MODEL-BASED FAULT DIAGNOSIS IN INFORMATION POOR PROCESSES

by

John Howell

Thesis submitted for
the Degree of Doctor of Philosophy

Department of Mechanical Engineering
University of Glasgow

March 1991
Glasgow, Scotland

ACKNOWLEDGEMENTS

I would like to thank Professor D Murray-Smith for reading Chapters 1 to 4 and Mr F J Walford (retired, formerly with the Nuclear Materials Control Team of AEA Technology) for his comments regarding the NRTMA application described in Chapters 5 and 6. Last, but no means least, my thanks to Isabelle for her considerable effort in the publication of this thesis.

CONTENTS

CONTENTS (cont.)

CONTENTS (cont.)

Page No.

CONTENTS (cont.)

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

Apart from the following reserved symbols used for the materials accountancy application

I_k the physical inventory measurement at the end of balance period k

MUF_k material unaccounted for at the end of period k

$CUMUF_k$ cumulative MUF_k

U_k the net input measurement during balance period k

V_k^2 the error variance on I_k

U_k^2 the error variance on U_k

the nomenclature will be as below:-

x_i, α lower case characters denote **scalar** quantities

$\underline{c}$, $\underline{y}_k$ lower case underlined characters denote **vectors**.
The presence of a subscript signifies that the vector pertains to a particular time period, k.

J, P_{k_1} upper case characters denote **matrices** or **sets**.
The presence of a subscript signifies a time period, a partition or a square dimension.

$\Delta\underline{c}$ a perturbation or deviation in vector $\underline{c}$

$\{\underline{\theta}\}_i$
$\{P_k\}_{ii}$ this notation is used to represent the i^{th} component of a vector or matrix

J' the symbol " ′ " is reserved for the transpose of a matrix

θ_i^* the symbol " * " is reserved for those elements that are significantly in error

I the identity matrix

$|x|$, $|T|$ the absolute value of a scalar quantity and the determinant of a square matrix

E{x}	the expected value of a random variable x
var{x}	the variance of a random variable x
cov{x,y}	the covariance between two random variables x and y
cor{x,y}	the correlation between two random variables x and y
Log	the natural logarithm
Max	the maximum element in a list or set of real or integer numbers

Quantities that frequently have the same meaning are listed below.

$\hat{\underline{y}}_k$,$\tilde{\underline{y}}_k$	the measurement vector and its model prediction at the end of time period k
$\underline{x}_k$	the state vector at the end of period k
U	the set of all non-path faults and measurement biases
Ψ	the set of all parameters and variables necessary to describe the re-distribution process
Θ	that sub-set of Ψ which the diagnostician is uncertain about
$\underline{\Psi}_k$	the vector of parameters and variables necessary to describe the re-distribution process
$\underline{\theta}_k$	the vector of those elements of θ that are currently of interest on period k
$\underline{w}_k$	a measurement bias
$\underline{v}_k$	zero-mean measurement noise
$\underline{n}_k$	non-path faults
$\underline{u}_k$	the combination of measurement biases and non-path faults
$\underline{e}_j$	the jth natural basis vector

P_k	the covariance matrix of $\underline{\Delta\theta}_k$
R_k	the covariance matrix of the measurements
J	a Jacobian-like matrix
ρ_i	the correlation between MUF_k and MUF_{k+1}
σ_i^2	the variance

ABBREVIATIONS

CUMUF	cumulative material unaccounted for
MUF	material unaccounted for
NRTMA	near-real time material accountancy

SUMMARY

A theory of model-based fault diagnosis is proposed which is suitable for non-linear plants that are information poor. That is, there are a bare minimum of sensors available to operate the process without recourse to analytical redundancy, the sensors output at frequencies which are likely to be low, relative to the dynamics of the plant, and there is considerable uncertainty surrounding any mathematical models that are available. Other approaches are likely to be more suitable for information rich plants. However, it should be of, at least, philosophical interest to the diagnostician who assumes that he is dealing with such a plant, if only because it should lead him to question whether his plant actually satisfies criteria necessary to support this assumption.

The theory argues against model-based fault diagnosis as a panacea for fault diagnosis in favour of a data fusion approach where model-based reasoning forms one input. One consequence of this is that a knowledge-based approach is proposed to implement the different inputs that are possible and to fuse their conclusions. Another is that a model-based alarm system is thought undesirable. Methods are therefore proposed, both to alarm that a fault has actually occurred and to perform a preliminary diagnosis, without recourse to models. Based on control charts, these seek to combine well-known detection theory with a qualitative approach to pattern recognition: the former performing the task of alarm generation, the latter diagnosis.

It is proposed that model-based reasoning be based on two principles, a Principle of Re-Distribution and a Principle of a Minimum Number of Explanations. The Principal of Re-distribution provides the diagnostician with a formal qualitative approach to explaining discrepancies between plant and model whilst maintaining quantitative rigour. This leads to the construction of a candidate space of all possible combinations of all possible explanations. The Principle of a Minimum Number of Explanations is then proposed as a strategy for searching this space. Based on common sense, it attempts to imitate the diagnostician.

A method is then described to appraise a particular candidate set of suspect faults and model inaccuracies. This assumes that the diagnostician has some subjective view of errors in both the model and measurement systems.

The application of both theory and methods to one particular process, that of near-real time material accountancy in fuel reprocessing plants, is described. This has been implemented in a hybrid lisp/FORTRAN environment: the alarm system, model-based reasoning and other knowledge sources being implemented in the lisp environment; plant simulation and candidate appraisal being implemented in FORTRAN. The lisp environment consists, essentially, of multiple production systems. Inference is by forward chaining. System performance to various fault scenarios is investigated, with encouraging results.

A great deal has yet to be done and various issues that are still outstanding are raised in the Conclusions.

1. INTRODUCTION

1.1 General

It is possible that, in the years to come, cheap, robust, reliable instruments will be available to measure every possible process variable that exists in any plant. It is also possible that we will develop an acute understanding of the physical processes that arise in any plant. With these capabilities we should be in a position to detect and locate any fault, almost immediately, by comparing the data collected with appropriate, valid mathematical plant models. Unfortunately, in some industries at least, we are a long way off this desirable state of affairs. Instruments are either not available, too expensive, in terms of either capital or maintenance costs, or unreliable making instrument failure a possibility. Valid models may also not exist.

Over the past two decades, considerable research (1-9) has been carried out into the detection and diagnosis of faults. Various approaches have evolved including analytical redundancy (10), constraint suspension (11) and qualitative reasoning(12). These have largely focused on plants or systems that are information rich: either in the sense that there is a proliferation of sensors throughout the plant, the outputs of which are of a sufficiently high quality and are recorded at frequencies much higher than that of underlying process fluctuations or in the availability of models that are accurate in terms of structure and parameters or both. Certain approaches discriminate between linear and non-linear plants (eg 13): in these cases, it appears (8) that work on non-linear aspects is still in its infancy. A large number of possible applications have been identified including electronics (11), nuclear power plants (14-16), process plants (13,17,18) and ships (19).

This research has taken a scientific approach in that it has attempted to seek systematic formulations that are generally, if not universally, applicable. Thus mathematical techniques have been proposed (13,84) to improve the robustness to uncertainty in the models. The question then arises as to what happens if these formulations are inappropriate? Should the diagnostician give in? Faced with such a situation, he would probably use his common sense (30) and whatever tools that are available to him, to make a judgement specific to the particular situation he is faced with. He would not aim for some grand rigorous statement about plant operation, but would confine himself to finding the fault. Faced with uncertainty, he would be prepared to be wrong.

This thesis describes a common sense approach to model-based fault diagnosis that is suitable for non-linear plants that are information poor. That is, there are a bare minimum of sensors available to observe the process without recourse to analytical redundancy, the sensors output at frequencies which are likely to be low relative to the dynamics of the plant, and there is considerable uncertainty surrounding any models that are available. Such plants may not only exist by design; for instance, the initial failure at Three-Mile Island *(20)* caused the plant to enter a state which was unlikely to have been modelled previously and where the diagnosis of subsequent failures was key to shutting the plant-down in an orderly manner. The work has been motivated by one specific process, that of fast reactor fuel reprocessing and in particular, in the application of near-real time materials accountancy *(48)* to a reprocessing plant. Near real time material accountancy (NRTMA) is a method of enhancing conventional material accountancy techniques to improve the sensitivity and timeliness of detection through the use of in-process instrumentation (generally operator equipment) to increase the frequency of the account. A ***fault*** is then deemed to be any ***significant*** error in the account. Difficulties may arise because the sensor systems and modes of plant operation are optimised for reprocessing and not for NRTMA.

The remainder of this Chapter is devoted to explaining the various terms: *fault, information poor, diagnosis* and *model-based* and outlining the contents of this thesis.

1.2 Faults

As defined by Isermann [(4)] a *fault* is any nonpermitted deviation of a characteristic property which leads to the inability to fulfil the intended purpose. Himmelblau [(2)] defines a *fault* as a synonym to designate the departure from an acceptable range of an observed variable or calculated parameter associated with a piece of equipment. Both are difficult to implement, the former because of difficulties in knowing which deviations are permitted, the latter because of the number of variables and parameters in a plant.

The control of any plant is usually hierarchical in nature in that the primary control objective is normally described by a few characteristic quantities which are affected by other characteristic quantities attributed to the next level down and so on. For instance the primary objective of a power station is to supply power as required, that of a factory energy management system is to minimise total energy consumption and that of a materials accountancy system in a nuclear fuel reprocessing plant is to account for the total throughput of nuclear material. These individual objectives are dependent on the operation of specific plant components which in turn can be viewed as being made up of sub-systems and so on. The net effect is often that the plant operator will tend to 'drive' the plant on data derived from the upper levels of the hierarchy and leave information pertaining to the lower levels to the plant maintenance engineers. Data at the lower levels may be used to trigger alarms or be presented to the operator for information; other items may not be communicated to the operator at all and may not even be recordable. Indeed there is often an economic case against over-instrumentation on the grounds of both capital and maintenance costs. Certain malfunctions may therefore only be detected if and when they affect the control variables monitored at the upper levels or during scheduled maintenance. If they do affect the upper levels then the effect is unlikely to be unique. [illegible]. One can therefore identify a corresponding hierarchy of malfunctions where those at the top are critical to plant operation whereas those at the bottom can wait until scheduled maintenance.

This thesis attempts to differentiate by defining a *fault* or *malfunction* as any occurrence in time that results in the plant deviating from its intended mode of operation. This may be as a result of the total failure of a component or a less serious misalignment or maladjustment; it could be as a result of measurement bias or of some external affect. For instance, an operator or plant engineer may take an incorrect action or there could be an unpredicted variation in the feedstock. Some malfunctions may only occur once and last only a short period of time, others could be more regular but still be intermittent whilst others could develop very slowly.

1.3 Information Poor Plants

It is difficult to define, succinctly, what one means by a plant being *information poor*. The term has largely been derived from inferring what the plant is **not**, that is *information rich*. Scarl et al [18] use the term *sensor-rich* to describe an environment with an abundance of on-line instrumentation. Himmelblau [2] describes the temporary need to install additional sensors and to perform special tests to diagnose certain faults in chemical plants, that is to increase the quantity and quality of the information flow from the plant. One of the aims of the philosophy proposed here is to provide the diagnostician with evidence to support the case for such resource intensive activities.

Consider a hypothetical plant which is *information poor*. The flowrate and composition of the feedstock may vary and although attempts may be made to monitor it, there may always be a possibility that the monitoring process might be *fooled* because certain properties, eg temperature, might be out of range or it might not be designed to monitor certain features or chemicals. The feedstock may also vary at a rate which is significant relative to the frequency at which it is being monitored. The operation of individual units in the plant may be varied with pumps being stopped/started, valves being opened/closed and so on. The operator may not need to record, accurately, the times at which each activity takes place but such information may be needed as input to a simulation for diagnosis purposes. Faults may not only arise because of instrument or actuator failures but also because of, for instance, the build-up of crud (ie solids). It is unlikely that a simulation would be able to predict all possibilities.

1.4 Fault Diagnosis

Fault *detection, diagnosis, isolation* and *location* are generic terms which are used either singly or combined [4,6,8,17] to describe the process of ascertaining that a fault has arisen and of determining its location and cause. Here we assume that the term *fault diagnosis* [6] subsumes them all.

A large number of model-based techniques have been published to detect faults and there is still considerable research in this area that is ongoing[83,85]. One or more of their assumptions are usually violated when applied to non-linear plants which are information poor. A common approach to circumventing any uncertainty is to introduce thresholds to distinguish a fault. The problem with thresholds is that they not only reduce sensitivity to faults, but also vary with variations in the plant inputs and disturbances. Choosing the threshold too small increases the false alarm rate; choosing it too large reduces the power to detect.

There are two main approaches to determining the location and cause of a fault: either a set of all possible faults may be formed and the effect of each individual fault compared with that observed or an argument may be derived on the basis of deviations from what is expected. Rasmussen [21] uses the term *symptomatic search* to denote the former and *topographic search* the latter.

The two main ways of performing symptomatic search appear to be the fault dictionary [2] and the diagnostic tree [24]. The fault dictionary is a list of causes and effects, and diagnosis is performed by looking up the effects in a cause-effect table to see what the cause was. The diagnostic tree is a way of constraining the search to go along different diagnosis paths. Both techniques use look-up tables. For instance, alarm procedures have been used to sort out alarms which are present on the basis of prestored data of association links between the alarms [24]. There are a number of difficulties with using these techniques in larger plants. A large number of entries may be needed in the look-up table. Sufficient data and resources must be made available to trace the faults through the plant. It is very important that all potential faults have been predicted in advance; this can be difficult, if not impossible, in large plants. If this *a priori* analysis is not complete and correct, then diagnosis may not only be impossible but also positively misleading. They are therefore of only limited applicability to plants that are information poor.

Topographic search methods rely on their ability to predict plant performance so that discrepancies or *residues* can be used to diagnose the fault. There appear to be two approaches, *analytical redundancy* [10] which derives from modern control theory and the method of *violated expectations* [11,22] which comes from Artificial Intelligence. Both appear to be applicable to diagnosis in plants which are information poor because they do not require *a priori* knowledge of possible faults.

The technique known as *constraint suspension* [11] is derived from the method of violated expectations. Faults are not hypothesised explicitly but are hypothesised in terms of their observed affect on the system. For instance, a control actuator failure could be identified by the hypothesis, 'a constant valve opening of X% would explain all the discrepancies or *symptoms* observed.' Faults may therefore be systematically isolated without recourse to the set of all possible faults. The technique lends itself to a hierarchical approach in that it should be possible to generate[17] discrepancies at component level, then sub-component level and so on. However this is likely to neglect common mode failures which can percolate through lower levels. For instance, control valves may be pneumatically driven from the same air supply or chemical samples may be analysed at the same laboratory station. A common mode failure may then be misinterpreted as being a multiplicity of different faults at the component level.

One approach [4,7], an analytical redundancy technique, is to recourse to on-line identification to evaluate the model parameters. Unexpected changes in these parameters can then be correlated with possible fault scenarios. This assumes that either the parameters are known explicitly or they can be identified prior to any fault developing. Problems arise with either assumption when dealing with an information poor plant, firstly because of uncertainty regarding the model structure and secondly, because the frequency of data collection is low relative to the variation in feedstock and changes in mode of operation. It is difficult to ensure that there will be sufficient information to perform identification in a reasonable timescale and that no fault develops whilst the initial identification proceeds.

1.5 Modelling

There is no such thing as a unique model of a process [2,82]. Any model-based approach which is applicable to information poor plants must tackle uncertainty surrounding model structure and parameters. Choosing a relevant model may not be straightforward, because the decision as to whether or not a '*valid*' model is available, must be taken in the context of the application. *Validity* is application specific. For instance, a model need not be perfectly accurate for the purposes of constraint suspension; it need only be sufficiently accurate to ensure the uniqueness of the mismatches.

This leads us to the fundamental philosophical question: can a model be proved to be correct when used in the context above? The basis for an answer to this question was first provided by the philosopher Sir Karl Popper [31] in 1934, who was interested in the characteristics of a scientific theory. In Popper's view, a theory can never be proved by any of its successes, since a new test, perhaps as yet not thought of, may come along that it will fail. Failure in any fair test, on the other hand, indicates a fault in the theory. One possible interpretation of this is that models may evolve or *learn* by their failures: a model may be deemed to be *valid* until it fails to detect and locate a particular fault when the model can be revised to take account of that particular failure. However this may be of limited utility because the same fault might never occur twice: for instance, it may have resulted in a design re-evaluation or a change in operational procedure. Clearly the credibility of any fault detection and diagnosis system would also depend on the frequency of model failure.

This thesis describes an alternative view which [illegible] is more pertinent to fault diagnosis. It requires that the diagnostician understand concepts of systems modelling and in particular the Laws of Conservation. That is, the diagnostician does not treat the *model* as a set of equations but rather as a mathematical description. Lind [25] has proposed one possible approach based on this theme. He uses a method of computer modelling he calls Flow Modelling to predict the distribution of mass and energy around a plant. This distribution can be compared with the measured distribution and faults hypothesised to explain the pattern of discrepancies [26].

1.6 Overview of Thesis

This thesis is divided into three parts: theory, method and practice. Chapter 2 describes a theory of model-based fault diagnosis applicable to any plant. The approach adopted is rather laborious but necessary to ensure general applicability. Chapters 3 & 4 outline methods which implement certain aspects of the theory. Chapters 5 & 6 then describe one particular application, that of NRTMA. Neither theory, method nor practice are complete; Chapter 7 therefore lists some aspects that are still outstanding.

2. A THEORY OF MODEL-BASED FAULT DIAGNOSIS

2.1 Introduction

The main aim of any fault diagnosis system is to hypothesise possible faults or classes of faults which *explain* the measurements observed. A model-based fault diagnosis system works on the premise that it is possible to generate suitable hypotheses by looking at the discrepancies between model predictions and these measurements.

- What techniques can be used to hypothesise these faults?
- What level of detail is required of the simulation?
- What is meant by *explains*?
- If more than one fault is hypothesised, can they be ordered in some way?

Any theory of model-based fault diagnosis must tackle these basic issues.

Model-based fault diagnosis is discussed and problems of model validity raised. These are circumvented by proposing that the measurements to be explained should be restricted primarily, to those measurements that have nominal values. That is those measurements that are usually capable of being alarmed. Such a system is unlikely to identify a specific fault unless the fault invokes a unique combination of symptoms. It is therefore proposed to reduce the number of possible candidates by requiring the simulation to *explain* other aspects of the data collected. Further questions then arise as to how the 'other aspects of the data collected' should be identified and handled.

The theory is presented as a set of propositions in *italics*.

2.2 A Strategy For Model-Based Fault Diagnosis

What do we mean by model-based fault diagnosis?

In its simplest form [8] it can be considered as consisting of a parallel simulation predicting plant measurements, a comparison between these predictions and the actual measurements and a means of generating fault hypotheses on the basis of any discrepancies obtained.

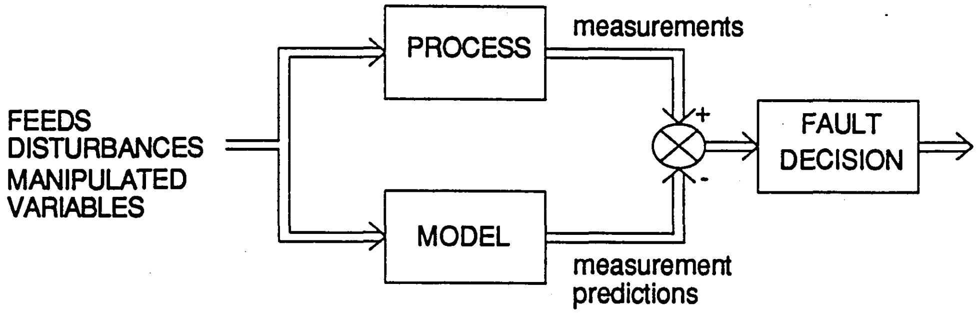


Figure 1: Model-Based Fault Diagnosis

A slightly more sophisticated arrangement[4] is where the simulation outputs model parameters instead.

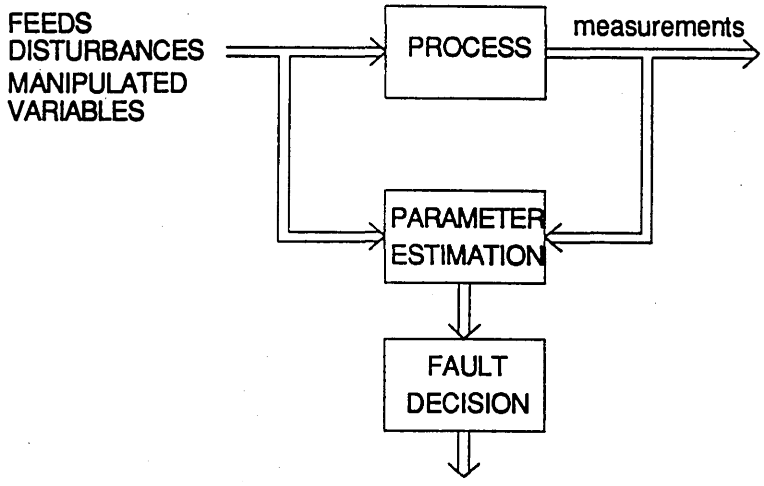


Figure 2: Model-Parameter-Based Fault Diagnosis

There are two ways of generating hypotheses: either the system can hypothesise a fault explicitly by correlating the discrepancies with those predicted *a priori* on the basis of a list of hypotheses or it can specify it in terms of that set of plant conditions that would explain the discrepancies. The former can suffer from problems of completeness, of validity, because models must be known *a priori*, and of practicability, whereas the latter suffers from problems of validity. Both have a role to play provided they are applied selectively.

If symptomatic search, ie the former, were to be pursued by itself, it would require that the list of possible hypotheses be complete even if only a small proportion of malfunctions are of direct interest. A malfunction can be thought to occur for one of three reasons: as a result of noise, of a malfunction in the measurement or alarm system, or of a plant malfunction or maloperation. For instance, a measurement system might malfunction because of a hardware fault, a violation of the physical model upon which the measurement is based (eg single phase flow), a parameter error (eg the device may not include any temperature compensation although it might be susceptible to changes in temperature) or an error in the actual recording process. Only a plant malfunction or maloperation is crucial to plant operation, a malfunction of a measurement system may be crucial, indirectly, because it might mislead the operator or provide him with insufficient information on which to operate the plant. The others merely affect the credibility of the alarm system. But all must be entered on the list. It is unlikely that the list will ever be complete. This does not mean that techniques based on previous observations, whether direct or through simulation, like fault dictionaries or diagnostic trees cannot be applied with success but rather that there is often no guarantee that they can be.

If topographic search were pursued by itself, then the diagnostician would be relying on his ability to simulate all effects. His chosen model would have to be robust in that it would have to be valid for any operating regime. This includes those scenarios not foreseen during the development of the model. It is unlikely that it will ever be possible to guarantee robustness in general even if modelling and fault diagnosis techniques were to be optimised with this in mind.

The following propositions seem reasonable given that neither approach is perfect.

Model-based fault diagnosis should pursue topographic search.

Given that there is a poor chance of success in producing a complete list of hypotheses for information poor plants, the only option appears to be to examine whether topographic search can be made sufficiently robust to diagnose *most* faults. However there is one *proviso.*

Model-based fault diagnosis must not exclude non-model-based approaches.

It would be a folly, against this background of uncertainty, to omit any technique capable of generating fault hypotheses whether they produce specific or classes of faults. For instance, hypotheses could be generated by applying heuristics which either pertain to past history or to any peculiarities that are known about current operation or even to some profound thought that the operator might have. Although the proposed system is notionally based on models, it must be able to combine or fuse the lists of hypotheses obtained by applying any technique available. That is, it must include non-model-based symptomatic search.

One possible approach, Figure 3, to implementing such a system is to hold knowledge pertaining to each technique in separate knowledge-sources arranged in a star network around a *supervisor* which gathers and combines the various hypotheses. The system is data-driven in that the various measurements are analysed by the various knowledge-sources.

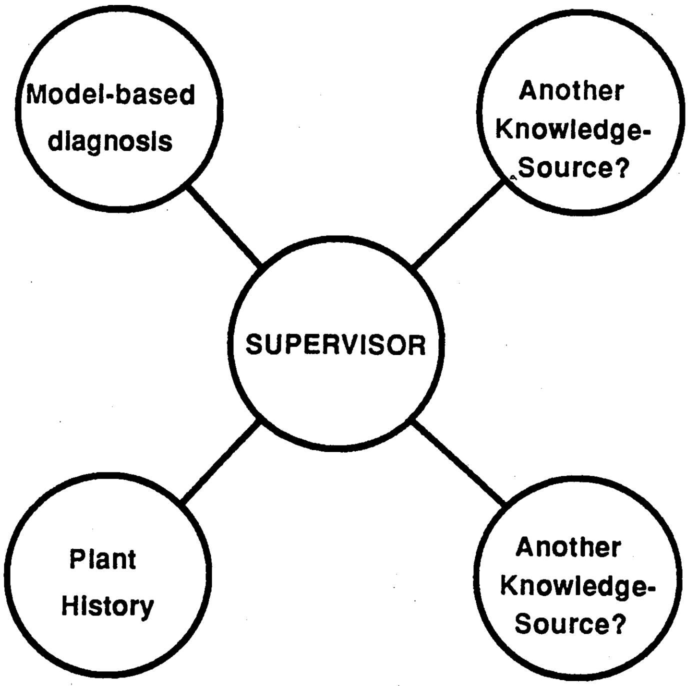


Figure 3: Outline Knowledge-Based Approach

A number of guidelines seem appropriate to regulate the generation of hypotheses from the various knowledge-sources.

Hypotheses must be justified

A fault diagnosis system cannot say 'this is **the** fault' unless it has actually observed it. For instance, an instrument may only be definitely at fault if it is outputting a signal which is not attainable irrespective of the actual state of the substance being measured. The best it can do is to present the operator with a list of hypotheses together with supporting evidence. It is then up to the operator to either choose between them or even decide on something quite lateral. What is important though is that the list of hypotheses, themselves, must be properly justified or validated. Clearly the credibility of a fault diagnosis system will depend on the validity of its suggestions.

Hypotheses must be valid

How can a diagnostic system be certain that it is producing valid or justifiable hypotheses. One of the central themes of Popper's work [31] is that a scientific theory must be falsifiable. The equivalent, in fault diagnosis would be that all hypotheses should be readily testable. Unfortunately, faults may be intermittent. Others may be difficult to test in that they may be due to a peculiar operating regime or, for instance, a build-up of sludge somewhere in the plant. Such a criterion may not, therefore, be practicable.

A looser criterion is needed in cases where the criterion of testability is not practicable. If the hypothesis cannot be tested directly, why not test the indicators used to generate the hypothesis instead. This concept has a considerable impact if models are used to generate hypotheses because the model structure, its parameters, inputs and the assumptions it is based on, must all be testable. A complicated model with a detailed structure, a large number of parameters identified on-line and a host of assumptions will be difficult to validate.

Testability tends to undermine credibility in that a fault diagnosis system will not be too popular if it either continually asks the operator to obtain information in addition to that already available or requires a large capital investment to increase the scope of the data collected. The criterion of testability encourages simplicity.

2.3 An Introduction to Modelling For Fault Diagnosis

It may be concluded from the above that the simplest, valid model possible would be the ideal choice for a model-based fault diagnosis system. What is meant by *valid*? Before discussing this, it may be worthwhile to reflect on some of the issues that surround the development and application of models.

2.3.1 Mathematical Models - Their Structure

The normal approach to developing a physical law based dynamic model of a system is to perform one or more balances of either the system, in its entirety, or of the system divided into a number of sub-systems. For instance, these balances may be of mass or energy or momentum and the sub-systems may be obtained by dividing the system into individual physical components. These balances may be used to estimate the change in internal state of the individual sub-systems over any period in time, k,

state at end of period k = state at beginning of period k
+ net change over period k

This assumes that the state within a particular sub-system is either uniform or varies in some pre-defined manner, for instance, linearly. The state at the end of period k will therefore be known provided both an initial condition and the net changes over the periods are known.

The balances that are applied to any specific situation are chosen to estimate certain properties of a particular sub-system directly. For instance, these properties may include volume, mass, density or enthalpy. These properties are often denoted by the term 'state variable' when applied to these sub-systems.

