ECONOMIC SCHEDULING IN ELECTRIC POWER SYSTEMS
A MATHEMATICAL MODEL FOR THE U.A.E.

BY

DARWISH M K F AL-GOBAISI BSc C.Eng MIEE

A doctoral thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of the Loughborough University of Technology, 1988

Supervisor: Professor A. C. Bajpai OBE C.Eng Companion IEE
Head of Department of Engineering Mathematics
Loughborough University of Technology

ABSTRACT

Research work reported in this thesis is directed at developing and applying mathematical modelling and optimization techniques to problems involved in the economic operation functions in electric power systems, with special reference to the Abu Dhabi Electric System in the United Arab Emirates. The three main areas treated in this thesis are; power system load forecasting, fuel cost model parameter estimation, and economic dispatching algorithms. The three areas are essential ingredients required for a successful implementation of economic operation strategies in systems such as that of Abu Dhabi.

Research work reported in the area of power system load forecasting involves short term prediction of the load on the system. Conventional techniques in the general area of forecasting are explored and their performance in forecasting the actual load on the system is evaluated. The Box-Jenkins methodology for time series analysis is discussed and appropriate models suited to the Abu Dhabi system are obtained. Advanced system-theoretic-based prediction techniques such as Kalman filtering and the Instrumental Variable method are treated. Particular attention is paid to the problem of noise statistics characterization for Kalman filtering implementation. The work reported involves a comparative evaluation of the performance of each method, based on the developed models of the Abu Dhabi Electric load characteristics.

Evaluating the fuel cost model parameters necessary for economic scheduling is the second ingredient treated in this thesis. Here conventional least squares, weighted least squares and recursive techniques are considered in dealing with the 56 units of the Abu Dhabi system. The thesis reports on the finding of infeasible parameter estimates which has not been reported before. Diagnostic tools are developed to identify the possible data pairs that cause the infeasibility. The thesis also suggests a constrained parameter estimation technique to overcome this problem.

A second important contribution in the area of fuel cost model parameter estimation is related to the application of robust parameter estimation techniques using the Iteratively Reweighted Least Squares techniques for the first time. In this thesis, results using eight weighting functions are reported. Evaluation of the results is carried out in terms of accuracy as well as in terms of the resulting economic scheduling strategy's cost.

Economic dispatching of the system is considered from an efficient implementation point of view. The Generalized Reduced Gradient method and the Lambda Iteration method are treated and results of their application are detailed. In the course of this investigation a newly developed method was discovered. The method, which is called the Lambda Aggregation method, has been shown to be an extremely fast technique for dispatching and does not require any initial guess for its implementation.

It is hoped that the work in this thesis will be of interest and significance to the neighbouring countries in the Gulf States. Finally, recommendations and suggestions for further research are made.

KEYWORDS

Economic Dispatching, Economic Scheduling, Power System Optimization, Electric Power Systems, Electric Load Forecasting, Load Characteristics, Fuel Cost Parameter Estimation, Recursive Parameter Estimation, Kalman Filtering, Instrumental Variable, Robust Estimation.

ACKNOWLEDGEMENT

The author wishes to express his deepest gratitude to Professor A. C. Bajpai for his constant guidance and for imparting much of his knowledge and wisdom during the course of developing and evolving this thesis. The author's interest in the elegance and usefulness of mathematics has been profoundly influenced to a significant extent since his undergraduate days by Bajpai's unique approach in the field of mathematical education of engineers. It is a privilege to learn and receive guidance from such a mentor and authority on mathematics and its applications. He inspired me and will always be remembered for leading me to a space from which I could think clearly. If it were not for the patience and understanding that he has shown me despite his extremely busy schedule I could not have completed this work, and for all of these I will be forever grateful.

The author greatly recognises the support of H.E. Sheik Suroor Al-Nahayan, Chairman of W.E.D., Mr. Shaiba Khamis, Member of the Executive Council, and Mr. Sayah Moussa, Under Secretary of W.E.D., during the course of this work.

The author sincerely thanks Mr. Bushara Makkawi and Mr. Adel Radif who had to undertake a major part of his administrative responsibilities during this work.

Discussions with Dr. P.N.P. Singh on some aspects of this work were useful and are hereby acknowledged.

Programming assistance given by Mr. I. Woldai and Mr. J. Goode was helpful in this work.

The author wishes to acknowledge the patience and understanding of his family while this work was in progress.

The author thanks Mr. M. Zaffar for his patience, diligence and perseverance in typing the final manuscript.

Over and above all I thank the Almighty who guided me along the right path.

In the quest for perfect knowledge we encounter problems for which we seek solutions. We are of imperfect nature so are our perceived solutions. However seemingly different are the paths, it is always the same goals that we seek to reach the unique optimum. The inquisitiveness gift of our Creator is bestowed upon us to achieve enlightenment.

D. Al-Gobaisi

Dedicated to Fatima

and

my beloved Children

TABLE OF CONTENTS

CHAPTER I
INTRODUCTION

1.1 The Scenario

The planning and operation of an electric power system involves a number of challenging tasks that require a skilful blend of mathematical and system theoretic tools to arrive at strategies that ensure reliable and economic functioning of the system. These tasks invariably involve computer-aided computations that demand fast and efficient algorithms.

This thesis is concerned with the application of mathematical and system theoretic concepts to problems of load forecasting, fuel cost parameter estimation, and economic dispatching in an electric power system. While the techniques discussed and developed in this thesis are applicable to electric power systems in general, computational testing and results are reported with reference to the United Arab Emirate's Abu Dhabi electric power system.

In the course of this investigation, techniques from the areas of forecasting, time series analysis, state and parameter estimation, identification, optimization and computational methods for the solution of non-linear equations, have been investigated and employed to achieve the desired result. In a number of instances, new contributions to the state of the art in specific areas have been made. These are summarized in Chapter 7 of this work. An outline of the work conducted under this research programme is given in the next section.

The impetus for this thesis work is due to a number of advanced Seminars on "Engineering Mathematics and its Applications" conducted at WED under the direction of Professor Bajpai, an eminent authority on this subject. The author became convinced that the mathematical modelling approach is the best vehicle to improve the planning and operation of the generation system in Abu Dhabi.

1.2 Scope of the Thesis

In Chapter 2, a presentation of relevant background material is given. The general requirements of electric power systems planning and operation are detailed in this chapter. A brief description of the proposed Gulf-interconnection is given and this is followed by a discussion of the Abu Dhabi electric power system in terms of generation resources and load characteristics. Economic scheduling requirements in an electric power system are also presented.

A requirement of economic scheduling is an accurate prediction of the system power demand. Chapter 3, discusses "Load Forecasting", beginning with a detailed review of previous work in this area. A listing of the data base for forecasting power system load for the Abu Dhabi electric power system is given, and then we discuss General Exponential smoothing as applied to seasonal load forecasting using trigonometric functions, where a brief resume of the technique is given, and this is followed by results summarizing computational experience in both mid-term and short-term forecasting of power load on the Abu Dhabi (AD) system.

In Sections 3.5 and 3.6, the use of Winters' seasonal forecasting via the additive/multiplicative models is considered. Following a brief review of the development of the technique, computational experience with mid-term and short-term forecasting for the AD system is given. A similar treatment of the multiplicative model case is then given.

One powerful technique in forecasting is the Box-Jenkins methodology which is reviewed and computational results are reported.

Kalman filtering for prediction in systems described by state space models is considered, beginning with a brief summary. A state space model of the load forecasting function is then developed. A unique feature of the model is that it is a time-invariant model, and therefore offers computational time savings. One of the drawbacks of Kalman filtering is that it requires accurate knowledge of model parameters as well as the noise statistics. Many techniques of adaptive filtering have been proposed. A comprehensive review of the various techniques is offered, and we conclude with computational experience including a new approach to noise statistics evaluation using a simplified maximum likelihood approach.

Load forecasting by recursive weighted least squares using exponential weighting in the past is treated next. Moreover we deal with the method of Instrumental Variable for system identification as applied to the forecasting task. Finally in this chapter a comparative evaluation of the seven techniques is given.

Chapter 4, is devoted to the second ingredient of the economic dispatch function, namely that of fuel cost model parameter estimation. Here, following a review of previous work, the problem is formulated in Section 4.3. The three subsequent sections briefly treat theoretical foundations of the Weighted Least Squares technique and recursive parameter estimation.

In the course of applying parameter estimation techniques to data of the fifty-six units in the Abu Dhabi system, a challenging problem that has not been documented in the literature was discovered. The problem is that one of the parameters must be positive to obtain meaningful optimization results. The causes for this phenomenon were investigated using tools from the area of parameter estimation. We were also led to formulate the problem as one of constrained parameter estimation for which a non-linear programming approach was proposed. Computational experience is reported to conclude the chapter.

Under the title of "Economic Dispatch Studies", Chapter 5 initially includes a comprehensive historical survey of the area. A review of the economic dispatch optimization problem formulation and its variational solution is then given, and three techniques for solving the problem were considered. The application of the generalized reduced gradient method, the Lambda Iteration method and the Lambda Aggregation technique are discussed in this chapter. The latter technique was developed by the author and has proven to offer computational savings when applied to the Abu Dhabi system. A major contribution of this thesis is the proposal of many new efficient initial guess generating procedures that are also reported in this chapter.

In Chapter 6, the problem of parameter estimation for fuel cost representation is reconsidered. Here non-least squares objectives for estimation are discussed. The use of non-quadratic objectives such as least absolute residuals and robust estimation are considered and we discuss the iteratively reweighted least square method as well as the associated weighting functions. Computational results and a comparative evaluation of the techniques in terms of the economic dispatch results are given. It is suggested that this chapter contains significant new results not reported so far in the power system field.

Chapter 7 is devoted to concluding remarks summarizing the contributions of the thesis and recommendations for future work.

CHAPTER 2
GENERAL CONSIDERATIONS

2.1 Introduction

This chapter presents the background material which underpins this thesis. In Section 2.2, a discussion of power systems planning and operations requirements is given. Since the techniques developed in this thesis are applied to a utility in the Arabian Gulf area, a brief discussion of the proposed Gulf inter-connection, which was approved in the middle of 1987 by the Executive Council of the Arab Gulf Cooperation Council (AGCC), is presented in Section 2.3. This is followed by a short discussion of Abu Dhabi's electric power system, its load characteristics and unique features.

Section 2.6 is devoted to an introduction, concentrating mainly on definitions to topics within the broad area of economic scheduling of electric power systems, that are covered in this thesis. It should be noted that literature reviews, related to each topic discussed in this thesis, are found at the beginning of each chapter with the exception of this chapter.

2.2 Functions in Power Systems Planning and Operation

In an existing electric power system, the functions of expansion planning and operation offer the electric power systems engineer many challenging and complex tasks. A basic requirement is knowledge of the system power demand over the planning time horizon of interest [1]. The time horizon can vary in duration from as long as 10 years to as short an interval as 10 seconds, depending on the required lead time for the function to be performed. Planning and operation functions are designed so as to meet the demand for electric power in the most economical manner, while simultaneously ensuring safety and reliability (continuity) of service. These objectives are all inter-related and are of equal importance in the day-to-day operation of an electric power system.

Figure (2-1) shows the time horizons normally encountered in common power system planning and operational scheduling functions as well as the discretization intervals that are considered. Expansion planning for generation, transmission and distribution facilities requires committing capital investment and may need 10 years' lead time. Long term generation planning to determine the required fuel supply over a number of years in the future is another function that has to contend with economic as well as logistics aspects.

Operational planning of electric power systems may require a lead time of up to 5 years. A load forecast of up to 1 year is required for scheduling unit maintenance and generation targets. Accuracy in forecasting both at the planning and operational planning levels is important, especially in situations requiring long delivery periods and costly order processing.

Short term hourly load forecasting for up to one week is required to determine unit commitment strategies on an hourly basis [2]. Given a unit commitment schedule, the economic dispatch function is carried out for a time horizon of between 10 minutes to one hour to ensure that generation is allocated to achieve overall minimum fuel cost in the system. The shortest updating interval of 1 to 10 seconds is encountered in load frequency control where generation levels are adjusted continuously by the governor action to follow the load variations while minimizing the frequency deviations from a reference system frequency.

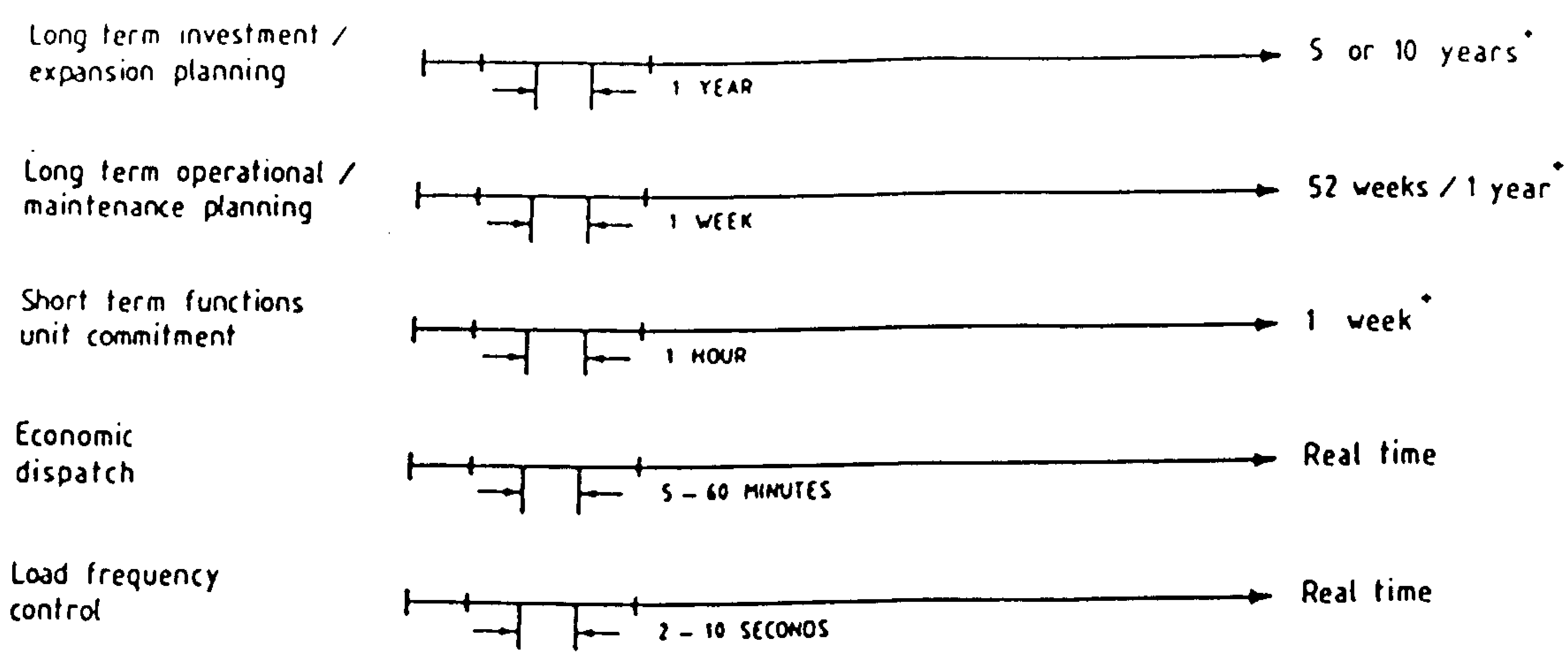


Figure (2-1) Time Horizons for Planning and Scheduling Activities

Electric power systems in the Arabian Gulf are in a unique situation, since the utility is invariably charged with both electricity production, and water production by desalination. This makes the planning and operation task a more complex one since the two activities are inter-dependent.

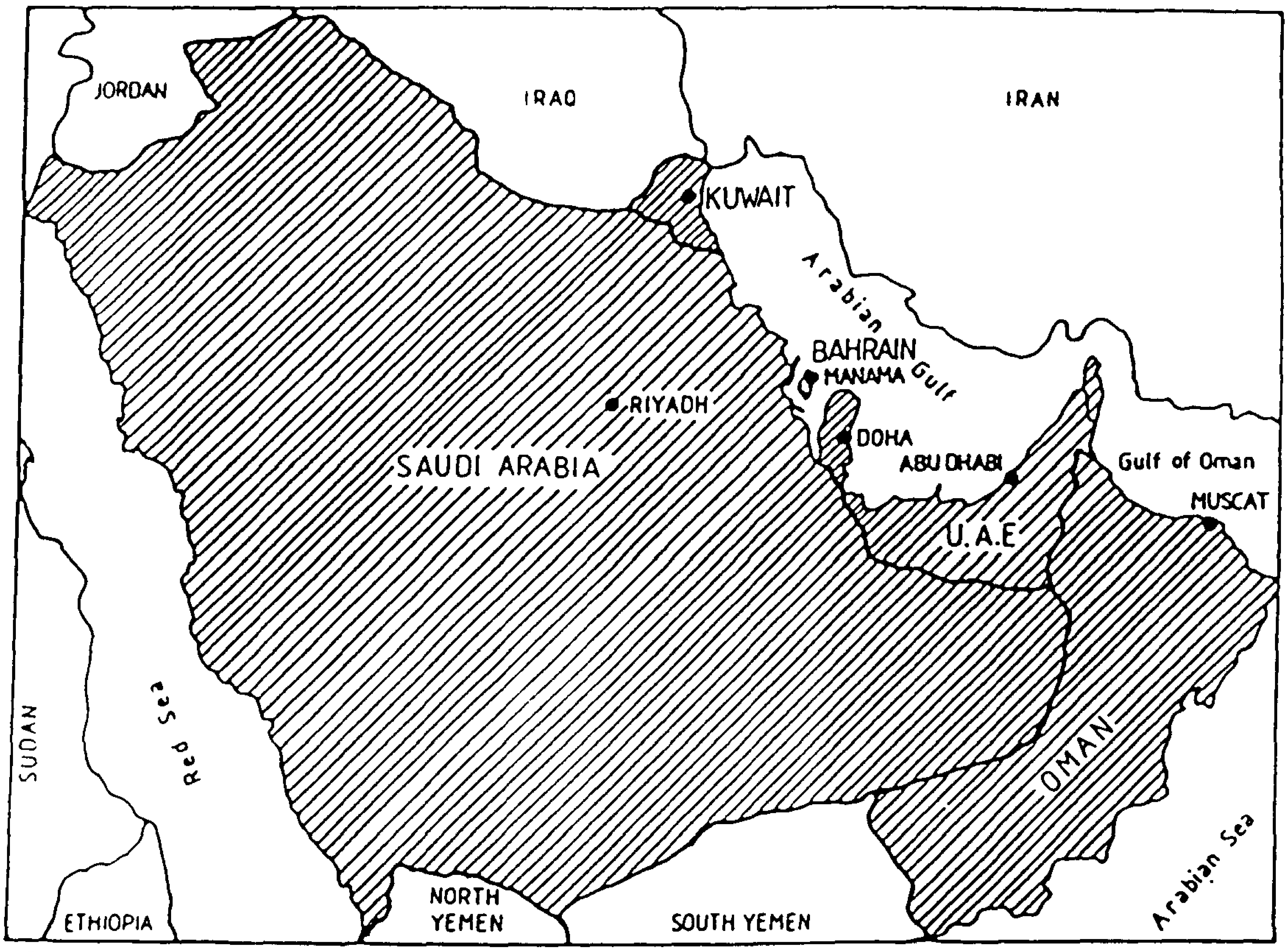
Due to the rapid growth of electric power systems in the Arabian Gulf region over a relatively short period of time, the long range load forecasting and planning of functions were predominantly conducted by outside consulting organizations. In the meantime, operational planning has not received sufficient attention and was left mainly to plant operators. It is important to note that the necessity of operating the electric power system most economically was not considered in an era of abundant fuel supply and oil-related income. This situation is rapidly changing following the realization that development funds are limited.

2.3 The Gulf Interconnection

On May 25, 1981, six states bordering on the Arabian Gulf announced the formation of the Arab Gulf Cooperation Council (AGCC) with the objective of preserving stability and security of the region and advancing the welfare of its people. The AGCC member states are; Bahrain, Kuwait, Oman, Qatar, Saudi Arabia and the United Arab Emirates, and lie between the longitude of 34°E & 60°E and latitude between 17°N & 33°N. The solar time difference between the extremities of the region is about $1\frac{3}{4}$hr. A map of the AGCC region is shown in Figure (2-2).

The AGCC region is predominantly a desert land without rivers. The winter climate is quite mild, and space heating is required only for a few months in some parts of Kuwait, Oman and Saudi Arabia. The summers are hot, with temperatures ranging from 35°C to 50°C and humidity can reach up to 100% in parts of the region. Air conditioning is used extensively throughout the region during summer, and this load component dictates the summer peak electric power demand.

ARAB GULF COOPERATION COUNCIL _ STATES



SAUDI ARABIA, KUWAIT, BAHRAIN, QATAR, UNITED ARAB EMIRATES, AND OMAN

Figure (2-2) A Map of the States of the
Arab Gulf Cooperation Council

The main sources of potable water in the area are underground aquifers and desalination of sea water. The load due to desalination plants is significant and non-seasonal because of the steady water requirements throughout the year.

The peak electric power demand in the region can thus be seen to take place in the summer period of July-August, during the day.

Water and electric power services are provided at highly subsidized rates by governments of the AGCC states. Electric power generation, transmission and distribution is done through a number of isolated systems. Within the same state, isolated service areas exist due to the physical distances between major generating sources (close to major population centres) and remote areas characterized by low demand. Interconnecting the regions' isolated systems has not been a major consideration in the planning of electric systems in the AGCC states until just recently.

Interconnecting the AGCC electric power systems can result in the following benefits;

1. The reliability of all systems is improved for the same generation configuration.

2. Lower installed generation capacity requirements due to the increase in system size. This results in cost savings both in terms of capital cost and operational and maintenance costs.

3. In the interconnected system lower spinning reserve is required than that required by the individual systems. This results in operational cost savings.

4. Strong interconnections can limit the effects of and facilitate recovery from certain multiple outages that can otherwise have pronounced effects on individual systems.

5. Normally, an interconnected system can be scheduled more economically than individual systems, since optimal allocation of generation to the most economic units is performed independently of the geographic location.

6. System interconnection allows long term supply contracts between neighbouring utilities. These contracts enable one utility to postpone generation expansion in favour of importing from a neighbour with surplus capacity. Similar arrangements for short duration can improve maintenance scheduling also.

The feasibility of interconnecting the electric power systems of the Arabian Gulf states was the subject of the "Gulf Interconnection Study" commissioned by the Gulf Cooperation Council in 1984 [3]. The study, concluded late in 1986, established the feasibility of the interconnection and recommended a scheme shown in Figure (2-3) for implementation. The AGCC accepted the study's recommendations in principle in mid 1987.

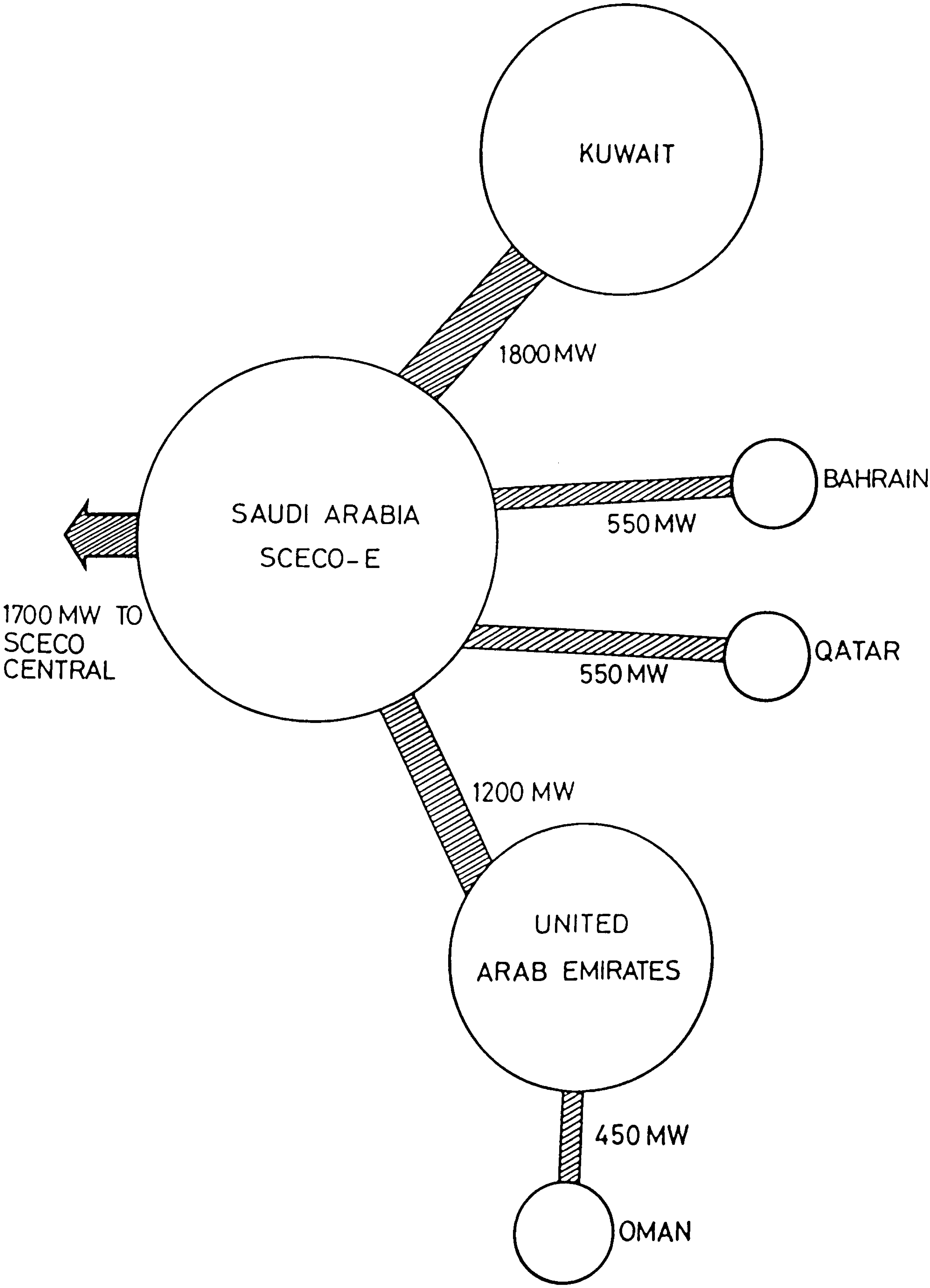


Figure (2-3) Gulf Interconnection Proposal

2.1 The Abu Dhabi System

The electric power system of Abu Dhabi of the UAE (see map of Figure 2-4) consists of; a main power transmission network of 180 km of 220 kV overhead lines, secondary transmission network of 132 kV including 138 km of underground cables, and about 32 km of overhead lines. A geographical map of the Abu Dhabi, Water and Electricity Department [WED] system is shown in Figure (2-5). A map of the projected UAE electric power system is shown in Figure (2-6).

The total power generation installed capacity reached 1955 MW in 1987. The largest site is at Umm Al Nar West which consists of ten steam turbine units with a total rating of 830 MW commissioned between 1979 and 1987. Umm Al Nar East has four gas turbines with a total rating of 253 MW. Abu Dhabi Power Station has 209 MW of gas turbine, 140 MW of steam turbine. The Diesel Power Station at Shabia has one 17.9 MW gas turbine unit, is an older site that was fully commissioned in 1966, and its capacity has been decreasing since 1973 from a peak rating of 30 MW. Al-Ain Power Station has a capacity of 373 MW, and that of Baniyas is 124 MW. A small and separate Diesel Power Station exists at Saadiyat, with a capacity of 7.0 MW, feeding a local network.

A new power station is planned at Al Taweelah to meet system demand in the late eighties. In the first phase a total of 255 MW (3x85 MW) gas turbines will be installed. In phase 2 and 3 a total of 2000 MW of steam turbine power will be installed by the year 2000.

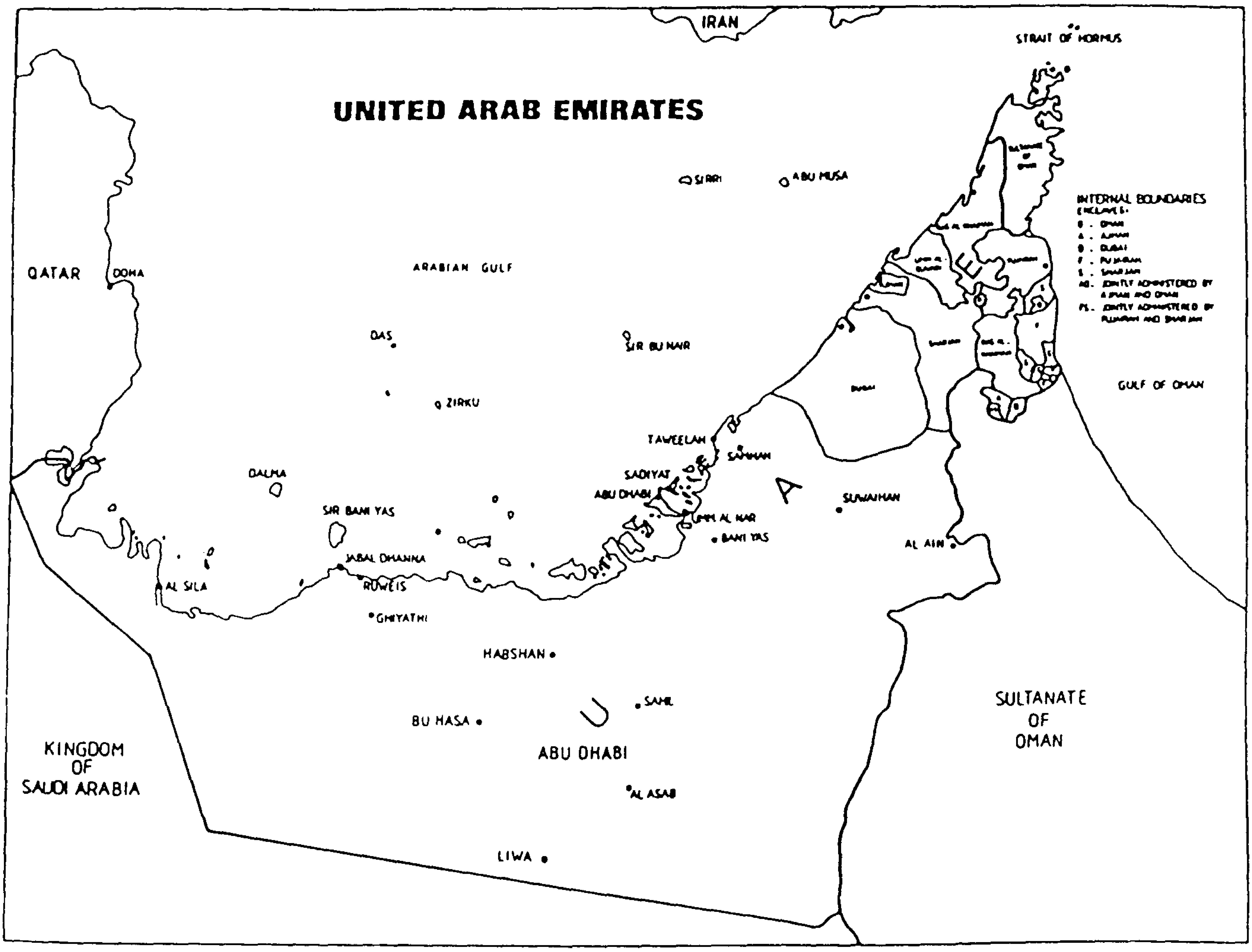


Figure (2-4) Geographical Map of the United Arab Emirates

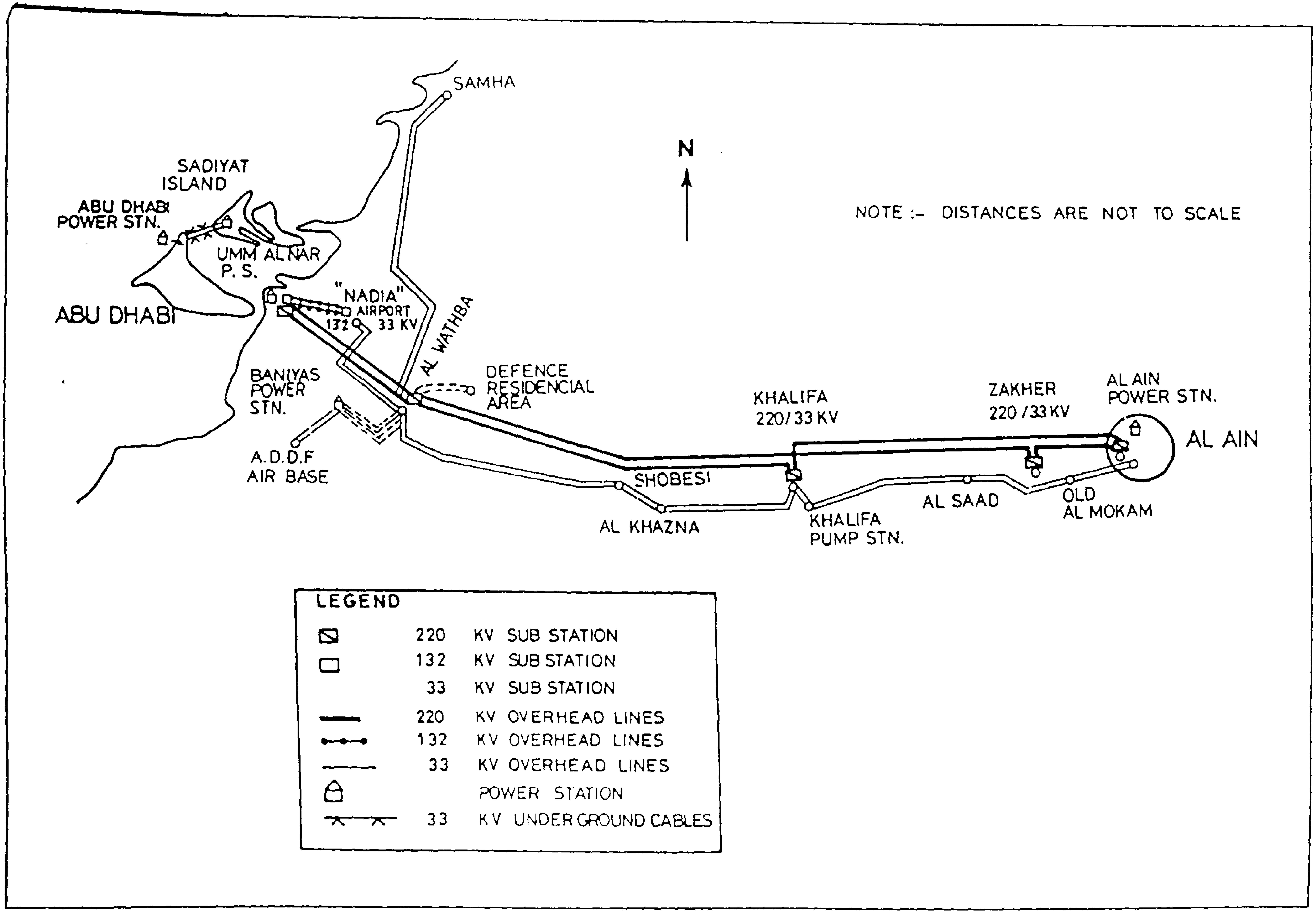


Figure (2-5) Abu Dhabi Electrical Power System

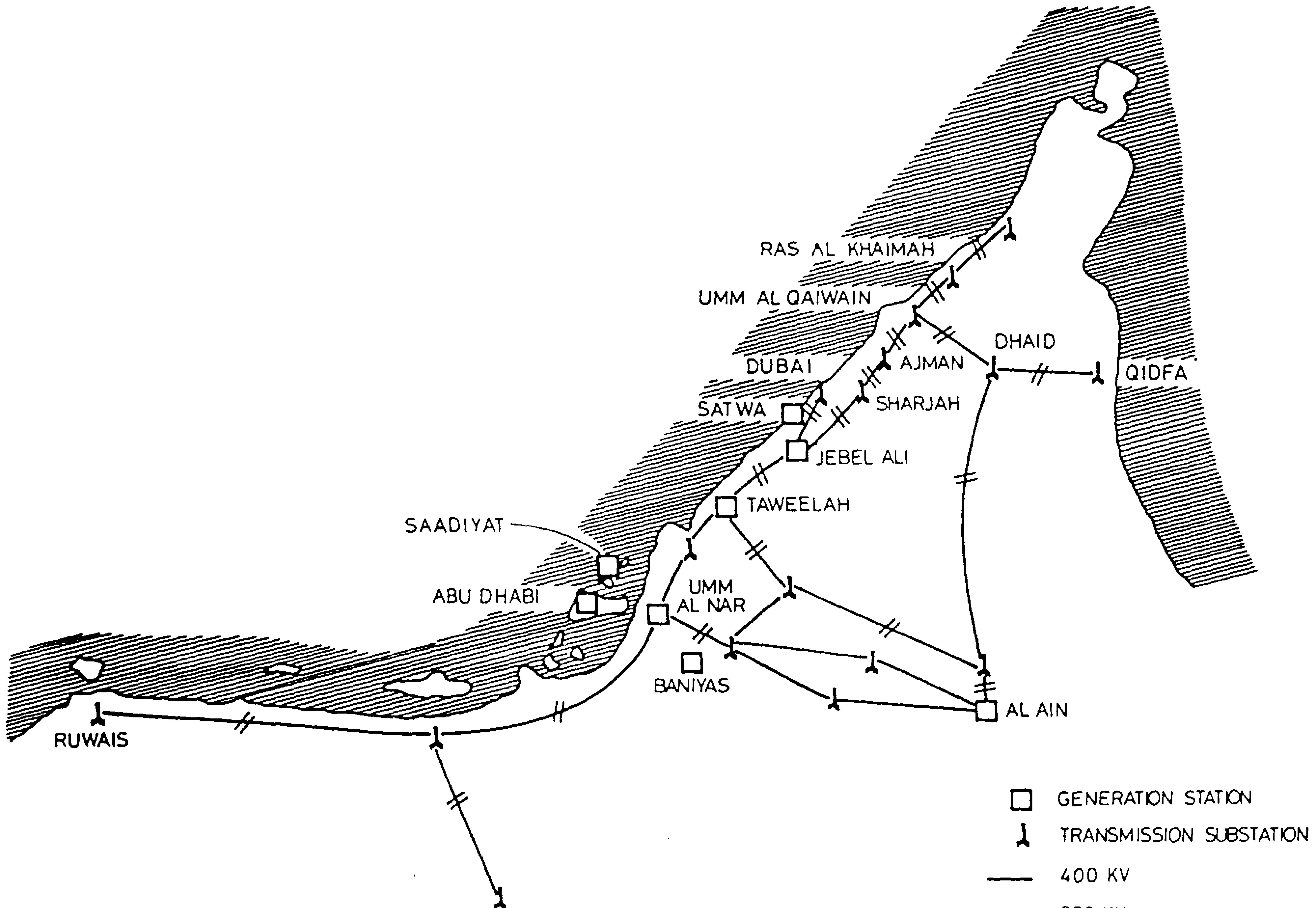


Figure (2-6) UAE Projected Electric Power System

As is the case with all AGCC systems, the WED system in Abu Dhabi is charged with the concurrent production of electric power and desalinated water. Water desalination plants use exhaust steam from steam turbines or exhaust gases from gas turbines. These lower temperature sources offer economic advantages. Three types of cogeneration schemes are commonly used:

1. Back pressure steam turbines with desalination plants. A typical schematic diagram is shown in Figure (2-7)(a).

2. Controlled extraction steam turbines with desalination plants combination as shown schematically in Figure (2-7)(b).

3. Gas turbines with exhaust heat recovery (EHR) boilers and desalination plants combination as shown in Figure (2-7)(c).

From an integrated system point of view, the operation of a typical AGCC power and water utility can be visualized as shown in the functional block diagram of Figure (2-8). The system consists of an energy input system feeding both the electric power system and the water system. As shown, the energy input system consists of gas turbines, steam generators, EHR boilers, and interacts with the production systems for electric power as well as water.

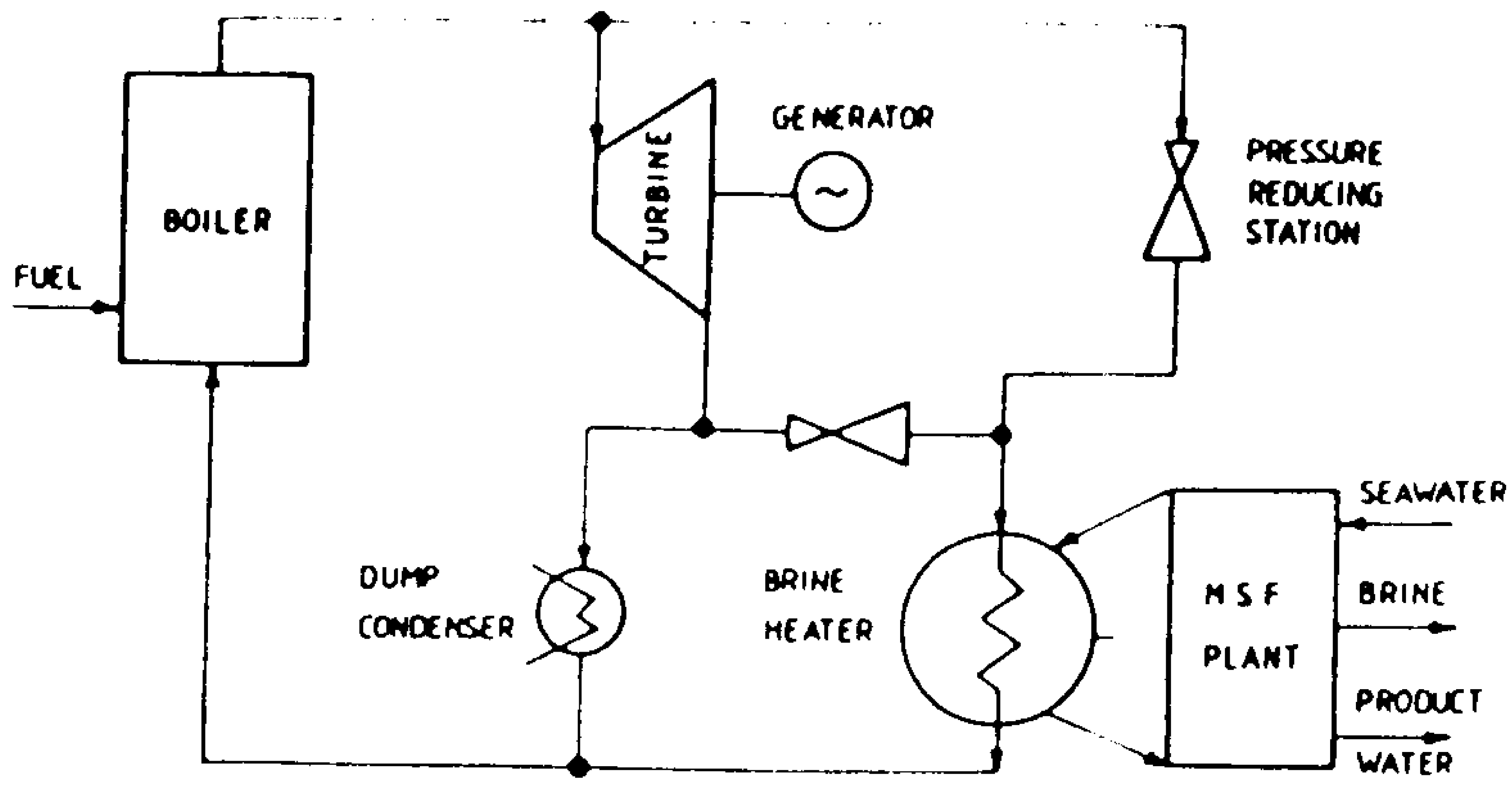


Figure (2-7) (a)
Back Pressure Turbine _ Desalination Plant Combination

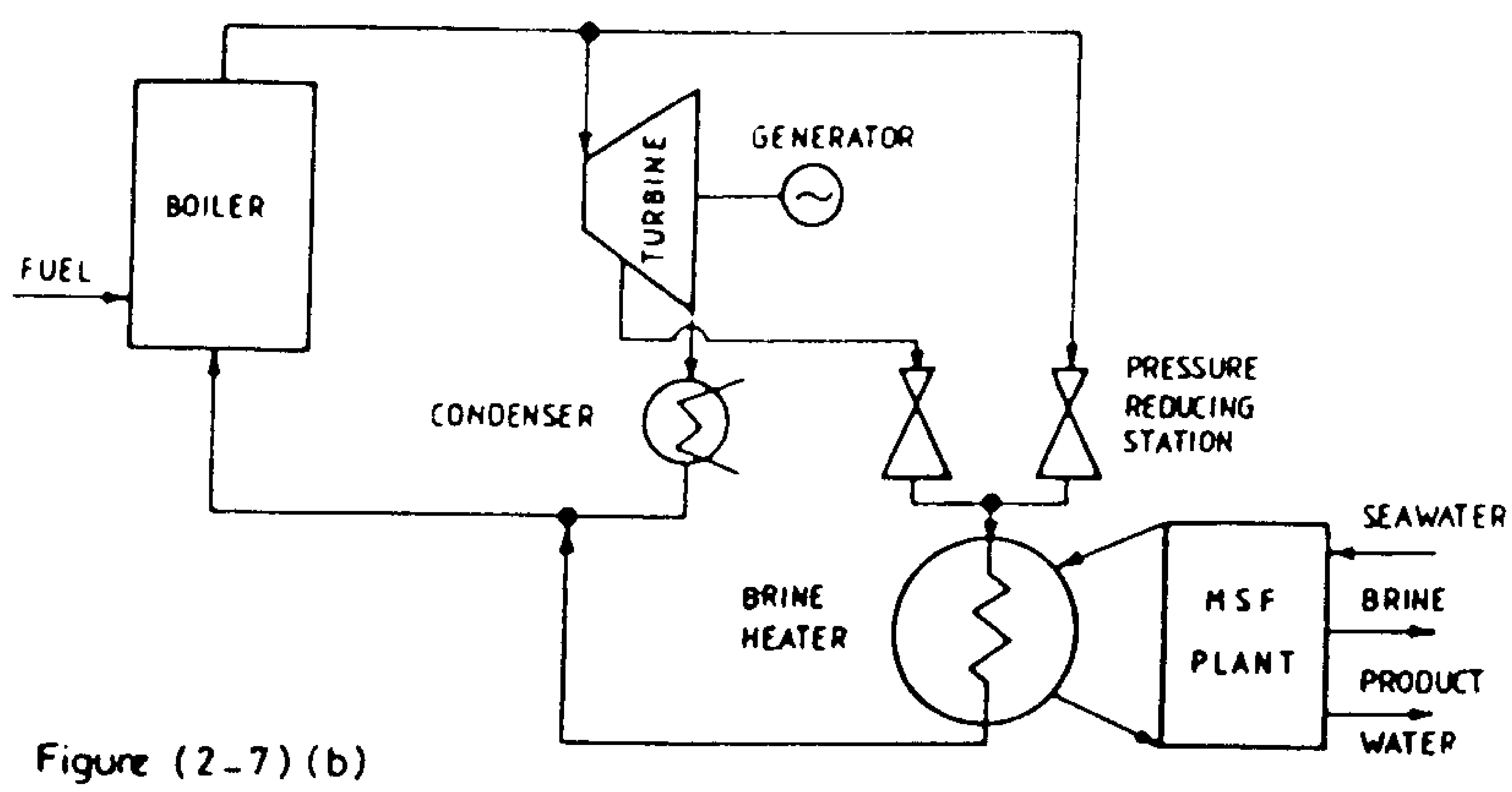


Figure (2-7) (b)
Controlled Extraction Steam Turbine _ Desalination Plant Combination

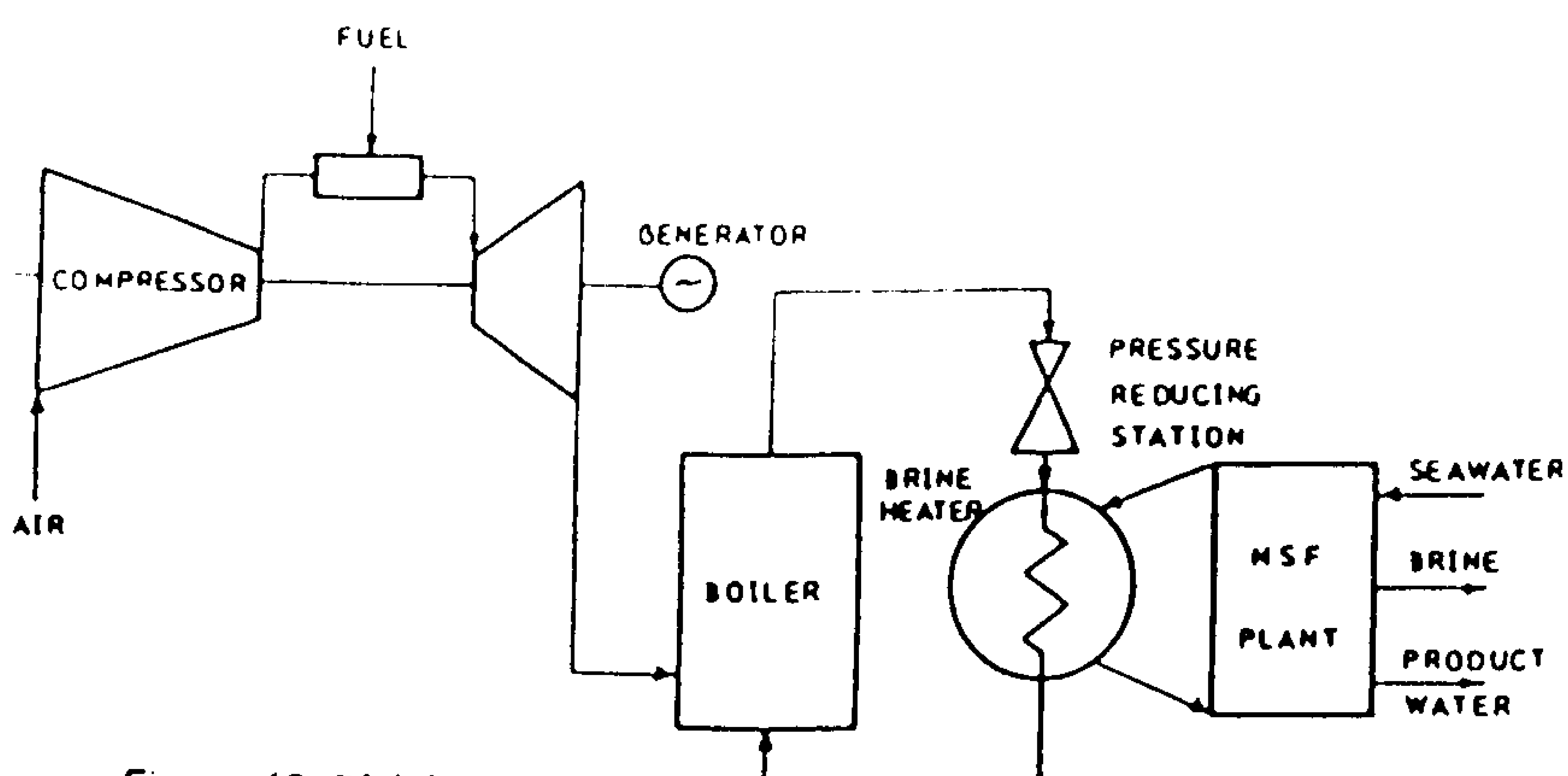


Figure (2-7) (c)
UNE / GTS Gas Turbine _ EHR Boiler _ Desalination Plant Combination

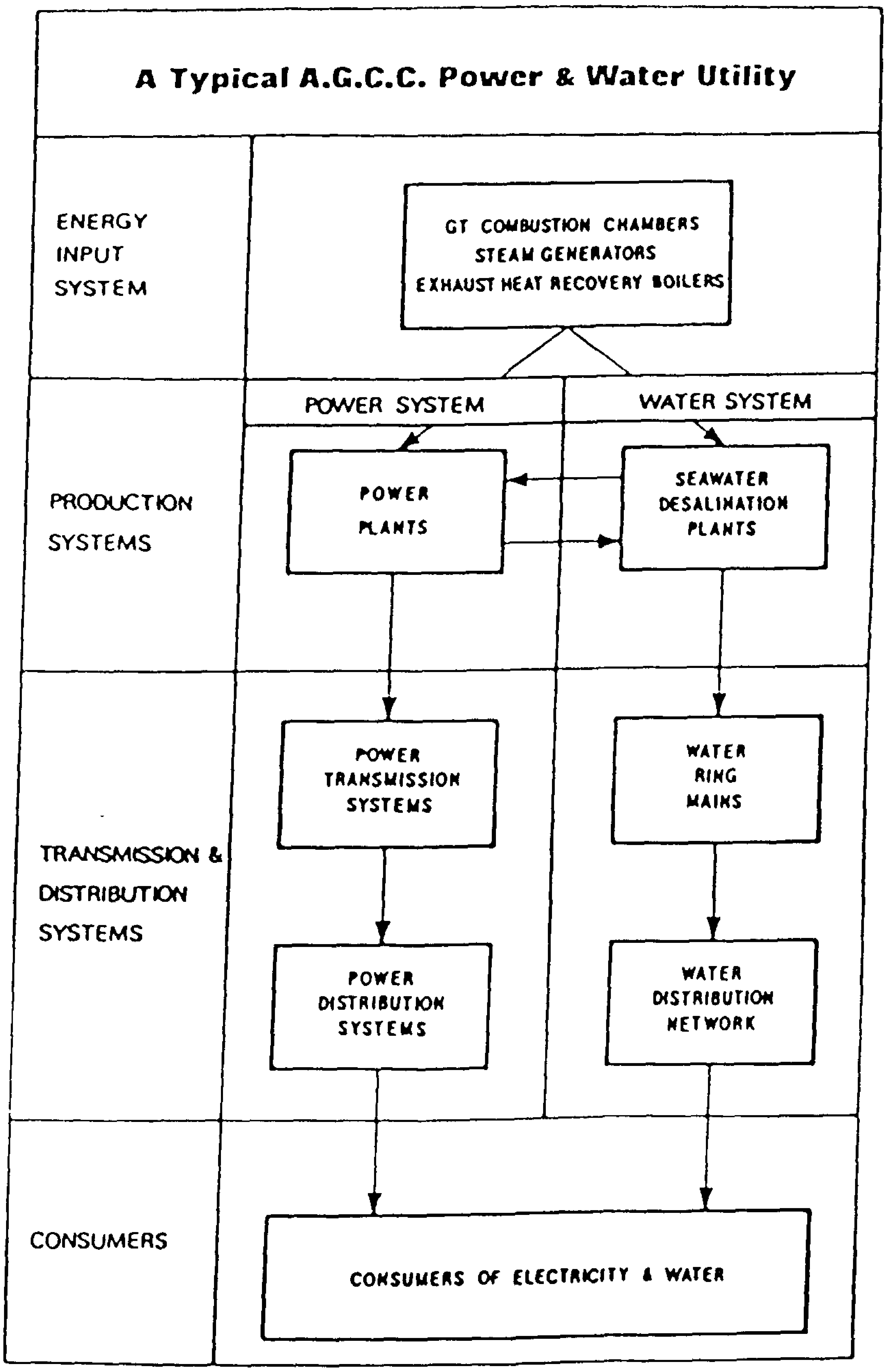


Figure (2-8) Functional Block Diagram

of a Typical AGCC Power and Water Utility

2.5 Abu Dhabi's Load Characteristics

As is expected in the climatic conditions of Abu Dhabi, the load characteristics of the WED electric power demand are such that summer is the season for high demand while winter is the season of low demand. During 1986, the minimum demand for electric power was 248 MW (in February), whereas the highest demand of 1136 MW was recorded in July. The weekly peak load pattern for the years 1984-1986 along with the mean weekly temperature profile are shown in Figure (2-9).

A typical hourly electric power demand (load curve) is shown in Figure (2-10) for a summer day and a winter day. The winter demand consists primarily of lighting, water heaters, domestic and industrial appliances. There is no space heating requirement unlike other Middle East Countries in the Gulf Area such as Kuwait, Bahrain and Saudi Arabia, where space heating is necessary for home and office use in the winter. The bulk of the summer load is due to air conditioning units for space cooling.

There are four distinct seasons that govern the electric power demand in Abu Dhabi. A detailed analysis of hourly loads from January 1983 to April 1987 shows the following patterns;

a. Summer - week # 17 -to- # 41 i.e. 25 weeks.
   Middle of April to end of September.
   Daily peak load occurs at hours 14:00, 15:00 or 16:00 but mostly at 15:00 hrs. Peak load is due to air-conditioners only. Daily mean drybulb temperature is 30°C or above, while the daily minimum drybulb temperature is around 27°C or above.

b. Fall i.e., summer to winter transition -
   week # 42 to # 46, i.e., 5 weeks October to early November.
   Daily peak starts to shift from 15:00 to 19:00/20:00 hrs. This shift commences with Thursdays and Fridays and eventually other weekdays depending on the drybulb temperature. Daily mean drybulb temperature continues to fall and the minimum drybulb temperature is below 25°C.

c. Winter - Week # 47 to # 52
   # 01 to # 11, i.e. 17 weeks. Early November to early March.

Daily peak load occurs mostly at 19:00 hrs but occasionally at 20:00 hrs. Air-conditioning load is minimal, while evening light and domestic appliances decide the daily peak demand. Daily mean drybulb temperature is below 25°C while the maximum drybulb temperature throughout the period is below 27°C.

d. Spring i.e., winter to summer transition -

Week # 12 -to- #16, i.e. 5 weeks.

Early March to mid April.

Daily mean drybulb temperature starts to rise and the maximum is above 27°C, so that afternoons are warmer and air-conditioning units, in increasing numbers, are turned on. The peak load on weekdays begins to occur at 15:00 hrs while at weekends it still occurs at 18:00/19:00 hrs. Mean daily drybulb temperature is nearly 25°C and rising.

Figure (2-11) shows the variation of the average minimum, mean, and maximum temperatures in Abu Dhabi in the period 1983-1986.

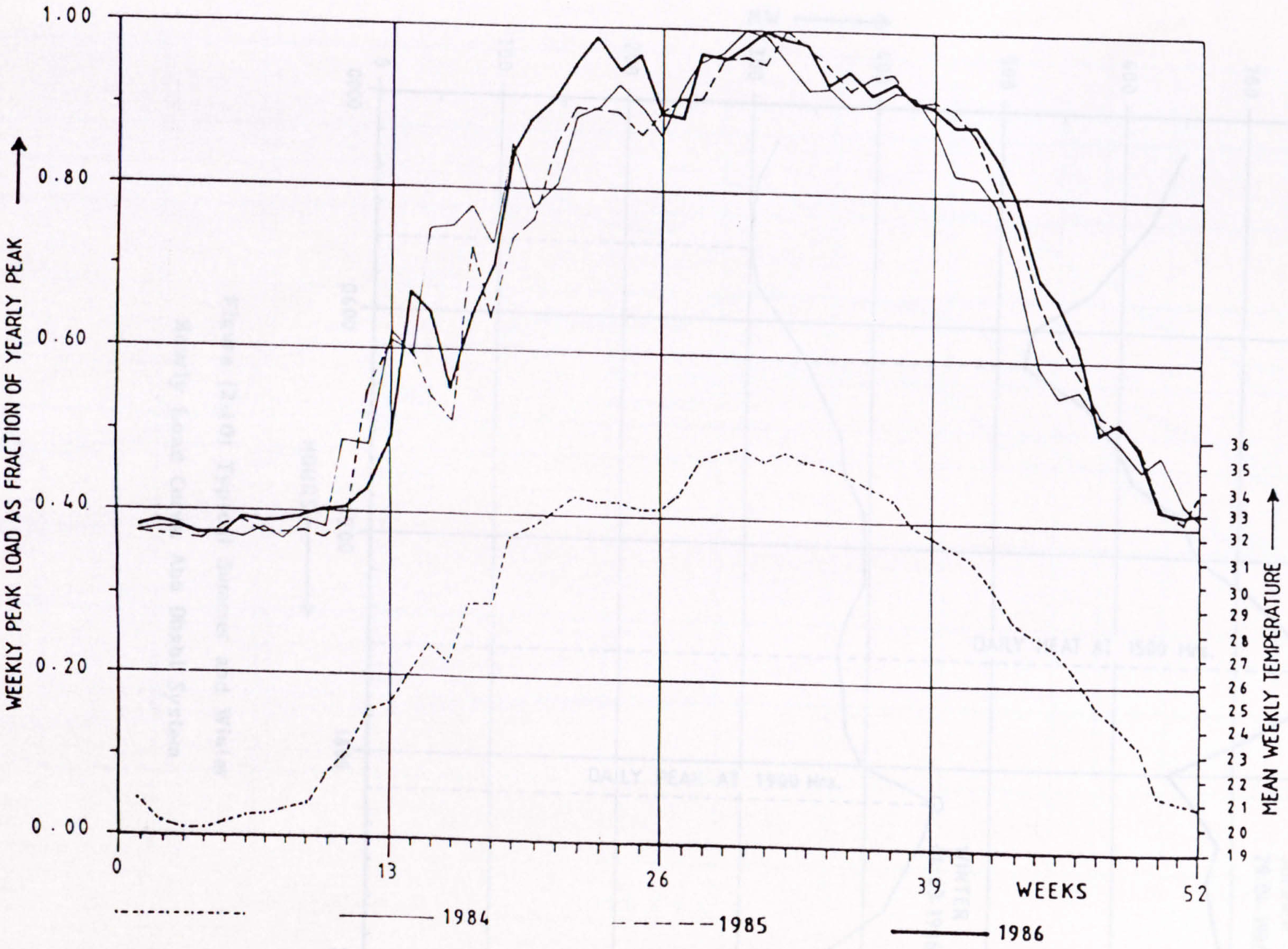


Figure (2-9) Weekly Peak Load Pattern and Mean Weekly Temperatures in WED System

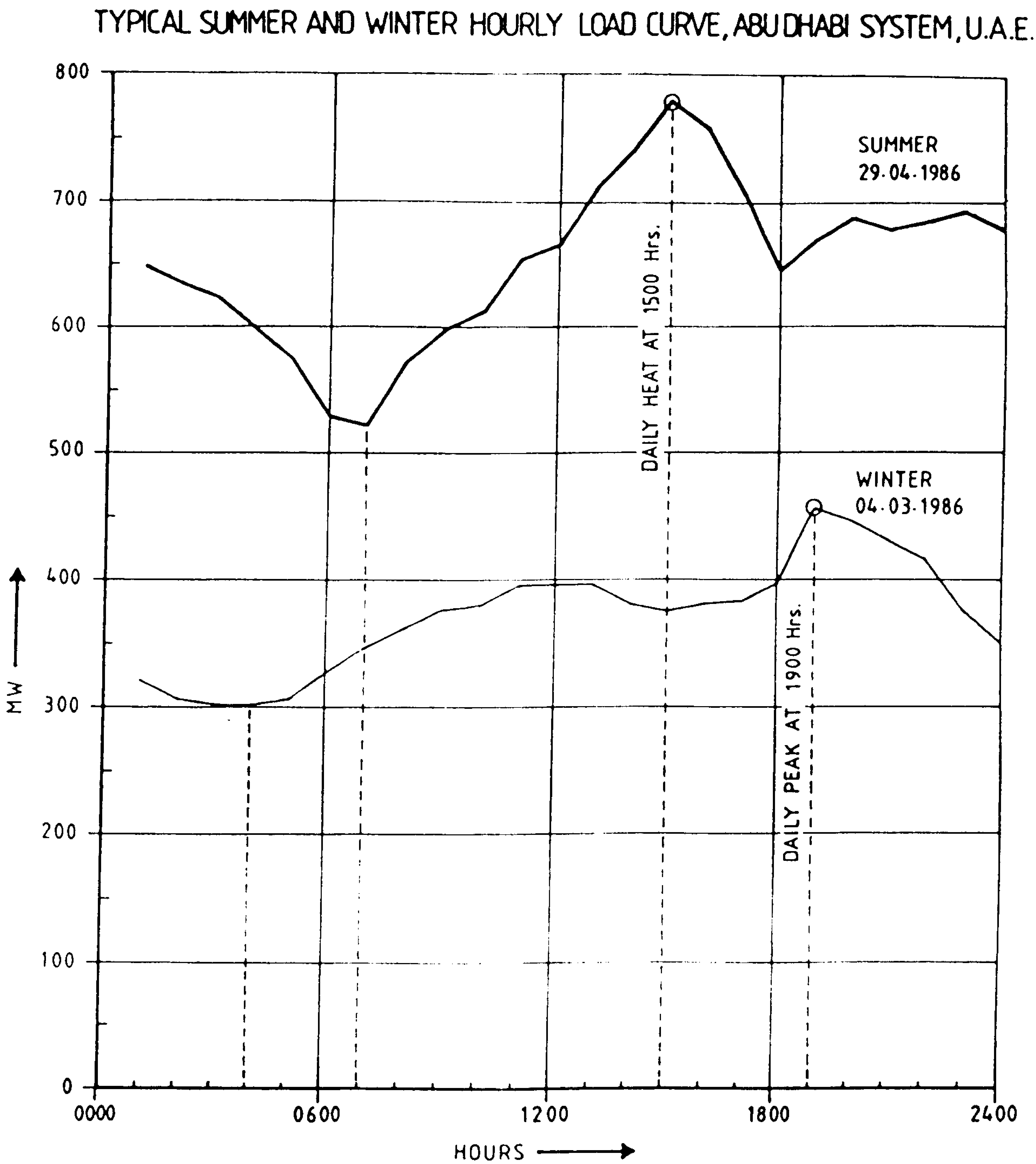


Figure (2-10) Typical Summer and Winter Hourly Load Curve, Abu Dhabi System

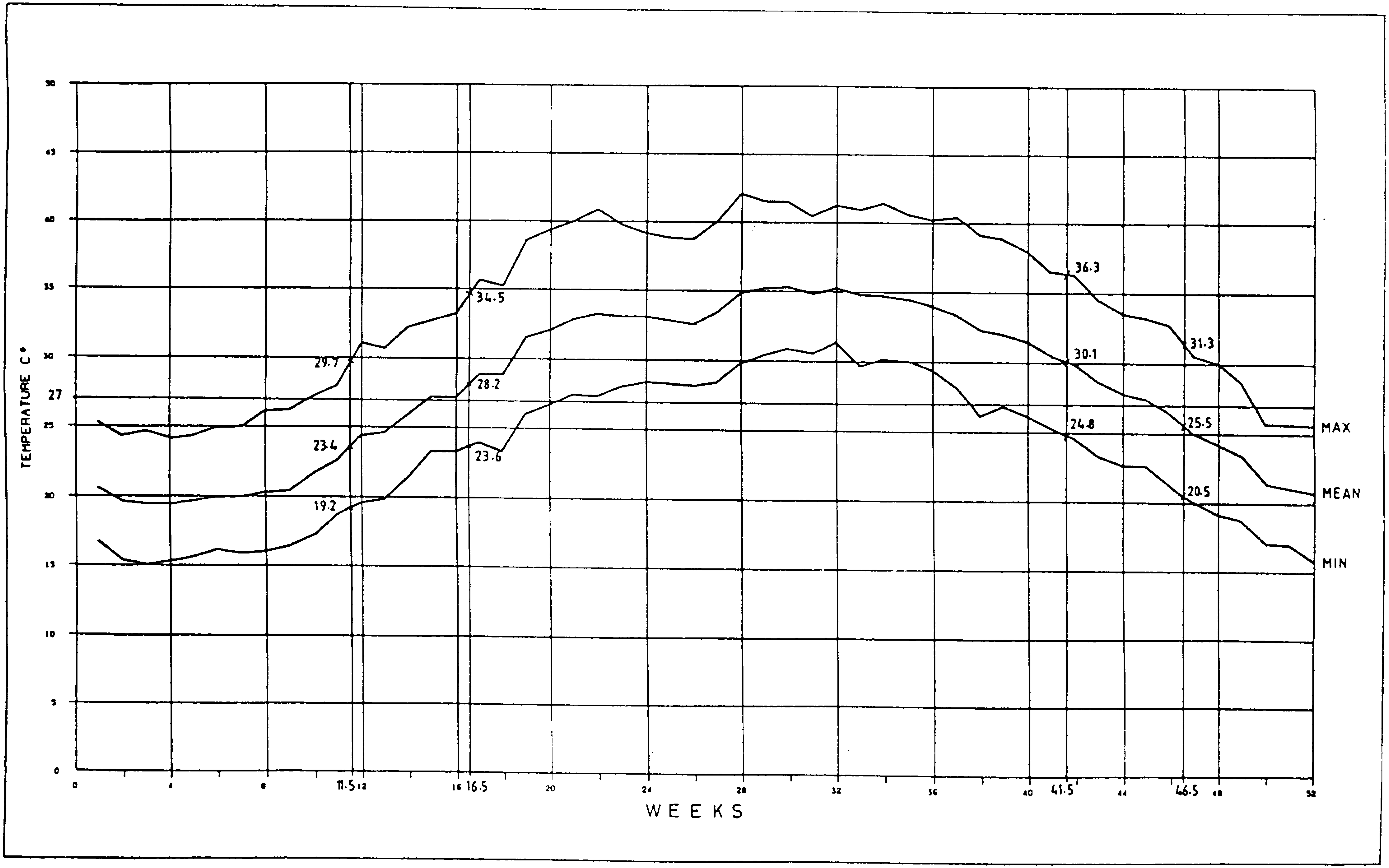


Figure (2-11) Average Minimum, Mean and Maximum Weekly Temperatures in Abu Dhabi, 1983-1986

In mid 1979, the author visualized the implementation of an integrated approach to planning and operation of the WED system. The concept of a computer based Engineering Information System (EIS) was envisaged. For scientific management, it is imperative that the right information is made available to the right person and at the right time in order to monitor and control power station performance to achieve maximum economy, continuity of service and reliability.

