AN APPLICATION OF SYSTEM IDENTIFICATION AND STATE PREDICTION TO ELECTRIC LOAD MODELLING AND FORECASTING

by

Francisco Daniel Galiana

B. Eng., McGill University
(1966)

S. M., Mass. Inst. Tech.
(1968)

Eng., Mass. Inst. Tech.
(1968)

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

March, 1971

Signature of Author ______________________________
Department of Electrical Engineering, March, 1971

Certified by ______________________________
Thesis Supervisor

Accepted by ______________________________
Chairman, Departmental Committee on Graduate Students

AN APPLICATION OF SYSTEM IDENTIFICATION AND STATE PREDICTION TO ELECTRIC LOAD MODELLING AND FORECASTING

by

Francisco Daniel Galiana

Submitted to the Department of Electrical Engineering on March 15th, 1971, in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

ABSTRACT

System modelling and identification techniques are applied in developing a probabilistic mathematical model for the load of an electric power system, for the purpose of short term load forecasting.

The model assumes the load is given by the sum of a periodic discrete time series with a period of 24 hours and a residual term. The latter is characterized by the output of a discrete time dynamical linear system driven by a white random process and a deterministic input, u, which is determined by a non-linear static function of the actual and normal temperatures.

System identification techniques are used to determine the model order and parameters which best fit the observed load behaviour for a wide range of conditions. These techniques are also utilized in adapting the model to seasonal variations.

Linear estimation and prediction is now used to determine load forecast curves for variable periods up to at least one week. Each load forecast is accompanied by a measure of its uncertainty in terms of its standard deviation. This allows us to devise a simple quantitative test to detect otherwise not-so-obvious-to-the-naked-eye abnormal behaviour.

The load forecast curve is updated once an hour as new load and temperature data is read through an optimum linear filter.

Tests are carried out with real load and temperature data to validate the proposed model's capability to forecast as suggested. The results are most satisfactory.

THESIS SUPERVISOR: Fred C. Schweppe
TITLE: Associate Professor of Electrical Engineering

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis supervisor Professor Fred C. Schweppe. Working with him has been a great privilege and experience. Below is a humble attempt to acknowledge his invaluable help throughout the thesis work:

" Thank you, illustrious Schweppe.
With your patience and great wisdom,
The awaited end at last has come,
And happy Galianas cry out: Whoopeee!"

Sincere thanks are extended to Prof. Sanjoy Mitter and Prof. Gerry Wilson for their helpful comments and advice.

Valuable discussions were also carried out with Dr. Duncan Glover and Martin Baughman whom I thank.

Special thanks are due to Mr. Thomas J. Kraynak and David L. Rosa of the Cleveland Electric Illuminating Company who very kindly supplied the load and temperature data used in this study.

A mis padres,Fernando y Maria, que nunca dudaron el resultado final y que me dieron coraje en momentos dificiles, mil gracias.

Last but not least my deepest love and gratitude go to my wife, Mimi, whose help was beyond words. Without her typing, drawing, proofreading, advice, and constant encouragement and trust, this would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

FIGURES: page

page

TABLES:

I.0 INTRODUCTION & BACKGROUND

I.1 Introduction:

We have two principal objectives in this study. The primary aim is to study the problem of short term electric load forecasting via the concepts of system modelling and identification, and state estimation and prediction. The second is to verify the extent of the validity of the available theory of system identification and modelling as applied to a real problem such as load forecasting, and if necessary and possible extend its range.

We first hypothesize a general mathematical model for the load of a power system based on physical reasoning and observed load behaviour under varying conditions. The main points of this model are the separation of load into two components, periodic and residual. The first depends on the time of the day and day of the week, while the residual term is a random process defined as the output of a discrete linear system driven by white noise and by a temperature deviation variable. The latter is in turn defined by a non-linear memoryless function of both actual and normal temperatures. This is described in sections II.1 through II.6.

The remaining sections of chapter II describe the main forecasting algorithms, which are essentially the Kalman filter-predictor equations (Ref. 11). The forecasting technique is computationally simple, and provides the operator with a number of possible interactions which can be used to complement each other and improve the reliability and efficiency of the operation of a power system. For example, the operator may be given a load forecast for any future time he requests, together with the expected error standard deviation. Alternatively, he may wish to see a display of predicted load values at discrete intervals with the corresponding statistical confidence level. This display may prove very valuable in scheduling generation and power exchanges.

Another main asset of this approach is the capability of detecting abnormal load behaviour otherwise not so easily noticeable by the naked eye. Under these conditions the operator is warned of the abnormality, and may introduce personal corrections to the forecast if necessary.

Chapter III describes a number of approaches for the identification of the number and set of unknown model parameters from a past observed data record. A least squares or maximum likelihood approach is used and a number of techniques are described and analyzed. For reasons of convenience, we choose one relying on the Fletcher-Powell algorithms for minimizing functions (Ref. 22).

Chapter IV describes a number of tests carried out on both simulated and real data to evaluate the proposed model, identification and forecasting techniques. The load data was obtained from the Cleveland Electric Illuminating Company while the weather data came from the U.S. Weather Bureau.

The simulated data tests are described in sections IV.1 through IV.5.

Results with real data were very successful and are described in IV.6. Models are determined for the months of July, August and January, each of which yields predictions from one hour to one week compatible with the predicted error standard deviation of from 1% to 2.5% of peak load. The model's ability to predict is particularly emphasized during periods of large temperature deviations from normal, such as heat waves. It must also be emphasized that actual temperature data was used in these predictions. Predicted weather data together with its confidence level can be used, resulting in worse load forecasts. The detection of anomalous load behaviour is tested by artificially introducing small disturbances into the load data and proves very successful.

Chapter V proposes a number of guidelines and recommendations for the practical implementation of the proposed technique in a real system.

Chapter VI summarizes the work done and makes a number of conclusions and recommendations for future research.

I.2 Background:

I.2.1 General Description of an Electric Power System:

The purpose of an electric power system is to generate and distribute the necessary power demanded by its customers. Furthermore, this must be done reliably and efficiently, that is with a minimum number of power interruptions or disturbances and least cost to the company and the customer.

The satisfaction of these conditions is a formidable task which is receiving considerable attention. Load forecasting is but one aspect of this problem, but it is sufficiently challenging and important to merit separate attention.

The electric load is defined as the real electric power demanded by the customers of a power company. This demand varies considerably over a period of 24 hours, so that power generation must be adjusted over this period to follow this variation as closely as possible. Small load changes, those occuring in the order of a few minutes or fractions of, can be tracked by small changes of generation by the units already in operation. This is the so-called load-frequency control problem whose main objective is to maintain the frequency at 60 Hz, as well as the power flow to interconnected companies constant (Ref.1, 2).

Larger load variations which occur over the period of 24 hours cannot be tracked by the limited capacity generators already on-line. Instead, new blocks of generation must be brought into the system, and be ready to supply additional power when demanded. From an economic point of view it makes sense not to start up and maintain unnecessary spinning reserve. On the other hand, lack of spinning reserve would necessitate shedding of load or some such more drastic undesirable measures.

A compromise must thus be reached in scheduling generation to optimize operational costs as well as system reliability. The satisfaction of this compromise is complicated by a number of constraints. Some of these are, the delay of from three to six hours to bring a block of generation up to running speeds, that is 3600 rpm; relative costs of operating different forms of generation in the system, e.g. steam, hydro, nuclear; transmission failure contingency reserve, generation failure contingency reserve, maintenance shutdowns, power flow to interconnected companies and others.

It is clear that load forecasting plays an important part in this aspect of a power system's operation. That is, economical as well as reliable tracking of the load demanded cannot be accomplished, in view of the above described delays and constraints, unless ample warning time of future load behaviour is provided.

I.2.2 Observed Load Behaviour:

In this section we discuss load behaviour as it is known from years of observation, as well as the behaviour of the consumers under varying time and weather effects.

The load, z(t), is the power consumed at time t by all industrial, commercial, public, domestic and agricultural consumers supplied by the particular power system. Generally these customers are distributed over a large area, e.g. a large city, or a group of cities, towns and rural areas. The behaviour of z with time is therefore determined by the time variation of the myriad of power consuming devices in this area.

It would then be an impossible task to attempt to understand load behaviour from the point of view of its individual constituents. A better possibility may be to monitor chosen buses (network nodes) which serve mainly industrial, residential, commercial or agricultural loads and analyze the behaviour of each such load. This monitoring is however not available, and one could still have areas which do not fit any of the above categories. Nevertheless such studies, together with surveys in these areas as to the nature of the power consumption in them, might be a significant approach in more ambitious projects.

The load is fortunately a reasonably well behaved time series and tends to follow recognizable patterns even when the different types of loads discussed above are not separated. This regularity is due to the large quantities of power consuming devices and consumers which tend to smooth out or average the total load, in addition to the regular patterns of consumption by the customers and how they are affected by certain factors such as time and weather conditions. Nevertheless, the load is an uncertain process in the sense that its value at any time cannot normally be exactly determined, except after it is observed, that is one cannot normally exactly model or predict the load. The aim of load forecasting is thus to model and predict the load, z, as closely as possible in the presence of this inherent uncertainty.

Short time load behaviour (up to 5 min.)

If we observe the load over the period of a few seconds to minutes, under normal conditions, small random fluctuations are clearly evident. In addition, we may have larger longer lasting variations which are significant from the point of view of load frequency control, that is variations which can be followed by the on-line generators. Under normal conditions load changes in this time interval are not large enough to necessitate bringing additional generation into line.

The small fast fluctuations are generally ignored by the load frequency controller as they could not be followed rapidly enough by the generation equipment in addition to normally being sufficiently small to be negligible when compared with the total load.

Daily load behaviour (5 min. to 24 hours)

The load behaviour over a 24 hour period acquires considerable regularity. A typical sample over a 48 hour period is shown in Fig. 1.

The most obvious characteristic is that of near periodicity over a 24 hour period. It generally rises very rapidly in the early morning, breakfast time, stays approximately constant over the morning hours and decays slowly after supper time, approximately repeating this cycle every 24 hour period. The reason for this behaviour is quite evident, that is, consumption approximately follows

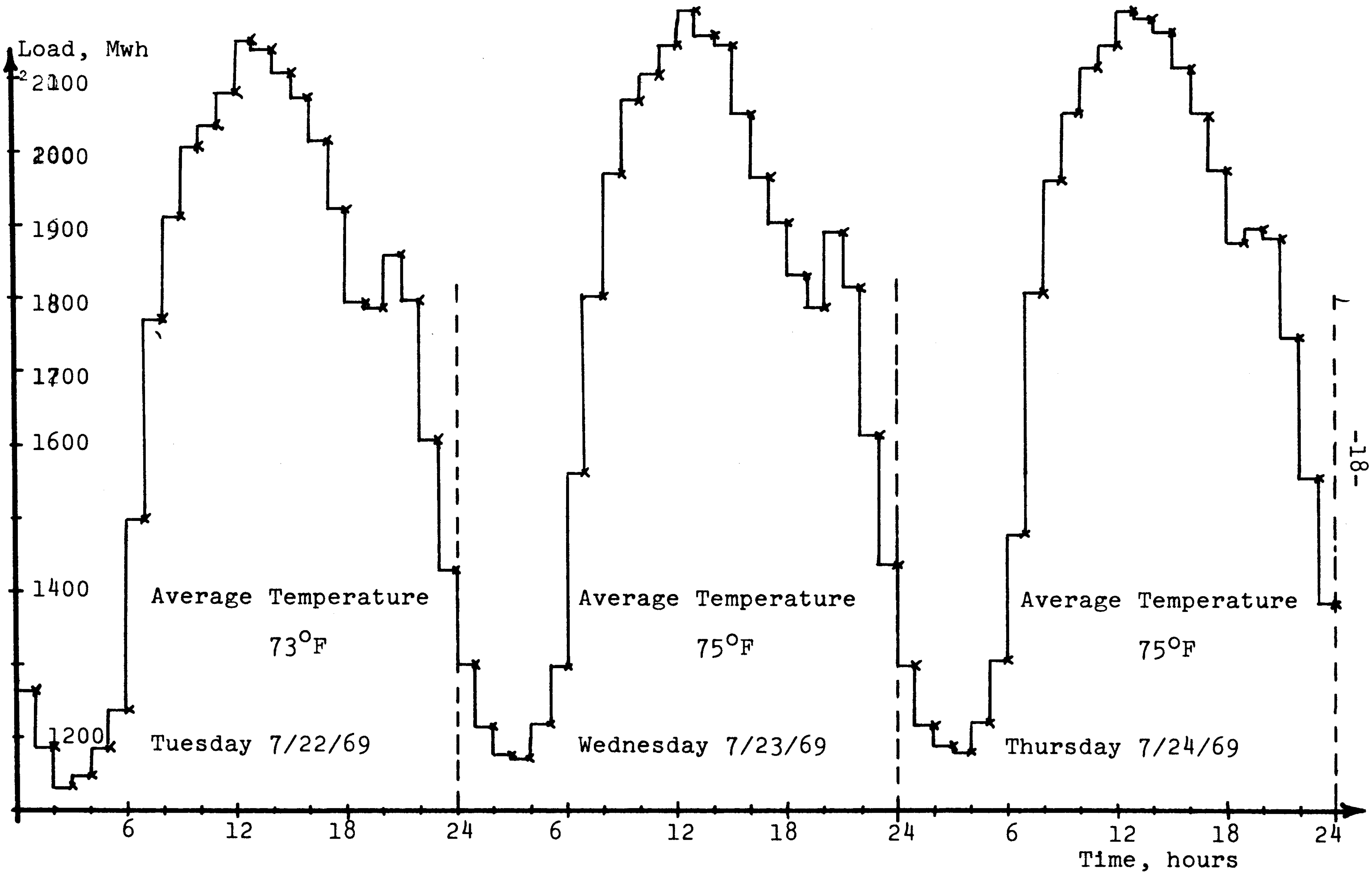


Fig. 1: Typical Weekday Load Behaviour (Cleveland Illuminating Co.)