There may be more than one way in dividing a particular system into sub-systems. The decision is usually made on the basis of the assumption surrounding the variation of the state within the sub-system, the need to calculate the appropriate net changes and in computational complexity. The larger the number of sub-systems, the greater the chance of internal uniformity. Conversely, the larger the number of sub-systems, the larger the number of net changes that need be determined and the greater the computational problem. The chance that these net changes can be estimated directly from available in-process measurements also reduces as the number of sub-systems increases: recourse must then be made to estimation on the basis of differences in properties between 'neighbouring' sub-systems; this evaluation often involves parameters which need identifying. On-line identification may then be necessary if some of these parameters vary with the state of the plant. Although this may not cause problems in particular applications, it does require that any identification be carried out either prior to any fault developing or at a pace slow enough to ensure that any fault is not identified as being a parameter variation. Alternatively, if discrepancies are based on parameter variations then these must be identified quickly enough and with no ambiguity.

There is therefore a case to minimise both the number of sub-systems involved and the number of parameters that need on-line identification. One approach to this is to accept that the assumption of internal uniformity may be violated provided that the internal variation can be shown to be within reasonable bounds. That is, to use a lumped parameter approach in general. This has the effect that a particular measurement of a sub-system may differ from that estimated by the model. The measurements are said to have *'systematic errors'*. In addition, the expressions used to determine the net changes are also likely to be affected.

2.3.2 Mathematical Models - Disturbances and Manipulated Variables

In reality there could also be some uncertainty surrounding the specification of the inputs to the model. These may be viewed as largely being the manual interventions that are used to drive the plant. For instance, set-point changes and times at which pumps are switched on or off. Times at which operational modes are changed may only be recorded approximately and may relate to the start, middle or end of the change-over sequence; certain variables, that are not critical for plant operation, may not be recorded reliably; it is not necessarily certain that what was recorded using one convention during one period, would be recorded in the same way during the next and so on. It may therefore be difficult to produce a quantitative model of these inaccuracies objectively: for instance, certain variables could be extremely accurate one period and in considerable error the next.

2.4 Defining the Stucture of a Model-Based Fault Diagnosis System

The following statements now seem evident,

Model-based fault diagnosis is really automatic model refinement

The decision as to whether or not a model is *valid* must be taken in the context of the application; the context here being its ability to *explain* the measurements. Model-based fault diagnosis would like to assume that all the discrepancies between the actual measurements and its own model predictions are as a result of either some fault or noise. That is, it would like to assume that the model is valid. Unfortunately this is often not possible because of the uncertainty described above. Either the process of model-based diagnosis must be viewed as being the process of diagnosing both model inaccuracies and faults, or robust methods must be evolved or thresholds must be introduced resulting in reduced performance as described in Chapter 1. If the faults are included in the model then diagnosis can be viewed as performing automatic model refinement to ensure model validity.

Automatic model refinement is not a panacea for fault diagnosis

The term *automatic* model refinement is largely a misnomer because it will always require some degree of interaction with the user: someone must tell the system what is, and is not, possible. The plant operator is unlikely to accept this change in emphasis if it requires him to investigate problems of model inaccuracy. Model-based reasoning cannot, in general, be viewed as being a panacea of fault diagnosis.

This leads to the following proposition.

Preferably, fault diagnosis should not be centred on model-based reasoning.

We have already argued that model-based fault diagnosis should be augmented by other knowledge. We now argue that model-based fault diagnosis should, itself, augment something else. If it is known that a fault exists and it can be classified in some way, then model-based reasoning can focus the model refinement on these aspects and effectively play-down problems of model validity. A procedure for doing this is described in Chapter 3.

The knowledge-based approach would then be revised as shown below.

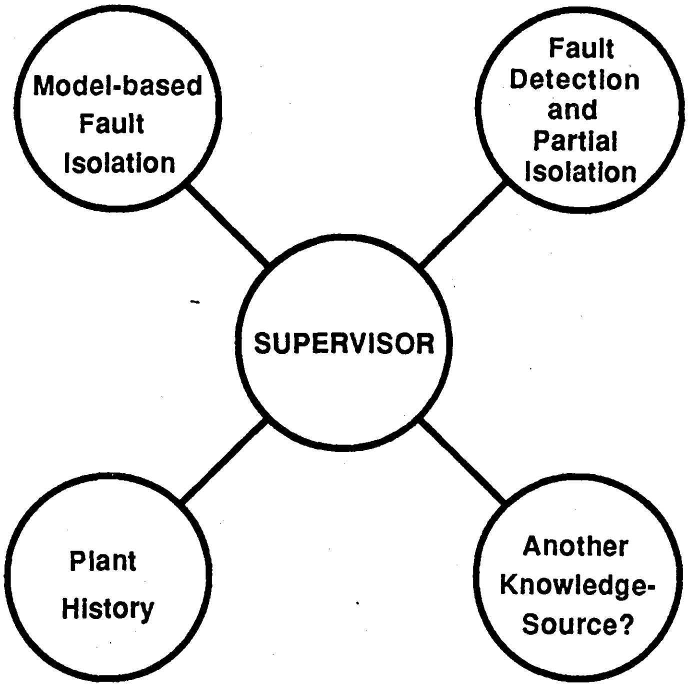


Figure 4: A More Realistic Knowledge-Based Approach

Clearly this does not rule out the possibility of a model-based approach remaining central, it merely stresses its undesirability.

2.5 Non-Model-Based Fault Detection and Partial Isolation

A plant is rarely operated precisely to a state determined by a model. A model may be used to derive, for instance, a flowsheet, so that plant control variables like flowrates and pressures can be input as controller setpoints but other variables will be allowed to take up there own values. However, certain variables, often called characteristic variables, will be key to satisfactory production and will be monitored closely in that a nominal value and tolerances will be specified and alarms will be set to alert the operator of a problem. These variables may have to be estimated on the basis of other measurements. For instance energy consumption must be obtained by integrating instantaneous power over the period or the position of the heavy metal front in the first stage of a solvent extraction system may have to be estimated by either looking at the rafinate or some other measurement.

It is commonplace for the operator to display charts of these variables to enable him to detect the occurrence of a fault. These charts may simply be plots of the individual instances in time or of cumulative instances in time or of some other variable[2]. Depending on the application they may be updated almost continuously, every minute or hour or even every day. The frequency largely depends on the methods of data collection available and the rates at which faults are expected to develop. Although quality control techniques deal primarily with open loop processes they may also be applicable to closed loop situations where, for instance, the quantities of interest could be the states of the final control elements.

It is therefore proposed that fault detection and partial isolation should be based on the deviations in these variables. Although they are likely to deviate during normal operation, techniques exist to accommodate these effects. One possible methodology is described in the next Chapter.

Unfortunately restricting the number of measurements in this way reduces the possibility of identifying a malfunction uniquely. Hence the need for model-based reasoning. There is now a difference though because it need only be applied once the deviations have first been used to say that there is a malfunction and secondly been used to initiate the diagnostic process.

Fault diagnosis must now be considered in terms of two separate contexts: of deviations and model-based reasoning.

A diagnosis will first be deemed to be valid if, with the fault hypothesised included, all deviations are returned to within tolerances. Having satisfied this context, the model must then predict all other measurements to within measurement tolerances.

2.6 Model-Based Fault Diagnosis

2.6.1 Modelling for Fault Diagnosis

Model invalidity can be tackled in two ways: either a model must be chosen carefully to ensure that it can never be invalidated or we must reason with multiple models. Although in Popper's view a scientific theory can never be said to be irrefutable, it is unlikely that this event will occur in the models of interest here. It is not the universal laws that will fail but the way they are applied. We therefore argue for a single model, examine the modelling issues necessary to achieve this and propose a single-model-based diagnostic system.

One may persist and still ask the question as to what happens if the model is found to be invalid. Is the system discedited? Although unlikely to occur, the assessment carried out on the basis of the deviations will still stand, only the conclusions based on the other symptoms will be put into doubt.

2.6.2 Handling Uncertainty With a Single Model

We tackle model uncertainty as far as it concerns fault diagnosis by examining how to ensure the successful application of the Laws of Conservation. We start by considering the simplest structure, that of a single open system. If the system is viewed as a whole, then Laws of Conservation can be applied both simply and accurately provided that transfers into and out of the system are monitored accurately. If the transfer errors can be assumed to be error free then we can infer global conservation.

One of the main concerns when applying the Laws at a more local level is to ensure that global conservation is maintained. This is often achieved computationally by using unique variables to define the flow from one sub-system to an adjoining one. If global conservation is maintained, then the effect of any simulation error, whether structural or due to parameter uncertainty, will be observed as a re-distribution of the values taken up by the various state variables defined when forming the balances. For instance, mass may be re-distributed.

If these sub-systems are chosen along physical boundaries, the flows between the sub-systems will align with the well-defined physical connections of the system. The effect of any simulation error whatsoever must then manifest itself at the physical ports of the sub-system. It follows that any simulation error in a particular sub-system will cause a re-distribution to neighbouring physical components.

Similar re-distributions will also be obtained if a process fault results in a change in the flow pattern. For instance, a faulty valve may cause a re-distribution of mass. If this type of fault is denoted by the term *path fault* then *non-path faults* will be those faults that only affect a single sub-system. A typical example is that of a measurement error. It follows that non-path faults will affect global conservation whereas path faults will not.

The question therefore arises as to whether a particular re-distribution can be attributed, uniquely, to a particular class of simulation errors or faults. It is envisaged that the only faults that would produce a similar re-distribution would be those that had a direct effect on the *paths* between sub-systems. Other faults would either have a more local effect, for instance measurement errors, or percolate through some other set of paths, for instance common mode failures.

It should therefore be possible to locate and to discriminate between simulation errors and path faults on the one hand and non-path faults on the other, by assessing these re-distributions. However some form of further investigation will be required to discriminate between simulation errors and path faults.

This leads to the suggestion that the state variables should be estimated, either directly or indirectly, from in-process measurements and compared with model predictions. Clearly, it may not be possible to estimate every state variable: the more states that can be estimated, the better the discrimination.

One obvious pitfall with this approach is that of initial conditions: the Laws of Conservation may be applied successfully but the end result may still be in error. Incorrect initial conditions may be viewed as being equivalent to non-path faults that only occur on the time period that the simulation is started. This must be accommodated in the fault diagnosis system.

The model structure may be suspect because of problems with internal uniformity. It is proposed to ignore this issue, explicitly. Inaccuracies with the model structure will then be mistaken for non-path faults because the state measurements may now appear to be biased and path faults because the inter-sub-system flows may now be calculated incorrectly.

These concepts can be loosely described as a principle, a Principle of Re-Distribution.

2.6.3 Model Formulation

Let $\underline{x}_k$ denote the vector of variables used to describe the actual state of the system at the end of a particular period in time k when measurements become available and let $\underline{\tilde{x}}_k$ denote the equivalent vector in the model. Let $\underline{\psi}_k$ denote the vector of parameters and variables necessary to describe the re-distribution process, then

$$\underline{\tilde{x}}_k = \underline{\tilde{x}}_{k-1} + \underline{g}(\underline{\tilde{x}}_{k-1}, \underline{\psi}_k, k)$$

where $\underline{g}$ is the vector of calculated net changes.

Let $\underline{\hat{y}}_k$ denote the vector of measurements obtained to generate the *symptoms*. Let these be related to the state measurements by

$$\underline{\hat{y}}_k = \underline{h}(\underline{\hat{x}}_k)$$

where ideally, $\underline{h}(\ )$ would be the identity matrix.

Then the *symptoms* are described by the vector $\underline{\hat{y}}_k - \underline{\tilde{y}}_k$ where $\underline{\tilde{y}}_k = \underline{h}(\underline{\tilde{x}}_k)$

It has already been suggested, Section 2.3, that there could be some uncertainty surrounding the simulation vector $\underline{\psi}_k$. The expressions used to determine the net changes, that is $\underline{g}$, may also be inaccurate; for instance, they may be affected by the assumption of individual uniformity: the model structure may therefore be suspect.

Although this uncertainty exists, it is still necessary to estimate $\underline{\psi}_k$ for the simulation to proceed. Let the set Ψ contain all scalar variables that are elements of $\underline{\psi}_k$. Then certain elements of Ψ will be known precisely whilst we might be less confident about others. Suppose that we attempt to identify that subset Θ we are uncertain about and form a vector containing them, ie.

$$\underline{\theta} \in \mathbf{R}^n$$

where $\{\underline{\theta}\}_i \in \Theta \in \Psi \quad \forall i, i = 1,\ldots n$

Clearly it is not possible to guarantee that this vector contains all elements that are actually in error. Let the estimates of $\underline{\theta}$ input to the simulation for period k be denoted by $\hat{\underline{\theta}}_k$ and let the errors in these estimates ie $(\underline{\theta}_k - \hat{\underline{\theta}}_k)$ be denoted by $\underline{\Delta\theta}_k$.

We define a new function $\underline{f}$ to relate the measured sub-system outputs at the end of one period to the sub-system outputs at the end of the subsequent period obtained by simulation:

$$\tilde{\underline{y}}_k = \hat{\underline{y}}_{k-1} + \underline{f}_k(\hat{\underline{x}}_{k-1}, \hat{\underline{\theta}}_k, k)$$

where the analytic form of this function need not be known.

Note that the initial conditions $\hat{\underline{x}}_{k-1}$ are assumed to be derived from measurements and not from the simulation unless it is necessary to do so because of a lack of suitable measurements. This reflects the belief that the measurements are more likely to be 'correct' than the simulation.

The measured sub-system outputs $\hat{\underline{y}}_k$, may be biased by $\underline{w}_k$ so that

$$E\ \{\hat{\underline{y}}_k\} = \underline{y}_k + \underline{w}_k$$

and corrupted by measurement noise, $\underline{v}_k$, assumed to be of zero-mean and uncorrelated in time, thus,

$$\hat{\underline{y}}_k = \underline{y}_k + \underline{w}_k + \underline{v}_k$$

$$\text{cov}\ \{\hat{\underline{y}}_j\ ,\ \hat{\underline{y}}_k\} \equiv R_k\ \delta_{kj} \quad \epsilon\ \mathbf{R}^m \times \mathbf{R}^m$$

where δ_{kj} is the Kronecker delta function,

$$\delta_{kj} = \begin{cases} 1 & \text{for } k = j \\ 0 & \text{for } k \neq j \end{cases}$$

The bias $\underline{w}_k$ will not only be affected by the actual physical measurement system but also by the assumptions on which the measurement is based, for instance perfect mixing. Let set U contain all these possible causes. Set U must also contain all possible non-path faults because these will have a similar affect on the discrepancy $\hat{\underline{y}}_k$-$\tilde{\underline{y}}_k$. Let the actual effect that these non-path faults have on the discrepancies on a particular period k be denoted by vector $\underline{n}_k$.

As a consequence of this, the initial conditions $\hat{\underline{x}}_{k-1}$ are likely to be in error. We accept the inevitability of this and aim to minimise its effect on the fault diagnosis process by ensuring that faults are identified and incorporated into the simulation sequentially in time. That is, we only start the simulation from a period that we are 'confident' is error free. This may mean that measurements pertaining to more than one period may have to be compared with the simulation in one go. Then

$$\bar{\underline{y}}_k = \bar{\underline{f}}_k\left(\hat{\underline{y}}_{k-1}, \hat{\underline{x}}_{k-1}, \hat{\underline{\theta}}_k, k\right)$$

$$\text{where} \quad \bar{\underline{y}}_k = \begin{bmatrix} \tilde{\underline{y}}_{k+n} \\ \tilde{\underline{y}}_{k+n-1} \\ \tilde{\underline{y}}_{k+n-2} \\ \vdots \\ \tilde{\underline{y}}_k \end{bmatrix}$$

must be used instead of the formulation above.

Clearly this confidence could be unfounded, so it is important to include this possibility in the reasoning process.

2.6.4 *Some Observations*

We imagine ourselves as the diagnostician and suggest that he might take the following viewpoint.

The 'symptoms' should be categorized, qualitatively.

The choice of categorization is open to discussion. For instance, it could be based on statistical or fuzzy measures [36]. A statistical representation is preferred here, primarily because of its obvious compatibility with the measurement statistics. The ith *symptom* could be deemed to be in error at level n, if n is the largest integer:

$$\left|\{\underline{\hat{y}}_k - \underline{\tilde{y}}_k\}_i\right| \preccurlyeq n\sigma_{ii}$$

where the parameter σ_{ii} is based on the some measure of the uncertainty surrounding the ith *symptom.*

A problem arises here because it is difficult to specify, in absolute terms, the uncertainty in $\underline{f}$. One possibility is to make the pragmatic decision to ignore, completely, errors associated with modelling uncertainty and let

$$\sigma^2_{ii} = \{R_{k-1} + R_k\}_{ii}$$

Bounds Can Be Placed On Simulation Errors.

It is likely that the diagnostician would have some subjective appreciation as to possible inaccuracies in $\underline{\hat{\theta}}_k$ provided he assumes no fault exists. Although imprecise, he would have some idea of the order of magnitude and probably 'err on the large side' to accommodate any unforeseen factors:

$$E\{\underline{\Delta\theta}_k\} = E\{\underline{\theta}_k - \underline{\hat{\theta}}_k\} = 0$$

$$\mathrm{cov}\{\underline{\Delta\theta}_k, \underline{\Delta\theta}_k'\} = P_k$$

and it is assumed that these inaccuracies are not correlated with the discrepancy measurement errors. Although not strictly necessary, it is likely that P_k would be assumed to be diagonal (P_k can always be diagonalized by re-defining $\underline{\theta}_k$ to accommodate these cross-correlations).

The diagnostician would probably agree that the *symptoms* could be explained by errors that were *insignificant*, if a $\underline{\Delta\theta}_k$ could be identified to reduce the *symptoms* to a specified level:

$$\left|\{\underline{\Delta\theta}_k\}_i\right| < m \quad \sqrt{\{P_k\}_{ii}} \qquad \forall i$$

> where integer m, the level of *significance*, would probably be chosen to be either 2 or 3 by direct analogy with hypothesis testing.

Conversely, if an explanation could only be obtained by allowing certain elements to disobey this condition, he would argue that these elements were *significantly* in error. Let us use a star to indicate these elements, for example, θ_1^* and let Θ^* denote the set of all elements that are significantly in error.

Catastrophic failures should be handled differently to non-catastrophic failures.

Before going any further it is worth pointing out that a slightly different approach would be taken if a fault, or inaccurate input data to the simulation, results in a catastrophic failure. For instance, a tank emptying whilst it is still feeding the process downstream. These failures are easier to detect because firstly a failure has definitely occurred and secondly it is usually clear whether it has occcurred in either the plant or simulation or both. For instance, an empty tank is unlikely to go unnoticed. This leads to the following line of reasoning:

```
    (catastrophic plant) & (catastrophic simulation) →
                                              (maloperation)

    not (catastrophic plant) & (catastrophic simulation) →
              ((simulation fault) (measured–plant–input incorrect))

    (catastrophic plant) & not (catastrophic simulation) →
   ((maloperation) (simulation fault) (measured–plant–input incorrect))
```

where → means *implies.*

However it may still be necessary to pursue the same line of diagnosis to determine the root cause in the second and third cases.

2.6.5 The Candidate Space

A candidate is a particular hypothesis for how the plant differs from the model. That is, it is a set of elements of $\Theta \cup U$ whose values could, hypothetically, be manipulated to explain the discrepancies. Figure 5 shows the candidate space for a simple example containing just two path and two non-path errors. Each set of explanations, indicated by [.], is possible giving 35 candidates in all. More realistic situations would involve much larger candidate spaces. If considerable uncertainty surrounds the diagnosis, then candidates at the top of the lattice are likely to be true; the greater the certainty, the more likely the candidates towards the bottom of the lattice become.

The number of explanations contained in each of the candidates towards the top of the lattice are likely to be far greater than the number of symptoms available to perform a diagnosis. It will therefore be difficult to discriminate between the various candidates unless it is possible to eliminate some of the explanations on grounds of inter-dependence or through non-symptom related arguments.

Fortunately, the diagnostic problem changes from being one that is underdefined to one that is overdefined as we move down the lattice. We therefore seek an argument to enable diagnosis to proceed by considering candidates towards the bottom of the lattice where techniques like regression can be applied. This is contrary to the *scientific* view that we should tackle uncertainty as it really is. Here we propose to use our common sense to avoid the issue.

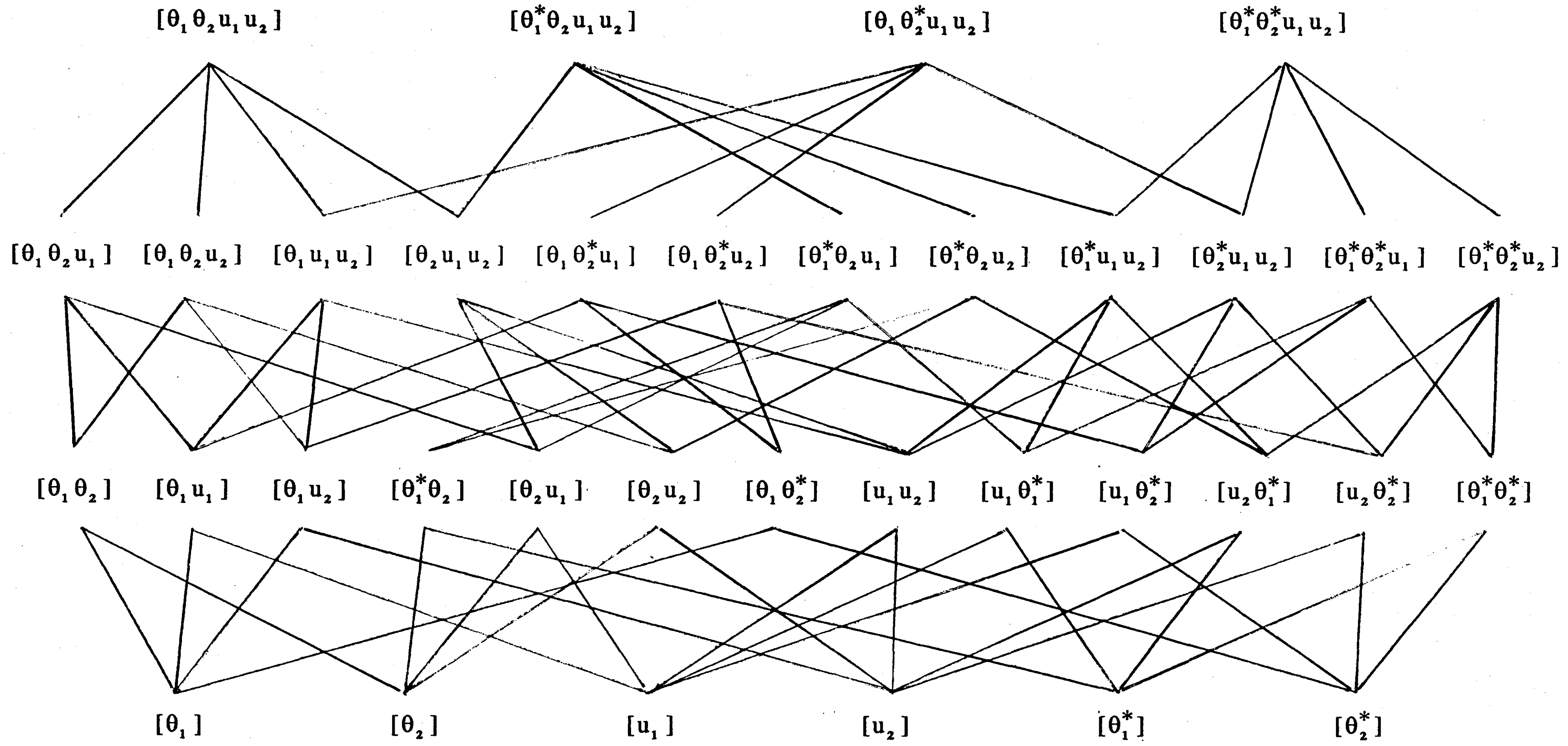


Figure 5: A Typical 4-element Candidate Space

2.6.6 Common Sense Reasoning

As to what is meant by *common sense* the reader is referred elsewhere[30]. Two aspects of importance here are that common sense 'is concerned with the concrete and particular' and that 'its function is to master each situation as it arises.'

The following posture is argued as being *common sense.*

1. Insight is not required into the state the plant is actually in but only into why the discrepancies have occurred.

2. Although a large number of errors may exist in the model structure, parameters and measurements, the diagnostician is only interested in those errors that would result in the observed symptoms. That is, he would not be interested in the true value of $\underline{y}_k$ nor of $\underline{\theta}_k$ but only of those elements of U, Θ^* and, if necessary Θ, that could provide an explanation.

3. If the term *image* is used to denote any plant state that is likely to exist, then the diagnostician will only be interested in those images that explain the symptoms.

4. Other insights gleaned from, for instance, history, fault detection and partial isolation should aid the search.

A heuristic which we will call The Principle of a Minimum Number of Explanations is proposed to encompass some of these ideas.

2.6.7 The Principle of a Minimum Number of Explanations

The symptoms can be explained by a minimum number of errors with all other errors pertaining to the symptoms being assumed to be zero.

That is, it is assumed that all elements of U, Θ^* and Θ can be ignored except those incriminated.

There is a fundamental difference between this and The Principle of Parsimony which Reiter (29) interprets as meaning that

"a diagnosis is a conjecture that some minimal set of components are faulty."

One refers to errors whereas the other refers to faults. The Principle of Parsimony merely proposes that diagnosis move upwards from the bottom of the lattice searching for faults, which either have, or have not, occurred, until a suitable candidate is identified. The Principle of a Minimum Number of Errors has no scientific basis; it also argues for a bottom-up search but now assumes a minimum uncertainty. It does, however, conform to the notion of testability described in Section 2.2.

2.6.8 Candidate Evaluation

There are likely to be a number of different techniques capable of evaluating the credibility of individual candidates. As such they are only techniques and including one particular method in the theory would tend to detract from its intended universal applicability. This is therefore left to a separate chapter, Chapter 4.

2.6.9 The Basic Strategy

It is important that the candidate space be searched prudently because of the large number of candidates that are likely.

Information external to the model-based diagnosis can be used to guide the search. If no faults are expected then the search should start with candidates containing insignificant path errors only, if path faults are expected then candidates with starred elements are to be preferred whereas if non-path faults are expected, candidates with non-path elements are of prime interest. Thus a combination of breadth-first and depth-first search is proposed. All candidates on the bottom row, which contain one particular type of error, are examined first, then the next row and so on. As will be seen in Chapter 4, computational efficiency may also be a factor in how the lattice is searched. A decision must be made as to whether to continue up the lattice, indefinitely, considering only those candidates that contain one particular type of error or whether to broaden the search: if the Principle of Parsimony were to be applied then at some point other candidates towards the bottom of the lattice would be preferred. Such a decision is application specific and one method of deciding when to broaden the search is described in the Application Chapter, Chapter 6.

Another aspect common sense would deem to be worthy of consideration is that of focusing on one particular part of the plant. If discrepancies are confined to one part then there may be an argument to eliminate candidates not pertaining to it. However caution should be exerted because of the problem of cancelling errors. For instance, a path error between two inter-connecting sub-systems A&B when combined with a non-path error, of similar magnitude, in B could be mistaken for a single non-path error in A.

Such candidates can therefore only be eliminated from the lower rows of the lattice where it is known that other errors cannot be present.

Efficient diagnostic procedures based on the Principle of Parsimony have been previously proposed in the literature [27-29]. These procedures aim at arriving at this minimal set by constructing conflict sets of components to explain the symptoms. For instance, in the above, a conflict set could be arrived at, which indicates that one or more of $< \theta_1 \;\; u_1 \;\; u_2 >$ is in error. They do not appear to be applicable, at least directly, here because they assume that every occurrence of a particular fault in the candidate space will be the same: true or false. Here the estimated values of the elements in $[u_1 \, u_2]$ may be quite different to those in $[\theta_1 \, u_1 \, u_2]$.

2.6.10 The Solution Candidate Set

If the Principle of a Minimum Number of Errors were to be applied, strictly, then all candidates, on the lowest row possible, that explain the symptoms are admissible. However, other factors like one type of error being preferred to another come into play. As will be seen in Chapter 4, the method of determining suitable values for the elements in the candidate set may also affect this admissible set.

2.7 The Role of the Supervisor

The question then arises as to when is one particular hypothesis more appropriate than any of its rivals? One answer could be that it is appropriate if it is correct. Unfortunately the only guarantee that a particular hypothesised fault is correct is physically to examine the component involved. This is relatively straightforward if the fault is catastrophic, for instance if a transducer is failing to transmit, but not so easy if, for instance, the component has to be taken out of service for re-calibration or if the malfunction is intermittent; for instance it may be difficult to reproduce the precise conditions again. When should a plant operator accept a particular hypothesis and act accordingly? There could be political as well as technical ramifications. For instance, production may have to be reduced if a particular component is taken out of service. He may be prepared to examine an alternative, marginally less likely, hypothesis if this has less effect on plant operation. However his decision must weigh-up the possibility that failure to act could increase the risk of a plant shut-down. There could also be resource implications regarding the scheduling of maintenance staff. Certain hypotheses may not need investigating at all. For instance, a single physical inventory measurement error will have no effect on the overall account in NRTMA; an entry as to a possible weakness in the measurement system may be all that is required. If this hypothesis is one of a number of alternatives, should it be chosen in preference to any of the others? There may be a case for delaying making a decision by one or more periods to see how a fault develops or to change the operation of a particular plant unit to investigate its affect.