The appropriate computer hardware, operating system, and applications software were selected to handle the following envisaged tasks:

1. Inventory Management System
2. Personnel Management System
3. Cost and Budget Analyses
4. Operational Planning System
5. Maintenance Planning and Scheduling System, and
6. Production Costing System.

The Operational Planning System (item 4), in turn, was to contain such applications as:

a. Create Operating Statistics
b. Forecast Demands (Power and Water)
c. Reliability Studies
d. Fuel Forecast and Related Statistics
e. Heat Balance Calculations
f. Unit Commitment and Economic Dispatch
g. Power System Studies
h. Expansion Planning

The computer hardware was installed in early 1984 and applications software development has proceeded from that time. Several modules of the application program are being developed simultaneously, the discussion of which is beyond the scope of this work.

2.6 Economic Scheduling Functions

The research reported in this thesis is in the broad area of economic scheduling of electric power systems. Here, given unit availability for scheduling and pertinent data on variables that allow a load forecast, the goal is to determine the optimal generation levels at each unit. Optimality is defined in terms of minimum fuel costs in an all-thermal system. A functional block diagram of economic scheduling functions is shown in Figure (2-12). Here the problem is decomposed into the following subproblems [4]:

- Load Forecasting
- Unit Commitment
- Economic Dispatch
- Loss Modelling
- Fuel Cost Parameter Estimation.

One is essentially dealing with economic dispatch (E.D.) and its supporting functions. Here one relies on the definition of E.D. given in the IEEE Standard Dictionary of Electrical and Electronics Terms [5] as;

"The distribution of total generation requirements among alternative sources for optimum system economy with due consideration of both incremental generating costs and incremental transmission losses".

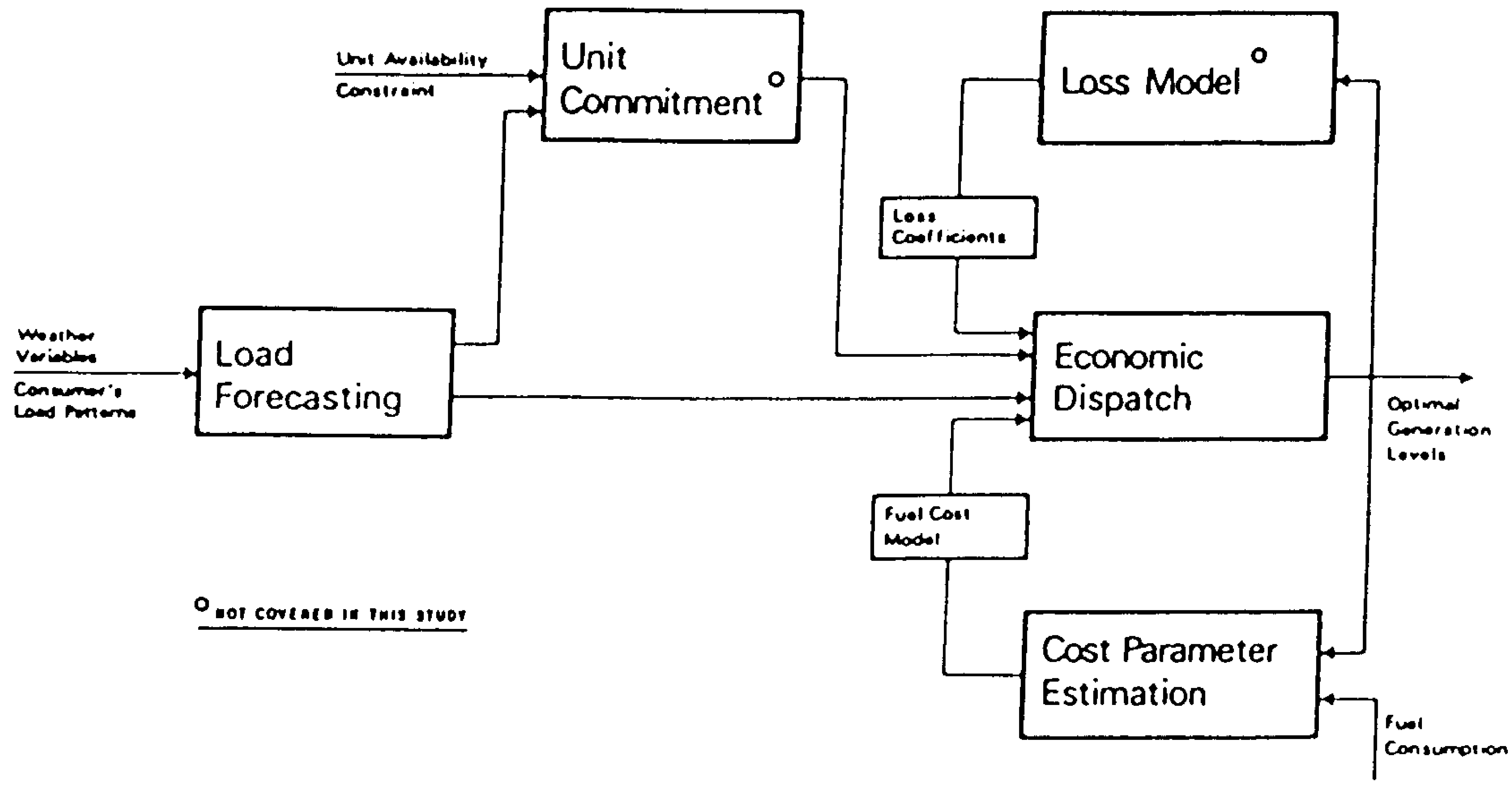


Figure (2-12) Functional Block Diagram of Economic Scheduling Functions

The total generation requirement is obtained from a load forecast to be discussed shortly. A unit commitment program uses the load forecast and information on units available for scheduling, recognizing outages due to maintenance schedules, to determine the selection of units to meet the demand. The duration of unit commitment function is usually one week conducted on an hourly basis (Figure 2-1). This function relies on start-up-shut-down and operational cost data to provide the input to the economic dispatch function.

Economic dispatch requires an accurate representation of both losses in the electric network and the costs of fuel as a function of the unit output. This latter aspect is covered in the cost parameter estimation function.

For an electric utility, load forecasting is of paramount importance, since a forecast is an essential input to many long-term and operational planning functions. A forecasting procedure analyzes past data in order to identify a pattern that can be used to describe the behaviour of existing data. This pattern is then extrapolated or extended into the future to obtain a forecast. The basic strategy employed in all forecasting techniques rests on the assumption that the identified pattern obtained on the basis of past data will continue in the future. If this pattern does not persist in the future, the forecasting technique will most likely yield results that are inaccurate. The components of a load forecast will invariably contain contributions from the following [6&7]:

- Trend (growth)
- Seasonality and cyclic components
- Weather related components
- Socio-economic variables and
- Random fluctuations.

A functional block diagram relating the inputs to a load forecasting model and subsequent output and monitoring functions is shown in Figure (2-13).

A load forecasting model takes as input past records as well as weather forecasts of variables that influence the demand for electricity. These variables, which are sometimes referred to as explanatory variables, include drybulb temperatures, dew point temperatures, relative humidity, and cloud cover. Such environmental data is more relevant to short-term rather than long-term forecasting. In many instances, only one weather variable (temperature) is included in the data base. Demand statistics are an essential part of the forecasting model since they provide the basis for identifying a pattern from past records. In long-range forecasts, a model of the demand growth based on trend analysis/socio-economic indicators/ analytical techniques is necessary for accurate forecasts. In short-range forecasts patterns of social behaviour including special events, holidays and customs may be included. As an example in Muslim Countries such as the UAE, the observance of the Holy Month of Ramadan changes the consumption pattern of electricity. This month is fixed in the Hijra Calendar (Islamic Lunar) and advances each year by nearly 11 days in relation to the Gregorian Calendar. Figure (2-14) shows the cyclic movement of the month of Ramadan for the years 1982-1990.

The current ad hoc method to forecast annual and weekly peak power demand for the WED system in Abu Dhabi is based on a combination of procedures. These include least squares polynomial analysis of past load data, annual percentage increase in demands, and the time diversity of the peaks occuring in Al Ain and in Abu Dhabi areas. The forecast so obtained is then adjusted to incorporate the most recent available information.

Having obtained a single value for the annual demand, taken as the most likely, the weekly expected peak loads are then determined from this load. The procedure is to obtain the normalized weekly peak load factors for the immediate past 5 years, take their averages, plot these and superimpose a smooth curve. This curve is then read off to determine the normalized weekly factors. Finally, an estimate of the expected weekly peak demand is obtained by multiplying the forecast of the annual peak demand with the normalized weekly factor.

Annual peak load data for Abu Dhabi, UAE, is available for the last 14 years i.e. from 1973, while more detailed weekly and monthly peak demand data is available from 1979 onwards. Figure (2-15) shows the weekly system peak load from January 1983 to December 1986.

2.7 Summary

In this chapter, introductory material related to topics treated in this thesis was presented. The setting of the problems discussed within the framework of power system planning and operations activities was outlined. Practical application of techniques developed in subsequent chapters is conducted for the electric power system of the Water and Electricity Department (WED) of Abu Dhabi, United Arab Emirates. Various unique aspects of both the Arab Gulf Cooperation Council's (AGCC) interconnection being contemplated were discussed together with the Abu Dhabi Electric System including its load characteristics. The chapter was concluded with a section on the main functions of economic scheduling via appropriate definitions and preliminary comments.

DEMAND FORECASTING MODEL

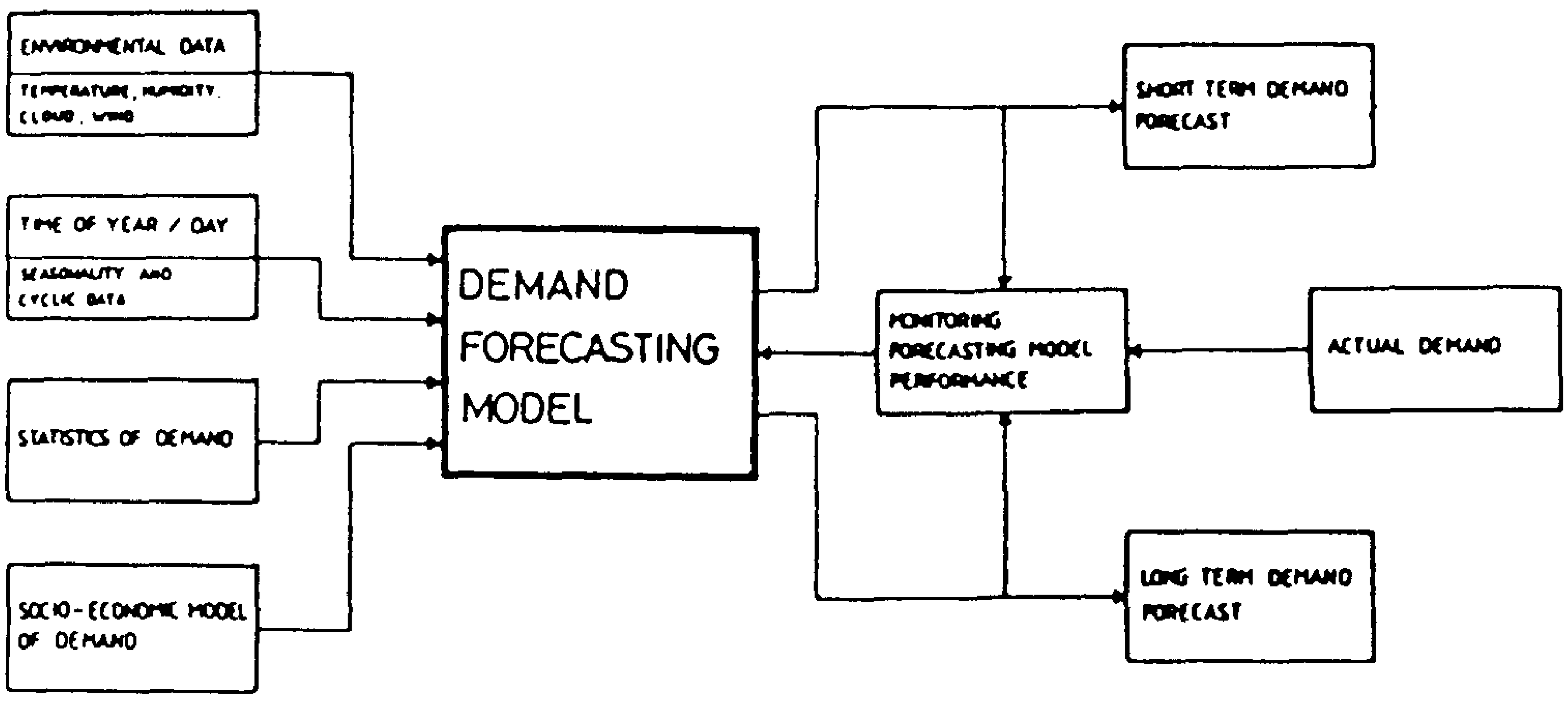


Figure (2-13) Functional Block Diagram

of Load Forecasting

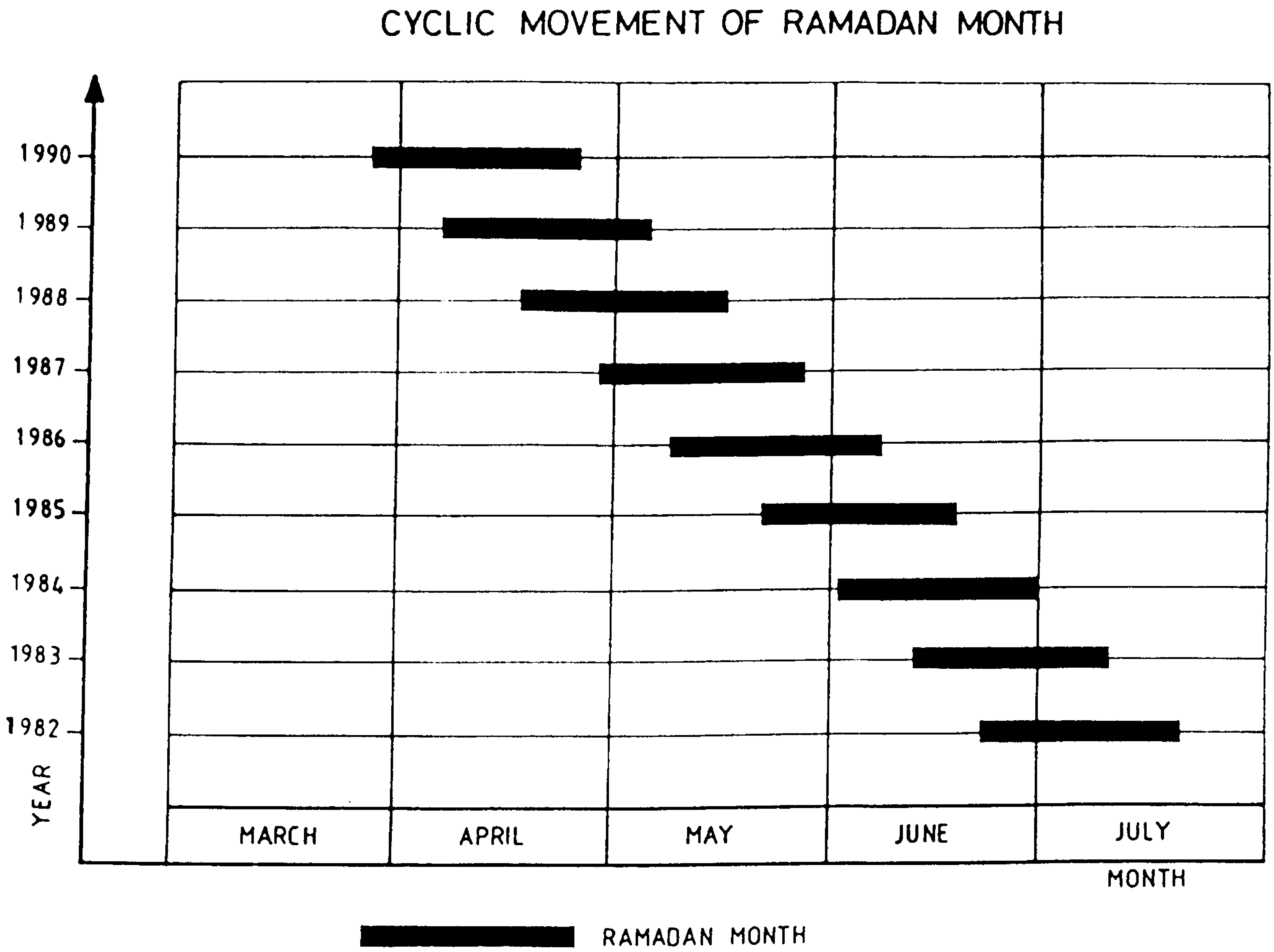


Figure (2-14) Cyclic Movement of the Month of Ramadan for the Years 1982-1990

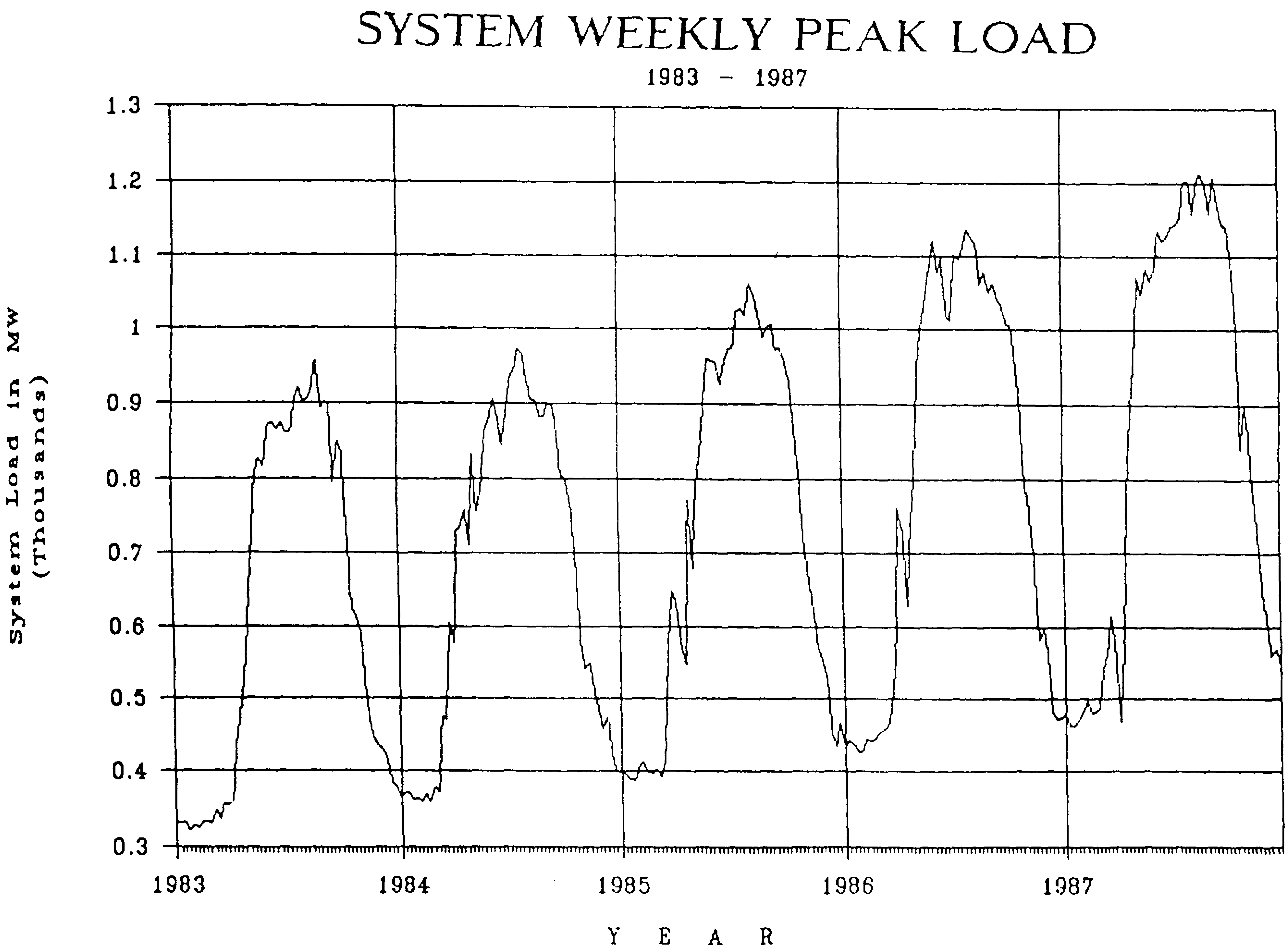


Figure (2-15) System Weekly Peak Load for WED Abu Dhabi

CHAPTER 3
LOAD FORECASTING IN POWER SYSTEMS
A COMPARATIVE EVALUATION

3.1 Introduction

Load Forecasting plays a major role in economy-security functions in any electric power system and this chapter is devoted to this topic. We begin in Section 3.2 with a comprehensive review of progress in this area, followed by a critical evaluation of a number of techniques for load forecasting, conducted using data bases from the Abu Dhabi electric system.

The techniques discussed in this chapter can be broadly classified into three categories:

- Conventional Forecasting Methodologies
- Time Series : Box-Jenkins Methodology
- System Theoretic Methodologies

Three major conventional forecasting techniques discussed in this chapter are:

1. General Exponential smoothing for seasonal load forecasting, which is treated in Section 3.4.
2. Winters' Additive model seasonal forecasting which is discussed in Section 3.5
3. Winters' Multiplicative model seasonal forecasting which is discussed in Section 3.6.

The Box-Jenkins approach is considered in Section 3.7. In each case a detailed discussion of salient features of the approach is given. Moreover, computational results pertaining to two forecasting problems are presented.

Advances in system theory have a significant impact on many power systems functions, and load forecasting is no exception. The powerful technique of Kalman filtering is reviewed in Section 3.8. Particular attention is paid to the problem of adaptive filtering and a comprehensive review of approaches to evaluating noise statistics is offered. A simple technique is used in conjunction with our computational investigations of the short-term forecasting problem.

In Section 3.9, a specific version of the Recursive Weighted Least Squares technique of identification is considered. Results for the short-term forecasting requirement are given. The last technique considered is that of Instrumental Variable and experience from the application of the method is presented in Section 3.10.

The chapter is concluded by a critical evaluation of the performance of each method and recommendations for the appropriate selection of techniques for forecasting load with special reference to the Abu Dhabi electric power system are submitted.

3.2 Literature Review of Methods on Load Forecasting

Several needs of an electric utility are met by load forecasts, including corporate planning, competitive strategy development, revenue planning and budgeting, fuel purchasing, marketing and rate making for energy sales, system planning, and operations and dispatch management. Load forecasting includes power and energy as the two major variables of interest. Energy forecasts are important in long range planning activities, whereas power forecasts play a dominant role in systems operations and dispatch studies. The chapter deals with power demand forecasting.

Classifying a large field such as load forecasting is equivalent to projecting an n-dimensional surface onto one or two dimensions at a time [8]. Certain classifications are quite distinct, while others may well intersect. A popular scheme is according to the load model structure in terms of the effects influencing the load behaviour:

1] Models which rely mainly on the time of day and the latest load behaviour.
2] Models which include weather effects but do not rely on the latest load behaviour for the forecast.
3] Models that include both weather, time of day and immediate past load information.

This classification scheme is employed in two bibliographic reviews of the area [9 & 10].

An alternative classification scheme is advocated by Abu El-Magd and Sinha [11], on the basis of the modelling approach used to represent the load, since the most important part of forecasting is in the system identification stage. Moreover, incorporating weather-sensitive components in the forecasting technique does not affect the methodology used to model the other load components.

A third classification is according to whether the entire load curve (hourly or a daily or weekly forecast or daily for a monthly or annual forecast) can be modelled, or whether only the peak load profile is modelled.

According to Happ [12], the history of electric load forecasting may be traced to the early days of economic dispatch prior to 1930, using methods such as classical decomposition. Dryar [13], correlates weather effects with electric power demand in forecasting daily peak loads. Multiple regression models are used in Davies [14] to correlate weather and demand, and in a similar manner Heinemann et al. [15] relate weather variables to a weather-sensitive load to calculate the daily peak load for the summer period.

In [16] to [18], Farmer and his associates use an additive model to represent the load $\mathfrak{z}(t,d,w)$ at time t on day d of week w as the sum of a long term trend $A(t,w)$, a component $B(t,d)$ that depends on the day of the week, and a residual or random component $\nu(t,d,w)$. A simple moving average is used to estimate the trend component and A is updated daily on the basis of an N_d day moving average. The day-of-the-week effect $B(t,d)$ is computed using an average over the number of weeks of the difference between the actual load and the trend component. This procedure [18] minimizes the mean square error of the random component averaged over several weeks of past data. The random component ν is estimated using a spectral decomposition. This method has been tested in [19] as well.

The use of an exponentially-weighted moving average has been advocated by Lijesen, and Rosing [20]. Here one updates $B(t,d)$ as:

$$B(t,d) = \alpha[\mathfrak{z}(t,d,w) - A(t,w)] + (1-\alpha)\, B(t,d-7)$$

where α is an empirically derived smoothing parameter. In addition, Lijesen and Rosing incorporate a weather-sensitive component into their residuals by means of allocating a daily energy requirement via a 24-element vector of constants to each hour.

Christiaanse [21] uses a procedure based on Brown's exponential smoothing [22] for hourly loads over an interval of one week. The model considered is of the Fourier Series form.

The selection of the fitting functions and the smoothing constant are major issues discussed in [21]. To improve the accuracy of the forecast, Christiaanse extends the method by forecasting the observed errors from the model already developed. Here the autocorrelation of the observed residuals for a period of time is studied and an autoregressive model with lags of one and twenty four hours is used to forecast the errors.

Sachdev and Ibrahim [23] propose an on-line technique using exponential smoothing in two stages. In the first the load $z(d,h)$ of the h^{th} hour on the d^{th} day is estimated from

$$\hat{z}(h,d) = z(d-1,h) + (1-\alpha_d)\,[\hat{z}(d-1,h) - z(d-1,h)]$$

where α_d is the daily smoothing constant. The error component is then estimated using a similar model. The adaptive estimation of $\upsilon(t)$, has been done using an autoregressive model AR(k) given by

$$\upsilon(t) = \sum_{i=1}^{k} \alpha_i\, \upsilon(t-i) + \epsilon(t)$$

where $\epsilon(t)$ is a white noise residual. In [24], Gupta and Yamada use AR(24) on hourly data. Galiana [25] on the other hand uses AR(1) and AR(2) for two different residual series.

In [26] to [29], a general time series model to forecast medium and long range power demand is used. This is adopted by Vemuri et al. to perform mid term forecasting in [30]. Here an Autoregressive Integrated Moving Average (ARIMA) $(0,1,1)X(0,1,1)_{12}$ model with a twelve month period was used to forecast monthly peak loads for lead times of up to 40 months. The very short term forecasts were first addressed by Keyhani and El-Abiad [31] using Autoregressive Moving Average (ARMA) models. They use ARMA (1,0) model on 1 minute load data, ARMA (1,1) and (2,1) models on 5 minute data and an ARMA (2,0) model on hourly load data. Later Keyhani and Eliassi Rad [32] use models which combined Autoregressive (AR) models with some weather inputs and trigonometric trend functions to forecast hourly loads up to one week ahead.

Hagan and Klien [33] use ARIMA models with a daily period to forecast hourly loads with one to four hour lead times. Different models were developed for each season. Meslier [34] later used ARIMA $(1,0,0)\times(0,1,1)_{365}$ models to forecast daily energy consumption one day ahead. Correction factors are added to the model to adjust for the holidays. In [35], Abu El-Magd and Sinha use multivariate AR models for forecasting load demands of a multinode system. They use an ARI (6,1) model to forecast the load demand at four substations at five minute intervals and an AR (2) model with 24 hour differencing to forecast at 1 hour intervals. In [36] Vemuri, Huang, and Nelson develop a method for on-line identification of AR models using sequential least squares. They use AR (10) models on 3 hour load data to forecast up to 21 hours ahead.

In [37], Hagan and Behr demonstrate that a simple polynomial regression analysis to describe the non-linear relationship between loads and temperatures, when combined with a Box and Jenkins transfer function model [59], can provide more accurate short term forecasts. Ross et al. [38] suggest a hierarchial structure to obtain load forecasts with five-minute, one-hour, and twenty-four hour lead times. The load is assumed to be the sum of a component predicted by regression of the load an all observed exogenous factors (base value and a weather-sensitive component) and an error term which is represented by an ARMA model.

Van Meeteren and Van Son [39] decompose the load into five components

$$z(t,d) = A(t,d) + B(t,d) + T(t,d) + W(t,d) + \nu(t,d)$$

Here T is a trend effect and W(t,d) is an additional weather-dependent component.

Srinivasan and Pronovost [40] use a model

$$\hat{z}(t,d) = \sum_{k=1}^{4} \alpha_k \, \hat{z}_k(t,d)$$

where

$\hat{z}_1(t,d)$ is a forecast of $z_2(t,d)$ based upon first order autoregression with a lag of one hour. $\hat{z}$ (t,d) is a forecast of $z(t,d)$ based upon first order autoregression with a lag of one day. $\hat{z}_3(t,d)$ and $\hat{z}_4(t,d)$ similarly involve a lag of one week and one year respectively.

The α_k are linear weights seeking an optimal combination of these four separate forecasts.

In [41], De Martiano et al. also divide the load into a seasonal component, a weekly component and a residual component. Exponential smoothing is used to identify the seasonal and weekly components. A mixed ARMA model is employed to identify the residual component.

Abu-Hussien et al. [42] present an adaptive model based on hourly loads and weather information. They use an individual power system load bus which is strongly dependent on weather variables to test the accuracy of their algorithm. Maximum error was found as 4% of the average hourly load.

Keyhani and Miri's [43] approach involves estimating the parameters of an ARIMA model fitted to historical load and weather data. With a strong correlation between load and temperature, they find good results with RMS error for hourly forecasts ranging from 2% to 4.4% of the daily peak load for one-step ahead load predictions.

Irisarri et al. [44] develop an hourly-recursive Kalman filter, which includes an exogenous temperature forecast, to formulate an equivalent deterministic component of the load value. They present results with average absolute errors ranging from 2.2% to 4.8%.

Krogh et al. [45] use a regression based technique to normalize daily peaks and troughs for weather variables and hence allow a univarate ARIMA model to be used to forecast the hourly normalized data.

The formulation of the load evolution model in state space form allows the application of Kalman fittering theory to conduct the forecasting task. This is ideally suited for on-line applications since it is recursive in nature. Toyoda et al. [46] are credited with the application of state estimation methodology to load forecasting. Here the state variables are the system load itself, the increment of the load, and short-term and long-term load patterns.

In [47], Sharma and Mahalanabis apply state estimation techniques to short-term load forecasting. Here, in contrast with Toyoda's assumption of a known system matrix, Sharma and Mahalanbis assume two properties of the load. In the first the load is approximated by a finite order polynomial of the time in hours. In the second assumption, the ratio of the change in the load demand to the actual load demand is constant for all days. The same model proposed by Christiaanse [21] is rearranged by Sharma and Mahalanabis [48] to apply adaptive state estimation to identify model parameters and noise covariance.

Singh et al. [49] assume a time series model and develop a predictor which identifies the coefficients of the time series. In [50], Galiana and Schweppe obtain load forecasts up to one week in advance by separating the load into a periodic term and a residual term. The periodic term is expressed as a Fourier Series expansion while the residual is an AR model with forcing terms that are weather dependent. Least squares estimation is used to find the model parameters off-line.

In [51], a multivariable state space model for multinode systems is proposed. Two sub-systems are used to model the state and parameter variations separately.

The use of maximum entropy method to identify an AR model to obtain 24 hour forecasts is advocated by Lu and Rao [52]. This can provide a good model even with a short data record. This advantage is not critical for load forecasting and the computations are quite lengthy.

In [53] the application of pattern recognition techniques to load forecasting is reported. It is interesting to note that in [54], Dehdashti et al. present results applying pattern recognition techniques to 24 hour load forecasting of load in a small town with a peak load of 123 MW. They state that this technique is not appropriate for large areas as the diversity of loads distorts the weather sensitive pattern of loads. In [55], Rahman and Bhatnagar present a knowledge-based system as an alternative to forecasting one-to-twenty-four loads with errors of up to 3.5%.

There is another approach where a load management view point is adopted. As an example McRae et al. [56] use regression analysis to deal with potential for management of load at peak times. For a collection of papers describing current state of the art in power system load forecasting, the edited book by Bunn and Farmer [57] is recommended.

At this point in the discussion, it is useful to note that no single most preferable forecasting method has emerged for load forecasting problems in electric utility systems. One reason is that each utility is presented with a unique set of external inputs that affect the composition and time evolution of its load (weather, cultural and special occasion related). As a result it appears that a considerable amount of "tailoring" (to use Bunn and Farmers' term [57]) has to be done to the forecasting technique in sometimes ad hoc ways to deal with the specific problem at hand.

Some methods, such as conventional techniques listed in Section 3.1, are characterized by simple computational algorithms that do not require a large amount of tuning as compared to advanced techniques such as Box-Jenkins, Kalman Filtering and the Instrumental Variable methods. In these latter cases, one applies some sophisticated procedures to arrive at optimum model orders and their structures.

One of the intentions of this chapter is to examine how many of the methods cited in the literature, and newer ones, fare in the application to forecasting the load on the Abu Dhabi electric system.

3.3 Abu Dhabi's Load Forecast Data Base

Load Forecasting in an electric power system is an ongoing activity that utilizes full data sets from past years' records. The testing of proposed forecasting algorithms should be done with data records that exhibit the most variability to allow evaluation of the efficiency of the proposed algorithms. In conducting research into load forecasting for the Abu Dhabi system, a main short-term load data set has been utilized and is listed in Table (A-1) and plotted in Figures (A-1) of Appendix A. The data set consists of the hourly load for 672 points covering the period from September 7th to October 5th, 1986 as plotted in Figure (A-1-a). This period straddles the boundary between summer and fall and therefore is representative of a typical segment where load patterns are less clearly defined. This period is critical for system maintenance planning and therefore requires a reliable load forecast. In Figure (A-1-b) data for one week of hourly load is shown.

The mid-term data base for our forecasting spans eleven years of monthly peak demands. This is listed in Table (A-2) and shown in Figure (A-2) of Appendix A.

3.4 Seasonal Load Forecasting via Trigonometric Functions : A General Exponential Smoothing Approach

Power system loads exhibit a cyclical pattern that is repeated over a period of time. This is referred to as seasonality and the length of the cycle as seasonal period power case s. An additional trend component is observed in the load time series. The traditional approach to modelling seasonal data is to decompose the series into three components : a Trend $T(k)$, a Seasonal Component $S(k)$, and a Noise (residual) Component $\nu(k)$ where k is the discrete time index.

In the additive decomposition approach, the series is modelled as a sum of the three components given by

$$z(k) = T(k) + S(k) + \nu(k) \tag{3-1}$$

The multiplicative decomposition approach assumes that the applicable model is the product of the three components

$$z(k) = T(k)\, S(k)\, \nu(k) \tag{3-2}$$

The multiplicative model is readily transformed into an additive model by taking the logorithm of both sides of the model equation and dealing with the new time series in $\ln[z(k)]$. Equations (3-1) and (3-2) form the basis for a number of seasonal time series forecasting techniques.

3.4.1 Seasonal Models Using Trigonometric Functions

Traditionally, the Trend Component is modelled by a low-order polynomial of time k

$$T(k) = \beta_0 + \sum_{i=1}^{K} \beta_i \frac{k^i}{i!} \tag{3-3}$$

K is the order of the polynomial. The Seasonal Component $S(k)$ can be described by seasonal indicators or by trigonometric functions

$$S(k) = \sum_{i=1}^{m} A_i \sin\left[\frac{2\pi i}{s} k + \Phi_i\right] \tag{3-4}$$

where A_i and Φ_i are the amplitude and the phase shift of the sine function with frequency $f_i = \frac{2\pi i}{s}$ and m is the number of harmonics included. Combining the three components using the additive model leads to seasonal models of the form;

$$z(k) = \beta_0 + \sum_{i=1}^{K} \beta_i \frac{k^i}{i!} + \sum_{i=1}^{m} A_i \sin\left[\frac{2\pi i}{s} k + \Phi_i\right] + v(k) \qquad (3\text{-}5)$$

This model incorporates the parameters in a non-linear fashion, and it is more convenient to write the model in the form;

$$z(k) = \beta_0 + \sum_{i=1}^{K} \beta_i \frac{k^i}{i!} + \sum_{i=1}^{m} [\beta_{1i} \sin f_i k + \beta_{2i} \cos f_i k] + v(k) \qquad (3\text{-}6)$$

The modelling problem is reduced to determining the parameters $\beta_0, \beta_1, \ldots, \beta_K, \beta_{11}, \beta_{21}, \ldots, \beta_{1m}, \beta_{2m}$, provided that f_i have been assigned. A direct least squares parameter estimator can be applied to obtain the parameters of the model on the basis of a number of observed loads $z(k)$. This assumes that, the parameters in the Trend Component $T(k)$ and in the Seasonal Component $S(k)$ are fixed constants and that the errors $\{v(k)\}$ are uncorrelated.

These models work well for trends and seasonal components with fixed amplitudes and phases [6]. It is more logical for load time series to allow for adaptive - time changing trend and seasonal components. One possible implementation technique involving the principle of general exponential smoothing via discounted least squares is discussed in Appendix B.

3.4.2 Computational Results: Mid-Term Forecasting

The data base given in Appendix A corresponding to monthly peak load history was used in an implementation of the seasonal forecasting using trigonometric functions and general exponential smoothing to forecast the monthly peak for 12 months ahead.

The Trend Component was modelled using a linear relationship, so that K = 1. The Seasonal Component was modelled using four harmonics so that m = 4, with seasonality s = 12 months. Thus we have

$$f_i = \frac{2\pi}{12} i$$

The estimated model parameters were found using the procedure of Appendix B as:

$\hat{X}(1) = \hat{\beta}_0 = 865.939$ $\hat{X}(2) = \hat{\beta}_1 = 5.763$

$\hat{X}(3) = \hat{\beta}_{11} = -349.463$ $\hat{X}(4) = \hat{\beta}_{21} = 37.171$

$\hat{X}(5) = \hat{\beta}_{12} = 25.297$ $\hat{X}(6) = \hat{\beta}_{22} = 52.300$

$\hat{X}(7) = \hat{\beta}_{13} = -24.853$ $\hat{X}(8) = \hat{\beta}_{23} = 17.101$

$\hat{X}(9) = \hat{\beta}_{14} = -4.335$ $\hat{X}(10) = \hat{\beta}_{24} = 34.370$

The estimation is based on 130 data points.

Table (3-1) lists the resulting forecasts compared to the actual values. Figure (3-1) depicts the results graphically. The following are the fit statistics:

Mean Squared Error (MSE) = 1695
Mean Absolute Percentage Error (MAPE) = 8.26

The forecast statistics are as follows;
MSE = 2924
MAPE = 5.312

The optimum value of the smoothing constant α is 0.150, which was found by a trial and error process to arrive at the lowest possible value of mean square prediction error.

Inspection of the forecast errors in Table (3-1) reveals that only three points involved a percentage error larger than 10%. The spread of errors around zero is reasonable, as can be seen from Figure (3-1).

Table (3-1) Forecasting Results
For Monthly Data Using Trigonometric
Functions and General Exponential Smoothing

MONTH (L)	ACTUAL	FORECAST	ERROR	PERCENT ERROR
131	762.000	720.433	41.567	5.455
132	529.000	593.003	-64.003	-12.099
133	493.000	501.982	-8.982	-1.822
134	528.000	523.372	4.628	0.877
135	616.000	694.817	-78.817	-12.795
136	965.000	864.173	100.827	10.448
137	1084.000	1153.662	-69.662	-6.426
138	1142.000	1188.018	-46.018	-4.030
139	1204.000	1194.449	9.551	0.793
140	1213.000	1193.069	19.931	1.643
141	1209.000	1137.748	71.252	5.893
142	1092.000	1076.033	15.967	1.462

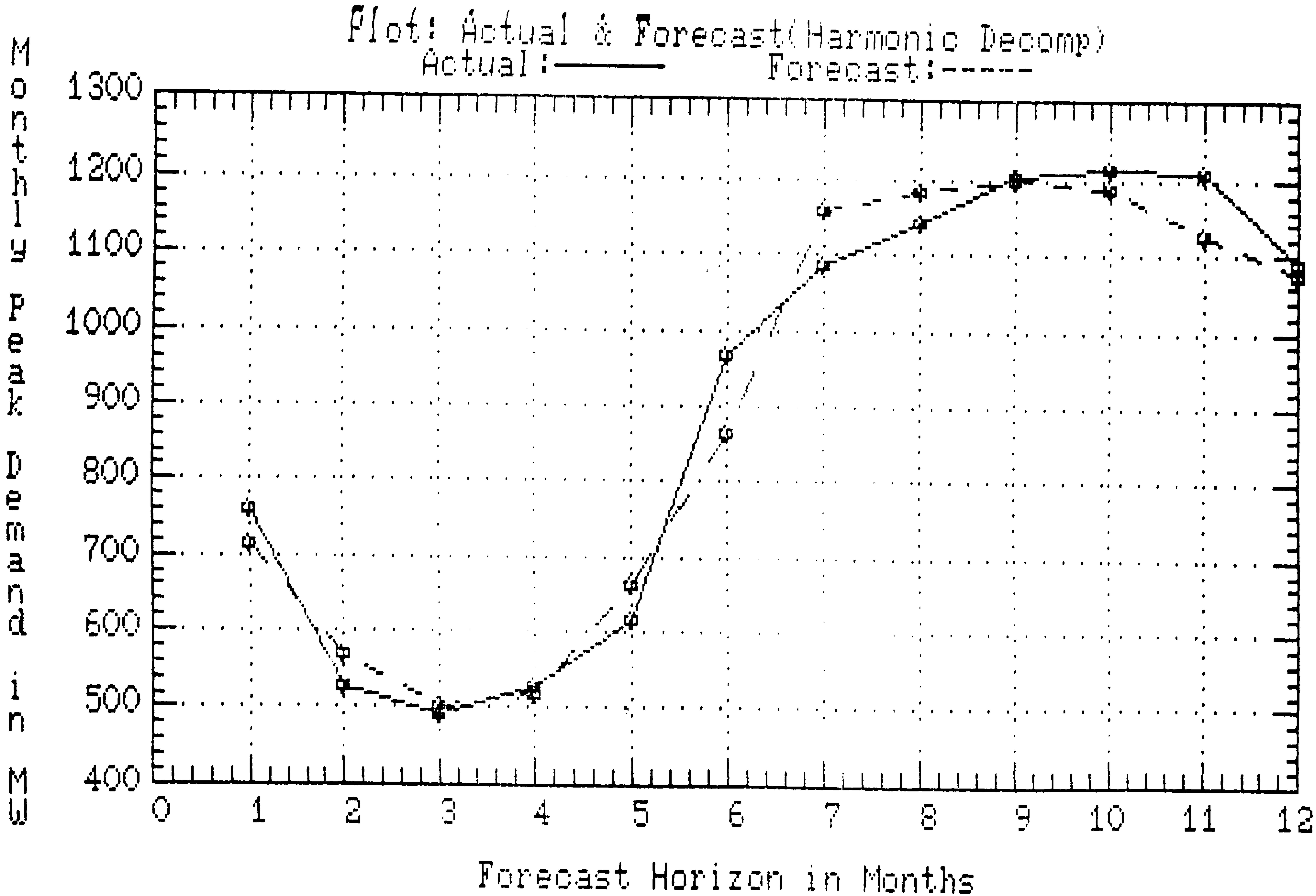


Figure (3-1) Forecasting Results for Mid-Term Problem, Seasonal Using Trigonometric Functions and General Exponential Smoothing

3.4.3 Computational Results: Short-Term Forecasting

The data base given in Appendix A corresponding to hourly load history was used in an implementation of the seasonal forecasting using trigonometric functions and general exponential smoothing to forecast hourly loads.

Here, once again we used a trend model such that K = 1. Five harmonics were required with seasonality of 24 hours. The estimated model parameters on the basis of 660 data points, were found to be:

$\hat{X}(1) = \hat{\beta}_0 = 815.872$ $\quad\hat{X}(2) = \hat{\beta}_1 = -0.446$

$\hat{X}(3) = \hat{\beta}_{11} = 31.710$ $\quad\hat{X}(4) = \hat{\beta}_{21} = 47.966$

$\hat{X}(5) = \hat{\beta}_{12} = 129.635$ $\quad\hat{X}(6) = \hat{\beta}_{22} = -33.591$

$\hat{X}(7) = \hat{\beta}_{13} = 36.327$ $\quad\hat{X}(8) = \hat{\beta}_{23} = 26.350$

$\hat{X}(9) = \hat{\beta}_{14} = 12.741$ $\quad\hat{X}(10) = \hat{\beta}_{24} = -8.777$

$\hat{X}(11) = \hat{\beta}_{15} = -21.135$ $\quad\hat{X}(12) = \hat{\beta}_{25} = 2.164$

Table (3-2) lists the resulting forecasts compared to actual values. Figure (3-2) shows graphically the forecast and actual hourly loads. The following are the fit statistics;

MSE = 555.97
MAPE = 2.17

The forecast statistics are as follows;

MSE = 195
MAPE = 1.097

The optimum value the smoothing constant α is 0.055, which was obtained using a trial and error procedure to obtain the lowest possible MSE.

From Table (3-2) it is clear that the percentage forecast error is relatively small compared to that involved in the mid-term problem. The errors do not exceed 3.5%.

Table (3-2) Forecasting Results
Hourly Data Using Trigonometric Functions
and General Exponential Smoothing

HOUR	ACTUAL	FORECAST	ERROR	PERCENT ERROR
661	893.000	892.215	0.785	0.088
662	943.000	938.900	4.100	0.435
663	966.000	983.041	-17.041	-1.764
664	1008.000	1006.000	2.000	0.198
665	933.000	995.117	-2.117	-0.213
666	936.000	959.723	-23.723	-2.534
667	930.000	927.779	2.221	0.239
668	956.000	923.750	32.250	3.373
669	956.000	947.335	8.665	0.906
670	962.000	973.232	-11.232	-1.168
671	962.000	972.865	-10.865	-1.129
672	928.000	938.375	-10.375	-1.118

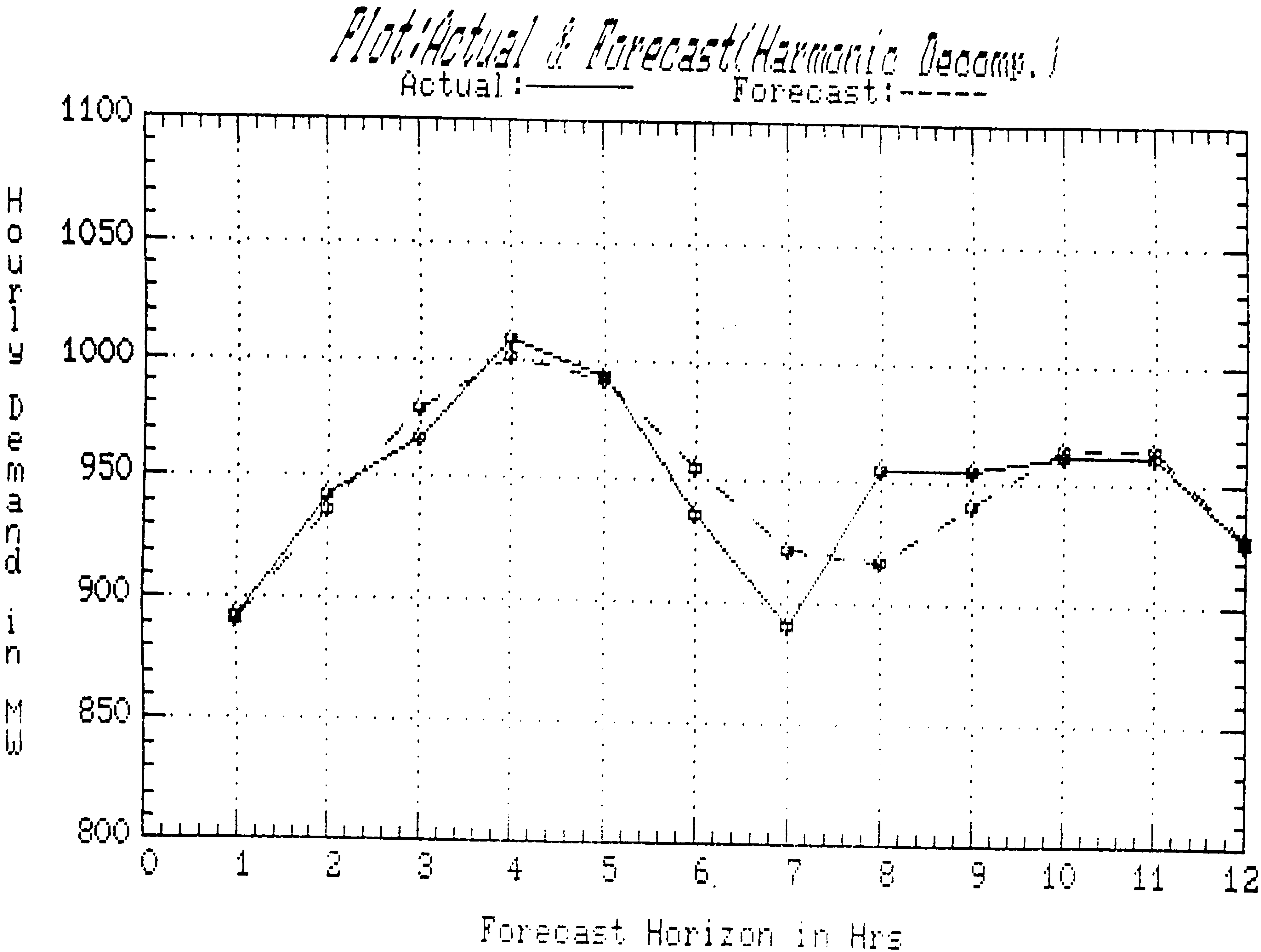


Figure (3-2) Forecasting Results for Short-Term Problem, Seasonal Using Trigonometric Functions and General Exponential Smoothing

3.5 Winters' Seasonal Forecasting: The Additive Model

An alternative technique for forecasting seasonal time series is offered by Winters' procedures [58]. These are two versions of the technique depending on the choice of model representing the evolution of the series as either additive or multiplicative. The present section is devoted to the additive case, beginning with a discussion of the origin of the procedures as traced back to the concept of general exponential smoothing via discounted least squares.

3.5.1 Background

In the formulation of the discounted least squares approach one usually assumes that the series follows a given pattern defined by the elements of the vector function $\underline{h}(j)$. If the series is non-seasonal, then only a trend component is present (m=0) and one has the following three important special cases:

1- Simple exponential smoothing : n=1 and $h_1(j)=1$

2- Double exponential smoothing : n=2 and the trend is assumed to vary as a straight line with the time index.

3- Triple exponential smoothing : n=3, and the trend is quadratic. This is sometimes called Brown's one-parameter quadratic smoothing method.

It is important to note that the smoothed estimates depend on the parameter α, which is usually chosen to minimize the forecast error.

The literature on time series analysis and forecasting includes a number of procedures that can be viewed as modifications to the basic exponential smoothing concept.

3.5.2 Winters' Additive Procedure

Winters' [6] considers the forecasting problem involving seasonal components. The one parameter exponential smoothing procedure can be disadvantageous in situations where the seasonal components are more stable than the trend. Thus in a manner similar to Holt's two parameter procedure, Winters considers updating equations with several smoothing constants. The models considered can be additive or multiplicative. The additive procedure is discussed in Appendix C.

3.5.3 Computational Results : Mid-Term Forecasting

The results of applying Winters' additive seasonal forecasting to the mid-term problem are listed in Table (3-3) and the corresponding graphical display is shown in Figure (3-3). Here with N=130, we obtain $\hat{X}_n(N) = 856.815$, and $\hat{X}_l(N) = 5.628$. With seasonal period s = 12, we get the following seasonal components.

ℓ	$\hat{S}(N+\ell-12)$
1	-140.931
2	-307.573
3	-365.248
4	-360.126
5	-209.364
6	-18.017
7	269.208
8	284.250
9	292.939
10	277.747
11	211.685
12	151.790

Inspection of the forecast errors in Table (3-3) shows that the absolute value of the percentage error is less than 10%. The forecast errors using Winters' additive model are less than those obtained using the general exponential smoothing, as can be seen from a comparison of Tables (3-1) and (3-3).

The fit statistics are;

MSE = 2092

MAPE = 9.206

The forecast statistics are;

MSE = 2,545

MAPE = 4.731

The optimum values of the method's smoothing constants are

$\alpha_1 = 0.05 \quad \alpha_2 = 0.5 \quad \alpha_3 = 0.95$

Table (3-3)

Forecasting Results Mid-Term

Using Winters' Additive Method

MONTH	ACTUAL	FORECAST	ERROR	PERCENT ERROR
131	762.000	721.512	40.488	5.313
132	529.000	560.498	-31.498	-5.954
133	493.000	508.452	-15.452	-3.134
134	528.000	519.202	8.798	1.666
135	616.000	675.592	-59.592	-9.674
136	965.000	872.567	92.433	9.674
137	1084.000	1165.419	-81.419	-7.511
138	1142.000	1186.090	-44.090	-3.861
139	1204.000	1200.407	3.593	0.298
140	1213.000	1190.843	22.157	1.827
141	1209.000	1130.409	78.591	6.500
142	1092.000	1076.143	15.857	1.452

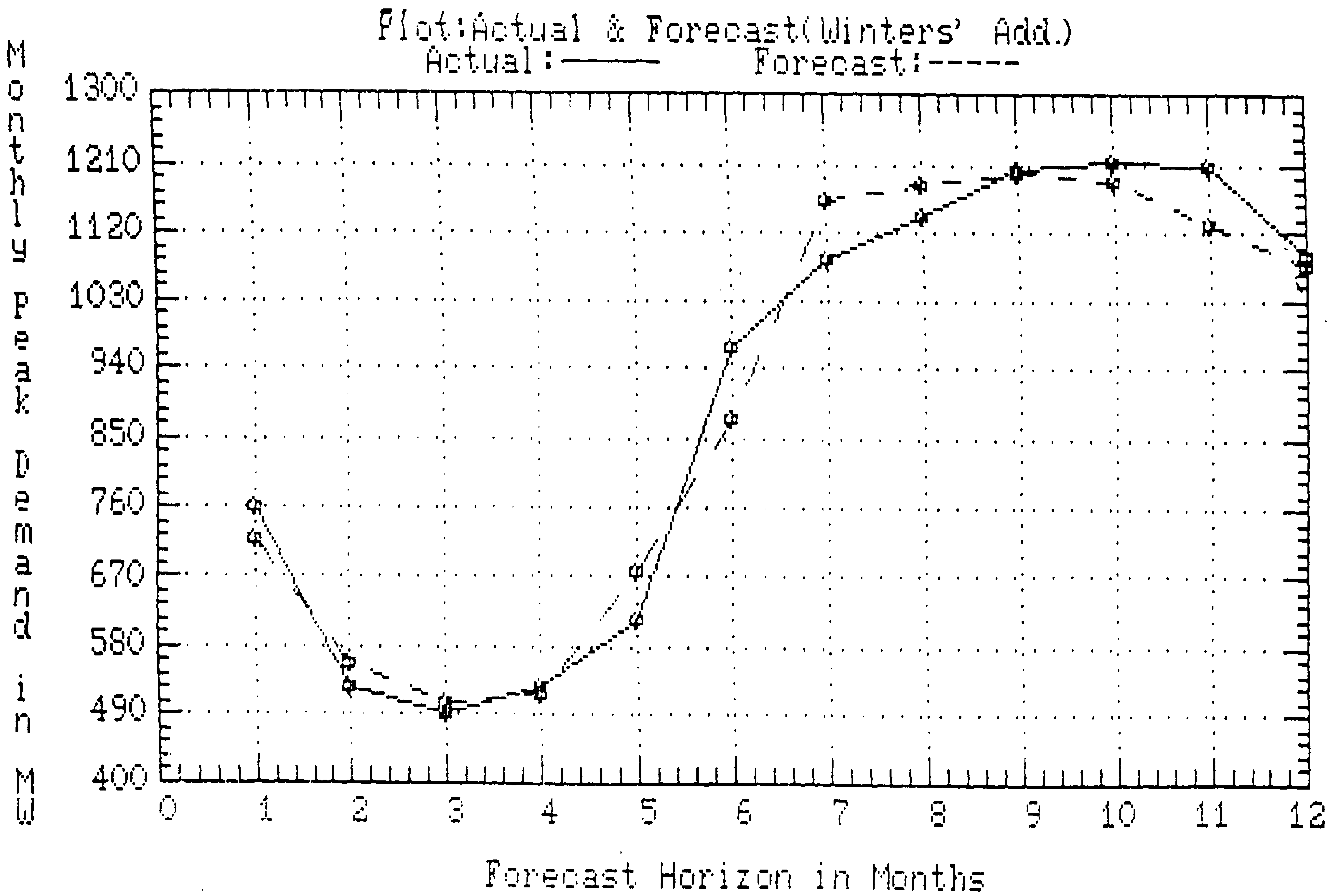


Figure (3-3) Forecasting Results for Mid-Term Problem, Using Winters' Additive Seasonal Forecasting

3.5.4 Computational Results : Short-Term Forecasting

The application of Winters' additive seasonal forecasting to the short-term problem resulted in forecasts listed in Table (3-4) and displayed in Figure (3.4). Here with N = 660, we obtain $\hat{X}_0(N) = 833.909$, and $\hat{X}_1(N) = -0.277$.

The fit statistics are;

MSE = 640.2, MAPE = 1.561

The forecast Statistics are as follows;

MSE = 401.6, MAPE = 1.575

The optimum values of the method's smoothing constants are

$\alpha_1 = 0.15$ $\alpha_2 = 0.1$ $\alpha_3 = 0.7$

From Table (3-4), one observes that the absolute value of the percentage forecast error is less than 5%. The forecast errors using Winters' additive model are consistently higher than those obtained using general exponential smoothing for the short-term problem. It is also noted that the forecast errors for the short-term problem are lower in absolute value than those for the mid-term problem since we have more data for model fitting in the short-term case.

Table (3-4) Forecasting Results

Short-Term Using Winters' Additive Model

HOUR	ACTUAL	FORECAST	ERROR	PERCENT ERROR
661	893.000	886.078	6.922	0.775
662	943.000	945.414	-2.414	-0.256
663	966.000	922.520	43.480	4.501
664	1008.000	1023.330	-15.330	-1.521
665	933.000	1019.938	-26.938	-2.713
666	936.000	967.395	-31.395	-3.354
667	891.000	887.828	3.172	0.356
668	956.000	954.622	1.378	0.144
669	956.000	966.134	-10.134	-1.060
670	962.000	964.704	-2.704	-0.281
671	962.000	988.000	-26.000	-2.703
672	928.000	939.422	-11.422	-1.231

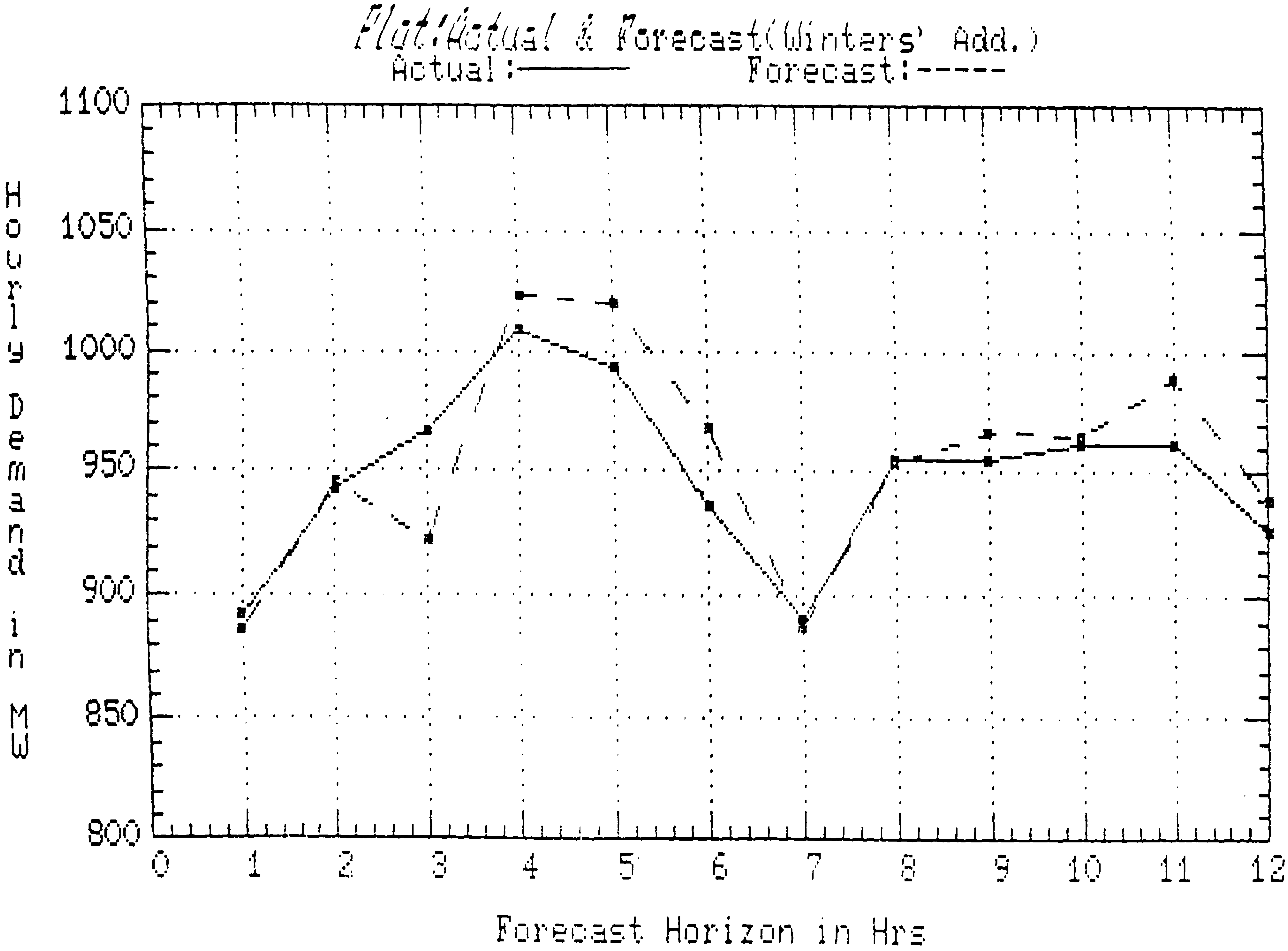


Figure (3-4) Forecasting Results for Short-Term Problem Using Winters' Additive Seasonal Forecasting

3.6 Winters' Seasonal Forecasting : The Multiplicative Model

In the case when seasonal swings and the variations in the errors are proportional to the level, one would transform the data and consider the additive model for the logorithmically transformed observations. On the other hand, if the seasonal swings are proportional to the level, but the errors are not, one can consider the application of Winters' multiplicative version of seasonal exponential smoothing [6] described in Appendix D.

3.6.1 Computational Results : Mid-Term Forecasting

Winters' multiplicative seasonal forecasts for the mid-term problem are listed in Table (3-5) and plotted in Figure (3-5). Here N = 130, yields $\hat{X}_0(N) = 872.117$, and $\hat{X}_1(N) = 5.094$. With seasonal period s = 12, we get the following seasonal components:

ℓ	$\hat{S}(N+\ell-12)$
1	0.799
2	0.601
3	0.526
4	0.527
5	0.732
6	0.973
7	1.243
8	1.286
9	1.328
10	1.337
11	1.261
12	1.161

The fit statistics are;

MSE = 1,422

MAPE = 5.507

The forecast statistics are;

MSE = 1,843

MAPE = 4.498

The optimum values of the method's parameters are

$\alpha_1 = 0.1 \quad \alpha_2 = 0.05 \quad \alpha_3 = 0.3$

Inspection of the forecast errors in Table (3-5), reveals that the absolute value of the percentage forecast error is less than 11%. This may suggest that this model gives an inferior forecast than the preceding two techniques. This is not true as one can verify that the multiplicative model's forecasts at points 132, 135, 137, 138, 141 and 142 are closer to the actual values than forecasts obtained by other methods. This fact is supported by observing that the MSE and MAPE of the forecast statistics using Winters' Multiplicative model are lower than their counterparts for other methods implemented for this problem.

Table (3-5) Mid-Term Forecasting Results
Using Winters' Seasonal Multiplicative Model

MONTH	ACTUAL	FORECAST	ERROR	PERCENT ERROR
131	762.000	701.299	60.701	7.966
132	529.000	530.020	-1.020	-0.193
133	493.000	466.463	26.537	5.383
134	528.000	470.140	57.860	10.958
135	616.000	675.036	-41.036	-6.662
136	965.000	878.464	86.536	8.967
137	1084.000	1128.281	-44.281	-4.085
138	1142.000	1173.678	-31.678	-2.774
139	1204.000	1219.234	-15.234	-1.265
140	1213.000	1234.374	-21.374	-1.762
141	1209.000	1170.436	38.564	3.190
142	1092.000	1083.619	8.381	0.767

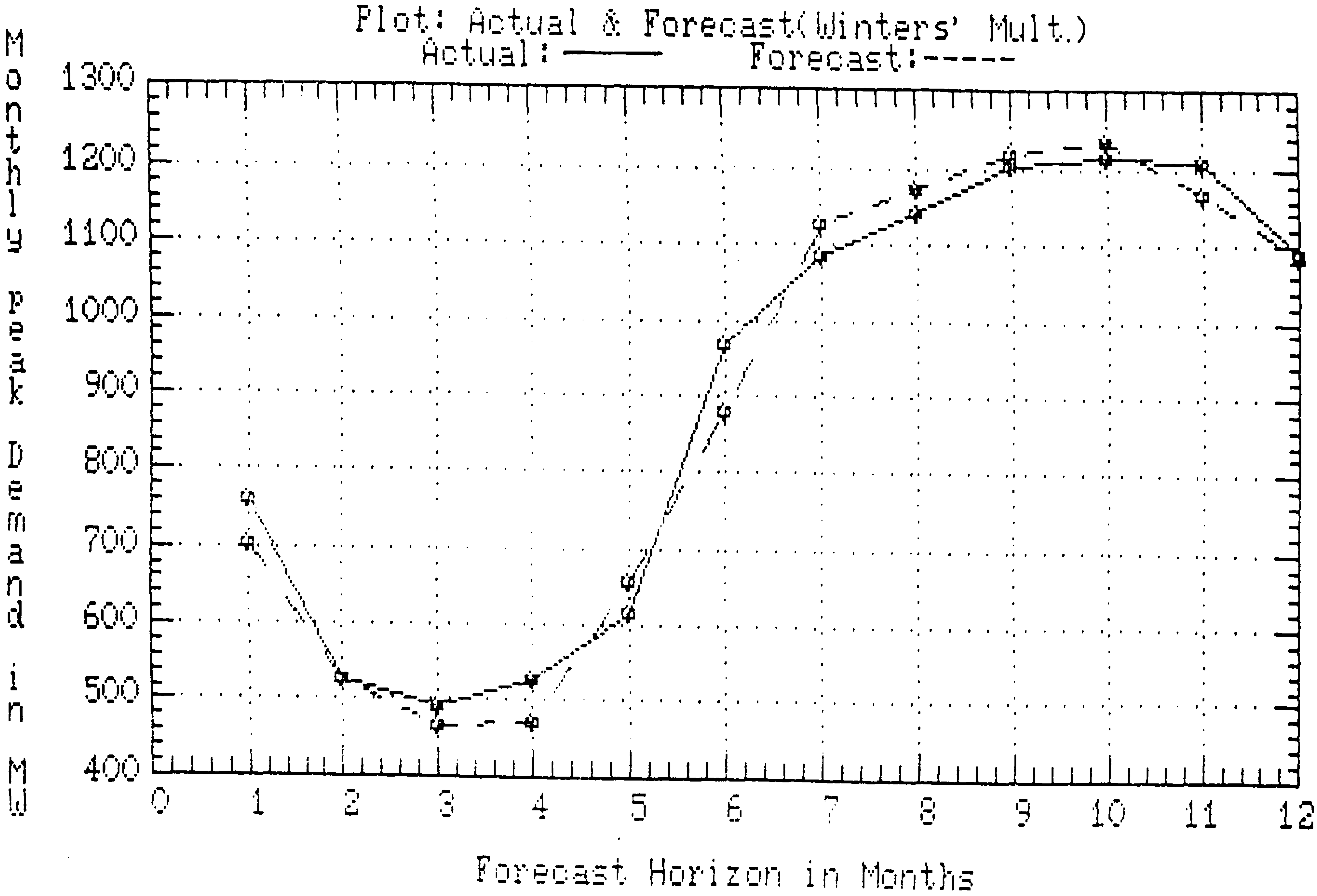


Figure (3-5) Forecasting Results for Mid-Term Problem Using Winters' Multiplicative Seasonal Forecasting

3.6.2 Computational Results : Short-Term Forecasting

The results of applying Winters' seasonal multiplicative forecasts to the short-term problem are given in Table (3-6) and shown in Figure (3-6). Here with N = 660, we get $\hat{X}_0(N) = 836.044$, and $\hat{X}_1(N) = 1.600$.