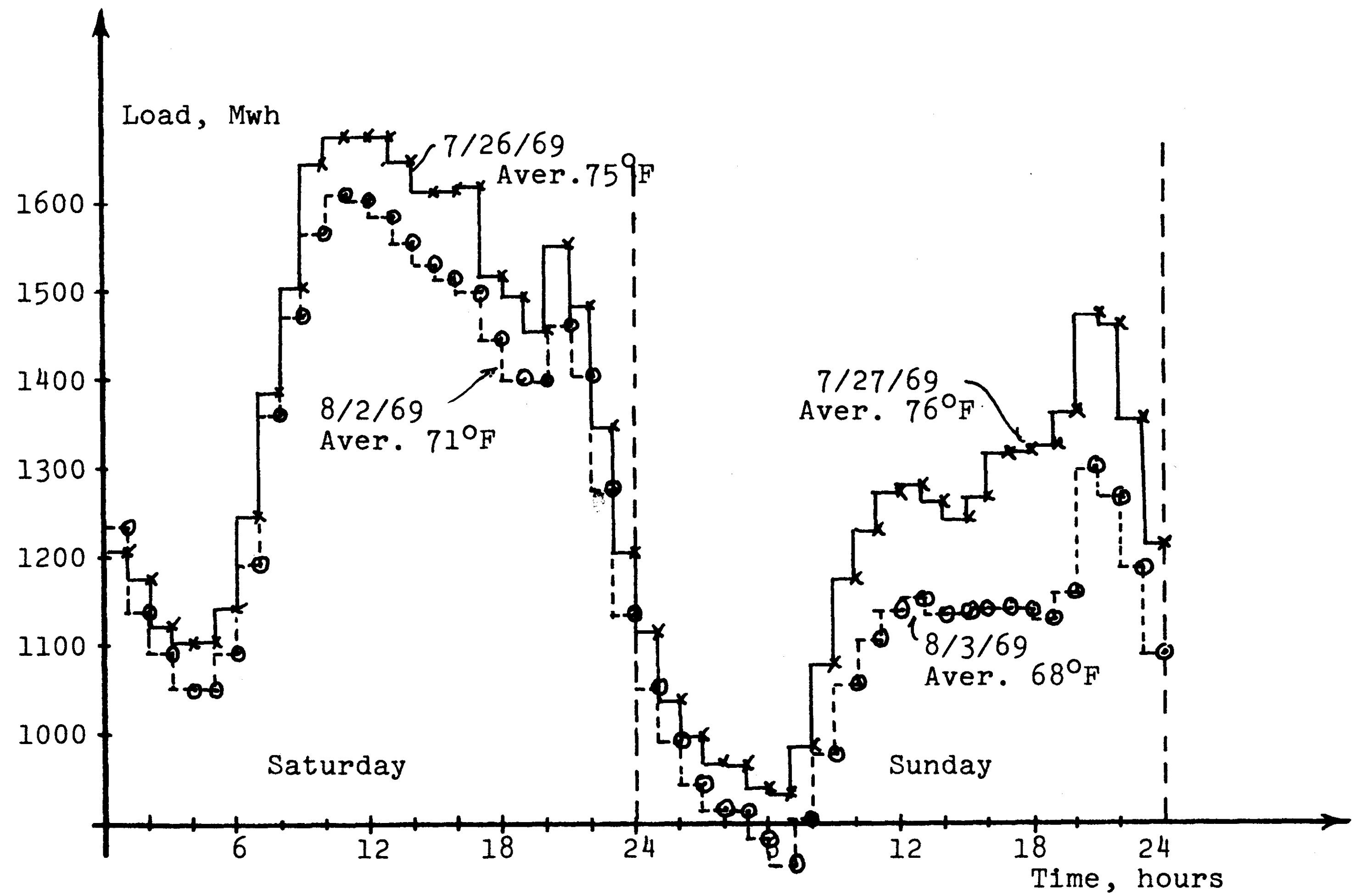


Fig. 2: Typical Weekend Load Behaviour (Cleveland Illuminating Co.)

the sleeping, working, resting cycles of its customers, which fortunately are fairly regular.

Effect of weekday on load behaviour

This basic daily load behaviour may vary from weekday to weekday to a small extent, given that all conditions are equal, in a manner not attributable to the random nature of the process. This is due to small changes in power consumption habits from day to day, e.g. from a Monday to a Friday. This difference is much more pronounced during the weekend days or a holiday period, for obvious reasons. See figure 2.

Given all other conditions equal, days separated by multiples of one week will show a behaviour approximately equal, with larger differences appearing as this interval increases.

Effect of seasons on load behaviour

The nature of the load changes considerably over the seasons. This is due to variations in weather, duration of day, and customers' consumption habits. Thus during the winter we may make use of more electric heating or cooking devices. Days are shorter hence lighting loads increase. During the summer, air conditioning and refrigeration loads become significant, but lighting loads may decrease. In addition consumers' habits change, thus in summer we go to bed later, watch less television, and use less lighting.

This seasonal changes in the behaviour of the daily load curve are fortunately slow and can be readily identified. However, although slow, they tend to change the daily load curve from week to week by a non-negligible amount.

Effect of years on load

Superimposed on the daily and seasonal patterns of load behaviour one has the inevitable growth factor due to the rapid growth of power consuming devices available.

Such dffects are however very slow and for the purpose of short term load forecasting not very difficult to take into account. For the purpose of long term planning of future power system needs this factor is significant.

Effect of weather on load

Analysis of daily load curves not too far apart in the calendar, e.g. one week apart, shows that, for the same weekday, weather plays a very significant role in the power consumption as shown in Fig. 3, 4. The major weather effects which influence the load behaviour are:

i) Temperature

ii) Light Intensity

iii) Humidity

iv) Wind Speed

v) Precipitation

The above factors are listed in approximately the order of

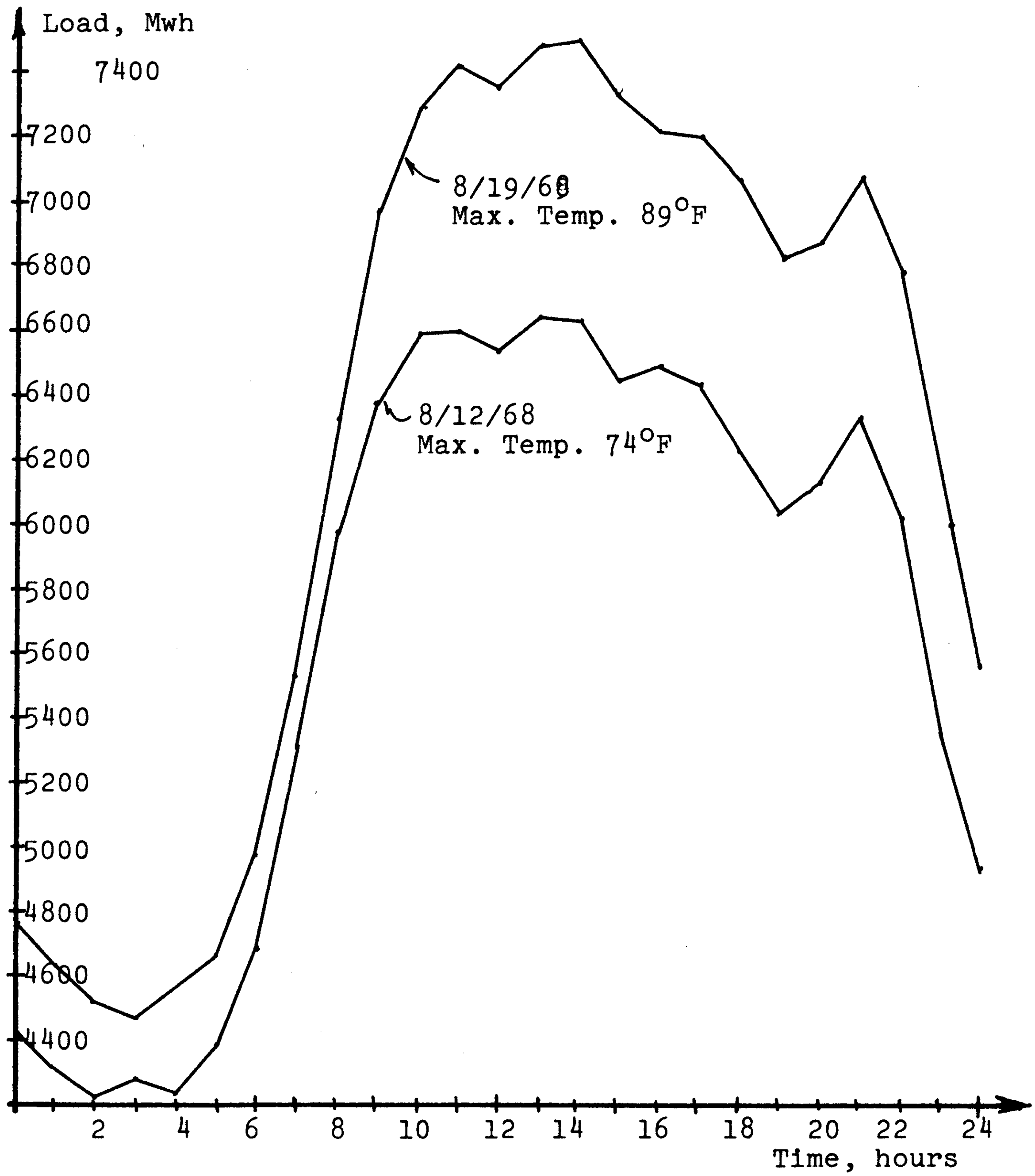


Fig. 3: Effect of Temperature on Load in Summer

(American Electric Power)