These considerations are beyond the scope of this thesis in that they do not pertain to fault diagnosis, *per se*, but to the adjudication process that may follow. Unfortunately it is difficult to separate the two tasks succinctly because the form the diagnosis takes will have a bearing on how it is used. It is therefore desirable to have some idea of the different types of assessment. I do not intend to digress too much into this because the information generated by the diagnostic system proposed here is of only one possible form, that of ordered lists. This derives from an assumption that if we accept that our knowledge of possible faults, of modelling errors and of the plant itself, is never complete, we must also accept that it is impossible to specify, in absolute terms, the probability of any individual error occurring. However each technique used to generate hypotheses should be able to order the list it outputs in some way. Such an ordering should be of benefit to the operator even if it cannot be viewed as being the absolute truth. Again one particular method of ordering is considered in Chapter 4.

However this is a moot point. The proposed theory would therefore not seem complete without, at least, providing a brief literature survey.

2.7.1 Adjudication

It must be evident from the above that the adjudicator must be utilitarian in nature. According to the traditional theory of utility [32], the approach should be to assign a utility to each hypothesis, to estimate the likelihood that each hypothesis will obtain, to calculate the utility of each act and to choose an act of maximum utility. The primary weakness with this is in the type and amount of information required. Even if it were possible to produce a fault dictionary with a utility assigned to each fault this can never be complete. There would probably be a lack of temporal factors; for instance the dictionary is unlikely to include the possibility of the maintenance engineer failing to carry out his duties properly.

We therefore leave the utilitarian aspect to the operator and concentrate on how to *infer* that one hypothesis is more likely than any other.

Inference is the process of argument where, on the basis of evidence, one or more hypotheses are proposed which are more probable than any rival. If it is unsuccessful then there is deemed to be insufficient evidence: either more evidence must be gathered or all hypotheses must be carried forward. The methods used in the inference process depend on the category of inference required [33]. An inference is a deductive one if all it does is draw out of the premises propositions already (albeit tacitly) contained in the premises taken together. So in deductive logic the truth of the premises makes certain the truth of the conclusion. In the inference from evidence to hypothesis, it is possible that the truth of the premises, ie the evidence, may only make certain to some degree the truth of the conclusion, ie the hypothesis. Hence it is natural to regard this type of inference as being similar but distinct from deduction and it is given the name of *induction.* Deduction is then regarded as being the limiting case of induction.

Deduction would be possible in fault diagnosis if perfect measurements could be recorded throughout a plant and compared with the perfect model. Alternatively an approximation to this may be obtained using signal processing and parameter identification. Induction becomes necessary when there is a lack of measurements and models and those that are available are uncertain. We therefore focus on induction.

There are two types of induction (34): ampliative and summative. Workers (eg 71) in expert systems generally use the term *induction* solely to refer to *'summative'* induction. Summative induction establishes a generalization on the basis of what are known to be all its instances, as when a railway inspector, passing down the whole train, establishes that every passenger on the train has a ticket. Induction is *'ampliative'* when it extrapolates beyond existing data. It is unlikely that all possible faults will be known so the process of inferring the existence of a particular fault on the basis of a set of indirect measurements is likely to be ampliative. Workers in expert systems use the term *uncertainty*, loosely, to refer to ampliative induction. For instance, 'The term "uncertainty" .... appears to be used whenever reasoning by strict logical implication is not considered possible' (72).

A number of reviews have been published into suitable techniques for ampliative induction (eg 72,73,74). Other comparisons can also be found in publications which describe particular expert systems (75). The techniques fall into two categories, numerical and non-numerical. These, in turn, divide into sub-categories where techniques based on probability (72), fuzzy sets and fuzzy measures (36) are examples of the former and methods of endorsements (73) and relevant variables (35) are examples of the latter.

The most obvious numerical technique is that of probability theory. The arguments concerning the appropriate role of probabilistic statements, of how they should be chosen and manipulated, have raged for centuries. Various articles [72,76] have been written to promote probability theory, others [33,34] to describe the attributes of six alternative theories of probability whilst others [72] to propose techniques for their implementation.

A number of non-numerical techniques have been proposed that set various levels of qualitative hurdle. The 'method of relevant variables' [35] fits naturally into systems that sequentially apply a standard set of increasingly stringent tests. The 'method of endorsements' [73] uses a ledger metaphor to represent evidence pro and con where the certainty of a hypothesis can be represented as its strongest endorsement.

2.8 The System Must Learn

[illegible] The operators learn by experience. A knowledge-base is therefore required which can [illegible] can [illegible] expand in time.

It is envisaged that, at least during the infancy of a plant, the system will alarm frequently. It is therefore proposed that the system should be invoked on every period so as to minimise the number of un-explained faults present at any one time. The data set would then be updated everytime a fault was diagnosed to the satisfaction of the operator. It is not intended that the original data set be overwritten unless the diagnosis is properly verified. Either a second data set should be created or the modification included as an update to the original. In either case the system would now restart and assume that the revised data set is now correct.

3. MAKING USE OF THE PRIMARY CHARACTERISTIC VARIABLES

3.1 Introduction

Control charts are used in process plants to provide a visual indication of a problem developing. Trends in the plots can not only provide information as to whether or not there is a fault but also as to its cause. The approach is qualitative rather than quantitative; the cognitive process is symptomatic because the operator attempts to correlate these trends with fault related patterns. Historically, plant control charts have had a disparate role in quality control [63]. This Chapter examines whether both roles may be incorporated into a model-based fault diagnosis system.

Considerable expertise exists in the application of control charts to processes with either known statistics or a reasonable flow of data [2,3,63], that is, to information rich processes. Less is available for processes which are information poor. This Chapter is devoted to providing a brief overview and proposing specific methods which have been found to be appropriate for information poor plants.

3.2 Traditional Plant Control

The objective of quality control is to check that the actual value of a quantity agrees with its expected value. This limits the number of control variables to those where a true expected value can be generated.

The traditional role of quality control [63] has been in fault detection and not in fault diagnosis. The approach is to apply statistical techniques to the time series obtained by measuring the process discrepancies at discrete intervals of time. Procedures [64] have been recommended if the relevant probability densities are not available. A number of test procedures are available to optimise the *power* to detect a fault against its *credibilty* (ie its false alarm rate) [2,65,66]. Quality control charts [64,67] provide graphical representations of test procedures applied to the time series. The simplest plot is that of the Shewart chart [68] suitable for uncorrelated time series of constant variance. Another common plot is that of the cum ulative sum. Both plots are against time.

Considerable effort has been expended in the development of techniques to improve both the power to detect and credibility when the time series is correlated in time. They usually assume that a statistical model is available and split naturally into 2 distinct categories, namely *estimators* and *detectors.* Estimators are solely concerned with producing new estimates of discrepancy values that have errors with reduced variances. On the other hand the objective of detectors is to manipulate the time series so that a more effective test procedure can be generated.

It should be appreciated that estimators only produce test statistics and it is necessary to apply a test procedure in order to obtain any results. This is chosen on the basis of its credibility and power to detect a particular error scenario. The reduced error variances achieved by the estimator will improve the performance of the test and so both categories have the same ultimate goal of increased detection probability.

As detectors are not concerned with producing a meaningful physical estimate it is possible that they will be more effective than their counterpart. However they have no role in diagnosis, as opposed to detection.

The main disadvantage of most estimators and detectors is that they require knowledge of the measurement errors involved. This may not be too problematic if the characteristic variables are measured directly but may be difficult if they are derived from a set of measurements. If it is not possible to produce a meaningful measurement model, then it is argued that, on the basis of this uncertainty, plant control should be viewed as being more qualitative than quantitative. This is more in line with that of supervisory control.

3.3 Viewing Control Charts Qualitatively

Plant control charts are used in supervisory control in two ways: to provide assurance that the control variables are behaving in a reasonable manner and to correlate the patterns observed in the charts with each other and with the various plant actions. If a fault develops the charts can be used as an aid in symptomatic search. The general principle is to develop rules to explain plant activity by relating the patterns on either a single or number of charts to possible fault scenarios. Although there is unlikely to be sufficient information to identify the fault uniquely, the charts provide a means of reducing the search space.

Considerable effort has been expended (3,26) into developing statistical techniques so that patterns may be attributed to possible fault scenarios in some optimal manner. In general, they adopt the same approach as for the detectors and estimators above and hence suffer from the same limitations.

This statistical approach is not the way the operator would infer from the charts. His rules would largely be heuristics and he would normally attempt to ignore any noise by visually filtering it out. That is, his view of noise would be restricted to any 'high frequency' fluctuations superimposed on the plots. The only statistical representation would then be deemed to be equivalent to that obtained when the same measurement is taken a number of times and would be defined as being its '*random error*'.

It is this qualitative approach that is considered here.

One of the issues is how often should the process of pattern recognition be repeated. Clearly the longer the delay, the more the information and the better the discrimination. Conversely, the longer the delay, the less timely the detection and diagnosis. Diagnosis may prove difficult if the state of the plant has changed, significantly, from when a particular incident was initiated. For instance, the chances of finding an intermittent fault are reduced. In addition operators are a great deal more aware of the current situation than that which took place some time ago.

The same approach can be applied to quality control charts of any physically meaningful variable, for instance a control variable or its cumulative sum. Although the odd result is general, for instance, a bias may be estimated by visual inspection of the cumulative sum, most results are application specific and are best examined by example.

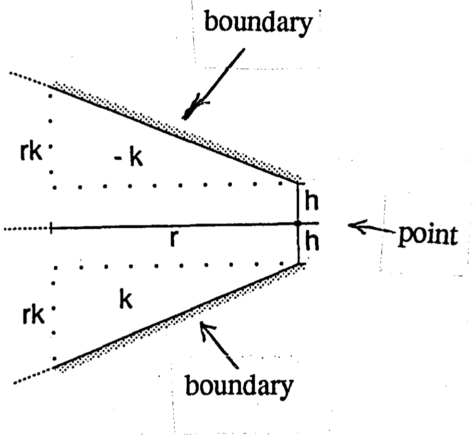
3.4 Automatic Extraction of Information From Control Charts

The primary role of the control charts is to detect the occurrence of a fault and having done so, to output two lists: a list of discrepancies and a list of assertions which point to possible classes of faults that could account for the patterns observed. It would be unusual for the charts to identify a fault uniquely; their role is to focus attention. Two issues complicate the recognition process, that of noise and of multiple faults. Both can largely be overcome by adding any fault, that is remotely likely, to the list. Care must then be taken to ensure that the most likely are considered first.

The pattern recognition exercise is largely one of detection and not of estimation because it seeks to determine the existence of various patterns. Hence it involves the application of tests. These need not be founded on rigorous statistical arguments.

3.4.1 The CUSUM Plot

It is common practice to use plots of the cumulative sum of particular discrepancies to examine drift. The approach has two main weaknesses. Firstly, the sum can build up with time so that special provision must be made when using it to detect events that are local in time. Secondly, although informative visually, this information is difficult to extract automatically primarily because changes from period to period are of interest and not absolute changes which are relative to some datum. These problems are usually overcome by applying the CUSUM test [69]. In its graphical form, the technique involves centring the V-mask



r - period

h, k - user specified parameters

on each point of a plot of the normalized cumulative sum,

$$S_n = \sum_{i=1}^{n} \chi_i$$

where χ_i is the normalised value of the discrepancy on period i, and alarming if ~~it~~ the plot crosses the boundary shown.

This is equivalent to the tests

$$\text{alarm if } S_n - S_{n-r} \geqslant h + rk$$

$$\text{alarm if } S_n - S_{n-r} \leqslant -h - rk$$

for any r = 1, ..., n.

The importance of the CUSUM test to quality control in general is perhaps reflected in the sheer number of papers (see, for instance 70) that present methods for choosing the two parameters h, k. This in turn is an indication that the choice may be a complicated one. Fortunately approximate values should be sufficient here because the approach does not require the test to achieve specific leve' of power and credibility.

The V-mask has great virtue, deriving from its easy visual interpretation. However, a numerical version is required for computer applications. It can be shown (65) that it is equivalent to the following algorithm.

$$\text{Alarm if} \quad \tau_n^+ \geqslant h \quad \text{or} \quad \tau_n^- \leqslant h$$

where

$$\tau_n^+ = \max\{\tau_{n-1}^+ + \chi_n - k, 0\}$$

$$\tau_n^- = \min\{\tau_{n-1}^- + \chi_n + k, 0\}$$

At first glance the test does not appear to be applicable for information poor plants because of the need to know the standard deviation to normalise the individual discrepancies. However if the posture of Section 2.4 is adopted, the control variables will be restricted to those variables which are either measured directly or obtained by performing some algebraic manipulation of a set of measurements. In either case it should be possible to obtain a reasonable estimate of random error.

3.4.2 *Visual Estimation of Drift*

The main role of the CUSUM plot is in detection because the process of normalizing the control variable visually distorts the plot. An un-normalized plot is more suitable for looking for trends such as biases. Consider the un-normalized plot shown in Figure 6 below and suppose that the process is deemed to be in *control* if the points lie, approximately, along a horizontal line.

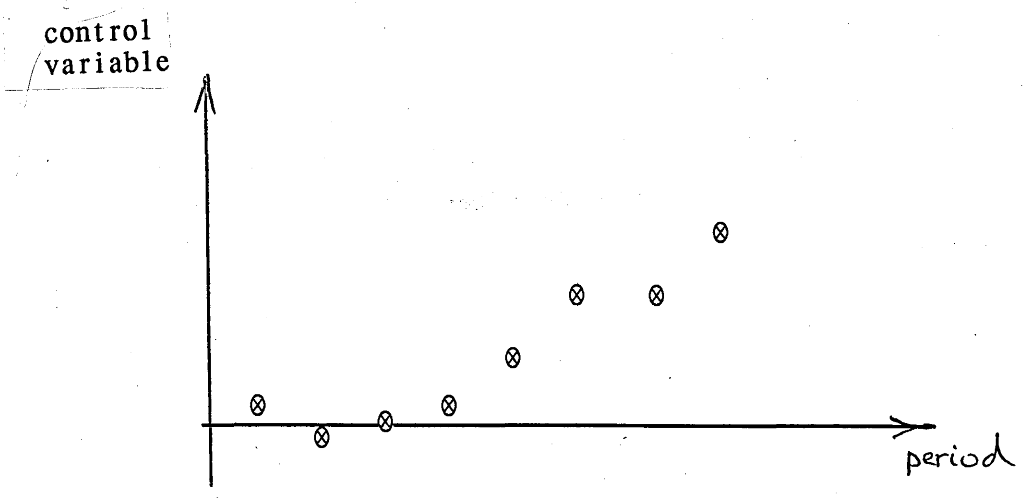


Figure 6: Un-normalized Cumulative Plot

Two main issues arise when trying to describe the drift. Did the problem causing the drift start at the second, third or fourth point? What is the underlying rate at which the drift occurred?

By direct analogy with the Cusum test, one possible approach could be to perform all linear regressions with a minimum of three points and then to choose the one with the greatest gradient. This would give a start point, final point, gradient, m, and variance, σ^2 (37):

$$\sigma^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - mx_i - c)^2$$

A decision could then be made to extend this *trend* to previous points by applying the test:

$$\left|(x_1 - x_0)\right| < (m-6\sigma)$$

where x_0 is the new start point hypothesised.

4. CANDIDATE APPRAISAL

4.1 Introduction

The notion of a candidate space was introduced in Chapter 2. However the chapter avoided proposing a method for calculating the values a particular candidate would need to take to explain the symptoms observed. This was because it was devoted to developing a universal theory. This situation is rectified here.

One possible approach is described which, again, has leanings towards the common sense rather than the scientific. The approach makes various assumptions which influence the way the candidate space is searched.

4.2 Intuitive Overview

We imagine ourselves once more as the diagnostician faced with the problem of determining the values a candidate set of errors should take to explain symptoms he views as being at various levels of significance. These errors may be dispersed throughout the plant, measurement system and model. He would be aware that this candidate would be one of many, a large proportion of which would be neither credible nor sensible. One approach he might adopt could be to try to determine *reasonable* values for one or more of the candidate elements which would explain the most significant symptom. Having done this, he would then move onto the next most significant symptom and so on. He would note that credible deviations in certain elements had little affect on any of the symptoms and would discard any candidate containing them.

The approach described here has been developed to reflect this point of view but with certain modifications to accommodate some of its shortcomings. Firstly, all elements of the candidate are assessed simultaneously to avoid the conflict that could arise if one or more elements needed to explain the first symptom are also needed to explain the second. Secondly, symptoms at a relatively low level of significance (for instance, 1) are included because of the possibility of two errors partially cancelling each other out. Thirdly, for computational efficiency, a single model is derived to relate deviations in all possible errors contained in the candidate space to the symptoms observed. Reduced versions of this model are then formed to determine possible values for a particular candidate.

4.3 Developing a Model for Value Estimation

Recall from Section 2.6.3 that

$$\underline{\tilde{y}}_k = \underline{\hat{y}}_{k-1} + \underline{f}_k(\underline{\hat{x}}_{k-1}, \underline{\hat{\theta}}_k, k)$$

and consider the case of path errors first.

If a *symptom*, $\{\underline{\hat{y}}-\underline{\tilde{y}}\}_i$, were to be attributed to a distortion in a single variable or parameter, j, contained in $\underline{\theta}_k$, then its size could be estimated deterministically by viewing estimates of all the other elements as being perfect so that

$$\{\underline{\hat{y}}_k\}_i = \{\underline{\hat{y}}_{k-1} + \underline{f}_k(\underline{\hat{x}}_k, \underline{\hat{\theta}}_k + \{\underline{\Delta\theta}_k\}_j . \underline{e}_j, k)\}_i$$

where $\underline{e}_j$ is the j^{th} natural basis vector.

That is, the required distortion could be obtained by determining that change in $\{\underline{\Delta\theta}_k\}_j$ needed to enable the simulation to predict the i^{th} measurement precisely. In practical terms, this could be achieved by applying any one of a number of standard numerical algorithms to the simulation. This calculation would simultaneously estimate the effect that the required distortion would have on the other *symptoms*. Let the resultant perturbation in all the *symptoms* be denoted by $\underline{\Delta y}_k$ where

$$\underline{\Delta y}_k \Big|_j = \underline{\hat{y}}_{k-1} + \underline{f}_k(\underline{\hat{x}}_k, \underline{\hat{\theta}}_k + \{\underline{\Delta\theta}_k\}_j . \underline{e}_j, k) - \underline{\tilde{y}}_k$$

A Jacobian-like matrix may now be obtained by forming a list, p_{list}, of all elements of Θ that are thought to be suspect and assembling the vectors obtained by solving for the most significant *symptom*, taking one element at a time, to form

$$J_1 \;=\; \left[\; \left.\frac{\Delta\underline{y}_k}{\{\Delta\underline{\theta}_k\}}\right|_{\ell_1} \;\vdots\; \left.\frac{\Delta\underline{y}_k}{\{\Delta\underline{\theta}_k\}}\right|_{\ell_2} \;\vdots\; \cdots\cdots\cdots\; \right]$$

where subscripts ℓ_i refer to the ith element of p_{list} and

$$J_1\,\underline{\Delta\theta_k} \;\triangleq\; \hat{\underline{y}}_{k-1} + \underline{f}_k(\hat{\underline{x}}_{k-1}\,,\ \underline{\theta}_k,\ k) - \tilde{\underline{y}}_k$$

It may not be possible to obtain a solution for every element; certain elements may simply not affect the symptom or may have only a limited affect, too small to eliminate it. However these elements may affect the other symptoms. The solution process is therefore re-focused onto the next most significant symptom and any relevant vectors obtained and so on until either all the columns of J_1 have been filled or all the symptoms, that are deemed to be significant, are exhausted.

The implication of the latter is that there could be certain elements which can only explain a small proportion of the discrepancy between the measurements and the simulation. In this case, an alternative approach must be taken to incorporate their effects in the model. These variations can be represented by considering infinitesimally small perturbations around the operating point. That is by adopting the usual procedure for determining a Jacobian. Having obtained such a Jacobian, J_2, the vector $\underline{\Delta\theta_k}$ can be re-arranged into two parts, $\underline{\Delta\theta_{k_1}}$ and $\underline{\Delta\theta_{k_2}}$: the first part containing all the successful elements whilst the second part containing those where no solution could be obtained. A composite Jacobian-like matrix J can now be defined so that it operates on the entire composite vector. That is

$$J\underline{\Delta\theta_k} \;=\; \begin{bmatrix} J_1 & : & J_2 \end{bmatrix} \begin{bmatrix} \underline{\Delta\theta_{k_1}} \\ \underline{\Delta\theta_{k_2}} \end{bmatrix}$$

If the covariance of $\underline{\Delta\theta}_k$, P_k, is assumed to be at least block diagonal, the above has covariance,

$$JP_k\, J' = \begin{bmatrix} J_1 & : & J_2 \end{bmatrix} \left[\begin{array}{c|c} P_{k1} & 0 \\ \hline 0 & P_{k2} \end{array}\right] \left[\begin{array}{c} J_1' \\ \hline J_2' \end{array}\right]$$

$$= J_1\, P_{k1}\, J_1' + J_2\, P_{k2}\, J_2'$$

On the basis that the previous period is error free and that inaccuracies in the model structure can be viewed as being path errors, and hence affecting $\underline{\theta}_k$, and non-path errors, and hence affecting $\underline{n}_k$, function $\underline{f}$ also represents the true change in sub-system output:

$$\underline{y}_k = \underline{y}_{k-1} + \underline{f}_k(\hat{\underline{x}}_{k-1}, \underline{\theta}_k, k) + \underline{n}_k$$

Then

$$\underline{y}_k = \underline{y}_{k-1} + \underline{f}_k(\hat{\underline{x}}_{k-1}, \hat{\underline{\theta}}_k + \underline{\Delta\theta}_k, k) + \underline{n}_k$$

and

$$\hat{\underline{y}}_k - \tilde{\underline{y}}_k \triangleq \underline{w}_k + \underline{v}_k + \underline{n}_k - \underline{w}_{k-1} - \underline{v}_{k-1} + J\,\underline{\Delta\theta}_k \quad .$$

For convenience, combine biases $\underline{n}_k$ and $\underline{w}_k$ into a single vector $\underline{u}_k$. This vector therefore represents biases in the measurement system, non-path faults and the effect of inaccuracies in the model structure. Considerable uncertainty surrounds this vector. If this were not the case, there would be little justification for it to be represented explicitly because the same information could be accommodated in the simulation. It is therefore proposed to handle this uncertainty in a similar way to the above by viewing the vector as being a random process probably correlated in time where

$$E\{\underline{\Delta u}_k\} = E\{\underline{u}_k\} = 0$$

and

$$\mathrm{cov}\{\underline{\Delta u}_k, \underline{\Delta u}_k\} = P_{uk}$$

A composite vector $[\underline{\Delta\theta}_k' | \underline{\Delta u}_k']'$ can then be formed with covariance

$$\left[\begin{array}{c|c} P_k & 0 \\ \hline 0 & P_{uk} \end{array}\right]$$

and the model formulation can be revised to

$$\hat{\underline{y}}_k - \tilde{\underline{y}}_k \triangleq \underline{v}_k - \underline{w}_{k-1} - \underline{v}_{k-1} + \left[\begin{array}{c|c} J & 0 \\ \hline 0 & I \end{array}\right] \left[\begin{array}{c} \underline{\Delta\theta}_k \\ \\ \underline{\Delta u}_k \end{array}\right]$$

Two aspects of the model still require clarification: the alignment of two or more vectors making it impossible to discriminate between them and the quantification of uncertainty, that is the specification of covariance matrices P_k and P_{uk}. The former is left until Section 4.6 where it is related to the actual estimation process whilst the latter is considered here.

The diagnostician can only specify a P_k if he assumes that there are not any errors that are significant and he may be reluctant to specify a P_{uk} at all. It is therefore proposed to resort to the pragmatic approach of assuming that a symptom can be explained by a single error. It therefore seems reasonable to hypothesise the standard deviation, ie $\sqrt{\{P_{uk}\}_{ii}}$, as being equal to the error in the appropriate *symptom.* Turning to the specification of P_k. This has already been specified in Section 2.6.4 for errors that are insignificant but not otherwise, that is for starred elements. A similar approach as for P_{uk} is therefore proposed.

Finally, the model most appropriate for estimating the values for a particular candidate is formed by simply eliminating all columns and rows that do not relate to either the elements of the candidate or the symptoms that would be affected by varying those elements.

4.4 Estimation

The mean and variance of the marginal distribution of $\underline{\Delta\theta}_k$ for a particular instance $(\hat{\underline{y}}_k - \tilde{\underline{y}}_k)$ may now be determined by applying the linear transform

$$\begin{bmatrix} I & -P_k\,J'(R_{k-1} + R_k + JP_kJ') \\ 0 & I \end{bmatrix}$$

to the composite vector, $\underset{\sim}{c}$:

$$\underset{\sim}{c} = \begin{bmatrix} \underline{\Delta\theta}_k \\ (\hat{\underline{y}}_k - \tilde{\underline{y}}_k) \end{bmatrix}$$

where

$$E\{\underset{\sim}{c}\} = \begin{bmatrix} E\{\underline{\Delta\theta}_k\} \\ (\ JE\{\underline{\Delta\theta}_k\} - \underline{w}_{k-1}) \end{bmatrix}$$

and

$$\mathrm{cov}\{\underline{c},\ \underline{c}'\} = \begin{bmatrix} P_k & P_kJ' \\ JP_k & R_{k-1} + R_k + JP_kJ' \end{bmatrix}$$

This block diagonalises cov{c, c'} and produces a marginal distribution with mean

$$E\{\underline{\Delta\theta}_k - P_kJ'(R_{k-1} + R_k + JP_kJ')^{-1}\ (\hat{\underline{y}}_k - \tilde{\underline{y}}_k)\}$$

$$= E\{\underline{\Delta\theta}_k\} - P_kJ'(R_{k-1} + R_k + JP_kJ')^{-1}\ (\underline{w}_{k-1} - E\{J\underline{\Delta\theta}_k\})$$

If $\underset{\sim}{c}$ is assumed to be jointly normally distributed then it can be shown that $\underline{\Delta\theta}_k - P_kJ'\ (R_{k-1} + R_k + JP_kJ')^{-1}(\hat{y}_k - \tilde{y}_k)$ and $(\hat{y}_k - \tilde{y}_k)$ will be independent.

It follows that the marginal distribution will be the same as the conditional distribution for a given value of $\hat{\underline{y}}_k - \tilde{\underline{y}}_k$ so that

$$E\{\underline{\Delta\theta}_k \,|\, (\hat{\underline{y}}_k - \tilde{\underline{y}}_k)\} = E\{\underline{\Delta\theta}_k\} + P_k J'(R_{k-1} + R_k + JP_kJ')^{-1} (\hat{\underline{y}}_k - \tilde{\underline{y}}_k - \underline{w}_{k-1} + E\{J\underline{\Delta\theta}_k\})$$

with covariance, $Q_k = P_k - P_k J'(R_{k-1} + R_k + JP_kJ')^{-1} JP_k$

It can be shown that this is the minimum variance unbiased estimate.

Hence if the prior estimate of $\underline{\theta}_k$ is taken to be $\hat{\underline{\theta}}_k$

$$\text{ie} \quad E\{\underline{\Delta\theta}_k\} = 0$$

and if non-path errors pertaining to the previous period have already been deemed to have been resolved (Section 2.6.3), then the posteriori estimate is given by

$$E\{\underline{\theta}_k \,|\, (\hat{\underline{y}}_k - \tilde{\underline{y}}_k)\} = \hat{\underline{\theta}}_k + P_k J'(R_{k-1} + R_k + JP_kJ')^{-1} [\hat{\underline{y}}_k - \tilde{\underline{y}}_k]$$

The above may be partitioned to produce separate estimates for $\underline{\Delta\theta}_{k_1}$ and $\underline{\Delta\theta}_{k_2}$. The estimation process now becomes,

$$\text{estimate} = E\{\underline{\Delta\theta}_k\} + P_{k_1} J_1' (R_{k-1} + R_k + JP_kJ')^{-1} (\hat{\underline{y}}_k - \tilde{\underline{y}}_k - \underline{w}_{k-1} + E\{J\underline{\Delta\theta}_k\})$$

with covariance,

$$Q_k = P_{k_1} - P_{k_1} J_1' (R_{k-1} + R_k + JP_kJ')^{-1} J_1 P_{k_1}$$

since only those elements of $\underline{\Delta\theta}_{k_1}$ are of interest.