The fit statistics are;

MSE = 145.5, MAPE = 1.007

The forecast statistics are;

MSE = 100.75, MAPE = 0.966

The optimum values of the method's smoothing constants are

$\alpha_1 = 0.8 \quad \alpha_2 = 0.15 \quad \alpha_3 = 0.2$

By inspecting Table (3-6), one can conclude that for this short term problem the absolute values of the percentage forecast errors are all less than 2%. The Winters' multiplicative model's forecasts are the lowest compared to the results of the previous two methods at data points 663, 666, 671. The MSE and MAPE of the present technique is the lowest compared with the previous two methods. Once again, one notes that the error measures for the short-term problem are considerably lower than those for the mid-term problem.

Table (3-6) Short-Term Forecasting Results
Using Winters' Seasonal Multiplicative Model

HOUR	ACTUAL	FORECAST	ERROR	PERCENT ERROR
661	893.000	900.668	-7.668	-0.859
662	943.000	952.191	-9.191	-0.975
663	966.000	972.940	-6.940	-0.718
664	1008.000	1013.500	-5.500	-0.546
665	933.000	1002.830	-9.830	-0.990
666	936.000	952.479	-16.479	-1.761
667	891.000	877.913	13.087	1.469
668	956.000	941.030	14.970	1.566
669	956.000	953.835	2.165	0.226
670	962.000	955.740	6.260	0.651
671	962.000	957.487	4.513	0.469
672	928.000	940.633	-12.633	-1.361

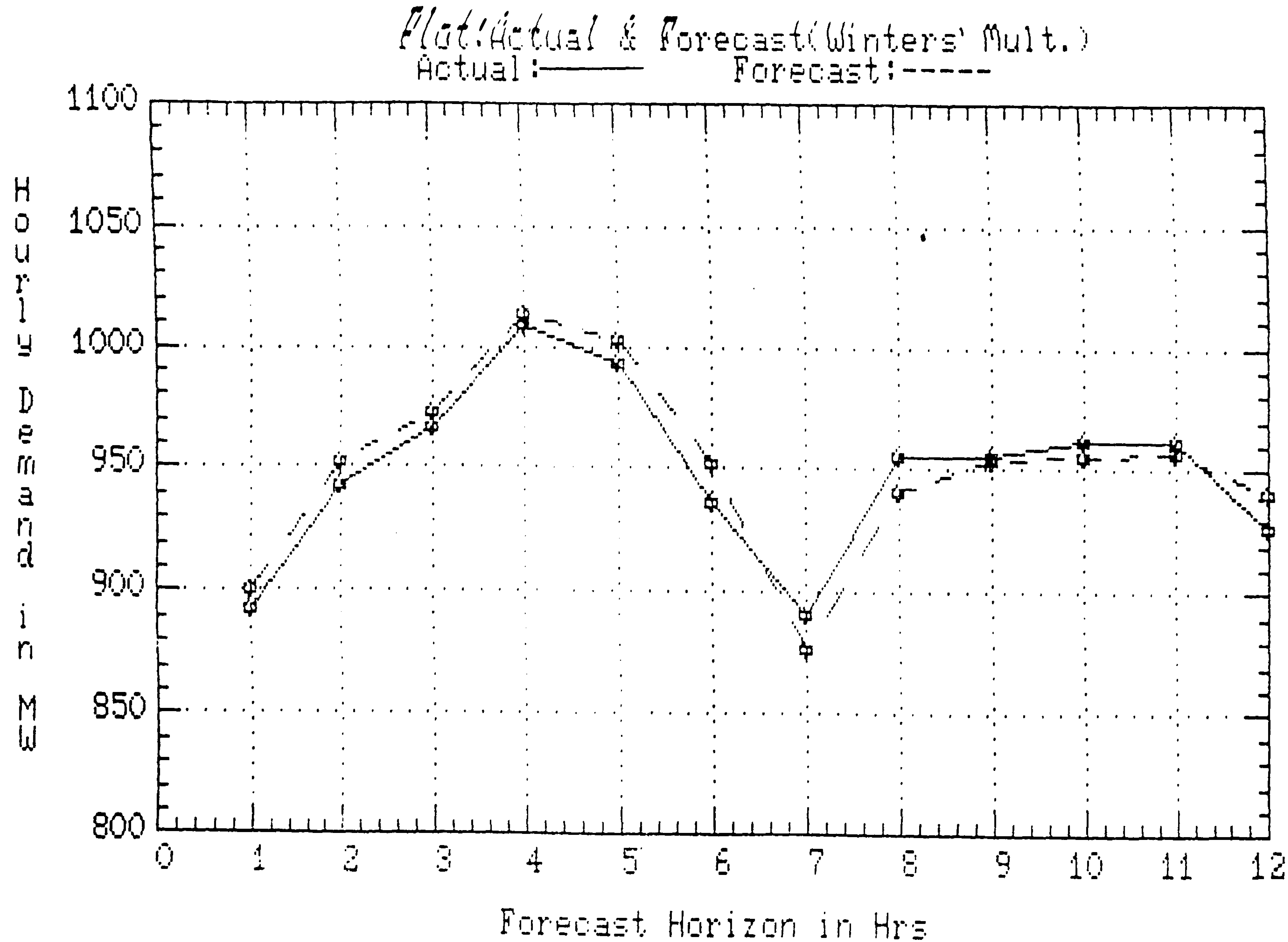


Figure (3-6) Forecasting Results for Short-Term Problem Using Winters' Multiplicative Seasonal Forecasting

3.7 Time Series Seasonal Forecasting via Box-Jenkins Methodology

This section is concerned with the application of time-series analysis forecasting methodology of Box and Jenkins considering the seasonality observed in the data base. The approach involves more general and statistical based techniques for time series analysis involving ARIMA (Autoregressive Integrated Moving Average) processes which have been studied extensively by George Box and Gwilym Jenkins [59] and their names have frequently been used synonymously with general ARIMA models. Autoregressive (AR) models were first intorduced by Yule in 1926 [60] and later generalized by Walker in 1931 [61], while moving average (MA) models were first used by Slutzky [59] in 1937. The theoretical foundations of combined ARMA processes was laid by Wolds' work [62] in 1938.

Box and Jenkins have effectively integrated in a comprehensive manner, the relevant information required to understand and use univariate time series ARIMA models. The basis of their approach is summarized in Figure (3-7) and consists of three stages: identification, estimation, diagnostic testing and application.

At the heart of model specification is the principle of parsimony advocated by Tukey [63] and Box-Jenkins, which can be paraphrased as;

"In a choice among competing hypotheses, other things being equal, the simplest is preferrable". Alternatively, parsimony implies the inclusion of only as many parameters as one really needs. Reasons for preferring simple models over models with a large number of parameters are:

(1) Simple models are easier to understand and interpret.

(2) The estimation of each unnecessary parameter will increase the variance of the prediction error.

A brief discussion of the ingredients of Box-Jenkins methodology is given in Appendix E.

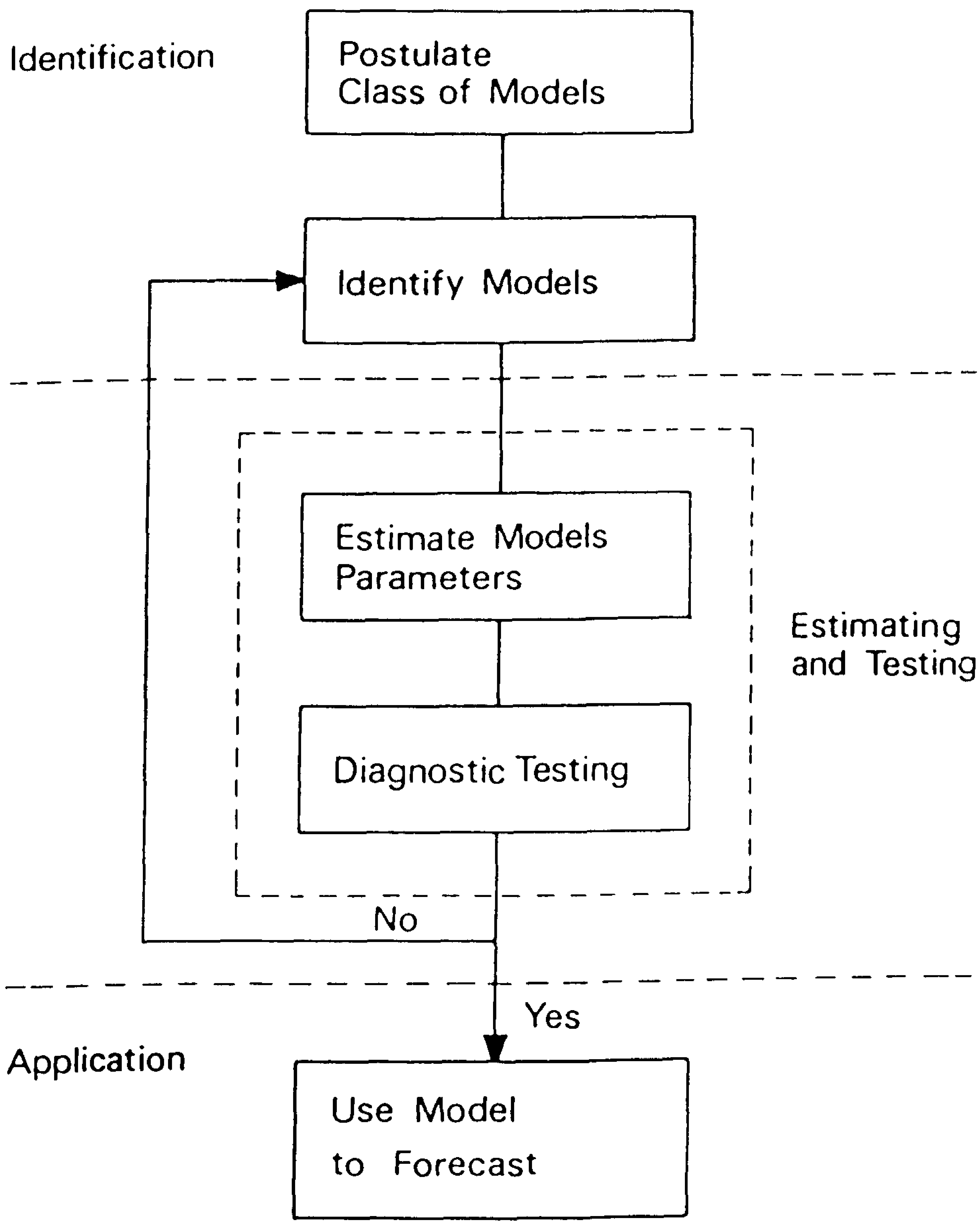


Figure (3-7) Elements of Box-Jenkins Forecasting Methodology

3.7.1 Computational Results : Mid-Term Forecasting

Applying Box-Jenkins forecasting methodology to the mid-term forecasting problem involved experimenting with data transformations in order to achieve a stationary time series as discussed in Appendix E. The transformed time series used is of the form

$$z(k) = [\text{Data}]^{\lambda} \qquad \lambda = 0.5.$$

A seasonal ARIMA model of the following form was found to yield the best results.

$$\Phi_p(B)\ \Phi_P(B^{12})\ \nabla^d\ \nabla_{12}^D\ z(k) = \theta_q(B)\ \theta_Q(B^{12})\ a(k)$$

The model particulars are given by:

D=1 Q=1 P=1

d=0 q=0 p=1

$\Phi_p(B) = 1 - \Phi_1 B$

$\Phi_1 = 0.5873$

$\Phi_P(B) = 1 - \Phi_{12} B^{12}$

$\Phi_{12} = 0.2865$

$\theta_q(B) = 1, \quad \theta_Q(B^{12}) = 1 - \theta_{12} B^{12}$

$\theta_{12} = 0.6522$

The fit statistics are;

MSE = 1384, MAPE = 5.24

The results of applying Box-Jenkins forecasting methodology to the mid-term forecasting problem are listed in Table (3-7) and shown in Figure (3-8).

The following are the forecast statistics;

MSE = 2551, MAPE = 5.30,

Inspection of Table (3-7) reveals that the forecast percent errors are all below 10% except for points 132 and 135. The forecast of points 131, 134, 136 and 141 using the Box-Jenkins model involves the lowest forecast error among the four methods discussed so far. In terms of MSE and MAPE the technique did not perform as well as did the Winters' multiplicative seasonal model.

Table (3-7) Forecast Results for Mid-Term Using Box-Jenkins Methodology

MONTH	ACTUAL	FORECAST	ERROR	PERCENT ERROR
131	762	735.59	26.45	3.47
132	529	583.08	-54.8	-10.22
133	493	519.37	-26.37	-5.34
134	528	524.6	3.39	0.64
135	616	690.8	-74.8	-12.144
136	965	885.18	79.81	8.27
137	1084	1154.39	-70.38	-6.49
138	1142	1209.42	-67.4	-5.9
139	1204	1252.35	-48.35	-4.01
140	1213	1249.56	-36.55	-3.01
141	1209	1182.96	26.04	2.15
142	1092	1113.5	-21.5	-1.96

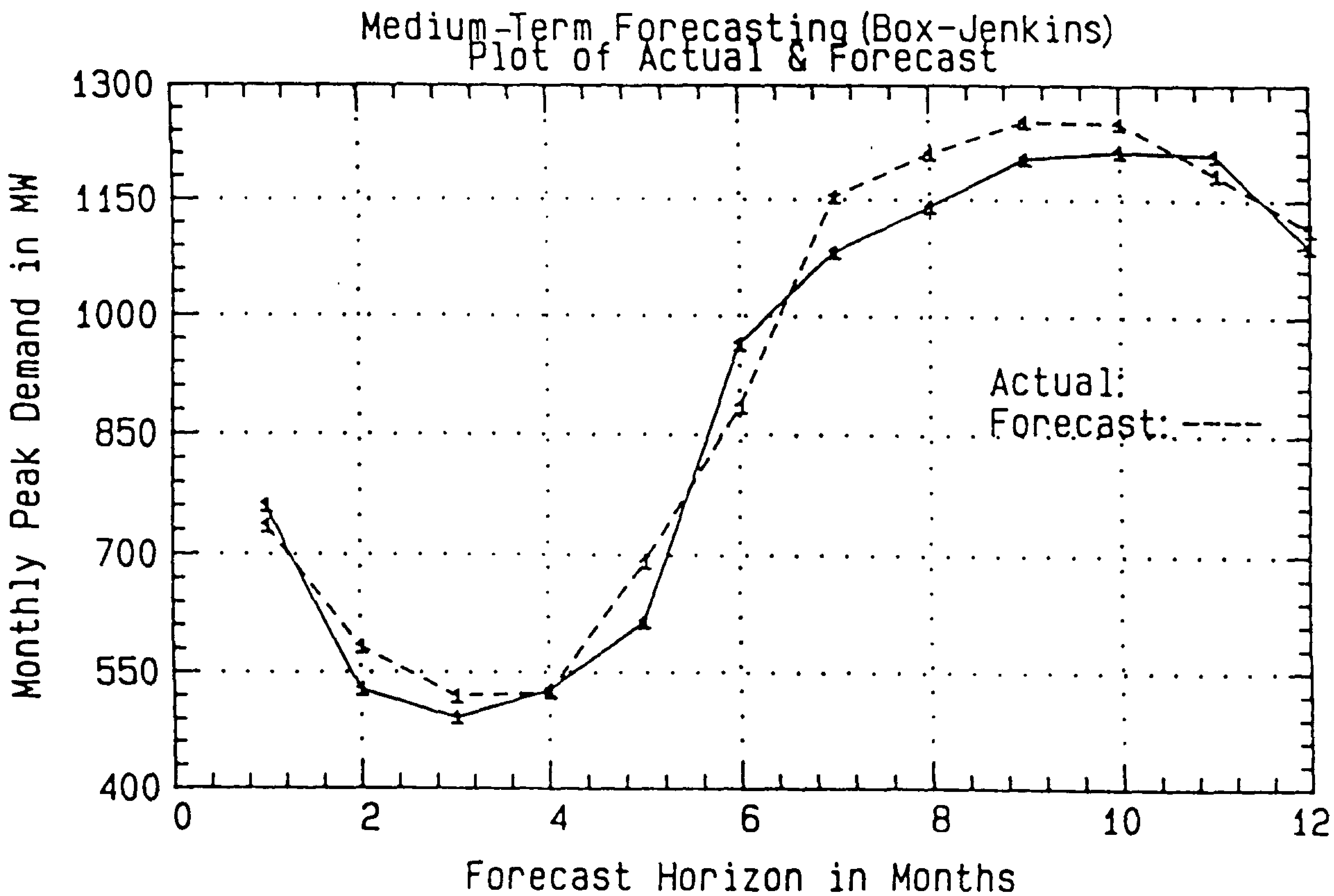


Figure (3-8) Forecast Results for the Mid-Term Problem Using Box-Jenkins Time Series Model

3.7.2 Computational Results : Short-Term Forecasting

The application of Box-Jenkins time series modelling to the short-term forecasting problem was conducted for two distinct applications. In the first, one deals with forecasting for one day, and thus a periodicity of 24 hours is assumed. On the other hand, the second application involves hourly forecasting for one week ahead. This requires a periodicity of 168 hours. The one-week-ahead forecasts are useful for unit commitment purposes. No transformations of the raw data was needed. A seasonal ARIMA model was assumed of the following form:

Model 1: Periodicity 24;

$$\Phi_p(B)\,\Phi_P(B^{24})\,\nabla^d\,\nabla_{24}^D\,z(k) = \theta_q(B)\,\theta_Q(B^{24})\,a(k)$$

The model particulars were obtained as:

$D=1 \quad Q=1 \quad P=0$

$d=1 \quad q=1 \quad p=0$

$\Phi_p(B) = 1 \qquad \theta_q(B) = 1 - \theta_1 B$

$\Phi_P(B) = 1 \qquad \theta_Q(B^{24}) = 1 - \theta_{24}B^{24}$

$\theta_1 = 0.3733 \qquad \theta_{24} = 0.9719$

The fit statistics are;

MSE = 245.88

MAPE = 1.236

The forecast results are listed in Table (3-8-a) and shown in Figure (3-9-a). Validation of Box-Jenkins short-term forecasting model with 24 hour periodicity is shown in Figure (3-9-c). The forecast statistics are;

MSE = 88.26

MAPE = 0.896

Inspection of Table (3-8-a) reveals that the percent forecast errors have all absolute values of less than 2%. This model has a superior performance over the previous methods in terms of MSE and MAPE.

Table (3-8-a) Short-Term Forecasting Results Using Box-Jenkins Methodology with 24 hour periodicity

HOUR	ACTUAL	FORECAST	ERROR	PERCENT ERROR
661	893	894	-1	-0.14
662	943	934	9	0.96
663	966	978	-12	-1.29
664	1008	1005	3	0.26
665	993	1002	-9	-0.86
666	936	948	-12	-1.24
667	891	881	10	1.12
668	956	940	16	1.70
669	956	949	7	0.78
670	962	953	9	0.95
671	962	957	5	0.47
672	928	937	-9	-0.97

The second model with periodicity of 168, has the following parameters:

$D=1 \quad Q=1 \quad P=0$

$d=0 \quad q=0 \quad p=3$

$\Phi_p(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \Phi_3 B^3$

$\Phi_P(B) = 1$

$\Phi_1 = 0.492 \qquad \Phi_2 = 0.1902$

$\Phi_3 = 0.1558$

$\theta(B^Q) = 1 - \theta_{168} B^{168}$

$\theta_q(B) = 1$

$\theta_{168} = 0.8836$

The fit statistics are;

MSE = 171.8

MAPE = 1.052

Table (3-8-b) lists the forecast results and Figure (3-9-b) show a graphical comparison of the forecast and actual values. The forecast statistics are;

MSE = 289.26

MAPE = 1.530

Inspecting Table (3-8-b) reveals that the forecast errors in percent are less than 3% which is inferior to that obtained using a model with 24 hour periodicity. This can also be seen from the MSE and MAPE values.

Table (3-8-b) Short-Term Forecasting Results Using Box-Jenkins Methodology Using the Model with 168 hour Periodicity

HOUR	ACTUAL	FORECAST	ERROR	PERCENT ERROR
661	893	885	8	0.95
662	943	912	31	3.25
663	966	968	-2	-0.26
664	1008	997	11	1.11
665	993	992	1	0.13
666	936	942	-6	-0.61
667	891	864	27	2.99
668	956	931	25	2.63
669	956	944	12	1.25
670	962	946	16	1.65
671	962	945	17	1.78
672	928	912	16	1.77

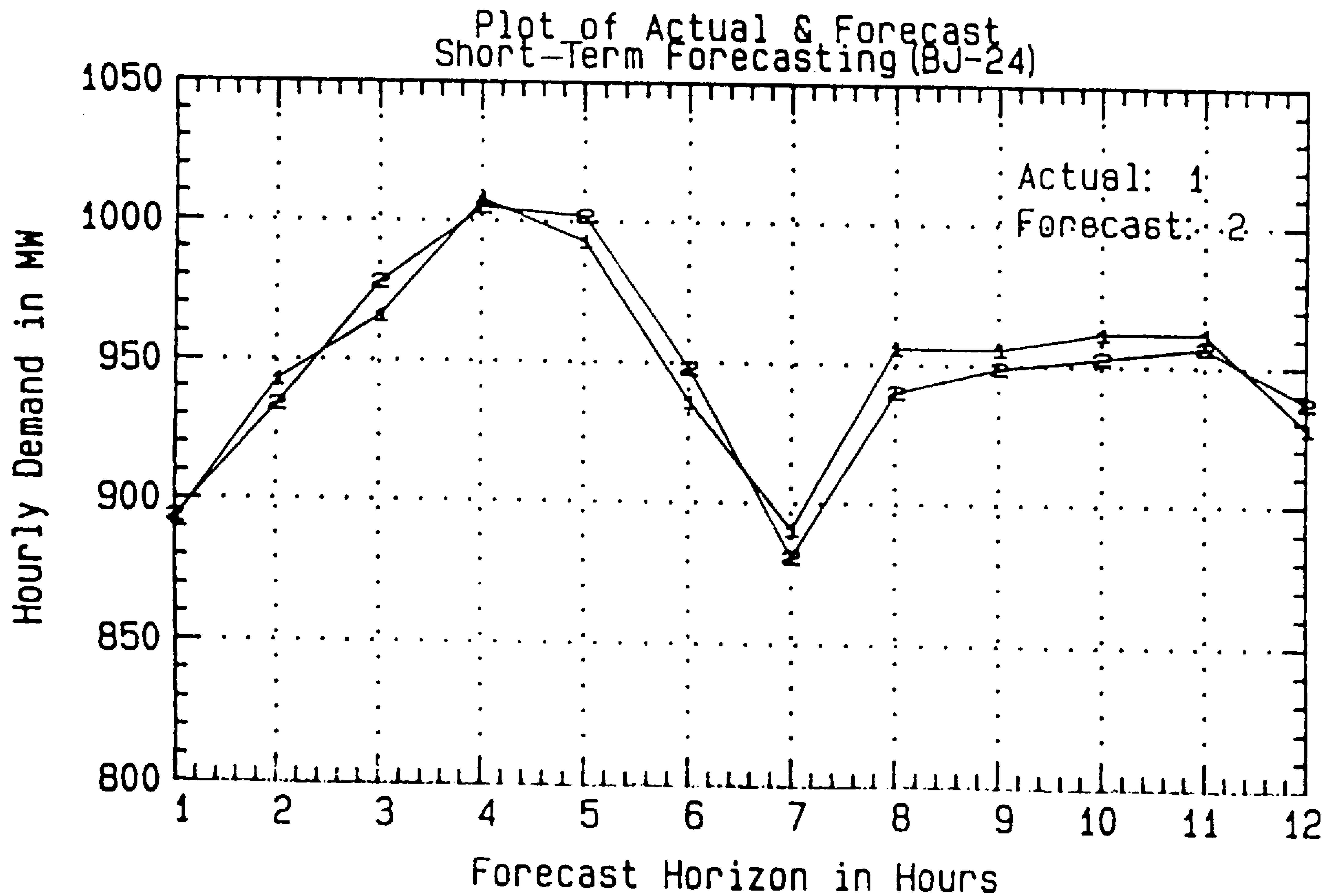


Figure (3-9-a) Forecast Results for the Short-Term Problem Using Box-Jenkins Time Series with 24 hour Periodicity

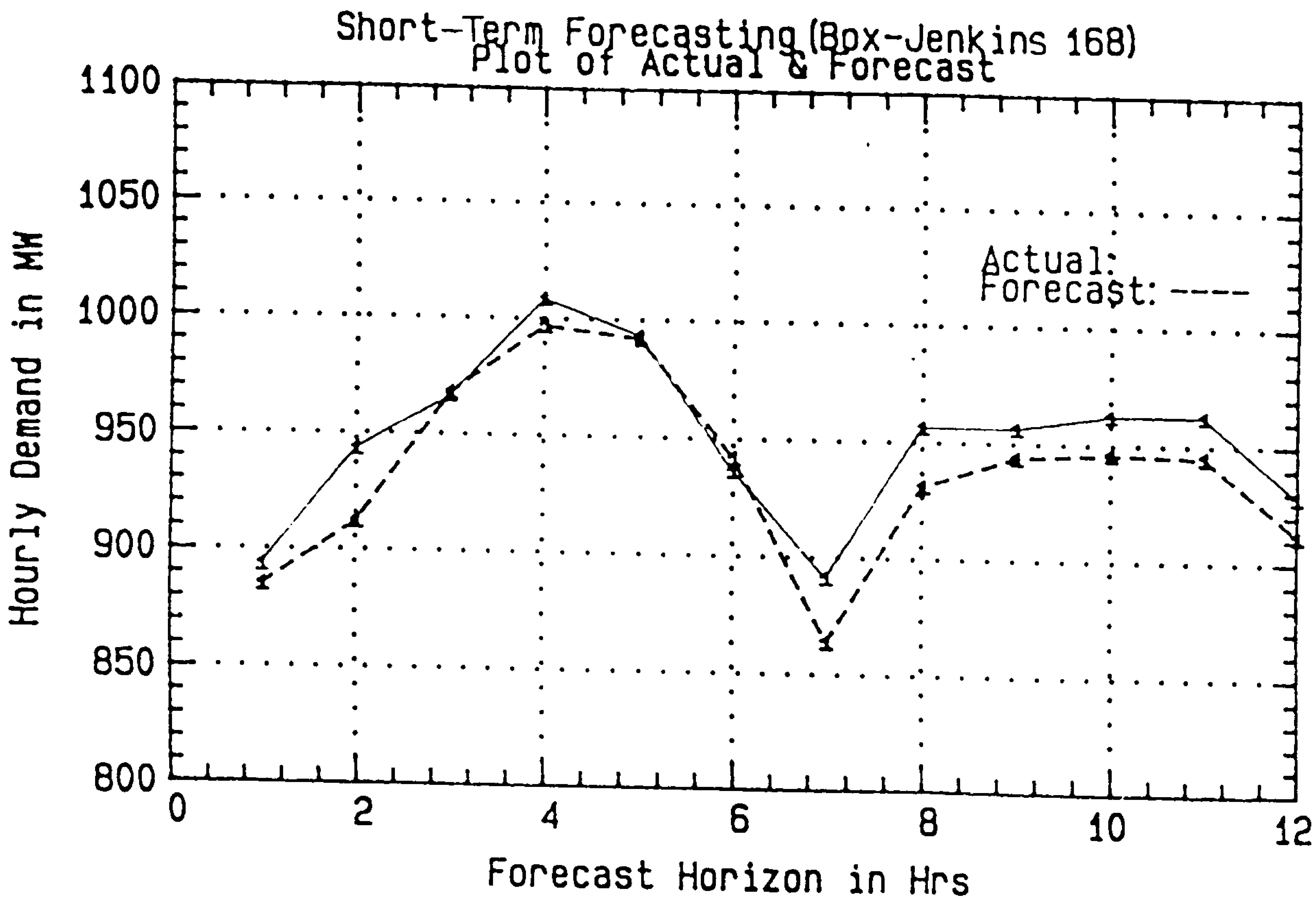


Figure (3-9-b) Forecast Results for the Short-Term Problem Using Box-Jenkins Time Series with 168 hour Periodicity

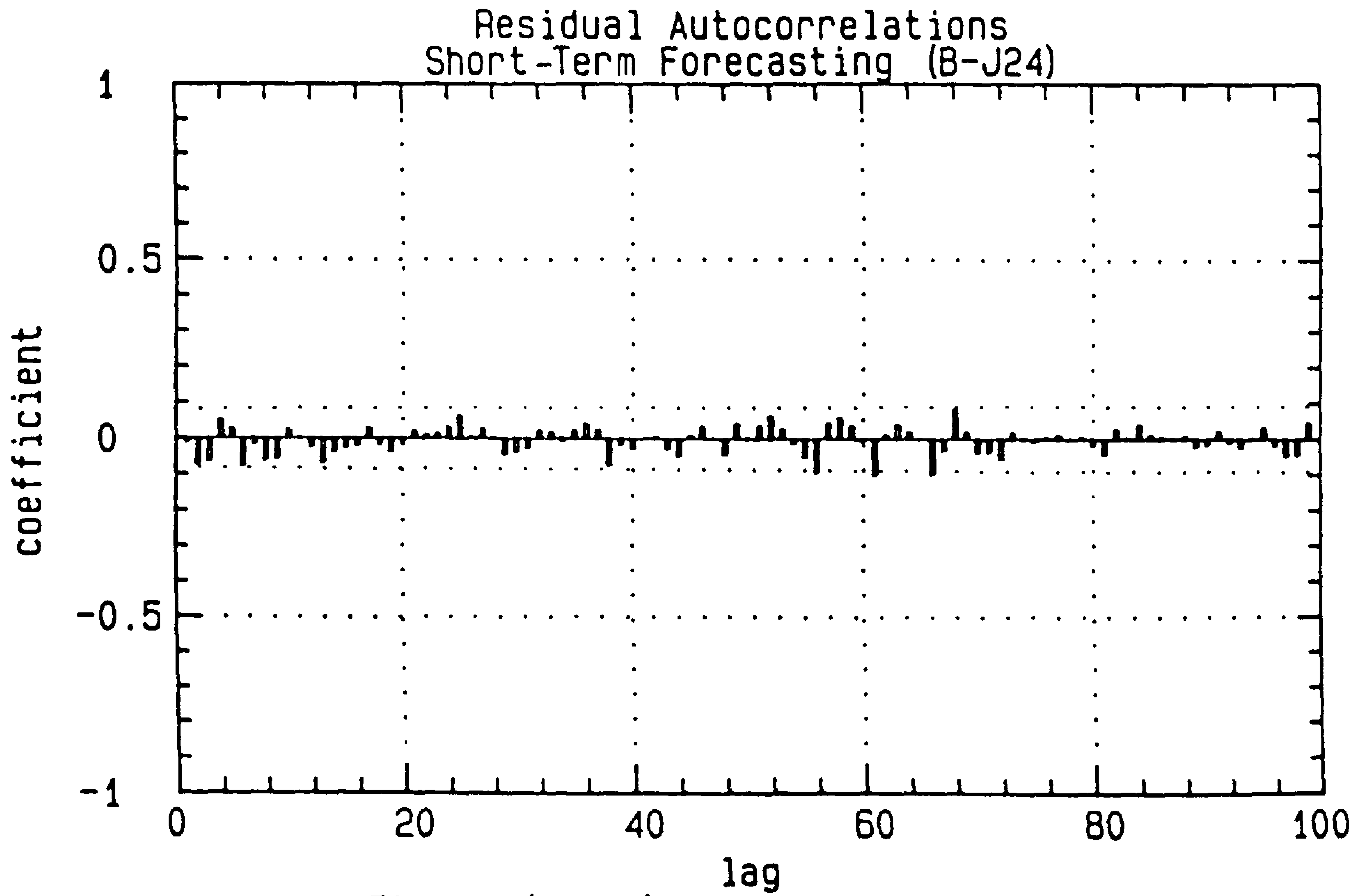


Figure (3-9-c) Validation of Box-Jenkins Short-Term Forecasting Model with 24 Hour Periodicity.

3.8 Load Forecasting via Kalman Filtering Methodology

A Kalman Filter is a recursive data processing algorithm that accepts noise - corrupted measurements to provide an estimate of the "state variables" that describe the behaviour of a dynamic system. The Kalman Filter (K.F.) combines the real-time data with the results of stochastic modelling of system dynamics and measurements device characteristics and the statistical description of the system noises, measurement errors and/or inadequacies in the mathematical model. The K.F. takes into consideration any available a priori statistical information about the system states.

Modern interest in recursive estimation for non-stationary processes over finite-time intervals is due to Kalman [66] and Kalman Bucy [67]. The great interest in recursive algorithms is due to their obvious computational advantages. The "standard" form of K.F. refers to the estimator first given by Kalman [66], from which the discrete optimal filter was derived and subsequently documented in many books [68 to 72].

The filter algorithm is usually given in two sets of equations - one for prediction (extrapolation) and the other for correction (updating). The basic algorithm is discussed in the next section.

3.8.1 Kalman Filtering Algorithm

The Kalman filter algorithm is stated for the linear system represented by the discrete state space model

$$X(k) = \Phi(k, k-1)\, X(k-1) + B_d(k-1)\, u(k-1) + G_d(k-1)\, W_d(k-1) \qquad (3\text{-}7)$$

where $X(\cdot)$ is an n-vector state process, $u(\cdot)$ is an r-vector of control input, $W_d(\cdot)$ is an s-vector. The matrix Φ is a state transition matrix which is an nxn, B_d is an nxr matrix, and G_d is an nxs matrix. The noise input $W_d(\cdot)$ is assumed to be discrete time zero-mean White Gaussian Noise Sequence with covariance kernel

$$E\{\, W_d(k)\, W_d^T(j)\,\} = \begin{cases} Q_d(k) & k=j \\ 0 & k\neq j \end{cases} \qquad (3\text{-}8)$$

Q_k is an sxs matrix which is symmetric and positive semi-definite for all k. Measurements are available at discrete time points 1, 2, ... k, ... , and are modelled by

$$z(k) = H(k)\, X(k) + v(k) \tag{3-9}$$

where $z(k)$ is an m_n-vector discrete time measurement process, $H(\cdot)$ is an mxn measurement matrix, and $v(\cdot)$ is an m_n-vector discrete time white Gaussian noise with statistics:

$$E\{v(k)\} = 0$$

$$E\{v(k)\, v^T(j)\} = \begin{cases} R(k) & k=j \\ 0 & k\neq j \end{cases} \tag{3-10}$$

In this description $R(k)$ is an $m_n x m_n$, symmetric, positive definite matrix for all k. It is further assumed that $v(k)$ and $W_d(k)$ are uncorrelated.

The state X at k=0, is modelled as a random vector that is normally distributed with mean $\hat{X}(0)$ and covariance $P(0)$:

$$E\{X(0)\} = \hat{X}(0) \tag{3-11}$$

$$E\left\{[X(0) - \hat{X}(0)]\,[X(0) - \hat{X}(0)]^T\right\} = P(0) \tag{3-12}$$

where P(0) is an nxn matrix that is symmetric and positive semidefinite.

The object of a Kalman Filter is to combine the measurement data taken from the actual system with the information provided by the system and measurement models and statistical description of uncertainties in order to obtain an optimal estimate of the system state.

It is convenient to introduce a composite vector which comprises the entire measurement history to the current time k, and denote it as $Z(k)$, where

$$Z(k) = \begin{bmatrix} z(1) \\ z(2) \\ \vdots \\ z(k) \end{bmatrix} \tag{3-13}$$

we consider two measurement times $(k-1)$ & k, and propagate optimal estimates from the point just after the measurement at time $(k-1)$ has been incorporated into the estimate to the point just after the measurement at time k is incorporated. This is shown in Figure (3-10) as propagating from time $(k-1)^+$ to time $(k)^+$. We define $\hat{X}_+(k-1)$ and $P_+(k-1)$ to be the conditional mean and conditional covariance of the optimal estimate:

$$\hat{X}_+(k-1) \triangleq E\{X(k-1) \mid Z(k-1)\} \tag{3-14}$$

$$P_+(k-1) \triangleq E\left\{ [X(k-1)-\hat{X}_+(k-1)] \, [X(k-1)-\hat{X}_+(k-1)]^T \middle| Z(k-1) \right\} \tag{3-15}$$

We also let $\hat{X}_-(k)$ denote the conditional mean of $X(k)$ before the measurement $z(k)$ is taken and processed:

$$\hat{X}_-(k) \triangleq E\{X(k) \mid Z(k-1)\} \tag{3-16}$$

The conditional covariance of $X(k)$ before the measurement $z(k)$ is taken and processed is thus

$$P_-(k) \triangleq E\left\{ [X(k)-\hat{X}_-(k)] \, [X(k)-\hat{X}_-(k)]^T \,\middle|\, Z(k-1) \right\} \tag{3-17}$$

The Kalman Filtering algorithm propagates the optimal state estimates from measurement time $(k-1)$ to time k by a set of recursive linear relations given by

$$\hat{X}_-(k) = \Phi(k,k-1)\,\hat{X}_+(k-1) + B_d(k-1)\,u(k-1) \tag{3-18}$$

$$P_-(k) = \Phi(k,k-1)\,P_+(k-1)\,\Phi^T(k,k-1) + G_d(k-1)\,Q_d(k-1)\,G_d^T(k-1) \tag{3-19}$$

This gives the mean and covariance of the optimal state estimate at time k, before incorporating the measurement $z(k)$. The estimate is updated by defining the Kalman filter gain $K(k)$ by

$$K(k) = P_-(k)\,H^T(k)\,A^{-1}(k) \tag{3-20}$$

where

$$A(k) = H(k)\,P_-(k)\,H^T(k) + R(k) \tag{3-21}$$

and the measurement residual $r(k)$ (usually called the innovations) is the difference between the measurement value $z(k)$ and its best prediction before it is actually taken as

$$r(k) = z(k) - H(k)\,\hat{X}_-(k). \tag{3-22}$$

As a result we have the updated estimate

$$\hat{X}_{+}(k) = \hat{X}_{-}(k) + K(k)\, r(k), \qquad (3\text{-}23)$$

where the error covariance is given by

$$P_{+}(k) = [I - K(k)\, H(k)]\, P_{-}(k). \qquad (3\text{-}24)$$

The mechanics of the Kalman filtering algorithm can best be understood in terms of two stages called the predictor and corrector stages. In the predictor stage, knowledge of the optimum state estimate $\hat{X}_{+}(k-1)$ and its error covariance matrix $P_{+}(k-1)$ allows us to compute the estimates $\hat{X}_{-}(k)$ and $P_{-}(k)$, based on the system model parameters $\Phi(k, k-1)$, $G_d(k-1)$, and $Q_d(k-1)$. In the corrector stage, we incorporate the measurement $z(k)$ by computing the residual $r(k)$, and correcting $\hat{X}_{-}(k)$ to obtain the optimal estimate $\hat{X}_{+}(k)$ using Kalman's gain matrix $K(k)$. The latter is obtained on the basis of $A(k)$ computed from results of the predictor stage. The optimal error covariance matrix $P_{+}(k)$ is obtained using $K(k)$ and $P_{-}(k)$ as depicted in Figures (3-10 and 3-11).

It is important to realize that a Kalman filter requires knowledge of the optimal initial state mean $X(0)$, its error covariance matrix $P(0)$, system and measurement model parameters $\Phi(k, k-1)$, $B_d(k-1)$, $G_d(k-1)$, $H(k)$, and the associated noise covariance matrices $Q_d(k)$ and $R(k)$.

$\hat{X}_+(k-1)$ $\quad\quad$ $P_+(k-1)$

↓ ↓

Predictor Equations

$$\hat{X}_-(k) = \Phi(k,k-1)\,\hat{X}_+(k-1)$$

$$P_-(k) = \Phi(k,k-1)\,P_+(k-1)\,\Phi^T(k,k-1) + G_d(k-1)\,Q_d(k-1)\,G_d^T(k-1)$$

↓ ↓

$\hat{X}_-(k)$ $\quad\quad$ $P_-(k)$

↓ ↓

Corrector Equations

$$A(k) = H(k)\,P_-(k)\,H^T(k) + R(k)$$

$$K(k) = P_-(k)\,H^T(k)\,A^{-1}(k)$$

$$r(k) = z(k) - H(k)\,\hat{X}_-(k)$$

$$\hat{X}_+(k) = \hat{X}_-(k) + K(k)\,r(k)$$

$$P_+(k) = [I - K(k)\,H(k)]\,P_-(k)$$

↓ ↓

$\hat{X}_+(k)$ $\quad\quad$ $P_+(k)$

Figure (3-10) The Predictor Corrector form of K.F.

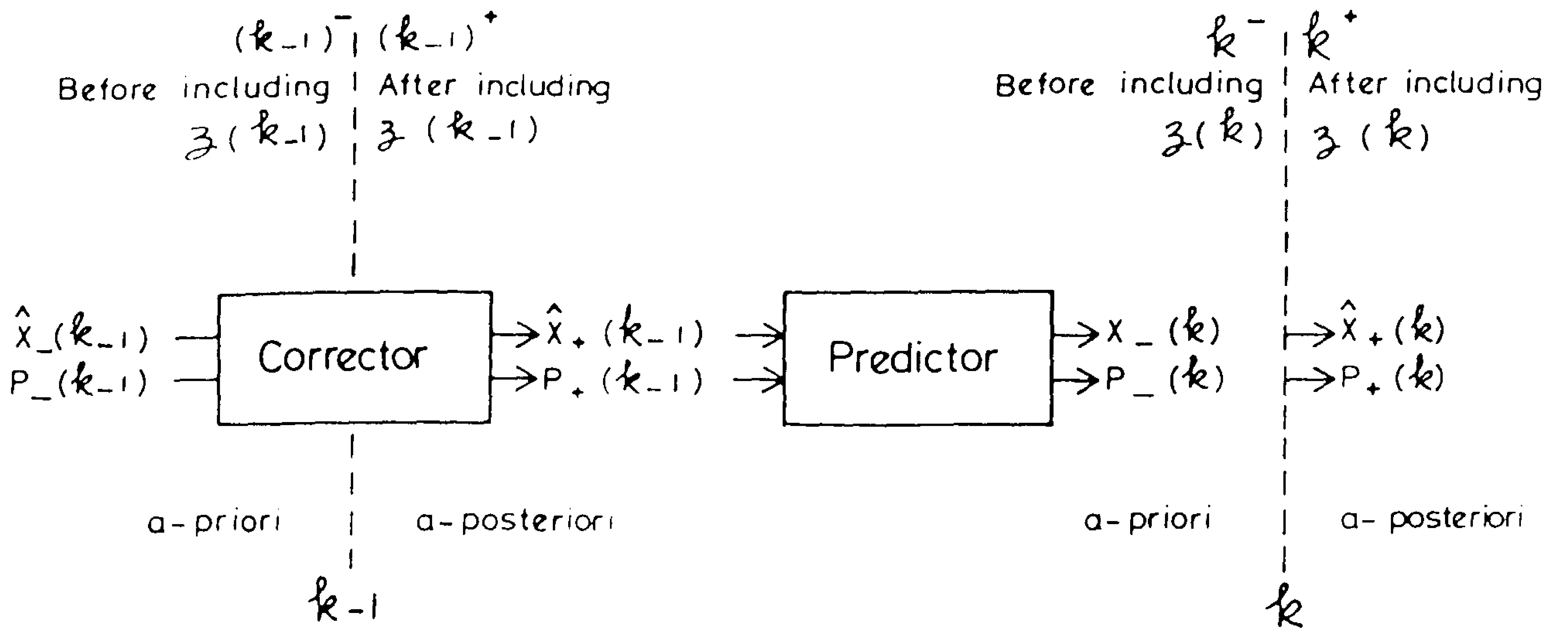


Figure (3-11)

Timing Diagram Illustration Of
The Predictor - Corrector From Of K.F.

3.8.2 State Space Model of Load Function

The successful implementation of a K.F. as a tool in forecasting power system load requires choosing an appropriate state space model that reflects the salient features of the system dynamics. There are a number of possibilities depending on the time series representation chosen. In this research work we adopted a seasonal model using trigonometric functions of the form of Equation (3-6), as it appeared to represent the load evolution closely. The load $z(k)$ therefore varies with time according to a linear trend plus harmonics to give a model of the form:

$$z(k) = \beta_0 + \beta_1 k + \sum_{i=1}^{m} B_{i1} \sin\omega_i k + \beta_{i2} \cos\omega_i k + v(k) \qquad (3\text{-}25)$$

As shown in Appendix B, this can be written in the compact form:

$$z(k) = f^T(k)\, a(k) + v(k) \qquad (3\text{-}26)$$

where the fitting function vector is introduced as

$$f^T(k) = [\,1 \;\; k \;\; \sin\omega_1 k \;\; \cos\omega_1 k \;\; \ldots \;\; \sin\omega_m k \;\; \cos\omega_m k\,] \qquad (3\text{-}27)$$

The parameter vector is defined by

$$a^T(k) = [\,\beta_0 \;\; \beta_1 \;\; \beta_{11} \;\; \beta_{12} \;\; \cdots \;\; \beta_{m1} \;\; \beta_{m2}\,] \qquad (3\text{-}28)$$

The dynamics of the parameter vector $a(k)$ are assumed to evolve according to a random walk model

$$a(k+1) = a(k) + \omega(k) \qquad (3\text{-}29)$$

where $\omega(k)$ is a noise vector. Note that Equations (3-26) & (3-28) are of the same form as the measurement Equation (3-9) and the system dynamic Equation (3-7) respectively.

As it stands the state space model (3-26) & (3-29) is characterized by time varying matrices $\Phi(k, k-1)$ and $H(k)$. Recall that the fitting functions vector defined in Equation (3-28), satisfies the transition property given in Equation (3-30) as

$$f(k+1) = L f(k) \qquad (3\text{-}30)$$

One can show that

$$f(k) = L^k f(0) \qquad (3\text{-}31)$$

This fact allows us to convert the state space model to a form that has time invariant Φ and H as follows.

Using Equation (3-31) in (3-26), we obtain

$$z(k) = f^T(0)\,[L^T]^k\, a(k) + v(k) \tag{3-32}$$

We define the state vector $X(k)$ in terms of $a(k)$ and the L matrix by:

$$X(k) = [L^T]^k\, a(k) \tag{3-33}$$

We also substitute

$$H = f^T(0) \tag{3-34}$$

As a result, the measurement model (3-32) is written as

$$z(k) = H\,X(k) + v(k) \tag{3-35}$$

We now take (3-29) and pre-multiply both sides by $[L^T]^{k+1}$ to obtain

$$[L^T]^{k+1}\, a(k+1) = L^T\,[L^T]^k\, a(k) + [L^T]^{k+1}\,\omega(k)$$

Using Equation (3-33) we therefore get

$$X(k+1) = L^T\,X(k) + [L^T]^{k+1}\,\omega(k) \tag{3-36}$$

We now let:

$$\Phi = L^T \tag{3-37}$$

$$W_1(k) = [L^T]^{k+1}\,\omega(k) \tag{3-38}$$

As a result, our state model is

$$X(k+1) = \Phi\,X(k) + W_1(k) \tag{3-39}$$

The desired representation is therefore given by (3-35) and (3-39).
To summarize, we have

$$\Phi = \mathrm{diag}\,[L_1^T, L_2^T] \tag{3-40}$$

with

$$L_1^T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \tag{3-41}$$

$$L_2^T = \mathrm{diag}\,[L_{2i}^T] \tag{3-42}$$

$$L^T_{2i} = \begin{bmatrix} \cos\omega_i & -\sin\omega_i \\ \sin\omega_i & \cos\omega_i \end{bmatrix} \tag{3-43}$$

$$H = f^T(0) = [10\ 01\ 0\ \ldots\ 01] \tag{3-44}$$

The conversion from the original parameter vector $a(k)$ to the state vector $X(k)$ is governed by Equation (3-33) and the conversion of the noise vector $\omega(k)$ to the new noise vector $W_d(k)$ is governed by Equation (3-38). In both cases the k^{th} power of L is required. It is easy to verify that

$$|L^T|^k = \text{diag}\left[|L_1^T|^k, |L_2^T|^k \right] \tag{3-45}$$

with $$|L_1^T|^k = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \tag{3-46}$$

$$|L_{2i}^T|^k = \begin{bmatrix} \cos k\omega_i & -\sin k\omega_i \\ \sin k\omega_i & \cos k\omega_i \end{bmatrix} \tag{3-47}$$

3.8.3 Approaches to Adaptive Filtering

Kalman Filtering requires the availability of lumped model parameters represented by elements of the matrices Φ, B_d, G_d and H specified in Equation (3-7) and (3-9) as well as statistical characterization of the system and measurement model noises represented by Q_d and R of (3-8) and (3-10). If these parameters are specified accurately, then the Kalman Filtering algorithm provides an optimal state estimate. Any errors in the modelling effort will yield a sub-optimal filter. During the second half of the 1960s, a number of papers investigated errors in K.F. when it is implemented using various approximations. For example, Heffes [73] studied the actual error covariance matrix for the discrete case by considering uncertainties in P_0, Q_d, and R_i. Nishimura [74] extended this study to include the continuous time case. Price [75] derived recursive equations for the sub-optimal estimation error covariance matrix in the discrete-time case.

Adaptive filtering is the term used to describe techniques which deal with filtering when one or more of $\alpha = \{\Phi, B_d, G_d, Q_d, R\}$ are unknown. The object of adaptive filtering is to attain an on-line capability that learns the unknown data based not only on measurements, but also on measurement residuals as they are generated in real-time.

The process essentially uses the new information to adapt the filter gains and model coefficients to the measurements. A parameter-adaptive filter estimates elements of Φ, B_d, and G_d that are unknown. On the other hand, a noise-adaptive filter is one that estimates unknown Q_d and/or R.

An early survey of adaptive filtering techniques can be found in Mehra [76]. The subject is treated in Chin's review [77]; Maybeck's monograph [78], Isaksson's paper [79], and briefly in Stengel [80].

It is important to note that our formulation of the load forecasting problem as a Kalman filtering problem as given by Equation (3-7) to (3-10), with defining Equations (3-35) and (3-39), does not require parameters in H and Φ that are related to the physical process. Here it is assumed that the radian frequencies ω_i are precisely known. The uncertainties involved in our model are related to Q_d and R entries.

As a result, the present review concentrates on the noise-adaptive filtering problem.

Approaches to adaptive filtering can be broadly classified into the following categories:

1- Bayesian Based Approach
2- Maximum Likelihood Approach
3- Correlation Based Techniques
4- Covariance Matching.

As a general rule, if the number of unknowns is small then satisfactory results may be obtained. The problem of identifying R appears to be easier than that of identifying Q_d using all approaches. A review of the main techniques in each category is given in Appendix F.

3.8.4 Application of K.F. Methodology to Load Forecast

In this section, we report on computational experience with the application of Kalman filtering methodology to forecasting the hourly load on the Abu Dhabi power system using the state space model detailed in Section 3.8.2. The data set used is that listed in Appendix A which consists of 672 data points. Forecasts for 12 hours ahead were obtained and the evaluation of the method is based on comparing the forecast based on the preceeding 660 points to the actual load in each of the hours considered.

To specify the model, we need the values of angular frequecies ω_i and their number, as indicated in Equation (3-25). The weekly pattern was considered as the basis and therefore a fundamental frequency ω_1 was chosen as

$$\omega_1 = \frac{2\pi}{168}$$

On the basis of a spectral analysis of the data set, it was determined that the fundamental, the seventh, fourteenth, twenty-first and twenty-eighth harmonics were dominant in terms of their energy content. As a result, we have

$$\omega_2 = 7\omega_1 \qquad \omega_3 = 14\omega_1$$
$$\omega_4 = 21\omega_1 \qquad \omega_5 = 28\omega_1$$

The number of parameters to estimate is therefore n=12.

Table (3-9) Elements of the Matrix P_0

POSITION	ENTRY	POSITION	ENTRY
(1,1)	4.5548	(7,7)	2.2053
(2,2)	3.0489×10^{5}	(8,8)	2.2050
(3,3)	2.2921	(9,9)	2.2051
(4,4)	2.2050	(10,10)	2.2050
(5,5)	2.2067	(11,11)	2.2050
(6,6)	2.2050	(12,12)	2.2050

To initialize the filtering process, one needs the error covariance matrix P(0). This was obtained for the data set using conventional least squares techniques. The diagonal elements of P(0) are listed in Table (3-9). The measurement noise covariance was obtained as R = 740.

To complete the model specification for Kalman filtering implementation, we need to specify Q_d, if one chooses a fixed R. It is appropriate at this point to comment on some physically based interpretations of the process model covariance matrix Q_d.

Inspection of Equation (3-19) reveals that $P(k)$ is greater than Q_d, and therefore $P(k)$ is prevented from getting too small. From Equation (3-20) we see that as a result the gain K is prevented from going to zero. A large value of Q_d causes a large gain to weigh newer innovations $r(k)$ more heavily at the expense of old data, as can be seen from Equation (3-23). We therefore conclude that Q_d controls the rate at which old data are forgotten. Moreover, it is on the basis of Q_d, that the tracking speed (or rate of change of state values) of the estimator does change.

Since Q_d affects $P(k)$, it therefore affects estimation accuracy. A large value of Q_d is reflected as a large error covariance $P(k)$ and therefore estimates become more erratic. On the other hand, a small Q_d results in lower estimates $\hat{X}$ and P. For very small values of Q_d, $\hat{X}$ will lag behind X and a biased estimate will result.

In light of the preceding discussion, it is clear that care must be taken in selecting Q_d. Ideally, one would choose one of the approaches reviewed in Appendix F, provided that satisfactory results can be obtained. It is noted, however, that of the many approaches reported, the reliable ones require a considerable computational effort involving iterative non-linear equation solution. It seems appropriate for our purposes to adopt a simpler approach that can provide reliable results while minimizing the computational burden. We take a maximum likelihood approach, and employ a simple search technique as discussed next.

Optimal Likelihood Choice of Q_d

We have chosen to experiment with the effect of varying Q_d as a multiple of the covariance matrix P_n. For each value considered, the following log likelihood function was evaluated

$$\ln(f) = -0.5N\,\ln 2\pi - 0.5\,\ln\left|A(k)\right| - \left[\sum_{i=1}^{N} r^T(i)\,A^{-1}(i)\,r(i)/N\right]$$

This log likelihood function is the same as that of Equation (F-4) except for the elimination of the a posteriori terms (containing $P_+(k)$ and $\hat{X}_+(k)$ and some constants). This a priori log likelihood function is consistent with Harvey's [101] and Sallas and Harville [102].

The variation of ln (f) with Q_d as multiple of P_0 is listed in Table (3-10) and is plotted in Figure (3-12). With this simple search technique, it was determined that with R=740, the maximum likelihood is attained for $Q_d=4P_0$.

Table (3-11) lists the 12-hour-ahead forecast results for three estimators. In the table the actual power demand is listed along with forecast values. Forecast 1 corresponds to $Q_d = 3P_0$, forecast 2 corresponds to $Q_d = 4P_0$, and forecast 3 corresponds to $Q_d = 5P_0$. Table (3-12) lists the corresponding percent forecast errors. Figure (3-13) shows the forecast demand in comparison with the actual demand.

The statistical summary of Table (3-13) gives the mean square error and maximum absolute percent error (MAPE) for fitted and forecast quantities. In terms of fitting the data, it is clear that the multiplier of 5 offers lower error statistics. On the other hand, in terms of forecast MSE, the multiplier of 3 is better. The maximum likelihood criterion selects the multiplier of 4, which offers a compromise between fitting and forecasting accuracy.

Figures (3-14) and (3-15) show the spectrum of data and the corresponding residuals for this filter. It is evident that the fitting and forecasting resulted in good reduction in the residuals spectrum.

From an overall comparison point of view, it appears that Kalman filtering offers an excellent forecasting error as measured by the MAPE criterion, being better then Winters' additive and Box-Jenkins with 168 hours periodicity. Kalman filtering results, as measured by the MSE criterion, do not rank as high as with the MAPE criterion.

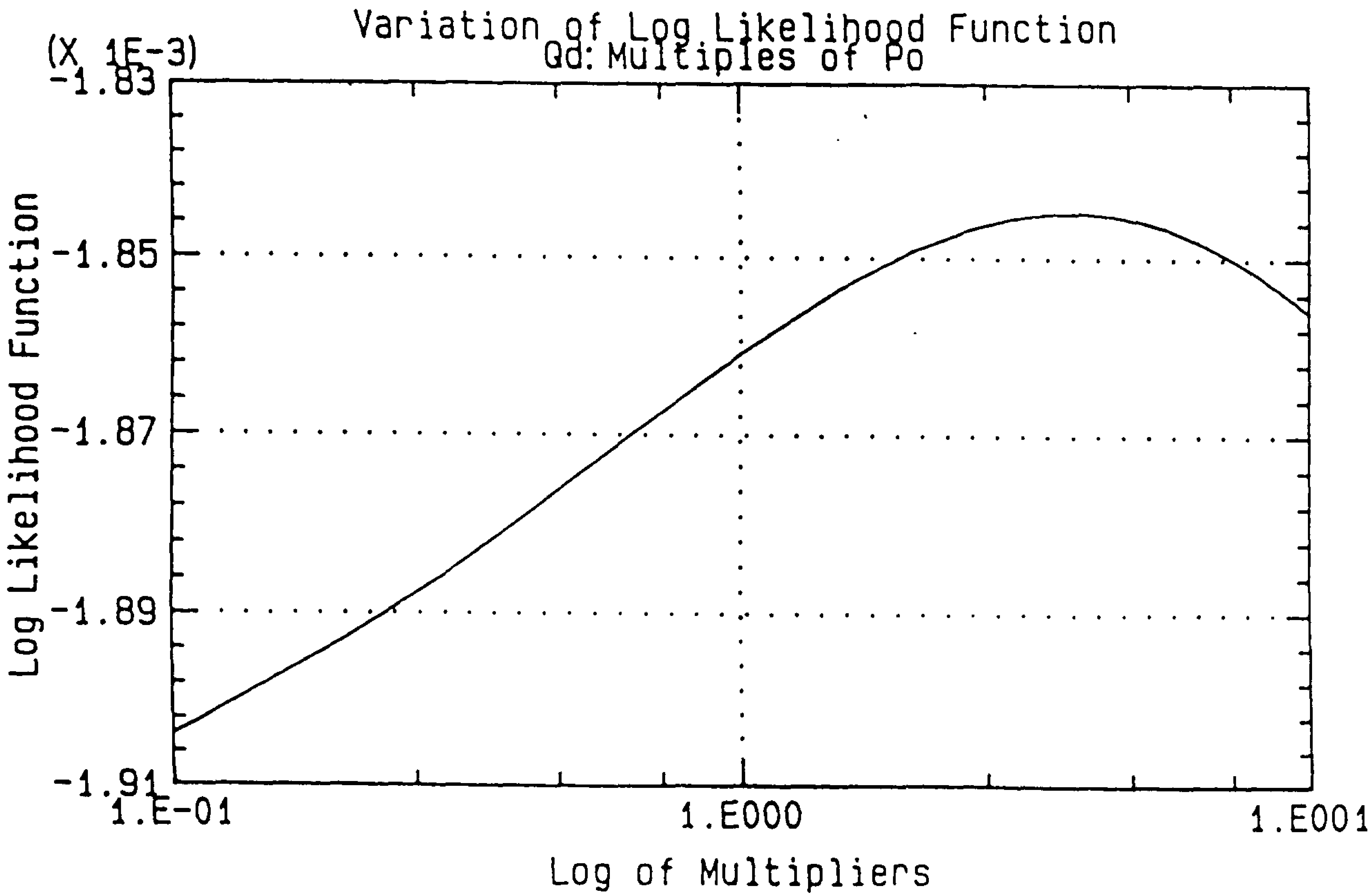


Figure (3-12) Variation of Log Likelihood Function with Q_d as Multiple of P_o

Table (3-10) Variation of Log Likelihood Function with Q_d as Multiple of P_n

Multiplier	10^{-3} (ln f)	Multiplier	10^{-3} (ln f)
0.0	-1.9471	5.5	-1.8455
0.1	-1.9040	6.0	-1.8474
0.2	-1.8932	6.5	-1.8484
0.3	-1.8859	7.0	-1.8494
0.4	-1.8802	7.5	-1.8504
0.5	-1.8755	8.0	-1.8515
0.6	-1.8716	8.5	-1.8527
0.7	-1.8683	9.0	-1.8538
0.8	-1.8655	9.5	-1.8549
0.9	-1.8631	10.0	-1.8560
1.0	-1.8609		
1.5	-1.8534		
2.0	-1.8492		
2.5	-1.8468		
3.0	-1.8455		
3.5	-1.8450		
4.0*	-1.8449*		
4.5	-1.8452		
5.0	-1.8458		

* denotes optimal value.

Table (3-11)

Comparison of Forecast Demand
Using Three Values of Q_d

HOUR	ACTUAL	FORECAST 1	FORECAST 2	FORECAST 3
661	893	891.2	892.0	892.6
662	943	936.9	937.3	937.6
663	966	980.7	980.6	980.6
664	1008	1004.1	1004.1	1004.0
665	993	994.4	994.7	994.9
666	936	960.0	960.8	961.4
667	891	928.8	929.8	930.5
668	956	925.2	925.9	926.4
669	956	949.2	949.4	949.5
670	962	975.7	975.4	975.1
671	962	976.2	975.5	974.9
672	928	942.5	941.5	940.7

Table (3-12)

Comparison of Percent Forecast Error
Using Three Values of Q_d

HOUR	ACTUAL DEMAND	% ERROR 1	% ERROR 2	% ERROR 3
661	893	0.201	0.111	0.045
662	943	0.647	0.604	0.573
663	966	-1.512	-1.512	-1.511
664	1008	0.387	0.387	0.397
665	993	-0.100	-0.171	-0.191
666	936	-2.564	-2.649	-2.714
667	891	-4.242	-4.355	-4.433
668	956	3.221	3.148	3.096
669	956	0.711	0.690	0.679
670	962	-1.424	-1.393	-1.362
671	962	-1.476	-1.403	-1.341
672	928	-1.562	-1.455	-1.368

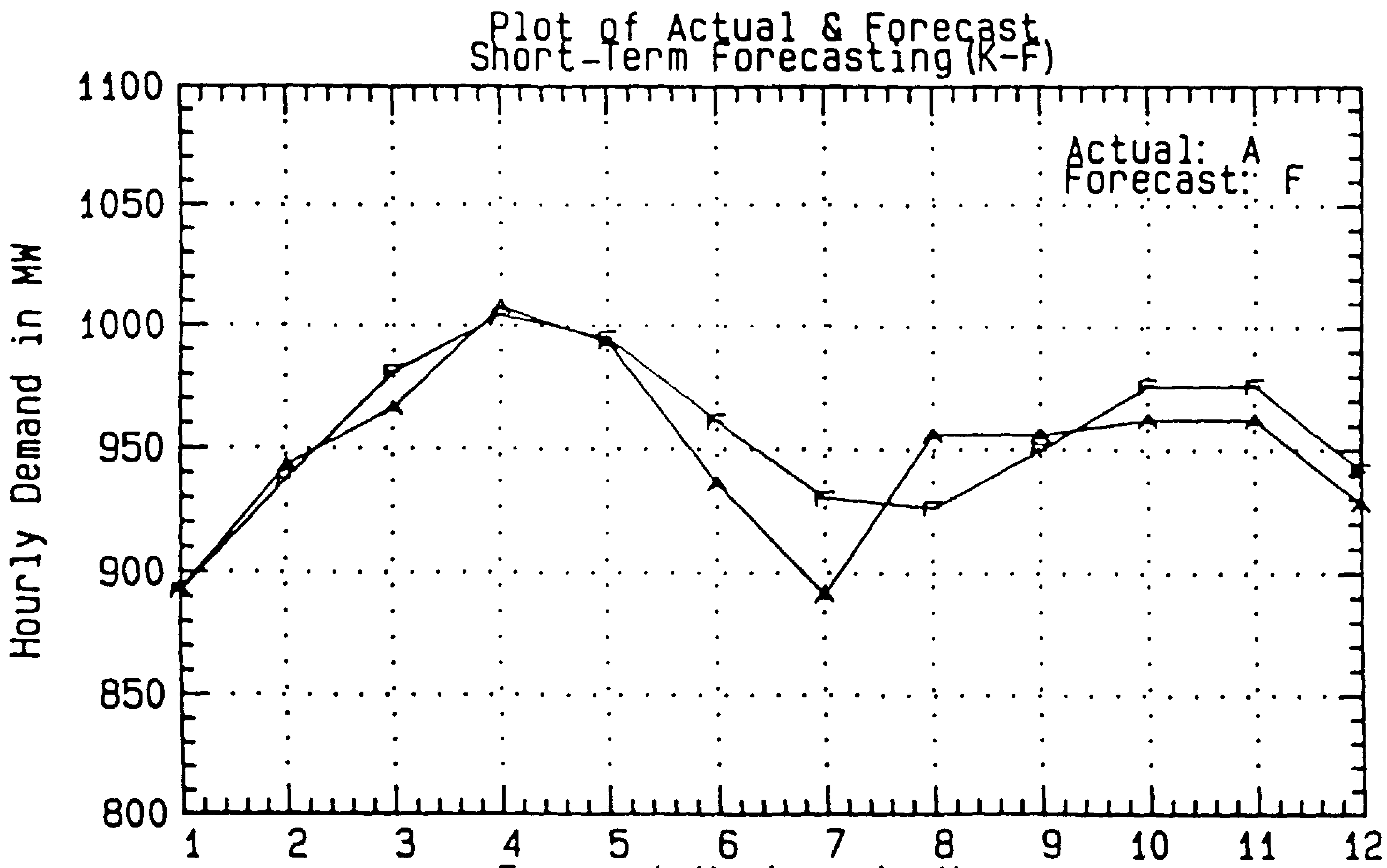


Figure (3-13) Comparison of Actual and Forecast Loads for Qd=4Po

Table (3-13)

Comparison of Fitting and Forecast Statistics
for Three Values of Q_d

	FITTING		FORECAST	
	MSE	MAPE	MSE	MAPE
$3P_n$	500.53	2.206	323.03	1.509
$4P_n$	496.33	2.200	323.42	1.490
$5P_n$	495.47	2.198	323.86	1.477

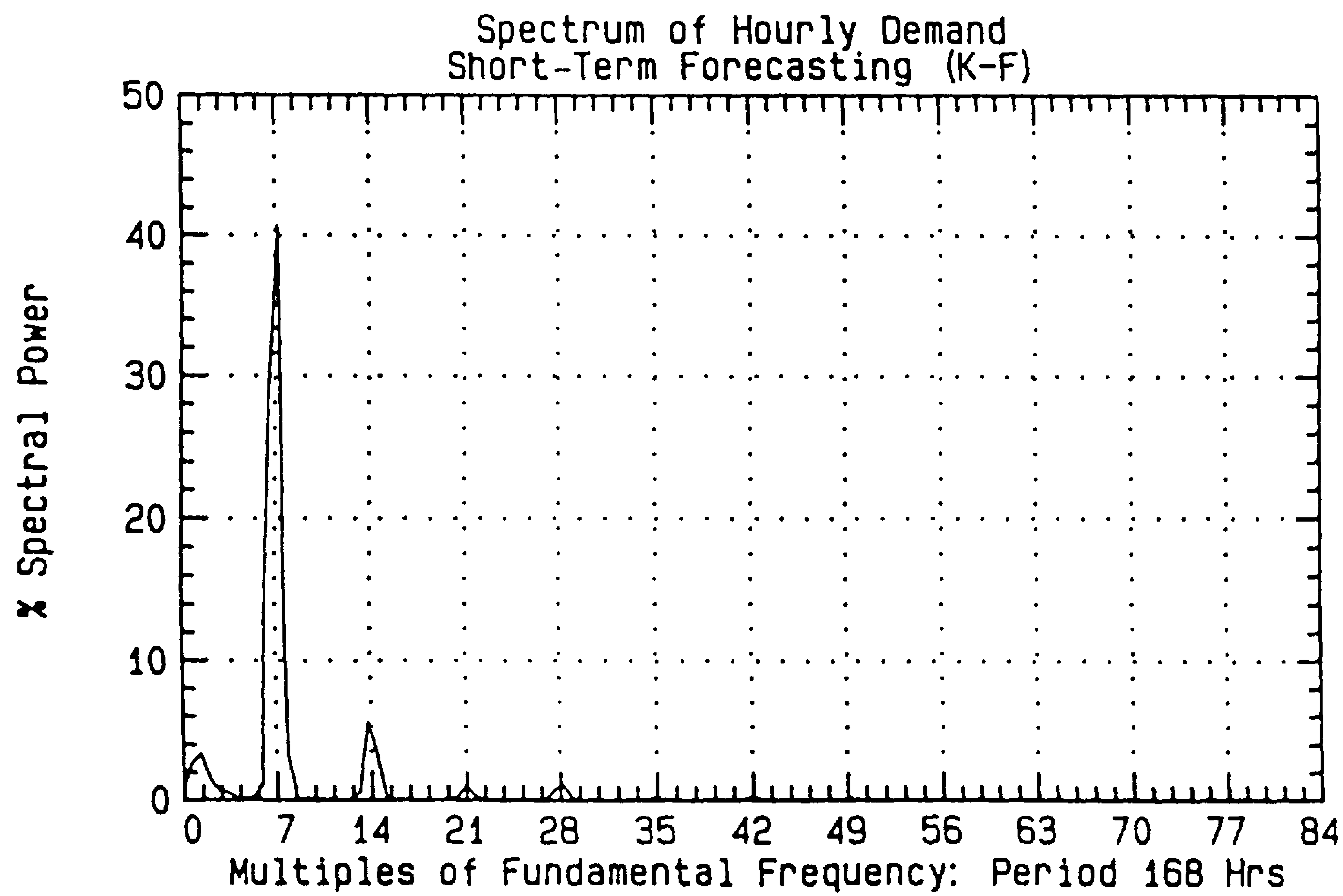


Figure (3-14) Spectrum of Hourly Data Qd=4Po

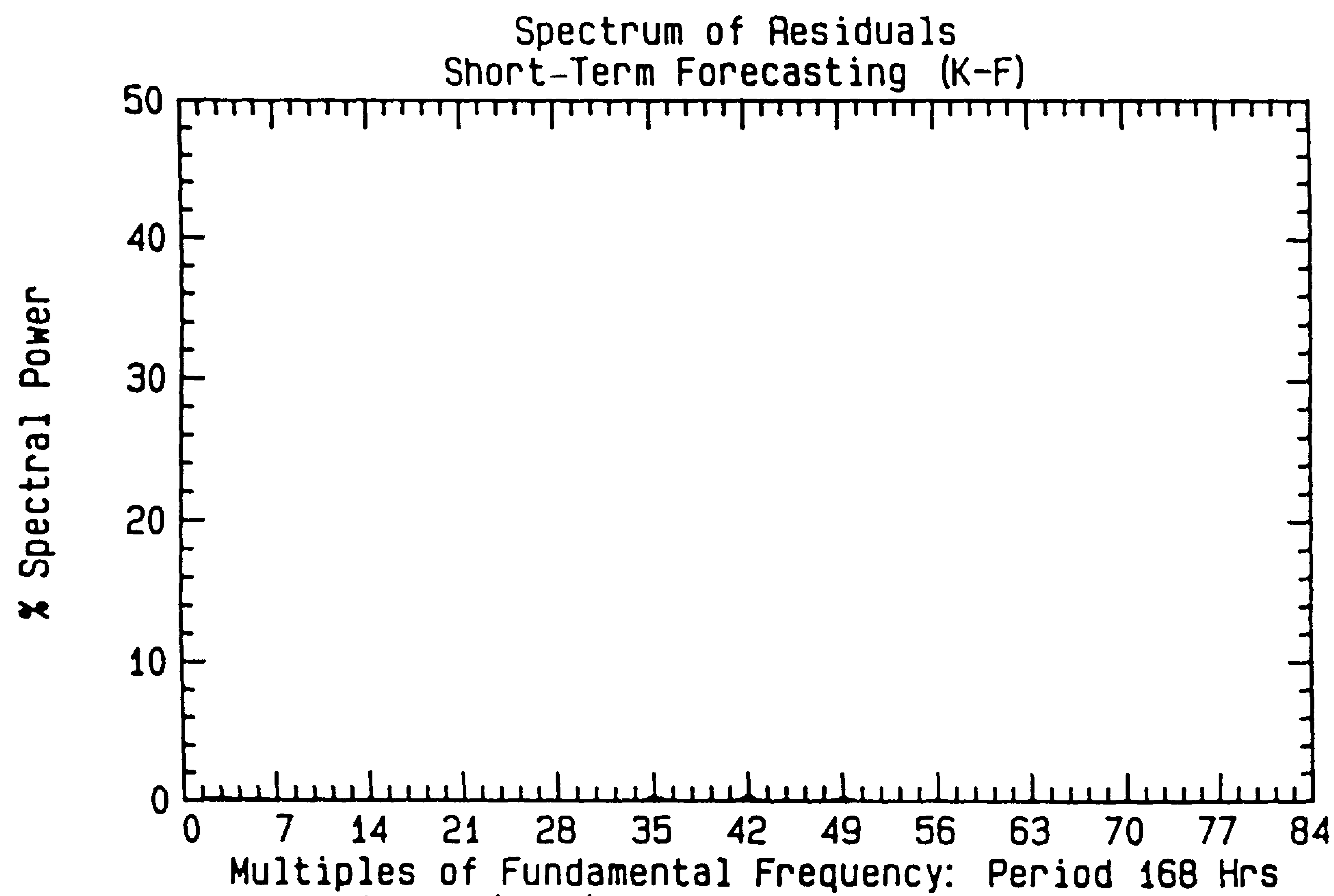


Figure (3-15) Validation of Kalman Filter
Short-Term Forecasting Model Qd=4Po

3.9 Recursive Weighted Least Squares Forecasting

An important class of estimation techniques employs the concept of weighted least squares in a recursive form that can be thought of as special cases of Kalman filtering. The class defines an optimality criterion of the form

$$J = \sum_{j=1}^{k} R^{k-j}(j)[y(j) - HX(j)]^2$$

The class is important in identifying time varying parameters, and is discussed extensively in the literature (see for example Young [103], Ljung [104], Ljung and Soderstrom [105], and Astrom and Wittenmark [106]).