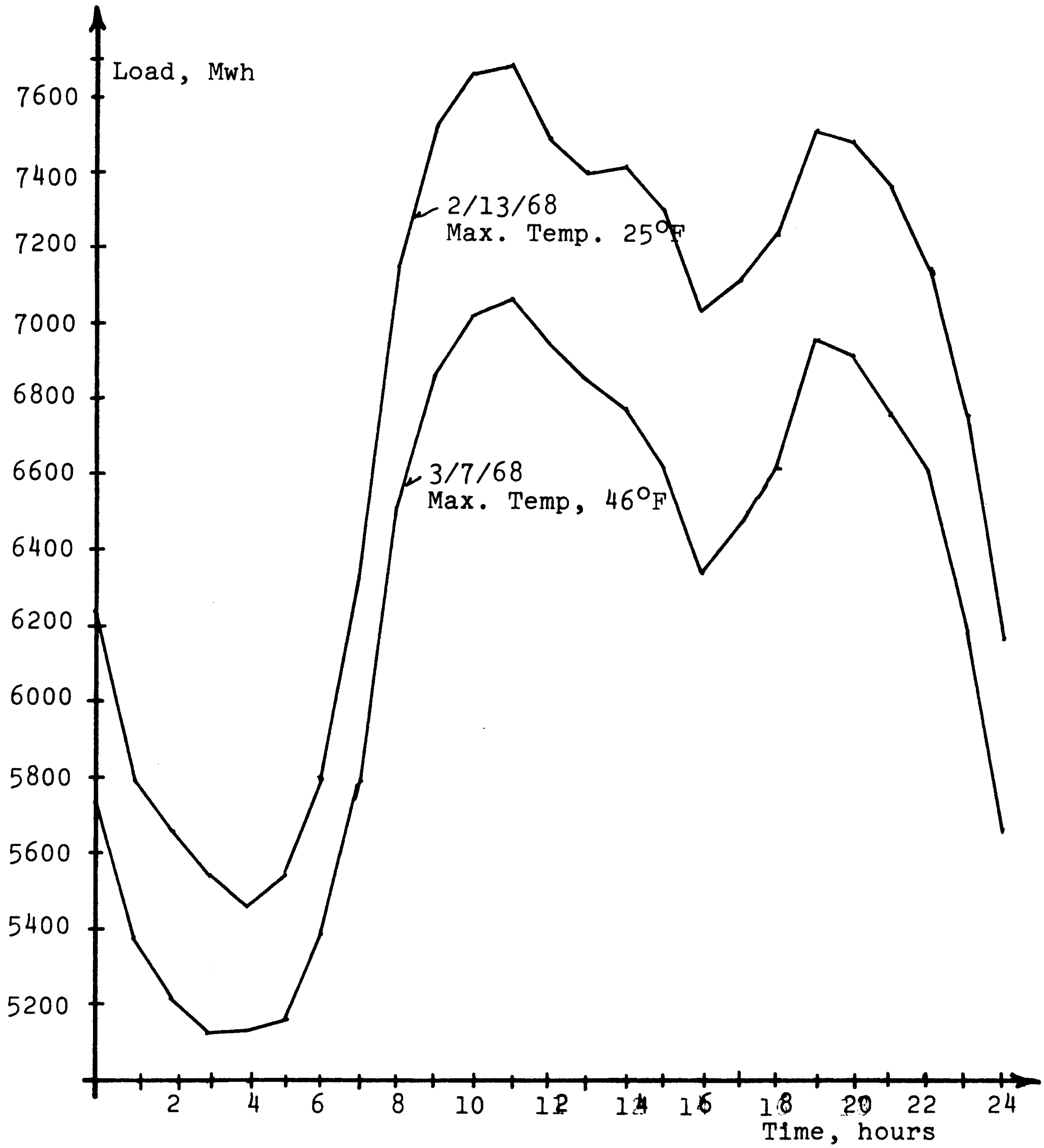


Fig. 4: Effect of Temperature on Load in Winter

(American Electric Power)

decreasing significance to the load behaviour.

The qualitative reason as to the influence of weather in load behaviour is quite clear. Our comfort, well-being and habits are closely related to the weather conditions. We surround ourselves with a number of electric devices which we use to maintain these living conditions at comfortable and desirable levels. Thus we have heating and cooling devices, refrigeration, cooking, cutting, washing, drying devices, lighting, radios, televisions and many others. The use and power consumption of all these is <u>to some degree</u> affected by the nature of the weather, through the consumer.

Clearly some loads are unaffected by weather. These are industrial and to some extent commercial loads. In these cases we have considerable power consumption which is independent of weather.

<u>Random load behaviour</u>

In spite of the load's regular behaviour with respect to time and weather, at any instant of time its value is uncertain to some degree. This behaviour is clear in view of the nature of the load. That is, the summation of the power consumed by literally thousands of devices, each of which is independently controlled by human beings, introduces inherent uncertainty in its value. However, because of the large number of devices involved, some regularity can be expected due to a smoothing or averaging effect, more or less like that of throwing a pair of dice a large number of times. In the case of the load the number of possible outcomes is infinite, time varying and depends on a number of factors both deterministic and uncertain.

This random behaviour may be of the type discussed earlier, that is very fast and small fluctuations superimposed on slower larger changes, or it may consist of a large discontinuity in the load as caused by the sudden gain or loss of a large load, such as factory or building.

A second form of randomness in the load is due to disturbances which although not well predictable in magnitude, can be predicted in time. Examples are effects like the World Series, Lunar landings, and election days. The worst cases of random disturbances unpredictable in both time and magnitude are caused by events such as national or local emergencies, assasinations of prominent figures, and natural disasters such as earthquakes, tornadoes and hurricanes.

I.2.3 <u>State of the Field</u>:

Load forecasting is still considered more of an art than a science. For this reason many power companies maintain a number of load forecasters who base their forecasts on experience and insight. Basically, they maintain records of past load behaviour, together with past weather conditions, industry strikes and any other factor which has been known to influence the load. Based on this information, forecast weather conditions for the area, and some other random effects, a load prediction is made by comparing these conditions with those of a day with similar ones, whose load behaviour is recorded in their files. Adding a fudge factor based on insight, a forecast is made which, more often than not, is sufficiently accurate.

The disadvantages here lie in the fact that large amounts of data must be stored, only one or two forecasts are made daily, and the method is highly subjective and therefore liable to human errors and biases. Alternatively this type of forecast may be advantageous, as the complexity of the problem is such that human insight or intuition may be the best way of arriving at a forecast when conditions deviate from the normal.

Over the past years a serious attempt has been made to develop mathematical models which could be implemented in a computer for automatic load forecasting. Good surveys on the subject have been written by Mattheuman and Nicholson, and Gupta (Ref. 3, 4).

There are two basic approaches. One is to establish a mathematical model based on the correlation between load and weather, and the second is to base this model on the relatioship between the present and past values of the load.

Arguments against weather-correlation models have been raised since these require weather data monitoring, as well as accurate weather prediction, leading to the complaint that companies do not have such monitoring facilities easily available, and that erroneous weather forecasts would lead to doubling the load forecast error. Techniques that make use of past load data only, have the advantage of being simpler to implement in the sense that this data is more readily available. Other arguments state that the response of the load to weather effects tends to be relatively slow (of the order of 24 hours) and therefore easily identifiable and predictable in terms of the most recent past load data (Ref. 5). Weather effects with faster responses such as cloud cover are difficult to predict and therefore are not included.

On the other hand ignoring this evident correlation would seem to be a rejection of very significant information abou the load behaviour. The use of past and present weather would seem to be a very significant in arriving at better load forecasts for lead times of 24 hours or more, than those based only on past load data. The argument being that past load data does not contain information about the present and most recent weather effects. In addition sudden changes in weather conditions could be incorporated into the load forecast. Also some correlation exists between the most recent weather and the present load, so that weather forecasts even if inexact should yield better forecasts than if altogether ignored. Finally, a weather dependent model can serve to carry out contingency predictions for different weather forecasts.

Weather correlation techniques

Work on load models based on weather correlation has been done by Dryar, Davies, and Heinemann (Ref. 6, 7, 8). The basic approach is to find a deterministic relation between the peak load and the average daily weather effects considered significant. Davies utilizes average temperature, wind effects, illumination index and precipitation index. Heinemann uses a similar approach but introduces a "dynamic effect" due to "heat build-up". In his model he considers a relation between load peaks now and average weather effects for the past three days. These effects are wet and dry bulb temperature as well as relative humidity.

These techniques are limited to the modelling of load at fixed points in time or to peak loads. Load variations with time of day are not considered, neither do these techniques provide a measure of the uncertainty associated with the prediction.

Load correlation techniques

This approach is favoured by some authors as discussed earlier. The jist here is to make use of the most recent load data to extrapolate in some sense into the future.

The more significant contribution along this line is made by Farmer and Potton (Ref. 5), who also introduce a probabilistic structure into the load model. Essentially, load observations over a period of six weeks in the past are used to estimate the value of the autocorrelation function of the load. A finite Karhoum-Loeve expansion based on the found autocorrelation function eigenvalues and vectors is then used to model the load. The parameters of the expansion can be recursively updated to best fit the mosr recent observations.

This approach provides the user with a continuous forecast and error standard deviation rather than just forecasting a few values during the day. It is however limited to the length of time it can accurately forecast by the fact that it does not consider weather effects. In addition, under varying weather conditions it is necessary to reevaluate the model eigenvectors, a task of considerable complexity.

A similar approach is attempted by Christiaanse (Ref. 9). He models the load behaviour at intervals of one hour by a time series of periodic functions with a period of one week. The free parameters are recursively updated by new observations. The advantage of this technique lies in its relative simplicity of implementation, however it contains the inherent drawback of not describing the effect of weather separately. Thus during fast changing weather conditions or periods of heat or cold, the model weakens.

Toyoda et al. (Ref. 10) suggest a state space model which incorporates both the effect of weather, temperature and humidity, as well as the effect of the latest load observation. This approach is somewhat similar to the author's, however we attempt to incorporate more complexity into the model and make considerable use of adaptive identification techniques to estimate the parameters of the model dynamics rather than assuming that these are apriori known. In addition we verify the validity of the technique with real data experimentation.

I.2.4 Important Factors in Load Forecasting Techniques:

Based on the results obtained by the various authors it is possible to arrive at a set of significant objectives one should aim for in designing a load forecasting technique.

The two primary aims are the accuracy and length of the forecast and the complexity involved in evaluating the forecast. Generally the industry leans toward techniques which are readily implementable on-line, i.e. will not burden the available computer excessively. Accuracy of forecast is a more obscure objective, exact numbers depending on the particular company's operational objectives. This criterion is also closely related to the length of the forecast. Shorter prediction times result in better accuracy, but prediction times of 24 hours or more can cause errors to deteriorate badly unless we go to more complex techniques.

As mentioned above techniques which collect and process data on-line to update the forecast are desirable. A technique, in addition to being accurate, should determine the extent of the model's confidence in the forecast or an estimate of the prediction error. This would provide us with a quantitative criterion to decide if the model has failed, that is, to detect anomalous load behaviour.

Since exact models are a practical impossibility, it is desirable to have a feedback type scheme for on-line corrections of forecasts as new observations are made. This will reduce the sensitivity of the forecast to modelling errors.

Finally, the model should be adaptible to seasonal variations in the load behaviour without drastic model changes, by for example adjusting a number of parameters. This would allow us to minimize the number of models which would have to be stored to be used under varying conditions.

II.0 PROPOSED LOAD MODEL & FORECASTING TECHNIQUE

II.1 Preliminary Remarks:

In this chapter we develop a mathematical model for the load of a large power system. This modelling step is based on the observed behaviour of a typical load curve under varying conditions, the behaviour of the consumers under the same conditions, as well as pure assumption. The model is kept as simple as possible at the same time incorporating the main hypothesized structure and different effects. A number of free coefficients parameterize the model, their exact value to be determined by model fitting techniques, that is system identification.

In addition, we describe the estimation-prediction algorithms which are used for on-line forecasting. This is essentially a Kalman estimator-predictor scheme (Ref. 11).

II.2 Load Modelling Concepts:

We are dealing here with a time varying process which is inherently uncertain. Its value at any one time is directly determined by the customers who turn switches on and off. These are in turn influenced by a number of factors such as weather conditions and living patterns to use electricity with a certain regularity. Since we are dealing with a very large ensemble of people it is expected that these various influencing effects will be felt in the load in some systematic or regular way. Indeed, observations show that such regularities exist.

We wish then to describe these regularities mathematically as simply as possible, yet including as many effects as considered significant.

The choice of a right model is very important if we wish it to be valid for future times and forecast accurately. There is however little more one can do at this stage except justify the model structure based on the underlying basic laws of the process, and experimentally verify its validity. If it should not be accurate, the identification step would give us an idea of how to alter the model, by suggesting that additional variables or non-linearities may be needed.

II.3 Load Model Structure:

The hypothesis is made that the load, z, at any time of the day, t, can be expressed as,

$$z(t) = y_p(t) + y(t) \qquad (2.1)$$

where y_p is a periodic component and y is a residual term.

The assumption is made that y_p is dependent only on the time of the day and the day of the week. We will also assume that y_p is a deterministic process, so that its exact value is determinable from its model.

The residual term, y, is assumed to be an uncertain process, time varying and correlated with itself over time as well as with certain weather effects.