4.5 Interpretation

The estimator can be applied to a number of different sets of candidate errors. The results obtained must be assessed, firstly to determine whether any particular estimate is *credible* and secondly to order them in some way.

The estimation process is deemed to be *credible* if

1. each individual estimate satisfies its *a priori* variance. That is,

$$\left[\, E\{\underline{\theta}_k \,\middle|\, (\hat{\underline{y}}_k - \tilde{\underline{y}}_k)\} \,\right]_i < n\sigma_i$$

where n is some subjective factor and σ_i relates to P_k or P_{uk} ;

2. the estimates reduce the *symptoms* to some specified level of significance when they are included in the simulation ie

$$\hat{\underline{y}}_k - \hat{\underline{y}}_{k-1} - \underline{f}\left[\hat{\underline{x}}_k,\ E\{\underline{\theta}_k \,\middle|\, (\hat{\underline{y}}_k - \tilde{\underline{y}}_k)\}, k\right]$$

replaces $(\tilde{\underline{y}}_k - \hat{\underline{y}}_k)$ in the test of Section 2.6.4.

4.5.1 *Ordering the Various Estimates*

Ordering must be viewed as being rather subjective because of the way the covariances P_k and P_{uk} have been specified. The outline order is largely dependent on whether or not a particular combination has necessitated either the revision of P_k to accommodate a significant error and/or the inclusion of P_{uk}. The diagnostician would probably prefer not to consider such aspects first and would therefore place any other combinations at the top of the list. This would be followed by combinations containing a single revision or element of P_{uk} and so on.

A method is proposed here which orders each part of the list by ordering the log likelihood function [38,39] evaluated for each candidate.

If we assume that $\underline{\Delta\theta}_{k_1}$ is jointly normally distributed with zero mean and covariance P_{k_1} then its likelihood function $L(\underline{\Delta\theta}_{k_1}, P_{k_1})$ is given by

$$L(\underline{\Delta\theta}_{k_1}, P_{k_1}) = \left|2\pi P_{k_1}\right|^{\frac{1}{2}} \exp\{-\tfrac{1}{2}(\underline{\Delta\theta}'_{k_1}\; P_{k_1}^{-1}\,\underline{\Delta\theta}_{k_1})\}$$

Similarly the log likelihood function $\ell(\underline{\Delta\theta}_{k_1}, P_{k_1})$ is given by

$$\ell(\underline{\Delta\theta}_{k_1}, P_{k_1}) = -\tfrac{1}{2}\log\{\left|2\pi P_{k_1}\right| - \tfrac{1}{2}(\underline{\Delta\theta}'_{k_1}\; P_{k_1}^{-1}\,\underline{\Delta\theta}_{k_1})\}$$

The term $|2\pi P_{k_1}|$ may be omitted from the above because it is common to all candidates and will therefore not affect the result. Hence the ordering process reduces to evaluating $(\underline{\Delta\theta}'_{k_1}\; P_{k_1}^{-1}\; \underline{\Delta\theta}_{k_1})$ for each candidate and ordering with the minimum at the top of the list.

4.6 Orthogonality and Reducing the Search Space

It is likely that variations in more than one element will be seen to affect the discrepancies in a similar way. This is because one sub-system can only affect other sub-systems through the interconnections between them. If two of the columns of J are aligned in the sense that

$$(\text{column } i) \;\triangleq\; \alpha\;(\text{column } j)$$

then the estimation process will have difficulty in discriminating between the i^th^ and j^th^ elements. It therefore seems sensible to reduce the number of elements in Θ by eliminating those aligned columns, and hence elements, that are less likely to affect the *other symptoms*. If an element with a vector which is aligned, is identified as being suspect then the other elements must also be suspect. Other factors must then be taken into account in deciding between them.

In practice this alignment can be tested for by applying the inner product property that two vectors are aligned if

$$< \text{column i}, \text{column j} > \; = \; \left\| \text{column i} \right\| \left\| \text{column j} \right\|$$

Hence a test can be applied of the form

$$\left| \; 1 - \frac{< \text{column i}, \text{column j} >}{\left\| \text{column i} \right\| \left\| \text{column j} \right\|} \; \right| < \epsilon$$

where ϵ is some specified tolerance.

Having produced a set of m aligned vectors, the most significant vector $\underline{j}$, is that associated with element i:

$$\max \left[\{\underline{j}\}_l \; \surd\{P_k\}_{ii} \right] \; , \quad \forall l \quad l: \; \{\underline{j}\}_l \neq 0$$

If $\underline{\Delta\theta_k}$ is ordered such that the aligned vectors relate to its first elements $\epsilon \; \Phi \subset \Theta$:

$$\underline{\Delta\theta_k}' \; = [\; \underline{\Delta\varphi_k}' \; \ldots \;]$$

then the aligned columns of J_1 can be eliminated by post multiplying J_1 by the operator M:

$$M = \left[\begin{array}{c|c} \underline{e}_1 & 0 \\ \hline \underline{0} & I_{n-m} \end{array} \right]$$

where $\underline{e}_1 \; \epsilon \; \mathbf{R}^m$ and is the first natural basis vector.

The problem with this approach is that the estimator does not take the eliminated elements into account. A variation may be worthy of consideration. If the principle of a minimum number of errors is allowed to be contradicted by combining similar elements using the operator L:

$$L = \left[\begin{array}{c|c} \underline{\alpha}' & \underline{0}' \\ \hline 0 & I_{n-m} \end{array}\right]$$

where $\underline{\alpha} \in \mathbf{R}^m$

and $$\{\underline{\alpha}\}_i = \frac{<\text{column } i,\ \text{column } i>}{\sqrt{<\text{column } 1,\ \text{column } 1>}}$$

then this will revise and shorten the vector by forming a composite element $\underline{\alpha}'\underline{\Delta\varphi_k}$ and giving it a revised covariance matrix LP_kL'.

The composite element will then be estimated as

$$E\,\{\underline{\Delta\theta_k} \mid (\hat{y} - \tilde{y})\}_1 = \underline{\alpha}'\ \underline{\Delta\varphi_k}$$

$$= \sum |\alpha_i|\, \mathrm{sign}(\alpha_i) \sqrt{\{P_k\}_{ii}} \left[\frac{\{\underline{\Delta\varphi_k}\}_i}{\sqrt{\{P_k\}_{ii}}} \right]$$

If each element of Φ is then assumed to contribute an *equal* amount in a sense that

$$\mathrm{sign}(\alpha_i)\ \frac{\{\underline{\Delta\varphi_K}\}_{ii}}{\sqrt{\{P_k\}_{ii}}} = k\ \forall i$$

then

$$\{\Delta\varphi_k\}_i = \frac{\sqrt{\{P_k\}_{ii}}\ \ \mathrm{sign}\ (\alpha_i)}{\sum |\alpha_i|\ \sqrt{\{P_k\}_{ii}}}\ .\ E\,\{\underline{\Delta\theta_k} \mid (\hat{y} - \tilde{y})\}_1$$

This assumption is obviously unlikely. The diagnostician would need to refer to additional information before he could identify a preference for a particular distribution.

4.7 Implications for Searching the Candidate Space

A considerable proportion of the computational effort needed to search the candidate space will be devoted to performing simulations because the regression exercise involves the manipulation of a relatively small number of matrices of low dimension. With large candidate spaces anticipated, it is important that the number of simulations be kept to a minimum. If the lattice defining the candidate space is viewed as being a number of sub-lattices, with each sub-lattice branching from a different node at the top of the lattice, then the same regression model (Section 4.3) can be used to estimate all candidates in any particular sub-lattice. It therefore seems sensible to form each regression model only once, especially since the number of simulations needed to form matrix J is typically 2 or 3 times the number of its columns.

Although this need not affect the method of search, searching the lattice, one sub-lattice at a time, would reduce computer memory requirements. Care should then be taken to ensure that individual candidates are not evaluated more than once.

Before describing how the method of candidate appraisal and search can be implemented, we digress a little to introduce the only application included in this thesis: that of near-real time materials accountancy.

5. APPLICATION TO NEAR REAL TIME MATERIALS ACCOUNTANCY: SYSTEM DEFINITION

5.1 Introduction

A model-based fault diagnosis system has been developed with the sole purpose of experimenting with the ideas described previously. A conscious decision was taken at the outset to focus on only one application, that of near real time materials accountancy (NRTMA) and in particular, on its role in nuclear fuel reprocessing. This was largely chosen because of the author's previous experience [46,47]. The application has the following important features:

1. the plants are inherently non-linear;

2. the boundaries around the plants are well-defined with the transfers across them being closely monitored;

3. materials accountancy is to do with ensuring that the measured or estimated flow of material through a plant *balances*; a *fault* is therefore deemed to be anything that upsets this balance;

4. the plant operator and materials accountant have very different objectives with the former largely dictating the type, quality and frequency, of the data collected;

5. the material of greatest interest is plutonium; failure to achieve a reasonable plutonium *balance* has political implications;

6. a significant proportion of the data needed to form an account is derived from chemical analyses performed off-line in laboratories. It may be some time before these results become available. Hence the term *near real time.* Obviously this also has a considerable bearing on the frequency at which the account can be taken.

This Chapter outlines the system developed; an initial assessment of its performance is given in the next Chapter. There are two main problems with developing even a pilot system in a university environment: evolution and testing. As has already been argued in Section 2.8, the system should evolve by interacting with the plant; experience or knowledge being accumulated incrementally. The system described here is therefore skeletal, its rules and simulated faults are merely representative.

5.2 Introduction to NRTMA

5.2.1 NRTMA Described

Nuclear materials safeguards[48] are the steps taken by the nuclear community to ensure the security of nuclear materials. The managers of plants handling nuclear material, national bodies overseeing the activities of such plants, and international agencies who are charged with implementing various international treaties all have an interest in safeguarding the use of nuclear materials. One of the main ways that this is achieved is through the application of material accountancy.

Nuclear materials accountancy is based on the following structure. The plant is divided into units called *materials balance areas*, which are used as a basis for balancing all transfers of nuclear material. The plant is usually operated continuously for 2 months to a year's duration, at the end of which the plant is completely cleaned out and a physical inventory is taken. This operational cycle is known as a *campaign*. A balance is now obtained and the *material unaccounted for*, denoted by *MUF*, derived on the basis of

$$\text{MUF} = \begin{array}{c}\text{Total}\\ \text{Net Transfer}\end{array} - \begin{array}{c}\text{Change in}\\ \text{Physical Inventory}\end{array}$$

This quantity should be zero if all the estimates are error free.

This procedure suffers from a lack of timeliness. The question therefore arises as to whether these balances could be formed more frequently by measuring the physical inventory *in situ*. This approach is known as *near real time materials accountancy (NRTMA)*.

There are essentially two distinct application areas, that of fuel fabrication and of fuel reprocessing. Primarily, the former involves powders and discrete solid components and the latter both solids and liquids. Research into NRTMA has largely been attracted to reprocessing because of inherent difficulties with measuring the inaccessible inventories and because of the benefits of early detection of 'faults'.

In the past near real time materials accountancy as applied to fuel reprocessing plants has taken two forms, that where material balances are made at relatively large intervals of time (eg every day or every batch)(49-51) and that where measurements are made almost continuously(52). The former has been fully implemented on operational plants whereas the latter, being both capital and resource intensive, has been tried experimentally.

5.2.2 *Statistical Approach*

A number of statistical techniques have been developed to detect whether there is a significant MUF over one or more balance periods(for instance 47,53,54) and a number of reviews(55-57) have been published. The approach is usually along the following lines.

Consider taking a balance at the end of period k, then

$$\mathrm{MUF}_k = U_k - (I_k - I_{k-1})$$

where I_k is the total physical inventory at the end of balance period k, and U_k is the net input during period k.

If it is assumed that the estimates of U_k, I_k and I_{k-1} are corrupted by random errors which are independent of each other then MUF_k can be viewed as being a random variable with variance given by

$$\sigma_k^2 = \mathrm{var}\ \{\mathrm{MUF}_k\} = V_{k-1}^2 + W_k^2 + V_k^2$$

where V_k^2 and W_k^2 are the measurement error variances of I_k and U_k.

It can easily be seen that

$$\sigma^2_{k+1} \;=\; \mathrm{var}\;\{MUF_{k+1}\} \;=\; V^2_k + W^2_{k+1} + V^2_{k+1}$$

The occurrence of $V_k{}^2$ in both these expressions causes the random variables MUF_k and MUF_{k+1} to be correlated. As

$$MUK_k + MUF_{k+1} \;=\; I_{k-1} + U_k + U_{k+1} - I_{k+1}$$

the result

$$\mathrm{var}\;\{MUF_k + MUF_{k+1}\} = V^2_{k-1} + W^2_k + W^2_{k+1} + V^2_{k+1}$$

follows immediately. Using the formula

$$\mathrm{var}\;\{x_k + x_{k+1}\} \;=\; \mathrm{var}\;\{x_k\} + 2\;\mathrm{cov}\;\{x_k,\; x_{k+1}\} + \mathrm{var}\;\{x_{k+1}\}$$

the covariance between MUF_k and MUF_{k+1} can be calculated,

$$\mathrm{cov}\;\{MUF_k,\; MUF_{k+1}\} \;=\; -\,V^2_k$$

Hence the covariance matrix has non-zero terms in the three leading diagonals and is therefore termed tri-diagonal.

The correlation between x_k and x_{k+1}, $\mathrm{cor}\{x_k,x_{k+1}\}$ is related to covariance by the definition

$$\mathrm{cor}\{x_k,x_{k+1}\} \equiv \frac{\mathrm{cov}\;\{x_k,\; x_{k+1}\}}{\sqrt{\mathrm{var}\;\{x_k\}\;.\;\mathrm{var}\;\{x_{k+1}\}}}$$

Hence, denoting correlation by ρ_k, it is clear that

$$\rho_k = \mathrm{cor}\;\{MUF_k,\; MUF_{k+1}\} \;=\; \frac{-\,V^2_k}{\sigma_k\;\sigma_{k+1}}$$

For the simplest case of all, (when $V_i = V$, $W_i = W$ for all i) the ρ_i will be constant with

$$\rho_i = \rho = \frac{-\,V^2}{W^2 + 2V^2} \qquad \forall i$$

Under these conditions it is easy to see that ρ: $-0.5 \leqslant \rho \leqslant 0$ depending on W^2 and V^2. For an inventory dominated plant (ie $V^2 >> W^2$) $\rho \rightarrow -0.5$, whereas for a net transfer dominated plant (ie $W^2 >> V^2$) $\rho \rightarrow 0$.

The MUF series and covariance matrix now form the input to a detector or estimator with the aim of determining whether or not a *diversion* has occurred. The IAEA (48) envisages two patterns of diversion, *abrupt* and *protracted.* The former is diversion of the significant quantity, 8 kg over a few days, while for the latter, the period would cover a large part of a campaign, perhaps extending over more than one campaign.

It has long been recognized (58) that this *naive* approach is flawed. In theory the measurement model should discriminate between systematic errors, so-called non-measurement errors and so-called random errors. A description of these errors in the context of nuclear materials accounting is given by Speed *et al* (56):

> 'Systematic errors can arise through a wide range of reasons such as plugged probes, solid buildup in tanks, miscalibration of measurement devices and so on, whilst non-measurement errors may include errors due to operators misreading, mistranscribing or miscalculating; and random errors are presumably the unavoidable errors that are left over after all other possible explanations have been exhausted.'

They point out that the usual approach to accommodating these non-random errors is simply to add an extra component of error, argue that this is far from satisfactory and conclude by specifying a number of extremely stringent pre-conditions needed to ensure the applicability of a statistical approach. In addition, diversions and estimation errors are not the only sources of deviation of MUF values from zero. A host of instrument and human errors, such as miscalibration of measuring devices, can produce effects which may persist over several material balance periods and may closely resemble the effects of diversions. Accountancy procedures may also have difficulty determining physical inventories of plant components involved in non-routine operations.

5.2.3 Diagnosis in NRTMA

Diagnosis has largely been neglected in NRTMA research because NRTMA has largely been viewed as a safeguards tool to detect and not as a tool to explain the alarms observed. This is presumably because the number of alarms are thought to be so few as not to warrant much *a priori* research and because there is little operational data available to highlight a need.

This neglect is not justified. The statistical approach is currently not practicable because of a lack of realistic measurement models and possibly reliable data collection procedures. The net effect is either that large errors must be hypothesised thus reducing the power to detect certain diversion scenarios or a large number of false alarms must be diagnosed. This is apparent from the limited operational experience that has been published. Jones *et al* (49) have written about their experience in performing NRTMA on a real plant. They describe some of the diagnostic procedures used to justify the data collected even when relatively large errors are hypothesised and argue that NRTMA is a contributor to instrumentation quality control.

This thesis examines a diagnostic system where only aspects of the measurement model that are truly known are incorporated in the test procedure. This results in smaller variances and a resultant higher false alarm rate. The system then improves credibility by eliminating most of these false alarms without recourse to operator intervention. This in turn will enable a case history of non-random errors to be built up thus improving the capabilities of the statistical tests.

5.3 Overview of Proposed Knowledge-Based System

5.3.1 Its Structure

The proposed system has the following outline:

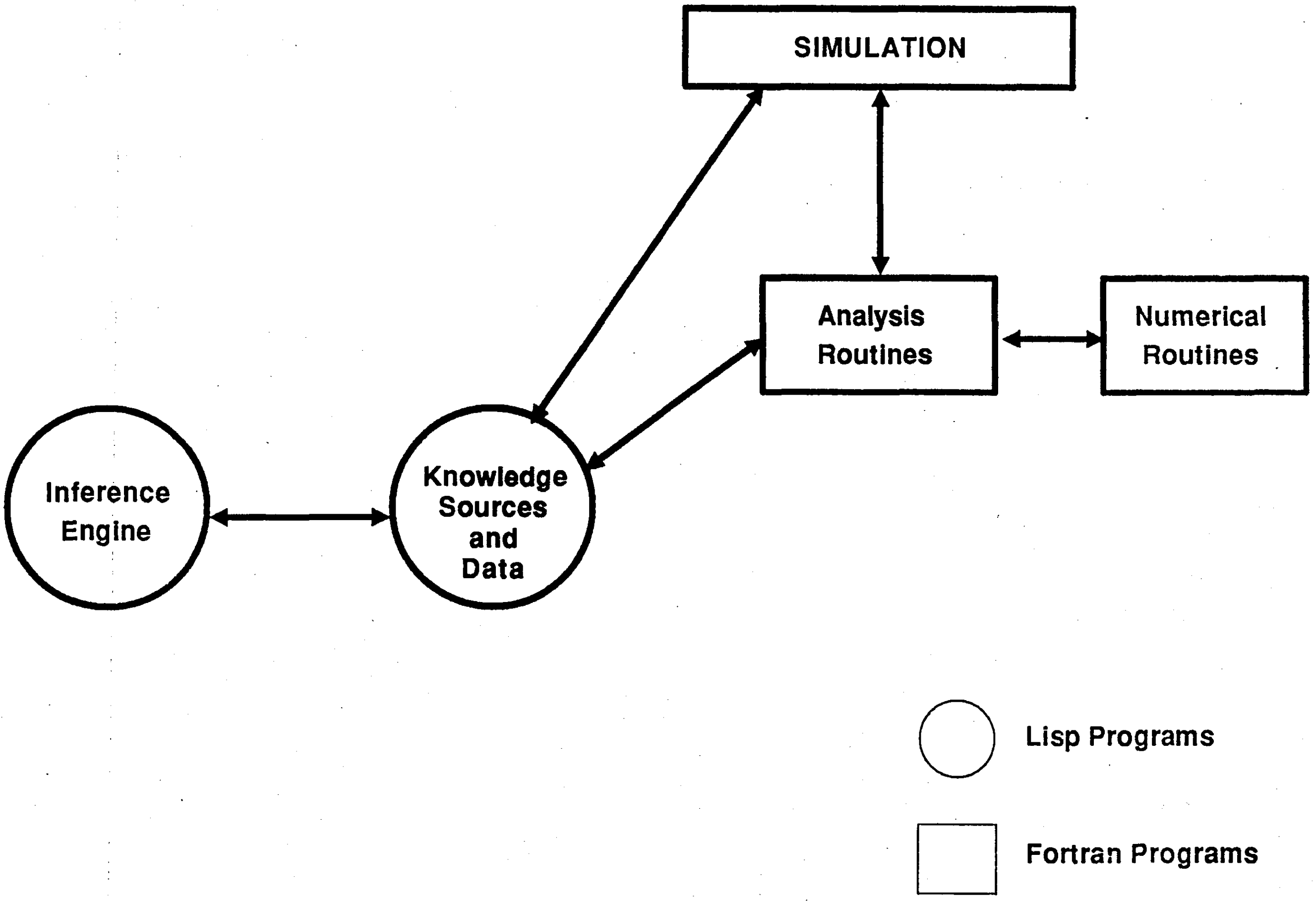


Figure 7: Proposed Knowledge-Based System

It consists of a hybrid lisp/Fortran environment with the lisp environment acting as host calling Fortran routines when necessary. A hybrid implementation is preferred because it combines the numerical affinity of Fortran with the list and symbolic processing powers of lisp. The Fortran environment is composed of simulation, analysis and numerical routines.

The numerical routines are a library of routines to perform matrix algebra, to solve a set of linear simultaneous equations and to generate random numbers for test purposes. The role of the analysis and simulation routines have largely been described elsewhere. The lisp environment consists of knowledge-sources, an inference engine and various global data structures and methods. There are four knowledge-sources arranged as shown in the Figure below.

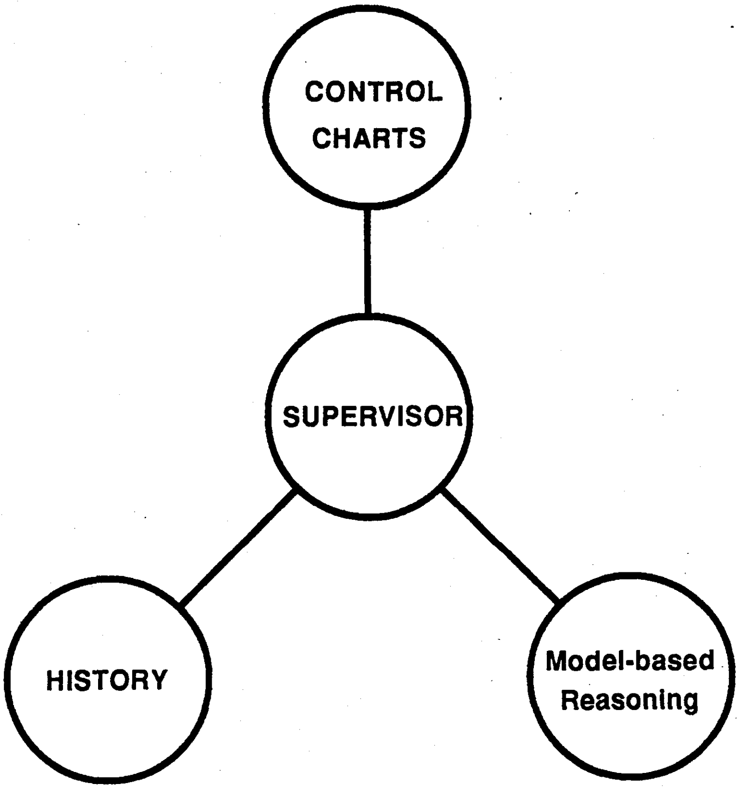


Figure 8: NRTMA Knowledge-Sources

The system is invoked everytime a set of plant measurements becomes available, that is at the end of every period. This data is input to the lisp environment and the control chart knowledge-source is activated.

The primary role of the control charts is to detect the occurrence of a fault and having done so, to output two lists: a list of discrepancies and a list of assertions which point to possible classes of faults that could account for the patterns observed. It would be unusual for the charts to identify a fault uniquely; their role is to focus attention. Two issues complicate the recognition process, that of noise and of multiple faults. Both can largely be overcome by adding any fault, that is remotely likely, to the list. Care must then be taken to ensure that the most likely are considered first.

The supervisor is driven by data flowing from the control charts where the overall objective is to detect and diagnose discrepancies in the control charts. Alternatively, if no discrepancies exist, the system can still be used either to identify malfunctions that do not give rise to discrepancies or to improve the simulation by learning. The Supervisor has two roles, that of an evidence gatherer and hypothesis generator and that of an *adjudicator*. We use the term *gatherer* to refer to the former.

The basic mechanism behind the Gatherer is as follows. On receipt of a list of fault scenarios, the Gatherer takes each fault scenario in turn and invokes one or more of three options: a simulation, a reference to history or its own assessment. In most situations it will invoke all three. Each option returns either a statement of its deliberations or nothing at all.

Finally the *adjudicator* is invoked. As explained in Section 2.7, little research has been carried out into the process of adjudication. The current state assumes that the operator will make his own assessment based on the hypotheses formed.

5.3.2 Knowledge Representation

Individual knowledge-sources are represented in lisp or productions or some combination of the two. Lisp tends to be used only where a specific task is procedural. The productions are of the form

if *antecedent* then *consequent*

or

antecedent → *consequent*

with the following syntax:

i) variables are always preceded by a '?';

ii) antecedants may be made up of one or more components linked by either '&' or 'or'; each component can either take the form of a lisp-like predicate, for instance (> ?a ?b), or a more general predicate, for instance (colour ?a ?b), where the list cannot be evaluated.

iii) consequences may contain a number of special symbols '!' and '$', the symbol '!' precedes any method and the symbol '$' is used to expand lists.

For instance, if the method (get_list x) returns a list (a b c) then

```
(conseq $ ! (get_list x))
```

will produce three consequenses (conseq a), (conseq b) and (conseq c).

There are two special cases: no consequences are produced if the list does not contain any elements and the use of multiple '$' symbols assumes that consequences are to be formed by extracting elements from corresponding positions of each list. Thus,

```
(conseq $(a b c) $(r s t))
```

returns (conseq a r), (conseq b s) and (conseq c t).

It is intended that the interfaces between the various knowledge sources should be custom built, that one knowledge source would usually invoke another via a method and that exit would usually result from either the search being exhausted or from invokation of the consequence (return ?information).

Global data is stored in one of three ways: the simple list, frame and lisp *structure*. Global data is kept to a minimum for ease of program development. Data that is amenable to hierarchical structuring and is relatively static is generally stored in frames. Most other data is stored in structures.

An example of a simple list is that of the list used to identify the individual periods the system currently has data for. These are stored in the list *periods* and take the form

(most_recent_period next_most_recent_period and so on)

where individual periods may be named batch1,batch2,...
or 01-JAN-90, 02-JAN-90, ...
and so on.

An example of a frame is that used to store static data pertaining to the state of the plant on a given period. These frames are automatically created as data is received from the 'plant' and have the form

```
      NAME: name of period
GENERATION: this is initially set to 'master'

          SLOT: plant component
               MEASUREMENT: type of measurement
                            VALUE:  floating point number
                            STATE:  recorded?
               MEASUREMENT: type of measurement
                            VALUE:  floating point number
                            STATE:  recorded?
               MEASUREMENT: type of measurement
                            VALUE:  floating point number
                            STATE:  recorded?
               OPERATION  : state
                            ON: (time_start time_stop)
                            others:
                    and so on
          SLOT: plant component
               MEASUREMENT: type of measurement
                            VALUE:  floating point number
                            STATE:  recorded?
               MEASUREMENT: type of measurement
                            VALUE:  floating point number
                            STATE:  recorded?
                    and so on
          SLOT: plant component
                            STATE:  recorded?
               MEASUREMENT: type of measurement
                            VALUE:  floating point number
                            STATE:  recorded?
                    and so on
```

The *generation* is used to ensure that the original data is not overwritten by any revisions that may be proposed.

An example of a structure when used to store static data is that of the plant component structure:

```
component - volume          : identifier (eg pftvol)
          - analysis        : identifier (eg pftanal)
          - initial-volume  : identifier
          - measurements    : list   (eg (volume analysis))
```

The same structure can also be used to store calculated variables pertaining to the operation of a particular component on a particular period. For instance,

```
component - inventory       : ((period value) (period ..
          - random          : ((period value) (period ..
```

5.3.3 The Inference Engine

It has already been argued that the inference process is one of building-up an explanation to describe the symptoms observed. As distinct from trying out various fault scenarios until one correlates with the symptoms. The former is amenable to forward chaining whereas the latter is amenable to backward chaining. A forward chainer has been developed specifically for this application. It is capable of inferring from productions with syntax described previously with the limitation that variable names can only be one of ?u, ?v ,?x, ?y or ?z. Although expandable this list is pre-defined. The unification algorithm used is that described in Charniak and McDermott (80). Its performance has been enhanced by Lam (81) who incorporated various pointers and data structures.