The weighting function is called an exponential weighting into the past (EWP) function. A typical form is given by

$$R(j) = \lambda_0 R(j-1) + (1-\lambda_0)R$$

The function tends to a steady state R as $k \rightarrow \infty$, with R being a positive constant of magnitude less than or equal to one. We will assume that R = 1, and λ_0 = 0.99 in subsequent work following Ljung [104].

The recursive algorithm for estimation in a Kalman-like form is given by:

$$\hat{X}_-(k) = \Phi(k-1)\hat{X}_+(k-1)$$

$$r(k) = y(k) - H(k)\hat{X}_-(k)$$

$$P_-(k) = \Phi(k-1)P_+(k-1)\Phi^T(k-1) + G_d(k-1)Q_d(k-1)G_d^T(k-1)$$

$$K(k) = P_-(k)H^T(k)[R(k) + H(k)P_-(k)H^T(k)]^{-1}$$

$$\hat{X}_+(k) = \hat{X}_-(k) + K(k)r(k)$$

$$P_+(k) = \{[I-K(k)H(k)]P_-(k)[I-K(k)H^T(k)]^T + K(k)R(k)K^T(k)\}/R(k)$$

In the computational experiments reported in this section we used initially a model of the same form as that of Equation (3-25), but with 10 parameters only with the following angular frequencies:

$$\omega_1 = \frac{2\pi}{168}$$

$$\omega_2 = 7\omega_1$$

$$\omega_3 = 14\omega_1$$

$$\omega_4 = 21\omega_1$$

The algorithm was applied to the short term data set consisting of 660 points and the last twelve were reserved for forecasting purposes.

Computational Results

Results of applying the Recursive Least Squares algorithm with EWP are reported here. The value of λ_0 was chosen as 0.99, and a range of values of R(0) from 0.8 to 0.98 were considered. Table (3-14) summarizes the fitting and forecast statistics for each case. It is clear that better performance is offered by R(0)=0.8.

Forecasting results with R(0)=0.8, are presented in Table (3-15). Note that the maximum absolute percent error is 3.3289, which is a reasonable forecast error. Figure (3-16) shows the actual and forecast demands for the 12-hours-ahead forecasting along with the residuals. The steady state values of the model parameters are obtained as:

$\beta_0 = 873.8189$	$\beta_1 = -0.0667$
$\beta_{11} = 19.6525$	$\beta_{12} = 13.123$
$\beta_{21} = -119.1135$	$\beta_{22} = 33.9418$
$\beta_{31} = 34.3972$	$\beta_{32} = 23.9104$
$\beta_{41} = -13.4597$	$\beta_{42} = 12.2493$

Table (3-14)
Comparison of Fitting and Forecast Statistics for Five Values of R(o)

	FITTING STATISTICS		FORECAST STATISTICS	
R(o)	MSE	MAPE	MSE	MAPE
0.8	1105	3.1244	1243	3.3289
0.85	1119	3.1494	1231	3.3054
0.9	1142	3.1719	1230	3.2955
0.95	1182	3.2221	1289	3.3666
0.98	1228	3.2979	1429	3.5369

Table (3-15)
Actual, Forecast and Errors Using R(o)=0.8 for 10 Parameter Model Realization

HOUR	ACTUAL	FORECAST	ERROR	% ERROR
661	893	870.7	22.3	2.5
662	943	920.1	22.9	2.4
663	966	948.3	17.7	1.8
664	1008	952.6	55.4	5.5
665	993	939.0	54.0	5.4
666	936	919.6	16.4	1.8
667	891	906.3	-15.3	-1.7
668	956	905.5	50.5	5.3
669	956	914.7	41.3	4.3
670	962	925.3	36.7	3.8
671	962	926.8	35.2	3.7
672	928	912.2	15.8	1.7

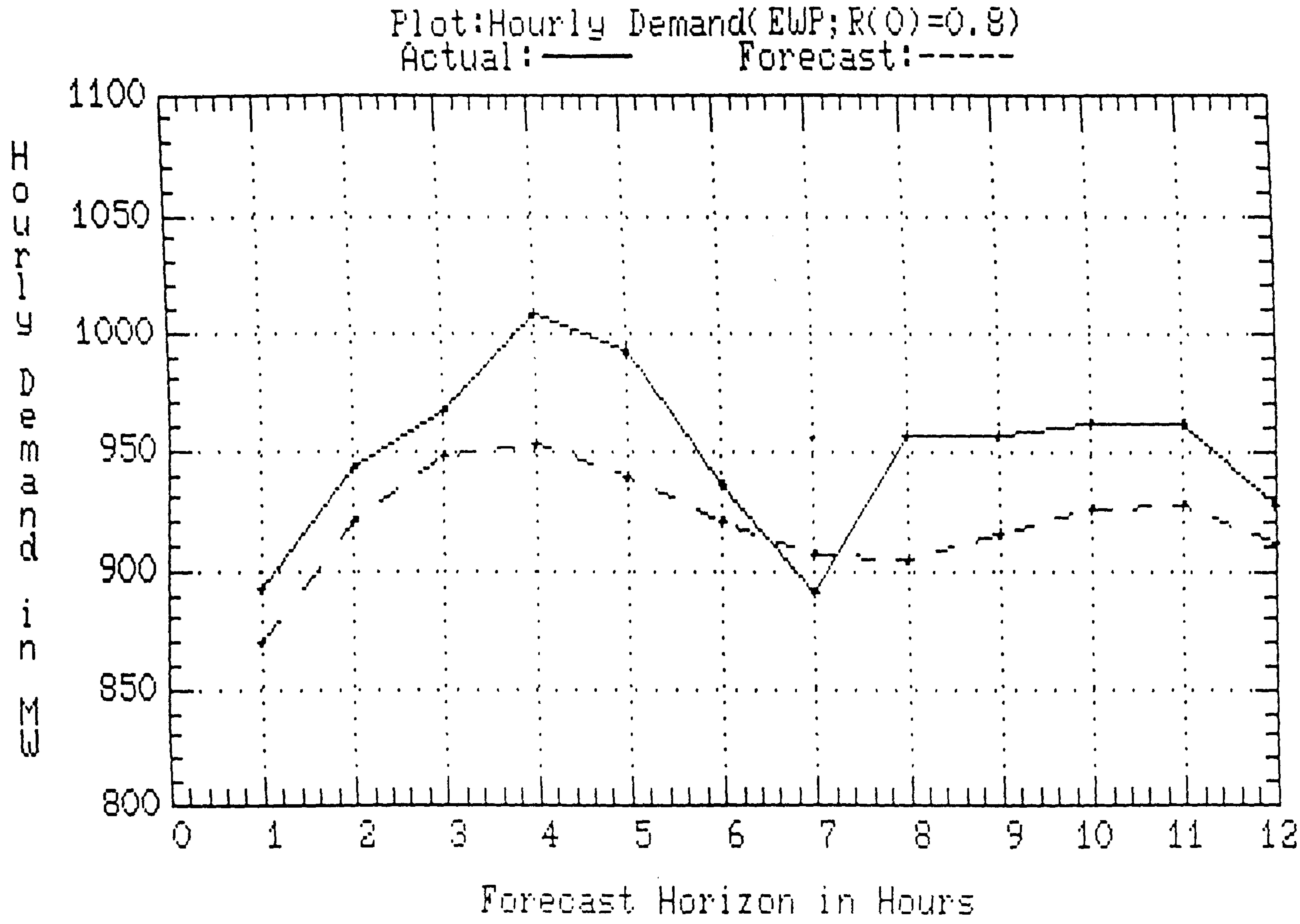


Figure (3-16) Forecast and Demand Using EWP with R(o) = 0.8

An improvement in the forecasts was obtained by adopting a twelve parameter model with R(0)=0.8, R=1 and λ_n retained at 0.99. The optimum parameter values were found to be:

β_0 = 874.8837 $\quad\beta_1$ = -0.0682 $\quad\beta_{11}$ = 19.4362

β_{12} = 13.5683 $\quad\beta_{21}$ = -119.667 $\quad\beta_{22}$ = 33.6816

β_{31} = 34.9120 $\quad\beta_{32}$ = 23.6402 $\quad\beta_{41}$ = -13.7688

β_{42} = 12.8282 $\quad\beta_{51}$ = -20.474 $\quad\beta_{52}$ = -0.3936

The following statistics were obtained

Fitting Statistics

MSE = 831, MAPE = 2.7334

Forecast Statistics

MSE = 1,098, MAPE = 2.9853

Table (3-16) lists the forecast values and errors involved, and Figure (3-17) shows a comparison of actual and forecast demands for 12 hours.

Table (3-16) Actual, Forecast and Errors
Using R(o)=0.8 for 12 Parameter Model

HOUR	ACTUAL	FORECAST	ERROR	% ERROR
661	893	853.5	39.5	4.4
662	943	904.2	38.8	4.1
663	966	950.9	15.1	1.6
664	1008	972.8	35.2	3.5
665	993	958.3	34.7	3.5
666	936	920.2	15.8	1.7
667	891	888.7	2.3	0.3
668	956	887.9	68.1	7.1
669	956	915.1	40.9	4.3
670	962	943.5	18.5	1.9
671	962	944.8	17.2	1.8
672	928	912.4	15.6	1.7

Inspecting the percent forecast errors associated with each point given in Table (3-16) shows that the maximum error is lower in absolute value than 8%. This technique did not perform well in terms of the MSE and MAPE in comparison with other methods. It must be noted, however, that this method involves less computational effort than other methods tested in this work.

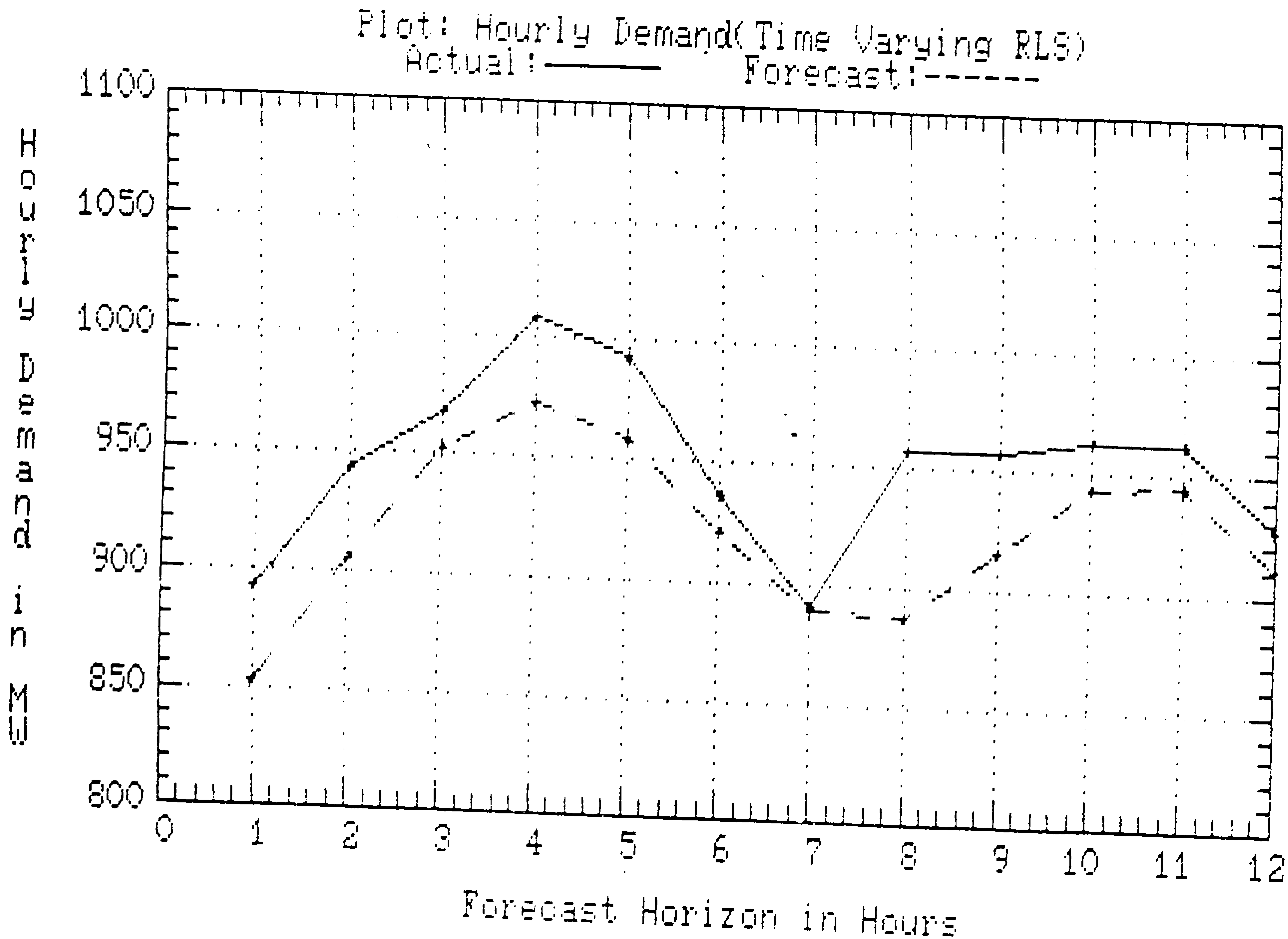


Figure (3-17) Forecast and Demand
for 12 Parameter Model

3.10 Instrumental Variable Approach

The Instrumental Variable Estimation technique is an excellent tool for time-series analysis and forecasting [103 - 106]. The theoretical foundations of the technique are described in Appendix G. The application of the method relies on the fact that there is a definite relationship between the hourly power demand on the Abu Dhabi system and the hourly temperatures. In this data, there is a definite 24 hours cyclic trend with fluctuations around an almost stationary mean. As a result, we carried out a pre-processing step of removing the respective means from the data sets.

From a preliminary investigation, it was concluded that model orders of N(B) and D(B) of up to 5 and 10 are appropriate. A search to define the optimum model order was conducted. The process model is given by

$$D(B)\,\tilde{z}(k) = N(B)\,\tilde{u}(k) + \upsilon(k)$$

$$L(B)\,\upsilon(k) = e(k)$$

with $\tilde{z}(k) = z(k) - z_m$

$$\tilde{u}(k) = u(k) - u_m$$

Here z_m and u_m denote the mean of load and temperature over an interval of 660 hours. The optimum model orders are found by trial and error to be

D(B) of order 2

N(B) of order 2

Prefilter L(B) order of 32

$$D(B) = 1 - 1.6228\,B + 0.69403\,B^2$$

$$N(B) = 1.0161\,B - 1.0256\,B^2$$

The choice of the L(B) order was based on the Akaike Information Criterion (AIC) values as listed in Table (3-17). The model of order 32 involves a minimum AIC. The parameters of L(B) are listed in Table (3-18) and were obtained by a least squares estimation. The forecast values are listed in Table (3-19) and plotted in Figure (3-18).

The following is a summary of the forecast statistics.

Fitting MSE = 275

Forecasting MSE = 224

Inspecting the forecast percent error associated with each load reveals that all absolute values are less than 3%. The forecasts of points 662, 663 are better than any obtained by other methods for short term forecasting. In terms of MSE, the results of the Instrumental Variable forecasting rank ahead of Winters' additive, Box-Jenkins model with periodicity of 168 hours, Kalman filtering, and Recursive Weighted Least Squares. The process of obtaining the forecast appears to involve more computations than many other methods. Most of the computational effort is spent on determining the prefilter's order.

Table (3-17)
Variation of AIC
with L(B) Model Order

MODEL ORDER	AIC	MODEL ORDER	AIC
20	1533.5	20	565.53
21	1532.9	29	562.54
22	1491.0	30	557.67
23	1486.1	31	552.56
24	700.52	32*	550.11*
25	594.62	33	551.13
26	572.25	34	551.49
27	570.73	35	552.35

Table (3-18)
Parameters of the AR(32) L Filter

INDEX i	COEFFICIENT (l_i)	INDEX i	COEFFICIENT (l_i)
1	0.4898	17	0.022528
2	0.27481	18	0.041865
3	0.064388	19	0.02218
4	-0.13496	20	0.049199
5	-0.13334	21	0.033055
6	-0.15091	22	0.044463
7	-0.13039	23	-0.058275
8	-0.040368	24	-0.7941
9	0.049585	25	-0.46953
10	0.057595	26	-0.2268
11	-0.0014735	27	-0.043121
12	0.030166	28	0.16544
13	0.028168	29	0.16106
14	0.016	30	0.18686
15	0.050917	31	0.13983
16	0.026413	32	0.066378

Table (3-19)
Actual, Forecast Demand and
Error Using Instrumental Variable

HOUR	ACTUAL	FORECAST	ERROR	%ERROR
661	893	900.5	-7.5	-0.84
662	943	940.7	2.3	0.24
663	966	973.5	-7.5	-0.78
664	1008	1026.2	-18.2	-1.81
665	993	1020.9	-27.9	-2.81
666	936	956.5	-20.5	-2.19
667	891	914.8	-22.8	-2.56
668	956	951.8	4.2	0.44
669	956	951.2	4.8	0.50
670	962	947.3	14.7	1.53
671	962	945.7	16.3	1.69
672	928	915.1	12.9	1.39

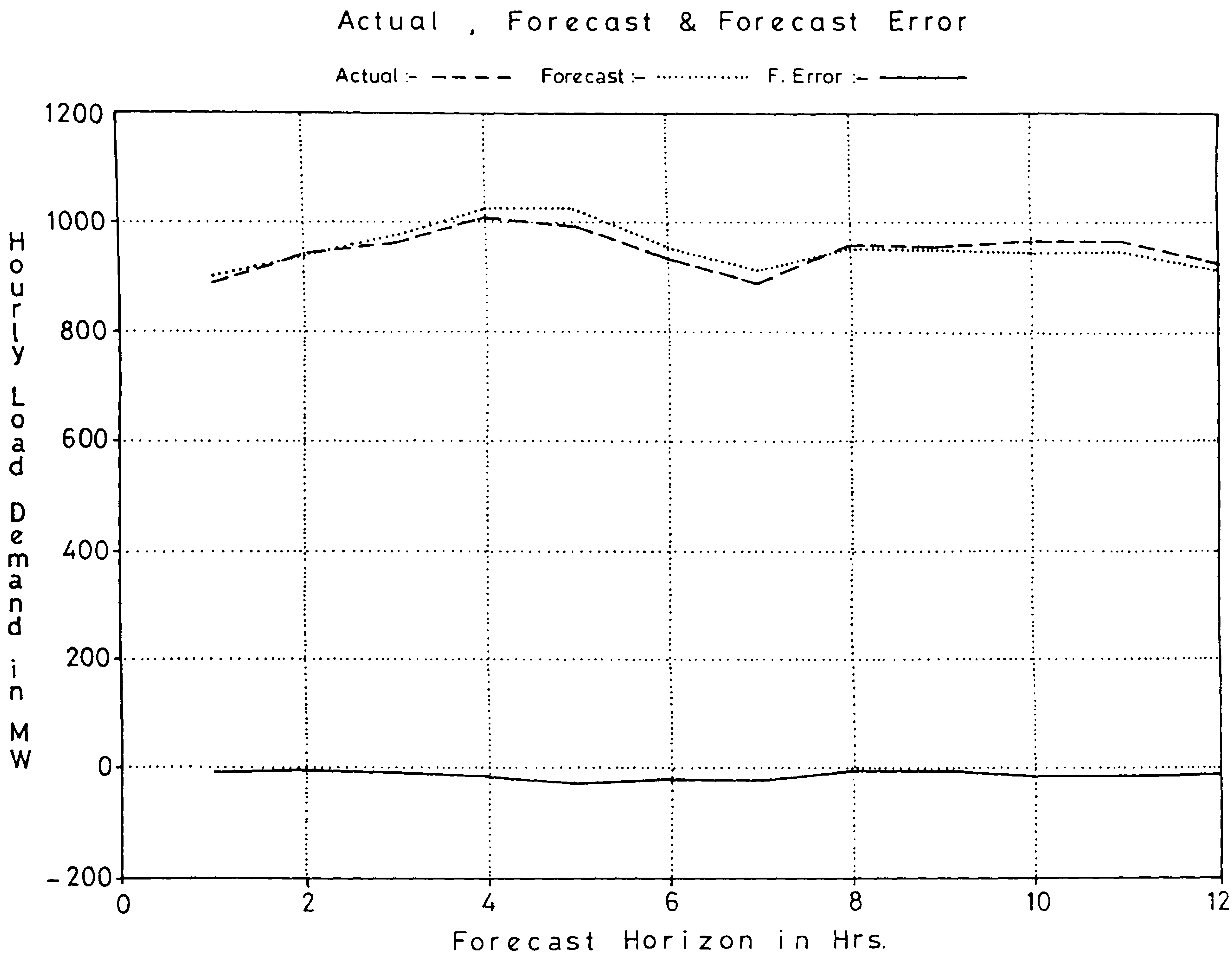


Figure (3-18) Forecasting Results for the Short-Term Problem Using the Instrumental Variable Method

3.11 A Comparative Evaluation of Approaches

In the present section comparisons between the various forecasting techniques are given for both the short-term and mid-term problems. The comparisons are based on individual point of forecasts, fitting, and forecast statistics.

Short-Term Forecasts

Considering the short-term forecasting problem first, Table (3-20) summarizes the fitting and forecast statistics for the techniques explored. Figures (3-19) and (3-20) give the same information graphically. The following rank order of the techniques is given on the basis of Table (3-20) for fitting and forecast criterion.

Fitting MSE

1- Winters' Multiplicative Model (WMM)
2- Box-Jenkins - 2, model with 168 hours seasonality (BJ2)
3- Box-Jenkins -1, model with 24 hours seasonality (BJ1)
4- Instrumental Variable (IV)
5- Kalman Filtering (KF)
6- Generalized Exponential Smoothing (GES)
7- Winters' Additive Model (WAM)
8- Recursive Weighted Least Squares (RWLS).

Fitting MAPE

1- Box-Jenkins 2
2- Winters' Multiplicative Model
3- Box-Jenkins - 1
4- Instrumental Variable
5- Winters' Additive Model
6- Generalized Exponential Smoothing
7- Kalman Filtering
8- Recursive Weighted Least Squares.

Forecast MSE

1- Box Jenkins - 1
2- Winters' Multiplicative Model
3- Generalized Exponential Smoothing
4- Instrumental Variable
5- Box-Jenkins - 2
6- Kalman Filtering
7- Winters' Additive Model
8- Recursive Weighted Least Squares.

Forecast MAPE

1- Box-Jenkins - 1
2- Winters' Multiplicative Model
3- Generalized Exponential Smoothing
4- Instrumental Variable
5- Kalman Filtering
6- Box-Jenkins - 2
7- Winters' Additive Model
8- Recursive Weighted Least Squares.

We observe from the ranking that models that fit the data better do not necessarily provide the best forecast. The Box-Jenkins model 1 is the best in terms of forecast MSE and MAPE, yet it is consistently placed third in the fitting statistics. This may not be totally true for Winters' Multiplicative model which is placed second consistently in the forecast error category but was either first or second in the fitting error category. The Generalized Exponential Smoothing approach is consistently the third in the forecast errors, yet it is consistently sixth in the fitting category. Note that the Instrumental Variable method is consistently placed fourth in all categories. The Recursive Weighted Least Squares method consistently came in last, indicating that improvements in the weighting mechanism must be sought.

Table (3-20)
Compararison of Performance of
Forecasting Techniques for Short-Term Problem

	FITTING STATISTICS		FORECAST STATISTICS	
	MSE	MAPE	MSE	MAPE
GES	555.97	2.17	195	1.097
WA	640	1.561	401.5	1.575
WM	145.5	1.07	100.75	0.966
B-J-1	245.87	1.236	88.26	0.896
B-J-2	171.87	1.053	289.26	1.53
KF	495.7	2.198	323.86	1.477
RWLS	831	2.73	1098	2.985
IV	275.65	1.435	241.75	1.4075

One can also extract useful information about the performance of each method if recourse is made to the nature of the error at each point to be forecast in the time series. To do this, we recall that the forecast error is the difference between the actual load and the forecast value. A negative forecast error means a higher forecast than actual, or over-forecasting in the forecasters' terminology. This situation is called "optimistic" in the power systems planning and operation, since one is on the safe side when using the results of an overforecast. The converse terminology would be "underforecasting, pessimistic", for a positive forecasting error. Based on these definitions, we can now state a "score" for each method based on inspection of Table (3-21) summarizing the forecast values over the 12 hour forecast horizon.

Short-Term Forecasting
MSE vs Forecasting Methods

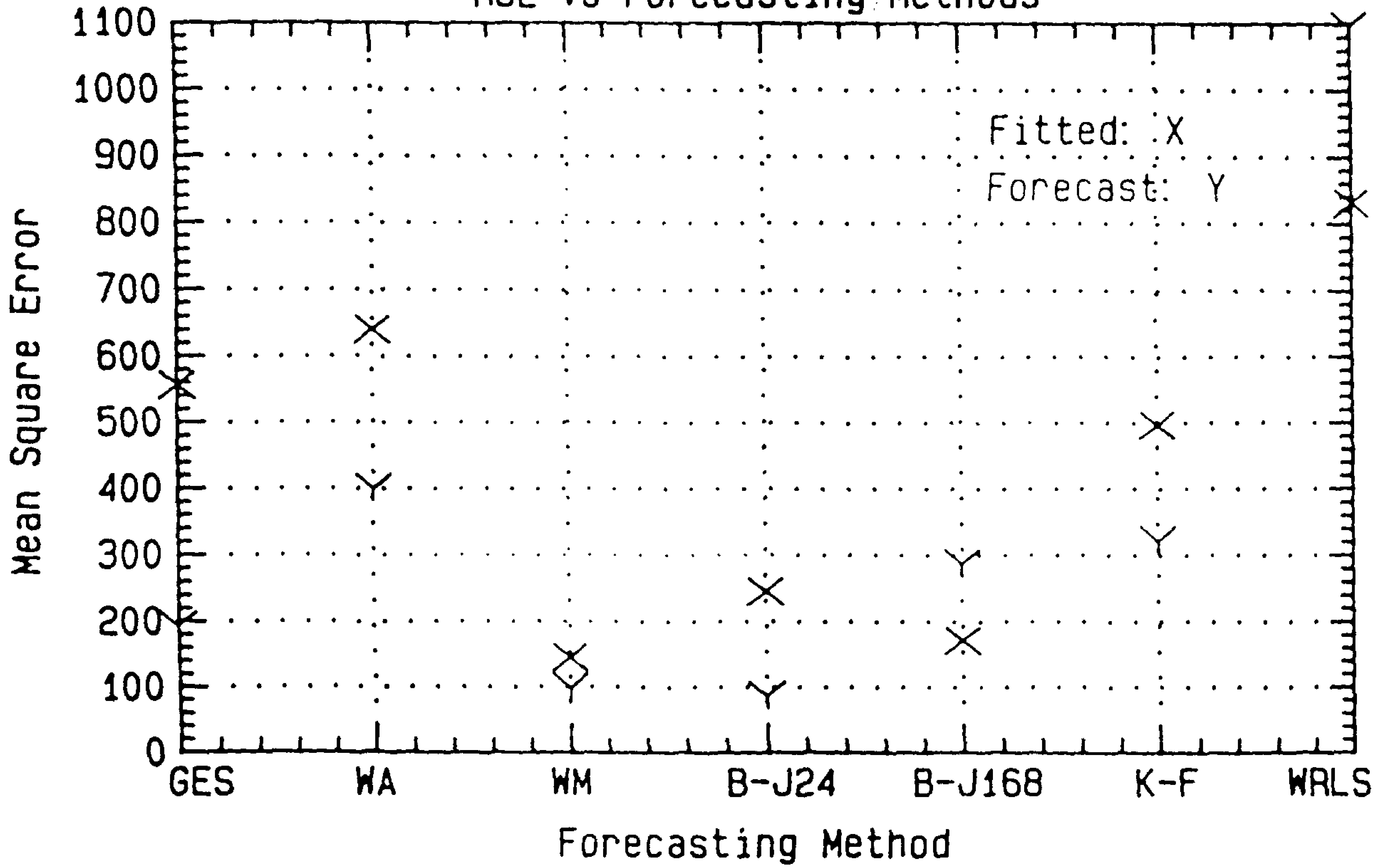


Figure: 3-19

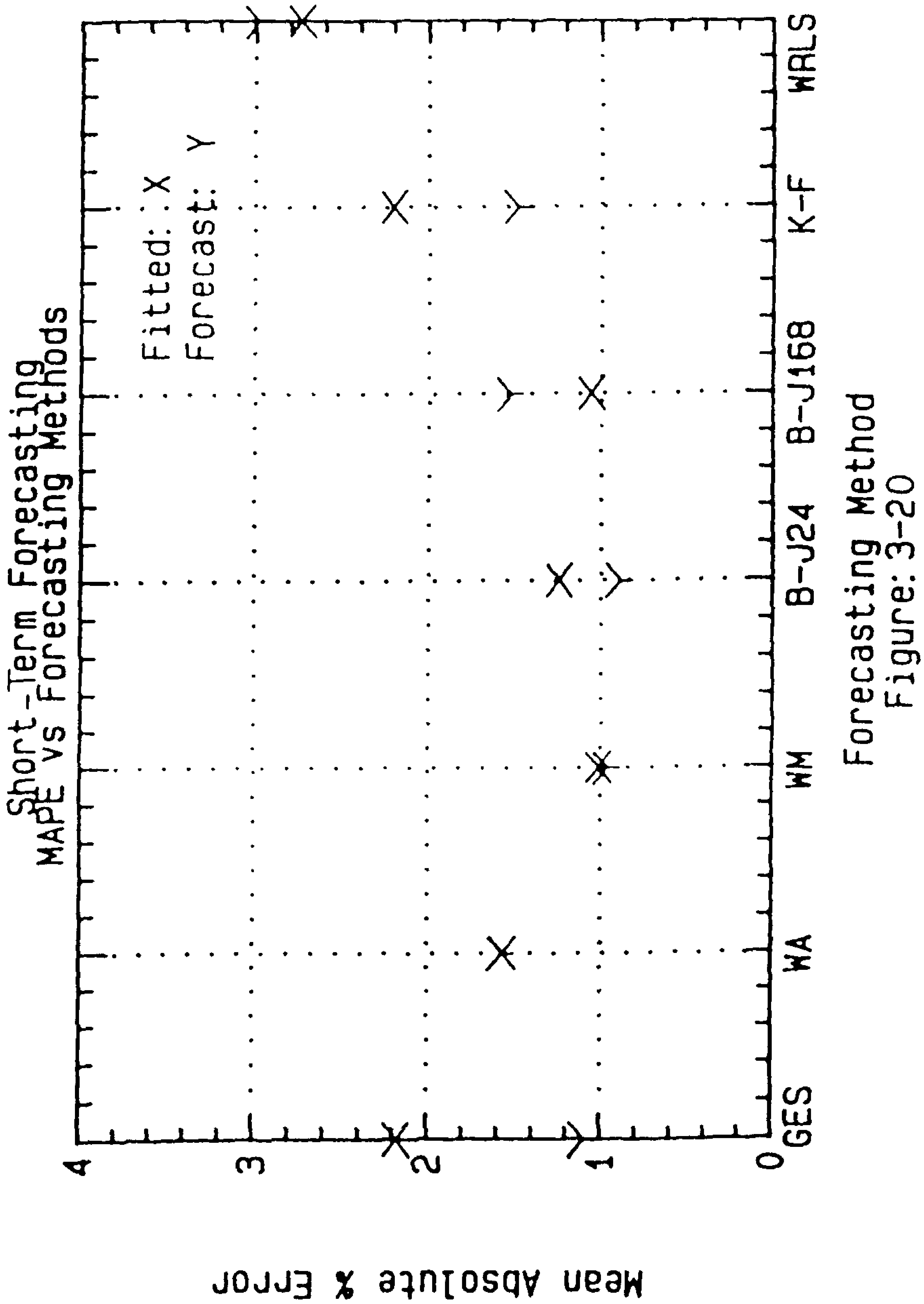

Short-Term Forecasting
MAPE vs Forecasting Methods
Fitted: X
Forecast: Y
0
1
2
3
4
GES
WA
WM
B-J24
B-J168
K-F
WRLS
Mean Absolute % Error
Forecasting Method

Figure: 3-20

1- Box-Jenkins, with 24 hour seasonality has 7 pessimistic forecasts and 5 optimistic ones.

2- Winters' Multiplicative Model forecasting has 5 pessimistic and 7 optimistic forecasts.

3- Generalized Exponential Smoothing has an equal division of 6 forecasts on each side.

4- The Instrumental Variable method has an equal division of 6 forecasts on each side.

5- Kalman filtering has 5 pessimistic and 7 optimistic forecasts.

6- Box Jenkins, with 168 hours seasonality has 10 pessimistic and 2 optimistic forecasts.

7- Winters' additive model has 4 pessimistic and 8 optimistic forecasts.

8- Recursive Weighted Least Squares has 12 pessimistic forecasts.

It is clear from the preceding discussion that a good spread of over and under-forecasts is an indication of the merit of the method. This information sometimes is masked by the MSE and MAPE which eliminate the sign of the error in their computation.

Table (3-21)
Comparison of Forecast Values for
the Short-Term Problem

METHOD	FORECAST HORIZON												STATISTICS			
													FITTED		FORECAST	
	1	2	3	4	5	6	7	8	9	10	11	12	MSE	MAPE	MSE	MAPE
Generalised Exponential Smoothing	892.2	938.9	938.0	1006.0	995.1	959.7	927.8	923.7	947.3	973.2	972.8	938.4	555.97	2.17	195.00	1.097
Winters' Additive	886.1	945.4	922.5	1023.3	1019.9	967.4	887.8	954.6	966.1	964.7	988.0	939.4	640.00	1.56	401.50	1.575
Winters' Multiplicative	900.7	952.2	972.9	1013.5	1002.8	952.5	877.5	941.0	953.8	955.7	957.5	940.6	145.50	1.007	100.75	0.966
Box-Jenkins S = 24 Hours	894.0	934.0	978.0	1005.0	1002.0	940.0	881.0	940.0	949.0	953.0	957.0	937.0	245.88	1.236	88.26	0.896
Box-Jenkins S = 168 Hours	885.0	912.0	968.0	997.0	992.0	942.0	864.0	931.0	944.0	946.0	945.0	912.0	171.80	1.053	289.26	1.530
Kalman Filter	892.2	937.3	980.6	1005.1	994.7	960.8	929.8	925.9	949.4	975.4	975.5	941.5	496.33	2.200	323.42	1.490
Recursive EWP Least Square	853.5	904.2	950.9	972.8	958.3	920.2	888.7	887.9	915.1	943.5	944.8	912.4	831.00	2.733	1098.00	2.985
Instrumental Variables	900.5	940.7	973.5	1026.2	1020.9	956.5	914.8	951.8	951.2	947.3	945.7	915.1	275	1.435	241.75	1.408
ACTUAL DEMAND	893	943	966	1008	933	936	930	956	956	962	962	928				

Mid-Term Forecast

Considering now the mid-term forecasting problem, we have Table (3-22) showing a comparison of fitting and forecasting statistics for the four methods considered.

Table (3-22)
Comparison of Performance of
Forecasting Techniques for Mid-Term Problem

TECHNIQUE	FITTING STATISTICS		FORECAST STATISTICS	
	MSE	MAPE	MSE	MAPE
GES	1995	8.26	2924	5.312
WA	2092	9.206	2545	4.731
WM	1422	5.507	1843	4.498
B-J	1384	5.24	2551	5.3

The rank ordering of the techniques is given as follows.

Fitting MSE & MAPE

1- Box-Jenkins
2- Winters' Multiplicative Model
3- General Exponential Smoothing
4- Winters' Additive Model.

Note that identical rankings are obtained from both MSE & MAPE

Forecasting MSE & MAPE

1- Winters' Multiplicative Model
2- Winters' Additive Model
3- Box-Jenkins
4- General Exponential Smoothing.

Note again ranking is the same for MSE & MAPE.

One observes once again that the best fitting model does not give the best forecast.

In terms of the forecast score for each method, one can conclude from the results given in this chapter that:

1- Winters' Multiplicative model has an even split of 6 pessimistic and six optimistic forecasts.
2- Winters' Additive model has 7 pessimistic and 5 optimistic forecasts.
3- Box-Jenkins has 4 pessimistic and 8 optimistic forecasts.
4- General Exponential Smoothing has 7 pessimistic and 5 optimistic forecasts.

Again we note that the more successful methods show a good spread of forecast errors.

The errors ecountered in the mid-term problem are larger than those for the short-term problem, since we have more data points in the latter case.

The Choice of a Preferred Technique

In the short-term problem, it is tempting to state that any of the top four methods would give satisfactory results. It is important to realize that for practical application the complexity and effort involved in a technique are important considerations.

Here we note that Box-Jenkins, Kalman filtering and the Instrumental Variable method involve more effort and decision making than those required for Winters' and the Generalized Exponential Smoothing.

3.12 Summary

This Chapter was devoted to comprehensive review and a Comparative Analysis of performance of major load forecasting techniques, both conventional and modern, with emphasis on Abu Dhabi's short-term and mid-term forecasting tasks. Among the unique contributions of this Chapter are:

- New Kalman-filtering based model of the forecasting function.
- An investigation into the optimal choice of covariance matrices for Kalman's forecasting.
- The application of the Instrumental Variable method to the load forecasting function.

The work reported in this chapter is hoped to help provide more insight into the practical and algorithmic considerations in dealing with Abu Dhabi's load forecasting problems.

CHAPTER 4
FUEL COST MODEL PARAMETER ESTIMATION

4.1 Introduction

The intent of this chapter is to offer a discussion of the important topic of parameter estimation of thermal generation production cost models. As mentioned in Section 2.6, thermal cost modelling is the second ingredient of the development of an economic dispatch strategy. Needless to say, accurate representations are the corner-stone of successful implementation.

The chapter begins with a background review of this important area fuel cost model parameter estimation, and is followed by a discussion of the parameter estimation problem formulation. The well-established approaches of the weighted least squares parameter estimation, recursive parameter estimation, are discussed as the basis for computational work reported in this chapter.

As a result of experimentation with data for our utility system, it was realized that in a number of instances, the estimated parameters were unsatisfactory from an end use point of view. One of the parameters of the quadratic model is required to be positive to guarantee a minimum in subsequent optimization work. The author was led to investigate approaches that might meet this requirement. The requirement of a positive coefficient is rightfully called a constrained parameter estimation problem and, therefore, the author considered investigating a nonlinear programming approach to the problem. This is believed be a new application. To conclude this chapter a summary of computational experience for one set of data in the 56-unit system is given, comparing the performance of all models discussed in this chapter.

4.2 Background

The object of economic dispatch is to determine the most economic pattern of load sharing amongst available units in an electric power system to meet a specified system power demand. The goal of such a study is to minimize the total cost of production to satisfy the given demand. In a modern fossil fuelled plant, the cost of fuel is the most important component in determining the total operating cost. This fuel related component accounts for up to 85% of the total cost of electricity production [107].

From an economic operation point of view, the fuel consumption of a thermal unit is a function of its active power output. This function is generally referred to as the input/output curve, with input (F) being measured in terms of hourly thermal energy consumption in (Mcal/hr) and output in net unit's active power send out (P) in MW. A typical input/output curve is shown in Figure (4-1).

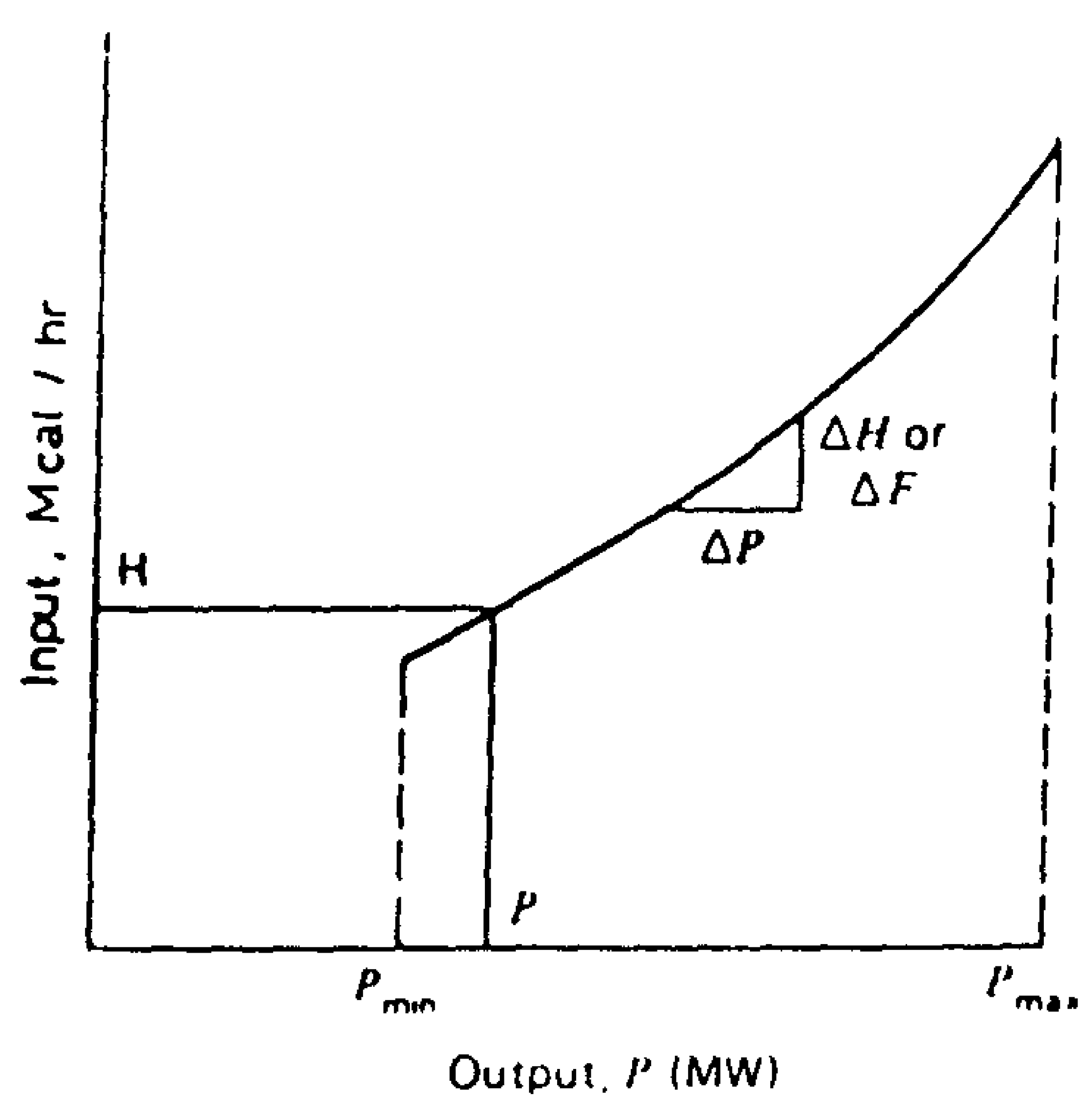


Figure (4-1) Input_Output Curve of a Steam Turbine Generator

Following commissioning of a power plant, acceptance tests are conducted to verify that each unit meets minimum guaranteed design specifications. These tests require high precision instrumentation and are conducted by specialists. Test results yield a set of data relating input (Mcal/hr) to output energy [MWhr]. These form the basis of an input/output curve referred to as manufacturer's guaranteed characteristics.

Unit performance characteristics deteriorate with time due to a number of factors that are classified as controllable and uncontrollable. Uncontrollable factors are those over which the utility has no control. It is not possible for the operator to act to improve plant performance degradation due to fuel type and quality, and set load points required by the grid demand. Controllable factors are those that can be altered to reverse deterioration. These in turn can be classified as short, medium and long term factors. Short term factors are ones that are under continuing operator control such as control settings. Medium term variables are influenced by decisions made on an hourly or daily basis such as soot blowing. Long term variables are influenced by decisions made on a monthly or yearly basis such as chemical cleaning of boilers.

In the long term, degradation of the performance characteristic can be attributed to individual components of a unit as follows [107].

1. **Boilers**

Extent of deposit accumulation on heat absorbing surfaces in the boiler affects the gas stack temperature. A rise of 10°C in fuel gas temperature is known to increase fuel consumption by about 0.5%

2. **Turbo-Generator**

Deposits on turbine blades, clearance of blade tips, roughness of blade surface and gland seals all affect the stage efficiency of a turbine. A roughness of turbine blade surface equivalent to a fine sand paper grade can cause a loss in stage efficiency of about 3%.

3. **Condenser**

Variations in the condenser's vacuum influences the heat consumption of a unit. A drop of 10 mbar of the condenser's vacuum increases the fuel consumption by up to 1%.

Developing accurate models to represent the thermal cost of generation is important for economic operation studies. In view of the preceding discussion, it is evident that maintaining up-to-date model parameters is of equal importance since the unit's performance continually changes due to the factors discussed earlier.

The subject of thermal cost model parameter estimation is part of the broader area of model parameter estimation in electric power systems security-economy functions. The evaluation of system model parameters is of continuing interest to the electric power industry. Most of the available literature, however, assumes the availability of such parameters. It is noted however that Sasson [108] reports on the use of the least squares error estimation approach as a viable option to obtain loss formula coefficients for economic dispatch purposes.

As a general rule, the evaluation of model parameters in electric power system studies has been conducted using physically- based arguments and therefore it has not been necessary to approach the problem on a system-theoretic basis. It was not until the late sixties that power systems engineers realized the limitations of a completely physically-based approach to the solution of important economy-security functions.

The introduction of the concept of a "Static State Estimator" [109] to evaluate the best estimate of the network state variables in terms of voltage profiles and power injections heralded the use of system theoretic estimation techniques in the electric power industry. A static state estimator is now an integral software tool that is in use in many modern energy control centres.

It has been shown [110] that parameter inaccuracies can severely limit the on-line state estimation function's accuracy in comparison with the actual measured value. As a consequence, Debs [110] presented a recursive filtering approach for on-line network parameter estimation.

A number of researchers have established the sensitivity of economic operation strategies to modelling errors as far back as Ringlee's Study [111]. More recently, Vemuri and Hill [112] and Dillon and Tan [113], demonstrate the potential loss in economy due to implementing optimal strategies with deviations from the assumed operating conditions. In [4] El-Hawary and Christensen highlight the problems caused by using erroneous model parameters, even for the smallest size system. In two related papers El-Hawary and Mansour [114 and 115] consider some basic models commonly used in conjunction with optimal economic operation studies.

The techniques of weighted least squares, Gauss-Newton, Powell's regression and Levenburgh- Marquardt for estimating model parameters were evaluated for determining the parameters of fuel cost models, active power loss and active reactive power loss models. It is concluded in the study that the weighted least squares approach is preferable. It is noted that El-Hawary and Mansour employed conventional least squares formulations requiring the solution of the normal equations in batch processing.

In [116] El-Hawary and Kumar apply a recursive least squares technique that uses one additional measurement in building the optimal estimates recursively. These studies were conducted using data contained in a report on common thermal unit types and their performance [117].

4.3 Formulation

In considering the problem of modelling the input/output curve for a thermal unit, one is guided by common practice adopting a quadratic polynominal in the active power generation P (in MW) to represent the fuel input F (in Mcal/hr). Therefore, one writes:

$$F(P) = a_0 + a_1 P + a_2 P^2 \tag{4-1}$$

Each unit consists of a thermal source (combustion-chamber, boiler-furnace), a turbine and a generator. A unit may either operate with full steam extraction for water co-production or no steam extraction (i.e. zero pass-out) for no water co-production. Equation (4-1) can be written in compact matrix form as:

$$y = \underline{h}^T \underline{X} \tag{4-2}$$

where one replaces F by y and defines

$$\underline{h}^T = [1 \quad P \quad P^2] \tag{4-3}$$

$$X^T = [a_0 \quad a_1 \quad a_2] \tag{4-4}$$

The identification of the parameters a_0, a_1 and a_2 requires the availability of data relating the output P to the input F. Normally, one would be tempted to use data obtained from high cost performance tests, similar to those conducted on commissioning, acceptance tests. Operating records documenting F versus P for each unit provide an alternative option that uses parameter estimation techniques to determine the required model.

Let us assume that each measurement of F is denoted by z to recognize the effects of measurement errors denoted by ν. Therefore one has:

$$z = y + \nu$$

Using eq. (4-2) one thus has the measurement model given by:

$$z = \underline{h}^T \underline{X} + \nu$$

In particular for the k^{th} measurement, one writes:

$$z(k) = \underline{h}^T(k)\underline{X} + \nu(k) \tag{4-5}$$

In general, if (N+1) measurements are available, then one has:

$$\left\{\begin{array}{l} z(k) = \underline{h}^T(k)\,\underline{X} + \nu(k) \\ z(k-1) = \underline{h}^T(k-1)\,\underline{X} + \nu(k-1) \\ \quad\vdots \\ (k-N) = h^T(k-N)\,\underline{X} + \nu(k-N) \end{array}\right\} \tag{4-6}$$

It is convenient to introduce the concatenated measurement vector $\underline{Z}(k)$ which has a dimension (N+1) x 1 and rewrite eq (4-6) in the compact form

$$\underline{Z}(k) = \underline{H}(k)\,\underline{X} + \underline{\nu}(k) \tag{4-7}$$

The matrix H(k) is of dimension (N+1) x n and is given by:

$$H(k) = \begin{bmatrix} h^T(k) \\ h^T(k-1) \\ \vdots \\ h^T(k-N) \end{bmatrix}$$

Here n is the number of unknown parameters and is equal to 3 in the present application.

The notation employed here follows closely that of Mendel [118], which is designed for time-oriented parameter estimation. In the power systems application the data pairs (F,P) are listed such that P values are increasing as the index increases. An example listing of a typical data set is given in Table (4-1). It should be noted that available data sets for thermal units studied in this investigation range from 5 to 23 data points. This is contrasted with only 5 data points available in [117].

Table (4-1)

A typical F - P data set

Count Index	LOAD (MW)	Thermal input (Mcal/hr)
1	10	12.92
2	11	14.20
3	12	15.48
4	13	16.77
5	14	18.06
6	15	19.34
7	16	20.59
8	17	21.85
9	18	23.06
10	19	24.26
11	20	25.50

4.4 Weighted Least Squares Parameter Estimation

The aim of an optimal parameter estimation algorithm is to find an optimal estimate $\hat{X}(k)$ of the parameter vector based on (N+1) data samples. The parameter estimation error $\tilde{X}(k)$ is defined as:

$$\tilde{X}(k) = \underline{X} - \hat{X}(k) \qquad (4\text{-}8)$$

The associated optimal estimate of $Z(k)$ is denoted by $\hat{Z}(k)$ and is related to $\hat{X}(k)$ by:

$$\hat{Z}(k) = H(k)\,\hat{X}(k)$$

The measurement error vector is denoted by $\tilde{Z}(k)$ and is defined by:

$$\tilde{Z}(k) = Z(k) - \hat{Z}(k)$$

clearly, one has:

$$\underline{\tilde{Z}}(k) = \underline{H}(k)\,\underline{\tilde{X}}(k) + \underline{\nu}(k) \qquad (4\text{-}9)$$

The optimal estimate vector $\hat{X}(k)$ is obtained in the weighted (Generalized) Least Squares of errors approach so as to minimize the objective $J[\hat{X}(k)]$ defined by:

$$J[\hat{X}(k)] = \underline{\tilde{Z}}^T(k)\,\underline{W}(k)\,\underline{\tilde{Z}}(k) \qquad (4\text{-}10)$$

The weighting matrix $W(k)$ is assumed to be symmetric and positive definite of dimension $(N+1) \times (N+1)$. If one carries out the minimization with respect to $\hat{X}(k)$, the following linear set of n equations in n unknowns, commonly referred to as the normal equations, is obtained.

$$[\underline{H}^T(k)\,\underline{W}(k)\,\underline{H}(k)]\underline{\hat{X}}(k) = \underline{H}^T(k)\,\underline{W}(k)\,\underline{Z}(k)$$

It is convenient at this point to introduce the matrix $P(k)$ defined by

$$\underline{P}(k) = [\underline{H}^T(k)\,\underline{W}(k)\,H(k)]^{-1} \qquad (4\text{-}11)$$

As a result, the normal equations are written as:

$$\hat{X}(k) = P(k)\,H^T(k)\,W(k)\,Z(k) \qquad (4\text{-}12)$$

Eq. (4-11) & (4-12) completely specify the optimal estimate vector $\hat{X}(k)$.

Error Statistics

The estimate $\hat{X}(k)$ is called an unbiased estimate of X, if $E\{\hat{X}(k)\} = \underline{X}$ The generalised least squares estimate of $\underline{X}$ given in eq.(4-12) is unbiased if $\underline{\upsilon}(k)$ has zero mean and if $\underline{\upsilon}(k)$ and H(k) are statistically independent.

In the case of a zero mean $\underline{\upsilon}(k)$, deterministic $\underline{H}(k)$, $\underline{\upsilon}(k)$ and H(k) being statistically independent and the covariance matrix of the measurement errors given by:

$$R(k) = E\{\upsilon(k)\ \upsilon^T(k)\} \tag{4-13}$$

the covariance matrix of the parameter estimation error denoted by Cov $[\tilde{X}(k)]$ is obtained as:

$$\text{Cov}\ [\tilde{X}(k)] = P(k)\ [H^T(k)\ W(k)\ R(k)\ W(k)\ H(k)]\ P(k) \tag{4-14}$$

Moreover, if the components of V(k) are equally distributed with zero mean and variance σ^2, then

$$\text{Cov}\ [\tilde{X}(k)] = \sigma^2\ P(k)\ [H^T(k)\ W^2(k)\ H(k)]\ P(k)$$

It is important to note that if one chooses the weighting matrix W(k) to be the measurement error covariance matrix inverse $W(k) = R^{-1}(k)$, then it follows from eq.(4-14) that Cov $[\tilde{X}(k)] = P(k)$. In this case the matrix P(k) is the parameter error covariance matrix.

The Choice of Weighting Matrix

The optimal estimate of the parameters and the corresponding error statistics depend on the choice of the weighting matrix W(k). There are two restrictions on this (N+1)x(N+1) matrix. The matrix must be positive definite and the inverse $[H^T(k)W(k)H(k)]^{-1}$ must exist. This latter condition is satisfied if H(k) is of maximum rank and W(k) is positive definite. With this, one still has a wide choice of elements of W(k). In the least squares method, the weighting matrix is taken as unity $\underline{W}(k)=I$

If k denotes discrete time, then one can employ a scheme to weight earlier measurements more heavily than latter measurements or vice versa. A weighting matrix can thus be used in both of these situations:

$$W(k) = \begin{bmatrix} \mu^{k-N} & & & \\ & \mu^{k-1+N} & & \\ & & \ddots & \\ & & & \mu^{k} \end{bmatrix} \qquad (4\text{-}15)$$

When $0 < \mu < 1$ one weighs recent measurements more heavily than past ones. In this case one discounts earlier measurements in favour of latter ones.

In the fuel cost model parameter estimation problem, the count index is not discrete time. In this case, one reorders the data pairs such that the least important loading points are placed early with the most important ones placed latter in the ordering scheme.

In the situation when one knows the statistics of the measurement error process, it is good policy to use all available a priori information. This is achieved in the minimum variance method. Assuming that the measurements have zero error;

$$E\{\upsilon(k)\} = 0$$

and $$E\{\underline{\upsilon}(k)\,\underline{\upsilon}(j)\} = 0 \qquad k \neq j$$

$$E\{\underline{\upsilon}(k)\,V^T(k)\} = R(k)$$

$R(k)$ is an $(N+1)\times(N+1)$ positive definite covariance matrix that is assumed known a priori. Minimizing the error covariance of each parameter subject to the requirement that the estimate must be unbiased is referred to as the unbiased minimum. It can be shown that the unbiased minimum variance estimate of $\underline{X}$ is a special case of the weighted least squares estimate $\underline{X}$, when

$$W(k) = R^{-1}(k) \qquad (4\text{-}16)$$

In this case the weighting matrix $R^{-1}(k)$ stresses the contributions of precise measurements and down-plays the contribution of less precise measurements. It is noted that if the probability density function of the measurements vector $f[z(k)]$ is multivariate Gaussian, then the maximum likelihood estimator coincides with the unbiased minimum variance estimator.

4.5 Recursive Parameter Estimation

Let us assume that subsequent to performing an optimal parameter estimation using the weighted least squares technique described by equation (4-12) for (L+1) measurements, it was decided to add more measurements, increasing the total to (L+p+1) where p is the number of additional measurements. One may be tempted to re-estimate X from a fresh start using eq. (4-12) for old and new measurements taken together. It is more efficient, however, to adopt a recursive parameter estimation strategy that uses the available estimate $\hat{X}(k)$ to obtain a new estimate $\hat{X}(k+p)$. This, of course, is done recognizing the contribution of the new p measurements. Stated simply, the new (up-dated) estimate is equal to the sum of old estimate and a correction term.

In the one-step ahead, or one additional measurement recursive estimator, one begins by finding an up-dated value of P(k) according to:

$$P^{-1}(k+1) = P^{-1}(k) + h(k+1)\, W(k+1)\, h^T(k+1) \qquad (4\text{-}17)$$

The weighted (generalized) least squares gain matrix is computed as:

$$K^*(k+1) = P(k+1)\, h(k+1)\, W(k+1) \qquad (4\text{-}18)$$

As a result one obtains the up-dated estimate as [118]:

$$\hat{X}(k+1) = \hat{X}(k) + K^*(k+1)[z(k+1) - h^T(k+1)\,\hat{X}(k)] \qquad (4\text{-}19)$$

This strategy is sometimes referred to as the expanding memory identifier.

Recursive parameter estimation is useful in identifying changes in parameter estimates as new data is introduced.

Alternative Form

An alternative formulation of the recursive estimator makes use of a matrix inversion lemma to yield for the one-step ahead case [118]:

$$K^*(k+1) = P(k)\, h(k+1)\left[h^T(k+1)\, P(k)\, h(k+1) + \frac{1}{W(k+1)}\right]^{-1} \qquad (4\text{-}20)$$

$$P(k+1) = [I - K^*(k+1)\, h^T(k+1)]\, P(k) \qquad (4\text{-}21)$$

$$\hat{X}(k+1) = \hat{X}(k) + K^*(k+1)\,[z(k+1) - h^T(k+1)\,\hat{X}(k)] \qquad (4\text{-}22)$$

This alternative formulation has a significant advantage over the original formulation in terms of required matrix inversions. Other computational considerations such as accuracy and initialization difficulties have to be accounted for in comparing the two formulations. Recursive parameter estimation is sometimes referred to as the sequential weighted least square estimator.

Start-up Considerations

The recursive weighted least squares parameter estimator given in either eq. (4-17) to (4-19) or (4-20) to (4-22) can be started at time k=o either by setting

$X(o)=o$ and $P(o)=\frac{1}{a} I_n$ or $P(o) = a I_n$

where a is a very large number (say 10^6).

Relation to the Discrete Kalman Filter

The discrete Kalman filter provides estimates of the states for the dynamic system described by:

$$X(k+1) = \phi(k+1,k)\, X(k) + \Gamma(k+1,k)\, W(k)$$
$$Z(k+1) = H(k+1)\, X(k+1) + \upsilon(k+1)$$

with $E\{W(k)\}=0 \quad E\{W(j)\, W^T(k)\} = Q(k)\delta_{jk}$

$$E\{\upsilon(k+1)\}=0 \quad E\{\upsilon(k+1)\, \upsilon^T(k+1)\} = R(k+1)\delta_{jk}$$
$$E\{\upsilon(j)\, W^T(k)\}=0$$
$$E\{X(o)\}=0 \quad E\{X(o)\, X^T(o)\} = P(o)$$

The optimal filtered estimate of $X(k+1)$ is denoted by $\hat{X}(k+1|k+1)$ and is given by the prediction equation

$$\hat{X}(k+1|k) = \phi(k+1,k)\, \hat{X}(k|k)$$

Correction Equation

$$\hat{X}(k+1|k+1) = \hat{X}(k+1|k) + K(k+1)[Z(k+1) - H(k+1)\, \hat{X}(k+1|k)]$$

The Kalman weighting or gain matrix is given by:

$$K(k+1) = P(k+1|k)\, H^T(k+1)\, [H(k+1)\, P(k+1|k)\, H^T(k+1) + R(k+1)]^{-1}$$

$$P(k+1|k) = \phi(k+1,k)\, P(k|k)\, \phi^T(k+1,k) + \Gamma(k+1,k)\, Q(k)\, \Gamma^T(k+1,k)$$

$$P(k+1|k+1) = [I - K(k+1)\, H(k+1)]\, P(k+1|k)$$

The discrete Kalman filter can be shown to be equivalent to the recursive parameter estimator if one sets:

$$\underline{X}(k) = X$$

$$\underline{\phi}(k+1,k) = I$$

$$Q(k) = 0$$

This results in

$$\hat{X}(k+1|k+1) = \hat{X}(k|k) + K(k+1)[Z(k+1) - H(k+1)\, \hat{X}(k|k)]$$

$$P(k+1|k+1) = [I - K(k+1)\, H(k+1)]\, P(k|k)$$

$$K(k+1) = P(k|k)\, H^T(k+1)[H(k+1)\, P(k|k)\, H^T(k+1) + R(k+1)]^{-1}$$

These are equations (4-20) to (4-22), the alternative form of the recursive parameter estimator with minor notational differences.

4.6 Constrained Weighted Least Squares Parameter Estimation

In the course of this investigation, it became apparent that there is a need for a methodology, whereby the parameter estimation task may be carried out subject to constraints on the parameters sought. In particular, it is required that the second order coefficients be positive to guarantee a global minimum in economic dispatch terms.

In order to illustrate how this requirement arises, consider the problem of minimizing the total operating costs in an electric power system where each units fuel cost is given by eq. (4-1). Here we seek to minimize

$$F_T = \sum_{i \in R_d} a_{0_i} + a_{1_i} P_i + a_{2_i} P_i^2 \qquad (4\text{-}23)$$

The minimization is carried out while satisfying the power balance equation

$$\Phi = P_d - \sum_{i \in R_d} P_i = 0 \qquad (4\text{-}24)$$

Here R_d is the set of generating units available to dispatch, P_i is the output of the i^{th} unit and P_d is a specified power demand. The constrained problem can be solved by augmenting the cost F_T by the equality constraint Φ using the Lagrange multiplier λ, to obtain

$$\tilde{F}_T = F_T + \lambda\Phi \tag{4-25}$$

To achieve an extremum, we set the partial derivatives of $\tilde{F}_T$ with respect to P_i to zero to obtain

$$\frac{\partial \tilde{F}_T}{\partial P_i} = a_{1_i} + 2a_{2_i} P_i - \lambda = o \tag{4-26}$$

For the extremum to be a minimum, the second partial derivatives of $\tilde{F}_T$ must be positive. Therefore

$$\frac{\partial^2 \tilde{F}_T}{\partial P_i^2} = 2a_{2_i} > o \tag{4-27}$$

The present section deals with two logical approaches to solving this problem. The object of a weighted least squares estimator is to find the optimal estimator vector $\hat{X}(k)$ so as to minimize the weighted error criterion

$$J[\hat{X}(k)] = \tilde{Z}^T(k) \underline{W}(k) \tilde{Z}(k) \tag{4-28}$$

The measurement error vector is given by:

$$\tilde{Z}(k) = Z(k) - H(k)\hat{X}(k)$$

Therefore $J[\hat{X}(k)] = Z^T(k)\, W(k)\, Z(k) + \hat{X}^T H^T(k)\, W(k)\, H(k)\, \hat{X}(k)$

$$- 2\hat{X}^T(k)\, H^T(k)\, W(k)\, Z(k) \tag{4-29}$$

If the parameter estimates are not constrained, one obtains the normal equations as the solution to this problem. In the application to the economic dispatch problem the component $\hat{X}_3(k)$ is required to be positive.