Again it should be emphasized that this is simply a model which makes sense from the point of view of load and customer behaviour, that is underlying influencing factors, however other models or variations of this idea may very well yield better results.. We stick to the concept of first trying the simplest model form and work from there.

The existence and structure of y_p and y are now justified and described respectively from a heuristic point of view.

II.4 Periodic Component:

This component attempts to describe that part of the load behaviour which depends only on the time of the day and day of the week.

That this component exists is justifiable in terms of the total observed load behaviour. Thus the load daily goes through a 24 hour near-periodic cycle which rises in the early morning, reaches a peak at mid-morning, may dip and rise again until the late afternoon drop, rising again during the evening, finally dropping considerably at night.

This approximate behaviour is consistently repeated, exhibiting similar rises and falls at approximately the same hours. In addition the shape of this characteristic curve is essentially unaffected by changes in weather conditions. In the examples of Fig. 3 and 4, this behaviour is seen. That is, for considerably large variations in temperature, the magnitude of the load is affected but its structure with time stays approximately the same.

The behaviour of y_p with the day of the week is also observable. Distinct differences in the daily load behaviour occur over the weekdays and in particular over weekends and holidays. Slighter differences also occur between the same weekdays separated by various week integers, especially over the seasons.

The existence of this component can also be justified from physical reasoning. The daily behaviour is clearly attributable to the consumer habits of power consumption centered abou their 24 hour cycles of sleep, eating, work and rest periods. Thus in the morning residential, commercial and industrial consumption pick up very rapidly. After noon time, there may be a decrease due to smaller residential loads, and as commercial and industrial loads decrease. A rise is again encountered at supper time due to cooking, lighting and television loads, which rapidly decays as people turn in.

The weekday variations can also be explained on this basis. Thus the cyclic daily consumption habits may be slightly different over the weekdays, and certainly over the weekends and holidays when industrial and commercial loads are mainly off.

The change in behaviour during longer time periods is again due to small accumulated changes in consumption habits, changes in eating and sleeping habits over the year. In particular during vacation periods, considerable changes in load consumption may occur.

The assumed independence of y_p for short periods of time with weather conditions is based on the reasoning that the "normal" consumption habits which y_p describes are not affected by weather as suggested in Fig. 3 and 4. This is due to the fact that when weather conditions deviate from normal then the weather sensitive load is excited beyond a normal level which is itself assumed unaffected.

Any deviation of the load from its normal level, whether due to weather deviations from normal or due to random effects, will be modelled by the residual term, y.

The periodic oomponent of load, y_p, is therefore a hypothetical deterministic process which is justified on the basis of observed load near-periodicity. The residual term, y, is in turn another hypothetical process whose structure will be discussed below. Their existence will be more definitely justified after experimental testing of the model with real data.

The structure of the periodic component, y_p, can be expressed as a time series,

$$y_p(t) = x_p^0 + \sum_{i=1}^{n_p}\{x_p^i \sin[2\pi i/24]t + x_p^{n_p+i}\cos[2\pi i/24]t\} \quad (2.2)$$

which we can rewrite in vector form as,

$$y_p(t) = \underline{\phi}^T(t)\,\underline{x}_p \qquad (2.3)$$

where defining,

$$\omega_o = 2\pi/24 \qquad (2.4)$$

then,

$$\underline{\phi}(t) = \begin{pmatrix} 1 \\ \sin \omega_o t \\ \cdot \\ \cdot \\ \cdot \\ \sin n_p \omega_o t \\ \cos \omega_o t \\ \cdot \\ \cdot \\ \cdot \\ \cos n_p \omega_o t \end{pmatrix} \qquad (2.5)$$

$$\underline{x}_p = \begin{pmatrix} x_p^0 \\ \cdot \\ \cdot \\ \cdot \\ x_p^{2n_p} \end{pmatrix} \qquad (2.6)$$

while t stands for the time of the day.

The vector $\underline{x}_p$ is assumed constant, however as suggested earlier small variations over the weekdays are possible. These are particularly noticeable on Saturday and Sunday. In this study we first consider three sets of $\underline{x}_p$, one for Monday through Friday, one for Saturday and one for Sunday, but later experiments indicate that a separate Monday model may be desirable.

Over the span of two or three weeks the value of $\underline{x}_p$ may remain constant for a given weekday, but since normal load consumption does vary over the seasons we will expect that $\underline{x}_p$ should vary accordingly. For these reasons this parameter will require periodic readjustment, for example once a week.

The number of elements in the vector $\underline{x}_p$, $2n_p+1$, is presently uncertain, its value to be determined by identification techniques. However it will later be shown by experimentation to lie between 9 and 15.

II.5 <u>Residual Component, y</u>:

Under "normal" conditions our model would say that the load, z, should equal the periodic component, y_p. These "normal" conditions are however highly hypothetical since the load is never exactly periodic. Thus even if all the weather effects considered significant are at their normal level, it is reasonable to expect a random variation around the periodic behaviour. This random variation is provided by the residual component, y, which we assume to be a random process depending on the deviations of significant weather variables from their normal level. More about this dependence will follow.

The significant weather variables are those that influence the behaviour of the load. These are in order of significance (Ref. 12) temperature, humidity, precipitation, light intensity, and wind speed, respectively denoted by T, H, P, L and S. These variables have different influences depending on the region and power company.

Temperature is the single most important effect as a great portion of the load is temperature sensitive, e.g. refrigeration, heating and air-conditioning. This is particularly important in large urban areas. The basic influence of T is to increase the load with its increase, but this behaviour is reversed during the cold months. The remaining variables have similar effects, all of considerably less importance, the exact value being difficult to establish, depending very much on the specific company.

The problem with analyzing the effect of weather conditions other than temperature is the difficulty in obtaining significant data. For example, wind speed, precipitation and light intensity may vary considerably over the area of the power company affecting certain portions of the load more than others. Such data is not presently available.

Temperature and humidity are much more uniformly distributed over the load region and can be periodically monitored without great problems. In addition this data is readily available from the weather bureau.

In this study we restrict ourselves to the effect of temperature only for simplicity's sake, but its extension to other weather effects could follow along similar lines in more ambitious investigations.

II.5.1 The Effect of Temperature:

In this study we try to analyze the effect of the temperature profile with time on the load, rather than that of some average daily temperature. The reasoning being that different temperatures during the course of the day will cause the load to behave differently even if the average daily temperature stays normal.

Now, since y_p models the normal load behaviour while y models deviations from normal, we should define the normal temperature level. A possibility is to choose some constant level, such as 65°F, or some normal average daily temperature. This choice would however yield on the average a temperature deviation profile which will be periodic, due to the warmer temperatures during the day and cooler ones at night. Such effects should however be describable by the term y_p. Instead we have chosen a normal daily temperature curve, $\hat{T}$, as the normal. This curve is averaged from weather bureau measurements over a period of 10 years or more, yielding monthly average daily temperatures at hourly intervals.. As shown in Fig. 5 it is approximately periodic over a day and deviations from this level are small and normally non-periodic. Interpolation allows us to calculate the normal daily temperature curve for any day of the year.

Deviations of the actual temperature, T, from the normal temperature, $\hat{T}$, for every interval of the day chosen (e.g. hours), are much smaller now, which gives us considerably more confidence in hypothesizing linear models between y and the temperature deviations. This assumption should make y_p dependent on normal temperature levels, but since these are very slowly varying as seen in Fig. 5, the corresponding variations in y_p should be identifiable.

The inputs to our model of y, the residual load, will then be the deviations in T from $\hat{T}$, $\Delta T = T - \hat{T}$, for every chosen interval of the day (e.g. hours). However some additional

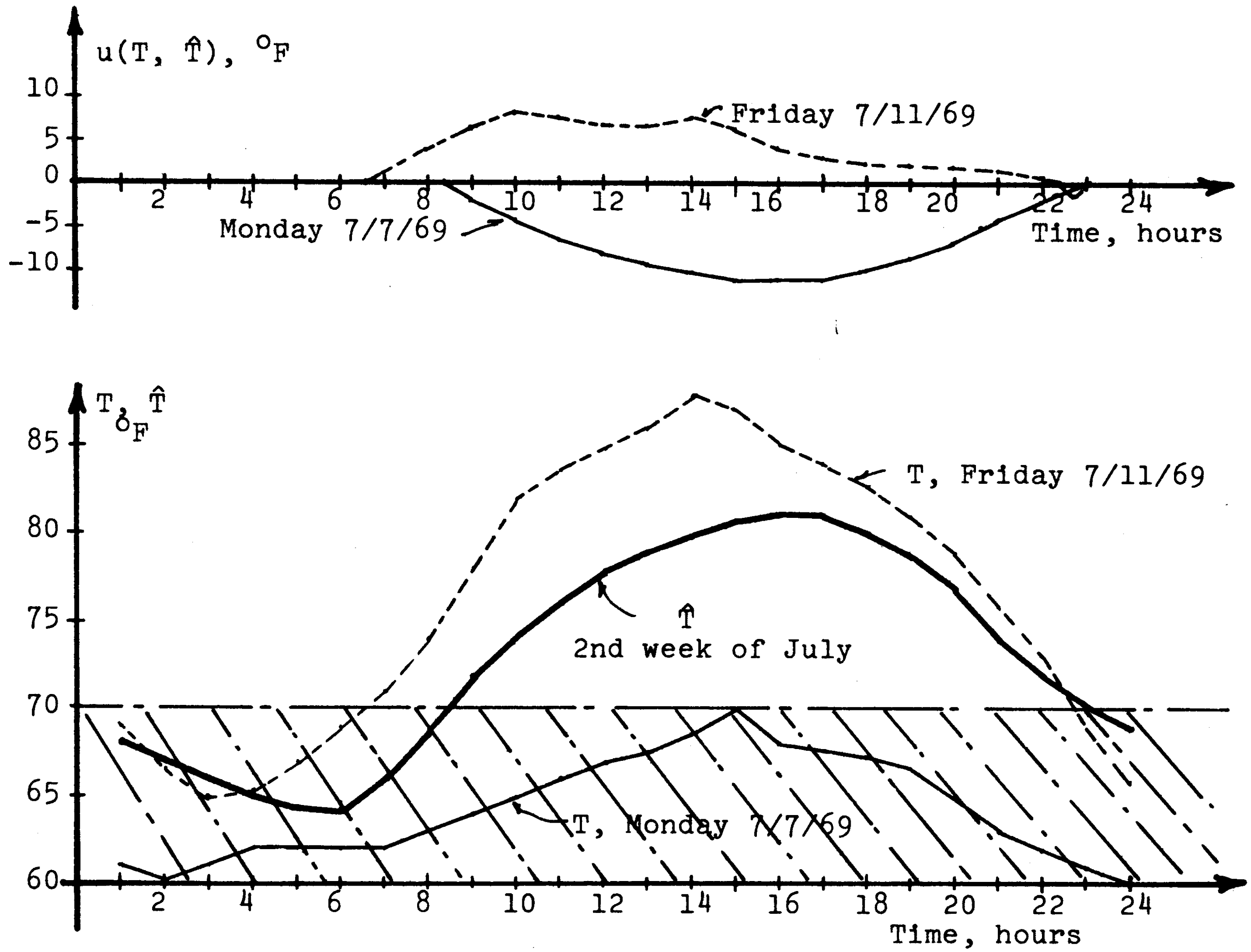


Fig. 5: Examples of Normal and Actual Daily Temperature Curves, and Resulting Temperature Deviations, u.

discussion on the inputs is needed before we propose a temperature dependent model.

Various possibilities come to mind. If the normal temperature $\hat{T}$ is high, air conditioning and refrigeration form the predominant temperature sensitive load so that we would expect positive deviations in T from $\hat{T}$ to induce an increase in load; alternatively negative deviations will induce a decrease in load. At the other extreme, if $\hat{T}$ is low then the heating load becomes dominant so that now negative deviations from normal will induce an increase in load while positive ones the opposite.

How to separate the two regions? Studies on human comfort zones with respect to temperature indicate that between 60°F and 70°F most people will be comfortable (Ref. 13). Above this region heat discomfort will predominate while below it cold discomfort is most important.

Thus, depending on the value of $\hat{T}$ the effect of temperature will be either to increase load with increasing T or viceversa. In order to use only one model at a time for the effect of temperature on load we then define the input to this model by $u(T, \hat{T})$, given by the following chart.