5.4 The Control Chart Knowledge-Source

5.4.1 Specification of Control Charts and Associated Patterns

It is envisaged that, at least during the infancy of the plant, the only sensible plots will be those of MUF versus time and its cumulative sum because of uncertainty surrounding the measurement models. As described in Section 3.4 the latter can take two forms,

$$\mathrm{CUMUF}_k = \sum_{i=1}^{k} \mathrm{MUF}_k$$

with variance

$$\mathrm{var}(\mathrm{CUMUF}_k) = V_0^2 + \sum_{i=1}^{k} W_i^2 + V_k^2$$

or

$$\mathrm{CUSUM}_k = \sum_{i=1}^{k} \frac{\mathrm{MUF}_k}{\sigma_k}$$

Although NRTMA has tended to concentrate on the accountancy of plutonium, Jones [49] has argued that uranium should also be accounted because this provides additional information for diagnosis. This would then result in 2 sets of plots which could be compared.

A number of heuristics can be derived to explain the patterns that arise when particular categories of error occur. These are based, primarily, on the serial correlation caused by the presence of the same physical inventory in consecutive balance periods. Thus the effect of a physical inventory error should be observed on more than one period.

Five patterns are apparent.

Pattern 1: The effect of an error in a transfer either into or out of the balance area on a single period, m. If this is the only error and has magnitude e, then

$$\mathrm{MUF}_k = \begin{cases} 0 \; ; \\ e \; ; \\ 0 \; ; \end{cases} \qquad \mathrm{CUMUF}_k = \begin{cases} 0 \; ; & k < m \\ e \; ; & k = m \\ e \; ; & k > m \end{cases}$$

Pattern 2: The effect of a physical inventory measurement error on a single period, m, will be observed on both that period and the next. If this is the only error and has magnitude e, then

$$\mathrm{MUF}_k = \begin{cases} 0 \; ; \\ e \; ; \\ -e \; ; \end{cases} \qquad \mathrm{CUMUF}_k = \begin{cases} 0 \; ; & k < m \text{ or } k > m+1 \\ e \; ; & k = m \\ 0 \; ; & k = m+1 \end{cases}$$

That is a reflection will be observed in the MUF chart and a single spike in the CUMUF chart.

Pattern 3: A constant, additive bias in a transfer measurement will be observed as a linear shift in the MUF plot and a non-zero gradient (ie an incline) in the CUMUF plot. If this is the only error, it has magnitude e and it only occurs between periods m to n inclusive then,

$$MUF_k = \begin{cases} 0 \ ; \\ e \ ; \\ 0 \ ; \end{cases} \qquad CUMUF_k = \begin{cases} 0 \ ; & k < m \\ (k-m+1)\ e \ ; & m < k < n \\ (n-m+1)\ e \ ; & n < k \end{cases}$$

Pattern 4: A constant, additive bias in a physical inventory measurement will be observed in the MUF plot as a single spike on the period the bias develops with a reflected spike on the period the bias stops. This will have the effect of creating a plateau on the CUMUF plot. If this is the only error, it has magnitude e and it only occurs during periods m to n inclusive then,

$$MUF_k = \begin{cases} -e \ ; & k = m \\ 0 \ ; & k \neq m \ \text{ or } \ k \neq n+1 \\ e \ ; & k = n+1 \end{cases}$$

$$CUMUF_k = \begin{cases} -e \ ; & k = m \\ -e \ ; & m < k < n \\ 0 \ ; & k < m \ \text{ or } \ k > n \end{cases}$$

Pattern 5: A single loss on period m from a physical inventory would produce Pattern 1 provided that the resultant reduction in output does not occur on the same period. If it does occur on the same period then,

$$\mathrm{MUF}_k = \begin{cases} 0\ ; & k < m \\ 2e\ ; & k = m \\ -e\ ; & k = m+1 \\ 0\ ; & k > m+1 \end{cases}$$

$$\mathrm{CUMUF}_k = \begin{cases} 0\ ; & k < m \\ 2e\ ; & k = m \\ e\ ; & k > m \end{cases}$$

Clearly the patterns will be less well-defined than the above suggests because of the effect of random noise. In addition patterns 3 and 4 may be less clearly defined because neither the net transfer nor the physical inventory are measured directly: a constant bias in a particular measurement will have a more subtle effect on the plots. However the underlying patterns will often be the same with the additional complications often being observed as random fluctuations. Figure 9 shows a typical set of MUF plots corrupted by the various error scenarios. The errors have been applied on Period 10 and, where appropriate, revoked on Period 15.

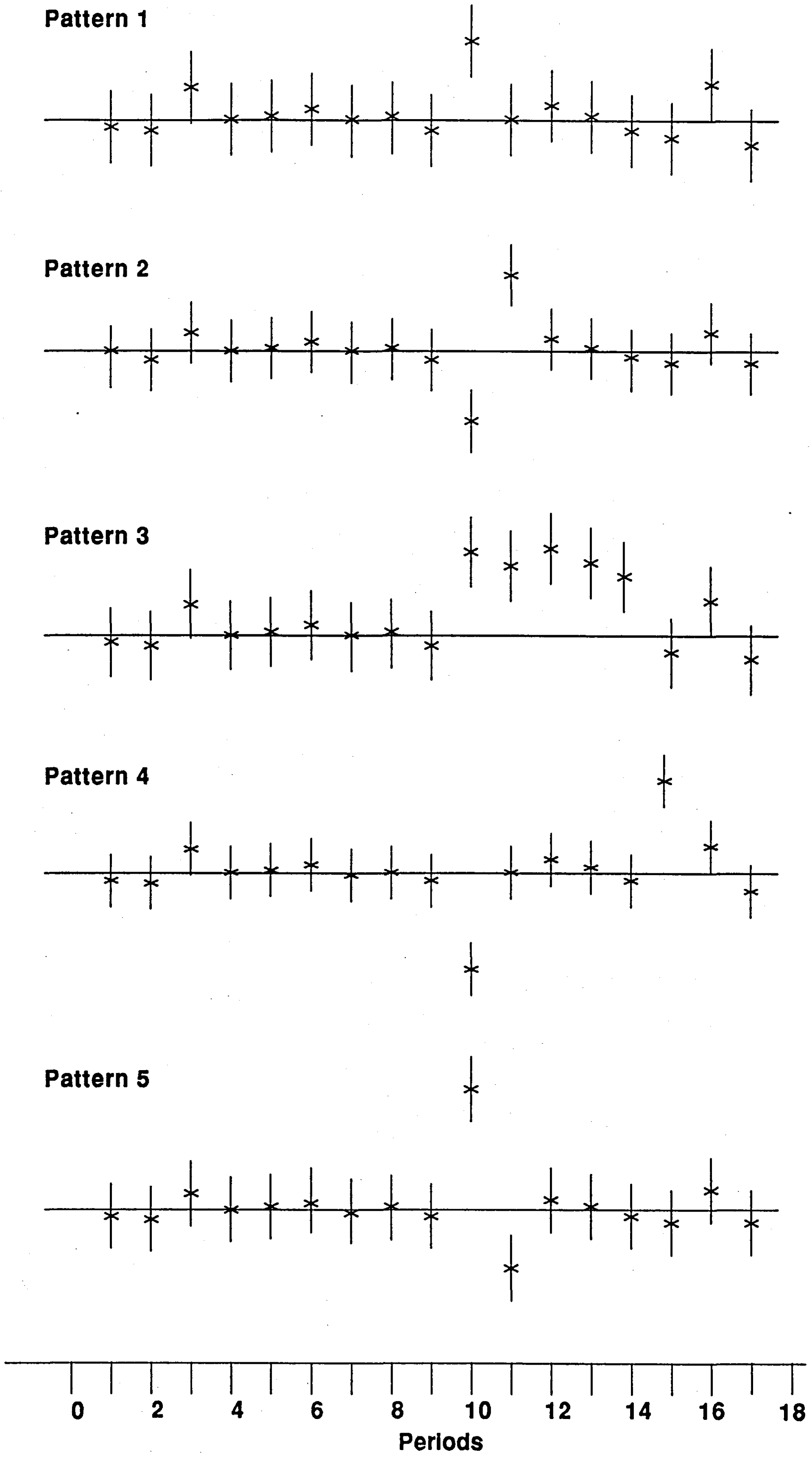


Figure 9: Simulated MUF Plots

5.4.2 Timeliness

With reference to Figure 9, a question arises as to whether all data points should be examined at the end of each period. If the process were to be repeated every period then, initially at least, all the above patterns would be identified as pertaining to Pattern 1 on period m. This has been examined by Russell *et al*[47] in the context of optimal detection. They argue on both practical and statistical grounds that, although the process should be repeated every period, the detector should only be applied to all periods up to, but not including, the current period. That is current data is used only to improve the power to detect something that has arisen on previous periods. If this approach were to be applied to the recognition process then Patterns 1 and 4 would be viewed as one possibility whilst all other patterns would be viewed as separate entities.

The question naturally arises as to whether an even longer delay would be of benefit. Clearly the longer the delay, the more the information and the better the discrimination. Conversely, the longer the delay, the less timely the detection and diagnosis. There is therefore an argument for a recognition process which is applied to all periods (ie with no delay) but with less power to identify patterns on current periods than on previous ones.

5.4.3 Pattern Identification

The approach is to apply the four tests,

test A: alarm if $MUF_k > h_m \sqrt{var (MUF_k)}$

test B: alarm if $MUF_k < h_m \sqrt{var (MUF_k)}$

test C: alarm if $\tau_n^+ \geqslant h$

test D: alarm if $\tau_n^- \leqslant h$

where τ_n^+ and τ_n^- are as defined in Section 3.4, to the relevant time series and to correlate the resulting (binary) sequencies with the (binary) sequencies that would arise if the same tests were applied to the patterns above.

The method favoured in NRTMA [55,59] for choosing the parameters h, k and h_m is to initially use simulation and then to revise with operational experience. The selection process is largely one of balancing credibility, that is, the possibility of an alarm never being caused by random fluctuations against power to detect. Russell [55] has calculated that for the amount of serial correlation expected in NRTMA (ρ=-0.4), the CUSUM test with h = 3.059 and k = 0.063 will ensure 96% credibility over 10 periods (95% over 20) with a power to detect a 0.5σ constant bias within 6 periods of 52.8% (99.3% within 15). This was on the basis that the CUSUM test was the sole test applied. However it is well-known [59,70] that the results are sensitive to the serial correlation present.

A detailed investigation is needed to ensure that reasonable parameters are chosen for a particular plant. For computational reasons, the investigation is best carried out in two stages: an assessment of performance on the basis of series of random numbers followed by an assessment using time series derived from a more realistic simulation. This is because considerable insight can be gleaned by examining the performance of tests on time series with statistics which approximate to reality. Some of the results obtained by considering time series which approximate to those expected in NRTMA are given in the Appendix.

As a result of this insight, the following test combinations appear to provide a method of ordering the various patterns:

1: A_k or B_k

2: $(A_{k-1}$ & $D_k)$ or $(B_{k-1}$ & $C_k)$ or $(A_{k-1}$ & $B_k)$ or
$(A_k$ & $B_{k-1})$ or $(C_k$ & $D_{k-1})$ or $(C_{k-1}$ & $D_k)$

3: $(\bar{A}_k$ & $\bar{B}_k)$ & $(C_k$ or $D_k)$

4: $(A_k$ & $B_{k-1})$ or $(A_{k-1}$ & $B_k)$

since they can be used as follows

combination 1: (pattern 2)$_k$, (pattern 1)$_k$, (pattern 5)$_k$

combination 2: (pattern 2)$_{k-1}$, (pattern 5)$_{k-1}$

combination 3: (pattern 3)$_{k-1}$, (pattern 2)$_k$, (pattern 1)$_k$

combination 4: (pattern 2)$_{k-1}$, (pattern 5)$_{k-1}$

combinations 1 & 2: (pattern 2)$_{k-1}$, (pattern 5)$_{k-1}$

Note that pattern 4 is omitted here. It can be identified, indirectly, by searching for multiple occurrences of pattern 1.

Table 1 shows the results obtained when the four tests were applied to 100,000 series of random numbers; 10,000 for each of the 10 different error scenarios. The random numbers were generated on the basis of an ideal, constant throughput plant with the ratio of the physical inventory to throughput being chosen to produce a serial correlation, ρ, of -0.4 in the MUF time series. The magnitudes of the faults were specified as multiples of the plant throughput and not of σ_{MUF}, as is the convention in the assessment of detectors. The equivalent proportions of σ_{MUF} can be obtained by multiplying the proportion of plant throughput by the factor $\surd(0.5+\rho)$. Thus the results in the first column represent a transfer error of 6% of the throughput on a particular period and so on. The tests assumed that h_m=3.0, h=3.059 and k=0.063.

It is recommended that combinations 2, 3 and 4 should be sought prior to combination 1 because they require more alarms. That is they are compound tests. As described in Section 3.4.2, additional information can be extracted if Pattern 3 is suspected.

Once the pattern recognition process is completed, relevant classes of faults may be added to the list for output. The patterns may not define uniquely a particular class of faults. For instance, Pattern 3 not only describes a net transfer bias but also a temporary hold-up which is slowly building up; for instance as a result of a tank not being monitored. More than one possible class may therefore be output per pattern.

From now on any reference to these patterns may be replaced by more physically meaningful terms to make what follows more readable. Thus,

Pattern 1 - single_transfer_error or inv_loss
Pattern 2 - single_inv_error
Pattern 3 - transfer_bias or inc/dec_holdup
Pattern 4 - temp_holdup
Pattern 5 - inv_loss + sim_out

ALARM Combinations	e Period	Pattern 1		Pattern 2		Pattern 3			Pattern 5		No Fault
		6	12	6	12	1	2	3	6	12	
1	10	13	78	14	78	0	1	2	79	100	0.3
	11	0	0	13	79	0	1	2	14	79	0.3
	12	0	0	0	0	0	1	2	0	0	0.2
	13	0	0	0	0	0	1	2	0	0	0.3
	15	0	0	0	0	0	1	2	0	0	0.3
2	10	1	1	1	2	0	0	0	2	2	0.0
	11	0	0	8	74	0	0	0	13	79	0.0
	12	0	0	0	0	0	0	0	0	0	0.0
	13	0	0	0	0	0	0	0	0	0	0.0
	15	0	0	0	0	0	0	0	0	0	0.0
3	10	25	15	24	15	5	6	9	14	0	3.1
	11	14	4	1	0	6	14	26	0	0	3.5
	12	7	1	9	11	8	20	29	8	11	3.7
	13	5	1	6	4	10	20	22	6	4	3.7
	15	3	2	5	2	10	15	23	4	2	3.8
1 & 2	10	1	1	1	2	0	0	0	2	2	0.0
	11	0	0	8	74	0	0	0	13	79	0.0
	12	0	0	0	0	0	0	0	0	0	0.0
	13	0	0	0	0	0	0	0	0	0	0.0
	15	0	0	0	0	0	0	0	0	0	0.0
1 & 3	10	0	0	0	0	0	0	0	0	0	0.0
	11	0	0	0	0	0	0	0	0	0	0.0
	12	0	0	0	0	0	0	0	0	0	0.0
	13	0	0	0	0	0	0	0	0	0	0.0
	15	0	0	0	0	0	0	0	0	0	0.0
2 & 3	10	0	0	0	0	0	0	0	0	0	0.0
	11	0	0	0	0	0	0	0	0	0	0.0
	12	0	0	0	0	0	0	0	0	0	0.0
	13	0	0	0	0	0	0	0	0	0	0.0
	15	0	0	0	0	0	0	0	0	0	0.0
4	10	0	0	0	0	0	0	0	2	2	0.0
	11	0	0	4	65	0	0	0	13	79	0.0
	12	0	0	0	0	0	0	0	0	0	0.0
	13	0	0	0	0	0	0	0	0	0	0.0
	15	0	0	0	0	0	0	0	0	0	0.0
No alarm in 6 periods	-	27	1	37	1	34	2	0	4	0	70.4

0 denotes < 0.5 %

Table 1 : Test Performance with Time Series

5.4.4 Its Form

The knowledge-source has two parts: a lisp section where the various test statistics (MUF, CUMUF and CUSUM) are both formed and tested, and a production section where the list of alarms are interpreted. Both alarms and interpretations are then entered into the assertion base.

5.5 The Supervisor

The Supervisor is driven by the data added to *goals*. The overall objective is to detect and diagnose discrepancies in the control charts. Alternatively, if no discrepancies exist, the system can still be used either to identify malfunctions that do not give rise to discrepancies or to improve the simulation by learning.

On receipt of *goals*, the Gatherer takes each interpretation in turn and invokes one or more of three options: model-based reasoning, a reference to history or its own assessment. In most situations it will invoke all three. Each option returns either a statement of its deliberations or nothing at all.

Assessments made by the Gatherer can be divided into two parts: those heuristics peculiar to plant operation and those which attempt to interpret the hypotheses generated.

5.5.1 *Typical Plant Operation Heuristics*

A number of rules may be derived from the fact that a tank is simply a storage device. Having determined both the maximum and minimum analyses of the liquor entering the tank over the past n period, rules of the form

```
[  (minimum_input_analysis ?tank ?period ?min)
 & (maximum_input_analysis ?tank ?period ?max)
 & {(analysis_measured < ?min) or (analysis_measured > ?max)} ]
        → (suspect_analysis_measurement ?tank ?period)
```

may be applied. In practice, these assertions may also be used in model-based reasoning so the heuristics are applied here.

Another heuristic [46] for diagnosing transfer errors, is to combine the material balance area with a connecting one so that an external transfer point becomes internal. Data pertaining to this point will not be included in the resulting balance. A comparison of the various balances may then reveal whether this data is suspect.

5.5.2 Hypothesis Interpretation

Most of the hypotheses generated thus far have been subjective because they have been related to one's experience, models with all their uncertainty or plant operation heuristics. It is therefore important to ensure that they will actually explain the alarms ie the *goals*.

Of central importance here are a set of methods or procedures which calculate that change in any particular variable that would eliminate one or more discrepancies that are specified. These methods are denoted here by the single name *perturbation_required*. The objective is then achieved if a set of faults are hypothesised that result in perturbations which explain all the evidence generated provided that the set is valid.

For instance, a single transfer error hypothesis can be corroborated by referring to the following productions,

```
(single_transfer_error ?period) →
  (transfer_occured ?period $ ! {get_transfers ?period})

(transfer_occured ?period ?transfer) →
  (perts_req s_t_e ! {perturbation_required ?period ?transfer})
```

5.5.3 Its Form

The Supervisor (Figure 8) is based solely on productions.

5.6 The History Knowledge-Source

This knowledge source contains data describing any peculiarities pertaining to the plant, either currently or recently and rules based on basic physical principles which relate this type of data to possibile fault scenarios. It is largely application specific and is likely to expand with time.

General assertions are typically of the form,

```
(measurement_maintained period measurement)
(suspect feedstock period)
```

whilst rules are of the form,

```
if [ and (net_transfer_bias_from period)
         (measurement_maintained period measurement)
         (measurement ε net_transfer_measurements) ]
 → (poor_maintenance period measurement)
```

A typical application specific rule relates to problems in the determination of the physical inventory of the solvent-extraction plant. Fortunately it is usual operational policy to run at one flowsheet, ie one load, for extensive periods so that physical inventory changes are infrequent. Pattern 4 will then be observed when a flowsheet change is invoked and the physical inventory is estimated incorrectly. Rules can therefore be applied to correlate this pattern with known plant activity, thus

```
if (Pattern 4 m n) → ! (look_at_solvent_ext_history m n+1)
if [and (flowsheet_change ?sol_ex_a m)
        (flowsheet_change ?sol_ex_b n+1)
        (no_flowsheet_change m+1 n) ]
 → (sol_ex? ?sol_ex_a ?sol_ex_b m n)
```

where look_at_solvent_ext_history is a method which generates the assertions needed by the subsequent rule.

5.7 Model-based Reasoning

The purpose of model-based reasoning is to explain the fault scenario that is hypothesised. This explanation can be given in terms of any level of the physical hierarchy that is feasible. The knowledge-source is written in two parts: the analytical approach of Chapters 2 and 4 is invoked and controlled by lispcode whilst additional heuristics are written as productions. The knowledge-source is invoked through the method look_for_fault with the production system being invoked first.

The method is also used to specify the periods over which the simulation is to be performed. The policy adopted here is to use the category of fault hypothesised to do this: if the category refers to information pertaining to periods (k-i) through to k then it would seem appropriate to start at period (k-i) and stop at either period k or k+1. For instance, one way of discriminating between a single error, on period k, involving a transfer out of the system and a single inventory measurement error is to simulate over two periods k and k+1. The inventory error will only give rise to discrepancies on period k whereas the transfer error will continue to have an effect on subsequent periods.

Hence

```
(single_transfer_error_k) → ! (look_for_fault s_t_e (k) {k k+1})

(single_inv_error k) → ! (look_for_fault s_i_e (k) {k})
```

where the first list denotes the periods to look for a fault and the second list denotes the simulation runlength.

The following assertions are returned to the production system from the analytical approach when the search has been completed.

1. (interpret_model {template*1* template2 ....})

where template*n* contains all successful candidates found in a given sub-lattice as defined by its significant path elements. It has the form

[(number_of_significant_path_elements

(list_of_significant_path_elements)

(candidate*1* candidate2 ...)

(score*1* score2 ...)]

where score*1*>score2>score*3*....
and candidate*n* is a list of all its elements and their estimated values and has the form

[(element*1* value*1*) (element2 value2) .....]

2. (?fault element value list_of_significant_path_elements i:candidatei)

These are components of interpret__model where only elements which are significant are included and scores, which are only consistent, internally, are ignored. The variable ?fault is replaced by the first element of look_for_fault.

5.7.1 The Locality Heuristic

The possibility of focusing on one part of the plant, as a means of reducing the candidate space, was mooted in Section 2.6.9 where it was proposed that only errors associated with those components significantly in error need be considered. However it was appreciated that care must be taken to minimise the possibility of the effects of multiple errors being screened by partial cancellation. Since complete cancellation is unlikely, a pragmatic approach would be to consider all components in error down to a low level of significance, say 1.

Slots containing the related path and non-path errors are therefore assigned to each component structure and these are accessed when constructing the candidate space.

5.7.2 General Framework of the Analytical Approach

Figure 10 shows the primary dataflow through the principal lisp functions of the knowledge-source. Going from top to bottom, these functions are

interface - runs the simulation and identifies suspect components, *focus–inv*, by applying the locality heuristic;

focus - identifies the candidate set on the basis of *focus–inv*, and controls inter_sub_lattice search by outputting sets of significant paths, starting with the Empty Set, then the Universal Set, then the set of single errors, followed by double errors and so on. Section 6.6.4 gives an explanation as to why this strategy is adopted. The search can be terminated, after completing any sub–lattice, if a suitable candidate has been returned from

search_sub_lat - which controls intra-sub-lattice search by identifying all candidates, one row at a time, terminating prematurely if a successful candidate containing solely insignificant path elements and, if variable *non-path* is set .true., non-path elements, is returned from focus12 - which assimilates the results obtained by examining one candidate at a time using

do_it2 - to call out to the Fortran analysis routines and to test the estimates that are returned by updating the simulation variables via

calc–value - and calling on

modify&simulate- to re-run the simulation with the revised values.

Figure 11 shows the primary dataflows through the Fortran routines. The routines are accessed through points A to D. Of central importance are the two routines FOCUS and SIMU: the former controls the construction and application of the regression model whilst the latter performs the simulation. Routines *makeA* and *makeJ* construct the two perturbation matrices J_1 and J_2. Simulation variables are updated through *setV* whereas measurement errors are introduced through *corect.*

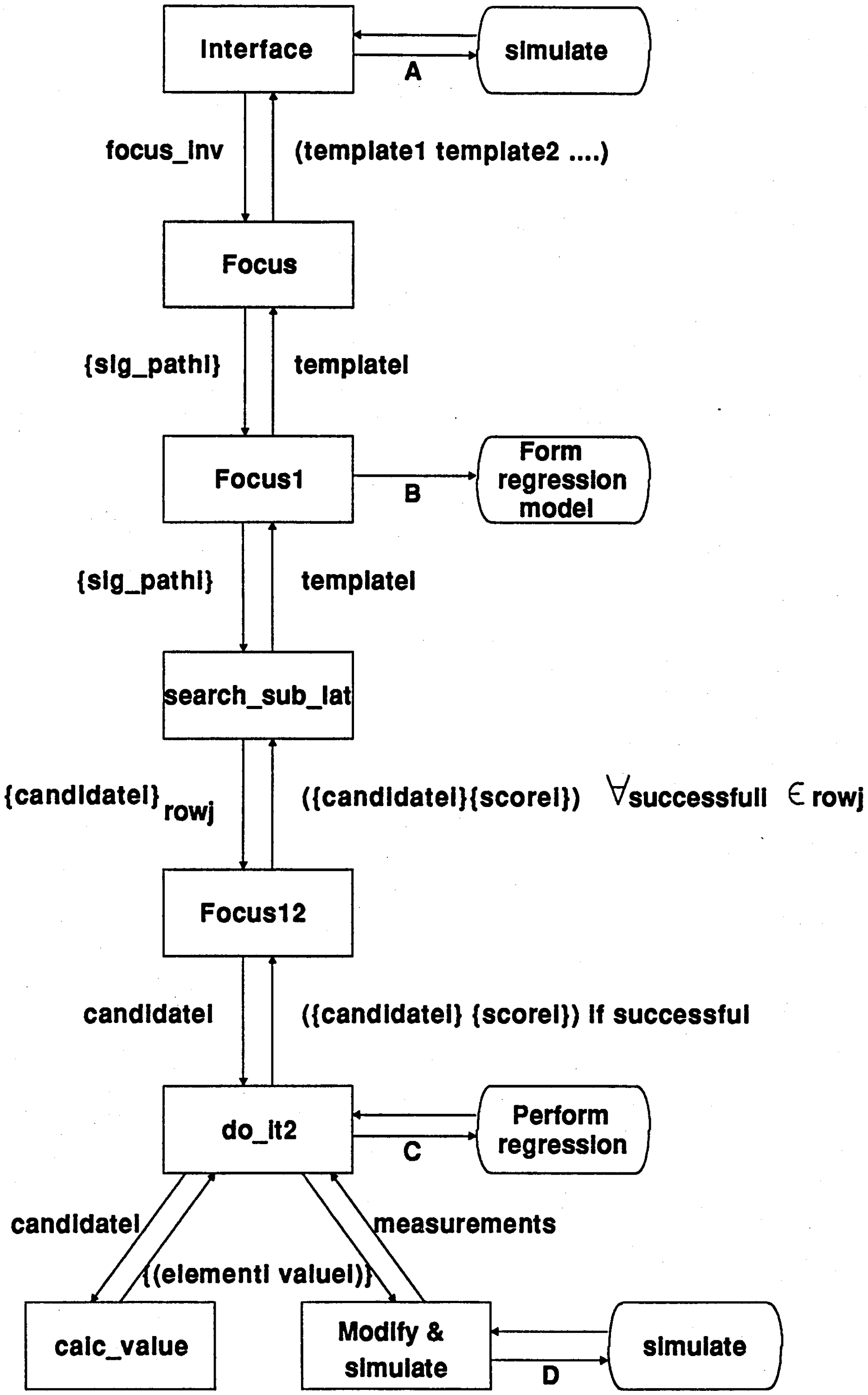


Figure 10: The Lisp Domain in Model-based Reasoning

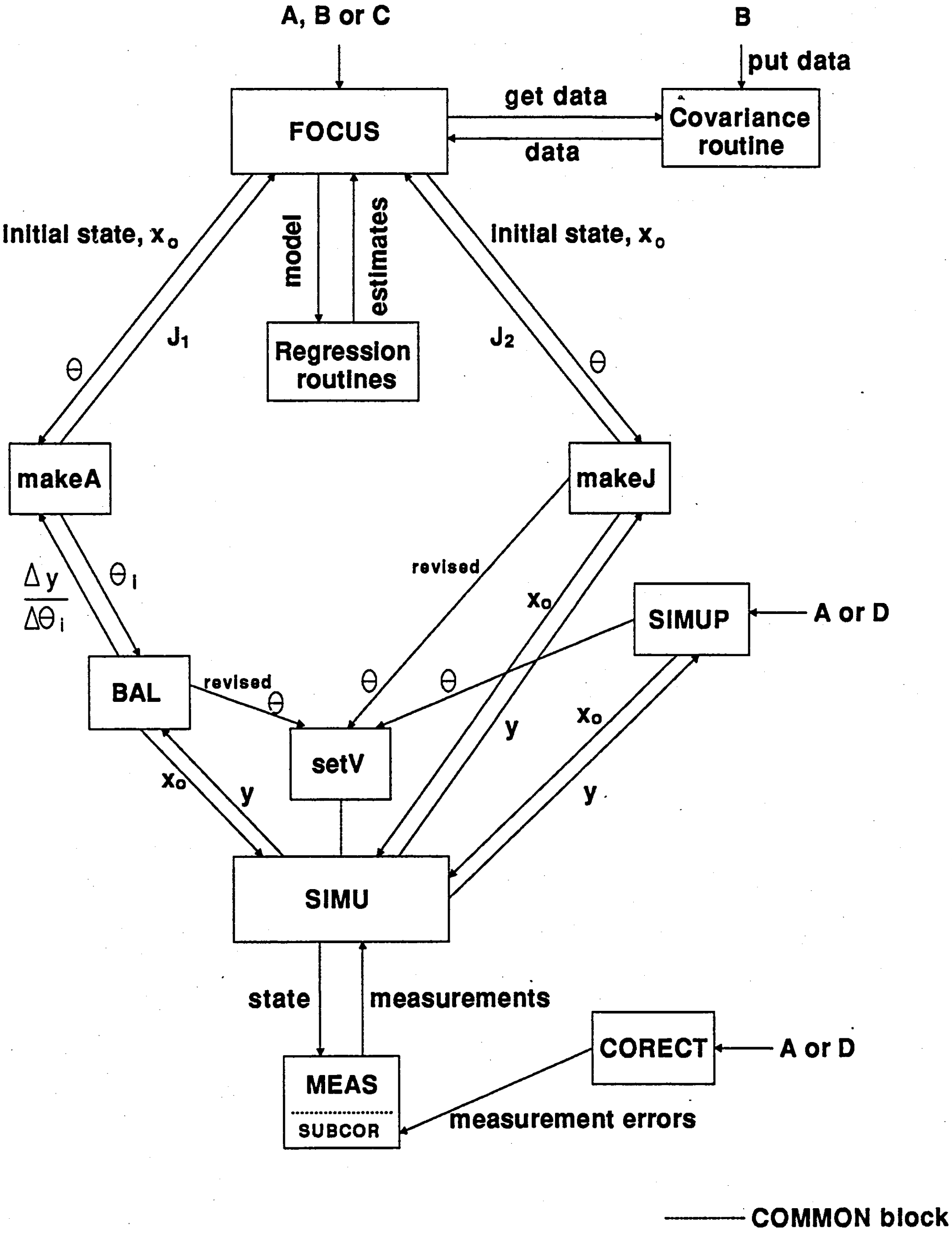


Figure 11: The Principal Fortran Routines

5.7.3 Heuristics

As above, a number of rules may be derived from the fact that a tank is simply a storage device. If a particular tank is suspected on a particular period and if the simulation is accepted as being correct, then the following can be applied to investigate whether the measurement model is in error,

```
[  (simulation_analysis > measured_analysis)
 & (measured_tank_inventory < simulated_tank_inventory)
 & (simulation_analysis > maximum_analysis_input) ]
    → (stratified?)
```

This is of little use by itself because of uncertainty over the simulation. However if

```
[  (stratified?)
 & (! (perturbation_required 'analysis) < simulation_analysis)]
      → (measurement_model_error)
```

also holds then both (stratified?) and (measurement_model_error) can be returned to the Supervisor as evidence.