One therefore has to solve a constrained optimization problem. The matrix formulation of the objective function is manipulated to obtain the following form:

$$\mathfrak{J}[\hat{X}(k)] = a_{11}X_1^2 + a_{22}X_2^2 + a_{33}X_3^2 + 2a_{12}X_1X_2 + 2a_{13}X_1X_3$$
$$+ 2a_{23}X_2X_3 - 2b_1X_1 - 2b_2X_2 - 2b_3X_3 \tag{4-30}$$

The minimization is carried out subject to the inequality constraint $X_3 > 0$

The coefficients required for the objective function are given by;

$$a_{11} = \sum_{j=1}^{N} W(j) \qquad (4\text{-}31\text{-}1)$$

$$a_{12} = \sum_{j=1}^{N} W(j)\, P(j) \qquad (4\text{-}31\text{-}2)$$

$$a_{13} = \sum_{j=1}^{N} W(j)\, P^2(j) \qquad (4\text{-}31\text{-}3)$$

$$a_{22} = \sum_{j=1}^{N} W(j)\, P^2(j) \qquad (4\text{-}31\text{-}4)$$

$$a_{23} = \sum_{j=1}^{N} W(j)\, P^3(j) \qquad (4\text{-}31\text{-}5)$$

$$a_{33} = \sum_{j=1}^{N} W(j)\, P^4(j) \qquad (4\text{-}31\text{-}6)$$

$$b_1 = \sum_{j=1}^{N} W(j)\, Z(j) \qquad (4\text{-}32\text{-}1)$$

$$b_2 = \sum_{j=1}^{N} W(J)\, P(j)\, Z(j) \qquad (4\text{-}32\text{-}2)$$

$$b_3 = \sum_{j=1}^{N} W(j)\, P^2(j)\, Z(j) \qquad (4\text{-}32\text{-}3)$$

The number of data points is N. Details of the derivations are given in Appendix H.

The problem of constrained parameter estimation [119] is well recognized in many fields of science and engineering, but has not been treated in the power systems area before.

4.7 Computational Experience

Parameter estimation computations for data of 56 units in the utility system were conducted. Parameter estimates are listed in Table (4-2) and these results are useful for the economic dispatch studies of Chapter 5. Presently an important finding pertaining to data of four identical units is discussed. The results of this investigation bring together all estimation approaches discussed in the preceeding sections.

The least square estimate of the unit parameters of the unit whose data is given in Table (4-1) is obtained as:

$a_0 = -0.7261 \qquad a_1 = 1.414 \qquad a_2 = -0.0051$

Table (4-2)

56 Unit System Data

UNIT NO	P_{MIN}	P_{MAX}	a_0	a_1	a_2
1	8	28	4.05127	2.56441	8.13716×10^{-3}
2	18	18	0.984	1.23	3×10^{-3}
3	18	18	0.984	1.23	3×10^{-3}
4	18	18	0.984	1.23	3×10^{-3}
5	18	18	0.984	1.23	3×10^{-3}
6	8	28	4.05127	2.56441	8.13716×10^{-3}
7	5	12	15.329	0.44088	6.45779×10^{-2}
8	5	12	12.8579	1.11871	3.51052×10^{-2}
9	5	12	11.7029	1.44421	2.73719×10^{-2}
10	8	12	15.1093	2.16889	6.15996×10^{-2}
11	8	12	15.1093	2.16889	6.15996×10^{-2}
12	8	15	1.45384	1.54264	0.95546×10^{-2}
13	10	18	26.8755	1.44404	5.45426×10^{-2}
14	10	18	30.347	0.44969	9.94145×10^{-2}
15	10	18	25.2581	1.4525	4.87114×10^{-2}
16	10	12	21.97844	2.10979	3.80244×10^{-3}
17	10	12	21.9048	2.01706	1.28266×10^{-2}
18	10	18	28.7014	1.28539	5.44844×10^{-2}
19	10	18	25.6835	1.63675	3.92341×10^{-2}
20	8	18	24.5833	1.90378	2.45608×10^{-2}
21	20	60	-2.11152	2.06606	3.66639×10^{-3}
22	20	35	-2.11152	2.06606	3.66639×10^{-3}
23	20	60	-2.11152	2.06606	3.66639×10^{-3}
24	20	35	-2.11152	2.06606	3.66639×10^{-3}
25	20	22	-2.11152	2.06606	3.66639×10^{-3}
26	20	22	-2.11152	2.06606	3.66639×10^{-3}
27	20	60	18.4888	1.89252	4.40305×10^{-3}
28	20	75	23.4884	1.69801	6.09390×10^{-3}
29	40	65	60.8187	0.38813	4.46988×10^{-3}
30	40	65	60.8187	0.38813	4.46988×10^{-3}
31	25	75	68.6577	1.9829	1.78273×10^{-3}
32	25	75	68.6577	1.9829	1.78273×10^{-3}
33	8	30	20.2877	2.02209	7.72012×10^{-3}
34	8	30	21.4153	1.94077	1.10727×10^{-2}
35	8	30	19.5795	2.04247	8.59541×10^{-3}
36	8	30	20.2877	2.02209	7.72012×10^{-3}
37	2	7	3.31764	1.60332	5.49957×10^{-2}
38	2	7	3.31764	1.60332	5.49957×10^{-2}
39	2	7	3.31764	1.60332	5.49957×10^{-2}
40	2	7	3.31764	1.60332	5.49957×10^{-2}
41	6	16	1.68581	1.90197	1.56548×10^{-2}
42	6	16	2.50403	1.81245	2.11244×10^{-2}
43	6	16	1.28414	1.98009	1.43057×10^{-2}
44	6	16	2.02671	1.84021	2.02094×10^{-2}
45	5	15	30.1168	0.87605	1.12108×10^{-1}
46	5	15	30.1168	0.87605	1.12108×10^{-1}
47	5	15	30.1168	0.87605	1.12108×10^{-1}
48	5	15	30.1168	0.87605	1.12108×10^{-1}
49	5	15	30.1168	0.87605	1.12108×10^{-1}
50	5	15	30.1168	0.87605	1.12108×10^{-1}
51	8	22	47.3067	0.81433	9.09956×10^{-2}
52	8	22	47.3067	0.81433	9.09956×10^{-2}
53	8	22	47.3067	0.81433	9.09956×10^{-2}
54	8	22	47.3067	0.81433	9.09956×10^{-2}
55	8	26	22.479	3.43752	6.94301×10^{-3}
56	8	26	22.479	3.43752	6.94301×10^{-3}

The appearance of a negative value of a_2 leads to no guarantee of optimality in economic dispatch studies, since under certain conditions it may lead to a non-positive definite Hessian. In fact, when losses are neglected in the dispatch formulation, the Hessian is a diagonal matrix of second order coefficients of the individual units a_2. A parameter estimate solution with this property is called an infeasible parameter estimate.

In order to determine the cause of this anomaly and to devise a remedy, a number of computational experiments have been conducted. One can classify the experiments as belonging to a diagnostic phase followed by a corrective phase. The purpose of this investigation is to detect the effect of each of the data pairs on the value of the estimate of the parameter a_2. Such data pairs can then either be excluded or down-weighted since it would appear that their measurements involved an extraordinary error.

Phase A: Diagnostic Phase

The objective of this phase is to attempt to determine the data point or points that lead to the infeasible parameter estimates. Two experiments were conducted in phase A for this purpose.

Experiment A-1

In this experiment, the technique of recursive least squares parameter estimation was employed as a diagnostic tool. The object is to determine the particular data point or points that cause/force a negative value of a_2 to appear. The first three measurements were used to obtain a least squares estimate (this is a perfect fit, since the number of experiments is equal to the number of estimated parameters). The estimates are listed in row 1 of Table (4-3) corresponding to measurement index 3. These estimates were used as starting values to obtain new estimates by including the fourth measurement information. The recursive procedure was carried out and the full sets of results is shown in Table (4-3).

Table (4-3)
Recursive Parameter Estimation Results

Measurement Index	Power Generation	a_0	a_1	a_2
3	12	0.4303	1.2228	0.0026
4	13	0.4310	1.2226	0.0026
5	14	0.3644	1.2345	0.0021
6	15	0.1713	1.2680	0.0007
7	16	-0.1869	1.3281	-0.0018
8	17	-0.3039	1.3473	-0.0026
9	18	-0.5618	1.3883	-0.0042
10	19	-0.7645	1.4197	-0.0053
11	20	-0.7361	1.4154	-0.0052

It is noted from the table that values of a_2 are negative for measurements starting at the 7th experiment, corresponding to 16MW unit loading. A closer examination reveals that the introduction of the 6th experiment (15MW) results in relatively significant decrease in the value of a_2. Thus one is led to conclude that experiments 6 upwards introduce the negative value of the second order coefficients a_2 for this unit. The fuel cost model follows a convex quadratic up to 15MW, and then turns concave. Note that the negative estimate of a_0 implies that a negative cost is associated with generating zero power. This is of no consequence to the optimization as the values of generation are not allowed to be below a positive minimum.

Experiment A-2

In this investigation, the coefficients a_0, a_1, a_2 are determined by taking 3 measurements at a time and employing a sliding window. The results are given in Table (4-4). It is clear that the introduction of the 6th experiment results in a negative value of a_2. This effect prevails up to the last

Table (4-4)
Sliding Window Parameter Estimation Results

Indices of Measurements Included	a0	a1	a2
1-3	0.12	1.28	8.8×10^{-16}
2-4	0.78	1.165	5×10^{-3}
3-5	1.07×10^{-4}	1.29	0
4-6	- 0.91	1.425	-5×10^{-3}
5-7	- 3.01	1.715	-1.5×10^{-2}
6-8	1.79	1.095	5×10^{-3}
7-9	- 6.37	2.085	-2.5×10^{-2}
8-10	- 0.25	1.385	-5×10^{-3}
9-11	8.3	0.46	2×10^{-2}

set (experiments 9-11). This confirms the conclusions of the investigation of experiment 1. Another way to arrive at a conclusion as to which measurements cause a negative a_2 is to consider the results of finite differencing to approximate a_2 from the measurement records. It is clear from inspection of Table (4-5) that experiments 6,7,9,and 10 correspond to a negative second difference.

Table (4-5)
Finite Difference Results

Data Point i	1	2	3	4	5	6	7	8	9	10	11
$z(i)$	12.92	14.2	15.48	16.77	18.06	19.34	20.59	21.85	23.06	24.26	25.5
$\Delta z(i)$		1.28	1.28	1.29	1.29	1.28	1.25	1.26	1.21	1.20	1.24
$\Delta^2 z(i)$			0	0.01	0	-0.01	-0.03	0.01	-0.05	-0.01	0.04

Phase B: Corrective Phase

Here, a number of techniques that could possibly lead to a positive value of a_2 were attempted.

Experiment B-1

A straight-forward least squares parameter estimation computation is carried out:

a) The measurements for P=15, and 16 MW corresponding to indices 6 and 7, were deleted and a least squares estimation procedure was applied. The result is

$a_0 = -0.6538 \qquad a_1 = 1.4036 \qquad a_2 = -0.0048$

It is clear that no improvement took place.

b) Dropping points 6,7,9,and 10 resulted in still negative values

$a_0 = -0.6358 \qquad a_1 = 1.4002 \qquad a_2 = -0.0046$

c) Dropping points 5,6,7,8,9, and 10 gave

$a_0 = -0.3535 \qquad a_1 = 1.3611 \qquad a_2 = -0.0034$

The significant change in a_2 towards zero lead to eliminating point 11 to obtain

$a_0 = 0.3703 \qquad a_1 = 1.2336 \qquad a_2 = 0.0021$

This uses only 4 points but yields a positive a_2. Note that including up to point 6 as suggested by the recursive algorithm yields still a positive value of a_2.

Experiment B-2

Here the idea of a weighting matrix using a discount factor b_i, was tried. The diagonal elements of $\underline{W}$ are given by

$$W(i,i) = b^{N-i}$$

Table (4-6) lists the results. It was found that for low values of b up to 0.312, a positive a_2 can be obtained

$a_0 = 0.98366 \qquad a_1 = 1.2255 \qquad a_2 = 1.3717\times10^{-5}$

The corresponding weighting factors for this case:

$w(1,1) = 8.7407\times10^{-6}$

$w(2,2) = 2.8015\times10^{-5}$

$w(3,3) = 8.0792\times10^{-5}$

$w(4,4) = 2.8779\times10^{-4}$

$w(5,5) = 9.2242\times10^{-4}$

$w(6,6) = 2.9565 \times 10^{-3}$

$w(7,7) = 9.4759 \times 10^{-3}$

$w(8,8) = 3.0371 \times 10^{-2}$

$w(9,9) = 9.7344 \times 10^{-2}$

$w(10,10) = 3.12 \times 10^{-1}$

$w(11,11) = 1$

This weighting sequence favours the last five data points.

Table (4-6)

Parameter Estimates with Weighting $w(i,i) = b^{N-i}$

b	a_0	a_1	a_2
0.1	5.449	0.7543	1.2412×10^{-2}
0.2	2.9709	1.0136	5.6417×10^{-3}
0.3	1.1529	1.2072	5.04×10^{-4}
0.4	4.185×10^{-2}	1.3283	-2.7794×10^{-3}

The weighting matrix was changed to $w(i,i) = b^{i-1}$. This weighs earlier measurements heavier than latter ones. Table (4-7) lists the results. For a discount factor upto 0.3 the value of a_2 is positive.

Table (4-7)

Parameter Estimates with Weighting $w(i,i) = b^{i-1}$

b	a_0	a_1	a_2
0.1	0.2364	1.258	0.001
0.2	0.2480	1.2558	0.0011
0.3	0.1575	1.2721	0.0004
0.4	0.0004	1.2996	-0.0008

Experiment B-3

In this case the following weighting matrix was attempted.

$$\omega(i,i) = b^{N-i} + a$$

An extensive computational experiment was conducted with the result that negative values of a_2 appeared in all cases. As a sample of the results, one finds:

For b = 1 $\quad$ a = 0.1 : a_0 = -0.7355 $\quad a_1$ = 1.4153 $\quad a_2$ = -0.0052

For b = 0.025 $\quad$ a = 0.2 : a_0 = -0.719 $\quad a_1$ = 1.419 $\quad a_2$ = - 0.0052

The estimates do not appear to be sensitive to the variations in this particular weighting matrix. This is a discount factor plus a bias term weighting.

Experiment B-4

In this case a weighting matrix which is unity up to the 5th measurement followed by a discount factor weighting was attempted.

$$\omega(i,i) = 1 \qquad i=1,\dots,5$$
$$= b^{N-i} \qquad i=6,\dots,11$$

The discount factor b was varied from b=1 to b=0.025 in steps of 0.025. The results have consistently contained a negative a_2 as indicated in Table (4-8).

Table (4-8)

Parameter Estimates with Weighting $w(i,i) = b^{N-i}$ for $i > 5$

b	a_0	a_1	a_2
0.9	-0.696	1.4097	-0.005
0.8	-0.6538	1.4037	-0.0048
0.7	-0.6123	1.3979	-0.0046
0.6	-0.5750	1.3927	-0.0044
0.5	-0.5451	1.3885	-0.0043
0.4	-0.5240	1.3855	-0.0042
0.3	-0.5112	1.3837	-0.0042
0.2	-0.5052	1.3828	-0.0041
0.1	-0.5039	1.3826	-0.0041

Experiment B-5

This experiment is concerned with use of the constrained parameter estimation algorithm, as an implementation of the Generalized Reduced Gradient method detailed in Chapter 5. Using this nonlinear programming approach, we seek to minimize the objective function given by eq.(4-31) subject to the requirement that $X_3>0$. With strict inequality the optimization algorithm indicated that there is no feasible solution. It was thus decided to investigate the problem with the inequality constraint $X_3 \geq \epsilon$ where ϵ is a small positive number representing a chosen lower bound on X_3.

In Table (4-9), the optimal estimates are listed for the choice of lower bounds as

$a_0 \geq -50 \qquad a_1 \geq 50 \qquad a_2 \geq 0.001$

Many starting guess values were tried as shown with the result that the lower bound on a_2 was obtained as its optimum value.

Table (4-9)
Starting Guess Effects

STARTING GUESS			SOLUTION		
a_0	a_1	a_2	a_0	a_1	a_2
0.00	0.00	0.00	0.6083	1.229	0.001
5.00	5.00	5.00	0.6002	1.229	0.001
-1	-1	-1	0.6118	1.229	0.001
100	100	100	0.6002	1.229	0.001

Table (4-10) corresponds to a different set of lower bounds given by

$a_0 \geq -50 \qquad a_1 \geq 50 \qquad a_2 \geq 0.005$

The same result was obtained. Note that in this case the optimum values of a_0 and a_1 differ slightly depending on the initial guess values as indicated. This is due to the nature of the iterative method and can be expected.

Table (4-10)
Starting Guess Effects

STARTING GUESS			SOLUTION		
a_0	a_1	a_2	a_0	a_1	a_2
0.00	0.00	0.00	1.469	1.109	0.005
5.00	5.00	5.00	1.457	1.11	0.005
-1	-1	-1	1.470	1.109	0.005
100	100	100	1.457	1.11	0.005

Table (4-11) lists the results with different values of the lower bound on a_2. Again, the optimum values of a_2 is its specified lower bound. As a_2 is increased we see that a_0 increases with a corresponding decrease in a_1.

One can conclude that the most appropriate way for the present constrained estimation problem is to fix a_2 at a specified lower bound and carry out a linear estimation procedure for a_0 and a_2. In this case we obtain a unique solution for a_1 and a_2.

Table (4-11)
Lower Bound Effects

LOWER BOUNDS			SOLUTION		
a_0	a_1	a_2	a_0	a_1	a_2
-50	-50	0.001	0.6002	1.229	0.001
-50	-50	0.005	1.457	1.11	0.005
-50	-50	0.01	2.531	0.9597	0.01
-50	-50	0.015	3.667	0.8054	0.015
-50	-50	0.02	4.702	0.6566	0.02007
-50	-50	0.025	5.766	0.5081	0.02502

4.8 Summary

This chapter dealt with the important area of parameter estimation of thermal generation production cost models. Beginning with a background review of this important area, and a discussion of the parameter estimation problem formulation, the chapter proceeded with brief outlines of well established approaches to solving the problem. In particular, the weighted least squares parameter estimation and recursive parameter estimation were discussed as the basis for computational work reported in this chapter.

As a result of experimentation with data for the utility system, it was recognized that in a number of instances, the estimated parameters were unsatisfactory from an end use point of view. One of the parameters of the quadratic model is required to be positive to guarantee a minimum in subsequent optimization work. This aspect of the problem appears not to have been previously addressed in the literature. The chapter included results of an investigation of approaches that might meet this requirement. In the course of this work, the technique of iteratively reweighted least squares parameter estimation was used and will therefore be documented in Chapter 6. The requirement of a positive coefficient is rightfully called a constrained parameter estimation problem and as a result of this, the chapter reported on the use of a nonlinear programming approach to the problem. This again is believed be a new application.

In the course of the investigation, the utility of weightings and recursive parameter estimation as diagnostic and remedial tools was realized. These aspects were discussed in the concluding section of this chapter.

CHAPTER 5
ECONOMIC DISPATCH STUDIES

5.1 Introduction

This chapter is devoted to a study of efficient algorithms for optimal economic dispatching of electric power systems. The problem considered deals with a system that includes thermal generation sources, and it is desired to determine the optimum loading to satisfy given load requirements and to simultaneously observe system and equipment constraints.

The solution of the problem considered is important in a number of economy-security functions of electric power systems. It is used primarily as a stand-alone function for dispatch in systems with concentrated load and negligible losses. Results of the dispatch are used to establish merit ordering. The optimal schedules and associated optimal costs are also used as a basis for power interchange and pool-transaction cost evaluation in large electric power interconnections. Finally, the solutions are used to establish initial guesses for more complex scheduling problems.

As a result of an extensive study of available optimization techniques, three methods emerged as the most practical tools to carry out the dispatching function. The methods discussed in this chapter are the Generalized Reduced Gradient (GRG) algorithm, the Lambda Iteration (LI) algorithm, and a recently developed technique called the Lambda Aggregation (LA) algorithm.

The LA algorithm is the result of the author's investigations and is based on a specific implementation of the principle of equal incremental cost of power delivered loading, for systems with quadratic objective functions. The implementation uses a set of equivalency relations to solve the problem in conjunction with the Kuhn-Tucker optimality conditions. Among its advantages is that no initial guess of the unknowns is required. All other methods require and are sensitive to the choice of the initial guess. The proposed technique is applicable, with slight modifications, to more complex problems where losses are accounted for using a linear model. It can also be applied to systems with fuel cost models of exponential form.

The theoretical foundations and details of the three methods are discussed in this chapter, and computational experience related to the 56 units in the existing UAE's Abu Dhabi utility system is presented. Guidelines for selecting initial guess values for the GRG and LI algorithm are discussed. Experience with other available methods shows the techniques discussed to be extremely fast and robust. A comparison of execution times of the three techniques on a DEC micro VAX-II running under VMS with 6 Mb of memory is given. A critical comparative evaluation of the methods is also detailed.

5.2 Historical Survey

Starting early in the Twentieth Century, electric utility operators realized the importance of scheduling generating resources in the most economic manner as is evident from Arismunander and Noakes' bibliography [120]. Interest in this area continues due to the potential savings as well as the development of new and efficient solution methods.

In an IEEE Working Group Report [121], economic dispatch is defined as:
The allocation or change in allocation of the power resources which are connected to the system at a particular time to meet the system load at that time in a manner which minimizes the overall cost to the system. Economic dispatch normally takes into account generating unit incremental input/output characteristics which include the effects of power station auxiliaries, fuel costs.

Two companion papers [122] and [123] chart in bibliographical form the developments of ideas in the areas of economy-security functions up to 1979. The solution to the original problem is commonly known as the equal incremental cost principle.

The principle of equal incremental cost of power delivered for optimal scheduling of electric power systems is attributed to Steinberg and Smith in their two papers [124] published in 1934. The principle was further developed in their book entitled "Economy Loading of Power Plants and Electric Systems" [125] in 1943. Interest in practical implementation of the optimal strategies continued in 1953 by Ward [126], Early, Phillips, and Shreve, [127], where an incremental cost of power developed device is described in 1955. Kirchmayer's classic book on "Economic Operation of Power Systems", [128] summarizes developments in this area. More recently, books such as El-Hawary and Christensen [4] in 1979, Wood and Wollenberg [2] in 1984, and Heydt [129] in 1986 treat practical aspects of the overall problem of economic operation of power systems.

It is important to realize that the inclusion of the inequality constraints is a complicating factor that makes it necessary to use special algorithms such as the Lambda Iteration method [2] and [129], and nonlinear programming techniques [130]. The LA technique described here uses an efficient combination of results of Kuhn-Tucker [131] and special reduced forms of solution given by El-Hawary [132] to yield the Lambda Aggregation technique.

5.3 Problem Formulation

Optimal economic dispatching of an electric power system requires finding the optimal power generation levels of each unit in the system to cover a system load (power demand) and minimizing the system production cost. The optimization is carried out such that all system and equipment constraints are observed. System constraints include network balance equations (represented by the power flow equations) as in the optimal power flow formulation or the active power balance equation in the conventional dispatch function.

The aim of the optimization process is to minimize the operating cost function F_T defined by:

$$F_T = \sum_{i \in R_d} F(i) \qquad \text{(Mcal/hr)} \qquad (5\text{-}1)$$

The set of dispatchable units in the system is denoted by R_d. This is essentially all units committed for generation. The fuel cost functions F(i) represent the variation of the input to the unit with its active power generation.

The active power generation of the system is constrained by the system load requirements. When transmission losses are neglected this constraint is represented by the power balance equation:

$$P_{dem} = \sum_{i \in R_d} P(i) \qquad \text{(MW)} \qquad (5\text{-}2)$$

This is a simple equality constraint with P_{dem} denoting the system load.

The active power generation of each unit P(i) is constrained by upper and lower bounds representing unit capacity and system spinning reserve requirements. This is stated as:

$$b_l(i) \leq P(i) \leq b_u(i) \qquad (5\text{-}3)$$

Equations (5-3) is a set of inequality constraints that requires special consideration in all implementations discussed.

5.4 The Variational Solution

The variational solution to the dispatching problem ignoring the inequality constraints leads to the well known equal incremental cost of loading principle [2] and [4] which states that the incremental cost of power generation at each unit should be equal to the system incremental cost λ. This is stated as:

$$\frac{dF(i)}{dP(i)} = \lambda \qquad \text{(Mcal/hr/MW)} \qquad (5\text{-}4)$$

To account for inequality constraints, the Kuhn-Tucker conditions [4] are used. The process involves solving Eq.(5-4) and then examining the resulting solution for violations to the inequality constraints. Variables that violate a bound are set to that bound and are excluded from subsequent optimization.

So far, nothing has been stated about the nature of the fuel cost model F(i). There are a number of representations for the variation of F with the active power generation P. If one uses a linear model, then the optimization problem reduces to a linear programming problem and available software can be used directly to obtain the optimal dispatch strategy.

It is well known that a quadratic cost model is preferred in the power industry to represent the variation of the fuel cost F with the active power generation. Here the model is of the form:

$$F(i) = a_0(i) + a_1(i)P(i) + a_2(i)P^2(i) \qquad (5\text{-}5)$$

The parameters $a_0(i)$, $a_1(i)$, and $a_2(i)$ are assumed available for each unit. In this case one has:

$$\frac{dF(i)}{dP(i)} = a_1(i) + 2a_2(i)P(i) \qquad (5\text{-}6)$$

This particular formulation allows one to design an extremely fast and efficient algorithm for optimum economic dispatch called the Lambda Aggregation method. The foundations of the technique are discussed now.

If inequality constraints are not present, one can easily show that the system operating in an optimal sense can be represented by one equivalent unit with optimal cost function given by:

$$F_{eq} = a_{0_{eq}} + a_{1_{eq}} P_{dem} + a_{2_{eq}} P_{dem}^2 \qquad (5\text{-}7)$$

The subscript (eq) is used to denote an equivalent quantity.

The equivalent cost parameters are given by:

$$a_{2_{eq}}^{-1} = \sum_{i \in R_d} a_2^{-1}(i) \qquad \text{(Mcal/MWhr x MW)} \quad (5\text{-}8)$$

$$a_{1_{eq}} = a_{2_{eq}} \sum_{i \in R_d} a_1(i)\ a_2^{-1}(i) \qquad \text{(Mcal/MWhr)} \quad (5\text{-}9)$$

The optimal system incremental cost of power delivered in this case is given by:

$$\lambda = a_{1_{eq}} + 2a_{2_{eq}} P_{dem} \qquad (5\text{-}10)$$

The individual optimum power generation are obtained as:

$$P(i) = \frac{\lambda - a_1(i)}{2a_2(i)} \qquad (5\text{-}11)$$

The derivation of the preceeding equations can be found in [132]. The Lambda Aggregation technique uses equations (5-8) to (5-11) and the Kuhn-Tucker conditions to solve the problem accounting for inequality constraints. Note that equations (5-8) to (5-11) give a closed form solution in the absence of inequality constraints.

5.5 The Generalized Reduced Gradient Method

The concept of the GRG method goes back to 1964-65, when it was proposed by Abadie and Carpentier [133 and 134] as a generalization of Wolfe's Reduced Gradient Method [135 and 136]. The first implementation of the method was ranked first in Colville's 1968 study, comparing some 30 methods using 8 nonlinear programming problems [137]. Subsequent improvements to the code resulted in many software implementations including Lasdon and Waren's GRG2 [138 to 140], Murtagh and Saunder's MINOS [141 to 143]. The latter is designed specifically for large sparse linearly constrained problems.

The reduced gradient method was adopted by Dommel and Tinney [144] in solving the optimal power flow problem in 1968. The first application of the original GRG technique to the OPF problem is due to Peschon et al. [145], in 1972. Further refinements have been reported by Burchett et al. [146 and 147]. The algorithm reported in [147] is an adaptation of the MINOS implementation of the GRG concept.

The purpose of GRG algorithms is to solve nonlinear programming problems of the following canonical form:

$$\text{Minimize } f(x)$$

subject to

$$g(x) = b$$
$$Lb \le h(x) \le Ub \qquad (5\text{-}12)$$
$$L \le x \le U$$

where x is an nx1 column vector of the decision variables, f(x) is the objective function, g is an mx1 column vector of equality constraint, h is a px1 column vector of inequality constraint functions, and Lb, Ub, L, and U are vectors of corresponding lower and upper bounds. Appendix I briefly reviews the GRG algorithm.

5.6 The Lambda Iteration Method

The Lambda Iterations method [2] and [129] is based on an intuitive approach that recognizes the fact that if the value of the system incremental production cost is known, then with the availability of the incremental cost characteristics for each unit, the corresponding power output of each unit is found. In the case of negligible losses, the sum of the resulting power generations is obtained. The sum is compared with the required power demand, and if they are equal, an optimal solution has been found. If the sum of generations is higher than the given power demand, then the value of lambda is reduced and the process is repeated. If, on the other hand, the value of the sum of power generations is less than the power demand, then the value of lambda is increased, and the process is repeated.

Dealing with inequality constraints on the output of the units is simple in the Lambda iteration method, since one sets the generation of the unit to the value corresponding to the violated limit, as it occurs.

There are a number of practical implementation aspects that need particular attention in the Lambda iteration method. The first is the amount of reduction (or increase) in lambda value for the next iteration. A recommended procedure involves a Golden Section search. The second and most important aspect is that of the initial guess value of lambda. A number of possibilities exists, and these are discussed in the section on computational results.

5.7 The Lambda Aggregation Technique

The theoretical background given in section (5-4) is the basis of the Lambda Aggregation technique for optimal economic dispatch algorithm described here. The algorithm is designed as a main routine comprising three steps. The first is called the aggregation step and solves a dispatch problem without inequality constraints using the concepts of equivalent cost function for the dispatchable units. The second step is called the reset step and is an implementation of the Kuhn-Tucker conditions. The third step is a verification step that directs the flow of subsequent computations.

There are two auxiliary routines, called the overdispatch and underdispatch routines which are used whenever the resulting optimization sets all units to their respective upper and lower bounds.

As a preparatory step, the elements of an array called Mark (i) are set to zero. The value of Mark (i) corresponds to the status of the i^{th} unit. A zero value signifies that the unit is available to dispatch. If a unit is to remain at its upper bound, the corresponding value is Mark (i) = +1. In a similar manner Mark (i) = -1, indicates that the unit is set to its lower bound.

The term dispatchable demand, denoted by P_{dd} is used to describe the portion of the system power demand to be covered by the available dispatchable units. It is equal to the difference between the system power demand p_{dem} and the sum of the generations of units set to their upper or lower bounds. Initially the value of P_{dd} is equal to P_{dem}, since all units are available to dispatch.

Lambda Aggregation Step

In this step, the parameters a_{eq1} and a_{eq2} of the equivalent cost function for the system composed of all dispatchable units are obtained using equations (5-8) and (5-9). The dispatchable system incremental cost of power generated λ is computed using:

$$\lambda = a_{eq1} + 2a_{eq2}P_{dd} \qquad (5\text{-}13)$$

The optimum power generation of each dispatchable unit can now be computed as:

$$P(i) = \frac{\lambda - a_1(i)}{2a_2(i)} \qquad (5\text{-}14)$$

It should be noted that the outcome of this step is not guaranteed to yield optimum generations that satisfy the imposed inequality constraints. This aspect is treated in the second (reset) step.

Reset Step

This step is a direct implementation of the Kuhn-Tucker conditions. Here the values of each optimum generation P(i) obtained in the Lambda Aggregation step is checked for violations to the upper and lower bounds. The generations and values of the array Mark (i) are reset to respect the constraints and indicate unit status in the following manner.

Lower Bound Violation: If $P(i) < b_l(i)$ Set $P(i) = b_l(i)$ Set Mark(i) = -1
Upper Bound Violation: If $P(i) > b_u(i)$ Set $P(i) = b_u(i)$ Set Mark(i) = +1
No Bound Violation: If P(i) is within bounds Set P(i)=P(i) Set Mark(i)=0

Verification Step

The result of the Reset Step is a set of power generations and associated array value Mark(i). Two Verification computations are necessary to determine subsequent action. The difference between the power demand P_{dem} and the sum of P(i) is calculated as P_{dif}, given by:

$$P_{dif} = P_{dem} - \sum P(i) \qquad (5\text{-}15)$$

If the magnitude of P_{dif} is less than a specified tolerance limit ϵ, then an optimum dispatch satisfying the demand constraint and the upper and lower bound constraints has been obtained and the program is successfully terminated. If the magnitude of P_{dif} is not less than the specified tolerance limit ϵ, then an optimum dispatch satisfying the demand constraint and the upper and lower bound constraints has not been obtained. The sum of unit generations set at the upper bounds denoted by σ_{bu}, and the sum of generations set at the lower bounds denoted by σ_{bl} are calculated as:

$$\sigma_{bu} = \sum P(i) \tag{5-16}$$

The sum is over all i such that Mark(i) = +1

$$\sigma_{bl} = \sum P(i) \tag{5-17}$$

The sum is over all i such that Mark(i) = -1

The dispatchable power demand P_{dd} is computed as:

$$P_{dd} = P_{dem} - \sigma_{bu} - \sigma_{bl} \tag{5-18}$$

The array values Mark(i) are examined. If some values of the array are zero, this indicates that not all units have been set to either their respective upper or lower bounds. The algorithm goes back to the first (aggregation) step, to repeat the process.

If all values of the array Mark(i) are either + 1 or -1, this indicates that all units have been set to either their respective upper or lower bounds. No dispatchable units are available. In this case one needs to go to either the overdispatch routine or the underdispatch routine according to the value of power difference P_{dif} calculated earlier in this step.

<u>**Overdispatch Routine**</u>

This routine is entered from the verification step for the case with all units having been set to their upper and lower bounds and $dP_{dif} < 0$, i.e. the sum of generations P(i) is greater than the power demand. One needs here to reduce the generated powers.

For maximum economy, it is clear that in this case the units set at their lower bounds should remain there, and allow all units set to their upper bounds to be scheduled below their upper bounds. As a result in the overdispatch routine one sets Mark(i) = 0 for all units with Mark (i) = +1 from the main routine. The value of P_{dd} is calculated as:

$$P_{dd} = P_{dem} - \sigma_{bl} \tag{5-19}$$

The next steps involve repeating the aggregation and reset steps of the main routine with the object of reducing the power level of units at their upper bound enough to make the generation match the demand. The reduction is done in terms of a merit ordering scheme reducing generation from units with highest incremental cost. The process is repeated in accordance with a merit order until P_{dif} is reduced to zero.

The verification step of this routine calculates:

$$P_{dif} = P_{dem} - \sum P(i) \tag{5-20}$$

If the magnitude of P_{dif} is less than the specified tolerance, the program successfully terminates. If P_{dif} is less than zero, the algorithm repeats the overdispatch function. If P_{dif} is greater than zero, the algorithm branches to the underdispatch function.

Underdispatch Routine

This routine is entered from the verification step for the case with all units having been set to the upper and lower bounds and $P_{dif} > 0$, i.e the sum of generations P(i) is less than the power demand. One needs here to increase the generated powers.

For maximum economy, it is clear that in this case the units set at their upper bounds should remain there, and allow all units set to their lower bounds to be scheduled above their lower bounds. As a result in the underdispatch routine one sets Mark(i) = 0 for all units with Mark(i) = -1 from the main routine. The value of P_{dd} is calculated as:

$$P_{dd} = P_{dem} - \sigma_{bu} \tag{5-21}$$

The next step involve repeating the aggregation and reset steps of the main routine.
The verification step of this routine calculates:

$$P_{dif} = P_{dem} - \sum P(i) \tag{5-22}$$

If the magnitude of P_{dif} is less than the specified tolerance, the program successfully terminates. If P_{dif} is greater than zero, the algorithm repeats the underdispatch function. If on the other hand P_{dif} is less than zero, the algorithm branches to the overdispatch function.

5.8 Computational Results

The algorithms described in the preceding sections were implemented on a DEC Micro-VAX-II in Fortran-77. The programs were used to obtain the most economic schedule of generation for a utility system consisting of 56 units with maximum installed capacity of 1426 MW. The system load varied from 700 to 1400 MW in the computational experiments conducted.

5.8.1 System Data

Data of the system are given in Table (5-1) listing unit lower and upper bounds and the quadratic fuel cost parameters. The system considered consists of a mix of steam units, gas turbines, and diesel units, that vary in capacity from 7 MW to 75 MW. The evolution of the system leads to a mix of upper and lower bounds on each unit that proved to be a complicating factor for other software tested. From Table (5-1), it can been seen that the following upper and lower bound overlaps exists:

1- The upper bound of units 37-40 is 7 MW, which is below the lower bounds of units 1-6,10-36, and 51-56.

2- The upper bound of units 7-11 and 16-17 is 12 MW, which is below the lower bounds of units 2-6 and 21-32.

3- The upper bound of units 12, and 45-46 is 15MW, which is below the lower bounds of units 2-5 and 21-32.

4- The upper bound of units 41-44 is 16 MW, which is below the lower bounds of units 2-6 and 21-32.

5- The upper bound of units 2-5, 13-15 and 20 is 18 MW, which is below the lower bounds of units 21-32.

6- The upper bound of units 25-26 and 51-54 is 22 MW, which is below the lower bounds of units 29-32.

7- The upper bounds of units 55-56 is 26MW, which is below the lower bounds of units 29-30.

8- The upper bounds of units 6 is 28 MW, which is below the lower bounds of units 29-30.

9- The upper bounds of units 33-36 is 30 MW, which is below the lower bounds of units 29-30.

10- The upper bounds of units 22 and 24 is 35 MW, which is below the lower bounds of units 29-30.

One also notes that units 2-5 are required to operate at 18 MW generation level. Moreover, there are many identical units in the system in terms of their fuel cost model parameters. From Table (5-1), the following can be observed:

1- Units 2-5 are identical
2- Units 10 and 11 identical
3- Units 25 and 26 are identical
4- Units 29 and 30 are identical
5- Units 31 and 32 are identical
6- Units 37-40 are identical
7- Units 45-50 are identical
8- Units 51-54 are identical
9- Units 55-56 are identical

Table (5-1)
56 Unit System Data

UNIT NO	P MIN	P MAX	a 0	a 1	a 2
1	8	28	4.05127	2.56441	8.13716×10^{-3}
2	18	18	0.984	1.23	3×10^{-3}
3	18	18	0.984	1.23	3×10^{-3}
4	18	18	0.984	1.23	3×10^{-3}
5	18	18	0.984	1.23	3×10^{-3}
6	8	28	4.05127	2.56441	8.13716×10^{-3}
7	5	12	15.329	0.44088	6.45779×10^{-2}
8	5	12	12.8579	1.11871	3.51052×10^{-2}
9	5	12	11.7029	1.44421	2.73719×10^{-2}
10	8	12	15.1093	2.16889	6.15996×10^{-2}
11	8	12	15.1093	2.16889	6.15996×10^{-2}
12	8	15	1.45384	1.54264	0.95546×10^{-2}
13	10	18	26.8755	1.44404	5.45426×10^{-2}
14	10	18	30.347	0.44969	9.94145×10^{-2}
15	10	18	25.2581	1.4525	4.87114×10^{-2}
16	10	12	21.97844	2.10979	3.80244×10^{-3}
17	10	12	21.9048	2.01706	1.28266×10^{-2}
18	10	18	28.7014	1.28539	5.44844×10^{-2}
19	10	18	25.6835	1.63675	3.92341×10^{-2}
20	8	18	24.5833	1.90378	2.45608×10^{-2}
21	20	60	-2.11152	2.06606	3.66639×10^{-3}
22	20	35	-2.11152	2.06606	3.66639×10^{-3}
23	20	60	-2.11152	2.06606	3.66639×10^{-3}
24	20	35	-2.11152	2.06606	3.66639×10^{-3}
25	20	22	-2.11152	2.06606	3.66639×10^{-3}
26	20	22	-2.11152	2.06606	3.66639×10^{-3}
27	20	60	18.4888	1.89252	4.40305×10^{-3}
28	20	75	23.4884	1.69801	6.09390×10^{-3}
29	40	65	60.8187	0.38813	4.46988×10^{-3}
30	40	65	60.8187	0.38813	4.46988×10^{-3}
31	25	75	68.6577	1.9829	1.78273×10^{-3}
32	25	75	68.6577	1.9829	1.78273×10^{-3}
33	8	30	20.2877	2.02209	7.72012×10^{-3}
34	8	30	21.4153	1.94077	1.10727×10^{-2}
35	8	30	19.5795	2.04247	8.59541×10^{-3}
36	8	30	20.2877	2.02209	7.72012×10^{-3}
37	2	7	3.31764	1.60332	5.49957×10^{-2}
38	2	7	3.31764	1.60332	5.49957×10^{-2}
39	2	7	3.31764	1.60332	5.49957×10^{-2}
40	2	7	3.31764	1.60332	5.49957×10^{-2}
41	6	16	1.68581	1.90197	1.56548×10^{-2}
42	6	16	2.50403	1.81245	2.11244×10^{-2}
43	6	16	1.28414	1.98009	1.43057×10^{-2}
44	6	16	2.02671	1.84021	2.02094×10^{-2}
45	5	15	30.1168	0.87605	1.12108×10^{-1}
46	5	15	30.1168	0.87605	1.12108×10^{-1}
47	5	15	30.1168	0.87605	1.12108×10^{-1}
48	5	15	30.1168	0.87605	1.12108×10^{-1}
49	5	15	30.1168	0.87605	1.12108×10^{-1}
50	5	15	30.1168	0.87605	1.12108×10^{-1}
51	8	22	47.3067	0.81433	9.09956×10^{-2}
52	8	22	47.3067	0.81433	9.09956×10^{-2}
53	8	22	47.3067	0.81433	9.09956×10^{-2}
54	8	22	47.3067	0.81433	9.09956×10^{-2}
55	8	26	22.479	3.43752	6.94301×10^{-3}
56	8	26	22.479	3.43752	6.94301×10^{-3}

5.8.2 Experience with GRG Method

The application of the GRG technique to the dispatching problem using the GRG2 software proved the robustness of the method. Table (5-2) lists the optimal schedules for a selection of system power demands. The first column in Table (5-2) gives unit index, and subsequent columns list the optimal power generation of each unit corresponding to system power demand between 700 and 1400 MW. Row 57 of Table (5-2) gives the minimum operating cost denoted by F for each power demand, and the last row gives the number of iterations (ITER) required to arrive at the optimal strategy using the GRG Method.

It is important to examine the optimization results to detect patterns of change and to verify that the strategies obtained conform to rational expectations based on practical considerations. From inspection of the results in Table (5-2), the following are some interesting observations:

1- Units 10 & 11, are more expensive units and the optimal strategy requires that their output remain at the lower bound of 8 MW up to and including a power demand of 1300 MW. At 1400 MW demand, the two units output is the maximum allowed of 12 MW.

2- Unit 15, is required to stay at the minimum output level of 10 MW up to 1200 MW, then the GRG optimization requires an optimum generation of 10.46 MW, but then at 1300 MW, the output is required to drop back to 10 MW. At 1400 MW the required output is 18 MW. From a practical point of view, one would expect the optimal unit generation to increase (or stay the same) with an increase in power system demand.

3- Units 29 & 30, are least expensive to operate and therefore the optimization outcome requires that their output remain at the maximum allowed of 65 MW, as expected from a practical point of view.

4- Units 31 & 32, are identical and it is expected that their optimal generation be the same. However, the optimal values returned from GRG 2 show that at a power demand of 800 MW, the output of unit 31 is required to be 44.61 MW, whereas that of unit 32 is 44.03 MW.

5- Units 37-40, are identical and the GRG 2 results require identical outputs as expected.

6- Units 45-50, are identical but at 800 MW the GRG 2 results for units 45-47 are slightly less than that for units 48-50. The optimal generations, however, increase with an increase in power demand as expected.

7- Units 51-54, are identical. The schedule at 1000 MW demand is higher for unit 51 at 8.13 MW than these for subsequent units. Similar observations can be made about higher demands.

The preceding observations lead us to conclude that deviations from expected practical results can take place when using GRG 2. A possible explanation is that the search mechanism in the approach attempts to carry out the optimization by sequentially increasing the values of the decision variables until a minimum is reached. No attempt is made to equalize outputs that should be equal, since these constraints are not specified in the problem formulation.

A typical execution time of 17s on a DEC Micro-VAX-II was obtained. In the implementation of the method, the Hessian was found to be too large for use of the variable metric method and the algorithm switched to the conjugate gradient method using Fletcher-Reeve's direction. Termination of the computation was based on the determination that the fractional change in the objective function was less than 10^{-4}.

The algorithm displayed a sensitivity to the initial guess values in terms of convergence behaviour as discussed next.

Table (5-2)

Optimal Schedule for 56 Unit System
Using the GRG Method

UNIT	700	800	900	1000	1100	1200	1300	1400
1	8	8	8	8	8	8	8	28
2	18	18	18	18	18	18	18	18
3	18	18	18	18	18	18	18	18
4	18	18	18	18	18	18	18	18
5	18	18	18	18	18	18	18	18
6	8	8	8	8	8	8	8	28
7	10.24	12	12	12	12	12	12	12
8	9.07	12	12	12	12	12	12	12
9	5.57	12	12	12	12	12	12	12
10	8.00	8	8	8	8	8	8	12
11	8.00	8	8	8	8	8	8	12
12	11.12	15	15	15	15	15	15	15
13	10	10	10	10	10	10.01	10	18
14	10	10	10	10	10	10.31	10	18
15	10	10	10	10	10	10.46	10	18
16	10	10	12	12	12	12.00	12	12
17	10	10	10	10	12	12.00	12	12
18	10	10	10	10	10.06	10.9	10	18
19	10	10	10	10	10	10.7	10	18
20	8	8	8	8	9.84	13.43	18	18
21	20	20	20.71	30.84	43.74	60	59	60
22	20	20	20.71	30.84	35	35	35	35
23	20	20	20.71	30.84	43.74	60	59	60
24	20	20	20.71	30.84	35	35	35	35
25	20	20	22	22	22	22	22	22
26	20	20	22	22	22	22	22	22
27	20	29.55	36.46	45.52	56.55	60	60	60
28	20	36.42	40.04	48.75	57.49	73.35	75	75
29	65	65	65	65	65	65	65	65
30	65	65	65	65	65	65	65	65
31	25	44.61	75	75	75	75	75	75
32	25	44.03	75	75	75	75	75	75
33	8	8.09	12.11	17.74	23.91	30	30	30
34	8	9.20	12.15	16.01	20.09	30	30	30
35	8	8.0	9.70	14.74	20.19	30	30	30
36	8	8.09	12.11	17.74	23.91	30	30	30
37	2	4.97	5.59	6.28	7	7	7	7
38	2	4.97	5.59	6.28	7	7	7	7
39	2	4.97	5.59	6.28	7	7	7	7
40	2	4.98	5.59	6.28	7	7	7	7
41	6	6	6	12.49	16	16	16	16
42	6	7.85	9.46	11.35	13.62	15.79	16	16
43	6	6	6	10.96	15.03	16	16	16

Table (5-2) (Cont'd)

Optimal Schedule for 56 Unit System
Using the GRG Method

UNIT	700	800	900	1000	1100	1200	1300	1400
44	6	7.44	8.62	11.18	13.55	15.77	16	16
45	5	5.63	5.86	6.32	6.78	7.26	15	15
46	5	5.63	5.87	6.32	6.78	7.26	15	15
47	5	5.63	5.87	6.32	6.78	7.26	15	15
48	5	5.64	5.86	6.32	6.78	7.26	15	15
49	5	5.67	5.85	6.32	6.78	7.26	15	15
50	5	5.62	5.84	6.32	6.78	7.26	15	15
51	8	8	8	8.13	8.69	9.23	22	18
52	8	8	8	8	8.69	9.23	22	18
53	8	8	8	8	8.69	9.23	22	22
54	8	8	8	8	8.53	9.01	22	22
55	8	8	8	8	8	8	8	26
56	8	8	8	8	8	8	8	8
F Dh/hr	36,089.07	39,336.	42,817.	46,380.	50,092.	53,995.6	59,560.	64,219.6
ITER	33	32	30	31	14	30	16	29

Initial Guess Effects

The sensitivity of the convergence pattern of the GRG algorithm to initial guesses of the unknowns was investigated. The initial values of power generated were required to be feasible. Three sets of initial values were chosen. In the first the generation was set at the lower bound (lower guess) of the unit. The second set assumes that units are operating at a level midway between the lower and upper bounds (Midway guess). In the third the generation was set at the upper bound (upper guess) of the unit. A comparison of the performance of the algorithm in terms of the final value of the objective function, using the three sets, was conducted for a sequence of demand values as shown in Table (5-3).

It is clear from inspecting Table (5-3) that the midway guess is by far superior at all demand levels. As a general conclusion, in terms of speed of convergence, on computational experience one shows that the Midway guess results in faster convergence for demand values at 800 MW and higher. At low demand level the lower bound guess provided for faster convergence. It should be noted, however, that for a high demand such as 1000 MW, the algorithm failed to converge with the lower bound guess.

An alternative means of generating a reliable, feasible initial guess has been devised in the course of this investigation. Here one uses the value of the optimal power generations obtained neglecting the inequality constraints. The value of the variables are adjusted to respect the upper and lower bounds on the generations. The resulting initial guess values constitute a feasible solution but do not satisfy the power balance equation. In our computational experiments this guess resulted in the following reductions in number of iterations to convergence relative to those for the midway guess.

800 MW Case: From 31 iterations to 13 iterations
1000 MW Case: From 31 iterations to 8 iterations
1200 MW Case: From 30 iterations to 21 iterations

All convergence criteria were set at 10^{-4} relative error in the objective function.

Table (5-3)

Effect of Initial Guess on Final Value of Objective Using the GRG Method

DEMAND	LOW Mcal/hr	MIDWAY Mcal/hr	HIGH Mcal/hr
700	2292.8	2292.8	2292.8
750	2393.7	2393.6	2393.7
800	2499.6	2499.6	2499.6
850	2638.7 (*)	2608.2	2608.2
900	2745.6 (*)	2718.8	2804.2 (*)
950	2862.2 (*)	2831.5	2872.6 (*)
1000	3056.4 (*)	2944.4	3141.5 (*)
1100	3280.2 (*)	3178.9	3291.7 (*)
1200	3499.8 (*)	3425.8	3425.8
1300	3753.4 (*)	3708.9	3737.2
1400	4075.4 (*)	4065.8	4065.8

(*) indicates final convergence not attained.

5.8.3 Experience with Lambda Iteration Method

The application of the Lambda Iterations (LI) technique to the dispatching problem proved the robustness of the method. The code for the LI algorithm was provided by B. Wollenberg as share-ware for the IBM personal computers and was subsequently adapted to the DEC Micro-VAX II. Table (5-4) lists the optimal schedules for a selection of system power demands using the Lambda iteration algorithm with stopping criterion set at 0.001. A typical execution time of 2s on a DEC Micro-VAX II was obtained.

Entries in Table (5-4) are the same as those used in Table (5-2). It must be noted that the results obtained using the LI method conformed with the following two practical considerations:

1- The optimal generation of identical units must be identical.

2- For an increasing power demand, the optimal generations for each unit must be at least non-decreasing.

This is not surprising, since the program's formulation recognizes these two requirements. The algorithm appears to display sensitivity to the values chosen for the initial guess on λ. This aspect is discussed next.

Table (5-4)

Optimal Schedule for 56 Unit System
Using Lambda Iterations Method

UNIT	700	800	900	1000	1100	1200	1300	1400
1	8	8	8	8	8	8	8	28
2	18	18	18	18	18	18	18	18
3	18	18	18	18	18	18	18	18
4	18	18	18	18	18	18	18	18
5	18	18	18	18	18	18	18	18
6	8	8	8	8	8	8	8	28
7	10.21	12	12	12	12	12	12	12
8	9.1	12	12	12	12	12	12	12
9	5.7	12	12	12	12	12	12	12
10	8	8	8	8	8	8	8	12
11	8	8	8	8	8	8	8	12
12	11.1	15	15	15	15	15	15	15
13	10	10	10	10	10	10.2	14.8	18
14	10	10	10	10	10	10.6	13.1	18
15	10	10	10	10	10	11.3	16.4	18
16	10	10	12	12	12	12	12	12
17	10	10	10	10	12	12.00	12	12
18	10	10	10	10	10.1	11.6	16.2	18
19	10	10	10	10	10	11.7	18	18
20	8	8	8	8	9.9	13.2	18	18
21	20	20	21.2	30.8	44.3	60	60	60
22	20	20	21.2	30.8	35	35	35	35
23	20	20	21.2	30.8	44.3	60	60	60
24	20	20	21.2	30.8	35	35	35	35
25	20	20	21.2	22	22	22	22	22
26	20	20	21.2	22	22	22	22	22
27	20	28.3	37.4	45.4	56.55	60	60	60
28	20	36.4	43	48.8	56.9	70.1	75	75
29	65	65	65	65	65	65	65	65
30	65	65	65	65	65	65	65	65
31	25	44.5	67	75	75	75	75	75
32	25	44.5	67	75	75	75	75	75
33	8	8	12.9	17.5	23.9	30	30	30
34	8	9.1	12.7	15 9	20.3	27.6	30	30
35	8	8.0	10.4	14.5	20.3	29.7	30	30
36	8	8	12.9	17.5	23.9	30	30	30
37	2	4.9	5.6	6.3	7	7	7	7
38	2	4.9	5.6	6.3	7	7	7	7
39	2	4.9	5.6	6.3	7	7	7	7
40	2	4.9	5.69	6.3	7	7	7	7

Table (5-4) (Cont'd)

Optimal Schedule for 56 Unit System
Using Lambda Iterations Method

UNIT	700	800	900	1000	1100	1200	1300	1400
41	6	7.6	10.2	12.5	15.6	16	16	16
42	6	7.8	9.7	11.4	13.7	16	16	16
43	6	6	8.4	10.9	14.4	16	16	16
44	6	7.5	9 4	11.2	13.6	16	16	16
45	5	5.6	6	6.3	6.88	7.5	9.7	13.8
46	5	5.6	6	6.3	6.8	7.5	9.7	13.8
47	5	5.6	6	6.3	6.8	7.5	9.7	13.8
48	5	5.6	6	6.3	6.8	7.5	9.7	13.8
49	5	5.6	6	6.3	6.8	7.5	9.7	13.8
50	5	5.6	6	6.3	6.8	7.5	9.7	13.8
51	8	8	8	8.1	8.7	9.6	12.3	17.4
52	8	8	8	8.1	8.7	9.6	12.3	17.4
53	8	8	8	8	8.7	9.6	12.3	17.4
54	8	8	8	8	8.7	9.6	12.3	17.4
55	8	8	8	8	8	8	8	26
56	8	8	8	8	8	8	8	26
F Dh/hr	36,084.5	39,330.4	42,798.0	46,374.7	50,085.4	53,983.2	58,428.9	64,047.3
ITER	23	26	40	24	40	23	22	40

Initial Guess Effects

In the GRG method an initial guess on all problem variables (56 power generation levels) is required. In the LI method, however, one is required to start with an initial guess of the incremental cost of power production lambda only. The choice of lambda influences the convergence pattern of the algorithm.

The sensitivity of the convergence behaviour of the lambda iterations algorithm to initial guess of the unknown lambda was investigated. The initial values of power generated were required to be feasible. This results in a number of feasible ranges of lambda corresponding to the feasible ranges of the system units. Recall that for the quadratic cost case, the individual optimum power generations are obtained by (5-11) as:

$$P(i) = \frac{\lambda - a_1(i)}{2a_2(i)}$$

As a result the value of lambda is given in terms of the power generation as:

$$\lambda = a_1(i) + 2a_2(i)P(i)$$

Corresponding to the lower and upper bounds on each P(i), one obtains lower and upper bounds on lambda as follows:

$$\lambda_{min}(i) = a_1(i) + 2a_2(i)P_{min}(i)$$

$$\lambda_{max}(i) = a_1(i) + 2a_2(i)P_{max}(i)$$

Ideally a feasible range of lambda can be obtained by finding:

$$\lambda_{inf} = \underset{i}{Max}[\lambda_{min}(i)]$$

$$\lambda_{sup} = \underset{i}{Min}[\lambda_{max}(i)]$$

In this case the feasible range of lambda is defined by:

$$\lambda_{inf} \leq \lambda \leq \lambda_{sup}$$

One can further define the lowest value of lambda as λ_L and the highest value of Lambda as λ_H. Here one has:

$$\lambda_L = \underset{i}{Min}[\lambda_{min}(i)]$$

$$\lambda_H = \underset{i}{Max}[\lambda_{max}(i)]$$

For the system studied the computations yield the following:

$$\lambda_L = 11.83458$$

λ_H = 76.46383

λ_{inf} = 56.31642

λ_{sup} = 15.38143

Note that λ_{inf} is greater than λ_{sup}. If one defines a midway λ by $\lambda_{mid} = [\lambda_{inf} + \lambda_{sup}]/2$, the following value is obtained.

λ_{mid} = 35.84892

These values are good candidates for an initial guess on λ.

One can also use the value of λ computed from the first step of a lambda aggregation computation for the power demand considered using eq. (5-10). It is clear that there are six options for an initial guess. Results of testing are summarized in Table (5-5) for system demands of 800, 1000, and 1200 respectively.

Table (5-5)
Effect of Initial Guess on Number of Iterations Using the LI Method

DEMAND	LOW	SUP	HIGH	INF	LA	MID
800	27	20	27	23	25	26
1000	25	24	25	22	24	24
1200	24	21	24	23	23	23

From the table, one can see that the method is not very sensitive to the initial guess. In terms of preference of possible start, the supremum value of λ appears to give lowest number of iterations overall.

5.8.1 Experience With Lambda Aggregation Method

The coding of the Lambda Aggregation technique was tested for the system considered. Table (5-6) shows the optimal schedules for the system for load levels from 700 MW to 1400 MW, using the Lambda Aggregation algorithm, with stopping criterion set at 0.001. The algorithm converged in 3 iterations for load values of 700 and 1000 MW. Four iterations were required at 1400 MW and 5 iterations for 1100 MW and 1200 MW loads. It is clear that the algorithm correctly schedules identical power generations for identical units as expected. It is noted that other software tested for the same system sometimes failed to produce this result. In terms of computational time, the algorithm typically required 1 second to obtain a solution as compared to 17 seconds required to obtain a similar solution using the Reduced Gradient Method for optimization.

Consider for example unit 50, the optimal active power generation at the low power demand of 700 MW is set by the strategy at the minimum value of 5 MW. The optimal active power generation of the unit increases as the power demand is increased up to a demand of 1300 MW where the optimal generation is set at its maximum allowed of 15 MW.

It has been noted that the system considered contains a number of identical units and one would expect that the optimal generations of such units be identical. The GRG2 code failed to produce this result, whereas LI and LA software produced this expected result, as can be seen from inspection of Table (5-6).

The algorithm failed to converge (oscillated), for test cases involving negative values of the cost parameter a_2. Note that this case corresponds to a non-convex objective and the Hessian of the problem is not positive definite.

It should be noted that a faster implementation of the present algorithm can be affected by simply using equivalents for identical units. In the present application this was not attempted as the resulting solution was obtained in a very fast time.

Table (5-6)

Optimal Schedule for 56 Unit System
Using Lambda Aggregation Method

UNIT	700	800	900	1000	1100	1200	1300	1400
1	8	8	8	8	8	8	8	28
2	18	18	18	18	18	18	18	18
3	18	18	18	18	18	18	18	18
4	18	18	18	18	18	18	18	18
5	18	18	18	18	18	18	18	18
6	8	8	8	8	8	8	8	28
7	12	12	12	12	12	12	12	12
8	11	12	12	12	12	12	12	12
9	5	12	12	12	12	12	12	12
10	8	8	8	8	8	8	8	12
11	8	8	8	8	8	8	8	12
12	8	15	15	15	15	15	15	15
13	10	10	10	10	10	10	10	18
14	10	10	10	10	10	10	10	18
15	10	10	10	10	10	10	10	18
16	10	10	12	12	12	12	12	12
17	10	10	10	10	12	12	12	12
18	10	10	10	10	10	10	10	18
19	10	10	10	10	10	10	10	18
20	8	8	8	8	8	13.72	18	18
21	20	20	20	30.99	44.63	60	60	60
22	20	20	20	30.99	35	35	35	35
23	20	20	20	30.99	44.63	60	60	60
24	20	20	20	30.99	35	35	35	35
25	20	20	22	22	22	22	22	22
26	20	20	22	22	22	22	22	22
27	20	28.27	38.11	45.51	56.87	60	60	60
28	20	36.39	43.49	48.84	57.05	72.20	75	75
29	65	65	65	65	65	65	65	65
30	65	65	65	65	65	65	65	65
31	25	44.48	68.77	75	75	75	75	75
32	25	44.48	68.77	75	75	75	75	75
33	8	8	13.34	17.56	24.04	30	30	30
34	8	9.06	12.97	15.92	20.44	28.77	30	30
35	8	8	10.8	14.59	20.41	30	30	30
36	8	8	13.34	17.56	24.04	30	30	30
37	2	4.89	5.68	6.27	7	7	7	7
38	2	4.89	5.68	6.27	7	7	7	7
39	2	4.89	5.68	6.27	7	7	7	7
40	2	4.89	5.68	6.27	7	7	7	7

Table (5-6) (Cont'd)

Optimal Schedule for 56 Unit System
Using Lambda Aggregation Method

UNIT	700	800	900	1000	1100	1200	1300	1400
41	6	7.65	10.42	12.50	15.69	16	16	16
42	6	7.79	9.84	11.38	13.75	16	16	16
43	6	6	8.67	10.95	14.44	16	16	16
44	6	7.45	9.60	11.21	13.68	16	16	16
45	5	5.64	6.03	6.32	6.77	7.59	15	15
46	5	5.64	6.03	6.32	6.77	7.59	15	15
47	5	5.64	6.03	6.32	6.77	7.59	15	15
48	5	5.64	6.03	6.32	6.77	7.59	15	15
49	5	5.64	6.03	6.32	6.77	7.59	15	15
50	5	5.64	6.03	6.32	6.77	7.59	15	15
51	8	8	8	8	8.68	9.69	21.50	22
52	8	8	8	8	8.68	9.69	21.50	22
53	8	8	8	8	8.68	9.69	21.50	22
54	8	8	8	8	8.68	9.69	21.50	22
55	8	8	8	8	8	8	8	13
56	8	8	8	8	8	8	8	13
F Dh/hr	36,283.66	39,522.8	42,922.0	46,568.0	50,280.0	54,184.3	59,676.0	64,491.8
ITER	3	3	2	3	3	5	4	4

5.8.5 A Comparison of Methods

The performance comparison of the three methods tested shows that the Lambda Aggregation technique provides the fastest execution times to obtain power system optimal schedules. In the set of Tables (5-7-A,B,C and D) the optimal generation levels for demand values of 700 MW to 1400 MW are displayed for the three methods tested. In Table (5-8) the corresponding optimal objective functions are listed.

The Lambda Aggregation method does not require an initial guess. It failed to converge for cases with negative a_2. On the other hand the GRG method is sensitive to the selection of an initial guess. Midway values were effective, but a drastic improvement in GRG method's speed of convergence was obtained by the use of results of the first step of the Lambda Aggregation method. The Lambda Iteration method proved to be insensitive to the initial guess values.

The Lambda Iteration method consistently resulted in optimal schedules with the lowest attainable values. It should be noted, however, from Table (5-8), that the differences in values of optimum objective are less than 0.5%.

An interesting observation is that the GRG method sometimes resulted in optimal generations that were not identical for units with identical model parameters and upper and lower bounds. This behaviour was not found with either of the other methods. As stated earlier, it is the very nature of the GRG algorithm to adjust the unknown values in sequential manner until a minimum is reached. This is a possible explanation for the behaviour of the optimal solutions, since the units with lower index number showed a higher optimal generation than those identical units, but with a higher index number.

Table (5-7-A)

Comparison of Optimal Loadings
700 and 800 MW Demands

	700 MW			800 MW		
	GRG	LA	LI	GRG	LA	LI
1	8	8	8	8	8	8
2	18	18	18	18	18	18
3	18	18	18	18	18	18
4	18	18	18	18	18	18
5	18	18	18	18	18	18
6	8	8	8	8	8	8
7	10.24	12	10.2	12	12	12
8	9.07	11	9.1	12	12	12
9	5.57	5	5.7	12	12	12
10	8	8	8	8	8	8
11	8	8	8	8	8	8
12	11.12	8	11.5	15	15	15
13	10	10	10	10	10	10
14	10	10	10	10	10	10
15	10	10	10	10	10	10
16	10	10	10	10	10	10
17	10	10	10	10	10	10
18	10	10	10	10	10	10
19	10	10	10	10	10	10
20	8	8	8	8	8	8
21	20	20	20	20	20	20
22	20	20	20	20	20	20
23	20	20	20	20	20	20
24	20	20	20	20	20	20
25	20	20	20	20	20	20
26	20	20	20	20	20	20
27	20	20	20	29.55	28.27	28.3
28	20	20	20	36.42	36.39	36.4
29	65	65	65	65	65	65
30	65	65	65	65	65	65
31	25	25	25	44.61	44.48	44.5
32	25	25	25	44.03	44.48	44.5
33	8	8	8	8.09	8	8
34	8	8	8	9.20	9.06	9.1
35	8	8	8	8	8	8
36	8	8	8	8.09	8	8
37	2	2	2	4.97	4.89	4.9
38	2	2	2	4.97	4.89	4.9
39	2	2	2	4.97	4.89	4.9
40	2	2	2	4.98	4.89	4.9
41	6	6	6	6	7.65	7.6
42	6	6	6	7.85	7.79	7.8
43	6	6	6	6	6	6

Table (5-7-A) (Cont'd)
Comparison of Optimal Loadings
700 and 800 MW Demands

	700 MW			800 MW		
	GRG	LA	LI	GRG	LA	LI
44	6	6	6	7.44	7.45	7.5
45	5	5	5	5.63	5.64	5.6
46	5	5	5	5.63	5.64	5.6
47	5	5	5	5.63	5.64	5.6
48	5	5	5	5.64	5.64	5.6
49	5	5	5	5.67	5.64	5.6
50	5	5	5	5.62	5.64	5.6
51	8	8	8	8	8	8
52	8	8	8	8	8	8
53	8	8	8	8	8	8
54	8	8	8	8	8	8
55	8	8	8	8	8	8
56	8	8	8	8	8	8
F Dh/hr	36,089.07	36,283.66	36,084.5	39,336.2	39,522.8	39,330.4

Table (5-7-B)

Comparison of Optimal Loadings
900 and 1000 MW Demands

	900 MW			1000 MW		
	GRG	LA	LI	GRG	LA	LI
1	8	8	8	8	8	8
2	18	18	18	18	18	18
3	18	18	18	18	18	18
4	18	18	18	18	18	18
5	18	18	18	18	18	18
6	8	8	8	8	8	8
7	12	12	12	12	12	12
8	12	12	12	12	12	12
9	12	12	12	12	12	12
10	8	8	8	8	8	8
11	8	8	8	8	8	8
12	15	15	15	15	15	15
13	10	10	10	10	10	10
14	10	10	10	10	10	10
15	10	10	10	10	10	10
16	12	12	12	12	12	12

Table (5-7-B) (Cont'd)
Comparison of Optimal Loadings
900 and 1000 MW Demands

	900 MW			1000 MW		
	GRG	LA	LI	GRG	LA	LI
17	10	10	10	10	10	10.7
18	10	10	10	10	10	10
19	10	10	10	10	10	10
20	8	8	8	8	8	8
21	20.71	20	21.2	30.84	30.99	30.8
22	20.71	20	21.2	30.84	30.99	30.8
23	20.71	20	21.2	30.84	30.99	30.8
24	20.71	20	21.2	30.84	30.99	30.8
25	22	20	21.2	22	22	22
26	22	20	21.2	22	22	22
27	36.46	38.11	37.4	45.52	45.51	45.4
28	40.04	43.49	43	48.75	48.84	48.8
29	65	65	65	65	65	65
30	65	65	65	65	65	65
31	75	68.77	67	75	75	75
32	75	68.77	67	75	75	75
33	12.11	13.34	12.9	17.74	17.56	17.5
34	12.15	12.97	12.7	16.01	15.92	15.9
35	9.70	10.8	10.4	14.74	14.59	14.5
36	12.11	13.34	12.9	17.74	17.56	17.5
37	5.59	5.68	5.6	6.28	6.27	6.3
38	5.59	5.68	5.6	6.28	6.27	6.3
39	5.59	5.68	5.6	6.28	6.27	6.3
40	5.59	5.68	5.6	6.28	6.27	6.3
41	6	10.42	10.2	12.49	12.50	12.5
42	9.46	9.84	9.7	11.35	11.38	11.4
43	6	8.67	8.4	10.96	10.95	10.9
44	8.62	9.60	9.4	11.18	11.21	11.2
45	5.86	6.03	6	6.32	6.32	6.3
46	5.87	6.03	6	6.32	6.32	6.3
47	5.87	6.03	6	6.32	6.32	6.3
48	5.86	6.03	6	6.32	6.32	6.3
49	5.85	6.03	6	6.32	6.32	6.3
50	5.84	6.03	6	6.32	6.32	6.3
51	8	8	8	8.13	8	8.1
52	8	8	8	8	8	8.1
53	8	8	8	8	8	8
54	8	8	8	8	8	8
55	8	8	8	8	8	8
56	8	8	8	8	8	8
F Dh/hr	42,817.	42,992.	42,789.	46,380.	46,568.	46,374.7

Table (5-7-C)

Comparison of Optimal Loadings
1100 and 1200 MW Demands

	1100 MW			1200 MW		
	GRG	LA	LI	GRG	LA	LI
1	8	8	8	8	8	8
2	18	18	18	18	18	18
3	18	18	18	18	18	18
4	18	18	18	18	18	18
5	18	18	18	18	18	18
6	8	8	8	8	8	8
7	12	12	12	12	12	12
8	12	12	12	12	12	12
9	12	5	12	12	12	12
10	8	8	8	8	8	8
11	8	8	8	8	8	8
12	15	15	15	15	15	15
13	10	10	10	10.01	10	10.2
14	10	10	10	10.31	10	10.6
15	10	10	10	10.46	10	11.3
16	12	12	12	12	12	12
17	12	12	12	12	12	12
18	10.06	10	10.1	10.9	10	11.6
19	10	10	10	10.7	10	11.7
20	9.84	8	9.9	13.43	13.72	13.2
21	43.74	44.63	44.3	60	60	60
22	35	35	35	35	35	35
23	43.74	44.63	44.3	60	60	60
24	35	35	35	35	35	35
25	22	22	22	22	22	22
26	22	22	22	22	22	22
27	56.55	56.87	56.6	60	60	60
28	57.49	57.05	56.9	73.35	72.20	70.1
29	65	65	65	65	65	65
30	65	65	65	65	65	65
31	75	75	75	75	75	75
32	75	75	75	75	75	75
33	23.91	24.04	23.9	30	30	30
34	20.09	20.44	20.3	30	28.77	27.6
35	20.19	20.41	20.3	30	30	29.7
36	23.91	24.04	23.9	30	30	30
37	7	7	7	7	7	7
38	7	7	7	7	7	7
39	7	7	7	7	7	7
40	7	7	7	7	7	7
41	16	15.69	15.6	16	16	16
42	13.62	13.75	13.7	15.79	16	16
43	15.03	14.44	14.4	16	16	16

Table (5-7-C) (Cont'd)
Comparison of Optimal Loadings
1100 and 1200 MW Demands

	1100 MW			1200 MW		
	GRG	LA	LI	GRG	LA	LI
44	13.55	13.68	13.6	15.77	16	16
45	6.78	6.77	6.8	7.26	7.59	7.5
46	6.78	6.77	6.8	7.26	7.59	7.5
47	6.78	6.77	6.8	7.26	7.59	7.5
48	6.78	6.77	6.8	7.26	7.59	7.5
49	6.78	6.77	6.8	7.26	7.59	7.5
50	6.78	6.77	6.8	7.26	7.59	7.5
51	8.69	8.68	8.7	9.23	9.69	9.6
52	8.69	8.68	8.7	9.23	9.69	9.6
53	8.69	8.68	8.7	9.23	9.69	9.6
54	8.53	8.68	8.7	9.01	9.69	9.6
55	8	8	8	8	8	8
56	8	8	8	8	8	8
F Dh/hr	50,092.	50,280.	50,085.4	53,995.6	54,184.3	53,983.2

Table (5-7-D)

Comparison of Optimal Loadings
1300 and 1400 MW Demands

	1300 MW			1400 MW		
	GRG	LA	LI	GRG	LA	LI
1	8	8	28	28	28	28
2	18	18	18	18	18	18
3	18	18	18	18	18	18
4	18	18	18	18	18	18
5	18	18	18	18	18	18
6	8	8	28	28	28	28
7	12	12	12	12	12	12
8	12	12	12	12	12	12
9	12	12	12	12	12	12
10	8	8	8	12	12	12
11	8	8	8	12	12	12
12	15	15	15	15	15	15
13	10	10	14.8	18	18	18
14	10	10	13.1	18	18	17.7
15	10	10	16.4	18	18	18
16	12	12	12	12	12	12

Table (5-7-D) (Cont'd)
Comparison of Optimal Loadings
1300 and 1400 MW Demands

	1300 MW			1400 MW		
	GRG	LA	LI	GRG	LA	LI
17	12	12	12	12	12	12
18	10	10	16.2	18	18	18
19	10	10	18	18	18	18
20	18	18	18	18	18	18
21	59	60	60	60	60	60
22	35	35	35	35	35	35
23	59	60	60	60	60	60
24	35	35	35	35	35	35
25	22	22	22	22	22	22
26	22	22	22	22	22	22
27	60	60	60	60	60	60
28	75	75	75	75	75	75
29	65	65	65	65	65	65
30	65	65	65	65	65	65
31	75	75	75	75	75	75
32	75	75	75	75	75	75
33	30	30	30	30	30	30
34	30	30	30	30	30	30
35	30	30	30	30	30	30
36	30	30	30	30	30	30
37	7	7	7	7	7	7
38	7	7	7	7	7	7
39	7	7	7	7	7	7
40	7	7	7	7	7	7
41	16	16	16	16	16	16
42	16	16	16	16	16	16
43	16	16	16	16	16	16
44	16	16	16	16	16	16
45	15	15	9.7	15	15	13.8
46	15	15	9.7	15	15	13.8
47	15	15	9.7	15	15	13.8
48	15	15	9.7	15	15	13.8
49	15	15	9.7	15	15	13.8
50	15	15	9.7	15	15	13.8
51	18	21.50	12.3	18	22	17.4
52	18	21.50	12.3	18	22	17.4
53	22	21.50	12.3	22	22	17.4
54	22	21.50	12.3	22	22	17.4
55	8	8	8	26	13	26
56	8	8	8	8	13	26
F Dh/hr	59,560.	59,676.	58,428.92	64,219.6	64,491.85	64,047.

