see 10) of using model - "pragmatic value". (§4.3.3.4)

   2.3 Heuristic validity criteria. Tests associated with the assessment of the potential of the model for scientific explanation, discovery, hypothesis testing, etc. - "heuristic potential". (§4.3.3.5)