The above may also be repeated for the case where the measured tank inventory is greater than the simulation tank inventory.

6. APPLICATION TO NEAR REAL TIME MATERIALS ACCOUNTANCY: A Specific Example

6.1 Introduction

This Chapter discusses the steps that should be taken to develop a model-based diagnosis system for a particular plant; in this case a solvent-extraction and concentration plant. There are three stages to such a development:

1. form a skeletal knowledge-base as described in the previous Chapter and produce simulation and analysis routines;

2. test the functionality of the resulting system using data output from the simulation;

3. assess its performance on the real plant.

Unfortunately neither a plant nor resources were available to perform the third stage. A relatively superficial assessment of its performance was therefore made by testing it against a simulation designed to reflect some of the uncertainty surrounding the model.

The plant, its layout and operation, are first described. Models are then proposed both for including in the system and for performing the assessment.

6.2 The Reprocessing Plant

A nuclear fuel reprocessing facility takes spent fuel assemblies as its input and produces separate streams of plutonium nitrate, uranium nitrate, high active, medium active and low active wastes as its outputs. Conventionally the fuel assemblies are first broken apart enabling the individual fuel pins to be extracted. These pins are then cut up into small lengths before being immersed, as batches, into nitric acid. The solution produced from a single batch is first centrifuged to remove any solids, then transferred to a tank where the quantity of plutonium and uranium present is measured. This tank is sometimes called the 'accountancy tank'. Once measured, the batch is emptied into a buffer tank which forms the start of the, continuous, solvent-extraction plant where the separate output streams are formed. Finally, the plutonium nitrate may be concentrated by evaporating off water so as to ease transportation.

Materials balance areas are usually identified as being from the input to and including the accountancy tank and from there to the various outputs. Separate accounts are struck for plutonium and for uranium. It is difficult to simulate the first materials balance area because little published data exists as to its operation. For instance the rate of loss of nuclear material from the main stream as a result of, for instance, small particles being formed at the fuel pin chopping stage is difficult to predict. However sufficient information (50,60) is available to enable a reasonable simulation of the solvent-extraction and concentration plant to be made.

The reprocessing plant examined here is shown in Figure 12. It is assumed that

i) the account is taken every 24 hours and that only one batch can be input during this period;

ii) either Buffer Tank B is connected to the solvent-extraction plant whilst Buffer Tank C feeds the concentrator or vice versa;

iii) Solvent-Extraction Plants A and B operate separately;

iv) Product Storage A is filling whilst Product Storage B is emptied after a more accurate inventory is taken and vice versa;

v) the feed to the Concentrator is switched-off at least 8 hours prior to its inventory being taken. This is similar to the mode of operation identified as being most suitable for NRTMA at Tokai in Japan (51).

Note that the account is taken relatively infrequently. For instance, both the input and concentrator are operated on a 24 hour cycle giving only one set of measurements per account. Although certain measurements may be taken more frequently, for example tank volumes may be recorded hourly, these are omitted for reasons of simplicity. It would obviously be beneficial to include them in any real system and this is raised in the Conclusions.

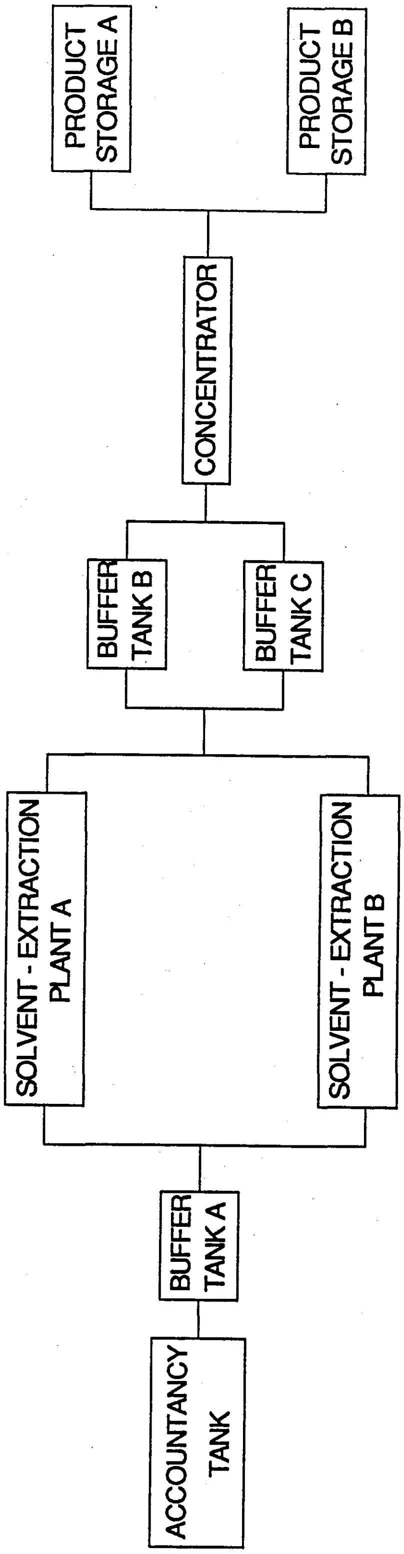


Figure 12: Reprocessing Plant Examined

6.3 The Simulation

A realistic simulation of a nuclear fuel reprocessing plant is not available in the public domain. However sufficient information does exist [49,50,52] to enable a simulation to be constructed which produces a similar level of modelling uncertainty. It must be stressed that the objective is to produce typical effects and not to predict the precise state of a particular plant. The latter would require considerable effort in its development and validation.

The simulation described here, only accounts plutonium as this is thought to be a large enough problem for our purposes. Jones *et al* [49] point out that considerable insight can be gleaned from correlating the plutonium account with the uranium account. Although this aspect should have a place in any fully operational implementation, it represents another level of complexity.

There are potentially two independent variables that are of any significance: the mass of liquor (less heavy metals dissolved) and the mass of the plutonium. However there would be considerable complexity in applying mass balances to the liquor passing through the solvent-extraction and concentration processes. This is because of the different streams entering and leaving the former and because of the evaporation process in the latter. Since the plutonium balance is of primary importance, liquor mass balances are only applied when it is straightforward to do so.

The model consists of simple mass and volume balances of the form:

$$\text{Mass}_j = \text{Mass}_{j-1} + \text{Net–transfer}_{j-1,j}$$

where $\text{Mass}_j \equiv$ mass at end of time step j,

to ensure an accurate account is maintained throughout. This is applied 96 times per accountancy period where the actual time interval used is allowed to vary to ensure a balance is taken whenever there is a change in plant operation.

6.3.1 Buffer Tanks

Modelling the contents of tanks is not straightforward because of difficulties in predicting the degree of inhomogeneity that is present. The task is made more difficult by the fact that one of the key features of a reprocessing plant is that the geometry of each plant item is designed to minimise the risk of criticality. The resulting designs may introduce complicated geometries which can increase the possibility of the liquor being stratified. Apparently [62] these tanks incorporate continuous mixing techniques to ameliorate this situation. Nevertheless this mixing does not produce homogeneity, resulting in biases in the analytical samples. It is debatable as to whether or not these biases are deterministic. Since the aim here is not one of reality and because it was felt that errors in model structures are of interest, a simplified deterministic approach has been adopted.

The tanks are modelled on the assumption that liquor enters at the top and leaves at the bottom. If stratification were to take place, then it would probably result in horizontal bands, of different chemical composition, being present in the tank. The depth of these bands would depend on variation in the feed. A simplified model is obtained by assuming that the liquor flowing into the tank forms horizontal zones of volume one-twentieth of the total volume of the tank where the liquor contained in each zone is perfectly mixed. These zones move down as liquor leaves the tank. The amount of inter-zonal mixing can be varied for individual simulations. The approach is to allow the N-zones closest to the free surface to mix, perfectly, where N is chosen to be between 1 and 20. In the case of Buffer Tank A where relatively large batches are input relatively infrequently, N is made to vary with the volume of each batch input.

An inventory of the tank contents is obtained by multiplying an estimate of the liquor volume by an estimate of it's plutonium content (per unit volume). In practice the volume is usually obtained by measuring differential pressure using pneumercator diptubes and density and relating to volume via calibration tables. The plutonium content is estimated by performing a volumetric analysis on a single sample taken from the tank. Good accountancy procedure dictates that the contents are properly homogenized prior to the sample being taken [51]. It is debatable to what extent this would be practicable if carried out frequently.

A further parameter, similar to N above, is therefore provided to enable different levels of inter-zonal mixing to be simulated prior to a volumetric analysis being made. The sample is then assumed to be taken from the zone at the free surface.

This uncertainty provides a useful scenario for performing a simulation-based assessment. Can a diagnostic system which assumes that the individual buffer tanks are perfectly mixed analyse a simulation with stratified tanks? We will return to this scenario later.

6.3.2 The Solvent-Extraction Plant

The plutonium inventory in a solvent-extraction plant is largely determined by the flowrates of its various inputs [(60)]. These tend to be varied together to maintain the heavy metal front in a fixed position. Under these circumstances and to a first approximation,

$$\text{plutonium inventory} \propto \text{plant throughput}$$

Although it may be possible to determine the plutonium inventory in low to medium active parts of the plant by direct measurement, there exists at present no satisfactory way of measuring the inventory in the first, ie high active, cycle. Here the high level of fission products and higher actinides mask any proven plutonium detection techniques. Computer predictions carried out by Walford *et al* [(60)] suggest that the inventory in the first cycle can be varied by between 0.8 to 4 times the design inventory by manipulating the solvent and scrub feeds. They publish a graph showing the steady state variation in inventory obtained by varying these two feeds separately.

This uncertainty provides another useful scenario for performing a model-based assessment. Can a system which adopts the simple proportionality above analyse a simulation where load and first cycle solvent and scrub feeds are not known precisely?

The detailed simulation is therefore based on the afore mentioned graph, assuming that the two effects can be summed together. Specifically,

$$\left[\begin{array}{l} \text{inventory of} \\ \text{first cycle} \end{array} \right] \quad \alpha \quad \left[\frac{L}{L_0} \right] \left(1 + \alpha_{sol} + \alpha_{scr} \right)$$

where $\frac{L}{L_0}$ - percentage load

$$\alpha_{sol} = \begin{cases} -\,3.\ \Delta Q_{sol} & \Delta Q_{sol} \leqslant 0 \\ 150.\ (\Delta Q_{sol})^2 & \Delta Q_{sol} < 0 \end{cases}$$

$$\alpha_{scr} = \max\left(0.5.\ \Delta Q_{scr}\ ,\ 16.\ \Delta Q_{scr} - 5.9 \right)$$

and ΔQ_{sol}, ΔQ_{scr} are deviations in feeds from their nominal values.

It is assumed that any feed or load changes are invoked at the beginning of a period, gradually over 8 hours, giving the plant time to settle before an inventory is taken. Deviations from this design inventory may then be imposed by perturbing the various feeds.

6.3.3 The Concentrator

The proposed mode of operation should result in the product liquor being homogeneous and of approximately the same volume and concentration whenever the inventory is taken. This is because it is assumed that the concentrator continues to evaporate and produce product after its feed has been stopped approximately 8 hours prior to the inventory being taken. By this time a steady state should be reached. This steady state can be viewed as being determined by volume and concentration 'setpoints'.

Assuming that the Concentrator performs as predicted then the only issue that arises in modelling for accountancy is in determining the rate at which liquor is produced. The simplest approach, and that adopted here, is to assume a constant rate of production.

6.3.4 *Product Storage*

Stratification is less of a problem here than in the buffer tanks provided the input concentration is maintained relatively constant. Tanks with perfect mixing are therefore adopted.

6.3.5 *Random Measurement Errors*

In accordance with convention, measurements are assumed to be corrupted by random, gaussian distributed errors.

6.3.6 *Plant Operation Timing Errors*

There appears to be little justification in recording changes in plant operation automatically. The operators are more likely to enter the time, for instance, at which the feed was switched from Product Storage A to Product Storage B, manually. Some form of rounding error must therefore arise, its extent being dependent on the particular operator. For instance, he could round to minutes, 5-minutes, 10-minutes, quarter-hours and so on.

Gaussian distributed random number generators are used to simulate these effects.

6.3.7 Some Typical Faults

System performance was assessed by examining its ability to detect and diagnose the following *faults*:

A. an erroneous measurement of volume (equivalent to 3.7 σ_{MUF}) of the plant feed tank (buffer tank A) on period 5;

B. an additative bias of approximately 2.3% in the accountancy tank volume measurement from period 3;

C. a diversion from the concentrator (equivalent to 4.2 σ_{MUF}) on period 5;

D. solvent_extraction_A plant load incorrectly specified on period 5 as 60% instead of 100% load; this will cause the simulation to either over or under predict the inventory in the plant feed tank with an associated under or over prediction of the inventories of buffer tanks B and C. In addition, the estimated inventory of the relevant solvent extraction plant will be in error.

6.4 Base Simulations

With the absence of a real plant for comparison, it is assumed that the structure of the model incorporated into the fault diagnosis system is the simplest possible. That is, one with perfect mixing in all the buffer tanks and the solvent-extraction plant as described above with the ratios of the various feeds maintained constant. The only random errors are those assumed to exist in the individual measurements.

System performance can then be examined by diagnosing faults generated from a simulation in which non-perfect mixing and variations in a number of feeds and parameters may or may not be imposed. To be concise, only two plant simulations are considered here,

Plant A:

i) random errors applied to individual measurements;

ii) fluctuations in each of the elements initially contained in the set of simulation variables, Θ:

Plant B: as above plus

iii) only the top quarter ,of the total possible volume, of each buffer tank is mixed;

iv) 0% mixing in the buffer tanks prior to sampling;

v) 0.5% fluctuation in the nominal load of the solvent-extraction plants;

vi) 0.5% fluctuations in each of the feeds to the first cycle of the solvent-extraction plant:

where an n% fluctuation is modelled as a multiplicative gaussian random error with standard deviation, n/100. In all, 16 different random sources are included in Plant B in addition to those used to corrupt the individual measurements. It must be emphasised again that these plant simulations have been performed for the purposes of experimentation and not to depict reality.

The simulation runs are restricted to 16 periods in all cases because of the not insignificant amount of data required like switching times. There are ten sets of data in all. Multiple runs are then obtained by repeating this 10*16 period sequence using different sets of random numbers. Although an attempt has been made to make the 10 sets different, a certain amount of similarity still exists because of difficulties in generating operational times which do not result in tanks being emptied.

It is difficult to give a reasonable impression of the difference in results generated from Plants A&B because

i) the approach is statistical: no single set of graphs will be truly representative;

ii) individual physical inventories may not be startlingly different: a difference of only 5% in the physical inventory of a particular component may typically represent one σ_{MUF}.

Figure 13 compares the MUF plot generated by Plants A (x) & B (o) using the same set of random numbers and simulating Fault A but applied on Period 10. Three σ_{MUF} error bars are superimposed. The results that are presented here are intended to give a general impression and no more.

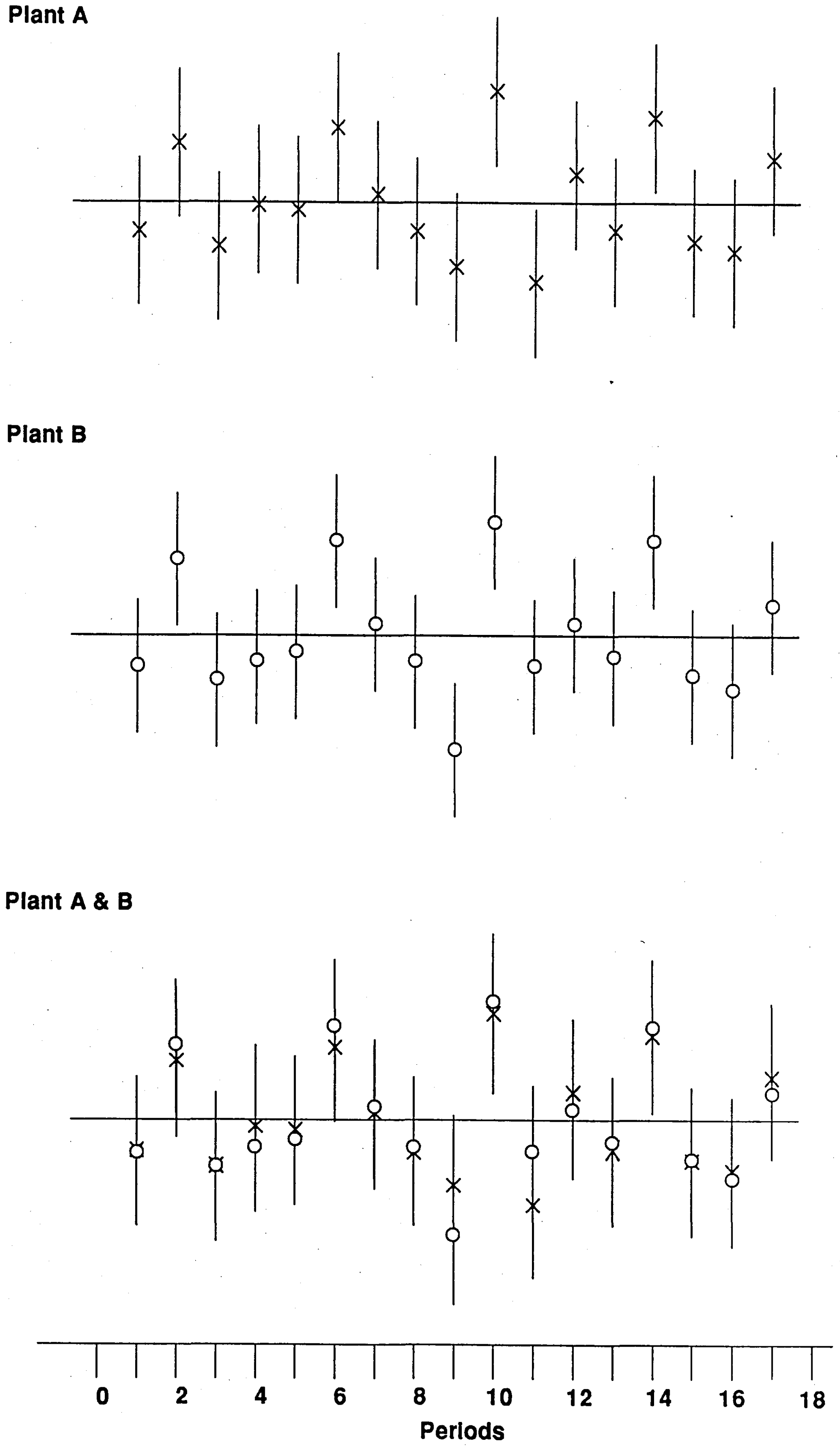


Figure 13 : Typical MUF plots

6.5 Control Charts

6.5.1 Their Form

The control charts incorporated into the system are as defined in the previous Chapter. These require some estimate of the error distributions of the individual transfers, inventories and hence, MUFs. The conventional approach to determining these error distributions would be to propagate the measurement errors through the various calculations. Since these calculations are a combination of additions and multiplications, this procedure would involve obtaining corresponding expressions for their error distributions. In particular [40], if x and y are two independent random variables with probabily densities f(x) and g(y) respectively, the probability density function p(w) of random variable w: w=xy is given by

$$p(w) = \int_{-\infty}^{\infty} \frac{1}{|z|} \; . \; f(z) \; . \; g\left(\frac{w}{z}\right) dz$$

Fortunately, this procedure need not be adopted in practice because of the central limit theorem [37]. The resulting error distribution should be approximately gaussian because the MUF calculation is a summation of a number of random variables with similar probability densities.

Of importance here then are the two results,

$$\text{var}\ \{ax+by\} = a^2\ \text{var}\ \{x\} + b^2\ \text{var}\ \{y\}$$

$$\begin{aligned}
\text{var}\ \{xy\} &\equiv E[(xy - E[xy])^2] \\
&= E[(xy)^2] - E[xy]^2 \\
&= E[x^2]E[y^2] - E[x]^2E[y]^2 \\
&= \text{var}\ \{x\}\ \text{var}\ \{y\} + E[x]^2\ \text{var}\ \{y\} + E[y]^2\ \text{var}\ \{x\}
\end{aligned}$$

6.5.2 Some Results

The results described in Section 5.3.2 were derived from studies using series of random numbers. This Section examines whether the conclusions drawn can be extrapolated to more realistic simulations. The method of combining alarms to infer particular patterns should be universally applicable because it is based on qualitative rather than quantitative arguments. It is the test parameters that are suspect.

The effect of varying serial correlation was first examined by applying identical statistical tests to those described in that section, this time to 100,000 repeated simulations of the plant assuming 100% mixing in the buffer tanks, no fluctuations in feeds to the solvent-extraction plant and the precise recording of operational changes. That is, the only noise stemmed from the measurement system. The fault scenarios which would cause the various patterns were again simulated by corrupting the resultant MUF sequences, 10,000 simulations per scenario. The results obtained are very similar as can be seen from Table 2.

The experiments were then repeated but this time using realistic Plant B. The results obtained are given in Table 3. Note that there are generally more alarms as would be expected with the increased variation.

ALARM Combinations	Period	Pattern 1		Pattern 2		Pattern 3			Pattern 5		No Fault
		6	12	6	12	1	2	3	6	12	
1	10	13	78	13	79	0	1	2	78	100	0.2
	11	0	0	13	77	0	1	2	13	78	0.2
	12	0	0	0	0	0	1	2	0	0	0.3
	13	0	0	0	0	0	1	2	0	0	0.3
	15	0	0	0	0	0	1	2	0	0	0.2
2	10	0	1	1	1	0	0	0	0	1	0.0
	11	0	0	8	75	0	0	0	13	78	0.0
	12	0	0	0	0	0	0	0	0	0	0.0
	13	0	0	0	0	0	0	0	0	0	0.0
	15	0	0	0	0	0	0	0	0	0	0.0
3	10	22	18	22	18	1	3	5	18	0	1.0
	11	17	3	0	0	4	11	28	0	0	1.5
	12	8	1	6	13	5	21	37	6	13	0.9
	13	5	0	5	5	8	27	20	4	5	1.0
	15	3	0	3	1	12	13	23	3	2	1.3
1 & 2	10	0	0	1	1	0	0	0	1	1	0.0
	11	0	0	8	75	0	0	0	13	78	0.0
	12	0	0	0	0	0	0	0	0	0	0.0
	13	0	0	0	0	0	0	0	0	0	0.0
	15	0	0	0	0	0	0	0	0	0	0.0
4	10	0	1	0	0	0	0	0	0	0	0.0
	11	0	0	4	65	0	0	0	12	78	0.0
	12	0	0	0	0	0	0	0	0	0	0.0
	13	0	0	0	0	0	0	0	0	0	0.0
	15	0	0	0	0	0	0	0	0	0	0.0

Table 2: Test Performance on Plant A

		Pattern 1		Pattern 2		Pattern 3			Pattern 5		No Faults		
		6	12	6	12	1	2	3	6	12	3σ/ 3.059	4σ/ 3.059	
	Period												
1	10	44	96	52	98	4	8	14	43	100	2	0	0
	11	0	0	23	87	0	0	1	8	68	0	0	0
	12	0	0	0	0	1	1	3	0	0	0	0	0
	13	0	0	0	0	1	3	5	0	0	0	0	0
	15	0	0	0	0	0	0	1	0	0	0	0	0
2	10	3	3	3	3	1	1	2	0	0	0	0	0
	11	0	0	21	87	0	0	0	7	68	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	1
	15	0	0	0	0	0	0	0	0	0	0	0	0
3	10	27	4	27	2	7	11	15	40	0	4	5	0
	11	10	0	0	0	8	19	33	1	0	3	3	0
	12	6	0	8	8	12	30	29	5	18	2	3	0
	13	6	0	3	2	26	25	16	9	10	6	6	0
	15	1	0	3	1	7	6	22	2	1	2	2	0
1&2	10	3	3	3	3	1	1	2	0	0	0	0	0
	11	0	0	21	87	0	0	0	7	68	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0
4	10	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	17	86	0	0	0	6	68	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0
No alarm in 15 periods	-	3	-	13	0	18	0	0	4	0	56	57	93

Table 3: Test Performance on Plant B

6.6 Model-based Reasoning

The experiments described here have largely been confined to reasoning about a single period primarily because of a lack of resources. There has been insufficient time to consider the added dimension of errors in the simulation variables being correlated and the increased complexity resulting from multiple periods would have stretched the current computational facility of a VaxStation 3100 with 8Mb of memory.

6.6.1 Its Form

The *symptoms* are generated by comparing all the measurements output for the purpose of performing the near real time accountancy of plutonium with those output from the simulation. This, essentially, consists of volumes and volumetric analyses. There are thirteen such measurements output every period from the simulation here. Each measurement must have a measurement error, in terms of a standard deviation, attached to it.

Initially the set Θ of parameters and other variables necessary to describe the re-distribution process is restricted to those elements that form the boundary conditions of the simulation. That is,

```
accountancy_tank_volume_transfered
accountancy_tank_analysis_transfered
time_of_accountancy_tank_transfer
start_time_of_feed_to_Product_tank_A
start_time_of_feed_to_Product_tank_B
time_concentrator_switched_off
start_time_of_feed_to_concentrate_tank_A
start_time_of_feed_to_concentrate_tank_B
solvent_extraction_plant_A_load
solvent_extraction_plant_B_load
```

Other elements can be added with experience. The uncertainty matrix P_k must be specified for each period k. Fortunately it is likely to be diagonal with those elements involving time being independent of k. The other elements are likely to be some multiple of either the volume, analysis or load involved.