Table (5-8)

Comparison of Minimum Costs for Methods Tested

METHOD	700 MW	800 MW	900 MW	1000 MW	1100 MW	1200 MW	1300 MW	1400 MW
GRG	36,089.1	39,336.2	42,817	46,380	50,092	43,995.6	59,560	64,219.6
LI	36,084.5	39,330.4	42,798.5	46,374.7	50,085.4	53,983.2	58,428.9	64.047
LA	36,283.6	39,522.8	42,922	46,568	50,280	54,184.3	59,676	64,491.8

5.9 Summary

In this chapter, three efficient algorithms for optimal economic dispatching of electric power systems were discussed. The problem treated involves a system with thermal generations, and the object is to determine the optimum loading to satisfy a given load requirement and observe system and equipment constraints.

As a result of an extensive study of available optimization techniques, the three methods emerged as most practical tools to carry out the dispatching function. The methods discussed in this paper are the Generalized Reduced Gradient (GRG) algorithm, the Lambda Iteration (LI) algorithm, and the Lambda Aggregation (LA) algorithm. The latter is believed to be a **NEW** approach to the problem and appears to have many advantages.

In this chapter, details of the algorithm were discussed and the computational results corresponding to a large size system with 56 units were summarized. Guidelines for selecting initial values for the GRG and LI algorithms were discussed. The experience with other available methods shows the techniques discussed to be extremely fast and robust. A comparison of execution times of the three techniques on a DEC micro VAX-II shows that the Lambda Aggregation method is extremely fast. A comparison of the optimal schedules and optimal objective obtained using the three methods shows reasonably close agreement.

The advantages of the newly developed Lambda Aggregation technique are its speed and reasonably competitive results. Another attractive feature of the technique is that it proved to be an efficient means for providing initial guess values for the GRG method to achieve fast convergence. The technique has been observed to fail when the Hessian has negative entries.

CHAPTER 6
NON-LEAST SQUARES PARAMETER ESTIMATES AND EFFECTS ON ECONOMIC DISPATCH

6.1 Introduction

The literature on regression analysis [148,149] indicates that Least Squares parameter estimators may not be appropriate to use with data whose measurement errors are not normally distributed. Alternative parameter estimation methods are available to deal with such cases. A generally accepted term for the process is that of robust regression (or parameter estimation).

The object of this chapter is to explore the application of robust parameter estimation methodology to fuel cost model parameter estimation for some units in the system. In this chapter, the Least Absolute Residual Estimation (LARE) or L_1-norm approximation estimators and the Iteratively Reweighted Least Squares (IRLS) estimators are discussed. A selection of system units is considered, and the estimation task is carried out using these techniques. The computational results are compared in terms of the effect of the model parameters on the outcome of the economic dispatch computation.

6.2 Least Absolute Residual Estimators

Least Squares parameter estimators work well when the measurement errors are normally distributed and in this case the maximum likelihood and Least Squares estimators are identical. When the observations follow a non-normal distribution, the method of least squares may not be appropriate [148].

A number of authors have proposed robust regression procedures designed to dampen the effect of measurements that would be highly influential if least squares were used. It can be shown that if the measurement errors follow a double exponential distribution, then the maximum likelihood estimate is equivalent to the least absolute residual estimator (LARE) whose objective is to minimize

$$J_{LAR} = \sum_{i=0}^{L} \left|\tilde{\gamma}\,(k-i)\right| \qquad (6\text{-}1)$$

The objective is essentially an L_1-norm of the measurement error vector where (L+1) is the number of available observations.

The L_1-norm parameter estimator can be formulated as a linear programming problem. If one lets ϵ_{+i} and ϵ_{-i} denote the positive and negative deviations about the estimated model, then the problem reduces to minimizing.

$$J_{LAR} = \sum_{i=0}^{L} (\epsilon_{+i} + \epsilon_{-i}) \qquad (6\text{-}2)$$

subject to

$$z(k-i) - \underline{h}^T(k-i)\underline{X} + \epsilon_{+i} - \epsilon_{-i} = 0 \quad i=0,\ldots,L \qquad (6\text{-}3)$$

with elements of $\underline{X}$ unrestricted in sign. The LP problem has (L+1) constraints [one for each observation] and [n+2(L+1)] variables (one variable for each model parameter and 2(L+1) variable representing the deviations).

6.3 Robust Parameter Estimation

The basis for the iteratively reweighted least squares estimator is the requirement of minimizing a function ρ of the individual measurement errors $\tilde{z}$. This type of estimator is called a robust estimator or an M-estimator, where M stands for maximum likelihood. The function ρ is related to the likelihood function for an appropriate choice of the error distribution. In the method of least squares (implying that the error distribution is normal) the function ρ is given by [149]

$$\rho(r) = \frac{1}{2}r^2 \qquad -\infty < r < \infty \qquad (6\text{-}4)$$

The estimation problem can be formulated as requiring minimizing the following robust objective criterion denoted by J_R over the components X_j of the nx1 unknown parameter vector x[149]

$$J_R(k) = \sum_{i=0}^{L} \rho(\tilde{z}(k-i)/s) \qquad (6\text{-}5)$$

The measurement error scale factor s is required since the M-estimator is not necessarily scale-invariant (if the measurement errors are multiplied by a constant, the new estimate may not be the same as the old one). Using the notation of chapter 4, the measurement error is defined by

$$\tilde{z}(k-i) = z(k-i) - h^T(k-i)\,\underline{X} \qquad (6\text{-}6)$$

There are (L+1) measurements available, and the overall measurement model of eq. (4.7) applies.

A robust estimate of X, denoted by $\hat{X}$ is obtained by setting the derivative of $J_R(k)$ with respect to the n components X_i to zero and solving the resulting set of n equations. To simplify the notation, one substitutes

$$r_i = \tilde{z}(k-i)/s \tag{6-7}$$

As a result the minimization of $J_R(k)$ requires solving

$$\frac{\partial J}{\partial x_j} = \sum_{i=0}^{L} \frac{\partial \rho}{\partial r_i} \frac{\partial r_i}{\partial x_j} = 0 \quad j = 1,\dots,n \tag{6-8}$$

Now $sr_i = z(k-i) - \sum_{j=0}^{n} h_j\ (k-i)\ x_j$ (6-9)

$$\frac{\partial r_i}{\partial x_j} = -h_j(k-i)/s \tag{6-10}$$

Therefore the M-estimation problem requires solving

$$\sum_{i=0}^{L} h_j(k-i) \frac{\partial \rho}{\partial r_i} = 0 \qquad j=1,\dots,n \tag{6-11}$$

It is customary to replace the partial derivative by the function $\Psi(r)$, called the influence function defined by

$$\Psi(r) = \frac{\partial \rho}{\partial r} \tag{6-12}$$

As a result, we need to solve

$$\Phi_j = \sum_{i=0}^{L} h_j(k-i)\Psi(r_i) = 0 \qquad j=1,\dots,n \tag{6-13}$$

in the unknown parameter $\hat{X}_j$ for j = 1,...,n. This is a set of nonlinear equations which require an iterative algorithm for its solution.

6.4 Iterative Solution of Robust Estimator Equations

A popular iterative technique for solving nonlinear equations is that of Newton's method [132]. Here one requires solving the set of equations.

$$\Phi_j(\hat{X}) = 0 \quad j=1,\dots,n. \tag{6-14}$$

Starting with an initial guess $\hat{X}_0$, one finds an improved solution $\hat{X}_1$ such that

$$\hat{X}_1 = \hat{X}_0 + \Delta \tag{6-15}$$

The improved estimate $\hat{X}_1$ is obtained as

$$\hat{X}_1 = \hat{X}_0 - J^{-1}\ \Phi\ (\hat{X}_0) \tag{6-16}$$

The Jacobian matrix $\underline{J}$ is nxn and is made of the partial derivatives of Φ with respect to x.

$$J = \begin{bmatrix} \frac{\partial\Phi_1}{\partial x_1} & \frac{\partial\Phi_1}{\partial x_2} & \cdots & \frac{\partial\Phi_1}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial\Phi_n}{\partial x_1} & \frac{\partial\Phi_n}{\partial x_2} & \cdots & \frac{\partial\Phi_n}{\partial x_n} \end{bmatrix} \tag{6-17}$$

In the application to the robust regression problem, one has:

$$\underline{\Phi} = \underline{H}^T \underline{\Psi} \tag{6-18}$$

Here Ψ is an (L+1)x1 vector defined by

$$\underline{\Psi} = [\Psi(r_0)\ \Psi(r_1)\ \ldots\ \Psi(r_L)]^T \tag{6-19}$$

The matrix H is as defined in Chapter 4. The elements of the Jacobian matrix are obtained from

$$\Phi_j = \sum_{i=0}^{L} h_j\ (k-i)\ \Psi(r_i) = 0 \qquad m=1,\ldots\ldots,n \tag{6-20}$$

$$\frac{\partial\Phi_j}{\partial x_m} = \sum_{i=0}^{L} h_j(k-i)\ \frac{\partial\Psi}{\partial r_i}\ \frac{\partial r_i}{\partial x_m} \tag{6-21}$$

Recalling that

$$\frac{\partial r_i}{\partial x_m} = -\frac{h_m(k-i)}{s} \tag{6-22}$$

One concludes that

$$\frac{\partial\Phi_j}{\partial x_m} = \frac{-1}{s}\sum_{i=0}^{L} h_j(k-i)h_m\ (k-i)\frac{\partial\Psi}{\partial r_i} \tag{6-23}$$

The Jacobian can be written in the compact vector-matrix notation as:

$$J = -\frac{1}{s}\underline{H}^T\ \underline{\Gamma}\ \underline{H} \tag{6-24}$$

Here one introduces the diagonal (L+1)x(L+1) matrix Γ defined by

$$\underline{\Gamma} = \begin{bmatrix} \frac{\partial\Psi}{\partial r_0} & 0 & 0 & \ldots & 0 \\ 0 & \frac{\partial\Psi}{\partial r_1} & 0 & \ldots & 0 \\ 0 & 0 & 0 & & 0 \\ 0 & 0 & 0 & & 0 \\ 0 & 0 & 0 & & 0 \\ 0 & 0 & 0 & \ldots & \frac{\partial\Psi}{\partial r_L} \end{bmatrix} \tag{6-25}$$

Newton's iterations for the robust regression problem can thus be written as:

$$\underline{X}_1 = \underline{X}_0 + s[\underline{H}^T\ \underline{\Gamma}\ \underline{H}]^{-1}\ \underline{H}^T\ \underline{\Psi} \tag{6-26}$$

Newton's method is difficult to implement since it requires the derivatives of the influence function Ψ and also because the matrix $\underline{H}^T \underline{\Gamma} \underline{H}$ may be negative definite.

An approximation discussed by Huber [150] and Bickel [151] assumes that $\underline{\Gamma}$ is a unity matrix and iterates according to the formula

$$\underline{X}_1 = \underline{X}_0 + s[\underline{H}^T \underline{H}]^{-1} \underline{H}^T \underline{\Psi} \qquad (6\text{-}27)$$

The Huber iterations are attractive since the inverse needs to be computed once, but it requires more iterations and is not as easy to use with existing least-squares regression packages.

An attractive iterations alternative is provided by the technique commonly known as the Iteratively Reweighted Least Squares estimator (IRLS), which is discussed in the next section.

6.5 Iteratively Reweighted Least Squares Estimation

In the Iteratively Reweighted Least Squares (IRLS) method a weighting function $w(r)$ is defined in terms of the influence function $\Psi(r)$ by

$$w(r) = \Psi(r)/r \qquad r \neq 0$$
$$= 1 \qquad r = 0 \qquad (6\text{-}28)$$

As a result, one has

$$\frac{\partial \Psi}{\partial r} = w(r) + r\,\frac{\partial w}{\partial r} \qquad (6\text{-}29)$$

As an approximation one obtains assuming that the second term is negligible

$$\frac{\partial \Psi}{\partial r} \cong w(r) \qquad (6\text{-}30)$$

This expression is exact for either $r=0$ or w being a constant.

As a consequence of the introduction of the weighting function approximation one has

$$\underline{\Gamma} = \underline{W}_0 \qquad (6\text{-}31)$$

Where $\underline{W}_0$ is a diagonal (L+1)x(L+1) matrix with elements $w(r_i)$. As a result, one writes [149]

$$\underline{\Psi} = \underline{W}_0\, \underline{\tilde{Z}}/s \qquad (6\text{-}32)$$

The iteration therefore progress according to

$$\underline{\hat{X}}_1 = \underline{\hat{X}}_0 + [\underline{H}^T \underline{W}_0 \underline{H}]^{-1} \underline{H}^T \underline{W}_0 [\underline{Z} - \underline{H} \underline{\hat{X}}_0] \quad (6\text{-}33)$$

It is clear that this iteration is equivalent to directly finding $\underline{X}_1$ using:

$$\underline{X}_1 = [\underline{H}^T \underline{W}_0 \underline{H}]^{-1} \underline{H}^T \underline{W}_0 \underline{Z} \quad (6\text{-}34)$$

This can be verified by expanding equation (6-33) to obtain equation (6-34), which is simply the solution to the normal equations with weighting matrix W_0 that depends on the measurement errors.

The IRLS method only requires knowing how to compute the weight function w(r) and then it is possible to use an existing weighted least squares algorithm or compute the square root of $\underline{W}(r)$, from the matrix $W^{1/2}H$ and $W^{1/2}Z$ and use a standard least squares program for each step. A fundamental issue in IRLS estimation is that of choosing the weight functions w(r) which is discussed in the next section.

6.6 The Weighting Functions

A number of robust criterion functions ρ have been proposed and robust estimation procedures can be classified by the behaviour of their influence function ψ. The influence function controls the weight given to each measurement error. It should be noted that the least squares method has

$$\rho(r) = \frac{1}{2}r^2 \quad (6\text{-}35)$$

and $\psi(r) = r$ (6-36)

Note that the ψ function in this case is unbounded and thus the least squares procedure becomes non-robust when used with data arising from a heavy tailed distribution.

Holland and Welsch [149] report on eight weight functions, chosen after some years of experience with robust estimation. The functions are designed so as not to weigh large measurement errors as heavily as least squares. Holland and Welsch give nominal values for tuning constants associated with each function to guide in the selection process. The tuning constant is used to define the region for which lower weighting of the errors takes place. The IRLS method is implemented in ROSEPACK [152] using the eight weighting functions.

We now give the definitions of the eight weighting functions and associated ρ and Ψ functions for the benefit of the power systems engineer.

1- Andrews Weighting Function

The Andrews weighting function is defined by the criterion function [153].

$$\rho(r) = A^2[1-\cos(r/A)] \qquad |r| \leq \pi A$$
$$= 2A^2 \qquad |r| > \pi A \qquad (6\text{-}37)$$

The associated influence function $\Psi(r)$ is given by the derivative of $\rho(r)$ as

$$\Psi(r) = A\ \sin(r/A) \qquad |r| \leq \pi A$$
$$= 0 \qquad |r| > \pi A \qquad (6\text{-}38)$$

The weight function therefore is given by

$$w(r) = (\tfrac{r}{A})^{-1}\ \sin(r/A) \qquad |r| \leq \pi A$$
$$= 0 \qquad |r| > \pi A \qquad (6\text{-}39)$$

Andrews weighting function is classified as a hard descender. The nominal tuning constant for this function is A = 1.339.

2- The Biweight Function

This is attributed to Beaton and Tuckey [154], and is classified as a hard descender. The defining functions are given by:

$$\rho(r) = \frac{B^2}{2}\left\{1-[1-(\tfrac{r}{B})^2]^3\right\} \qquad |r| \leq B$$
$$= \frac{B^2}{2} \qquad |r| > B \qquad (6\text{-}40)$$

$$\Psi(r) = r\left[1-(\tfrac{r}{B})^2\right]^2 \qquad |r| \leq B$$
$$= 0 \qquad |r| > B \qquad (6\text{-}41)$$

$$w(r) = \left[1-(\tfrac{r}{B})^2\right]^2 \qquad |r| \leq B$$
$$= 0 \qquad |r| > B \qquad (6\text{-}42)$$

The nominal tuning constant for the Biweight function is B = 4.685.

3- **The Cauchy Weighting Function**

The weighting function is sometimes referred to as the (t-likelihood function) and is classified as a soft redescender defined by [149]

$$\rho(r) = \frac{C^2}{2} \log\left[1 + \left(\frac{r}{C}\right)^2 \right] \tag{6-43}$$

$$\Psi(r) = r \left[1 + \left(\frac{r}{C}\right)^2 \right]^{-1} \tag{6-44}$$

$$w(r) = \left[1 + \left(\frac{r}{C}\right)^2 \right]^{-1} \tag{6-45}$$

The nominal tuning constant for the Cauchy weighting function is $C = 2.385$.

4- **The Fair Weighting Function**

This is a weighting function defined by [155]

$$\rho(r) = F^2 \left[\frac{|r|}{F} - \log\left(1+\frac{|r|}{F}\right) \right] \tag{6-46}$$

$$\Psi(r) = r \left[1 + \frac{|r|}{F} \right]^{-1} \tag{6-47}$$

$$w(r) = \left[1 + \frac{|r|}{F} \right]^{-1} \tag{6-48}$$

The Fair function is characterized by a monotone $\Psi(r)$ and is designed as an approximation to the least absolute residuals estimation procedure. The nominal tuning constant for this function is $F = 1.400$.

5- **The Huber Weighting Function**

The Huber weighting function is defined by [156]:

$$\rho(r) = \begin{cases} \frac{r^2}{2} & |r| \le H \\ H|r| - \frac{H^2}{2} & |r| > H \end{cases} \tag{6-49}$$

$$\Psi(r) = \begin{cases} r & |r| \le H \\ H \,\mathrm{sgn}(r) & |r| > H \end{cases} \tag{6-50}$$

$$w(r) = \begin{cases} 1 & |r| \le H \\ \frac{H}{|r|} & |r| > H \end{cases} \tag{6-51}$$

The nominal tuning constant H for the Huber weighting function is $H = 1.345$.

A comparison of the Hubert defining functions for the nominal tuning constant and the least squares functions is given in Figure (6-1). It is noted that as the measurement errors' magnitude increases, the weighting is decreased.

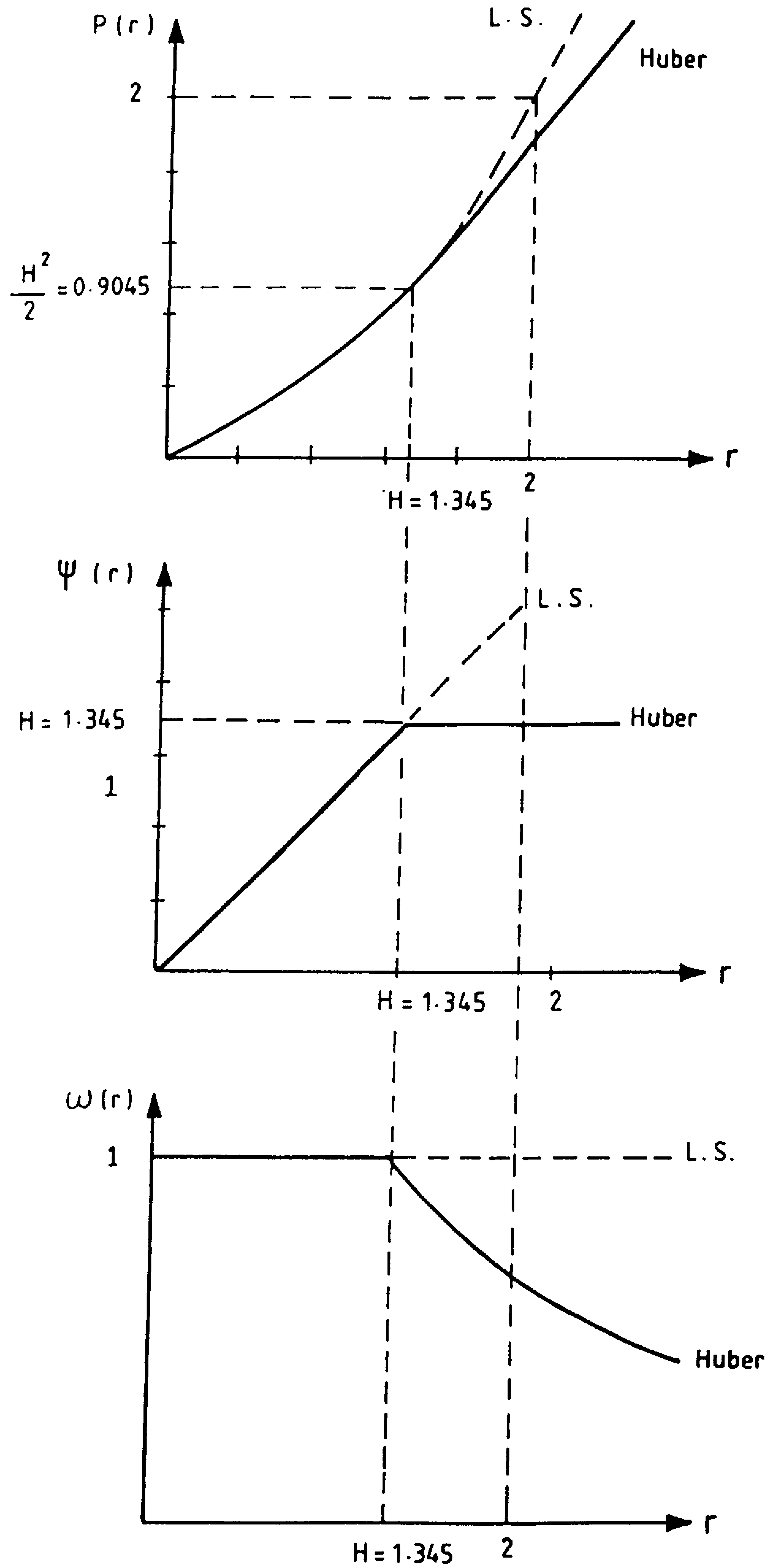


Fig.(6-1) Comparison of Least Squares and Huber's Defining Functions

6- The Logistic Function

This is defined by [149]

$$\rho(r) = L^2 \log\left[\cosh\left(\frac{r}{L}\right)\right] \tag{6-52}$$

$$\psi(r) = L \tanh\left(\frac{r}{L}\right) \tag{6-53}$$

$$w(r) = \left(\frac{r}{L}\right)^{-1} \tanh\left(\frac{r}{L}\right) \tag{6-54}$$

The nominal tuning costant L for this function is L = 1.205.

7- The Talwar Weighting Function

The Talwar function is defined by [157]:

$$\rho(r) = \begin{cases} \frac{1}{2}r^2 & |r| \leq T \\ \frac{T^2}{2} & |r| > T \end{cases} \tag{6-55}$$

$$\psi(r) = \begin{cases} r & |r| \leq T \\ 0 & |r| > T \end{cases} \tag{6-56}$$

$$w(r) = \begin{cases} 1 & |r| \leq T \\ 0 & |r| > T \end{cases} \tag{6-57}$$

The nominal tuning constant T for the Talwar function is T = 2.795.

8- The Welsch Weighting Function

This soft redescender function is defined by [158]:

$$\rho(r) = \frac{W^2}{2}\left[1 - \exp\left[-\left(\frac{r}{W}\right)^2\right]\right] \tag{6-58}$$

$$\psi(r) = r \exp\left[-\left(\frac{r}{W}\right)^2\right] \tag{6-59}$$

$$w(r) = \exp\left[-\left(\frac{r}{W}\right)^2\right] \tag{6-60}$$

The nominal tuning constant W for this function is W = 2.985.

It has been noted that the starting value $\hat{X}$ used in robust parameter estimation must be chosen carefully. Using the least squares solution can disguise the high residual points. The L_1-norm (LARE) estimates would be a good choice of starting values. In the ROSEPACK implementation the LARE solution technique of Bartels and Conn [159] is used.

Robust estimation procedures are extremely helpful in locating outliers and highly influential measurements. Whenever a least squares analysis is performed, it would be useful [148] to perform a robust fit also. If the results of the two procedures are in substantial agreement, then one should use the least squares results, because conclusions based on L.S. are at present better understood. However, if the results of the two analyses differ, then reasons for these differences should be identified. Measurements that are downweighted in the robust fit should be carefully examined. For a discussion of this aspect Mosteller and Tuckey [160] is recommended reading.

6.7 The Iterative Weighting Least Squares Estimator

In considering the problem of improving on the Least Squares parameter estimates for the fuel cost models in power systems, El-Hawary [161] and El-Hawary and Kumar [162] proposed the Iterative Weighted Least Squares (IWLS) method to deal with the problem. The basic concept of the technique is outlined in this section.

The weighted least squares error parameter estimator coincides with the unbiased minimum variance parameter estimator for the choice of the weighting matrix as [163]:

$$W = R^{-1} \tag{6-61}$$

The matrix R is the error covariance matrix of the measurements

$$R = E\{[v-\bar{v}]\ [v-\bar{v}]^T\} \tag{6-62}$$

where $\bar{v}$ is the mean value of the measurement noise v. This requires knowledge of the statistics of measurement errors, which involves repeating the measurements under the same conditions to establish the required means and covariances. In the lack of this information one may attempt to build this information from the weighted least square process itself.

One starts with a unit weighting matrix $W_0 = I$

An estimate $\hat{X}_0$ is obtained as

$$\hat{X}_0 = [H^T W_0 H]^{-1} H^T W_0 Z$$

The error V_0 is obtained as

$$V_0 = Z - H \hat{X}_0$$

From V_0, one approximates the covariance matrix R_1 by

$$R_1 = \underline{F}\ (V_0)$$

Now one takes a new weighting matrix

$$W_1 = R_1^{-1}$$

A new estimate vector $\hat{X}_1$ is calculated as

$$\hat{X}_1 = \left[H^T W_1 H \right]^{-1} H^T W_1 Z$$

A new measurement error V_1 is formed as

$$V_1 = Z - H \hat{X}_1$$

Before proceeding one checks the improvement in $\hat{X}_1$ over $\hat{X}_0$ by finding

$$\Delta_1 = \hat{X}_1 - \hat{X}_0$$

If Δ_1 is close to zero (a given tolerance) no further improvement will be required. Otherewise, R_2 is calculated and a new estimate is found $\hat{X}_2$. The process is repeated

On the j^{th} iteration, we have

$$\hat{X}_j = \left[H^T W_j H \right]^{-1} H^T W_j Z \qquad (6\text{-}63)$$

$$W_j = R_j^{-1}$$

The relation between the IWLS and the IRLS techniques can be seen from inspection of eq. (6-34) for the IRLS iterations and eq(6-63) for the IWLS iterations. The same form is adopted by both methods, but the difference is in the choice of the diagonal matrix elements. In the IRLS scheme, from eq.(6-31) the dependence of the new weightings is on the current iterations errors using one of the standard weighting functions. The IWLS is an expanded memory weighting as opposed to the one-step weighting of the IRLS.

6.8 Computational Results

An extensive computational experiment to explore the application of the LARE and IRLS estimators to the units of the system under consideration was conducted. On the basis of the system economic dispatch results, a number of units were selected for the study. Data for units 2-5 were selected to examine the aspect of an infeasible parameter estimate further. Four more data sets representing units 28, 42, 44 & (45-50) were considered, since these units experienced more changes in the optimal loading than all other units. In total 5 data sets were examined out of a possible 30 data sets.

The objective of the experiments was to ascertain the validity of the least squares parameter estimates and the underlying assumption of a Gaussian error distribution. The results of the experiment were used to determine the most economic mode of loading based on the newly obtained parameters of the units considered. Comparisons and conclusions in terms of various error measures, loadings, and minimum cost were also made in the work reported in this section.

The experiments were conducted using the Robust Statistical Estimation Package (ROSEPACK) implementations of LARE and IRLS estimators, on a DEC Micro-VAX II in Fortran 77. The convergence criterion used was 10^{-4} on the scaled errors in the parameter estimates. Scaling of elements of the vector Z and the matrix H was carried out using column maximum element scaling. Least absolute residual estimates were used to initiate the iterations.

Experiments using the eight weighting functions and the LARE estimators were conducted. Elements of the diagonal weighting matrix were examined on each iteration. Recommended default values of the tuning constant for each weighting function were used. The following error measures were considered and are reported in this section.

SSR - Sum of squares of residuals (measurement errors). This is L_2 norm.
SAR - Sum of the absolute value of the residuals. This is L_1 norm.
MAXRE - Maximum of the absolute value of the residuals. This is L_∞ norm.

This last measure is useful for ranking purposes using error measures. This is because the Least Squares Estimates (LSE) will always yield the lowest SSR and LARE will result in the lowest SAR by their very definition.

It should be noted in the application of the IRLS that whenever the scaled residuals are less than the tuning constant of the weighting function, the value of weighting will be unity for the measurement. If all scaled residuals are less than the tuning constant, then the algorithm returns the estimate to be precisely that of the LSE. This can be seen from eq. (6-34).

Results and conclusions given in this section can be divided into two parts. In the first, conclusions using statistical measures from ROSEPACK are drawn, and in the second results of economic dispatch using the GRG algorithm used in Chapter 5 are used to further evaluate the conclusions from an end-use point of view. This latter part is a unique contribution of this investigation.

Data for the four identical units (2-5) were listed in Table (4-1) and in Table (6-1) a listing of the results obtained using ROSEPACK are given. It is noted that out of the eight weighting functions, only Andrews resulted in estimates different from the LSE values. These are listed in Table (6-1) along with the error measures. Here LSE and Andrews gave almost identical parameters. The maximum of absolute residuals (MAXRE) is lowest for the Andrews estimates.

Table (6-1)
Comparison of Parameter
Estimates and Error Measures
for Units 2-5

	a_0	a_1	a_2	SSR	SAR	MAXRE
LSE	-0.7356019	1.415329	-0.0051867	0.0025847	0.1393057	0.023969
LARE	-0.7518344	1.416294	-0.0051850	0.0027186	0.131484	0.0274066
ANDREWS	-0.7359567	1.415346	-0.0051861	0.0025848	0.138885	0.023883

Table (6-2) lists fuel cost measurement data for unit 28 in this system. This data was used to obtain parameter estimates as shown in Table (6-3). It is noted that for this unit, three weighing functions resulted in new estimates. These are Andrews, Huber, and Talwar. The LSE values for a_0 and a_2 are the highest and the value of a_1 is the lowest. In terms of SSR, the LARE and Huber are closest to LSE measure, whereas, Huber and Andrews are closest to LARE in terms of SAR. Maximum value of the absolute residual is lowest for the LSE estimates.

Table (6-2)
Fuel Cost Measurement Data
for Unit 28

P (MW)	F (Mcal/hr)
20	61.20
25	69.38
30	78.60
35	89.78
40	101.00
45	112.50
50	124.00
55	135.85
60	147.60
65	159.90
70	172.20
75	184.50

Table (6-3)
Comparison of Parameter Estimates
and Error Measures for
Units 28

	a_0	a_1	a_2	SSR	SAR	MAXRE
LSE	23.48927	1.698061	0.00609200	5.027196	6.14531	1.3146
LARE	21.01563	1.809463	0.00500448	6.262388	5.047783	1.993324
ANDREWS	18.77898	1.883038	0.00439636	10.24764	5.659702	3.001726
HUBER	20.42926	1.832238	0.00477977	6.803525	5.262701	2.21407
TALWAR	19.07516	1.871445	0.00450209	9.609438	5.645875	2.895098

Table (6-4) gives the data for unit 42 and results of parameter estimation are given in Table (6-5). For this unit, only Andrews weightings resulted in parameter estimates different from the LSE parameters. The highest values of a_0 and a_2 are due to Andrews which can also be seen to be a good compromise on SAR and SSR between LARE estimates.

Table (6-4)
Fuel Cost Measurement Data
for Unit 42

P (MW)	F (Mcal/hr)
6	14.00
7	16.24
8	18.46
9	20.66
10	22.82
11	24.99
12	27.18
13	29.54
14	31.93
15	34.43
16	37.04

Table (6-5)
Comparison of Parameter Estimates
and Error Measures for
Units 42

	a_0	a_1	a_2	SSR	SAR	MAXRE
LSE	2.504698	1.812307	0.02113061	0.1015925	0.9186739	0.1392401
LARE	2.587843	1.799369	0.0215628	0.1031212	0.8775129	0.1603174
ANDREWS	2.51777	1.810364	0.02119283	0.1016335	0.9121832	0.1429066

Data for unit 44, are given in Table (6-6), and the corresponding parameter estimates and error measures are given in Table (6-7). Here one notes that Andrews and Hubers functions were active for this data set. One notes that the LSE estimates give highest a_0 and a_2. The best (lowest maximum residual) is due to the LSE as well. Andrews SSR is closest to LSE values and its SAR are closest to LARE's.

Table (6-6)
Fuel Cost Measurement Data
for Unit 44

P (MW)	F (Mcal/hr)
6	13.68
7	15.90
8	18.12
9	20.33
10	22.54
11	24.74
12	26.95
13	29.25
14	31.64
15	34.13
16	36.80

Table (6-7)
Comparison of Parameter Estimates
and Error Measures for
Units 44

	a_0	a_1	a_2	SSR	SAR	MAXRE
LSE	2.025374	1.840458	0.02019813	0.0960667	0.9137835	0.1565771
LARE	1.753449	1.900622	0.01718764	0.1081151	0.8275089	0.236557
ANDREWS	1.997309	1.847085	0.01984565	0.0963338	0.9003233	0.1688386
HUBER	1.453874	1.967021	0.01367429	0.1768975	0.9124443	0.3731728

The fifth data set is for units (45-50) and is listed in Table (6-8). The parameter estimation results are given in Table (6-9). One notes that Andrews parameter estimates are close to those for the LSE estimator.

Table (6-8)
Fuel Cost Measurement DatA
for Unit 45-50

P (MW)	F (Mcal/hr)
5	37.00
6	39.60
7	41.30
8	44.80
9	47.70
10	50.50
11	52.80
12	56.40
13	59.80
14	64.40
15	69.00

Table (6-9)
Comparison of Parameter Estimates
and Error Measures for
Units 45-50

	a_0	a_1	a_2	SSR	SAR	MAXRE
LSE	30.11814	0.875764	0.1121209	2.226672	4.569719	0.6515026
LARE	27.80556	1.408331	0.08611126	3.078162	4.472225	0.9166641
ANDREWS	30.10572	0.877533	0.112077	2.226759	4.565876	0.654669

It is important at this point to attempt to rank the parameter estimates using the various methods in terms of the statistical error measures. Earlier, it was stated that the SSR for the LSE would be the lowest, and the value of SAR would be lowest for LARE. The maximum absolute error (MAXRE) would be a good criterion for ranking and the listing of Table (6-10) gives such ranking for the five data sets for five sets of parameter estimates. The ranking gives a score of 1 to the top method with lowest MAXRE followed by 2 for the second lowest MAXRE and so on. It is clear that LSE values are superior, but those for Andrews are close second. The same conclusions can be drawn for Table (6-11) giving the ranking in terms of the sum of squares of error. In terms of the sum of absolute errors (SAR), Andrews estimates are close second to LARE's.

Table (6-10)
Ranking of Methods in Terms of
Maximum Absolute Error

	2-5	28	42	44	45-50
LSE	2	1	1	1	1
LARE	3	2	3	3	3
ANDREWS	1	5	2	2	2
TALWAR		4			
HUBER		3		4	

Table (6-11)
Ranking of Methods in Terms of
Sum of Squares of Errors

	2-5	28	42	44	45-50
LSE	1	1	1	1	1
LARE	3	2	3	3	3
ANDREWS	2	4	2	2	2
TALWAR		5			
HUBER		3		4	

Table (6-12)
Ranking of Methods in Terms of
Sum of Absolute Error

	2-5	28	42	44	45-50
LSE	3	5	3	4	3
LARE	1	1	1	1	1
ANDREWS	2	4	2	2	2
TALWAR		3			
HUBER		2		3	

It should be noted that one does not have enough information to properly rank Talwar's and Huber's results. It is therefore concluded on the basis of the foregoing analysis to proceed with further evaluation on the basis of the intended use of the acquired parameters in performing economic dispatch of the 56 unit system. Table (6-13) lists the minimum cost of the operation using the parameter estimates for power demand values ranging from 700 MW to 1400 MW. Andrews results appear to be superior, as can be seen from Table (6-14).

Table (6-13)
Comparison of Minimum Cost of Operation
Using Different Parameter Estimates

	LSE	TALWAR	HUBER	ANDREWS	LARE
700	36,089.7	36,063.97	36,074.03	36,051.81	36,078.00
800	39,336.2	39,372.47	39,382.53	39,272.89	39,336.00
900	42,817.0	42,807.61	42,818.72	42,745.38	42,815.00
1000	46,380.0	46,384.50	46,386.62	46,311.04	46,381.00
1100	50,092.0	50,097.05	50,095.26	50,017.77	50,092.00
1200	53,995.6	53,990.18	53,986.25	53,909.03	53,995.00
1300	59,560.0	59,554.95	59,551.12*	59,474.05*	59,561.00
1400	64,219.6	64,213.93	64,210.10*	64,133.03*	64,219.00

* The asterisk means that convergence was not obtained.

Table (6-14)
Ranking of Methods
in Terms of Minimum Cost of Operation

	LSE	TALWAR	HUBER	ANDREWS	LARE
700	5	2	3	1	4
800	3	4	5	1	2
900	5	2	4	1	3
1000	2	4	5	1	3
1100	2	5	4	1	3
1200	5	3	2	1	4
1300	3	2	1*	5*	4
1400	5	3	2*	1*	4

The experiments evaluating the performance of the parameter estimates were conducted using the GRG dispatch. Parameters used in Chapter 5 were retained and only those new ones for each method were changed in the runs reported.

Results comparing the optimal power generation of the 56 units for loads of 700 and 800 MW are given in Table (6-15). Here one notes that for the 700 MW demand the optimal generations are identical when using LSE and Andrews parameters. Lower generations for units 27, 28, 31, 36, 44 and 45-50 are required by Andrews parameters, whereas higher generation is required at unit 32, 37-40 and 42 for the 800 MW demand.

Table (6-16) lists the results for 900 and 1000 MW demands. Table (6-17) gives results for 1100 and 1200 MW demands. Finally Table (6-18) lists results for 1300 MW demands. Again the optimal generations for these latter demands are identical.

Comparing the minimum operating costs, one finds that values predicted by Andrews parameters are consistently lower than those obtained using the LSE values. This takes place even for 700 MW, 1300 MW and 1400 MW demands where the optimal generations are identical and logically one would expect the same actual fuel costs to be incurred. One can conclude that results of the comparison should be treated with more caution.

Table (6-15)
Comparison of Optimal Loadings
Using LSE and Andrews Parameters
for 700 and 800 MW Loads

	700 MW		800 MW	
	LSE	ANDREWS	LSE	ANDREWS
1	8	8	8	8
2	18	18	18	18
3	18	18	18	18
4	18	18	18	18
5	18	18	18	18
6	8	8	8	8
7	10.24	10.24	12	12
8	9.07	9.07	12	12
9	5.57	5.57	12	12
10	8	8	8	8
11	8	8	8	8
12	11.12	11.12	15	15
13	10	10	10	10
14	10	10	10	10
15	10	10	10	10
16	10	10	10	12
17	10	10	10	10
18	10	10	10	10
19	10	10	10	10
20	8	8	8	8
21	20	20	20	20
22	20	20	20	20
23	20	20	20	20
24	20	20	20	20
25	20	20	20	20
26	20	20	20	20
27	20	20	29.55	27.01
28	20	20	36.42	30.65
29	65	65	65	65
30	65	65	65	65
31	25	25	44.61	44.29
32	25	25	44.03	44.54
33	8	8	8.09	8
34	8	8	9.20	8
35	8	8	8	8
36	8	8	8.09	8
37	2	2	4.97	5.03
38	2	2	4.97	5.03
39	2	2	4.97	5.03
40	2	2	4.98	5.05
41	6	6	6	6

Table (6-15) Cpmt'd

	700 MW		800 MW	
	LSE	ANDREWS	LSE	ANDREWS
42	6	6	7.85	16
43	6	6	6	6
44	6	6	7.44	7
45	5	5	5.63	5.56
46	5	5	5.63	5.56
47	5	5	5.63	5.56
48	5	5	5.64	5.56
49	5	5	5.67	5.56
50	5	5	5.62	5.56
51	8	8	8	8
52	8	8	8	8
53	8	8	8	8
54	8	8	8	8
55	8	8	8	8
56	8	8	8	8
F Dh/hr	36,089.	36,051.8	39,336.2	39,272.9

Table (6-16)
Comparison of Optimal Loadings
Using LSE and Andrews Parameters
for 900 and 1000 MW Loads

	900 MW		1000 MW	
	LSE	ANDREWS	LSE	ANDREWS
1	8	8	8	8
2	18	18	18	18
3	18	18	18	18
4	18	18	18	18
5	18	18	18	18
6	8	8	8	8
7	12	12	12	12
8	12	12	12	12
9	12	12	12	12
10	8	8	8	8
11	8	8	8	8
12	15	15	15	15
13	10	10	10	10
14	10	10	10	10
15	10	10	10	10

Table (6-16) Cont'd

	900 MW		1000 MW	
	LSE	ANDREWS	LSE	ANDREWS
16	12	12	12	12
17	10	10	10	10
18	10	10	10	10
19	10	10	10	10
20	8	8	8	8
21	20.71	20.05	30.84	30.95
22	20.71	20.05	30.84	30.95
23	20.71	20.05	30.84	30.95
24	20.71	20.05	30.84	30.95
25	22	22	22	22
26	22	22	22	22
27	36.46	35.92	45.52	45.24
28	40.04	35.63	48.75	46.23
29	65	65	65	65
30	65	65	65	65
31	75	74.69	75	75
32	75	75	75	75
33	12.11	11.81	17.74	17.18
34	12.15	11.94	16.01	15.70
35	9.70	9.43	14.74	14.25
36	12.11	11.81	17.74	17.18
37	5.59	5.52	6.28	6.28
38	5.59	5.52	6.28	6.28
39	5.59	5.52	6.28	6.28
40	5.59	5.52	6.28	6.28
41	6	7.86	12.49	12.42
42	9.46	16	11.35	16
43	6	6	10.96	10.84
44	8.62	8.77	11.18	11.19
45	5.86	5.78	6.32	6.29
46	5.87	5.77	6.32	6.29
47	5.87	5.81	6.32	6.29
48	5.86	5.89	6.32	6.29
49	5.85	6.04	6.32	6.29
50	5.84	6.54	6.32	6.29
51	8	8	8.13	8.12
52	8	8	8	8
53	8	8	8	8
54	8	8	8	8
55	8	8	8	8
56	8	8	8	8
F Dh/hr	42,817.	42,745.38	46,380.	46,311.04

Table (6-17)
Comparison of Optimal Loadings
Using LSE and Andrews Parameters
for 1100 and 1200 MW Loads

	1100 MW		1200 MW	
	LSE	ANDREWS	LSE	ANDREWS
1	8	8	8	8
2	18	18	18	18
3	18	18	18	18
4	18	18	18	18
5	18	18	18	18
6	8	8	8	8
7	12	12	12	12
8	12	12	12	12
9	12	12	12	12
10	8	8	8	8
11	8	8	8	8
12	15	15	15	15
13	10	10	10.01	10
14	10	10	10.31	10
15	10	10	10.46	10
16	12	12	12.00	12
17	12	12	12.00	12
18	10.06	10	10.9	11.34
19	10	10	10.7	10.43
20	9.84	9.71	13.43	12.09
21	43.74	44.64	60	60
22	35	35	35	35
23	43.74	43.06	60	60
24	35	35	35	35
25	22	22	22	22
26	22	22	22	22
27	56.55	55.86	60	60
28	57.49	57.83	73.35	75
29	65	65	65	65
30	65	65	65	65
31	75	75	75	75
32	75	75	75	75
33	23.91	23.59	30	30
34	20.09	20.12	30	28.37
35	20.19	20	30	30
36	23.91	23.59	30	30
37	7	7	7	7
38	7	7	7	7
39	7	7	7	7
40	7	7	7	7
41	16	15.86	16	16

Table (6-17) Cont'd

	1100 MW		1200 MW	
	LSE	ANDREWS	LSE	ANDREWS
42	13.62	16	15.79	16
43	15.03	14.14	16	16
44	13.55	13.49	15.77	16
45	6.78	6.74	7.26	7.32
46	6.78	6.74	7.26	7.32
47	6.78	6.74	7.26	7.32
48	6.78	6.74	7.26	7.32
49	6.78	6.74	7.26	7.31
50	6.78	6.74	7.26	7.35
51	8.69	8.67	9.23	9.40
52	8.69	8.67	9.23	9.40
53	8.69	8.67	9.23	9.33
54	8.53	8.67	9.01	9.71
55	8	8	8	8
56	8	8	8	8
F Dh/hr	50,092.	50,017.77	53,995.6	53,909.03

Table (6-18)
Comparison of Optimal Loadings
Using LSE and Andrews Parameters
for 1300 and 1400 MW Loads

	1300 MW		1400 MW	
	LSE	ANDREWS	LSE	ANDREWS
1	8	8	28	28
2	18	18	18	18
3	18	18	18	18
4	18	18	18	18
5	18	18	18	18
6	18	18	18	18
7	12	12	12	12
8	12	12	12	12
9	12	12	12	12
10	8	8	12	12
11	8	8	12	12
12	15	15	15	15
13	10	10	18	18
14	10	10	18	18
15	10	10	18	18

Table (6-18) Cont'd

	1300 MW		1400 MW	
	LSE	ANDREWS	LSE	ANDREWS
16	12	12	12	12
17	12	12	12	12
18	10	10	18	18
19	10	10	18	18
20	18	18	18	18
21	59	59	60	60
22	35	35	35	35
23	59	59	60	60
24	35	35	35	35
25	22	22	22	22
26	22	22	22	22
27	60	60	60	60
28	75	75	75	75
29	65	65	65	65
30	65	65	65	65
31	75	75	75	75
32	75	75	75	75
33	30	30	30	30
34	30	30	30	30
35	30	30	30	30
36	30	30	30	30
37	7	7	7	7
38	7	7	7	7
39	7	7	7	7
40	7	7	7	7
41	16	16	16	16
42	16	16	16	16
43	16	16	16	16
44	16	16	16	16
45	15	15	15	15
46	15	15	15	15
47	15	15	15	15
48	15	15	15	15
49	15	15	15	15
50	15	15	15	15
51	22	22	18	18
52	22	22	18	18
53	22	22	22	2
54	22	22	22	2
55	8	8	26	26
56	8	8	8	8
F Dh/hr	59,560.	59,474.05	64,219.6	64,133.03

The percentage relative differences in the minimum operating costs for the power demands considered are calculated as

700 MW	0.1%
800 MW	0.16%
900 MW	0.167%
1000 MW	0.148%
1100 MW	0.148%
1200 MW	0.16%
1300 MW	0.1443%
1400 MW	0.1348%

Although all differences are below 0.2% and may be neglected from a numerical analysis point of view, it is important to realize that this can mean significant loss in economy over the span of years. Possible explanations of the discrepancy at 700 MW are effects of round off in listing solutions and that the Andrews fit underestimates the cost values at very low and high loadings on units modelled.

6.9 Summary

In this chapter parameter estimation techniques that offer an alternative to LSE were explored. The main features of the IRLS techniques is that they serve as a diagnostic tool to verify whether measurement errors are normally distributed or not. The least absolute residual estimator is ideally suited for measurement errors whose distribution is a double exponential. The IRLS estimates are a compromise between the two cases.

Application of the estimation techniques to data for a number of units in the system revealed that the measurement errors are almost Gaussian and that estimates based on Andrews weighting functions offer an attractive alternative. From an intended use point of view in terms of results of economic dispatch of the system this conclusion appears to be confirmed, although the improvements are slight.

CHAPTER 7
CONCLUSIONS, RECOMMENDATIONS, AND SUGGESTIONS FOR FURTHER RESEARCH

7.1 Introductory Comments

The intent of this chapter is to offer a summary of contributions made in this work as well as a number of proposed areas for future research that can be thought of as an off-shoot of this investigation. For conclusions and comparison of methods in each area, the final section of each chapter may be consulted.

7.2 Summary of this Research

In demonstrating that this thesis embodies certain contributions to advancing knowledge, it is important to categorize such contribution as follows:

1- Contributions that advance the state of the art in a specific area of applied mathematics with special reference to power system economy-security functions.

2- Contributions in terms of advancing state of knowledge of pertinent characteristics of the UAE's Abu Dhabi electric power system. In this category we include developing advanced techniques for efficient operation of this system. The application of these techniques is expected to result in about 4% savings in annual fuel costs for the system as well as enhancing the reliability of the system.

3- Contributions in terms comprehensive up-to-date literature reviews in areas encompassed by this research work. In this category we also have summaries of salient features of advanced techniques in the area treated.

7.2.1 Advancing State-of-the-Art

It is suggested that contributions have been made to advance state-of-the-art in the following areas:

a- Discovered the possibility of non-feasible parameters for the fuel cost model of thermal generating plants.

b- Devised a systematic procedure for identifying possible (bad data points) measurements pairs that cause non-feasible parameters. The procedure recognizes the role of recursive least squares as well as weighted least squares in achieving the desired result.

c- Presented a formulation of constrained parameter estimation for fuel cost models as a non-linear programming problem.

d- Proposed an extremely fast economic dispatching algorithm based on some realization of optimality conditions in closed form. Testing the method which is referred to as the Lambda aggregation technique proved a significant reduction in computing time over other techniques.

e- Suggested and implemented a number of initial guess estimators for use in conjunction with GRG implementation. A particular scheme based on the first result of a Lambda aggregation procedure proved to achieve considerable reduction in computing time.

f- Implemented for the first time robust parameter estimation techniques employing the iteratively reweighted least squares with a number of weighting functions to the fuel cost model parameter estimation problem.

g- Advocated the evaluation of parameter estimates on the basis of not only accuracy considerations, but also from an intended use point of view.

7.2.2 Advancing UAE's Abu Dhabi System Operation

The work reported in this thesis utilized system data from the UAE's Abu Dhabi electric system to test algorithms and perform comparative evaluations in the three major areas of load forecasting, parameter estimation and economic dispatch. As a result, the state of operation of the system will gain due to the following contributions:

a- Determined optimal smoothing constants for the mid-term and short-term problems of load forecasting using the following conventional forecasting techniques:
   - i- General Exponential Smoothing
   - ii- Winters' Additive Procedure
   - iii- Winters' Multiplicative Procedure

b- Explored and determined an optimal time series model structure for short-term and mid-term load forecasting using Box-Jenkins methodology.

c- Conducted extensive testing of Kalman filtering using a time-invariant state space representation employing a simple maximum likelihood noise adaptive filtering for short-term forecasting.

d- Applied a recursive weighted least squares technique with exponential weighting in the past with optimal coefficients to the short-term forecasting problem.

e- Implemented an instrumental variable procedure using temperature as an exogenous variable to obtain short-term forecasting.

f- Implemented GRG and Lambda iterations techniques with initial guess enhancements to obtain optimal dispatch strategies for the load profile of the system.

g- Conducted extensive testing of the newly proposed Lambda aggregations method.

h- Obtained a set of model of parameters for the 56 units of the Abu Dhabi system.

In all areas treated, comparative evaluations of possible alternatives were carried out. As a result it is felt that there is a sound basis for economic operation of the system which was not available before.

In particular, it is expected that fuel cost model parameter estimation for the majority of system units will be carried out using Andrews weighting function in conjunction with the iteratively reweighted least squares. The short-term forecasting function appears to be best dealt with using Box-Jenkins, Winters' multiplicative model, or the instrumental variable method. It may be advantageous to consider combinations of these forecasts to arrive at better forecasting accuracy. The results of the model parameters estimators and load forecasting are then used in an economic dispatching routine based on the Lambda aggregation method to schedule the system units to meet the forecast power system load.

It is to be emphasized that preliminary investigations show that the implementation of the results of economic scheduling of Chapter 5 to the Abu Dhabi System can potentially increase the savings by up to 4% over the present practice of generation scheduling at 1988 cost figures (Table 7-1).

7.2.3 Comprehensive Reviews

In terms of comprehensive up-to-date literature surveys, it is felt that the following are worth noting.

a- Survey of Load forecasting methods up to 1987.
b- Survey of Adaptive filtering techniques.
c- Survey of optimal economic dispatch area.
d- Survey of fuel cost model parameter estimation.
e- Survey of robust estimation techniques.

In addition, application-oriented summaries of procedures and techniques employed in the research work were also given.

7.3 Suggestions for Future Work

In conducting research work reported in this thesis, a number of areas of possible future research work emerge as deserving specific mention in this Section.

1- The work on load forecasting using general exponential smoothing involved a trial-and-error selection of the required smoothing constants for the various approaches tested. In the literature, this trial-and-error method is an accepted practice. It appears, however that the selection of optimal smoothing constants should be based on a systematic means for their evaluation. An attractive option to pursue is to investigate maximum likelihood choices as viable means for determining the smoothing constants.

2- The inquiry into available noise-adaptive filtering techniques pointed out the need for simple and fast techniques for evaluation of the noise statistics. It appears that improvements on innovation correlation methods hold an excellent promise. For example, the method of Carew-Belanger appears to be more amenable for faster solution techniques than those suggested in the original work.

3- The fuel cost parameter estimation problem appears to offer a fruitful area for applying techniques of adaptive filtering and this should be investigated.

4- The newly developed Lambda aggregation method should be pursued further. Possible extensions to systems with losses and those incorporating decoupled load flow are worth a closer examination for the potential computational efficiencies that might be realized.

5- The application of robust estimation techniques to the load forecasting problem is another area of possible future work.

6- Perhaps one of the most overlooked areas that should be considered is that of developing an integrated approach to adaptive economic dispatch. Here the objective would be to take the parameter variations (both in fuel cost as well as load evolution models) into consideration in an on-line optimizing function. A stochastic objective and formulation might be the route to follow towards this goal.

Before concluding this chapter, it is useful to mention that a number of publications based on some work reported in this thesis have been published [164-167].

Table (7-1)
Comparison of
Generation Costs

Hour	Load	Existing Practice (Dh)	Newly Developed System (Dh)
1	839	39,773.26	38,377.31
2	792	37,053.27	35,849.72
3	762	34,850.13	33,850.58
4	716	32,479.66	31,447.48
5	691	31,510.00	30,468.21
6	671	30,750.13	29,694.51
7	671	30,750.13	29,694.47
8	726	33,032.76	31,839.84
9	781	36,325.32	34,625.26
10	847	39,959.05	38,295.57
11	885	42,227.49	40,568.35
12	927	44,551.52	42,536.40
13	982	47,354.90	45,202.74
14	1,039	49,391.78	47,726.68
15	1,029	48,994.82	47,159.68
16	1,029	48,994.82	47,159.68
17	959	46,583.06	44,568.17
18	884	42,739.86	40,789.35
19	864	41,382.41	39,567.60
20	895	42,657.31	40,937.39
21	887	42,275.43	40,621.52
22	897	42,812.98	41,082.35
23	907	43,459.74	41,551.24
24	912	43,814.92	41,788.13
Total:		973,724.75	935,402.23
Savings:			38,322.52
% Savings:			3.94

REFERENCES

[1] Sullivan, R.L., "Power System Planning", New York, McGraw-Hill, 1978.

[2] Wood A.J., and Wollenberg, B., "Power System Generation and Control", New York, John Wiley & Sons, 1984.

[3] Feasibility Study for Interconnection of Arabian Gulf States Electrical Power Systems, Phase 5 report, August'1986.

[4] El-Hawary, M.E., and Christensen, G.S., Optimal Economic Operation of Electric Power Systems, New York, Academic Press, 1979.

[5] Jay, F., editor, "IEEE Standard Dictionary of Electrical & Electronics Terms", ANSI/IEEE Standard 100-1984, Wiley-Interscience, 1984.

[6] Makridakis, S., Wheelwright, S.C., and McGee V.E., Forecasting Methods and Applications, New York, John Willey & Sons, 1983.

[7] Bowerman, B.L., and O'Connell, Richard, T., Time Series Analysis & Forecasting, PWS Publishers, Duxbury Press, Massachusetts, USA, 1979.

[8] Galiana, F.D., "Short Term Load Forecasting", ERDA,76-66, Conf.750867, Systems Engineering for Power : Status and Prospects, pp. 105-115.

[9] IEEE Committee Report, "Load Forecasting Bibliography, Phase I", IEEE Trans. on Power Apparatus & Systems, Vol.PAS-99, No.1, pp.53-58, 1980.

[10] Sachdev, M.S., Billinton, R., and Peterson, C.A., "Representative Bibliography on Load Forecasting", IEEE Trans. on Power Apparatus and Systems, Vol. PAS-96, No.6, pp. 697-700, 1977.

[11] Abu El-Magd, M.A., and Sinha, N.K., "Short Term Load Demand Modeling and Forecasting : A Review", IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-12, No.3, pp.370-382, 1982.

[12] Happ, H.H., "Optimal Power Dispatch - A Comprehensive Survey", IEEE Trans. on Power Apparatus and Systems, Vol. PAS-96, pp. 841-854, 1977.

[13] Dryar, H.A., "The Effect of Weather on the System Load", AIEE Transactions, Vol. 63, pp.1006-1013, 1944.

[14] Davies, M., "The Relationship between Weather and Electricity Demand", Proc. IEEE, Vol. 106-C, pp. 27-37, 1958.

[15] Heinemann, G.T., Norman, D.A., and Plant, E.C., "The Relationship between Summer Weather and Summer Loads - A Regression Analysis", IEEE Trans. on Power Apparatus and Systems, Vol. PAS-85, No.11, pp. 1144-1154, 1966.

[16] Farmer, E.D., "A method of Prediction for Nonstationary of Processes and its Application to the Problem of Load Estimation", Proc. IFAC Congress, Basle, pp.47-54, 1963.

[17] Farmer, E.D., and Patton, M.J., "The Prediction of Load on a Power System", Proc. IFAC Congress, London, pp. 21F.1-21F.7., 1966.

[18] Farmer, E.D., and Patton, M.J.,"Development of On-line Load Prediction Techniques with Results from Trials in the South-West Region of CEGB", Proc. IEEE, Vol. 115, No.10, pp. 1549-1588, 1968.

[19] Mattewman, P.D., and Nicholson, H., "Techniques for Load Prediction in the Electric Supply Industry", Proc. IEEE, Vol. 115, No.11, pp.1451-1460, 1968.

[20] Lijesen, D.P., and Rosing, J., "Adaptive Forecasting of Hourly loads Based on Load Measurements and Weather Information",IEEE Trans. On Power Apparatus and Systems, Vol. PAS-90, No.4, pp. 1757-1767, 1971.

[21] Christiaanse, W.R., "Short-Term Load Forecasting using General Exponential Smoothing", IEEE Trans. On Power Apparatus and Systems, Vol. PAS-90, No.2, pp. 900-910, 1971.

[22] Brown, R.G., "Smoothing, Forecasting, and Prediction of Discrete Time Series", Englewood Cliffs, N.J., Prentice-Hall Inc., 1965.

[23] Sachdev, M.S., and Ibrahim, S.A., "Short-Term On-Line Load Forecasting", paper C72-454-7, IEEE PES Summer Meeting, July, 1972.

[24] Gupta, P.C., and Yamada, K., "Adaptive Short-Term Forecasting of Hourly Loads, Using Weather Information", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-91, pp. 2085-2094, 1972.

[25] Galiana, F.D., Handschin, E., and Fiechter, A., "Identification of Stochastic Electric Load Models From Physical Data", IEEE Trans. on Automatic Control, Vol. AC-19, pp. 887-893, 1974.

[26] Stanton, K.N., Gupta, P.C., and El-Abaid, A.H., "Long Range Demand Forecasting for Electric Utility Industry", Proc. of the 1969 Power Industry Computer Applications.

[27] Stanton, K.N., and Gupta, P.C, "Long Range Forecasting of Electrical Demand by Probability Methods", Proceedings of the American Power Conference, Vol. 31, pp. 964-969, 1969.

[28] Stanton, K.N., "Medium Range Weekly and Seasonal Peak Demand Forecasting by Probability Methods", IEEE Trans. on Power Apparatus and Systems, Vol. PAS-90, pp. 1183-1189, 1971.

[29] Gupta, P.C., "A Stohastic Approach to Peak Power Demand Forecasting in Electric Utility Systems", IEEE Trans. on Power Apparatus and Systems, Vol. PAS-90, pp. 824-832, 1971.

[30] Vemuri, S., Hill, E.F., and Balasubramanian, R., "Load Forecasting Using Stohastic Models", Proc. 1973 Power Industry Computer Applications Conference, paper TPI-A, pp. 31-37, 1973.

[31] Keyhani, A., and El-Abiad, A., "One-step-ahead Load Forecasting for On-line Applications", paper C75 027-8, IEEE PES Winter Power Meeting, 1975.

[32] Keyhani, A., and Eliassi Radi, T., "A Simulation Study for Recursive Prediction of Hourly Load Demands", Proc. 1977 PICA, pp. 228-236.

[33] Hagan, M.T., and Klein, R., "Identification Techniques of Box and Jenkins Applied to the Problem of Short-term Load Forecasting, Paper A77 618-2, IEEE PES Summer Meeting, 1977.

[34] Meslier, F., "New Advances in Short-Term Load Forecasting Using Box and Jenkins Approach", paper A78 051-5, IEEE Winter Meeting, 1978.

[35] Abu El-Magd, M.A., and Sinha, N.K., "Two New Algorithms for On-line Modeling and Forecasting of the Load Demand of a Multinode Power System", IEEE Trans. on Power Apparatus and Systems, Vol. PAS-100, No.7, pp. 3246-3253, 1981.

[36] Vemuri, S., Huang, W.L., and Nelson, D.J., "On-line Algorithms for Forecasting Hourly Loads of An Electric Utility", IEEE Trans. on Power Apparatus and Systems, Vol. PAS-100, No.8, pp. 3775-3784, 1981.

[37] Hagan, M.T., and Behr, S.M., "The Time Series Approach to Short-term Load Forecasting", paper 87 WM 044-1, IEEE Winter Power Meeting, 1987.

[38] Ross, D.W., Ackerman, G.B., Bischke, R., Podmore, R., and Wall, K.D., "Short-Term Load Prediction for Economic Dispatch of Generation", Proc. 1979, PIAC, pp. 198-204, 1979.

[39] Van Meeteren, H.P., and Van Son, P.J.M., "Short-Term Load Prediction with a Combination of Different Models", Proc. 1979 PICA, pp. 192-197, 1979.

[40] Srinivasan, K., and Pronovost, R., "Short-Term Load Forecasting using Multiple Correlation Models", IEEE Trans. on Power Apparatus and Systems, Vol. PAS-94, pp. 1854-1858, 1975.

[41] De Martino, B., Fusco, G.D., 'Mariani, E., Randino, E., and Ricci, P., "A Medium and Short-Term Load Forecasting Model for Electrical Industry", Proc. 1979, PICA, pp. 186-191, 1979.

[42] Abu-Hussien, M.S., Kandil, M.S., Tantawy, M.A., and Farghal, S.A., "An Accurate Model for Short-Term Load Forecasting", IEEE Trans. on Power Apparatus and Systems, Vol. PAS-100, pp. 4158-41, 1981.

[43] Keyhani, A., and Miri, S.M., "On-line Weather Sensitive and Industrial Group Bus Load Forecasting for Micro processor Based Applications", IEEE Trans. on Power Apparatus and Systems, Vol. PAS-102, pp. 3868-3876,1983.

[44] Irisarri, G.B., Widergren, S.E., and Yehsakul, P.D., "On Line Load Forecasting for Energy Control Centre Application", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-101, pp. 71-78, 1982.

[45] Krogh, B., De Llinas, E.S., and Lesser, B., "Design and Implementation of an On Line Load Forecasting Algorithm" IEEE Trans. on Power Apparatus and Systems, Vol. PAS-101, pp. 3284-3289, 1982.

[46] Toyoda, J., Chen, M., and Inouye, Y., "An Application of State Estimation to Short-Term Load Forecasting", IEE Trans. on Power Apparatus and System, Vol. PAS-89, pp. 1678-1688, 1970.

[47] Sharma, K.L.S., and Mahalanabis, A.K., "Application of Adaptive Estimation Techniques for Short-Term Load Prediction", paper C72 455-4,IEE PES Summer Meeting,1972.

[48] Sharma, K.L.S., and Mahalanabis, A.K., "Recursive Short-Term Load Forecasting Algorithm", proc. IEE Vol. 121, pp. 59-62, 1974.