T \ $\hat{T}$	below 60°F	60°F – 70°F	above 70°F
above 70°F	Cooling+Heating $[T-70]-[60-\hat{T}]$ $u = T+\hat{T}-130$	Cooling $u = T-70$	Cooling $u = T - \hat{T}$
60°F – 70°F	Heating $u = -[60-\hat{T}]$	no effect $u = 0$	Cooling $u = 70 - \hat{T}$
below 60°F	Heating $u = -[T-\hat{T}]$	Heating $u = -[T-60]$	Cooling+Heating $[70-\hat{T}]-[T-60]$ $u = 130-T-\hat{T}$

Table 1: Definition of $u(T, \hat{T})$ in terms of actual and normal Temperatures.

The chart is self-explanatory but we will describe a few cases:

A) $T \geq 70°$, $\hat{T} \geq 70°$. This is a purely cooling region. Positive deviations in T from $\hat{T}$ will increase load while negative ones will decrease it. The input to our model is thus $u=T-\hat{T}$. This will occur primarily in the summer.

B) $T \leq 60°$, $\hat{T} \leq 60°$. Here heating is predominant so that we have the reverse of case A. We have defined $u=-[T-\hat{T}]$ so that the same model as in part A may be used to yield the expected results. This is done in case that during a given day, T and $\hat{T}$ vary between regions A and B. Region B will however occur primarily during the winter.

C) $60° \leq T \leq 70°$, $60° \geq \hat{T} \geq 70°$. In this region variations of T from normal are considered negligible. Thus u=0.

D) $T \geq 70°$, $60° \leq \hat{T} \leq 70°$. Here cooling is expected predominant, however since $\hat{T}$ is in a "no effect" region, only deviations above this region will be significant. Thus u=T-70.

E) $60° \leq T \leq 70°$, $\hat{T} \leq 60°$. Here heating predominates. As temperature rises above $\hat{T}$, load will decrease, but only up to the boundary of the no effect region, i.e. 60°F. Thus $u=-[60-\hat{T}]$.

F) $T \leq 60°$, $\hat{T} \geq 70°$. Such situations are unusual but may happen, especially in the spring and fall. Since the normal temperature $\hat{T}$ is above 70°F we would expect to be in a cooling region. Cooling load will decrease with T but only down to the boundary of the no effect region, i.e. 70°F, hence the term $[70-\hat{T}]$. While T is in the no effect region no further cooling load decrease occurs, but below 60°F further decrease in temperature may increase heating load hence the term [60-T]. These two effects may have to be weighted differently, but in this study we weigh them equally.

It is quite clear that the relation between T and $\hat{T}$, and u, the weather variable influencing y, is highly nonlinear.

Examples: Consider some particular days when actual temperature and normal temperature profiles are as given below.

Hour / °F	1	4	7	10	13	16	19	22
T	61	62	62	65	67	68	67	62
$\hat{T}$	67	64	65	73	78	80	78	71
u	u=0	u=0	u=0	$u=70-\hat{T}$ $= -3$	$u=70-\hat{T}$ $= -8$	$u=70-\hat{T}$ $= -10$	$u=70-\hat{T}$ $= -8$	$u=70-\hat{T}$ $= -1$
T	78	78	75	81	75	79	80	75
$\hat{T}$	68	66	66	75	80	82	80	73
u	$u=T-70$ $= 8$	$u=T-70$ $= 8$	$u=T-70$ $= 5$	$u=T-\hat{T}$ $= 6$	$u=T-\hat{T}$ $= -5$	$u=T-\hat{T}$ $= -3$	$u=T-\hat{T}$ $= 0$	$u=T-\hat{T}$ $= 2$

Table 2: <u>Examples of Calculation of Temperature Deviation</u>.

II.5.2 <u>Relation Between u and y</u>:

In the previous section we argued the existence of a variable u depending on both actual temperature, T, and normal temperature, $\hat{T}$, whose value describes deviations in temperature from a normal level and is therefore related to the residual load, y. Now we will propose a relation between u and y.

Studies have shown that the effect of temperature on load is not instantaneous (Ref. 8), but that time constants of up to 48 hours may exist. This can be justified by the heat storage property of buildings which may take some time to cool off or warm up. This behaviour is thus clearly dynamical in nature.

Finally since the excitation, u, is a small deviation about a normal level we will hypothesize that y and u are the output and input of a linear dynamical system. Since we are dealing with discrete time processes we will then propose the following relation,

$$y(t) = \sum_{i=1}^{n} a_i y(t-i) + \sum_{j=0}^{m} b_j u(t-j) \qquad (2.7)$$

where a_i and b_j are constant coefficients. The dimensions of the model, n and m, as well as the parameters are unknown at this level, to be determined later by identification techniques. This type of model, called an auto-regressive-moving-average model, may include non-linear dynamical effects in u.

II.5.3 Uncertainty in Residual Load Model:

As mentioned earlier, the load behaviour is inherently uncertain, furthermore our assumptions are quite hypothetical so that modelling errors are likely. In such cases it would be naive to suggest a deterministic model. Instead we hypothesize one with modelling uncertainty. This means that the mathematics will not yield an exact value of y, but an uncertain measure of its value. This measure could be probabilistic, or unknown but bounded. With such a model we lose some knowledge of y, but get a more realistic description of its behaviour and much more confidence in the result even though uncertain. We may wish to view this uncertainty as the lumped effect of all the variables which we are not considering in our model, such as the weather variables H, S, L, P and other influences, as well as higher than first order effects or lack of sufficient parameters.

In this study we restrict ourselves to probabilistic models as the unknown but bounded ones are more difficult to identify and analyze (Ref. 14).

The simplest way to introduce uncertainty into the model for y is to alter equation 2.7 slightly, that is,

$$y(t) = \sum_{i=1}^{n} a_i y(t-i) + \sum_{j=0}^{m} b_j u(t-j) + c(t) \qquad (2.8)$$

where c(t) is a correlated process which has the following first and second order statistical properties,

$$E\{c(t)\} = 0 \quad ; \quad \forall\, t \qquad (2.9)$$

$$E\{c(t)\, c(\tau)\} = R_c(t-\tau) \quad ; \quad \forall\, t,\, \tau \qquad (2.10)$$

where E stands for the expected value while,

$$R_c(t-\tau) = 0 \quad ; \quad \forall\, |t-\tau| \geq T_c \qquad (2.11)$$

Thus c is correlated with itself but only up to a certain level, T_c.

It is convenient to express c as a moving average of another white process, w, as below,

$$c(t) = \sum_{k=1}^{p} d_k w(t-k) \qquad (2.12)$$

The constant parameters d_k, k=1, 2, ..., p, as well as p are chosen so that together with the relations,

$$E\,[w(t)] = 0 \qquad (2.13)$$

plus,

$$E\,[w(t)\, w(\tau)] = \begin{cases} 0 & ; \ t \neq \tau \\ Q & ; \ t = \tau \end{cases} \qquad (2.14)$$

both right and left sides of 2.12 have equal first and second order statistical properties. If c and w happen to be Gaussian then this implies that c and $\sum_{k=1}^{p} d_k w(t-k)$ have equal probability distributions.

The equalities require that,

$$R_c(t-\tau) = \sum_{i=1}^{p} \sum_{j=1}^{p} d_i d_j \; E\,\{w(t-i)\; w(\tau-j)\} \tag{2.15}$$

$$= \sum_{j=1}^{p} d_j d_{j+t-\tau} \; Q \tag{2.16}$$

We will then assume that the complete relation between y and u will be,

$$y(t) = \sum_{i=1}^{n} a_i y(t-i) + \sum_{j=0}^{m} b_j u(t-j) + \sum_{k=1}^{p} d_k w(t-k) \tag{2.17}$$

where the a_i, b_j and d_k parameters as well as their numbers are at this time unknown.

As with y_p, we will assume that the parameters of the model for y may be different over the weekdays. We will have a set of parameters for the weekdays and two others for Saturday and Sunday respectively. As before also we may expect these parameters to vary slowly with the seasons.

This model then describes the residual load, y, as a random process, correlated with its previous values whose mean value, $\hat{y}(t)$, depends on u and whose variance depends on the modelling uncertainty process w.

The identification of the model parameters as well as their number will be discussed in the next chapter. In the next section we assume that these are known and describe the application of Kalman filtering to the complete model for purposes of load prediction. First, however we express the complete model in state space form which is convenient for some of our objectives, e.g. computer applications.

II.6 State Space Model Description:

The complete load model can be written as a set of first order equations in a number of ways. Here we present but one,

$$\underline{x}_p(t+1) = \underline{x}_p(t) \qquad (2.18)$$

$$\underline{x}(t+1) = \underline{A}\,\underline{x}(t) + \underline{b}\,u(t) + \underline{d}\,w(t) \qquad (2.19)$$

$$z(t) = \underline{\phi}^T(t)\,\underline{x}_p(t) + \underline{c}^T\underline{x}(t) + b_o u(t) \qquad (2.20)$$

where,

$$\underline{A} = \begin{pmatrix} a_1 & 1 & 0 & . & . & . & 0 \\ a_2 & 0 & 1 & 0 & . & . & 0 \\ . & . & & & & & . \\ . & . & & & . & & . \\ . & . & & & & . & 0 \\ . & . & & & & & 1 \\ a_n & 0 & . & . & . & . & 0 \end{pmatrix} \qquad (2.21)$$

$$\underline{b} = \begin{pmatrix} b_1 + a_1 b_0 \\ b_2 + a_2 b_0 \\ . \\ . \\ . \\ b_n + a_n b_0 \end{pmatrix} \qquad (2.22)$$

$$\underline{d} = [\, d_1, d_2, \ldots, d_n]^T \qquad (2.23)$$

$$\underline{c} = [\, 1, 0, \ldots, 0]^T \qquad (2.24)$$

The fact that we have different models for Monday, Saturday, Sunday and the remaining weekdays can be interpreted as having a time varying model whose system parameters $\underline{A}$, $\underline{b}$,

$\underline{d}$, $\underline{c}$ and Q change to a new constant value at the end of each corresponding period.

The state of this load model is thus the vector $\underline{\tilde{x}}$ = $[\, \underline{x}_p, \underline{x}]^T$. Its knowledge at any one time completely characterizes its future behaviour and hence that of z, the load, given all future inputs.

II.7 Corrective Load Prediction Scheme:

Since we have argued for and developed a probabilistic model of the load, z, we may describe its behaviour in terms of the first and second statistics, mean and standard deviation. The mean gives us a deterministic measure of where z will on the average be found, whereas the standard deviation is a deterministic measure of how far from the mean will the actual value deviate on the average. Clearly for a good model this deviation should not be excessively large.

It is possible to develop a model for the propagation with time of the apriori* mean and variance of the state of the load model (Ref. 14) and hence of z, based on the apriori statistics of the systemnnoise, w, that is Q, as well as some apriori information about $\underline{\tilde{x}}$.

This is readily derived. From equation 2.18, 2.19, 2.20, taking expected values of both sides and letting,

$$E\{\underline{x}(t)\} = \underline{\hat{x}}(t) \qquad (2.25)$$

and,

$$E\{z(t)\} = \hat{z}(t) \qquad (2.26)$$

* Also referred to later as open loop estimation.

then we have for the mean values,

$$\underline{x}_p(t+1) = \underline{x}_p(t) \tag{2.27}$$

$$\hat{\underline{x}}(t+1) = \underline{A}\ \hat{\underline{x}}(t) + \underline{b}\ u(t) \tag{2.28}$$

$$\hat{z}(t) = \underline{\phi}^T(t)\ \underline{x}_p(t) + \underline{c}^T\hat{\underline{x}}(t) + b_o u(t) \tag{2.29}$$

Defining now $\underline{\Gamma}$, the state estimate error variance,

$$\underline{\Gamma}(t) = E\{\ [\underline{x}(t)-\hat{\underline{x}}(t)]\ [\underline{x}(t)-\hat{\underline{x}}(t)]^T\} \tag{2.30}$$

and Γ_z, the load estimate error variance,

$$\Gamma_z(t) = E\ \{[z(t)-\hat{z}(t)]^2\} \tag{2.31}$$

then it is easy to show that,

$$\underline{\Gamma}(t+1) = \underline{A}\ \underline{\Gamma}(t)\ \underline{A}^T + \underline{d}\ Q\ \underline{d}^T \tag{2.32}$$

and that,

$$\Gamma_z(t) = \underline{c}^T\ \underline{\Gamma}(t)\ \underline{c} \tag{2.33}$$

so that the model for $\hat{z}(t)$ and $\Gamma_z(t)$ could be used for modelling and prediction purposes.