In addition, the plutonium content estimates in both the solvent-extraction plants and the concentrator must be omitted from this assessment because the model used to estimate the content is also incorporated into the simulation. That is, $\hat{y}=\tilde{y}$ in both cases. Thus errors in the various feeds will have an identical effect on both $\hat{y}$ and $\tilde{y}$. Their presence must be inferred from their effect on the other measurements. These errors must therefore be included, explicitly, in Θ. Thus

```
solvent_extraction_plant_A_inventory_error ∊ Θ
solvent_extraction_plant_B_inventory_error ∊ Θ
concentrator_inventory_error ∊ Θ
```

6.6.2 The Candidate Space

A maximum of three explanations per period is chosen, at least initially, as being the most the diagnostician would be prepared to contemplate at any one time. This is obviously a moot point but is justified on the grounds that it represents a candidate space of some 9737 candidates which is more than sufficient for our purposes here. In addition and as suggested in Section 4.7, the lattice is searched one sub-lattice at a time to minimise computer storage.

6.6.3 Fault Free Studies

A number of experiments were carried out to optimise the credibility of the approach, that is its false alarm rate, to examine the effect of using composite as opposed to single elements (Section 4.6) and to assess whether the search space could be reduced by applying the locality heuristic as described in Section 2.6.9.

Each experiment consisted of diagnosing results from 10 simulations, of 16 periods each, of both Plants A and B. The same covariance matrices P_k, R_{k-1} and R_k as used to generate the data were assumed throughout. A particular candidate was deemed to be credible (Section 4.5) if each estimate satisfied its *apriori* variance at level 3 or less and the revised symptoms were explained at level 3 or below. In all, Plant A required some form of explanation on 108 periods whilst Plant B required explanations on 129 periods. (note that this false alarm rate cannot be compared directly with the results of Table 2 because different tests were applied.)

Ideally the diagnostician would prefer the system to only output path errors that are insignificant (Section 2.6.4) because no faults have arisen. Hence the search was restricted to that sub-lattice that contained solely insignificant path errors. The search adopted a depth first strategy, considering path errors first, and then non-path errors.

The following was observed.

1. Diagnosis was successful on all periods.

2. Only 3 non-path explanations were required if only candidates with composite elements were considered in the diagnosis of plant A. This compared with 10 if single elements were used instead. A decision was therefore made to use composite elements in all subsequent experiments.

3. It was felt that the demanded explanation rate of 129/160 for Plant B was rather excessive. The test on the revised symptoms was altered to be successful at level 4 or below. Plant A now required 80 explanations whilst Plant B required 118.

4. It was found that there was very little difference if the search space was restricted further by applying the locality heuristic of Section 5.7.1. This can be seen from Table 4 below which shows the distribution of candidates, providing successful explanations at level 4 or below, for Plants A and B with or without the locality heuristic being applied. This similarity was also observed in all subsequent experiments.

		CANDIDATES					
		Path elements only			Path & non-path elements		
		Number of elements			Number of elements		
Plant	Locality Heuristic	1	2	3	1	2	3
A	YES	64	14	0	0	2	0
	NO	64	13	1	0	2	0
B	YES	80	32	3	0	3	0
	NO	80	32	5	0	1	0

Table 4 : Distribution of Successful Candidates

6.6.4 Fault Studies

The simulations above were repeated but with either Fault A, C or D included creating 60 datasets in all, 3*10 pertaining to Plant A and 3*10 to Plant B. Again, that sub-lattice containing solely insignificant path errors was searched with elements being preferred to non-path ones. The results are given in Table 5. It can be seen that

Fault A: the correct measurement error is identified in all cases;

Fault C: diagnosis has been incorrect because it should have been performed over two periods (Section 5.7). However it has identified the correct plant component and, as will be shown in the next Section, still provides useful information;

Fault D: one diagnosis, per plant, failed completely whilst the other 9 required at least 2 non-path explanations. A question therefore arose as to whether or not these 2 explanations could be replaced by a single, significant path explanation. That sub-lattice containing no insignificant path errors was therefore examined for those datasets containing Fault D and the following results were obtained:

Plant A - $7*p^* \ \& \ 3*2p^*$

Plant B - $5*2p^* \ \& \ 1*(p^*+np) \ \& \ 4*3p^*$

where * means ‘cases had’ and a particular p^* in every case was that composite element that contained both solvent_extraction_plant_A_load and solvent_extraction_plant_B_load. Each sub-lattice containing a single significant path error which was identical to one of those identified was searched next. Only 2 significant path errors were succesful in the case of Plant A, the 2 loads and the following candidates were identified,

Plant A - solvent_extraction_plant_A_load* + (7 & 3*p)

Plant B - solvent_extraction_plant_A_load* + (4*p & 2*np & 4*2p)

This is the stategy adopted in function ’focus’ described in Section 5.7.2.

		CANDIDATES						
		p = path np = non-path						
Plant	Fault	np	p+np	2np	2p+np	p+2np	3np	np identified
A	A	6	4					pft_vol
	C	7	3					prod_storA_vol
	D					8	1	pft_vol +buf_tankB_vol (+prod_storA_vol)
B	A		7		3			pft_vol
	C	1	7		2			prod_storA_vol
	D					8	1	pft_vol +buf_tankB_vol (+prod_storA_vol)

Table 5 : Fault Studies Using Insignificant Path Sub-lattice

6.7 Overall System Performance

The system implemented to date is skeletal. As explained in Sections 5.1, 5.3.1 and 6.0, this is partly because of the lack of a real facility and partly because of a lack of time. The main limitations are,

1. the system only contains the bare minimum of heuristics needed to prove the interaction of the knowledge-sources;

2. only regression models based on the single period are formed;

3. the system lacks an adjudicator;

4. catastrophic failures (Section 2.6.4) are not accommodated.

The following examples are given merely to elaborate on system performance and to demonstrate its potential. Considerably more development is required before the system can be viewed as being operationally viable.

At present, the consequences generated by the control–chart and supervisor knowledge-sources are held in a common assertion base, *goals*. Model-based reasoning and history have their own assertion bases, *sim_asserts* and *corr_asserts*. The following assertion bases were generated when the system was invoked using the data described previously.

Plant A, Fault A: system invoked at the end of the day following the fault,

goals - as produced by the control charts

```
(INV–LOSS+SIM–OUT YESTERDAY)
(INV–LOSS YESTERDAY)
(SINGLE–TRANSFER–ERROR YESTERDAY)
(SINGLE–INV–ERROR YESTERDAY)
(LOOK–AT CUSUMP TODAY)
(LOOK–AT MUFTESTP TODAY)
(CUSUM–P YESTERDAY)
(MUFTEST–P YESTERDAY 316.498)
```

Note that the patterns were generated solely on the basis of combination 1. The LOOK-AT assertions are merely for connecting rules and are of no interest to the diagnostician.

goals - produced by the supervisor,

```
(VALUE–REQ S–T–E YESTERDAY PROD–STOR–B
  ((VOLUME 1.0e6 1.0e6)
   (ANALYSIS 73.8536 3.113602)
   (INI–VOL 106.1241 4.474106)))
(VALUE–REQ S–T–E YESTERDAY ACCN–T
  ((VOLUME -8.36588 9.865882)
   (ANALYSIS 29.20196 -2.87804)
   (INI–VOL 101.6041 -9.86589)))
(LOOK–FOR–FAULT S–I–E YESTERDAY (YESTERDAY YESTERDAY))
(LOOK–FOR–FAULT S–T–E YESTERDAY (YESTERDAY TODAY))
(TRANSFER–OCCURED YESTERDAY PROD–STOR–B)
(TRANSFER–OCCURED YESTERDAY ACCN–T)
(LOOK–FOR–FAULT I–L YESTERDAY (YESTERDAY TODAY))
```

The VALUE–REQ assertions specify the measurements together with their respective perturbations that would be needed to individually explain the MUFTEST–P. These would be of interest if model-based reasoning suspected either the accountancy tank or product storage tank B.

sim_asserts - as produced by the analytical approach,

```
(S–I–E ((PFTVOL–MEAS -9.20707)) NIL 1)
(INTERPRET–MODEL S–I–E
           ((0 NIL (((1 -0.492931) (991 -9.20707)))
                                  (0.965009))))
```

Only one possible explanation was generated on the second row of the insignificant path error sub-lattice and the search was terminated. (The first row having identified no suitable candidates.) A reduction of -9.2*l* in the simulated volume in the plant feed tank would explain the measurements. That is, there was a measurement error of -9.2*l* as compared to an actual error of -10*l*. An insignificant error in the volume input to the plant is also needed to explain the discrepancies.

sim–asserts - initial assertions'

```
(INV–TYPE PFT I–TANK)
(INV–TYPE SOL–EX–1 SPECIAL)
(INV–TYPE SOL–EX–2 SPECIAL)
(INV–TYPE BUF–TANK–B I–TANK)
(INV–TYPE BUF–TANK–C I–TANK)
(INV–TYPE EVAPORATOR SPECIAL)
(INV–TYPE PROD–STOR–A I–TANK)
(INV–TYPE PROD–STOR–B I–TANK)
```

sim–asserts - base data,

```
(ERROR YESTERDAY BUF–TANK–C -38.5993)
(ERROR YESTERDAY BUF–TANK–B 53.93359)
(ERROR YESTERDAY PFT -327.456)
(PROBLEM–INV YESTERDAY BUF–TANK–C)
(PROBLEM–INV YESTERDAY BUF–TANK–B)
(PROBLEM–INV YESTERDAY PFT)
```

3 suspect components have been identified by comparing the model output with the measurements where error=(actual-predicted). These form *focus–inv* (Section 5.7.2).

sim_asserts - produced by heuristics,

```
(ANAL–REQ PFT 30.28288)
(MIN–INPUT–ANAL PFT YESTERDAY 28.69)
(MAX–INPUT–ANAL PFT YESTERDAY 37.07)
(INPUT–ANAL–HIST PFT YESTERDAY ((32.08 37.07 28.69)))
(ANAL–REQ BUF–TANK–B 33.56335)
(MIN–INPUT–ANAL BUF–TANK–B YESTERDAY 28.84)
(MAX–INPUT–ANAL BUF–TANK–B YESTERDAY 35.68)
(INPUT–ANAL–HIST BUF–TANK–B YESTERDAY
     ((33.34999 35.68 28. 84) (33.18 35.29999 29.05999)))
(SIM–ANAL–HIGH?? YESTERDAY PFT -327.456 33.25375 33.20999)
(ANAL–REQ BUF–TANK–C 34.7864)
(MIN–INPUT–ANAL BUF–TANK–C YESTERDAY 28.84)
(MAX–INPUT–ANAL BUF–TANK–C YESTERDAY 35.68)
(INPUT–ANAL–HIST BUF–TANK–C YESTERDAY
     ((33.34999 35.68 28.84) (33.18 35.29999 29.05999)))
(SIM–ANAL–HIGH?? YESTERDAY BUF–TANK–C
     -38.5993 35.0352 35.18999)
(SIM–ANAL–LOW?? YESTERDAY BUF–TANK–B
     53.93359 32.76209 32.59)
```

The system attempted to assess whether any of the tanks were stratified by applying the heuristics of Section 5.5.1 but failed to identify any. The consequence (stratified ...) would have appeared if it had been successful.

Plant B, Fault A: consequences virtually identical to the above were added but with the following notable exceptions:

```
(S–I–E ((PFTVOL–MEAS -9.11259)) NIL 1)
(INTERPRET–MODEL S–I–E ((0 NIL (((1 -0.345518)
     (6 -0.634452) (991 -9.11259))) (2.227894))))
```

A single candidate was generated on the third row of the insignificant path error sub-lattice. Again, this contained a single significant error of -9.1*l* in the volume measurement.

Plant A, Fault B: This scenario was included primarily to examine the performance of the control charts because of the current limitation of a single period imposed on the model-based reasoning. It was assumed that the bias was applied three days ago with this diagnosis being performed at the end of 'TODAY'.

goals -

```
(IDH NOT-YET-DONE)
(LOOK-FOR-FAULT S-I-E TODAY (TODAY TODAY))
(LOOK-FOR-FAULT S-T-E TODAY (TODAY TOMORROW))
(TRANSFER-OCCURED TODAY ACCN-T)
(LOOK-FOR-FAULT I-L TODAY (TODAY TOMORROW))
(SIG-GRAD-ON (TWODAY YESTERDAY TODAY))
(CUMUF-GRADIENT 156.2182)
(VIS-GRAD UPTO (156.2182 (TWODAY YESTERDAY TODAY)))
(DO-VIS-GRAD UPTO TODAY (FIVEDAY FOURDAY THREEDAY TWODAY
     YESTERDAY TODAY))
(INV-LOSS TODAY)
(SINGLE-TRANSFER-ERROR TODAY)
(SINGLE-INV-ERROR TODAY)
(INC/DEC-HOLDUP NIL)
(TRANSFER-BIAS NIL)
(COMBIN-3 TODAY)
(LOOK-AT CUSUMP TOMORROW)
(CUSUM-P TODAY)
(CUMUF-TEST TODAY 423.7539)
```

Combination 3 is invoked which in turn requests that the possibility of a transfer_bias, inc/dec_hold-up, single_transfer_error or single_inventory_error be examined. A bias of about 0.16Kg/period is estimated starting one day after the actual occurrence.

sim–asserts - base data,

```
(ERROR TODAY PROD–STOR–A -55.6582)
(ERROR TODAY BUF–TANK–C -117.612)
(ERROR TODAY BUF–TANK–B 28.52734)
(ERROR TODAY PFT -102.767)
(PROBLEM–INV TODAY PROD–STOR–A)
(PROBLEM–INV TODAY BUF–TANK–C)
(PROBLEM–INV TODAY BUF–TANK–B)
(PROBLEM–INV TODAY PFT)
```

sim–asserts - produced by analytical approach,

```
(S–I–E (BUFTCANAL–MEAS -0.440304) NIL 1)
(INTERPRET–MODEL S–I–E
    ((0 NIL (((998 -0.440304))
    ((4 -0.09671) (5 0.09671))
    ((997 -2.84438)))
      (0.037795 0.074822 0.631933))))
```

The model-based reasoning is invoked but identifies a relatively insignificant measurement error of 0.44gm/*l*.

Plant B, Fault B: This has a similar output to the above, but with one major exception best explained by examining the alarms added to the *goals*,

```
(CUSUM–P YESTERDAY)
(CUMUF–TEST TODAY 390.6166)
(MUFTEST–P TODAY 327.1447)
```

Both the muf test and cusum test have alarmed on the same period so combination 3 is not invoked. The important indicator is the alarm generated by the cumuf test.

Plant A, Fault C: as explained in Section 6.6, it was expected that this fault scenario would be misinterpreted because diagnosis is currently restricted to a single period. However it can be seen that the fault can be inferred from the contradictory evidence in the *goals.*

goals:-

```
(VALUE-REQ S-T-E YESTERDAY ((VOLUME 1.0e6 1.0e6)
     (ANALYSIS 73.61585 2.875854)
     (INI-VOL 105.439 4.119049)))
(VALUE-REQ S-T-E YESTERDAY ((VOLUME 10.58294 9.082944)
     (ANALYSIS 29.43035 -2.64964)
     (INI-VOL 102.387 -9.08295)))
(LOOK-FOR-FAULT S-I-E YESTERDAY (YESTERDAY YESTERDAY))
(LOOK-FOR-FAULT S-T-E YESTERDAY (YESTERDAY TWODAY))
(TRANSFER-OCCURED YESTERDAY PROD-STOR-B)
(TRANSFER-OCCURED YESTERDAY ACCN-T)
(LOOK-FOR-FAULT I-L YESTERDAY (YESTERDAY TOMORROW))
(INV-LOSS+SIM-OUT YESTERDAY)
(INV-LOSS YESTERDAY)
(SINGLE-TRANSFER-ERROR YESTERDAY)
(SINGLE-INV-ERROR YESTERDAY)
(LOOK-AT CUSUMP TODAY)
(LOOK-AT MUFTESTP TODAY)
(CUSUM-P YESTERDAY)
(CUMUF-TEST YESTERDAY 283.7919)
(CUMUF-TEST TODAY 452.7478)
(CUMUF-TEST TOMORROW 461.236)
(MUFTEST-P YESTERDAY 336.875)
```

The CUMUF-TEST continues to alarm on subsequent periods. This would not happen if it was a single inventory error. Note that the diagnosis has been performed over an additional period, TOMORROW, to corroborate this.

sim–asserts:– base data,

```
(ERROR YESTERDAY PROD–STOR–A -362.353)
(ERROR YESTERDAY BUF–TANK–C -38.5993)
(ERROR YESTERDAY BUF–TANK–B 53.93359)
(PROBLEM–INV YESTERDAY PROD–STOR–A)
(PROBLEM–INV YESTERDAY BUF–TANK–C)
(PROBLEM–INV YESTERDAY BUF–TANK–B)
```

sim–asserts:- produced by analytical approach,

```
(S–I–E ((PRODTAVOL–MEAS -4.15504)) NIL 1)
   (INTERPRET–MODEL S–I–E
   ((0 NIL (((13 40.46071) (1000 -4.15504))) (0.830394))))
```

A single error is identified, that of a measurement error in product storage tank A. However from above, this is more likely to be a diversion.

Plant B, Fault C: the results obtained were very similar to the above and are therefore omitted.

Plant A, Fault D:

goals

```
(VALUE-REQ S-T-E YESTERDAY PROD-STOR-A
   ((VOLUME 1.0e6 1.0e6)
    (ANALYSIS 73.93459 3.194595)
    (INI-VOL 106.2404 4.590477)))
(VALUE-REQ S-T-E YESTERDAY ACCN-T
  ((VOLUME 11.6225 10.1225)
   (ANALYSIS 29.1271 -2.9529)
   (INI-VOL 101.3474 -10.1225)))
(LOOK-FOR-FAULT S-I-E YESTERDAY (YESTERDAY YESTERDAY))
(LOOK-FOR-FAULT S-T-E YESTERDAY (YESTERDAY TODAY))
(TRANSFER-OCCURED YESTERDAY PROD-STOR-B)
(TRANSFER-OCCURED YESTERDAY ACCN-T)
(LOOK-FOR-FAULT I-L YESTERDAY (YESTERDAY TODAY))
(INV-LOSS+SIM-OUT YESTERDAY)
(INV-LOSS YESTERDAY)
(SINGLE-TRANSFER-ERROR YESTERDAY)
(SINGLE-INV-ERROR YESTERDAY)
(LOOK-AT CUSUMP TODAY)
(LOOK-AT MUFTESTP TODAY)
(CUSUM-P YESTERDAY)
(CUMUF-TEST YESTERDAY 271.6474)
(MUFTEST-P YESTERDAY 324.7304)
```

sims_asserts:- base data,

```
(ERROR YESTERDAY PROD-STOR-A -91.2678)
(ERROR YESTERDAY BUF-TANK-B 401.3547)
(ERROR YESTERDAY PFT -668.361)
(PROBLEM-INV YESTERDAY PROD-STOR-A)
(PROBLEM-INV YESTERDAY BUF-TANK-B)
(PROBLEM-INV YESTERDAY PFT)
```

sim–asserts:- produced by analytical approach,

```
(S–I–E ((PFTVOL–MEAS -18.845) (BUFTBVOL–MEAS 12.44342))
     NIL 4)
(S–I–E ((PRODTAVOL–MEAS -0.563591) (PFTVOL–MEAS -18.8944)
     (BUFTBVOL–MEAS 12.49887)) NIL 3)
(S–I–E ((PFTVOL–MEAS -18.845) (BUFTBVOL–MEAS 12.44342))
     NIL 2)
(S–I–E ((PFTVOL–MEAS -18.8944) (PRODTAANAL–MEAS -0.739998)
     (BUFTBVOL–MEAS 12.49887)) NIL 1)
(S–I–E (((SOLEX1–LOAD 20.30077) (SOLEX2–LOAD 20.30077)))
     (7 8 4 5 6 13 1 2 3 9 10 11 12) 1)
(S–I–E (((SOLEX1–LOAD 38.62499))) (9) 1)
(S–I–E (((SOLEX2–LOAD 38.62499))) (10) 1)
(INTERPRET–MODEL S–I–E
 ((1 (10) (((9 1.943749) (10 38.62499))) (1.039366))
 (1 (9) (((9 38.62499) (10 1.943749))) (1.039366))
 (13 (7 8 4 5 6 13 1 2 3 9 10 11 12)
     (((9 20.30077) (10 20.30077))) (0.521842))
 (0 NIL (((1 -0.49561) (991 -18.8944) (1001 -0.739998)
     (995 12.49887)) ((7 0.203061) (1 -0.494316)
        (991 -18.845) (995 12.44342))
           ((13 21.78896) (1000 -0.563591) (1 -0.49561)
              (991-18.8944) (995 12.49887))
                 ((6 -0.446531) (1 -0.494316)
                    (991 -18.845) (995 12.44342)))
                       (2.060651 2.108644 2.306707
                          2.342489))))
```

The more important fault hypotheses have been added to the assertion base first and therefore follow immediately on from INTERPRET–MODEL. The first hypothesis is for a single significant error of 38.6% in the load of solvent extraction plant A. This is derived from that sub-lattice which only contains this single path error.

Note that the diagnostic system is unable to discriminate between the 3 possible scenarios of either or both of the sol_ex_loads being significantly in error. However a total increase of about 40% is consistently required to explain the measurements. Alternative fault hypotheses are then proposed which are based on more than one significant error.

Plant B, Fault D -

goals -

```
(VALUE-REQ S-T-E YESTERDAY PROD-STOR-B
  ((VOLUME 1.0e6 1.0e6)
   (ANALYSIS 73.61585 2.875854)
   (INI-VOL 105.439 4.119049)))
(VALUE-REQ S-T-E YESTERDAY ACCN-T
  ((VOLUME 10.58294 9.082944)
   (ANALYSIS 29.43035 -2.64964)
   (INI-VOL 102.387 -9.08295)))
(LOOK-FOR-FAULT S-I-E YESTERDAY (YESTERDAY YESTERDAY))
(LOOK-FOR-FAULT S-T-E YESTERDAY (YESTERDAY TODAY))
(TRANSFER-OCCURED YESTERDAY PROD-STOR-B)
(TRANSFER-OCCURED YESTERDAY ACCN-T)
(LOOK-FOR-FAULT I-L YESTERDAY (YESTERDAY TOMORROW))
(INV-LOSS+SIM-OUT YESTERDAY)
(INV-LOSS YESTERDAY)
(SINGLE-TRANSFER-ERROR YESTERDAY)
(SINGLE-INV-ERROR YESTERDAY)
(LOOK-AT CUSUMP TODAY)
(LOOK-AT MUFTESTP TODAY)
(CUSUM-P YESTERDAY)
(MUFTEST-P YESTERDAY 291.3813) 1
```

sim–asserts - base data,

```
    (ERROR YESTERDAY PROD–STOR–A -275.093)
    (ERROR YESTERDAY BUF–TANK–C 210.1196)
    (ERROR YESTERDAY BUF–TANK–B 436.0869)
    (ERROR YESTERDAY PFT -668.361)
    (PROBLEM–INV YESTERDAY PROD–STOR–A)
    (PROBLEM–INV YESTERDAY BUF–TANK–C)
    (PROBLEM–INV YESTERDAY BUF–TANK–B)
    (PROBLEM–INV YESTERDAY PFT)
```

sim–asserts:- produced by analytical approach,

```
(S–I–E ((PFTVOL–MEAS -18.845)
        (BUFTBVOL–MEAS 13.50509)) NIL 1)
(S–I–E (((BUFTBANAL -0.911213)
         (PFTVOL -9.78666)
         (PFTVOL–MEAS -9.51926) (BUFTBVOL–MEAS 13.50509)))
        (7 8 4 5 6 13 1 2 3 9 10 11 12) 3)
(S–I–E (((13 78.43061)
         (CONCT1VOL–MEAS -1.11111)
         (SOLEX1–LOAD 21.42752)
         (SOLEX2–LOAD 21.42752)
         (BUFTCANAL–MEAS 1.390303))
        ((BUFTBANAL -0.911213)
         (PFTVOL -9.78666)
         (PFTVOL–MEAS -9.51926)
         (BUFTBVOL–MEAS 13.50509)))
        (7 8 4 5 6 13 1 2 3 9 10 11 12) 2)
(S–I–E (((CONC–INV 108.1671)
         (PRODSTAVOL–MEAS -1.53238)
         (SOLEX1–LOAD 21.62524)
         (SOLEX2–LOAD 21.62524)
         (SOLEX1–INV -75.3563)
         (SOLEX2–INV -75.3569))
        ((CONC-INV 78.43061)
         (PRODSTAVOL-MEAS -1.11111)
         (SOLEX1-LOAD 21.42752)
         (SOLEX2-LOAD 21.42752)
         (BUFTCANAL-MEAS 1.390303))
        ((BUFTBANAL -0.911213)
         (PFTVOL -9.78666)
         (PFTVOL-MEAS -9.51926)
         (BUFTBVOL-MEAS 13.50509)))
       (7 8 4 5 6 13 1 2 3 9 10 11 12) 1)
(S-I-E (((PFTVOL-MEAS -18.845)
         (BUFTBVOL-MEAS 13.50509))) (13) 1)
(S-I-E (((PFTVOL-MEAS -18.845)
         (BUFTBVOL-MEAS 13.50509))) (9) 3)
(S-I-E (((SOLEX1-LOAD 36.32323))
        ((PFTVOL-MEAS -18.845)
         (BUFTBVOL-MEAS 13.50509))) (9) 2)
(S-I-E (((PRODSTAVOL-MEAS -1.86382)
         (SOLEX1-LOAD 40.77165) (BUFTCANAL-MEAS 1.388878))
        ((SOLEX1-LOAD 36.32323))
        ((PFTVOL-MEAS -18.845)
         (BUFTBVOL-MEAS 13.50509))) (9) 1)
(S-I-E (((PFTVOL-MEAS -18.845)
         (BUFTBVOL-MEAS 13.50509))) (10) 3)
(S-I-E (((SOLEX2-LOAD 36.32323))
        ((PFTVOL-MEAS -18.845)
         (BUFTBVOL-MEAS 13.50509))) (10) 2)
(S-I-E (((PRODSTAVOL-MEAS -1.86382)
         (SOLEX2-LOAD 40.77165)
         (BUFTCANAL-MEAS 1.388878))
        ((SOLEX2-LOAD 36.32323))
        ((PFTVOL-MEAS -18.845)
         (BUFTBVOL-MEAS 13.50509))) (10) 1)
(S-I-E (((PFTVOL-MEAS -18.845)
         (BUFTBVOL-MEAS 13.50509))) (11) 1)
```

```
(S-I-E (((PFTVOL-MEAS -18.845)
         (BUFTBVOL-MEAS 13.50509))) (12) 1)
(S-I-E (((BUFTBANAL -0.911213)
         (PFTVOL-MEAS -18.845)
         (BUFTBVOL-MEAS 13.50509))) (6) 1)
(S-I-E (((PFTVOL -9.78666)
         (PFTVOL-MEAS -9.51926)
         (BUFTBVOL-MEAS 13.50509))) (1) 1)
(INTERPRET-MODEL S-I-E
 ((1 (1) (((6 -0.909553) (1 -9.78666)
           (991 -9.51926) (995 13.50509) (3.130606))
  (1 (6) (((6 -0.911213) (1 -0.494316) (991 -18.845)
           (995 13.50509))) (3.395451))
  (1 (12) (((6 -0.909553) (1 -0.494316) (991 -18.845)
            (995 13.50509))) (3.606022))
  (1 (11) (((6 -0.909553) (1 -0.494316) (991 -18.845)
            (995 13.50509))) (3.606022))
  (1 (10) (((13 23.9064) (1000 -1.86382) (9 2.051777)
            (10 40.77165) (998 1.388878))
           ((4 -0.407746) (6 -0.67321) (9 1.827917)
            (10 36.32323))
           ((6 -0.909553) (1 -0.494316) (991 -18.845)
            (995 13.50509)))
           (1.881204 2.490629 3.606022))
   (1 (9) (((13 23.9064) (1000 -1.86382) (9 40.77165)
            (10 2.051777) (998 1.388878))
           ((4 -0.407745) (6 -0.67321) (9 36.32323)
            (10 1.827917))
           ((6 -0.909553) (1 -0.494316) (991 -18.845)
            (995 13.50509)))
           (1.881203 2.490628 3.606022))
   (1 (13) (((6 -0.909553) (1 -0.494316) (991 -18.845)
             (995 13.50509))) (3.606022))
   (13 (7 8 4 5 6 13 1 2 3 9 10 11 12)
           (((13 108.1671) (1000 -1.53238) (9 21.62524)
             (10 21.62524) (11 -75.3563) (12 -75.3569))
            ((13 78.43061) (1000 -1.11111) (9 21.42752)
             (10 21.42752) (998 1.390303))
            ((6 -0.911213) (1 -9.78666) (991 -9.51926)
             (995 13.50509))) (0.992551 1.159545 2.920035))
    (0 NIL (((6 -0.909553) (1 -0.494316) (991 -18.845)
             (995 13.50509))) (3.606022))))
```

The results are similar to the above but more complicated. The only candidates that contain a single significant error are still those that identify one of the solvent extraction plant loads and the same conclusions can be drawn.