[49] Singh, G., Biswas, K.K., and Mahalanabis, A.K., "Power System Load Forecasting Using Smoothing Techniques", International Journal of Systems Science, Vol. 9, pp. 363-368, 1978.

[50] Galiana, F.D., and Schweppe, F.C., "A Weather-dependent Probabilistic Model for Short-Term Load Forecasting", paper C72 171-2, IEEE PES Winter Meeting, 1972.

[51] Singh, G., Biswas, K.K., and Mahalanabis, A.K., "Load Modeling for Real-Time Monitoring of Power Systems", IEE Trans. on Power Apparatus and Systems, Vol. PAS-96, pp. 1908-1914, 1977.

[52] Lu, K.V., and Rao, N.D., "Short Range Load Forecasting by the Maximum Entropy Method", Proc. IFAC Symposium on Computer Applications in Large Scale Power Systems, Vol.1, pp. 261-268, 1979.

[53] Sebesta, D., "Pattern Recognition Applied to Forecasts", Electric World, Vol. 191, pp. 78-79, 1979.

[54] Dehdashti, A.S., "Forecasting of Hourly Load by Pattern Recognition - A Deterministic Approach", IEE Transactions on Power Apparatus and Systems, Vol.PAS-101, pp. 3290-3294, 1982.

[55] Rahman, S., and Bhatnagar, R., "An Expert System Based Algorithm for Short-Term Load Forecast", paper 87 MW 082-1, IEE PES Winter Meeting, 1987.

[56] McRae, M.R., "Integrating Load Management Programs into Utility Operations and Planning with a Load Reduction Forecasting System", IEEE Transactions on Power Apparatus and Systems, Vol.PAS-104, pp.1321-1325,1985.

[57] Bunn, D.W., and Farmer, E.D., "Comparative Models for Electrical Load Forecasting", New York, Wiley, 1985.

[58] Abraham, B., and Ledolter, J., "Statistical Methods for Forecasting", New York, Wiley, 1983.

[59] Box, G.E.P., and Jenkins, G.M., "Time Series Analysis : Forecasting and Control", Revised Edition, Oakland, CA, Holden-Day, 1976.

[60] Yule, G.U., "On the Time-Correlation Problems with Special Reference to the Variate Difference Correlation Methods", J. Roy Statistics Soc., Vol. 84, pp. 497-526.

[61] Walker, G., "On Periodicity in Series of Related Terms", Proc. Royal Soc., Vol. A131, p. 518, 1931.

[62] Wold, H., "A study in the Analysis of Stationary Time Series", Second Edition, Almquist and Wicksell, Uppsala, 1954.

[63] Tukey, J.W., "Discussion Emphasizing the Connection Between Analysis of Variance & Spectrum Analysis", Technometrics, Vol.3, pp.189-219, 1961.

[64] Pankratz, A., "Forecasting With Univariate Box-Jenkins Models : Concepts and Cases", New York, Wiley, 1983.

[65] Marquardt, D.W., "An Algorithm for Least Squares Estimation of Nonlinear Models", J. of Soc. of Ind. and Appl. Math, Vol. 11, pp 431-441, 1963.

[66] Kalman, R.E., "A New Approach to Linear Filtering and Prediction Problems", Trans. ASME, Vol.82, Ser.D, J. Basic Engineering, pp. 35-43, Mar., 1960.

[67] Kalman, R.E., and Bucy, R.S., "New Results in Linear Filtering and Prediction Theory", Trans. ASME., Vol.83, Ser, D., pp. 95-108, Mar. 1961.

[68] Gelb, A., (ed.), "Applied Optimal Estimation", MIT Press, Cambridge, Massachusetts, 1974.

[69] Meditch, J.S., "Stochastic Optimal and Linear Estimation and Control", McGraw-Hill, New York, 1971.

[70] Sage, A.P., and Melsa, J.L., "Estimation Theory with Applications to Communications and Control", McGraw-Hill, New York, 1971.

[71] Anderson, B.D.O., and Moore, J.B., "Optimal Filtering", Prentice-Hall, Englewood Chiffs, N.J., 1979.

[72] Maybeck, P.S., "Stochastic Models, Estimation and Control", Vol.1, Academic Press, New York, 1979.

[73] Heffes, H., "The Effect of Erroneous Models on the Kalman Filter Response", IEEE Trans. On Automatic Control, Vol. AC-11, pp. 541-543, 1966.

[74] Nishimura, T., "Error Bounds of Continuous Kalman Filters and the Application to Orbit Determination Problems", IEEE Trans. On Automatic Control, Vol. AC-12, pp.268-275, June, 1967.

[75] Price, C.F., "An Analysis of the Divergence Problem in the Kalman Filter", IEEE Trans. On Automatic Control, Vol.AC-13, pp.669-672, 1968.

[76] Mehra, R.K., Approaches to Adaptive Filtering", IEEE Trans. On Automatic Control, Vol. AC-17, pp. 693-698, 1972.

[77] Chin, L., "Advances in Adaptive Filtering", Control and Dynamic Systems (C.T. Leondes ed.), Vol.15, pp. 277-365, 1979.

[78] Maybeck, P.S., "Stochastic Models, Estimation and Control", Vol. 2, Academic Press, New York, 1982.

[79] Isaksson, A., "Identification of Time-Varying Systems and Application of System Identification to Signal Processing", Linkoping University, Department of Electrical Engineering, Thesis No. 75, 1986.

[80] Stengel, R.F., "Stochastic Optimal Control: Theory and Application", New York, Wiley-Interscience, 1986.

[81] Magill, D.T., "Optimal Adaptive Estimation of Sampled Stochastic Processes", IEEE Trans. On Automatic Control, Vol. AC-10, No. 4, pp. 434-439, October, 1965.

[82] Smith, G.L., "Sequential Estimation of Observation Error Variances in a Trajectory Estimation Problem", AIAA Journal, Vol. 5, No. 11, pp. 1964-1970, 1967.

[83] Hilborn Jr., C.G., and Lainiotis, D.G., "Optimal Estimation in the Presence of Unknown Parameters", IEEE Trans. On Systems, Science and Cybernetics, Vol. SSC-5, pp. 38-43, Jan. 1969.

[84] Sims, F.L., and Lainiotis, D.G., "Recursive Algorithm for the Calculation of the Adaptive Kalman Filter Weighting Co-efficients", IEEE Trans. On Automatic Control, Vol. AC-14, pp. 215-218, 1969.

[85] Hilborn, C.G., and D.G. Lainiotis, "Optimal Adaptive Filter Realizations for Sample Stochastic Processes with an Unknown Parameter", IEEE Trans. On Automatic Control, Vol. AC-14, pp. 767-770, 1969.

[86] Lainiotis, D.G., "Optimal Adaptive Estimation: Structure and Parameter Adaptation", IEEE Trans. On Automatic Control, AC-16, pp.160-170, 1971.

[87] Alspach, D.L., "A Parallel Filtering Algorithm for Linear System with Unknown Time Varying Noise Statistics", IEEE Trans. On Automatic Control, Vol. AC-19, pp. 552-556, 1974.

[88] Gauvrit, M., Fargeon, C., and Rodrigo, P., "Synthesis and Performance of Bayesian Estimators", Control and Dynamic Systems, (C.T. Leondes, ed.), Vol. 19, pp. 81-124, 1983.

[89] Abramson Jr., P.D., "Simultaneous Estimation of the State and Noise Statistics in Linear Dynamic Systems", M.I.T Rep. TE-25, May, 1968.

[90] Shellenbarger, J.C., "Estimation of Covariance Parameters for An Adaptive Kalman Filter", in Proc. Nat Electronics Conference, Vol. 22, pp. 698-702, 1966.

[91] Sage, A.P., and Husa, G.W., "Adaptive Filtering with Unknown Prior Statistics", in 1969 Proc. Joint Automatic Control Conference, pp. 760-769, 1969.

[92] Mehra, R.K., "On-Line Identification of Linear Dynamic Systems with Applications to Kalman Filtering", IEEE Trans. On Automatic Control, Vol. AC-16, pp.12-21, Feb. 1971.

[93] Mehra, R.K., "On the Identification of Variances and Adaptive Kalman Filtering", IEEE Trans. On Automatic Control, Vol. AC-15, pp. 175-184, Apr. 1970.

[94] Carew, B., and Belanger, P.R., "Identification of Optimum Filter Steady State Gains for Systems with Unknown Noise Covariances", IEEE Trans. On Automatic Control, Vol. AC-18, pp. 582-587, Dec. 1973.

[95] Ohap, R.F., and Stubberud, A.R., "Adaptive Minimum Variance Estimation in Discrete-Time Linear Systems", in Control and Dynamic Systems, (C.T Leondes ed.), Vol.12, pp.583-624, Academic Press, New York, 1976.

[96] Brewer, H.M., "Identification of the Noise Characteristics in a Kalman Filter", in Control and Dynamic Systems, (C.T. Leondes ed.), Vol. 12, pp. 491-581, Academic Press, New York, 1976.

[97] Martin, W.C., and Stubberud, A.R., "The Innovations Process with Applications to Identifications", in Control and Dynamic Systems, (C.T Leondes, ed.), Vol.12, pp.173-258, Academic Press, New York, 1976.

[98] Martin, W.C., and Stubberud, A.R., "An Uncoupling Method for Linear System Identification", IEEE Trans. On Automatic Control, Vol. AC-21, pp. 506-509, 1976.

[99] Ohnishi, K., "Direct Recursive Estimation of Noise Statistics", in Control and Dynamic Systems, (C.T. Leondes ed.), Vol. 16, pp. 249-297, 1980.

[100] Jazwinski, A.H., "Adaptive Filtering", Automatica, Vol.5, pp. 475-485, 1969.

[101] Harvey, A.C., "Time Series Models", Phillip Allan Publishers, Oxford, 1981.

[102] Sallas, W.M., and Harville, D.A., "Best Linear Recursive Estimation for Mixed Linear Models", Journal of the American Statistical Associations Theory & Methods Section, Vol. 76, pp. 860-869, 1981.

[103] Young, P., "Recursive Estimation and Time-Series Analysis", Berlin, Springer-Verlag, 1984.

[104] Ljung, L., "System Identification : Theory for the User" Englewood Cliffs, N.J., Prentice-Hall, 1987.

[105] Ljung, L., and Soderstrom, T., "Theory and Practice of Recursive Identification", Cambridge, Massachusetts, MIT Press, 1983.

[106] Astrom, K.J, and Wittenmark, B., "Computer Controlled Systems : Theory and Design", Englewood Cliffs, N.J., Prentice-Hall, 1984.

[107] Central Electricity Generating Board, "Modern Power Station Practice", Vol. 8 Pergamon Press, 1981.

[108] Sasson, A.M., "Optimal Load Flow - A Practical Outlook", Course Text on Applications of Optimization in Power Systems Engineering, IEEE Catalogue Number 76 CH 1107-2-PWR, New York, 1976.

[109] Schweppe, F.C., "Static State Estimation in Electric Power Systems", Proceedings of IEEE, Vol. 62, pp. 972-982, 1974.

[110] Debs, A.S., "Estimation of Steady State Power Systems Model Parameters", IEEE Transactions on PAS, Vol.PAS-93,pp.1260-1268, 1974.

[111] Ringlee, R.J., "Sensitivity Methods for Economic Dispatch of Hydro-electric Plants", IEEE Trans. On Automatic Control, Vol. AC-10, pp. 315-322, 1965.

[112] Vemuri, S., and Hill, E.F., "Sensitivity Analysis of Optimum Operation of Hydro-thermal Plants", IEEE Trans. On Power Apparatus and Systems, Vol. PAS-96, pp. 688-696, 1977.

[113] Dillon, T.S., and Tun, T., "Application of Sensitivity Methods to the Problem of Optimal Control of Hydro-thermal Power Systems", Optimal Control Applications and Methods, Vol.2, pp.117-143, 1981.

[114] El-Hawary, M.E., and Mansour, S.Y., "Performance Evaluation of Parameter Estimation Algorithms for Economic Operation of Electric Power Systems", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-101, pp. 574-582, 1982.

[115] El-Hawary, M.E., and Mansour, S.Y., "Optimal Parameter Estimation for Basic Problems of Economic Operation in Electric Power Systems", Optimal Control Applications and Methods, Vol.2, pp.269-287, 1981.

[116] El-Hawary, M.E., and Kumar, M., "Recursive Parameter Estimation for Economic Operation Models", Proceedings of the Fourth International Symposium on Large Engineering Systems, Calgary, 1982.

[117] Power Technology Inc., "Synthetic Electric Utility Systems for Evaluating in Advanced Technologies", Electric Power Research Institute Report EM-285, February, 1977.

[118] Mendel, J.M., Discrete Techniques of Parameter Estimation. The Equation Error Formulation, Marcel Dekker, New York, 1973.

[119] Lawson, C.L., and Hanson, R.J., "Solving Least Squares Problems", Prentice-Hall, Englewood Cliffs, N.J., 1974.

[120] Arismunader, A., and Noakes, F., "Bibliography on Optimum Operation of Power Systems : 1919-1959", AIEE Trans. PAS-81, 1963, pp.864-871.

[121] IEEE Working Group Report, "Description and Bibliography of Major Economy-Security Functions : Part I, Descriptions", IEEE Transactions of PAS, Vol. PAS-100, 1981, pp. 211-214

[122] IEEE Working Group Report, "Description and Bibliography of Major Economy-Security Functions : Part II, Bibliography, 1959-1972", IEEE Transaction of PAS, Vol. PAS-100, 1981, pp.214-223.

[123] IEEE Working Group Report, "Description and Bibliography of Major Economy-Security Functions : Part III, Bibliography, 1973-1979", IEEE Transaction of PAS, Vol. PAS-100, 1981, pp. 223-235.

[124] Steinberg, M.J., and Smith, T.H., "The Theory of Incremental Rates, Part I and II, Electrical Engineering, March-April, 1934.

[125] Steinberg, M.J., and Smith, T.H., "Economy Loading of Power Plants and Electric Systems", Wiley, New York, 1943.

[126] Ward, J.B., "Economy Loading Simplified", AIEE Trans. Vol. 72, Part III, pp. 1306-1311, 1953.

[127] Early, E.D., Phillips, W.E., and Shreve, W.T., "An Incremental Cost of Power Developed Computer", AIEE Transactions, Vol. 74, part III, pp. 529-534, 1955.

[128] Kirchmayer, L.K., "Economic Operation of Power Systems, "Wiley, New York, 1958.

[129] Heydt, G.T., "Computer Analysis Methods for Power Systems", MacMillan, New York, 1986.

[130] Sasson, A.M., and Merrill, H.M., "Some Applications of Optimization Techniques to Power System Problems", IEEE Proceedings, Vol. 62, pp.959-972, 1984.

[131] Kihn, H.W., and Tucker, A.W., "Nonlinear Programming", in Second Berkley Symposium on Mathematical Programming, Statistics, and Probability, 1950, University of California Press, Berkley, 1951.

[132] El-Hawary, M.E., "Electrical Power Systems design and Analysis", Reston-Prentice-Hall, N.J., 1983.

[133] Abadie, J., and Carpentier, "Generalization of the Wolfe Reduced Gradient Method for Nonlinear Constraints", Note HR6678, Electricite D.E. France, Paris, October, 1965.

[134] Abadie, J., and Carpentier, "Generalization of the Wolfe Reduced Gradient Method to the Case of Nonlinear Constraints", in : Optimization, R., Fletcher, (ed.), Academic Press, New York, pp.37-74, 1969.

[135] Wolfe, P., "Methods for Linear Constraints", in : Nonlinear Programming, Abadie, J., (ed.), North Holland Publishing, Amestradam, 1967, pp. 120-124.

[136] Wolfe, P., "Methods of Nonlinear Programming", in : Recent Advances in Mathematical Programming, Graves and Wolfe, (eds.) McGraw-Hill, New York, 1963, pp. 67-68.

[137] Colville, A.R., "A Comparative Study of Nonlinear Programming Codes", IBM, New York Scientific Center Report 320-2949, 1968.

[138] Lasdon, L.S., Waren, A.D., Jain A., and Ratner, M., "Design and Testing of a Generalized Reduced Gradient Code for Nonlinear Programming", ACM Transactions of Mathematical Software, Vol.4, No.1, March, 1978, pp. 34-50, 1978.

[139] Lasdon, L.S., and Waren, A.D., "Generalized Reduced Gradient Software for Linearly and Nonlinearly Constrained Problems", in : Design and Implementation of Optimization Software, H. Greenberg, (ed.), Sijthoff and Noordhoff Publishers, The Netherlands, 1979, pp. 363-369.

[140] Lasdon, L.S., and Waren, A.D., "GRG2 User's Guide", School of Business Administration, University of Texas at Austin, January, 1986.

[141] Murtagh, B., and Saunders, M., "Nonlinear Programming for Large Sparse Systems", Technical Report SOL 76-15, System Optimization Laboratory, Department of Operations Research, Standford University, August, 1976.

[142] Murtagh, B.A., and Saunders, M.A., "Large - Scale Linearly Constrained Optimization", Mathematical Programming, Vol. 14, No.1, January, 1978, pp. 41-72.

[143] Murtagh, B.A., and Saunders, M.A., "MINOS 5.1 User's Guide", Technical Report SOL 83-20R System Optimization Laboratory, Department of Operations Research, Stanford University, December, 1983, Revised January, 1987.

[144] Dommel, H.W., and Tinney, "Optimal Power Flow Solutions", IEEE Transactions on Power System Apparatus and Systems, Vol. PAS-87, pp. 1866-1876, 1968.

[145] Peschon, J., Bree, D.W., and Hajdu, L.P., "Optimal Power Flow Solutions for Power System Planning", Proceedings of IEEE, Vol, 60, No.1, pp. 64-70, 1972.

[146] Burchett, R.C., Happ, H.H., Vierath, D.R., and Wirgau, K., "Developments in Optimal Power Flow", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-101, No. 2, pp.406-414, 1982.

[147] Burchett, R.C., Happ, H.H., Vierath, D.R., and Wirgau, K., "Large Scale Optimal Power Flow", IEEE Transactions on Power Apparatus and Systems, Vol. PAS-101, No.10, pp. 3722-3723, 1982.

[148] Montgomery, D.C., and Peck, E.A., "Introduction to Linear Regression Analysis, Willey, New York, 1982.

[149] Holland, P.W., and Welsch, R.E., "Robust Regression Using Iteratively Reweighted Least-Squares", Communications in Statistical Theory and Methods, Vol. A6(9), pp. 813-827, 1977.

[150] Huber, P.J., Robust Methods of Estimating Regression Coefficients, Second International Summer School on Problems and Model Choice and Regression Analysis at Rheinhardsbrum, G.D.R., November, 1975.

[151] Bickel, P.J., One-Step Huber Estimates in the Linear Model, J. American Statistical Association, Vol. 70, pp. 428-34, 1975.

[152] Coleman, D., Holland, P.W., Kaden, N., Klema, V., and Peters, S.C., "A System of Subroutines for Iteratively Reweighted Least Squares Computation, ACM Trans. on Mathematical Software, Vol. 6, No. 3 pp. 327-336, 1980.

[153] Andrews, D., Bickel, P., Hampel, F., Huber, P., Rogers, W., and Tukey, J., Robust Estimated of Location: Survey and Advances, Pinceton University Press, Pinceton, 1972.

[154] Beaton, A.E., and Tukey, J.W., The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data, Technometrics, Vol. 16, pp. 147-85, 1974.

[155] Fair, R.C., "On the Robust Estimation of Econometric Models", Ann. Econ. Social Measurements, Vol.3, pp. 667-678, 1984.

[156] Huber, P.J., Robust Estimation of a Location Parameter, Ann. Math. Statistics, Vol. 35, pp. 73-101, 1964.

[157] Hinich, M.J., and Talwar, P.P., "A Simple Method for Robust Regression, J., Amer, Statistical Assoc., Vol. 70, pp. 113-119.

[158] Dennis, J.E., and Welsch, R.E., "Techniques for Nonlinear Least Squares and Robust Regression, Proc. American Statistical Association, Statistical Computing Section Washington, D.C., pp. 83-87, 1976.

[159] Bartels, R., and Conn, A., "Linearly Constrained Discrete, Problems", Tech. Rep. 248, Johns Hopkins University Baltimore, Md, June, 1976.

[160] Mosteller, F., and Tukey, J.W., "Data Analysis and Regression", Reading, Mass., Addison-Wesley, 1977.

[161] El-Hawary, M.E., "Parameter Estimation for Optimal Economic Operation of Power Systems", CEA Engineering and Operating Division Meeting, Montreal, March, 1982.

[162] El-Hawary, M.E., and Kumar, M., "Recursive Parameter Estimation for Economic Operation Models", Proceedings of the Fourth International Symposium on Large Engineering Systems, Calgary, 1982.

[163] Bard, Y., "Nonlinear Parameter Estimation", Academic Press New York, 1974.

[164] Al-Gobaisi, D.M.K., Bajpai, A.C., and Singh, P.N.P., "A Comparative Analysis of Optimization Techniques for Economic Dispatch", Electric Machines and Power Systems, Vol.14, 95/110, 1988.

[165] Al-Gobaisi, D.M.K., Bajpai, A.C., and Singh, P.N.P., "An Application of a Dual Algorithm for Quadratic Programming to Economic Dispatch of Power Systems", Symposium on Electrical Energy in the United Arab Emirates, U.A.E. University, April 3-5, 1988.

[166] Al-Gobaisi, D.M.K., Bajpai, A.C., and Singh, P.N.P., "Short Term Forecasting Techniques Comparative Analysis", Symposium on Electrical Energy in the United Arab Emirates, U.A.E. University, April 3-5, 1988.

[167] Al-Gobaisi, D.M.K., Bajpai, A.C., and Singh, P.N.P., "A Long and Medium Term Power Forecasting Scheme Economically Effective for Third World and Developing Countries", Strategic Power and Utilities Planning Symposium Institute for International Research, Singapore, March 22-23, 1988.

APPENDIX A

LOAD FORECASTING DATA

In this Appendix data used for the load forecasting experiments of Chapter 3 are listed. Short-term hourly load data are listed in Table (A-1) for 672 hours of operation. Figure (A-1-a) shows the same data set for Table (A-1). In Figure (A-1-b) an expanded view of a segment of the data for one week is shown. The data for mid-term load forecasting given as monthly peak demand for 142 months is listed in Table (A-2) and Figure (A-2) is a graphical presentation of the same set.

Table (A-1)
Short-term Load Data set

HOUR	MW	HOUR	MW	HOUR	MW	HOUR	MW	HOUR	MW
1	950	49	923	97	891	145	928	193	909
2	909	50	878	98	859	146	872	194	864
3	871	51	843	99	823	147	847	195	833
4	852	52	813	100	804	148	822	196	804
5	827	53	793	101	784	149	806	197	778
6	803	54	778	102	769	150	787	198	753
7	753	55	723	103	731	151	733	199	713
8	738	56	723	104	701	152	723	200	793
9	797	57	773	105	727	153	784	201	741
10	846	58	792	106	739	154	829	202	776
11	890	59	834	107	770	155	859	203	811
12	928	60	879	108	812	156	902	204	843
13	980	61	927	109	837	157	944	205	885
14	1025	62	982	110	848	158	998	206	950
15	1043	63	998	111	918	159	1011	207	987
16	1063	64	1023	112	940	160	1038	208	1017
17	1058	65	1023	113	940	161	1028	209	1010
18	1002	66	968	114	893	162	988	210	951
19	930	67	898	115	805	163	913	211	876
20	975	68	938	116	860	164	964	212	951
21	990	69	948	117	882	165	974	213	956
22	995	70	959	118	902	166	974	214	951
23	995	71	959	119	912	167	974	215	948
24	971	72	950	120	898	168	959	216	951
25	941	73	915	121	877	169	929	217	895
26	910	74	872	122	832	170	878	218	844
27	866	75	840	123	806	171	846	219	804
28	831	76	819	124	786	172	826	220	795
29	809	77	804	125	766	173	800	221	775
30	805	78	794	126	756	174	771	222	745
31	780	79	744	127	697	175	722	223	683
32	750	80	724	128	687	176	720	224	668
33	794	81	753	129	760	177	743	225	712
34	837	82	791	130	802	178	765	226	757
35	878	83	816	131	835	179	795	227	773
36	913	84	857	132	862	180	862	228	814
37	963	85	908	133	911	181	918	229	867
38	1032	86	954	134	972	182	963	230	918
39	1039	87	983	135	999	183	988	231	949
40	1058	88	1003	136	1041	184	1020	232	972
41	1028	89	983	137	1039	185	1020	233	957
42	978	90	928	138	999	186	993	234	911
43	908	91	863	139	913	187	887	235	851
44	955	92	923	140	958	188	954	236	906
45	975	93	928	141	978	189	957	237	916
46	985	94	919	142	968	190	972	238	921
47	985	95	916	143	978	191	972	239	935
48	968	96	899	144	963	192	934	240	923

Table (A-1) Continued

HOUR	MW	HOUR	MW	HOUR	MW	HOUR	MW	HOUR	MW
241	878	289	871	337	881	385	913	433	878
242	830	290	814	338	830	386	850	434	833
243	798	291	799	339	792	387	810	435	788
244	768	292	779	340	772	388	787	436	768
245	753	293	759	341	742	389	767	437	748
246	733	294	749	342	722	390	757	438	738
247	684	295	699	343	679	391	697	439	703
248	664	296	668	344	649	392	662	440	663
249	723	297	713	345	715	393	737	441	663
250	740	298	768	346	772	394	772	442	663
251	775	299	808	347	813	395	815	443	683
252	816	300	848	348	842	396	853	444	735
253	846	301	889	349	884	397	911	445	783
254	899	302	959	350	934	398	968	446	804
255	930	303	961	351	980	399	1005	447	882
256	953	304	999	352	1009	400	1022	448	823
257	949	305	985	353	1001	401	1015	449	920
258	895	306	931	354	946	402	975	450	863
259	838	307	863	355	869	403	918	451	798
260	888	308	923	356	945	404	988	452	858
261	888	309	928	357	944	405	988	453	868
262	882	310	928	358	954	406	988	454	903
263	887	311	941	359	954	407	988	455	923
264	880	312	920	360	934	408	950	456	902
265	864	313	860	361	888	409	905	457	862
266	824	314	824	362	840	410	846	458	812
267	784	315	781	363	815	411	826	459	784
268	764	316	768	364	801	412	802	460	759
269	749	317	748	365	777	413	782	461	749
270	739	318	728	366	757	414	762	462	729
271	694	319	675	367	697	415	712	463	685
272	679	320	650	368	672	416	687	464	657
273	689	321	707	369	742	417	722	465	715
274	719	322	758	370	808	418	778	466	774
275	754	323	796	371	847	419	805	467	795
276	789	324	838	372	881	420	837	468	827
277	814	325	890	373	918	421	872	469	883
278	821	326	835	374	964	422	935	470	910
279	908	327	968	375	1004	423	989	471	930
280	920	328	1003	376	1034	424	998	472	991
281	907	329	993	377	1024	425	987	473	996
282	857	330	930	378	968	426	932	474	933
283	797	331	852	379	903	427	857	475	863
284	849	332	922	380	960	428	917	476	923
285	874	333	934	381	978	429	932	477	923
286	899	334	934	382	979	430	912	478	920
287	916	335	934	383	979	431	912	479	935
288	913	336	914	384	950	432	893	480	912

Table (A-1) Continued

HOUR	MW	HOUR	MW	HOUR	MW	HOUR	MW
481	867	529	863	577	855	625	852
482	805	530	818	578	826	626	807
483	876	531	794	579	789	627	782
484	744	532	769	580	752	628	741
485	729	533	744	581	721	629	715
486	719	534	724	582	684	630	697
487	669	535	689	583	647	631	652
488	649	536	659	584	622	632	632
489	699	537	704	585	662	633	698
490	734	538	752	586	717	634	740
491	761	539	791	587	745	635	777
492	800	540	836	588	797	636	821
493	838	541	889	589	838	637	871
494	900	542	939	590	890	638	913
495	941	543	971	591	943	639	943
496	973	544	1008	592	963	640	1004
497	977	545	1003	593	953	641	999
498	932	546	948	594	883	642	929
499	857	547	884	595	829	643	884
500	932	548	945	596	884	644	944
501	952	549	948	597	885	645	944
502	959	550	946	598	892	646	944
503	959	551	943	599	885	647	944
504	912	552	909	600	873	648	914
505	863	553	859	601	843	649	876
506	832	554	819	602	789	650	830
507	819	555	784	603	764	651	795
508	787	556	774	604	729	652	777
509	767	557	749	605	703	653	757
510	747	558	729	606	678	654	727
511	707	559	674	607	648	655	682
512	677	560	647	608	619	656	657
513	727	561	696	609	634	657	703
514	772	562	735	610	671	658	758
515	804	563	743	611	701	659	801
516	827	564	808	612	729	660	856
517	879	565	863	613	775	661	893
518	828	566	911	614	809	662	943
519	962	567	943	615	886	663	966
520	984	568	975	616	904	664	1008
521	966	569	970	617	899	665	993
522	916	570	920	618	836	666	936
523	858	571	857	619	791	667	891
524	926	572	912	620	842	668	956
525	931	573	912	621	863	669	956
526	930	574	912	622	887	670	962
527	936	575	917	623	917	671	962
528	908	576	907	624	907	672	928

Table (A-2)
Monthly Peak Demand in MW
Jan 76 - Oct 87

MONTH	LOAD	MONTH	LOAD	MONTH	LOAD	MONTH	LOAD
1	61.2	37	143.3	73	301.0	109	405.0
2	61.0	38	169.7	74	293.0	110	413.0
3	69.5	39	204.3	75	371.5	111	649.0
4	118.8	40	333.5	76	572.0	112	773.0
5	177.2	41	424.5	77	748.0	113	961.0
6	201.8	42	465.5	78	799.0	114	973.0
7	211.9	43	488.0	79	822.0	115	1028.0
8	219.9	44	487.0	80	834.0	116	1063.0
9	219.5	45	456.0	81	806.0	117	1007.0
10	191.7	46	427.7	82	722.0	118	962.0
11	142.8	47	310.8	83	522.0	119	663.0
12	87.4	48	200.4	84	369.0	120	495.0
13	89.7	49	176.5	85	331.0	121	442.0
14	85.7	50	194.3	86	332.0	122	451.0
15	147.3	51	341.0	87	356.0	123	597.0
16	194.2	52	444.5	88	566.0	124	795.0
17	251.4	53	516.2	89	827.0	125	1097.0
18	271.6	54	550.4	90	883.0	126	1122.0
19	293.1	55	629.7	91	920.0	127	1136.0
20	293.7	56	607.1	92	953.0	128	1124.0
21	292.8	57	573.2	93	907.0	129	1063.0
22	262.0	58	543.5	94	835.0	130	1008.0
23	226.7	59	376.0	95	530.0	131	762.0
24	130.5	60	269.5	96	428.0	132	529.0
25	116.3	61	260.5	97	372.0	133	493.0
26	112.8	62	250.7	98	380.0	134	528.0
27	175.4	63	356.0	99	607.0	135	616.0
28	260.9	64	542.9	100	758.0	136	965.0
29	301.9	65	643.0	101	864.0	137	1084.0
30	345.6	66	674.0	102	906.0	138	1142.0
31	359.6	67	741.7	103	973.3	139	1204.0
32	381.8	68	728.1	104	962.0	140	1213.0
33	353.9	69	701.6	105	900.0	141	1209.0
34	324.6	70	635.0	106	800.0	142	1092.0
35	237.8	71	420.0	107	550.0		
36	168.4	72	345.5	108	475.0		

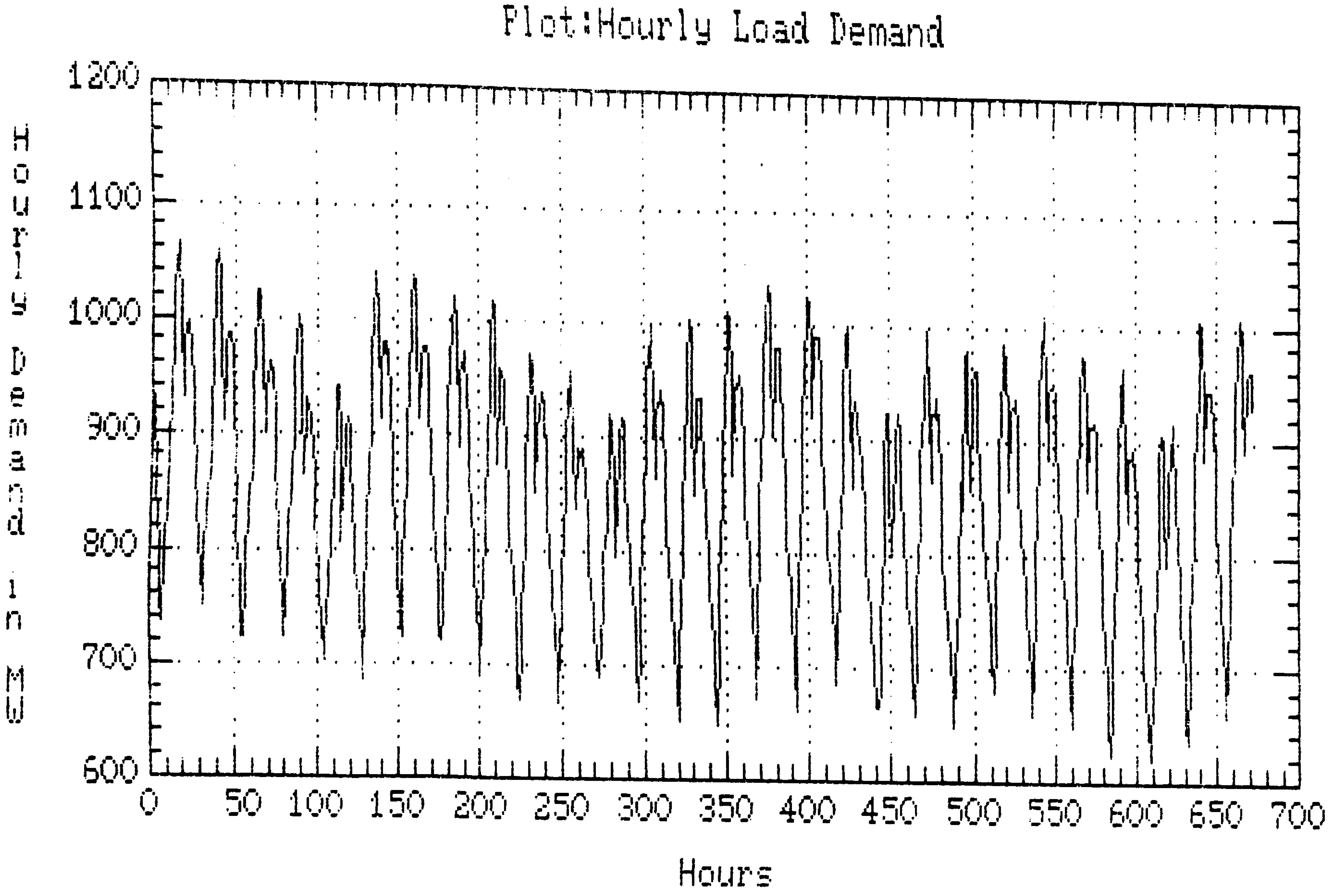


Figure (A-1-a)
Data Base for Short-Term Forecasting

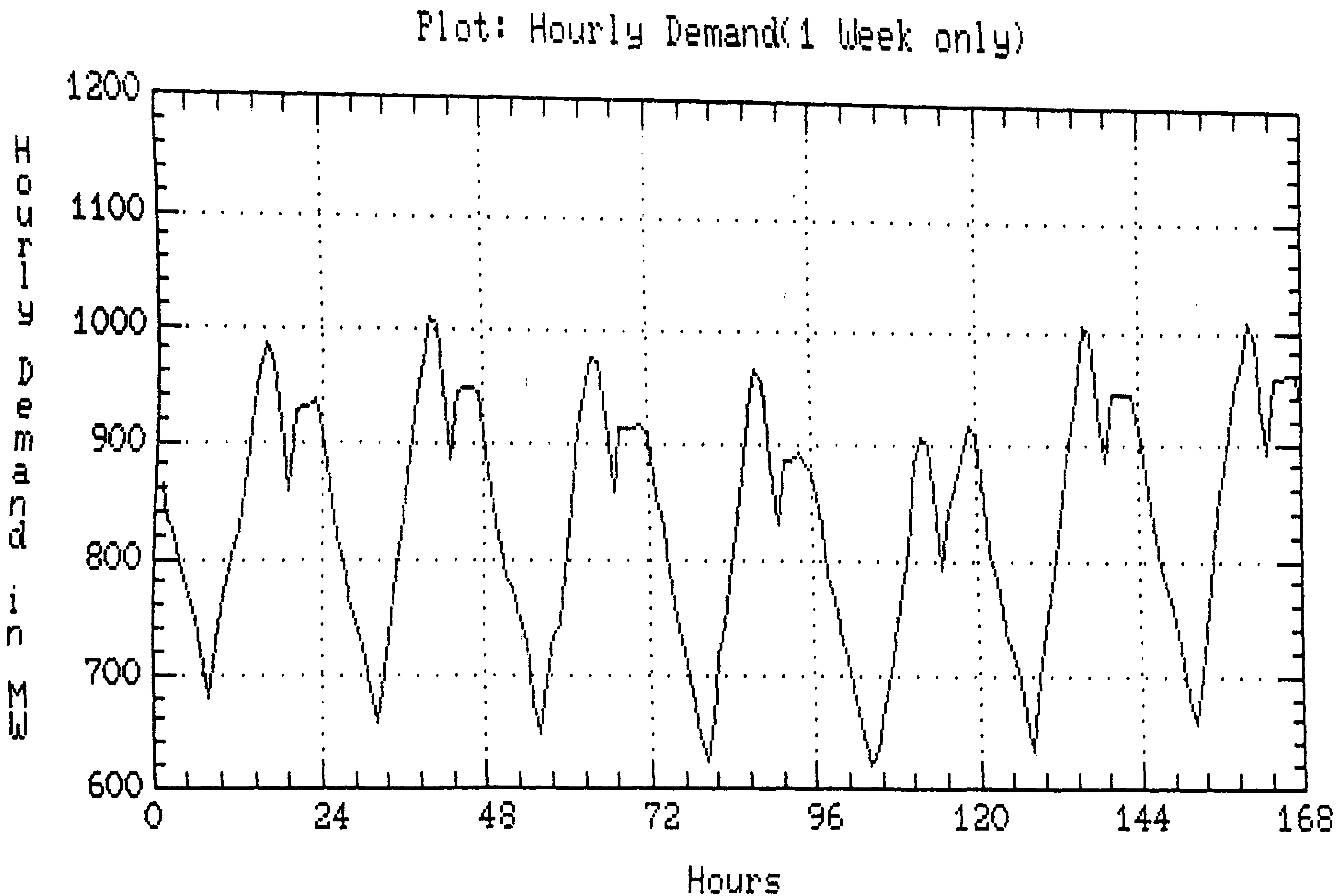


Figure (A-1-b)
Data Base for Short-Term Forecasting
(Summer '86)

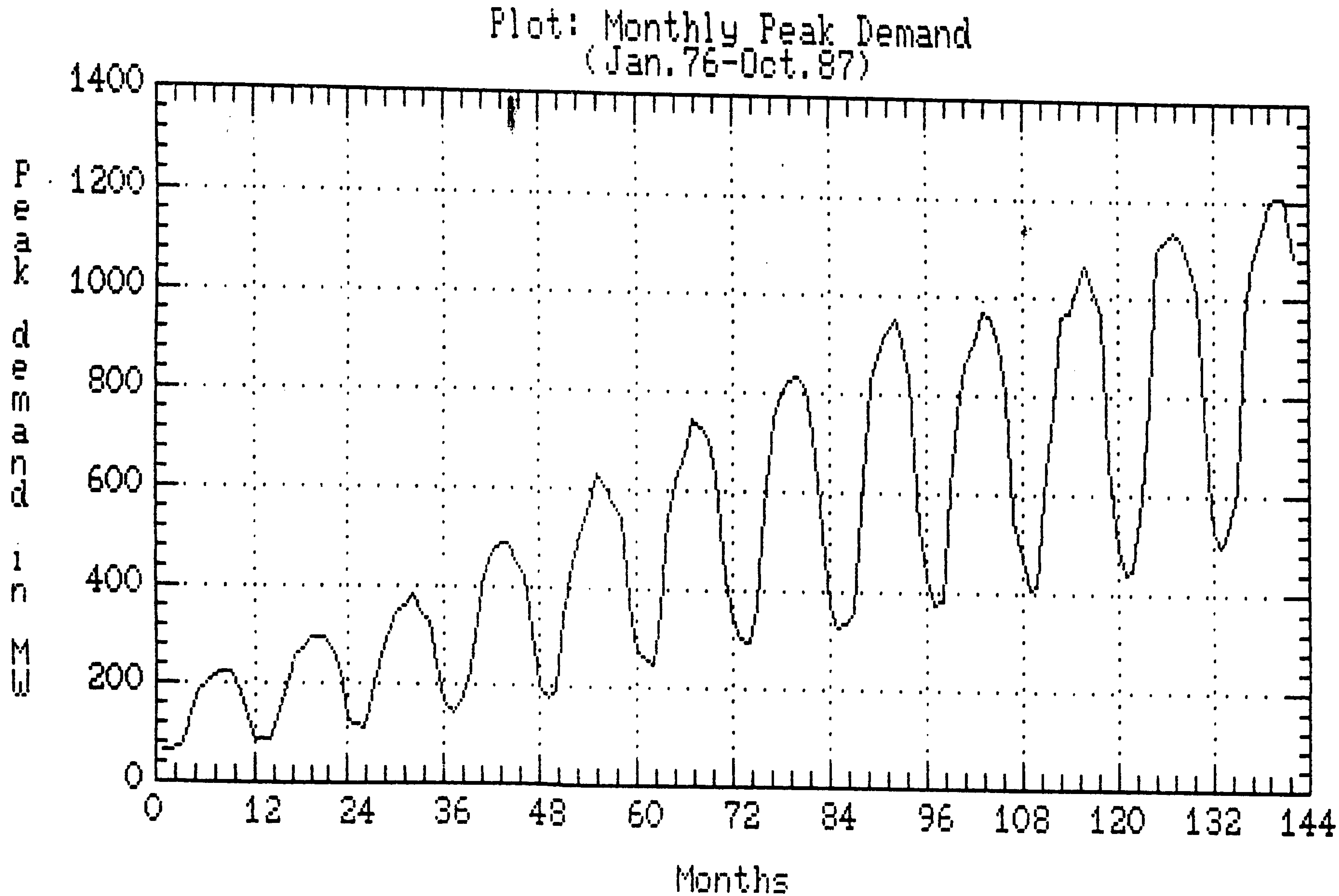


Figure (A-2)
Data Base for Mid-Term Forecasting

APPENDIX B
GENERAL EXPONENTIAL SMOOTHING VIA DISCOUNTED LEAST SQUARES

Equation (3-6) is cast in a form that is convenient for the application of parameter estimation techniques. The unknown parameter vector $\underline{X}$ is introduced by the following decomposition:

$$\underline{X}^T = [\underline{X}_1^T \ \underline{X}_2^T]$$

In this decomposition $\underline{X}_1$ denotes the trend parameter vector defined by

$$\underline{X}_1^T = [\beta_0, \beta_1, \ldots\ldots, \beta_K]$$

Here $\underline{X}_1$ is of dimension (K+1). The vector $\underline{X}_2$ is defined by the following partitioning

$$\underline{X}_2^T = [\underline{X}_{21}^T \ \underline{X}_{22}^T \ \ldots . \ \underline{X}_{2m}^T]$$

Each $\underline{X}_{2i}^T$ for i=1,....,m is defined by:

$$\underline{X}_{2i}^T = [\beta_{1i} \ \beta_{2i}]$$

It is clear that $\underline{X}_2$ is a 2m dimensional vector and corresponds to the seasonal component parameters.

In a similar manner one defines the vector $\underline{h}(.)$ as follows:

$$\underline{h}^T(.) = [\underline{h}_1^T(.) \ \underline{h}_2^T(.)]$$

$$\underline{h}_1^T(.) = \left[1, k, \frac{k^2}{2!}, \ldots\ldots, \frac{k^K}{K!}\right]$$

$$\underline{h}_2^T(.) = [\underline{h}_{21}^T(.) \ \underline{h}_{22}^T(.) \ \ldots . \ h_{2m}^T(.)]$$

where $\underline{h}^T_{2i}(.) = [\sin f_i k, \cos f_i k]$

It is therefore clear that the seasonal model using trigonometric functions equation (3.6) is written in the following compact form:

$$z(k) = \underline{h}^T(.)\,\underline{X} + v(k) \tag{B-1}$$

It is important to note that the vector $\underline{h}$ of the fitting functions in the present model satisfies the recursive relationship [22]:

$$\underline{h}(j) = \underline{L}\,\underline{h}\,(j-1) \tag{B-2}$$

Here the transition matrix $\underline{L}$ is defined by:

$$\underline{L} = \text{diag}\,[\underline{L}_1, \underline{L}_2]$$

The notation diag is used to denote a diagonal matrix of the arguments in brackets.

The (K+1)x(K+1) matrix $\underline{L}_1$ is the transition matrix of the polynomial trend model whose elements are:

$$\ell_{ij} = \frac{1}{(i-j)!} \quad \text{for } i \geq j$$
$$= 0 \qquad \text{otherwise}$$

The matrix $\underline{L}_2$ is defined by:

$$\underline{L}_2 = \text{diag}\,[\underline{L}_{21}, \ldots, \underline{L}_{2m}]$$

Here

$$\underline{L}_{2i} = \begin{bmatrix} \cos f_i & \sin f_i \\ -\sin f_i & \cos f_i \end{bmatrix}$$

The transition property (B-2) plays a central role in the general exponential smoothing methodology via discounted least squares generally attributed to Brown [22], discussed next.

It is assumed that N observations $z(1), z(2), \ldots, z(N)$ are available. The measurement model is assumed to be of the form:

$$z(k) = \sum_{i=1}^{n} x_i\, h_i(k-N) + v(k)$$

Alternatively

$$z(k) = \underline{h}^T(k-N)\,\underline{X} + v(k) \tag{B-3}$$

The n parameters x_i are unknown, but the n functions $h_i(j)$ are assumed to be available. The model Equation (B-3) can be written for smoothing $(k < N)$ purposes as

$$z(N-j) = h^T(-j)\,\underline{X} + v(N-j) \qquad j=0,1,\ldots,N-1 \tag{B-4}$$

The elements of h satisfy the transition rule (B-2)

$$\underline{h}(j+1) = \underline{L}\,\underline{h}(j)$$

To estimate the model parameters, one adopts a discounted least squares approach that seeks to minimize

$$J[N] = \sum_{j=0}^{N-1} W^j\,[z(N-j) - \underline{h}^T(-j)\,\underline{X}]^2 \tag{B-5}$$

The weighting is such that later measurements are weighted more heavily by taking $0 < W < 1$.

The optimal estimate vector $\hat{X}$ based on N measurements is obtained as [58]:

$$\underline{\hat{X}}(N) = \underline{P}(N)\,\underline{\xi}(N) \tag{B-6}$$

Here

$$\underline{P}^{-1}(N) = \sum_{j=0}^{N-1} W^{j}\,\underline{h}(-j)\,\underline{h}^{T}(-j) \tag{B-7}$$

$$\underline{\xi}(N) = \sum_{j=0}^{N-1} W^{j}\,h(-j)\,z(N-j) \tag{B-8}$$

This is a restatement of the normal equations.

From Equation (B-7) one has

$$\underline{P}^{-1}(N+1) = \sum_{j=0}^{N-1} W^{j}\,\underline{h}(-j)\,\underline{h}^{T}(-j)$$

$$= W^{N}\,\underline{h}(-N)\,\underline{h}^{T}(-N) + P^{-1}(N)$$

In the steady state with $N\to\infty$ and elements of $\underline{h}$ being polynomials in the time index, the first term vanishes

$$\lim_{N\to\infty} \underline{P}^{-1}(N+1) = \lim_{N\to\infty} \underline{P}(N) = \underline{P}^{-1}$$

$$P^{-1} = \sum_{j>0} W^{j}\,\underline{h}(-j)\,h^{T}(-j) \tag{B-9}$$

From Equation (B-8), one has

$$\underline{\xi}(N+1) = \sum_{j=0}^{N} W^{j}\,h(-j)\,z(N+1-j)$$

$$= h(o)\,z(N+1) + \sum_{j=1}^{N} W^{j}\,h(-j)\,z(N+1-j)$$

$$\underline{\xi}(N+1) = h(o)\,z(N+1) + \sum_{j=0}^{N-1} W^{j+1}\,h(-j-1)\,z(N-j)$$

Using the transition rule

$$\underline{\xi}(N+1) = h(o)\,z(N+1) + \sum_{j=0}^{N-1} WL^{-1}\,h(-j)\,z(N-j)$$

Thus, from Equation (B-8)

$$\underline{\xi}(N+1) = \underline{h}(o)\,z(N+1) + W\underline{L}^{-1}\,\underline{\xi}(N) \tag{B-10}$$

Equation (B-6) is written as

$$\hat{X}(N+1) = \underline{P}\,\underline{\xi}(N+1) \tag{B-11}$$

Using Equation (B-10) one obtains

$$\hat{X}(N+1) = \underline{P}\ \underline{h}(o)\ z(N+1) + W\ \underline{P}\ \underline{L}^{-1}\ \underline{\xi}(N)$$

$$\hat{X}(N+1) = \underline{P}\ \underline{h}(o)\ z(N+1) + W\ \underline{P}\ \underline{L}^{-1}\ P^{-1}\hat{X}(N) \qquad \text{(B-12)}$$

We can show that the elements of the last terms of (B-12) can be written as

$$W\ \underline{P}\ \underline{L}^{-1}\ \underline{P}^{-1} = \underline{P}[\underline{P}^{-1} - h(o)\ \underline{h}^{T}(o)]\underline{L}^{T} \qquad \text{(B-13)}$$

As a result, we rewrite (B-12) as:

$$\hat{X}(N+1) = P\ h(o)\ z(N+1) + P\ [P^{-1} - \underline{h}(o)\ h^{T}(o)]\ L^{T}\ \hat{X}(N)$$

$$\hat{X}(N+1) = L^{T}\ \hat{X}(N) + P\ h(o)\ [z(N+1) - h^{T}(o)\ L^{T}\ \hat{X}(N)]$$

$$= L^{T}\ \hat{X}(N)\ P\ h(o)\ [z(N+1) - h^{T}(1)\ \hat{X}(N)] \qquad \text{(B-14)}$$

The model Equation (B-3) is written for prediction (forecasting) purposes as:

$$z(N+\ell) = \underline{h}^{T}(\ell)\ \underline{X} + v(N+\ell)$$

The optimal forecast based on N measurements is

$$\hat{z}(N+\ell) = \underline{h}^{T}(\ell)\ \hat{X}(N) \qquad \text{(B-15)}$$

As a result

$$\hat{X}(N+1) = L^{T}\hat{X}(N) + P\ \underline{h}(o)\ [z(N+1) - \hat{z}(N+1)] \qquad \text{(B-16)}$$

This recursive formula updates the parameter estimate by a term corresponding to the one-step ahead forecast error and is used in the seasonal forecasting model using trigonometric functions.

APPENDIX C
WINTERS' ADDITIVE PROCEDURE

Assume that the measurement model is given by Equation (3-1) written in the form:

$$z(N+j) = T(N+j) + S(N+j) + \nu(N+j) \tag{C-1}$$

The trend component is assumed to be the sum

$$T(N+j) = X_0(N) + X_1(N)j \tag{C-2}$$

There are s seasonal factors

$$S_{(i)} = S_{(i+s)} = S_{(i+2s)} = \qquad i=1,\ldots,s \tag{C-3}$$

$$\sum_{i=1}^{s} S_{(i)}=0 \tag{C-4}$$

An estimate of the level $X_0(N)$ can be obtained from a combination of two components $\hat{X}_{0,1}(N)$ and $\hat{X}_{0,2}(N)$. The first component is obtained as:

$$\hat{X}_{0,1}(N+1) = z(N+1) - \hat{S}(N+1-s) \tag{C-5}$$

This is the most recent observation adjusted by its seasonal factor $\hat{S}(N+1-s)$ which is available before $z(N+1)$ becomes known. The second component is

$$\hat{X}_{0,2}(N+1) = \hat{X}_0(N) + \hat{X}_1(N) \tag{C-6}$$

This uses the observations at N. Winters' procedure takes a weighted average of these two components to obtain

$$\hat{X}_0(N+1) = \alpha_1[z(N+1) - \hat{S}(N+1-s)] + (1-\alpha_1)[\hat{X}_0(N) + \hat{X}_1(N)] \tag{C-7}$$

Here α_1 is a smoothing constant chosen by a trial end error procedure.

To estimate the slope $\hat{X}_1(N)$ two components are used

$$\hat{X}_{1,1}(N+1) = \hat{X}_0(N+1) - \hat{X}_0(N) \tag{C-8}$$

$$\hat{X}_{1,2}(N+1) = \hat{X}_1(N) \tag{C-9}$$

Using a smoothing constant α_2, one has

$$\hat{X}_1(N+1) = \alpha_2 \left[\hat{X}_0(N+1) - \hat{X}_0(N)\right] + (1-\alpha_2)\,\hat{X}_1(N) \tag{C-10}$$

The estimates of the seasonal factors are based on two components:

$$\hat{S}_1(N+1) = z(N+1) - \hat{X}_0(N+1) \tag{C-11}$$

$$\hat{S}_2(N+1) = \hat{S}(N+1-s) \tag{C-12}$$

Using the smoothing factor α_3, one has

$$\hat{S}(N+1) = \alpha_3\left[z(N+1) - \hat{X}_0(N+1)\right] + (1-\alpha_3)\,\hat{S}(N+1-s) \tag{C-13}$$

The forecast of future values of the variable of interest is now obtained as:

$$\hat{z}(N+\ell) = \hat{X}_0(N) + \hat{X}_1(N)\ell + \hat{S}(N+\ell-s) \qquad \ell=1,2,\ldots,s \tag{C-14}$$

$$\hat{z}(N+\ell) = \hat{X}_0(N) + \hat{X}_1(N)\ell + \hat{S}(N+\ell-2s) \qquad \ell=s+1,\ldots,2s \tag{C-15}$$

APPENDIX D
WINTERS' MULTIPLICATIVE PROCEDURE

Here the measurement model is given by:

$$z(N+j) = [X_0(N) + X_1(N)j]\ S(N+j) + v(N+j) \tag{D-1}$$

Estimating the level $X_0(N)$ is based on finding the two components:

$$\hat{X}_{0,1}(N+1) = \frac{z(N+1)}{\hat{S}(N+1-s)} \tag{D-2}$$

$$\hat{X}_{0,2}(N+1) = \hat{X}_0(N) + \hat{X}_1(N) \tag{D-3}$$

Thus using the weighting (smoothing) parameter α_1, one has the estimate of the level given by:

$$\hat{X}_0(N+1) = \alpha_1 \frac{z(N+1)}{\hat{S}(N+1-s)} + (1-\alpha_1)[\hat{X}_0(N) + \hat{X}_1(N)] \tag{D-4}$$

For the slope $X_1(N)$, one has to find the two components:

$$\hat{X}_{1,1}(N+1) = \hat{X}_0(N+1) - \hat{X}_0(N) \tag{D-5}$$

$$\hat{X}_{1,2}(N+1) = \hat{X}_1(N) \tag{D-6}$$

As a result, using the smoothing parameter α_2, one obtains an estimate of the slope given by:

$$\hat{X}_1(N+1) = \alpha_2[\hat{X}_0(N+1) - \hat{X}_0(N)] + (1-\alpha_2)\hat{X}_1(N) \tag{D-7}$$

The seasonal factors are obtained from the two components:

$$\hat{S}_1(N+1) = \frac{z(N+1)}{\hat{X}_0(N+1)} \tag{D-8}$$

$$\hat{S}_2(N+1) = \hat{S}(N+1-s) \tag{D-9}$$

As a result, using the smoothing parameter α_3, we obtain the estimate of seasonal factor as:

$$\hat{S}(N+1) = \alpha_3 \frac{z(N+1)}{\hat{X}_0(N+1)} + (1-\alpha_3)\hat{S}(N+1-s) \tag{D-10}$$

The forecast of future values of the variable of interest is therefore given by:

$$\hat{z}(N+\ell) = [\hat{X}_0(N) + \hat{X}_1(N)\ell]\ \hat{S}(N+\ell-s) \qquad \ell=1,2,\ldots,s. \tag{D-11}$$

$$\hat{z}(N+\ell) = [\hat{X}_0(N) + \hat{X}_1(N)\ell]\ \hat{S}(N+\ell-2s) \qquad \ell=s+1,\ldots,2s. \tag{D-12}$$

Winters' procedures are called three parameter trend and seasonality methods since one has to search for the parameters α_1, α_2, and α_3 that yield best estimates of the constituent factors.

APPENDIX E
ELEMENTS OF BOX-JENKINS FORECASTING METHODOLOGY

E.1 Background

It is convenient in dealing with time series, to introduce the backward shift operator B, defined by the relation:

$$Bz(k) = z(k-1)$$

From the definition, one can see that a backward shift by m samples is obtained by B^m operating on the original series

$$B^m z(k) = z(k-m)$$

In addition, a differencing operation is denoted by ∇ such that

$$\nabla z(k) = z(k) - z(k-1)$$

It is clear that

$$\nabla z(k) = (1-B)\, z(k)$$

Higher order differencing is obtained from

$$\nabla^d z(k) = [1-B]^d\, z(k)$$

The most fundamental time series models are the autoregressive (AR) model and the moving average (MA) model. In the autoregressive model AR(p), the current value of the variable is expressed as a linear combination of p previous values and a random component $a(k)$ described by:

$$z(k) = \Phi_1 z(k-1) + \ldots. + \Phi_p z(k-p) + a(k)$$

This is written compactly as:

$$\Phi_p(B)\, z(k) = a(k) \qquad \text{(E-1)}$$

where

$$\Phi_p(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \ldots. - \Phi_p B^p$$

In the moving average model MA(q), the current value of the process is expressed as a linear combination of q previous random components

$$z(k) = a(k) - \theta_1\, a(k-1) - \ldots. - \theta_q\, a(k-q)$$

or $\quad z(k) = \theta_q(B)\, a(k) \qquad \text{(E-2)}$

where $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots. - \theta_q B^q$

The more general model, called the autoregressive moving average model ARMA (p,q) is a combination of the AR and MA models defined by:

$$\Phi_p(B)\, z(k) = \theta_q(B)\, a(k) \tag{E-3}$$

The ARMA model can be used to represent a stationary process with finite variance, and it is assumed that the roots of $\Phi(B)$ and $\theta(B)$ lie outside the unit circle of the B-plane to obtain a stable time series.

One can model some types of non-stationary processes $z(k)$ by differencing the original process $z(k)$ to obtain a stationary process $w(k)$

$$w(k) = \nabla^d z(k)$$

The result is the autoregressive integrated moving average model ARIMA (p,d,q)

$$\Phi_p(B)\, \nabla^d z(k) = \theta_q(B)\, a(k) \tag{E-4}$$

Because of the periodic nature of the load curve, it is advantageous to use seasonal ARIMA models $(p,d,q) \times (P,D,Q)_s$, with seasonalities:

$$\Phi_p(B)\, \Phi_P(B^s)\, \nabla^d \nabla_s^D z(k) = \theta_q(B)\, \theta_Q(B^s)\, a(k) \tag{E-5}$$

In some cases, it is also useful to recognize another periodicity (such as weekly), and use a two period ARIMA $(p,d,q) \times (P,D,Q)_s \times (P',D',Q')_{s'}$ model defined by;

$$\Phi_p(B)\, \Phi_P(B^s)\, \Phi_{P'}(B^{s'})\, \nabla^d \nabla_s^D \nabla_{s'}^{D'} z(k) = \theta_q(B)\, \theta_Q(B^s)\, \theta_{Q'}(B^{s'})\, a(k) \tag{E-6}$$

Seasonal ARIMA models are used in this research for load forecasting in the short-term and mid-term problems. In the following section a synopsis of Box-Jenkins forecasting methodology is presented.

E.2 The Process

Preliminary identification of the appropriate time series model relies on the analysis of the autocorrelation function (ACF) and partial autocorrelation function (PACF). The method is systematic and very useful in the determination of the model order in the preliminary estimation of the model parameters and in diagnostic checking and model refinement.

ACF

The autocorrelation function (ACF) describes the inherent correlations between observations of a time series which are separated in time by some lag j

$$\gamma_j = E\,[z(k)\,z(k+j)] \tag{E-7}$$

$$\rho_j = \gamma_j/\gamma_0 \tag{E-8}$$

For a white noise process, in which there is no correlation in time, the ACF would be zero for all lags except at j=0.

For an AR process, the autocorrelation function satisfies the following equation

$$\Phi_p(B)\,\rho_k = 0 \tag{E-9}$$

$$\rho_k = \Phi_1\,\rho_{k-1} + \ldots + \Phi_p\,\rho_{k-p} \tag{E-10}$$

If we write

$$\Phi(B) = \prod_{i=1}^{p}(1-G_iB),$$

the general solution of (E-10) is

$$\rho_k = A_1G_1^k + A_2G_2^k + \ldots + A_pG_p^k$$

where G_1^{-1}, G_2^{-1} are the roots of the characteristic equation

$$\Phi(B) = 1 - \Phi_1B - \Phi_2B^2 - \ldots - \Phi_pB^p = 0$$

For the process to be stationary we require that, $|G_i| < 1$. Assuming that the roots are distinct, then the autocorrelation function of a stationary AR series will consist of a mixture of damped exponentials and damped sine waves.

Estimating the autoregressive parameters in terms of the autocorrelation can be done using the Yule-Walker equations [60-61]. These are obtained from (E-10) by setting k=1,2,....,p to obtain a set of linear equations

$$\underline{\Phi} = \underline{P}_p^{-1}\,\underline{\rho}_p \tag{E-11}$$

Here one has

$$\underline{\Phi} = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_p \end{bmatrix} \qquad \underline{\rho}_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \qquad \underline{P}_p = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & & \rho_{p-2} \\ \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & & 1 \end{bmatrix}$$

Note that the model order p must be known a priori in order to carry out the computation.

PACF

The determination of the model order p of an autoregressive model is facilitated by a device called the partial autocorrelation function. Briefly, the partial autocorrelation function (PACF) Φ_{kk} is zero for k greater than the model order.

The PACF is defined by considering an AR(k) process, so that equation (E-10) can be written as;

$$\rho_j = \Phi_{k1}\,\rho_{j-1} + \ldots + \Phi_{k(k-1)}\,\rho_{j-k+1} + \Phi_{kk}\,\rho_{j-k} \qquad j=1,2,\ \ldots\ k \tag{E-12}$$

Note that Φ_{kj} is the j^{th} coefficient in AR(k) so that Φ_{kk} is the last coefficient. The quantity Φ_{kk} regarded as a function of the lag k, is the partial autocorrelation function. Writing (E-12) in the Yule-Walker form one has

$$\underline{P}_k\,\underline{\Phi}_k = \underline{\rho}_k \tag{E-13}$$

The PACF is the last element in $\underline{\Phi}_k$. The PACF of a p^{th} order AR process has a cut-off after lag p.

There is a duality between AR and MA processes, while the ACF of an AR(p) process is infinite in extent, the PACF cuts off after a lag p. The ACF of an MA(q) process cuts off after a lag q, while the PACF is infinite in extent.

The ACF and PACF of a mixed ARMA(p,q) process are both infinite in extent and tail off (die down) as the lag k increases. Eventually (for k > q-p), the ACF is determined from the autoregressive part of the model. The PACF is eventually (for k>p-q) determined from the moving average part of the model.

Identification

From the foregoing discussion, it is clear that the ACF and PACF allow one to determine potentially useful model structures. They are also very useful in determining appropriate model adjustments when diagnostic checks indicate model inadequacy. Figure (E-1) illustrates the theoretical correlation functions for non-seasonal AR processes, and Figure (E-2) is the counterpart for seasonal processes. Figure (E-3) illustrates the theoretical correlation functions for non-seasonal MA processes and that for seasonal MA processes are shown in Figure (E-4).