This type of model does not however take full advantage of all available information in estimating the present state.

At all times we are observing the load z which contains some information about the state through the relation,

$$z(t) = \underline{\phi}^T(t)\ \underline{x}_p(t) + \underline{c}^T\underline{x}(t) \tag{2.34}$$

Instead of defining $\hat{\underline{x}}(t)$, the estimate of $\underline{x}(t)$, as its apriori expected value, one can define it as,

$$\hat{\underline{x}}(t) = E\ \{\underline{x}(t)\ /\ z(n)\ ;\ n\leq t\} \tag{2.35}$$

This estimate uses all available past observations and it can be shown (Ref. 14) that the corresponding error covariance, $\underline{P}(t)$, is a minimum if the noise drive, w, is Gaussian. If the

Gaussian assumption does not hold, as is normally the case in practice, we can define $\hat{\underline{x}}$ as, $\hat{\underline{x}}(t)$ = best linear estimate of $\underline{x}(t)$ given all $z(n)$; $n<t$ in the sense that $\underline{P}$, the error covariance matrix is minimized, where,

$$\underline{P}(t) = E\,\{[\underline{x}(t)-\hat{\underline{x}}(t)]\,[\underline{x}(t)-\hat{\underline{x}}(t)]^T \,/\, z(n)\ ;\ n<t\} \qquad (2.36)$$

It will be shown later that this yields a better state estimate than in the previous case and therefore that the resulting prediction is correspondingly better.

The set of equations which describes the updating of the state estimate with time as new observations are made, plus the propagation of the error covariance matrix is commonly known as a Kalman filter, and is well documented (Ref. 15). These are,

$$\hat{\underline{x}}(t+1) = \underline{A}\,\hat{\underline{x}}(t) + \underline{b}\,u(t) + \underline{g}(t+1)[z(t+1)-\hat{z}(t+1)] \qquad (2.37)$$

where,

$$\hat{z}(t+1) = \underline{c}^T[\underline{A}\,\hat{\underline{x}}(t)+\underline{b}\,u(t)]+b_o u(t+1)+\underline{\phi}^T(t+1)\underline{x}_p(t+1) \qquad (2.38)$$

and,

$$\underline{g}(t+1) = \underline{S}(t+1)\,\underline{c}\,[\underline{c}^T\underline{S}(t+1)\underline{c}]^{-1} \qquad (2.39)$$

$$\underline{S}(t+1) = \underline{A}\,\underline{P}(t)\underline{A}^T + \underline{d}\,Q\,\underline{d}^T \qquad (2.40)$$

$$\underline{P}(t+1) = \underline{S}(t+1)-\underline{S}(t+1)\underline{c}[\underline{c}^T\underline{S}(t+1)\underline{c}]^{-1}\underline{c}^T\underline{S}(t+1) \qquad (2.41)$$

This set of equations is depicted in Fig. 6 and 7.

II.7.1 Prediction Scheme:

The above discussion describes a set of equations which yield a best linear estimate of the state of the load model from observed data. In addition we are provided with a measure of the uncertainty associated with this estimate.

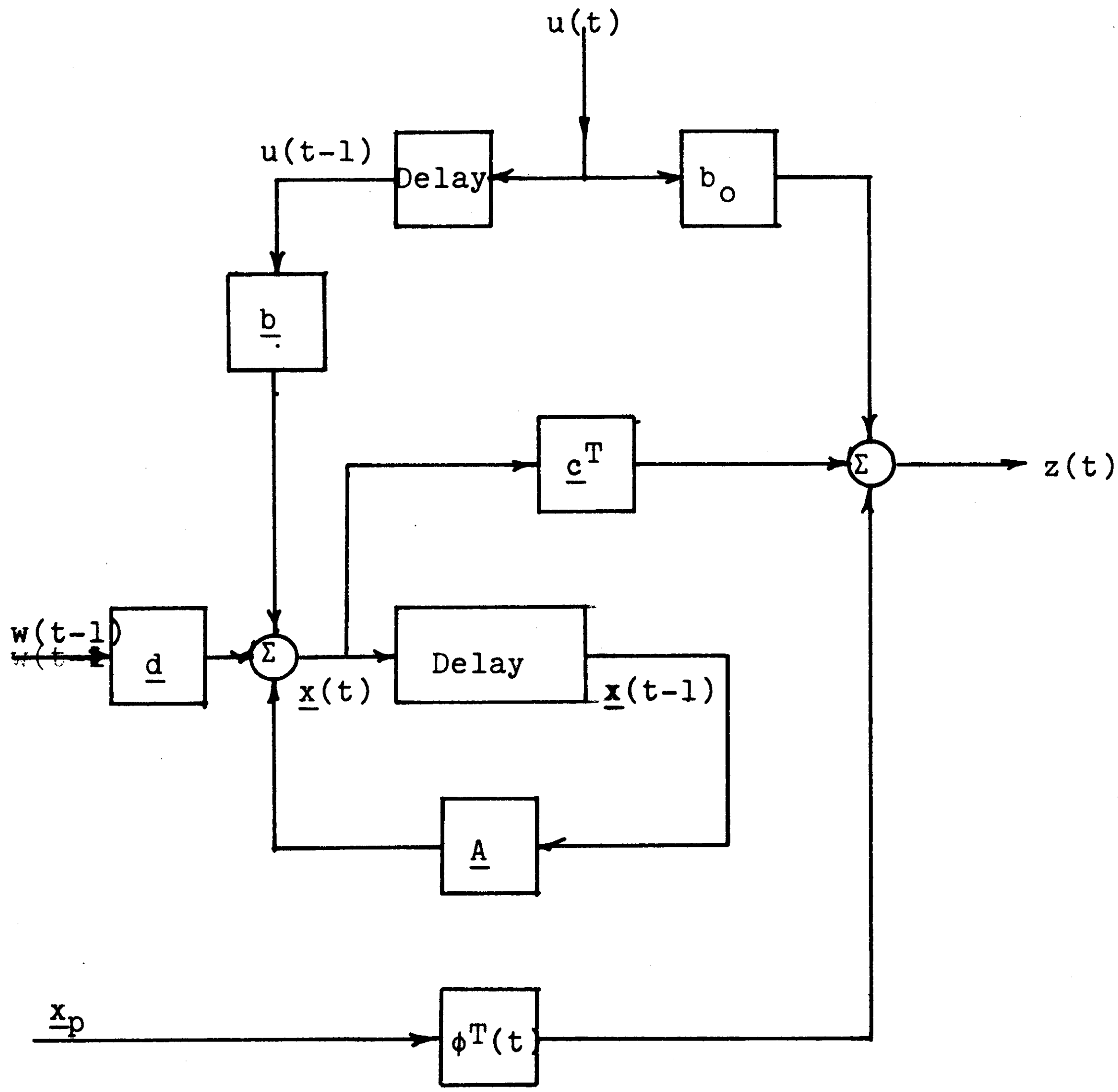


Fig. 6: <u>Assumed Stochastic Load Model</u>.

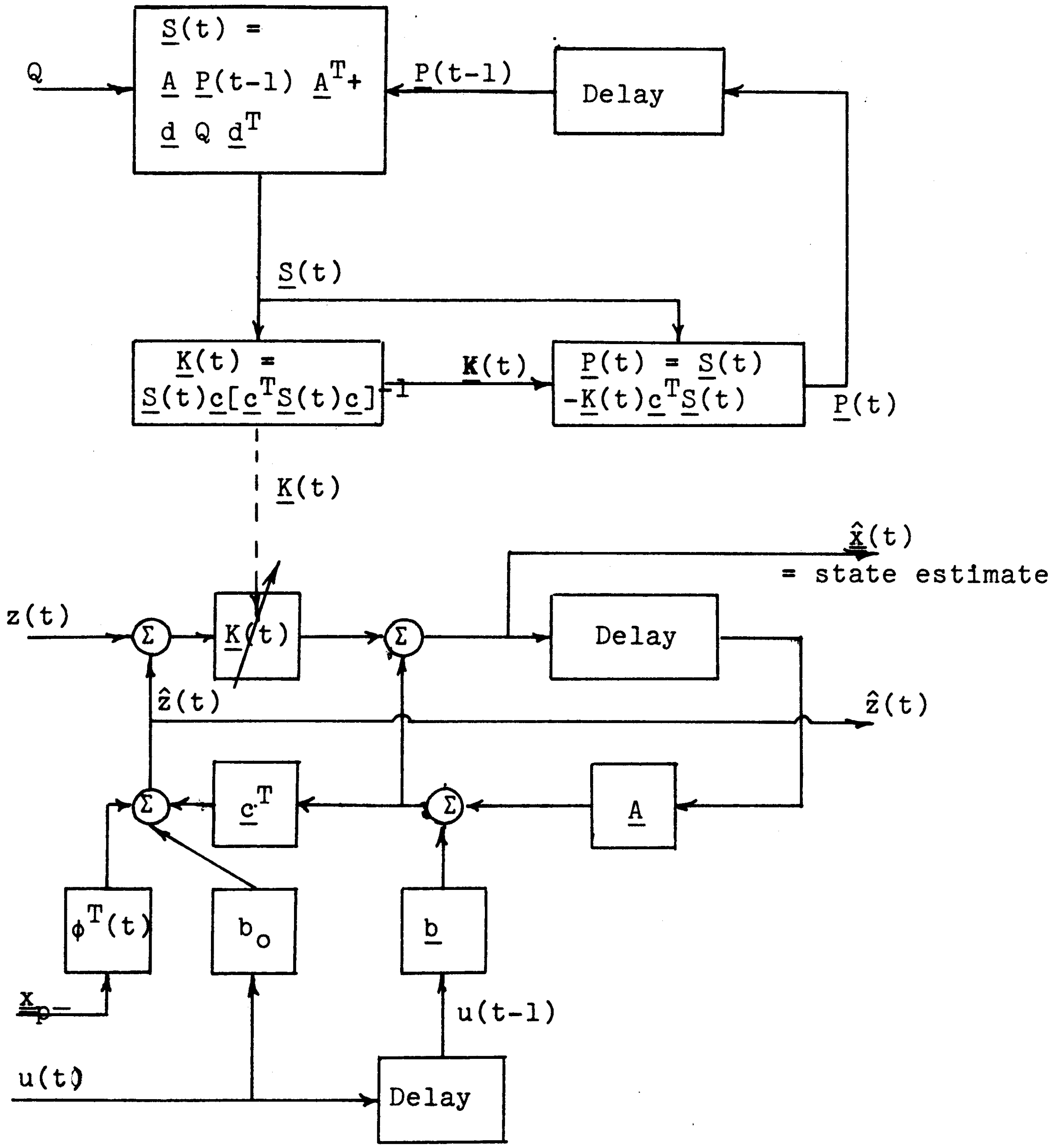


Fig. 7: Estimation of State - Deterministic Load Model.

This estimate of the state, at any time t, can now be used to predict future states and hence future load values. The equations here are given by, (Ref. 14),

$$\underline{\hat{x}}(t+n+1/t) = \underline{A}\ \underline{\hat{x}}(t+n/t) + \underline{b}\ u(t+n) \qquad (2.42)$$

with the initial condition,

$$\underline{\hat{x}}(t/t) = \underline{\hat{x}}(t) \qquad (2.43)$$

while the associated error variance,

$$\underline{\Sigma}(t+n/t) = E\{[\underline{x}(t+n)-\underline{\hat{x}}(t+n/t)][\underline{x}(t+n)-\underline{\hat{x}}(t+n/t)]^T\} \qquad (2.44)$$

is propagated by,

$$\underline{\Sigma}(t+n+1/t) = \underline{A}\ \underline{\Sigma}(t+n/t)\ \underline{A}^T + \underline{d}\ Q\ \underline{d}^T \qquad (2.45)$$

with initial condition,

$$\underline{\Sigma}(t/t) = \underline{P}(t) \qquad (2.46)$$

The predicted load for n units ahead of t, the present time, is,

$$\hat{z}(t+n/t) = \underline{\phi}^T(t+n)\underline{x}_p + \underline{c}^T\underline{\hat{x}}(t+n/t) + b_o u(t+n) \qquad (2.47)$$

while the forecast error variance is,

$$E\{[z(t+n)-\hat{z}(t+n/t)]^2\} = \underline{c}^T\ \underline{\Sigma}(t+n/t)\underline{c} \qquad (2.48)$$

The above equations are depicted in Fig. 8.