6.8 Sensitivity Analysis

The results described here have all been obtained on the assumption that the diagnostician's perception of variable uncertainty is correct. That is, the plant data sets have been generated with the same variance as those incorporated into the analysis. No formal sensitivity analysis has been performed, partly because of a lack of time but more importantly because of a difficulty in identifying a need. A detailed analysis based on experimentation cannot be extrapolated to another situation.

However a number of *ad hoc* experiments have been carried out where reasonable results have been obtained.

7. CONCLUSIONS

A theory of model-based fault diagnosis has been presented in Chapter 2. Various procedures to facilitate its application have been described in Chapters 3 and 4. Some issues of implementation have been explored in Chapter 5 in the context of a single specific application, that of NRTMA and some results pertaining to this application are given in Chapter 6.

Care has been taken in preparing this thesis to separate theory from methods and methods from application although the precise boundaries are still a moot point. It is intended that the theory should be generally applicable in the engineering domain whilst methods and implementation should be relevant to a reasonable cross-section of applications. However it is accepted that the theory has been developed from the author's own limited viewpoint so is unlikely to be complete.

One particular aspect where this is certainly the case is that it has largely been developed in response to the need to diagnose faults in information poor plants. Other approaches are likely to be more suitable for information rich plants. However it should be of, at least, philosophical interest to someone who thinks that he is diagnosing faults in an information rich plant if only because it should lead him to question whether his plant actually satisfies criteria necessary to support his assumption.

There are two main 'aspects' or 'strands' to the theory: the need for an integrated approach and the need for common sense reasoning about quantitative models. Both have an AI (artificial intelligence) flavour to them: the former through the acquisition and fusion of knowledge and the latter through common sense reasoning. Both are about the specific, making it difficult to draw any general conclusions. For instance, the approach has been shown to be successful in the limited simulations described in Chapter 6 but this cannot guarantee success on the real plant. A sensitivity analysis involving a large number of additional case studies would still not achieve this.

What matters more is that the theory should be deemed to have firm *foundations* both from a philosophical and a pragmatic point of view.

As highlighted by Charniak et al (80), the human being has evolved with well-developed senses whilst the computer has evolved with the ability to store and manipulate vast quantities of largely numerical data. A human being will have great difficulty multiplying 10-digit numbers whereas vision represents considerable complexity to a computer. Similarly, a diagnostician will tend to shy away from quantitative models because of his relatively poor numeracy but computers have no common sense. There is therefore a case to combine the quantitative powers of the computer with the common sense of the diagnostician. That is, there is a case for common sense reasoning about quantitative models since the suppleness of the human mind is well suited to handling uncertainty surrounding such models.

At a more pragmatic level, the theory argues against model-based fault diagnosis as a panacea for fault diagnosis in favour of a data fusion approach where model-based reasoning forms one input. That is, it accepts possible failings of a model-based approach. One outcome of this is that a model-based approach cannot be relied upon to alarm a fault.

The theory is based on two principles, a Principle of Re-Distristribution and a Principle of a Minimum Number of Explanations. By arguing that modelling inaccuracies manifest themselves as an erroneous re-distribution of mass, energy and so on throughout a plant, the former enables the diagnostician to relate to a plant simulation at a more qualitative level whilst maintaining the rigour of mathematical detail. Different faults and model inaccuracies will result in different re-distributions. Some will occur along known paths, others will not. Differences between distibutions observed in the plant and those in the model can then be related to these path and non-path errors. Of particular concern here may be that certain paths may be omitted from the analysis: the system should then identify two non-path faults, one at each end of the *hidden* path. Hence the need to take other knowledge into account and for the Supervisor to adjudicate (Section 2.7).

The Principle of a Minimum Number of Explanations argues that the diagnostician is not interested in estimating model uncertainty but in locating faults. His common sense would lead him to believe that gross re-distributions are more likely to be as a result of a few faults and inaccuracies than of a large number of small ones. He would then view his main task as being that of identifying suitable plausible candidate sets of a few faults and model inaccuracies. This is clearly contentious although arguably how a diagnostician would approach a problem.

The purpose of Chapter 3 is to identify methods of alarming that a fault has actually occurred, and of performing a preliminary diagnosis, without recourse to models. In practice, only one approach is considered, that relating to control charts of primary characteristic variables; quantitative methods are proposed to perform the alarm function whilst qualitative methods are proposed to perform a preliminary diagnosis. Only one primary characteristic variable is considered in the NRTMA application of Chapters 5 and 6, that of plutonium unaccounted for. Section 5.4 shows how various fault categories can be hypothesised by applying a set of rules to 4 different boolean time series derived by testing the time history of material unaccounted for or its cumulative sum. It is unlikely that a unique category will ever be identified; the rules merely identify an ordered list of possibilities. These possible categories must then be assessed one at a time. Some idea of the performance of the approach can be gleaned from the results from a large number of simulations tabulated in Tables 1, 2 and 3. However, again this should be viewed subjectively as the numbers, themselves, cannot be extrapolated to any real plant. Finally turning to the specific fault studies described in Section 6.7, one particular conclusion that can be drawn is that the control chart tests could be improved by incorporating the CUMUF test, explicitly.

The purpose of Chapter 4 is to identify methods of appraising a particular candidate set of suspect faults and model inaccuracies. Again only one method is considered here. This performs two separate tasks. Firstly, the determination of the *most likely* value each element of a particular candidate set would take if it was, indeed, the cause of the symptoms. That is those values that would *best* explain the re-distribution. Secondly, an assessment as to whether or not these values are more likely to occur than any other candidate set. The approach is conventional, being based on subjective probabilities. For computational efficiency a number of candidates are appraised simultaneously. One possible implementation is described in Section 5.7. Based on the limited experience to date (Section 6.6), the approach appears to be successful. This is obscured, in part, by deficiencies in the current implementation. The system was able to diagnose the correct fault on every occasion on which the implementation was complete. On all other occasions, it was able to diagnose something closely resembling the truth.

The main limitation up to now has been one of the computer: it has taken between 3 and 4 hours to appraise just 3 sub-lattices of the application candidate space. The 8Mb of memory available on the VaxStation currently in use, is suspected as being insufficient for the memory intensive lisp/FORTRAN environment installed.

Finally the question arises as to how the approach compares with others proposed elsewhere. The approach here is to tackle models, and their uncertainty, explicitly, whereas other approaches take a more implicit view, attempting to develop techniques which are robust to uncertainty. Clearly any comparison would have, somehow, to specify what is meant by *uncertainty* and this is beyond the scope of the work published here.

A great deal has yet to be done. The main contribution here has been the theory whilst everything else has been more exploratory. Certain avenues have been explored in depth, others at a superficial level whilst others, still remain untouched. For instance, the Supervisor requires considerable research especially into the process of *adjudication*; candidate appraisal over multiple periods needs to be implemented and assessed; the knowledge-base developed so far is only skeletal, considerable insight must be obtained by interacting with a working facility before a robust, practical implementation can be produced; such an implementation would require a diagnostician-computer interface and so on. As outlined in Section 6.2, the method of Chapter 4 may also need revising because it assumes that the symptoms arrive synchronously which need not be the case.

To conclude, the model-based fault diagnosis of information poor plants is still in its infancy, a great deal has yet to be done.

8. REFERENCES

Fault Diagnosis

1 *A. S. Willsky; A Survey of Design Methods for Failure Detection in Dynamic Systems. Automatica, 12, pp 601-611. (1976).*

2 *D. M. Himmelblau; Fault Detection and Diagnosis in Chemical and Petrochemical Processes. Elsevier. (1978).*

3 *L. F. Pau; Failure Diagnosis and Performance Monitoring. Marcel Derrer, New York. (1981).*

4 *R. Isermann; Process Fault Detection Based on Modelling and Estimation Methods - A Survey. Automatica, 20(4),pp 387-404. (1984).*

5 *L.F. Pau; Survey of Expert Systems for Fault Detection, Test Generation and Maintenance. Expert Systems, 3(2). (1986).*

6 *R. Milne; Strategies for Diagnosis. IEEE Transactions on Systems, Man, and Cybernetics, SMC-17(3), pp 333-339. (1987).*

7 *R. Isermann; Experiences with Process Fault Detection via Parameter Estimation, in S. Tzafestas, M. Singh, G. Schmidt (eds), System Fault Diagnostics, Reliability and Related Knowledge-Based Approaches, Volume 1. D. Reidal. (1987).*

8 *P. M. Frank; Fault Diagnosis in Dynamic Systems Using Analytical and Knowledge-based Redundancy - A Survey and Some New Results. Automatica, 26(3), pp 459-474. (1990).*

9 *R. Patton, P. M. Frank, R. N. Clark; Fault Diagnosis in Dynamic Systems: Theory and Applications. Prentice-Hall. (1989).*

10 *E. Y. Chow, A.S. Willsky; Analytical Redundancy and the Design of Robust Failure Detection Systems. IEEE Trans Automatic Control, AC-29(7), pp 603-614. (1984).*

11 *R. Davis; Diagnostic Reasoning Based on Structure and Behaviour. Artificial Intelligence. 24, pp 347-410. (1984).*

12 *J.R. Hobbs, R.C. Moore(eds); Formal Theories of the Common Sense World. Ablex Serieson Artificial Intelligence. (1985).*

13 *K. Watanabe, D. M. Himmelblau; Instrument Fault Detection in Systems with Uncertainties. Int. J. Systems Science, 13(2), pp 137-158. (1982).*

14 *Anil Nigam, R. Bhaskar; Qualitative Reasoning about Engineering Systems Using Dimensional Analysis. Proc. 4th Int Conf on Application of AI in Engineering, Cambridge, UK. (1989).*

15 *M. Herbert, G. Williams; An Examination of Qualitative Plant Modelling as a Basis for Knowledge-based Operator Aids in Nuclear Power Stations. Unicon Conference on Expert Systems, London. (1984).*

16 *A. Ray, M. Desai, J Deyst; Fault Dection and Isolation in a Nuclear Reactor. J. Energy, 7(1), pp 79-85. (1983).*

17 *J.J. Leary, P.J. Gawthrop; Process Fault Detection Using Constraint Suspension. IEE Proceedings, 134(4), Part D, pp 264-271. (1987).*

18 *E. A. Scarl, J. R. Jamieson, C. I. Delaune; Diagnosis and Sensor Validation through Knowledge of Stucture and Function. IEEE Trans Systems, Man, and Cybernetics, SMC-17(3), pp 360-368. (1987).*

19 *R. N. Clark; Instrument Fault Detection. IEEE Trans Aerospace and Electronic Systems, AES-14(3), pp 456-465. (1978).*

20 *J. R. Kemeny; The President's Commission on the Accident at Three Mile Island. Pergamon Press. (1979).*

21 *J. Rasmussen; Models of Mental Strategies in Process Plant Diagnosis, in J. Rasmussen, W. B. Rouse (eds), Human Detection and Diagnosis of System Failures. Plenum Press, New York. (1981).*

22 *J. DeKleer, G. J. Sussman; Propagation of Constraints Applied to Circuit Synthesis. International Journal of Circuit Theory. Vol 8, Part 2, pp 127-144. (1980).*

23 *E.J. Henley, H. Kumamoto; Reliability Engineering and Risk Assessment. Prentice-Hall. (1981).*

24 *D. Welbourne; Alarm Analysis and Display at Wylfa Nuclear Power Station. Proc IEE. 115. pp1726-1732. (1968).*

25 *M. Lind; The Use of Flow Models For Automated PLant Diagnosis, in J. Rasmussen, W. B. Rouse (eds), Human Detection and Diagnosis of System Failures. Editors J Rasmussen, W. B. Rouse. Plenum Press, New York. (1981).*

26 *L. F. Pau; Application of Pattern Recognition to Failure Analysis and Diagnosis, in J. Rasmussen, W. B. Rouse (eds), Human Detection and Diagnosis of System Failures. Editors J Rasmussen, W. B. Rouse. Plenum Press, New York. (1981).*

27 *J. deKleer, B.C. Williams; Reasoning About Multiple Faults. Proc. AAAI-86 Philadelphia,Pa, pp 132-139. (1986).*

28 *J. deKleer, B.C. Williams; Diagnosing Multiple Faults. Atrificial Intelligence, 32, pp 97-130. (1987).*

29 *R. Reiter; A Theory of Diagnosis from First Principles. Dept of Computer Science, University of Toronto, 187/86. (1985).*

Philosophy

30 *B.J.F. Lonergan; Insight - A Study of Human Understanding. Philosophical Library. (1957). phil NE90-lon2*

31 *K.R. Popper; The Logic of Scientific Discovery. Hutchinson, London. (1934, first English Edition 1959, Revised 1972)*

32 *A.W. Burks; Chance, Cause, Reason, 3rd Edition. The University of Chicago Press. (1977). phil LD200*

33 *R. Swinburne; An Introduction to Confirmation Theory. Methuen, London. (1973). phil LD200*

34 *L.J. Cohen; The Philosophy of Induction and Probability. Oxford: Clarendon Press. (1989). phil LD200*

35 *L.J. Cohen; The Probable and the Provable. Oxford: Clarendon Press. (1977). phil LD200*

Probability, Statistics and Filtering Theory

36 *G.J. Klir, T.A. Folger; Fuzzy Sets, Uncertainty and Information. (1988). Maths 0400 1988-k.*

37 *H.J. Larson; Introduction to Probability Theory and Statistical Inference. John Wiley. (1969).*

38 *K.V. Mardia, J.T. Kent, J.M. Bibby; Multivariate Analysis. Academic Press, London. (1979).*

39 *J.E. Freund; Mathematical Statistics, 2nd Edition. Prentice-Hall. (1971).*

40 *G.J. Hahn, S.S. Shapiro; Statistical Models in Engineering. John Wiley. (1967).*

41 *B.D.O. Anderson, J.B. Moore; Optimal Filtering. Prentice-Hall. (1979). Eng KT 170*

42 *G.H. Dunteman; Introduction to Multivariate Analysis. Sage Publications. (1984). Maths 6240*

43 *A.W.F. Edwards; Likelihood. Cambridge University Press. (1972). maths 6002*

44 *H. Jeffreys; Theory of Probability. Oxford Universiy Press. (2nd Edition, 1948). Maths 6000*

45 *G. Shafer; A Mathematical Theory of Evidence. Princeton University Press. (1976).*

Nuclear Materials Accountancy

48 *IAEA Safeguards Glossary; International Atomic Energy Agency, Vienna, IAEA/SG/INF/1 (rev. 1). (1987).*

49 *T.L. Jones, D Gordon; Near Real Time Nuclear Materials Accountancy at Dounreay. IAEA-SM-293/22, Nuclear Safeguards Technology, vol 1. (1986).*

58 *J.L. Jaech; Control Charts for MUF's. Journal of the Institute of Nuclear Materials Management, Vol 2, Part 4. (1974).*

60 *F.J. Walford, A.L. Mills, M.J. Waterman, S.A. Boler; Variations in the Plutonium Inventory of Solvent Extraction Contactors. 9th ESARDA Symposium on Safeguards and Nuclear Material Management. London. (1987).*

61 *J. Howell; An Intelligent Knowledge-Based Approach to Near-Real–time Materials Accountancy. 9th ESARDA Symposium on Safeguards and Nuclear Material Management. London. (1987).*

46 *J. Howell; Systematic Error Identification Via Kalman Filtering. ANS Conference on Safeguards Technology, Hilton Head Island, South Carolina (1983).*

59 *B. J. Jones; Near Real Time Materials Accountancy Using SITMUF and a Joint Page's Test: Comparison with MUF and CUMUF Tests. ESARDA Bulletin Number 15. (1988).*

50 *K. Ikawa, H. Ihara, H. Nishimura, M. Tsutsumi, T. Sawahata; A Near-Real-Time Materials Accountancy Model and its Preliminary Demonstration in Tokai Reprocessing Plant. IAEA-SM-260/136. (1982).*

51 *K. Ikawa, H. Ihara, H. Nishimura, M. Hirata, H. Sakuragi, M. Ido, T. Sawahate, M. Tsutsumi, N. Suyama, M. Iwanaga, J. Lovett; Study of the Application of Near-Real Time Materials Accountancy to Safeguards for Reprocessing Facilities. Japan Atomic Energy Research Institute PNCT N841-33-26. (1983).*

47 *N.S. Russell, M.H. Butterfield, J Howell; Comparison of Retrospective Testing with Statistical Tests in Near-Real–Time Materials Accountancy. Sixth ESARDA Symposium on Safeguards and Nuclear Material Management. Venice, Italy. (1984).*

55 *N. S. Russell; Stochastic Techniques for Time Series with Applications to Materials Accountancy. PhD Thesis, University of Southampton. (1985).*

56 *J.P. Shipley; Decision Analysis for Nuclear Safeguards, in Nuclear Safeguards Analysis: Non-Destructive and Analytical Chemical Techniques. American Chemical Society symposium series, No. 79. (1978).*

57 *T.P. Speed, D. Culpin; The Role of Statistics in Nuclear Materials Accounting: Issues and Problems. Journal of the Royal Statistical Society A, 149, part 4. (1986).*

62 *F.J. Walford; Private Communication (1991).*

Control Charts

63 *I.N. Gibra; Recent Developments in Control Chart Techniques. Journal Quality Technology, 7, pp 183-192. (1975).*

68 *W.A. Shewart; Economic Control of Quality of Manufactured Product. Macmillan, New York. (1931).*

66 *E.S. Pearson; The Application of Statistical Methods to Industrial Standardization and Quality Control. British Standards Institute. BSI 600. (1934).*

67 *B.P. Dudding, W.J. Jennett; Quality Control Charts. British Standards Institute. BSI 600R. (1942).*

64 *B.P. Dudding, W.J. Jennett; Control Chart Technique When Manufacturing to Specification. British Standards Institute. BSI 2564. (1955).*

65 *C. S. Van Dobben De Bruyn; Cumulative Sum Tests: Theory and Practice. Griffin, London. (1968).* *Maths 6200*

69 *E.S. Page; A Test For a Change in a Parameter Occurring at an Unknown Point. Biometrika, 42, pp 523-527.* (1955).

70 *R.A. Johnson, M. Bagshaw; The Effect of Serial Correlation on the Performance of CUSUM Tests. Technometrics, 16(1), pp 103-112. (1975).*

Artificial Intelligence

74 *R.K. Bhatnagar, L.V. Kanal; Handling Uncertain Information: A Review of Numeric and Non-numeric Methods, in L.N. Kanal, J. Lemmer (Eds), Uncertainty in Artificial Intelligence. North-Holland, Amsterdam. (1986).*

80 *E. Charniak, D McDermott; Introduction to Artificial Intelligence. Addison-Wesley. (1985).*

72 *D.J. Spiegelhalter; A Statistical View of Uncertainty in Expert Systems, in W. A. Gale (ed), Artificial Intelligence and Statistics. Addison-Wesley. (1986). comp J21*

79 *J. Fox; Knowledge, Decision Making and Uncertainty, in W. A. Gale (ed), Artificial Intelligence and Statistics. Addison-Wesley. (1986). comp J21*

77 *P.R. Cohen, M.R. Griberg; A Framework for Heuristic Reasoning About Uncertainty. Eighth International Conference on Artificial Intelligence. Karlsruhe. (1983).*

78 *D.J. Spiegelhalter; Probabilistic Reasoning in Predictive Expert Systems, in L.N. Kanal, J. Lemmer (Eds), Uncertainty in Artificial Intelligence. North-Holland, Amsterdam. (1986).*

76 *P. Cheeseman; Probabilistic versus Fuzzy Reasoning, in L.N. Kanal, J. Lemmer (Eds), Uncertainty in Artificial Intelligence. North-Holland, Amsterdam. (1986).*

73 *P.R. Cohen; Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach. Pitman. (1985). Comp J 28*

75 *B.G. Buchanan, E. H. Shortliffe; Rule-Based Expert Systems. Addison-Wesley. (1984).*

71 *A.D. Shapiro; Structured Induction in Expert Systems. Addison-Wiley. (1987).*

81 *L.Y Lam; Lisp Production System. Submitted as part of an MSc in Information Technology, University of Glasogw. (1989).*

Additional

82 *R. Aris; Method in the Modelling of Chemical Engineering Systems. Control and Dynamic Systems, 29. Academic Press. (1979).*

86 *G. L. Steele; Common Lisp: The Language. Digital Press. (1984).*

83 *P.M. Frank; Enhancement of Robustness in Observer-Based Fault Detection. The Safeprocess Conference, Baden-Baden. (1991).*

84 *X.C. Lou, A.S. Willsky and G.C. Verghese. Optimally Robust Redundancy Relations For Failure Detection in Uncertain Systems. Automatica (22), pp 333-344. (1986).*

85 *J. Wuennenberg, P.M. Frank. Model-based Residual Generation for Dynamic Systems with Unknown Inputs. Proc. 12th IMACS World Congress on Scientific Computation. Paris, (2), pp 435 -437. (1988).*

52 *H.A. Dayem et al; Demonstration of Near-Real Time Accounting: The AGNS 1980-81 Miniruns. Los Alamos Report LA-9942. (1984).*

53 *D.J. Pike, A.J. Woods; Statistical Methods in Nuclear Materials Accountancy. Nuclear Safeguards Technology, IAEA-SM-260/135, pp.359-372. (1982).*

54 *D. Sellinschegg; A Statistic Sensitive to Deviations from the Zero-loss Condition in a Sequence of Material Balances. J.Inst. Nucl. Mater. Manag., 11 (4), pp 48-59. (1982).*

APPENDIX: Time Series Studies

A number of simulations have been carried out to gain insight into how the CUSUM test performs by itself, firstly with no fault and then with fault scenarios which would cause the patterns described in Section 5.4. A constant throughput plant was chosen with 1% random errors and a ratio of the physical inventory to throughput to produce a serial correlation, ρ, of either -0.4 or -0.45 in the MUF time series. The magnitudes of the faults were specified as multiples of the error (W) in the plant throughput and not of σ_{MUF}, as is the convention in the assessment of detectors. The equivalent proportions of σ_{MUF} can be obtained by multiplying the proportion of plant throughput by the factor $\surd(0.5+\rho)$.

The no fault situation was examined by performing 300,000 simulations for each of 3 sets of h, k and ρ. A particular test sequence was stopped immediately it alarmed. The following results were obtained

	h = 3.059 k = 0.063 ρ = – 0.4		h = 5.00 k = 0.00 ρ = – 0.4		h = 3.059 k = 0.063 ρ = – 0.45	
Period	% Alarms	Credibility	% Alarms	Credibility	% Alarms	Credibility
1	0.17	1.00	0.00	1.00	0.18	1.00
2	0.49	1.00	0.00	1.00	0.42	1.00
3	0.95	0.99	0.00	1.00	0.58	0.99
4	1.47	0.98	0.03	1.00	0.81	0.99
5	1.97	0.98	0.09	1.00	1.02	0.99
10	3.33	0.96	1.17	1.00	1.85	0.98
15	3.21	0.95	1.03	0.99	2.09	0.97
20	2.55	0.95	-	-	1.96	0.97

Table A1: False Alarm Rates in Terms of % Alarms to occur for the 1st time and Credibility. (300,000 simulations).

The '% Alarms' column is the percentage of tests that alarm for the **first** time on that particular period and 'Credibility' is the proportion of the tests, that are actually applied on a particular period, that do not alarm. It can be seen that, initially, the false alarm rate increases until it reaches a maximum. The power to alarm any fault scenario then remains constant. That is it does not depend on how much later the fault occurs [55]. (The apparent decrease in the % Alarm rate is as a result of fewer tests being applied.) As is to be expected increasing h, increases the credibility of the test and that the effect of serial correlation is considerable.

A number of simulations (100,000/case) were carried out to examine the performance of the test whilst the false alarm rate is increasing. Faults relating to both Pattern 1 and Pattern 2 were examined and the same effect was observed in both sets of results. A summary of the results obtained for Pattern 2 are given below in Table A2. A single physical inventory measurement error was simulated, firstly on period 1, then on periods 5, 10 and 15. The percentage of tests that alarmed for the **first** time on a particular period were recorded. Note that the test was restarted if it alarmed prior to the fault being applied.

PATTERN 2						
Magnitude % throughput	Period Period Applied	+ 0	+ 1	+ 2	+ 3	+ 5
4	1	3.3	3.0	3.1	3.3	3.3
	5	13.2	3.3	3.6	3.8	3.7
	10	17.7	4.1	4.0	3.9	3.5
	15	18.7	4.4	4.1	-	-
6	1	11.1	7.9	6.3	5.2	3.8
	5	31.3	5.4	4.7	4.0	3.1
	10	37.0	5.5	4.6	3.9	2.9
	15	37.5	5.6	4.5	-	-
8	1	27.6	15.2	8.7	5.6	3.0
	5	55.3	7.4	4.8	3.4	2.0
	10	60.5	6.7	4.3	3.0	2.3
	15	61.2	6.6	4.3	-	-
12	1	74.7	14.5	3.9	1.6	0.8
	5	92.1	3.7	1.3	0.6	0.2
	10	93.0	3.4	1.1	0.5	0.2
	15	92.9	3.4	1.1	-	-

Table A2

It can be seen that the power increases with increasing false alarm rate and that the majority of alarms arise on the period the fault is applied.

A general survey was then carried out to examine the test's performance when the various faults occur. All faults were applied from period 10 to give the false alarm rate time to settle. The assessment was based on 10,000 simulations in most cases because of the computational load. The results obtained, Tables A3, A4 and A5, are therefore only suitable for qualitative comparison.

	PATTERN 1						PATTERN 2						PATTERN 5	
h	3.059		3.059		5.00		3.059		3.059		5.00		3.059	
k	0.063		0.063		0.00		0.063		0.063		0.00		0.063	
ρ	– 0.4		– 0.45		– 0.4		– 0.4		– 0.45		– 0.4		– 0.4	
e period	6*	12*	6	12	6	12	6*	12*	9	18	6	12	6	12
10	37.1	93.1	17.0	68.1	6.7	52.4	37.0	93.0	39.5	96.7	6.4	51.7	92.3	100
11	13.6	4.0	9.9	14.0	4.3	13.9	5.5	3.4	4.4	1.4	0.3	3.0	0.6	0
12	7.2	1.0	6.8	5.3	3.9	7.2	4.6	1.1	3.8	0.6	0.5	2.9	0.4	0
13	4.6	0.4	5.0	2.5	3.7	4.3	3.9	0.5	3.0	0.3	0.8	2.9	0.4	0
15	2.6	0.2	3.0	0.9	3.0	2.0	2.9	0.2	2.2	0.1	1.2	2.3	0.3	0

* 100,000 simulations

Table A3

	PATTERN 3											
h	5.00				3.059				3.059			
k	0.00				0.063				0.063			
ρ	– 0.4				– 0.4				– 0.45			
e period	0.5	1.0	2.0	3.0	0.5	1.0	2.0	3.0	0.5	1.0	2.0	3.0
10	0.4	0.4	0.8	1.6	3.9	4.6	6.8	12.2	2.1	2.4	3.1	5.5
11	0.6	0.8	3.0	6.8	4.7	6.3	14.3	27.4	2.3	3.5	7.4	14.7
12	0.9	1.7	6.7	19.3	5.0	7.8	20.5	31.4	2.7	4.1	12.0	24.3
13	1.2	2.9	12.6	28.2	5.2	9.9	20.2	18.7	3.2	5.8	15.7	24.7
14	1.6	4.6	16.7	23.8	5.5	10.0	16.0	7.7	3.3	7.0	16.9	17.2
15	2.2	6.2	18.1	13.6	5.6	9.5	10.2	2.1	3.7	7.0	14.1	8.5
16	2.6	6.7	15.7	5.0	5.8	9.2	6.2	0.4	3.8	6.2	11.4	3.6
17	2.8	8.4	11.2	1.3	5.2	8.4	3.0	0.1	3.8	5.3	8.2	1.0

Table A4

	PATTERN 4								
	period e	10	11	12	13	14	15	16	17
h = 5.00	4	2.6	2.0	2.2	2.2	2.0	2.1	1.6	1.7
k = 0.00	6	7.0	4.4	3.7	3.2	3.3	2.8	3.6	3.1
h = 3.059	4	17.7	9.3	7.1	5.3	4.7	4.0	15.7	9.5
k = 0.063	6	37.4	13.0	7.1	5.5	3.8	4.0	33.7	13.3

↑ ON

↑ OFF

Fault applied on Period 10
retracted on Period 16

Table A5