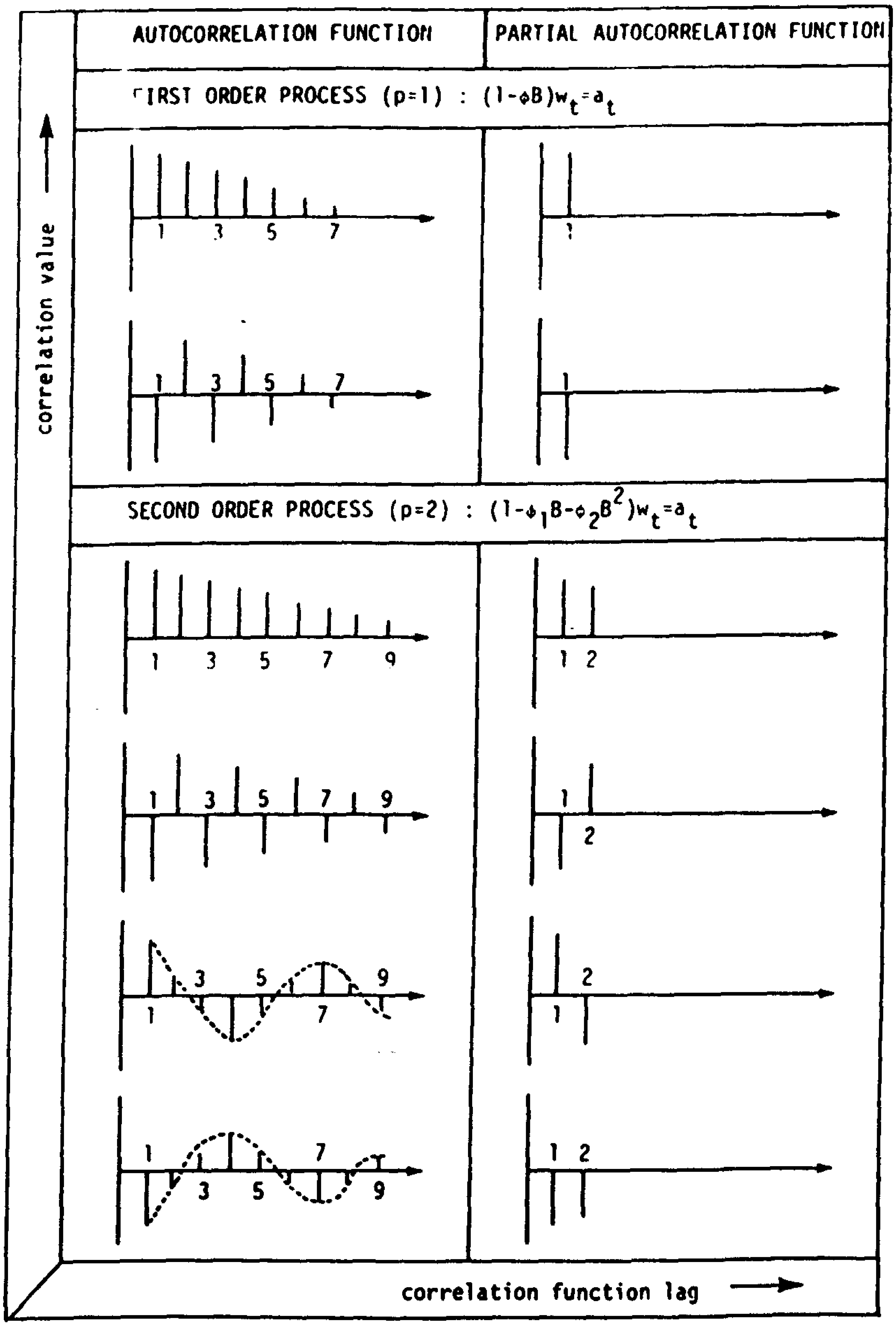


Figure (E-1)
Theoretical Correlation Functions
of Non-Seasonal Autoregressive Processes

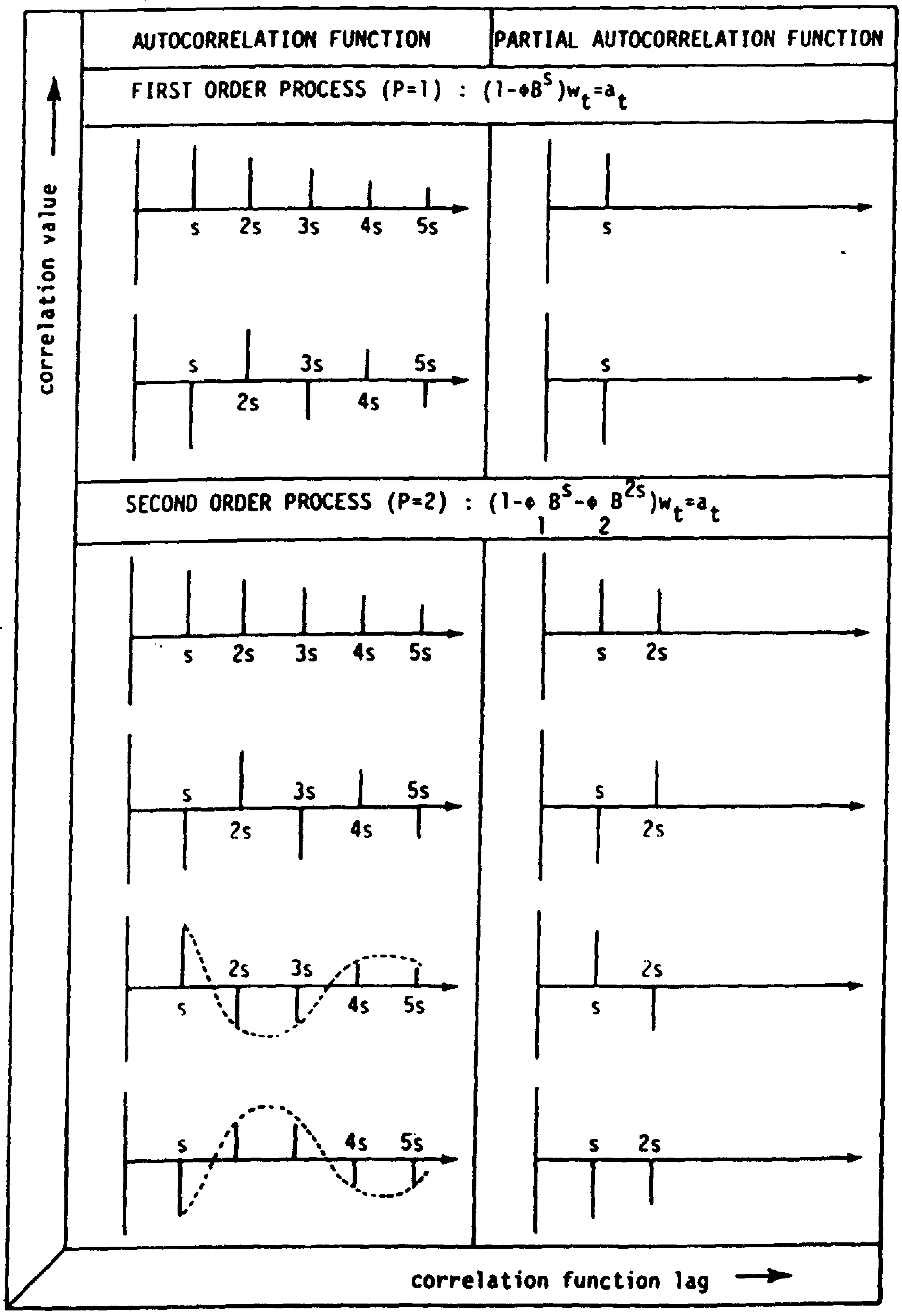


Figure (E-2)
Theoretical Correlation Functions
of Seasonal Autoregressive Processes

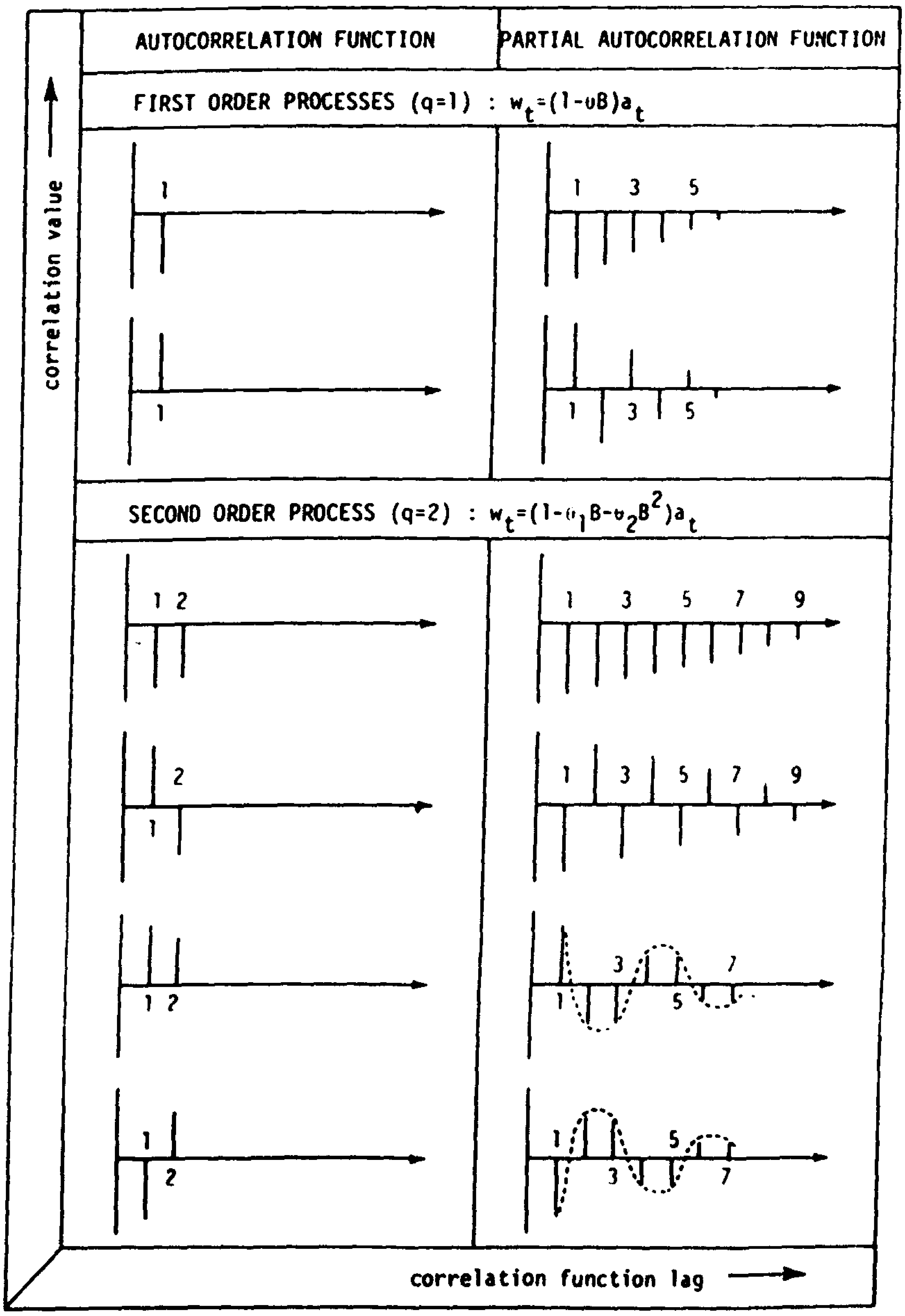


Figure (E-3)
Theoretical Correlation Functions
of Non-Seasonal Moving Average Processes

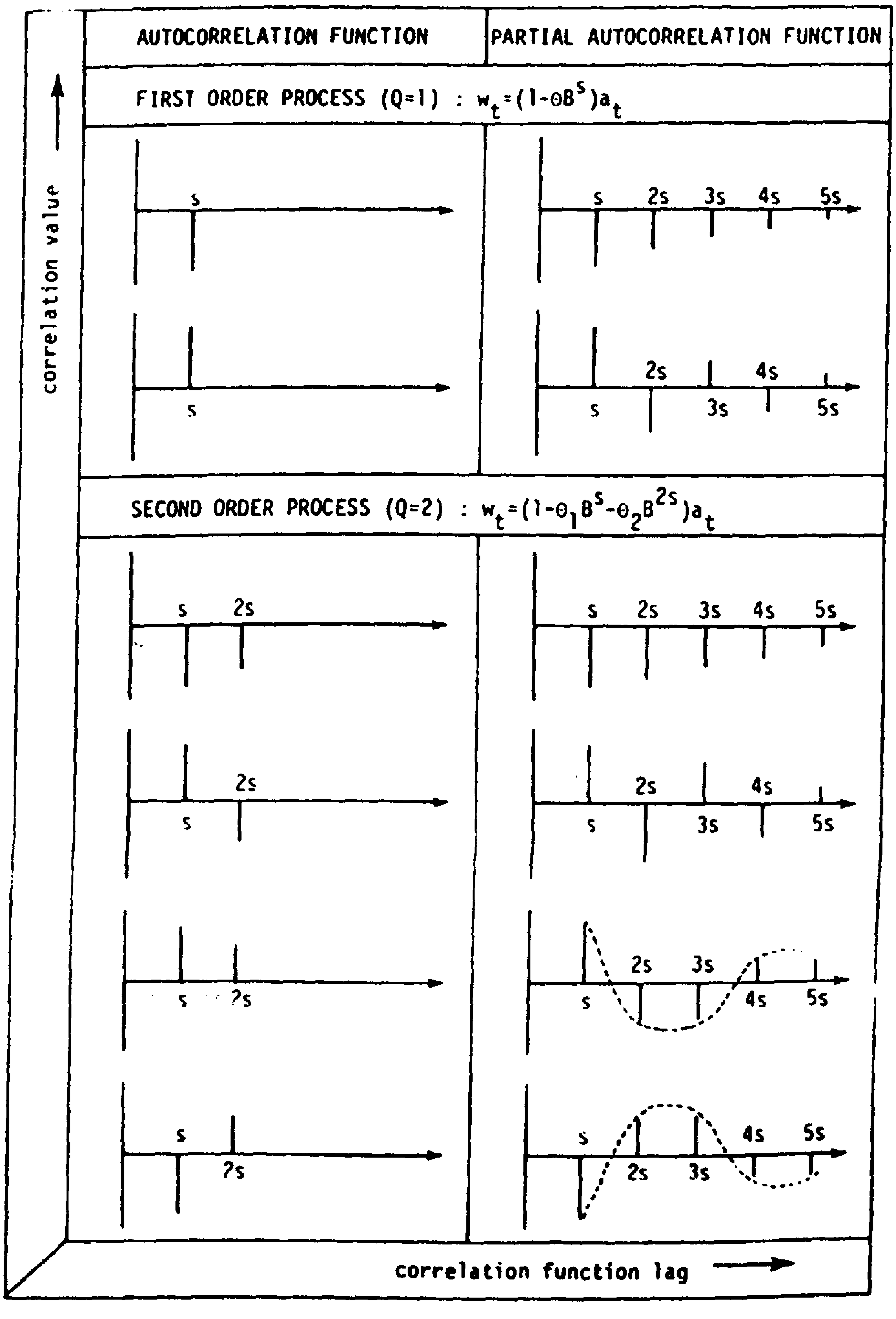


Figure (E-4)
Theoretical Correlation Functions
of Seasonal Moving Average Processes

The foregoing discussion is based on the assumption that the time series is stationary. If the random series $a(k)$ is normally distributed then the process is stationary if its mean, variance and autocorrelation are constant. In practice, many time series are nonstationary. Fortunately, nonstationary time series can sometimes be transformed into stationary time series. If the mean changes with time such that different parts of the series behave in a similar fashion except for changes in level and slope, then the series can be transformed to a stationary series by differencing. Some time series have a variance that changes with time and in such cases a constant variance may be obtained by using natural logarithms. It is possible that no suitable transformation will be found [64].

Estimation

The second stage of the Box-Jenkins methodology is that of estimation. Here the task is to fit the tentative model (obtained in the identification stage) to the data to get precise estimates of the model parameters. Here the coefficient values must be chosen in accordance with a certain criterion defining optimality. From an information content point of view, estimating the coefficients should be done so that a maximum likelihood (ML) criterion is minimized. From a practical point of view, finding ML estimates involves considerable computational difficulties. Observing that when the random components are normally distributed, then least squares (LS) estimates coincide with ML estimates, thus one prefers LS estimates in practice. In this case one seeks the model parameters that minimize the sum of squares of the residuals. The residual is an estimate of the random component defined as the difference between the observed value $z(k)$ and a computed value $\hat{z}(k)$. Linear least squares may be used to estimate parameters of only pure AR models without multiplicative seasonal terms. All other models require a non-linear least squares (NLS) iterative solution method.

Solving NLS parameter estimation problems relies on iterative algorithms that start with an initial guess of the unknown parameters and progressively improve on the solution until no further improvements are possible. There are many iterative procedures for solving NLS problems, but the most commonly used algorithms are Gauss-Newton (GN) Linearization and the Gradient Method. A favoured algorithm is the Lenvenburg- Marquardt method [65], which is a compromise between GN and the gradient methods.

Diagnostic Checking

The results of the estimation stage may be used to check the model for stationarity and invertibility as part of the third stage referred to as diagnostic checking.

The main object of the diagnostic checking stage is to verify the statistical adequacy of the model as completely specified through the identification and estimation stages. When diagnostic checking shows a model to be inadequate, one must return to the identification stage to tentatively select one or more other models. Moreover, diagnostic checking is extremely helpful in providing clues about how an inadequate model might be reformulated.

The basic assumption in ARIMA models is that $a(k)$ are uncorrelated random variables with zero mean and constant variance. Thus we would expect the behaviour of the residuals $\hat{a}(k)$ to be similar to that of the errors $a(k)$ for a long time series. One thus requires:

1- The mean of the residuals should be close to zero.
2- The variance of the residuals should be approximately constant.
3- The autocorrelations of the residuals should be negligible.

To check whether the mean of the residuals is zero, we can compare the sample mean $\bar{\hat{a}}$ with its standard error. To check whether the variance is constant, we examine the residuals. To check whether the residuals are uncorrelated, we compute the sample autocorrelations defined below and compare them with their standard errors:

$$r_{\hat{a}}(j) = \frac{\sum_{k=j+1}^{N} [\hat{a}(k) - \bar{\hat{a}}][\hat{a}(k-j) - \bar{\hat{a}}]}{\sum_{k=1}^{N} [\hat{a}(k) - \bar{\hat{a}}]^2} \tag{E-14}$$

The standard error of $r_{\hat{a}}(j)$ is usually approximated as [59]:

$$s[r_{\hat{a}}(j)] \cong N^{-\frac{1}{2}} \tag{E-15}$$

For small j, the true standard error can be much smaller.

An alternative test statistic is based on all residual autocorrelations as a set. This is defined by:

$$Q^{*} = N(N+2) \sum_{j=1}^{K} [N-j]^{-1} r_{\hat{a}}^{2}(j) \tag{E-16}$$

where N is the number of observations used to estimate the model. The Q' approximately follows a chi-squared distribution with (K-m) degrees of freedom, where m is the number of parameters. If Q' is large, then the residual ACF's as a set are significantly different from zero and one should then consider reformulating the model.

The residuals from a fitted model present a new time series that can be studied in the same manner as one did with the original series. Visual analysis of the plot of the residuals can be helpful in detecting unusual events that impact a time series, detecting data errors, and detecting problems with the fitted model.

An additional diagnostic check is that of overfitting. Here one adds another coefficient to see if the resulting model is better.

A final diagnostic test often cited is to divide the data set into subsets and estimate the model for each subset. This is useful if the model parameters are time varying.

Model Reformulation

When diagnostic checking reveals that a model is statistically inadequate, one then must return to the identification stage to tentatively select one or more other models. There are a number of possible ways to reformulate an apparently inadequate model, as discussed here.

Re-examining the estimated ACF and PACF computed from the original series might suggest one or more alternative models that did not initially seem obvious. Alternatively, one uses the residual ACF as a guide. This allows one to model the residuals as an implicit time series and therefore modify the original model.

Forecasting

The final outcome of the process is to forecast future values of a time series on the basis of an ARIMA model. Consider the process described by equation (E-4) rewritten as:

$$z(k) = \frac{\theta_q(B)}{\Phi_p(B)\nabla^d} a(k) \qquad \text{(E-17)}$$

One can define the function $\Pi(B)$ by:

$$\Pi(B) = \frac{\Phi_p(B)\ (1-B)^d}{\theta_q(B)} \tag{E-18}$$

It is assumed that one can expand $\Pi(B)$ as:

$$\Pi(B) = 1 - \Pi_1 B - \Pi_2 B^2 - \ldots.$$

As a result we have

$$\left[1 - \Pi_1 B - \Pi_2 B^2 - \ldots. \right] z(k) = a(k)$$

Therefore, an ARIMA model can be written in an equivalent AR form:

$$z(k) = \Pi_1 z(k-1) + \Pi_2 z(k-2) + \ldots. + a(k) \tag{E-19}$$

An alternative formulation is obtained by defining

$$\Psi(B) = \frac{\theta_q(B)}{\Phi_p(B)(1-B)^d} \tag{E-20}$$

As a result

$$z(k) = \Psi(B)\ a(k) \tag{E-21}$$

One assumes that Ψ is given by the expansion

$$\Psi(B) = 1 + \Psi_1 B + \Psi_2 B^2 + \ldots. \tag{E-22}$$

Thus the ARIMA model can be written in the equivalent MA form:

$$z(k) = a(k) + \Psi_1 a(k-1) + \ldots. \tag{E-23}$$

In forecasting, one is concerned with finding an optimum value of $z(N+\ell)$ denoted by $\hat{z}_N(\ell)$, $\ell \geq 1$, when one is at time N, such that the minimum of mean square error (MSE) is achieved. Here the mean square error is defined by:

$$MSE = E\left[z(N+\ell) - \hat{z}_N(\ell) \right]^2 \tag{E-24}$$

Using the MA form (E-23) we can show that

$$\hat{z}_N(\ell) = \Psi_\ell\ a(N) + \Psi_{\ell+1} a(N-1) + \ldots. \tag{E-25}$$

The forecast error is

$$\begin{aligned} e_N(\ell) &= z(N+\ell) - \hat{z}_N(\ell) \\ &= a(N+\ell) + \Psi_1\ a(N+\ell-1) + \ldots. + \Psi_{\ell-1}\ a(N+1) \end{aligned} \tag{E-26}$$

and its variance is given by

$$V[e_N(\ell)] = \sigma^2 \left[1 + \Psi_1^2 + \Psi_2^2 + \dots + \Psi_{\ell-1}^2 \right] \qquad \text{(E-27)}$$

If the random components are normally distributed and if one has an appropriate model, then the forecasts and the associated forecast errors are approximately normally distributed.

APPENDIX F
APPROACHES TO ADAPTIVE FILTERING

This Appendix is devoted to a review of the major techniques used in adaptive filtering discussed in Section 3.8.

A. Bayesian Based Approach

This approach is sometimes referred to as the multiple model filtering algorithm originated in the work of Magill [81]. One basic assumption is that the measurements $z(k)$ are formed from a number L of elemental stochastic processes $\{z_i(k) : k=1,2, \ldots, i=1,2,\ldots, L\}$. The measurement sequence is obtained by connecting a switch to one of L possible positions as shown in Figure (F-1). Let α_j denote that the switch is in position j. One assumes that the a priori probabilities $\{P(\alpha_i) : i=1,2, \ldots ,L\}$ of the switch being in each of the L positions are known.

The optimal Bayesian estimator $\hat{\omega}$ of some state ω can be shown [81] to be given by

$$\hat{\omega} = \sum_{A} \hat{\omega}(\alpha_i)\, \rho\left(\alpha_i \middle| Z(k)\right)$$

Here A is the space of all α_i, $\rho\left(\alpha_i \middle| Z(k)\right)$ is the conditional probability density function of α_i given the data vector $Z(k)$. The optimal estimate $\hat{\omega}(\alpha_i)$ is the optimal estimate conditional on α_i. This is the central result used in the Bayesian approach.

The optimal estimate is formed by taking the complete set of conditional estimates, weighting each with the conditional probability that the appropriate parameter vector is true and summing over the space of all possible parameter values. The algorithm essentially uses a number (L) of parallel Kalman filters whose optimal estimates $\hat{X}_+(k|\alpha_i)$ are weighted by $\rho\left(\alpha_i \middle| Z(k)\right)$. The optimal estimate is the sum shown in Figure (F-2):

$$\hat{x}_+(k) = \sum_{i=1}^{L} \hat{x}_+(k|\alpha_i)\, \rho\left(\alpha_i \middle| Z(k)\right) \tag{F-1}$$

The weighting coefficients are given by

$$\rho\left(\alpha_i \middle| Z(k)\right) = \frac{\rho\left(Z(k) \middle| \alpha_i\right) \rho(\alpha_i)}{\sum_{j=1}^{L} \rho\left(Z(k) \middle| \alpha_j\right) \rho(\alpha_j)} \tag{F-2}$$

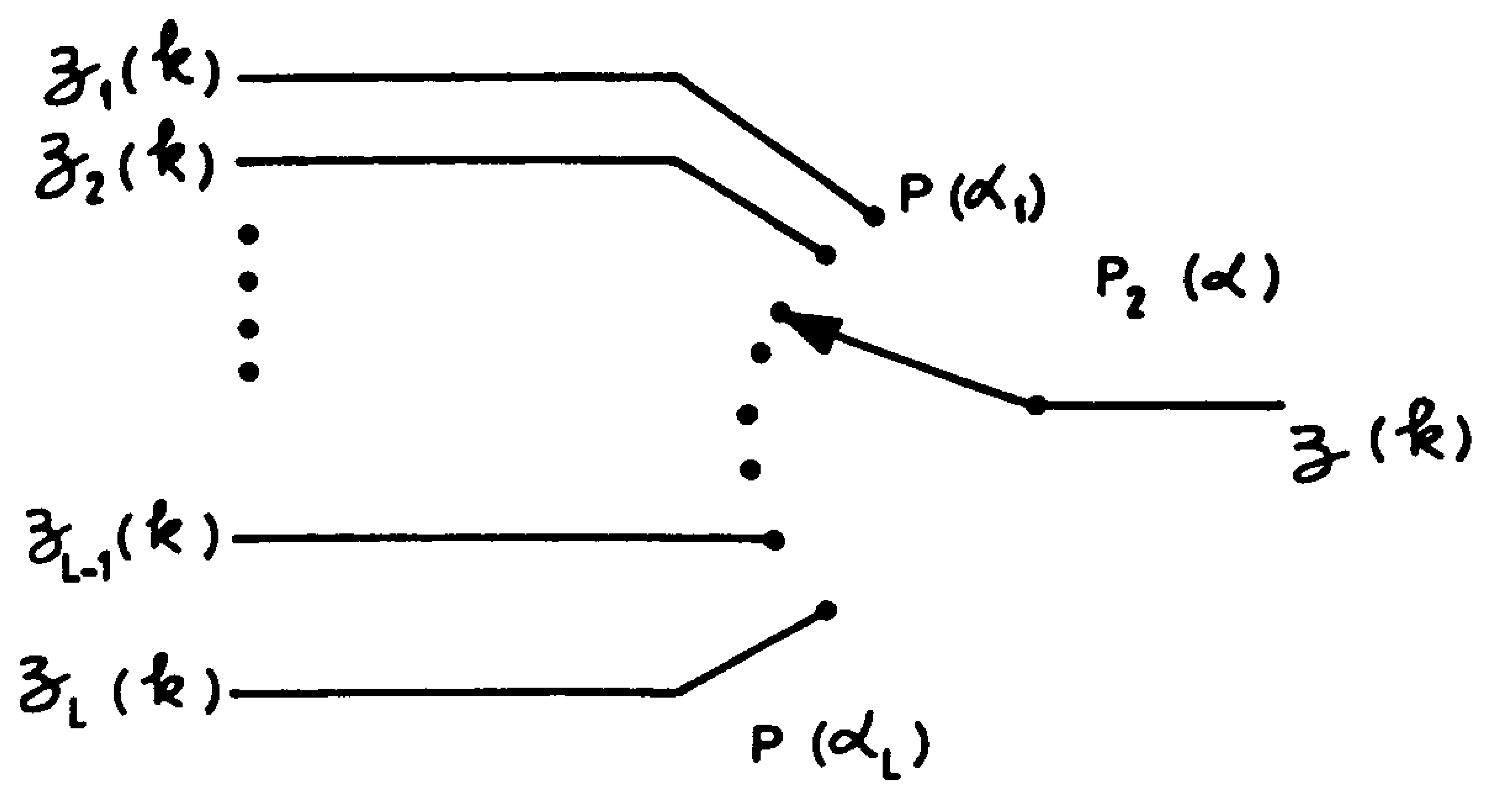


Figure (F-1) Bayesian Interpretation of Measurement Sequence

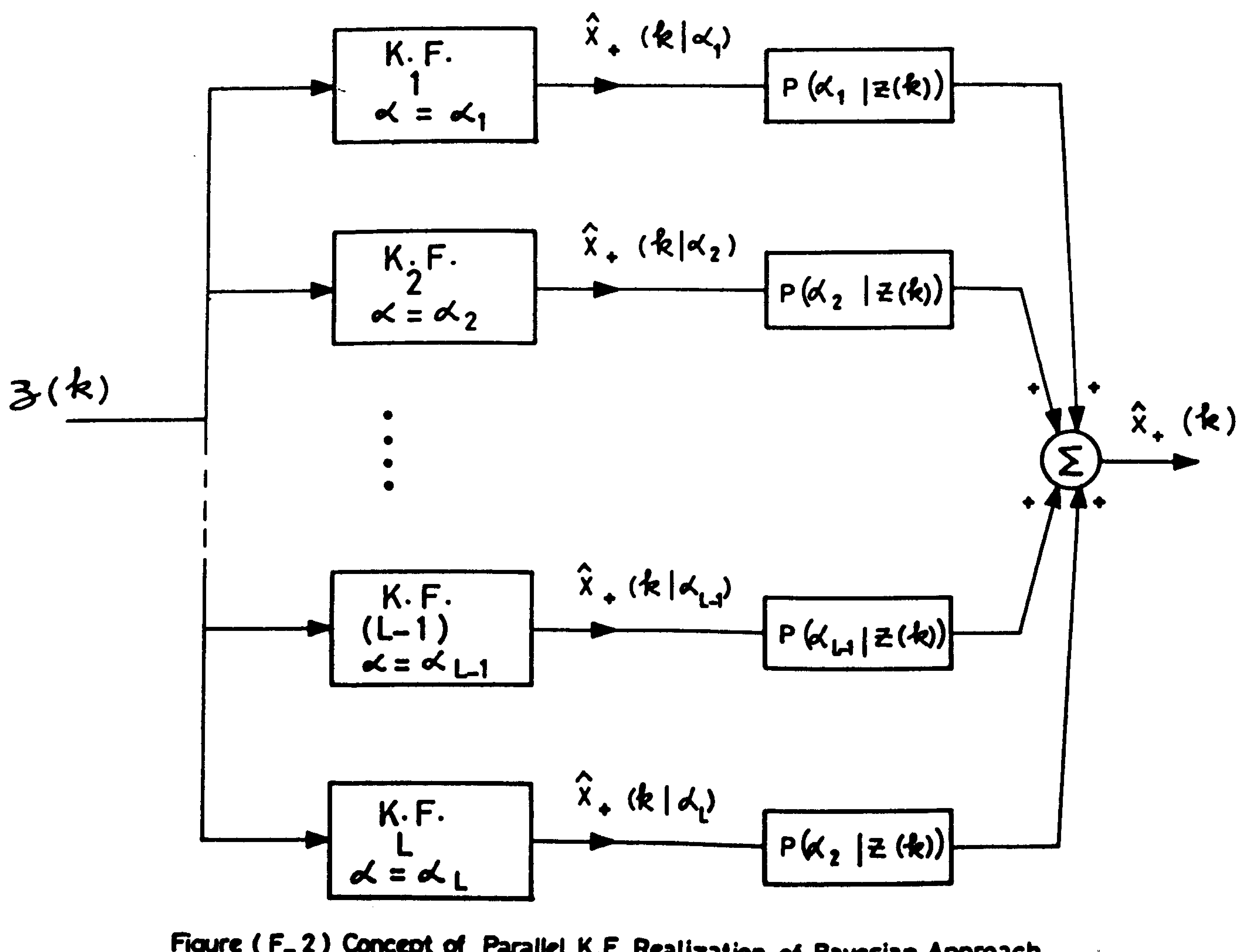


Figure (F-2) Concept of Parallel K.F. Realization of Bayesian Approach

In [82], Smith obtains results that relax the assumption that the distributions are known. The distributions are assumed normal but with unknown variances. Hilborn and Lainiotis [83] extend Magill's solution to the vector case. In [84] an improved recursive algorithm is proposed by Sims and Lainitios that offers computational and memory savings compared to Magill's original proposal. The case of time varying parameters is considered in Hilborn and Lainiotis [85]. The general problem concerning simultaneous adaptation of system structure and parameters as well as adaptation of system structure alone is treated in [86]. Time varying statistics are considered in Alspach's work [87]. The asymptotic properties of Bayesian estimates and convergence characteristics are given in [88].

B. Maximum Likelihood - Based Approaches

The maximum likelihood approach relies on maximizing a conditional probability density function such as $f\left(X(k), Z(k) \mid a\right)$, where a is the vector of uncertain elements. By applying Bayes' rule, the likelihood function is expressed as the following product:

$$f\left(X(k), Z(k) \mid a\right) = f\left(X(k) \mid Z(k), a\right) \prod_{j=1}^{k} f\left(z(j) \mid Z(j-1), a\right) \qquad \text{(F-3)}$$

It is usual to work with the log likelihood function for the case of Gaussian distributions which turns out to be

$$\begin{aligned} Ln\ f(X, Z \mid a) = & \frac{-(n+km)}{2}\ Ln\ (2\Pi) \\ & - \frac{1}{2}\ Ln\ (|P_+(k)|) \\ & - \frac{1}{2} \sum_{j=1}^{k} Ln\ (|A(j)|) \\ & - \frac{1}{2} [\xi - \hat{x}_+(k)]^T P_+^{-1}(k) [\xi - \hat{x}_+(k)] \\ & - \frac{1}{2} \sum_{j=1}^{k} [\xi(j) - H(j)\hat{X}_-(j)]^T A^{-1}(j) \\ & [\xi(j) - H(j)\hat{X}_-(j)] \end{aligned} \qquad \text{(F-4)}$$

The requirement is to determine the optimal state X^* and optimal parameter vector a^* to maximize the log likelihood function.

The maximization process consists of solving two sets of equations. The first, which is the result of setting the derivative of the log likelihood function with respect to ξ (the dummy variable associated with X) to zero, results in asserting the optimality of the Kalman Filter, provided that α is set to the optimal value of the parameter vector a^*. The second set of equations is related to taking derivatives with respect to α. The resulting set of equations is non-linear and is called the likelihood equations, or score vector [78].

For off-line applications, Newton-Raphson's iterative method is used and in this case an approximation to the negative of the Jacobian is termed the conditional information matrix. For on-line applications several approximating assumptions are made [78] to arrive at a fixed memory algorithm of length N to compute an estimate of R given by

$$\hat{R}(k) \triangleq \left[\frac{1}{N} \sum_{j=k-N+1}^{k} r(j)\, r^T(j) \right] - H(k)\, P_-(k)\, H^T(k) \qquad \text{(F-5)}$$

It should be recognized that R must be positive definite. An alternative result that guarantees a positive definite estimate of R is given by

$$\hat{R}(k) = \frac{1}{N} \sum_{j=k-N+1}^{k} \{ r_+(j)\, r_+^T(j) + H(j) P_+(j) H^T(j) \} \qquad \text{(F-6)}$$

where

$$r_+(j) = z(j) - H(j)\, \hat{X}_+(j) \qquad \text{(F-7)}$$

This last expression requires more computations than that required by (F-5).

Obtaining an estimate of Q_d is much more involved, but in the steady state with an invertible G_d, we have [78]

$$Q_d(k) = \frac{1}{N} \sum_{j=k-N+1}^{k}$$

$$\left\{ G_d^{-1}(j-1) [\Delta X(j)\, \Delta X^T(j) + P_+(j) - \Phi(j,j-1)\, P_+(j-1)\, \Phi^T(j,j-1)] G_d^{-1^T}(j-1) \right\} \qquad \text{(F-8)}$$

with

$$\Delta X(j) = \hat{X}_+(j) - \hat{X}_-(j) \qquad \text{(F-9)}$$

The estimation results cited in the preceding discussion is based on developments reported in Maybeck [78]. It is appropriate to mention alternative approximations cited in the literature.

In 1968, Abramson [89] derives formula for diagonal Q and R given by

$$Q_d^{ii}(k) = \frac{n-1}{n} Q_d^{ii}(k-1) + \frac{1}{n}[\Delta X(k)\, \Delta X^T(k) - P_+(k) - \Phi(k)\, P_+(k-1) \Phi^T]^{ii} \qquad \text{(F-10)}$$

$$R^{ii}(k) = \frac{n-1}{n} R^{ii}(k-1) + \frac{1}{n}[r_+(k)\, r_+^T(k) + H(k) P_+(k) H^T(k)]^{ii} \qquad \text{(F-11)}$$

Abramson's results are based on a conditional probability density function of the form $f\left(X(k), z(k) \mid Q_d, R \right)$.

Shellenbarger [90] maximizes the likelihood based on the conditional probability density function $f\left(X(k), z(k) \mid Q_d, R \right)$. Holding Q_d constant, Shellenbarger's estimate of R is given by

$$\hat{R}(k) = r(k)\, r^T(k) - H(k)\, P_-(k)\, H^T(k) \tag{F-12}$$

In 1969, Sage and Husa [91] present a suboptimal approximation using the discrete minimum principle. Here one has the following results

$$\hat{Q}(k) = \frac{1}{k}\left[(k-1)\hat{Q}(k-1) + K(k)r(k)r^T(k)K^T(k) + P_+(k) - \Phi(k,k-1)P_+(k-1)\Phi^T(k,k-1)\right] \tag{F-13}$$

$$\hat{R}(k) = \frac{1}{k-1}\left[(k-2)\, \hat{R}(k-1) + r(k)\, r^T(k) - H(k-1)\, P_-(k-1)\, H^T(k-1)\right] \tag{F-14}$$

It is important to note that Sage and Husa's results were given for non-zero mean noise terms.

C. Correlation Based Approaches

It has been pointed out earlier that correlation methods play an important role in time series analysis. It is therefore no surprise that this technique would be valuable in adaptive filtering when second-order statistics of the system and measurement noises are the only unknowns. There are two variants to correlation methods:

A- Output Correlation Method.

B- Innovation Correlation Methods.

The estimates obtained from the second method are more efficient than those obtained from the first method since the innovations $r(k)$ are less correlated than the output $z(k)$.

C.1 The Output Correlation Method

This method is reported in Mehra [76 & 92], and consists of obtaining estimates of correlation of the output using

$$\hat{C}(k) = \frac{1}{N} \sum_{i=k}^{N} z(i)\, z^T(i-k) \tag{F-15}$$

The sequence of matrices M(j) is formed by

$$M(j) = H \Phi^{j} \tag{F-16}$$

An optimal estimate of a matrix S is obtained as

$$\hat{S} = \left[\sum_{j=1}^{n} M^{T}(j)\, M(j) \right]^{-1} \left[\sum_{j=1}^{n} M^{T}(j)\, C(j) \right] \tag{F-17}$$

Note that

$$S = \S\, H^{T} \tag{F-18}$$

with

$$\S = E\{x(i)\, x^{T}(i)\} \tag{F-19}$$

The estimate of R is obtained as

$$\hat{R} = C(0) - H\hat{S} \tag{F-20}$$

To find an estimate of K, we solve the following algebraic Riccati equation in the matrix

$$\hat{\Pi} = \Phi\left[\hat{\Pi} + (\hat{S} - \hat{\Pi} H^{T}) (C(0) - H \hat{\Pi} H^{T})^{-1} (\hat{S}^{T} - H \hat{\Pi})\right]\Phi^{T} \tag{F-21}$$

The optimal estimate of the gain matrix K is thus

$$\hat{K} = [\hat{S} - P H^{T}]\, [C(0) - H \hat{\Pi} H^{T}]^{-1} \tag{F-22}$$

C.2 The Innovation Correlation Method

This method is based on using the autocorrelation of the filter's residuals (innovations) and is given in Mehra's 1970 paper [93], and in summary form in Mehra's review of 1972 [76], and Maybeck [78].

The method performs a correlation test on the observed residuals r(i) of the filter to determine statistically whether adaptation is required or not. If so, asymptotically Gaussian unbiased and consistent estimates of R are generated. If the number of unknowns in Q_d is less than (n m), then an estimate of Q_d can be obtained. Otherwise the steady state gain of the filter is estimated without an explicit estimate of Q_d.

It is important to note that this method is developed for the case when Φ, H and G_d of the system and measurement models have constant parameters.

The filter is initialized with a guess of $Q^{(0)}$ and $R^{(0)}$. From the K.F. result one computes the residual (innovation) sequence for i=1, ... ,N, given by:

$$r(i) = z(i) - H\hat{X}_{-}(i)$$

An estimate of the k-lag autocorrelations of the residuals is computed using

$$\hat{A}_k = \frac{1}{N}\sum_{i=k+1}^{N} r(i)\, r^T(i-k) \tag{F-23}$$

If the correlation test reveals that $R^{(0)}$ and $Q^{(0)}$ are not optimal, then a best estimate of $P^- H^T$ can be computed using

$$\hat{S} = \left[\sum_{k=1}^{n} M_k^T M_k\right]^{-1} \sum_{k=1}^{n} M_k^T \tilde{A}_k \tag{F-24}$$

with $M_k = H\,\Phi^k$ (F-25)

$$\tilde{A}_k = \hat{A}_k + \sum_{i=0}^{k-1} M_{k-i}\, K\, \tilde{A}_i \tag{F-26}$$

In (F-24) we define

$$S = P^- H^T \tag{F-27}$$

Here P^- is the steady state error covariance matrix.

An estimate of R denoted by $\hat{R}$ is now obtained as

$$\hat{R} = \hat{A}_0 - H\hat{S} \tag{F-28}$$

Note that if $\hat{R} = R^{(0)}$, the filter may be considered optimal.

To obtain an improved error covariance matrix P_-^*, we have to find $\hat{\Delta}$ to satisfy the following non-linear equation

$$\hat{\Delta} = \Phi\left[\hat{\Delta} - [\hat{S} + \hat{\Delta}H^T][\hat{A}_0 + H\hat{\Delta}H^T]^{-1}[\hat{S}^T + H\hat{\Delta}] + C\right]\Phi^T \tag{F-29}$$

Here $C = K\hat{S}^T + \hat{S}K^T - K\hat{A}_0K^T$ (F-30)

The improved value of P_- denoted by P_-^* is now given by

$$P_-^* = P_- + \hat{\Delta} \tag{F-31}$$

The improved gain matrix K^* is obtained from

$$K^* = [\hat{S} + \hat{\Delta}H^T]\,[\hat{A}_0 + H\hat{\Delta}H^T]^{-1} \tag{F-32}$$

To obtain an estimate of $G_d\, Q_d\, G_d^T$, we have

$$G_d\, Q_d\, G_d^T = P_-^* - \Phi[P_-^* - K^*HP_-^*]\Phi^T \tag{F-33}$$

A number of variations to the basic innovation correlation method exist. In [94], Carew and Belanger propose an alternative that uses a contraction mapping algorithm to solve for a matrix P^*. The steady state matrix P^* is defined by

$$P^* = E\left\{ [\hat{x}_-(i) - x^*_-(i)][\hat{x}_-(i) - x^*_-(i)]^T \right\} \qquad \text{(F-34)}$$

where x^* is the estimate from the suboptimal filter.

Let us denote the primary unknown P^* by X

$$X = P^* \qquad \text{(L-35)}$$

The algorithm begins by a suitable initial guess of X, say X_m, and calculates

$$W(X_m) = A(0) - H X_m H^T \qquad \text{(F-36)}$$

$$\Gamma(X_m) = \left\{ [\tilde{M}^T\tilde{M}]^{-1} \tilde{M}^T\tilde{A} - \Phi X_m H^T \right\} W^{-1}(X_m) \qquad \text{(F-37)}$$

The new improved estimate is thus

$$X_{m+1} = T(X_m) \qquad \text{(F-38)}$$

where

$$T(X) = [\Phi - \Gamma_D H]\, X\, [\Phi - \Gamma_D H]^T + [\Gamma_D - \Gamma(X)]\, W(X)[\Gamma_D - \Gamma(X)]^T \qquad \text{(F-39)}$$

The following matrices are used

$$\Gamma = \Phi K \longrightarrow \text{Optimum Filter Gain} \qquad \text{(F-40)}$$

$$\Gamma_D = \Phi K_D \longrightarrow \text{Suboptimum Filter Gain} \qquad \text{(F-41)}$$

The subscript D denotes a quantity for a suboptimum filter.

The matrix $\tilde{M}$ is defined by

$$\tilde{M}^T = \left[H^T \; (H\Phi)^T \; \ldots (H\Phi^{n-1})^T \right] \qquad \text{(F-42)}$$

The matrix $\tilde{A}$ is defined by

$$\tilde{A} = \begin{bmatrix} A(1) + H\Gamma_D A(0) \\ A(2) + H\Gamma_D A(1) + H\Phi\Gamma_D C(0) \\ \vdots \\ A(n) + H\Gamma_D A(n-1) + \ldots + H\Phi^{n-1}\Gamma_D C(0) \end{bmatrix} \qquad \text{(F-43)}$$

Here $A(k)$ is the approximate correlation of the innovation sequence.

The algorithm is shown [94] to converge uniquely to a solution P^* and thus the matrix Γ and hence the optimal filter gain K can be obtained. In contrast to the basic innovation correlation method which requires solving the algebraic Riccati equation (F-29) at every step, the Carew-Belanger method involves only matrix multiplications and inversions.

Another alternative is offered in Ohap and Stubberud's work [95], where it is observed that the matrix $\tilde{S}$ should be zero for an optimal filter.

$$\tilde{S} = P_- H^T - KA(0) \qquad \text{(F-44)}$$

The method estimates $\tilde{S}$ on the basis of the innovations correlation. If improvements are needed, Ohap and Stubberud suggest using the steepest descent method for its solution.

In [96], Brewer proposes the use of Kalman filtering to estimate the unknown statistical properties represented by elements of R and Q_d. A state space model for the unknown means and covariances is proposed. The original filter's residuals provide the necessary measurements for the second K.F.

In [97] and [98], Martin and Stubberud show that the identification of Q and R can be decoupled from identifying H and Φ. A stochastic approximation algorithm is employed to carry out the identification task.

Ohnishi, [99] suggests that a process called the ξ-process is advantageous in identifying the noise statistics, since it does not involve estimating the state. The method involves estimating $Q_d Q_d^T$ and RR^T from the ξ-process, obtained by a series of transformations on the original system and measurement models. A unique solution for RR^T is given, whereas for QQ^T, a set of linear equations with respect to the elements of QQ^T, are obtained. If this set consists of $(n^2+n)/2$ linearly independent equations, QQ^T can be uniquely identified.

D. Covariance Matching Approaches

This approach is based on requiring that the covariances of the residuals be consistent with their theoretical values. In Mehra's review [76] one computes from the K.F. results the theoretical value of the residual covariance

$$E[r(k)\, r^T(k)] = HP_-(k)H^T + R \qquad \text{(F-45)}$$

The actual covariance matrix of the residuals is approximated by a sample covariance for an empirically chosen N, given by

$$E_a(r(k)\, r^T(k)) = \frac{1}{N}\sum_{i=1}^{N} r(i)\, r^T(i) \qquad \text{(F-46)}$$

If the actual covariance is larger than the theoretical value, the process noise Q_d is increased to increase $P_-(k)$ and to bring the actual covariance of the residual closer to the theoretical value.

To find Q_d, one has

$$H\, Q_d\, H^T = E_a(r(k)\, r^T(k)) - H\, \Phi\, P_+(k-1)\, \Phi^T\, H^T - R \qquad \text{(F-47)}$$

If the rank of H is less than n, there is no unique solution for Q_d. We may restrict the number of unknowns in Q_d to obtain a unique solution. The convergence of the process is not guaranteed.

If Q_d is known, but R is unknown, the estimate of R given by Equation (F-5) is reported to yield a good approximation [78].

Isaksson [79], summarizes Shellenbarger's approach [90] as one of using the residual covariance matrix to obtain a linear regression in the elements of Q_d and R. The result is a recursive scheme employing the actual covariances of the optimal estimates.

The original concept of covariance matching has been attributed to Jazwinski [100]. Here to estimate Q with $\hat{X}_+(k)$ and $P_+(k)$ available we pick a time lag N and determine the optimum $Q(k, N)$ by requiring that

$$r^2(k+\ell) = E(r^2(k+\ell)) \quad \ell = 1, \ldots, N \qquad \text{(F-48)}$$

This produces consistency between the residuals and their statistics. The predicted residuals for $\ell > 0$ are defined by

$$r(k+\ell) \triangleq z(k+\ell) - \hat{z}(k+\ell|k)$$

One can show that

$$\begin{aligned} E(r(k+\ell|k)\, r(k+m|k)) = {} & H(k+\ell)\, \Phi(k+\ell, k)\, P_+(k)\, \Phi^T(k+m, k)\, H^T(k+m) \\ & + H(k+\ell) \sum_{i=1}^{\ell} [\Phi(k+\ell, k+i)\, G(k+i-1) \\ & \times Q\, G^T(k+i-1)\, \Phi^T(k+m, k+i)\, H^T(k+m)] \\ & + R(k+\ell)\, \delta(\ell, m) \qquad m \geq \ell \end{aligned} \qquad \text{(F-49)}$$

The basic idea can be seen from the simple case of an input noise that is *uncorrelated and identically* distributed $Q=qI$, for $N=1$; the optimal estimate $\hat{q}(k,1)$ is given by

$$\hat{q}(k,1) = \frac{r^2(k+1) - E\{r^2(k+1)|q=0\}}{H(k+1)\,G(k)\,G^T(k)\,H^T(k+1)} \tag{F-50}$$

This is taken only if its value is positive. In the case where the numerator is negative we set $\hat{q}(k,1)=0$. Jazwinski's paper treats cases where a sample mean of N predicted residuals is used, as well as the case of independent noise input.

APPENDIX G
THE INSTRUMENTAL VARIABLE METHOD

A. The Concept

Consider the measurement model given by

$$z(k) = H(k)\,X + v(k) \tag{G-1}$$

The least squares estimate of the vector X, given N measurements is given by

$$\hat{X}_{LS} = \left[\sum_{k=1}^{N} H^T(k)\,H(k)\right]^{-1}\left[\sum_{k=1}^{N} H^T(k)\,z(k)\right] \tag{G-2}$$

Let us assume that the data originated from a true model given by

$$z(k) = H(k)\,X_0 + v_0(k) \tag{G-3}$$

By direct substitution one can show that

$$\hat{X}_{LS} = X_0 + \left[\sum_{k=1}^{N} H^T(k)\,H(k)\right]^{-1}\left[\sum_{k=1}^{N} H^T(k)\,v_0(k)\right] \tag{G-4}$$

It is clear that a desired property of the estimate $\hat{X}_{LS}$ is that it is close to X_0 and that it converges to X_0 as N tends to infinity. This does not take place in practice since $H(k)$ and $v(k)$ are correlated.

Let us now assume that the vector $H(k)$ is observed with some error and that its true value is $H_0(k)$. If $e(k)$ denotes the vector of measurement noise associated with the observation of $H(k)$, then we write

$$H(k) = H_0(k) + e(k) \tag{G-5}$$

The least squares estimate of equation (G-2) is modified to the form

$$\hat{X}_{IV} = \left[\sum_{k=1}^{N} \hat{H}_0^T(k)\,H(k)\right]^{-1}\left[\sum_{k=1}^{N} \hat{H}_0^T(k)\,z(k)\right] \tag{G-6}$$

where $\hat{H}_0$ is a best estimate of $H_0(k)$. This is the basis for the instrumental variable (I.V) method.

The method tries a general correlation vector $Y(k)$ to replace $\hat{H}_0^T(k)$ so that the I.V. estimate is obtained using

$$\hat{X}_{IV} = \left[\sum_{k=1}^{N} Y(k)\, H(k) \right]^{-1} \left[\sum_{k=1}^{N} Y(k)\, z(k) \right] \tag{G-7}$$

provided that the indicated inverse exists.

One can show by direct substitution that

$$\hat{X}_{IV} = X_0 + \left[\sum_{k=1}^{N} Y(k)\, H(k) \right]^{-1} \left[\sum_{k=1}^{N} Y(k)\, v_0(k) \right] \tag{G-8}$$

Therefore, the choice of the elements of the $Y(k)$ vector should be such that $Y(k)$ and $v_0(k)$ be uncorrelated as N tends to infinity.

B. Application to Time Series

Assume that we are given the following time series:

$$z(k) + a_1 z(k-1) + \dots + a_n z(k-n) = b_0 \mu(k) + \dots + b_n \mu(k-1) + v(k) \tag{G-9}$$

The series can be expressed in the form of equation (G-1), where

$$H(k) = [\ -z(k-1),\ -z(k-2),\ \dots,\ -z(k-n),\ \mu(k),\ \dots,\ \mu(k-n)] \tag{G-10}$$

$$X^T = [a_1,\ a_2,\ \dots,\ a_n,\ b_0,\ b_1,\ \dots,\ b_n] \tag{G-11}$$

The series can also be represented in transfer function form as

$$z(k) = y(k) + v'(k) \tag{G-12}$$

where

$$y(k) = \frac{N(B)}{D(B)}\, \mu(k) \tag{G-13}$$

$$v'(k) = \frac{v(k)}{D(B)} \tag{G-14}$$

with

$$D(B) = 1 + a_1 B + a_2 B^2 + \dots + a_n B^n \tag{G-15}$$

$$N(B) = b_0 + b_1 B + \dots + b_n B^n \tag{G-16}$$

We note that the input sequence $\mu(k)$ is available directly for measurements and is uncorrelated with the noise. Therefore $\mu(k)$ satisfies the requirements of an instrumental variable.

Since $y(k)$ is obtained from the input $\mu(k)$ using the transfer function $N(B)/D(B)$, the two variables $y(k)$ and $\mu(k)$ are correlated. The degree of correlation depends on the form of the model and the values of its parameters. The transfer function causes a dynamic lag between $\mu(k)$ and $y(k)$, and it is expected that $y(k)$ is more highly correlated with $\mu(k-\tau)$ than with $\mu(k)$. Here τ is a pure time delay. We thus consider the instrumental variable vector to be chosen as

$$\gamma^T(k) = [\ -\mu(k-\tau-1),\ \ldots,\ -\mu(k-\tau-n),\ \mu(k),\ \ldots,\ \mu(k-n)] \qquad \text{(G-17)}$$

The ideal value of τ is chosen such that the covariance matrix $R_{Y,Y}$ is maximized.

One may obtain instrumental variables that are even more correlated with $y(k)$ by letting $\mu(k)$ drive an "Auxiliary model" of the process to obtain an output $\hat{y}(i)$ for $i=k-1, \ldots, k-n$, that is used to define an I.V. vector of the form

$$\gamma^T(k) = [\ -\hat{y}(k-1),\ \ldots,\ -\hat{y}(k-n),\ \mu(k),\ \ldots,\ \mu(k-n)] \qquad \text{(G-18)}$$

The auxiliary model is chosen to be of the same form as the transfer function

$$D(B)\,\hat{y}(k) = N(B)\,\mu(k) \qquad \text{(G-19)}$$

Here, we have

$$D(B) = 1 + \alpha_1 B + \ldots + \alpha_n B^n \qquad \text{(G-20)}$$

$$N(B) = \beta_0 + \beta_1 B + \ldots + \beta_m B^m \qquad \text{(G-21)}$$

The auxiliary model parameters α_i, β_i are chosen in some sensible manner. Figure (G-1) shows in block diagram form the IV estimation process using the Auxiliary model concept.

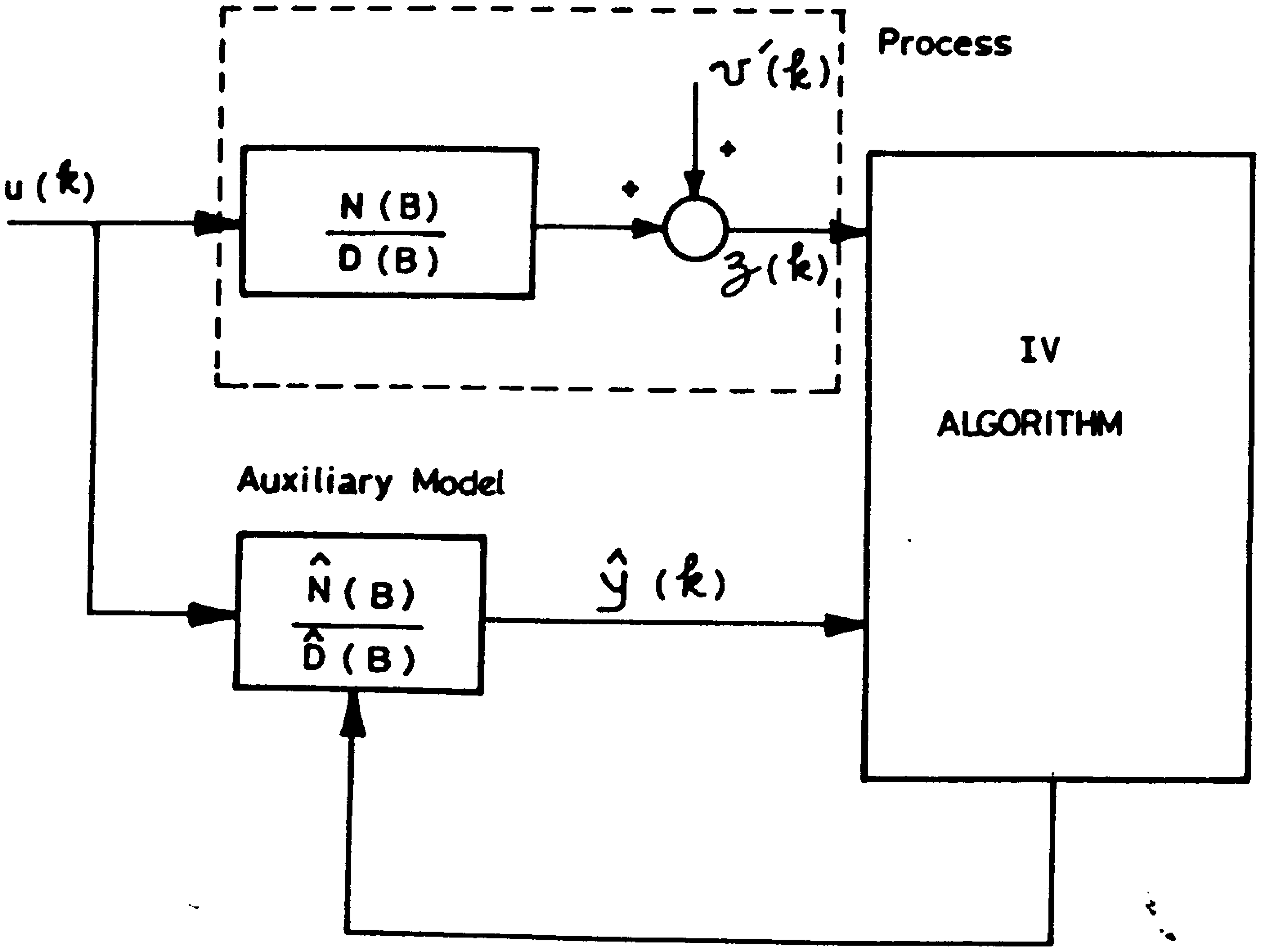


Figure (G-1) IV Estimation Via the Auxiliary Model Concept

C. Implementations Using the Auxiliary Model Concept

The instrumental variable method can be implemented in either an off-line or on-line fashion. Two possible off-line implementations discussed here differ only in the choice of the initialization estimates.

Off-line Method 1

Here one assumes that the instrumental variable vector is defined by equation (G-17), with an appropriate choice of ζ. An initial instrumental variable estimation is carried out using equation (G-7). The elements of $\hat{X}_{IV}$ obtained in this manner, define the parameters of the auxiliary model, given by (G-20) and (G-21) are

$$\hat{X}_{IV}^{T} = [\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_0, \beta_1, \ldots, \beta_n] \tag{G-22}$$

As a result the series $\hat{y}(k)$ can be computed using (G-19). A second pass can then be performed using the new instrumental variable vector defined by (G-18).

Off-line Method 2

In this method, the initial estimate of the auxiliary model parameters is obtained using a least squares estimator of the form (G-2). Thus we have

$$\hat{X}_{LS}^{T} = [\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_0, \beta_1, \ldots, \beta_n] \tag{G-23}$$

The series $\hat{y}(k)$ is now computed using (G-19). A second pass is performed using the new instrumental variable vector defined by (G-18).

The iterative procedure yields asymptotically unbiased estimates at each iteration if the IV's generated have the desired properties of high correlation with $y(k)$ and statistical independence from the noise. The process is continued until there is no significant change in the resulting estimates.

The iterative procedure can use either the en-bloc solution of the IV equations (G-7), or a recursive solution given by the Kalman filtering form:

$$\hat{X}(k) = \hat{X}(k-1) - \hat{K}(k)\,[H(k)\,\hat{X}(k-1) - z(k)] \tag{G-24}$$

$$\hat{K}(k) = \hat{P}(k-1)\ \hat{Y}(k)[\ 1 + H(k)\ \hat{P}(k-1)\ \hat{Y}(k)]^{-1} \qquad \text{(G-25)}$$

$$\hat{P}(k) = \hat{P}(k-1) - \hat{P}(k-1)\hat{Y}(k)\ [\ 1+H(k)\ \hat{P}(k-1)\ \hat{Y}(k)]^{-1}\ H(k)\ \hat{P}(k-1) \qquad \text{(G-26)}$$

On-line Method

An on-line method is obtained by using the recursive solution (G-24) to (G-26) and updating the auxiliary model continuously on the basis of the recursive estimates.

The parameters of the auxiliary model are updated in such a manner as to ensure that the instrumental variables satisfy the validity requirements. Since the estimate $\hat{X}(k)$ is correlated with $\upsilon'(k)$ at the same time instant, it is passed through a time delay filter and a discrete low pass filter as described in Young [103].

4. Refined Instrumental Variable Method

Based on maximum likelihood estimation arguments, Young [103] concludes that a good theoretical foundation of the I.V. method is obtained if data pre-filtering is carried out. The idea is to use the I.V. normal equations with pre-filtering variables replacing the raw variables. Here we obtain

$$z_f(k) = \frac{1}{D(B)}\ z(k)$$

$$\mu_f(k) = \frac{1}{D(B)}\ \mu(k)$$

$$\hat{y}_f(k) = \frac{N(B)}{D(B)}\ \mu_f(k)$$

The subscript f denotes pre-filtering.

Noting that the additional pre-filtering operation requires knowledge of the process N(B) and D(B), a common approach is to pre-select a filter L(B) to conduct the filtering operation

$$z_f(k) = L(B)\ z(k) \qquad \text{(G-27)}$$

$$\mu_f(k) = L(B)\ \mu(k) \qquad \text{(G-28)}$$

$$\hat{y}_f(k) = \frac{\hat{N}(B)}{\hat{D}(B)}\ \mu_f(k) \qquad \text{(G-29)}$$

The procedure of selecting the filter $L(B)$ is embedded in the refined I.V. algorithm discussed now.

Step 1: Select the order of $N(B)$ and $D(B)$.

Step 2: Compute the estimate of the parameter vector a $\hat{X}$ using least squares minimization process.

Step 3: Form the polynomials $\hat{D}_{LS}(B)$ and $\hat{N}_{LS}(B)$ from parameters of Step 2.

Step 4: Generate IV, variable $\hat{y}(k)$ using (G-19)

$$\hat{y}(k) = \frac{\hat{N}(B)}{\hat{D}(B)}$$

Step 5: Compute the estimate $\hat{X}_{IV}$ of the vector X.

Step 6: Form $\hat{N}_{IV}(B)$ and $\hat{D}_{IV}(B)$ from $\hat{X}_{IV}$.

Step 7: Generate the estimate of noise sequence

$$V(k) = \hat{D}_{IV}(B)\, z(k) - \hat{N}_{IV}(B)\, \mu(k).$$

Stpe 8: Search for an n^{th} order AR model that transforms the sequence $\upsilon(k)$ into a white noise sequence. This defines the pre-filter $L(B)$.

Step 9: Form the pre-filter $L(B)$ from the parameters of the n^{th} order AR model of Step 8.

Step 10: Pre-filter input data to generate filtered input and output sequences $\mu_f(k)$ and $z_f(k)$.

Step 11: Generate a new set of instrumental variables using the sequence $\mu_f(k)$ and $z_f(k)$.

Step 12: Compute the refined IV estimates $\hat{X}_{RIV}$ of the parameter vector X producing the optimum estimate of the system dynamics.

Step 13: Generate the noise sequence $\upsilon(k)$ corresponding to the estimates of Step 12.

Step 14: Use IV technique to model $\upsilon(k)$ to determine the optimum noise model.

APPENDIX H
CONSTRAINED WEIGHTED LEAST SQUARES DERIVATION

The objective function of the weighted least squares given by eq. (4-28) is written as

$$J = [Z - H\hat{X}]^T W [Z - H\hat{X}]$$

This can be expanded and rearranged to yield

$$J = Z^T WZ - 2\hat{X}^T H^T WZ + \hat{X}^T H^T WH\hat{X}$$

The first term in J does not depend on $\hat{X}$, and thus one needs to consider minimizing

$$\tilde{J} = \hat{X}^T H^T WH\hat{X} - 2\hat{X}^T H^T WZ \qquad \text{(H-1)}$$

We now define

$$A = H^T WH \qquad \text{(H-2)}$$

$$b = H^T WZ \qquad \text{(H-3)}$$

As a result $\tilde{J}$ of eq. (H-1) is written as

$$\tilde{J} = \hat{X}^T A\hat{X} - 2\hat{X}^T b \qquad \text{(H-4)}$$

In the present application, we have

$$\hat{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \qquad \text{(H-5)} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \qquad \text{(H-6)}$$

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \tag{H-7}$$

Expanding eq. (H-4), one obtains

$$\tilde{J} = a_{11}X_1^2 + a_{22}X_2^2 + a_{33}X_3^2 + 2a_{12}X_1X_2 + 2a_{13}X_1X_3 + 2a_{23}X_2X_3 - 2b_1X_1 - 2b_2X_2 - 2b_3X_3 \tag{H-8}$$

This is eq. (4-30).

To obtain the elements of the matrix A, we recall that

$$H = \begin{bmatrix} 1 & P(1) & P^2(1) \\ 1 & P(2) & P^2(2) \\ \vdots & & \\ 1 & P(N) & P^2(N) \end{bmatrix}$$

$$W = \text{diag}\ [W(1), W(2), \ldots, W(N)]$$

As a result, using eq. (H-2) and equating the result to the elements of eq. (H-7), we obtain the six expressions of eq. (4-31). In a similar fashion, the elements of the vector b of eq. (H-6) are evaluated using eq. (H-3) to obtain the three expressions of eq. (4-32).

APPENDIX I
THE GRG METHOD

The GRG algorithm is intended to solve the problem stated in eq. (5-12) iteratively. By introducing appropriate slack variables, one can convert the inequality constraints to equalities, and therefore GRG algorithms can be described in terms of the following nonlinear programming problem:

$$\text{Minimize } f(x) \tag{I-1}$$

subject to

$$g(x) = 0 \tag{I-2}$$

$$L \leq x \leq U \tag{I-3}$$

In a manner similar to that employed in Linear Programming, GRG algorithms use the m equality constraints (I-2) to solve for m basic variables in terms of the remaining (n-m) non-basic variables. One thus partitions the nx1 column vector x into the two vectors x_b (basic) and x_{nb} (non-basic). As a result the constraint equation (I-2) is written as:

$$g(x_b, x_{nb}) = 0 \tag{I-4}$$

Let J denote the Jacobian matrix associated with (5-16), defined by:

$$J = \frac{\partial g}{\partial x}$$

This Jacobian matrix is partitioned as:

$$\frac{\partial g}{\partial x} = \left[\frac{\partial g}{\partial x_b} \; \frac{\partial g}{\partial x_{nb}} \right]$$

The Jacobian submatrices are renamed as follows:

$$\frac{\partial g}{\partial x_b} = J_b$$

$$\frac{\partial g}{\partial x_{nb}} = J_{nb}$$

As a result, one writes:

$$\frac{\partial g}{\partial x} = [J_b, J_{nb}]$$

Alternatively, the Jacobian is written as:

$$J = [J_b, J_{nb}] \tag{I-5}$$

Note that one assumes that the basic variables have been renumbered as the first m components of x. One also assumes that the selection of basic variables is such that the Jacobian submatrix J_b evaluated at the feasible point $\bar{x}$ is non-singular. This allows the solution of (I-2) for the basic variables x_b in terms of non-basic variables x_{nb}. This is valid for all x_{nb} sufficiently near $\bar{x}_{nb}$.

To solve (I-2) one uses a Taylor expansion of vector function g as follows:

$$g(\bar{x}+\Delta) = g(\bar{x})+J\Delta$$

Assuming that $\bar{x}$ is chosen such that $g(\bar{x})=0$ and that $g(\bar{x}+\Delta)=0$, then it follows that:

$$J\Delta = 0$$

In terms of the partitions one thus requires:

$$J_b\Delta_b + J_{nb}\Delta_{nb} = 0$$

As a result:

$$\Delta_b = -J_b^{-1} J_{nb}\Delta_{nb}$$

Note that:

$$\Delta = x - \bar{x}$$

$$\Delta_b = x_b - \bar{x}_b$$

$$\Delta_{nb} = x_{nb} - \bar{x}_{nb}$$

One can write:

$$x_b = \bar{x}_b - J_b^{-1} J_{nb}[x_{nb} - \bar{x}_{nb}] \qquad \text{(I-6)}$$

The objective function is thus transformed into a function of the non-basic variables x_{nb} only as expressed by:

$$f(x_b(x_{nb}), x_{nb}) = F(x_{nb}) \qquad \text{(I-7)}$$

One can thus conclude that the original problem (I-1),(I-2) & (I-3), in the neighbourhood of $\bar{x}$ has been transformed to a simpler <u>reduced</u> problem:

$$\text{Minimize} \quad F(x_{nb}) \qquad \text{(I-8)}$$

subject to the bounds on x_{nb}. The function F is called the reduced objective, and its gradient ∇F is the reduced gradient.

Using the chain rule, one writes:

$$\frac{\partial F}{\partial x_{nb}} = \frac{\partial f}{\partial x_{nb}} + \frac{\partial f}{\partial x_b} \cdot \frac{\partial x_b}{\partial x_{nb}}$$

Recall from (5-18) that:

$$\frac{\partial x_b}{\partial x_{nb}} = -J_b^{-1} J_{nb}$$

Therefore, one writes the reduced gradient as:

$$\frac{\partial F}{\partial x_{nb}} = \frac{\partial f}{\partial x_{nb}} - \frac{\partial f}{\partial x_b} \cdot J_b^{-1} J_{nb}$$

One now defines:

$$\Pi^{T} = \frac{\partial f}{\partial X_b} \cdot J_b^{-1}$$

It is clear that Π^{T} is obtained as the solution to the set of linear equations:

$$J_b \Pi^{T} = \frac{\partial f}{\partial X_b}$$

As a result, the reduced gradient is written as:

$$\frac{\partial F}{\partial X_{nb}} = \frac{\partial f}{\partial X_{nb}} - \Pi^{T} \cdot J_{nb}$$

All partial derivatives are evaluated at $\bar{X}$. The first order necessary condition for optimality is that the reduced gradient is zero.

The (n-m) non-basic variables vector x_{nb} is further partitioned into a vector of s super-basic variables x_s and a vector of the (n-m-s) remaining non-basic variables X_r. The super-basic variables are strictly between their bounds and the variables of X_r are those which are at one of their bounds. The reduced gradient with respect to the remaining non-basic variables $\frac{\partial F}{\partial X_r}$ is used only to determine if one of these variables should be released from a bound to join the super-basic set. This decision is made at each iteration in the GRG2 implementation. On the other hand in the MINOS implementation, this decision is made after an optimization over the current-basics is completed. In either case, the reduced gradient with respect to the super-basic variables $\frac{\partial F}{\partial X_s}$ is used to form a search direction d. Both conjugate and variable metric methods have been used to determine $\bar{d}$. Then a one dimensional search is conducted to solve the problem of minimizing $F(\bar{X}_{nb}+\alpha d)$ over values of $\alpha \geq 0$. The vector $\bar{d}$ is extended to include zero components for the non-basic variables at bounds.

The minimization is carried out by choosing a sequence of positive values of α_i. For each value of α, the reduced objective $F(\bar{X}_{nb} + \alpha_i \bar{d})$ must be evaluated. The basic variables $(\bar{X}_{ab}+\alpha_i \bar{d})$ have to be determined, since

$$F(\bar{X}_{nb}+\alpha_i \bar{d}) = f(x_b(\bar{X}_{nb}+\alpha_i \bar{d}), (\bar{X}_{nb}+\alpha_i \bar{d}))$$

The basic variables satisfy the system of equations

$$g(x_b, \bar{X}_{nb}+\alpha_i \bar{d}) = 0$$

Here $\bar{x}_{nb}$, α_i, $\bar{d}$ are known and x_b is to be determined. If x_b appears nonlinearly in any constraint, then the system is solved iteratively using the modified Newton's method obtained by keeping the Jacobian fixed throughout the iterations.

In the case of nonlinear constraints, the one-dimensional search can terminate in three different ways:

1- Newton's method may not converge. If this takes place on the first step, α_i is reduced and the process is repeated. Otherwise, the search is terminated.

2- If Newton's method converges, some basic variables may be outside their bounds. In this case, a new value of α is determined such that at least one such variable is at its bound and all others are within their bounds. If at this new point, the objective is less than at all previous points, the one-dimensional search is terminated. A new set of basic variables is determined and solution of a new reduced problem is initiated.

3- The search may continue until an objective value is found which is larger than the previous value. Then a quadratic fit to the three α_i values close to the minimum is determined. F is evaluated at the minimum of this quadratic, and the search terminates with the lowest F value obtained. The reduced problem remains the same.

It is important to realize that the algorithm attempts to return to the constraint surface at each step in the one-dimensional search.