II.7.2 Effect of Uncertainty in Weather Forecast:

The above set of equations are derived based on the assumption that future weather, i.e. u, is perfectly known. Now we discuss the very realistic possibility of prediction uncertainty in u.

Suppose we are given by the weather bureau a set of future values $\hat{u}$ such that,

$$\hat{u}(t) = u(t) + v(t) \qquad (2.49)$$

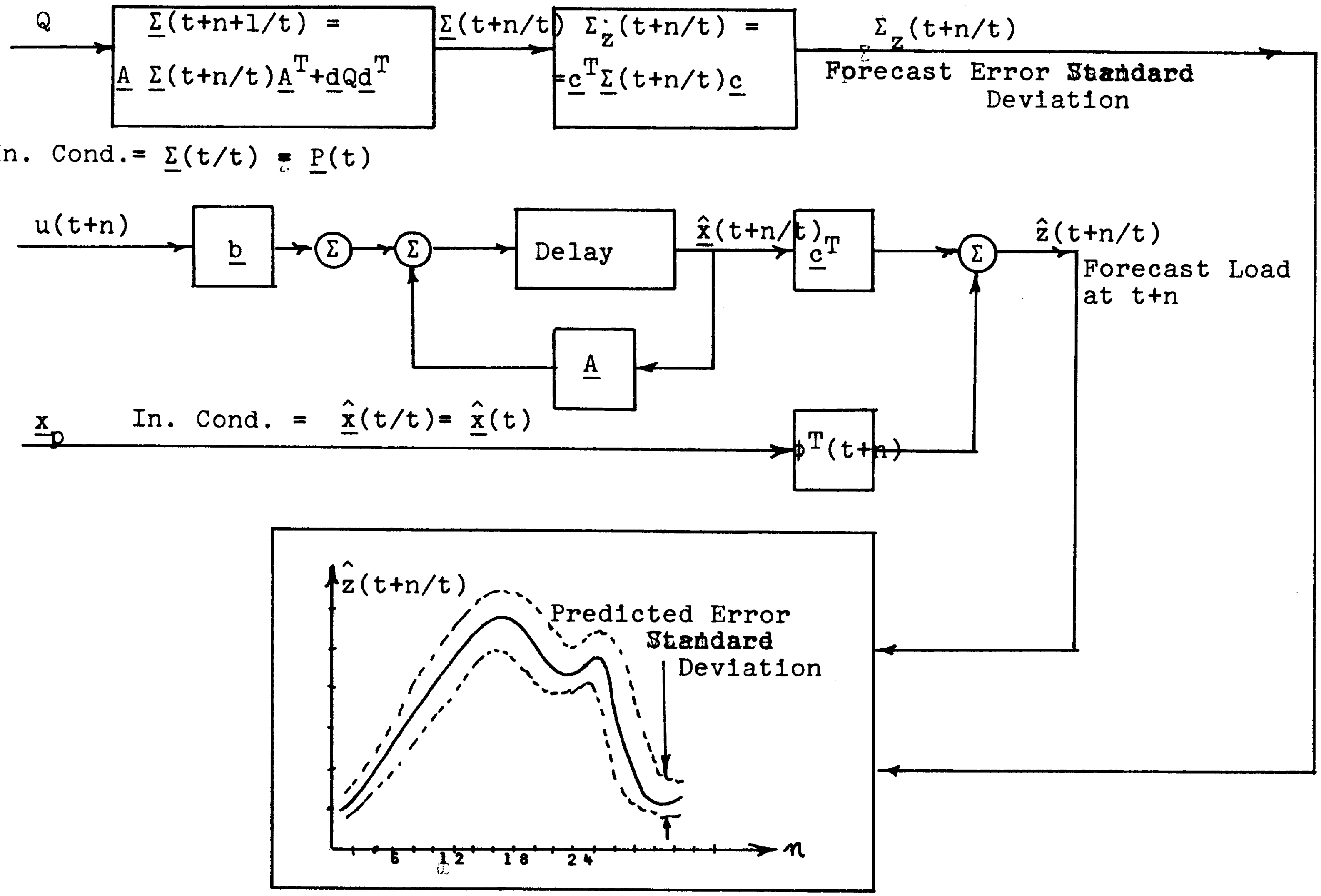


Fig. 8: Display of Forecast Load at Time t plus Associated Error Variance.

where v is some error quantity and u is the true value. We will assume a simple model for errors in the temperature deviation forecast, v, although more or less complex ones can also be treated. Suppose that,

$$v(t) = (t-t_o)\ v_o \tag{2.50}$$

where v_o is a random variable with,

$$E\{v_o\} = 0 \tag{2.51}$$

$$E\{v_o^2\} = R \tag{2.52}$$

Suppose now that we are at time t_o. The present estimate of the state, $\hat{\underline{x}}(t_o)$, and its covariance, $\underline{P}(t_o)$, are independent of future values of u or of their uncertainty since u is assumed perfectly known at the present time. Our stochastic model for y, the residual load, says that for $t>t_o$,

$$\underline{x}(t+1) = \underline{A}\ \underline{x}(t) + \underline{b}\ u(t) + \underline{d}\ w(t) \tag{2.53}$$

$$y(t+1) = \underline{c}^T\underline{x}(t+1) + b_o u(t+1) \tag{2.54}$$

Since,

$$u(t) = \hat{u}(t) - v(t) \tag{2.55}$$

$$= \hat{u}(t) - (t-t_o)\ v_o \tag{2.56}$$

we can write 2.53 and 2.54 as,

$$\underline{x}(t+1) = \underline{A}\ \underline{x}(t) + \underline{b}\ \hat{u}(t) - [t-t_o]\underline{b}\ v_o + \underline{d}\ w(t) \tag{2.57}$$

$$y(t+1) = \underline{c}^T\underline{x}(t+1) + b_o\hat{u}(t+1) - [t+1-t_o]\ b_o v_o \tag{2.58}$$

Now the best linear prediction n steps ahead of t_o for $\underline{x}$ and y, given observations of z up to time t_o, is given by,

$$\hat{\underline{x}}(t_o+k+1/t_o) = \underline{A}\ \hat{\underline{x}}(t_o+k/t_o) + \underline{b}\ \hat{u}(t_o+k)\ ;\ k=0,\ 1,\ldots n-1 \tag{2.59}$$

with initial condition $\underline{x}(t_o)$, while the predicted load is,

$$\hat{z}(t_o+n/t_o) = \underline{\phi}^T(t_o+n)\ \underline{x}_p + \underline{c}^T\hat{\underline{x}}(t_o+n/t_o) + b_o\hat{u}(t_o+n) \tag{2.60}$$

The confidence in this prediction is again given by the covariance of the state prediction error. The corresponding equations are now (assuming w and v are independent),

$$\underline{\Sigma}(t_o+n+1/t_o) = \underline{A}\ \underline{\Sigma}(t_o+n/t_o)\ \underline{A}^T + \underline{d}\ Q\ \underline{d}^T + \underline{b}n^2R\underline{b}^T \quad (2.61)$$

with initial condition $\underline{P}(t_o)$.

We can see that equation 2.61 is similar to 2.45 except for the added term $\underline{b}n^2R\underline{b}^T$ due to weather forecasting errors. It can also be seen that due to our model for weather forecasting errors, the load prediction uncertainty becomes progressively worse.

II.8 Uncertainty in System Parameters:

The previous equations are derived under the assumption that the system parameters $\underline{x}_p$, $\underline{A}$, $\underline{b}$, $\underline{d}$ and Q are known quantities. In the next chapters we will discuss methods of estimating these.

Note that the effect of uncertainty in the initial estimate of $\underline{x}$, $\hat{\underline{x}}(0)$, is not assumed important. This is due to the fact that the Kalman filter is asymptotically stable, so that in steady state the effect of initial condition uncertainty in $\underline{x}$ becomes negligible (Ref. 11).

Uncertainty in the parameters $\underline{A}$, $\underline{b}$, $\underline{d}$ and Q is quite important, however the feedback corrective nature of the filter is such as to make the result relatively insensitive to small deviations in these parameters (Ref. 15).

Throughout the earlier discussions we also assumed that $\underline{x}_p(t)=\underline{x}_p(0)\equiv\underline{x}_p$ is known, and correspondingly no corrections on that part of the state are made by the Kalman filter.

In actuality we don't know the value of $\underline{x}_p$, perhaps some initial guess close to its true value. In the next chapter we discuss various methods of getting at $\underline{x}_p$, however here it is significant to show how this parameter could be found using Kalman filtering techniques.

We will assume that $\underline{A}$, $\underline{b}$, $\underline{d}$ and Q are known and that $\underline{x}_p$ is as described by equation 2.18, that is a constant, except that its initial value is a random variable with a certain covariance matrix (uncertainty) associated with its initial guess or estimate. We will call the latter $\hat{\underline{x}}_p(0)$, and the former $\underline{S}_p(0)$.

Under this assumption we can estimate the entire state by a Kalman filter, as new data is observed,

$$\hat{\underline{x}}_p(t+1) = \hat{\underline{x}}_p(t) + \underline{g}_p(t+1)\,[z(t+1)-\hat{z}(t+1)] \tag{2.62}$$

$$\hat{\underline{x}}(t+1) = \underline{A}\,\hat{\underline{x}}(t) + \underline{b}\,u(t) + \underline{g}(t+1)\,[z(t+1)-\hat{z}(t+1)] \tag{2.63}$$

$$\hat{z}(t+1) = \underline{\phi}^T(t+1)\hat{\underline{x}}_p(t+1)+\underline{c}^T[\underline{A}\hat{\underline{x}}(t)+\underline{b}u(t)]+b_o u(t+1) \tag{2.64}$$

Defining $\underline{\Xi}$ as the covariance matrix of the total error, that is,

$$\underline{\Xi}(t) = E\left\{\left(\begin{pmatrix}\underline{x}_p(t)\\ \underline{x}(t)\end{pmatrix} - \begin{pmatrix}\hat{\underline{x}}_p(t)\\ \hat{\underline{x}}(t)\end{pmatrix}\right)\left(\begin{pmatrix}\underline{x}_p(t)\\ \underline{x}(t)\end{pmatrix} - \begin{pmatrix}\hat{\underline{x}}_p(t)\\ \hat{\underline{x}}(t)\end{pmatrix}\right)^T\right\} \tag{2.65}$$

then we define,

$$\underline{\Xi}(t) \equiv \begin{pmatrix}\underline{P}_p(t) & \underline{P}_{p1}(t)\\ \underline{P}_{p1}^T & \underline{P}_1(t)\end{pmatrix} \tag{2.66}$$

and defining $\underline{\P}$ as,

$$\underline{\P}(t) \equiv \begin{pmatrix} \underline{S}_p(t) & \underline{S}_{p1}(t) \\ \underline{S}_{p1}(t) & \underline{S}_1(t) \end{pmatrix} \tag{2.67}$$

then the covariance equations become,

$$\underline{\Xi}(t) = \underline{\P}(t)-\underline{\P}(t)\begin{pmatrix} \underline{\phi}(t) \\ \underline{c} \end{pmatrix}\left([\underline{\phi}^T(t),\underline{c}^T]\underline{\P}(t)\begin{pmatrix} \underline{\phi}(t) \\ \underline{c} \end{pmatrix}\right)^{-1}[\underline{\phi}^T(t),\underline{c}^T]\underline{\P}(t) \tag{2.68}$$

with,

$$\underline{\P}(t+1) = \begin{bmatrix} \underline{I} & \underline{0} \\ \underline{0} & \underline{A} \end{bmatrix} \underline{\Xi}(t) \begin{bmatrix} \underline{I} & \underline{0} \\ \underline{0} & \underline{A} \end{bmatrix}^T + \begin{bmatrix} \underline{0} \\ \underline{d} \end{bmatrix} Q \begin{bmatrix} \underline{0} \\ \underline{d} \end{bmatrix}^T \tag{2.69}$$

Now, we know (Ref. 15) that the covariance matrix for the entire state, $\underline{\Xi}(t)$, converges to a constant value as $t\to\infty$, $\underline{\Xi}$. Using this result we can show that $\underline{P}_p$, the corresponding limit error covariance for $\underline{x}_p$, can be exactly determined via linear estimation procedures.

<u>Theorem</u>: The estimate of $\underline{x}_p$, $\hat{\underline{x}}_p$, and its error covariance, $\underline{P}_p(t)$, as given by equations 2.62 and 2.66 respectively, satisfy,

$$\lim_{t\to\infty} \hat{\underline{x}}_p(t) = \underline{x}_p \tag{2.70}$$

that is,

$$\lim_{t\to\infty} E\left\{[\underline{x}_p-\hat{\underline{x}}_p(t)][\underline{x}_p-\hat{\underline{x}}_p(t)]^T\right\} = \underline{0}$$

or,

$$\lim_{t\to\infty} \underline{P}_p(t) = \underline{0} \tag{2.71}$$

<u>Proof</u>: By multiplying both sides of 2.68 by $\begin{pmatrix} \underline{\phi}(t) \\ \underline{c} \end{pmatrix}$ we get,

$$\underline{\Xi}(t) \begin{pmatrix} \underline{\phi}(t) \\ \underline{c} \end{pmatrix} = \underline{0} \quad ; \quad \forall\ t \tag{2.72}$$

Since $\underline{\Xi}(t)$ becomes a constant in steady state, say $\underline{\Xi}$,

$$\underline{\Xi} = \begin{pmatrix} \underline{P}_p & \underline{P}_{p1} \\ \underline{P}_{p1} & \underline{P}_1 \end{pmatrix} \tag{2.73}$$

then 2.72 implies that,

$$\underline{P}_p \underline{\phi}(t) + \underline{P}_{p1} \underline{c} = \underline{0} \quad ; \quad \forall\, t \tag{2.74}$$

and,

$$\underline{P}_{p1}^T \, \underline{\phi}(t) + \underline{P}_1 \underline{c} = \underline{0} \quad ; \quad \forall\, t \tag{2.75}$$

Now we first want to show that the vectors $\underline{\phi}(t)$ for all time, t, are either linearly independent or span $2n_p+1$ space, where $2n_p+1$ is the dimension of $\underline{\phi}(t)$.

We will assume that the number of observations made on the load is at least 24 per day, i.e. hourly, and that this number exceeds $2n_p+1$. Actually the dimension of $\underline{\phi}$, $2n_p+1$, will never exceed 15 so that this assumption is valid. Since $\underline{\phi}$ is periodic of period 24 hours, this requirement may be viewed as an observality condition on $\underline{x}_p$ (requirement which is normally needed in these proofs).

A sufficient condition to show that the set of $\underline{\phi}$'s, $\{\underline{\phi}(t),\ t=1,\ 2,\ \ldots,\ 24\}$, (for hourly observations) span $2n_p+1$ space, is that the matrix $\underline{M}$ have an inverse, where,

$$\underline{M} = \sum_{t=1}^{24} \underline{\phi}(t)\, \underline{\phi}^T(t) \tag{2.76}$$

To show this consider an example with $2n_p+1=3$. Extensions to higher order cases are straightforward.

In this case,

$$\underline{\phi}(t) \quad = \quad [\ 1,\ \sin\omega_o t,\ \cos\omega_o t]^T \qquad (2.77)$$

where, $\omega_o \quad = \quad 2\pi/24$ (2.78)

Thus,

$$\underline{M} = \sum_{t=1}^{24} \underline{\phi}(t)\underline{\phi}^T(t) = \sum_{t=1}^{24} \begin{pmatrix} 1 & \sin\omega_o t & \cos\omega_o t \\ \sin\omega_o t & \sin^2\omega_o t & \sin\omega_o t\cdot\cos\omega_o t \\ \cos\omega_o t & \cos\omega_o t\cdot\sin\omega_o t & \cos^2\omega_o t \end{pmatrix} \qquad (2.79)$$

Due to the 24 hour periodicity of the terms in $\underline{M}$, it easily follows that,

$$\underline{M} \quad = \quad \begin{pmatrix} 24 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 12 \end{pmatrix} \qquad (2.80)$$

which is clearly invertible. Thus the set of $\underline{\phi}$'s span $2n_p+1$.

It easily follows that,

$$\underline{m} \quad = \quad \sum_{t=1}^{24} \underline{\phi}(t) \quad = \quad \begin{pmatrix} 24 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \qquad (2.81)$$

Consider now 2.74,

$$\underline{P}_p\ \underline{\phi}(t) \quad = \quad -\ \underline{P}_{p1}\ \underline{c} \qquad (2.82)$$

Multiplying both sides by $\underline{\phi}^T(t)$ and adding we get,

$$\underline{P}_p\ \underline{M} \quad = \quad -\ \underline{P}_{p1}\ \underline{c}\ \underline{m}^T \qquad (2.83)$$

Thus,

$$\underline{P}_p \quad = \quad -\ \underline{P}_{p1}\ \underline{c}\ \underline{m}^T\ \underline{M}^{-1} \qquad (2.84)$$

Now if we can prove that $\underline{P}_{p1}$ is the zero vector we will have reached our goal. To do this consider equations 2.68 and 2.69 in steady state. These yield,

$$\underline{P}_p = \underline{P}_p - \frac{[\underline{P}_p\underline{\phi}(t) + \underline{P}_{p1}\underline{A}^T\underline{c}][\underline{P}_p\underline{\phi}(t) + \underline{P}_{p1}\underline{A}^T\underline{c}]^T}{\Delta(t)} \quad (2.85)$$

and,

$$\underline{P}_{p1} = \underline{P}_{p1}\underline{A}^T - \frac{[\underline{P}_p\underline{\phi}(t) + \underline{P}_{p1}\underline{A}^T\underline{c}][\underline{\phi}^T(t)\underline{P}_{p1}\underline{A}^T + \underline{c}^T\underline{S}_1]}{\Delta(t)} \quad (2.86)$$

where,

$$\Delta(t) = [\underline{\phi}^T(t), \underline{c}^T] \, \underline{\P} \begin{pmatrix} \underline{\phi}(t) \\ \underline{c} \end{pmatrix} \quad (2.87)$$

Now if $\Delta(t)$ is greater than zero then it follows from 2.85 that,

$$[\underline{P}_p\underline{\phi}(t)+\underline{P}_{p1}\underline{A}^T\underline{c}][\underline{P}_p\underline{\phi}(t)+\underline{P}_{p1}\underline{A}^T\underline{c}]^T = \underline{0} \quad (2.88)$$

and hence that,

$$[\underline{P}_p\underline{\phi}(t) + \underline{P}_{p1}\underline{A}^T\underline{c}] = \underline{0} \quad (2.89)$$

Substituting 2.89 into 2.86 we get,

$$\underline{P}_{p1} = \underline{P}_{p1}\,\underline{A}^T \quad (2.90)$$

or,

$$\underline{P}_{p1}\,[\underline{I}-\underline{A}^T] = \underline{0} \quad (2.91)$$

But since we assume that all the eigenvalues of $\underline{A}$ are <u>less than</u> <u>one</u>, that is that part of the model corresponding to the residual component is asymptotically stable, then $[\underline{I}-\underline{A}^T]^{-1}$ exists (Ref. 26) thus,

$$\underline{P}_{p1} = \underline{0} \quad (2.92)$$

which together with 2.84 yields,

$$\underline{P}_p = \underline{0} \quad (2.93)$$

that is,

$$\lim_{t\to\infty} E \{[\underline{x}_p-\hat{\underline{x}}_p(t)][\underline{x}_p-\hat{\underline{x}}_p(t)]^T\} = \underline{0} \qquad (2.94)$$

which proves our result.

Since $\Delta(t)$ is positive semidefinite then the only other alternative is that it is equal to zero. But since we assume that $\underline{\Xi}$ is bounded, it then follows after some algebra and boundedness arguments that even if $\Delta=0$, equation 2.90 holds true, and thus our desired result.

Q.E.D.

A little calculation easily shows that part of the gain matrix corresponding to $\underline{x}_p$ in the steady state, that is $\underline{g}_p$, equals $\underline{0}$, as expected.

Thus we can see from this result that $\underline{x}_p$ can be obtained via linear filtering concepts.

II.9 Discussion of Proposed Model and Load Forecasting Technique:

At this point it should be noted that the primary contribution made thus far is the development of a load model as a discrete time uncertain (probabilistic) process, justifying its specific structure from previous experience, observations, and hypothesis to a certain degree.

The description of this model in state space form and the resulting estimation and prediction scheme is simply a convenient way of making use of the model for prediction purposes. If the original model is not valid then the use of state space concepts will not help. However, assuming that the model is sufficiently valid, then the use of probabilistic concepts as well as the estimation scheme has a number of advantages which we now list:

1) The load is modelled as a discrete time process instead of at only a few points.

2) This process is uncertain, a very realistic assumption, due to inherent load uncertainty as well modelling uncertainty. This makes the load values uncertain, but we have more confidence in the model.

3) The state at time t, $\underline{x}(t)$, a relatively small small set of numbers, contains all past information about the load and weather behaviour which may influence the future load behaviour. Since $\underline{x}(t)$ cannot be exactly determined, as it is a random variable, we recursively estimate its value by $\hat{\underline{x}}(t)$. This estimate contains an "optimum" amount of information about $\underline{x}(t)$ and hence about the future load.

4) Past weather and load data need not be stored. Only present observed load and weather data, as well as weather forecasts are needed.

5) The state estimate, $\hat{\underline{x}}(t)$, and hence the load forecasts, are continuously being updated on-line as new load and weather data is observed.

6) The estimation-prediction algorithms are computationally quite simple to implement as they consist of discrete equations requiring only multiplications and additions.

7) The state estimate, $\hat{\underline{x}}(t)$, is "optimum" at all times in the sense that the error covariance matrix,

$$\underline{P}(t) = E\{[\underline{x}(t)-\hat{\underline{x}}(t)][\underline{x}(t)-\hat{\underline{x}}(t)]^T\}$$

is minimum. This implies that the load forecast errors are also optimum in the same sense.

8) Each state estimate, $\hat{\underline{x}}(t)$, is accompanied by its corresponding error covariance matrix, $\underline{P}(t)$, and similarly for the load forecast. The evaluation of this state error covariance and hence of the load forecast error variance can also be done on-line via equation 2.45.

9) Load forecasts can be updated on-line as better revised weather forecasts become available, by simply changing the input sequence $\underline{u}(t+i)$, $i=1, 2, \ldots$, into the load forecasting algorithm. It must be emphasized that the structure and initial condition, $\hat{\underline{x}}(t)$, of the load forecasting algorithm are both optimum, independent of whether the forecast weather isiin error. That is, given the expected weather forecast, the procedure used for predicting load is the best under the circumstances. As shown earlier we can make use of expected weather forecast uncertainty to adjust the load forecast uncertainty. Finally, contingency studies can be carried out by trying different forecasted weather sequences.

10) Having the state estimate, $\hat{\underline{x}}(t)$, the load forecasting algorithm allows us to optimally estimate the load for any future time through $\hat{z}(t+i/t)$. As indicated in Fig.8 we could have a display of $\hat{z}(t+i/t)$ for a series of points,

i=1, 2, ..., up to say 24 hours in advance, or simply determine the forecast load for whatever time the operator requires.

11) The feedback corrective nature of the estimation algorithm is such that the estimate is relatively insensitive to modelling errors. It also has the ability to return the prediction errors to normal after some anomaly which had disturbed the scheme has died out.

12) The estimate of the load, $\hat{z}(t)$, could be used to test for abnormal load behaviour. That is, behaviour other than that expected from the hypothesized model. Such behaviour could be detected by comparing the term $[z(t)-\hat{z}(t)]^2$ with its expected variance, $\underline{c}^T\underline{P}(t)\,\underline{c}$, and verifying whether there is a noticeable deviation for more than one value of t (see Fig. 9).

The best approach during a major disturbance is to revert to the purely subjective operator action and discontinue use of the automatic forecasting mode. One of the reasons for this procedure is that when power is restored to a load block which was formerly cut-off, unusually high consumption levels occur which are abnormal in nature and therefore are not described by our model. The operator will carry out the load forecasting operation until normality is reestablished when the automatic mode will take over again. The value of the state used to initiate the automatic mode will have been updated in an open-loop manner during the disturbance by the observed weather variables, $\underline{u}$, and the last state estimate before the disturbance occured.

The capability of going from automatic to operator mode during abnormal conditions is a very important characteristic as it would be impractical and difficult to develop a load model which would be applicable to all conditions and replace the operator completely. In addition this capability allows the operator to add to the forecast the effect of predictable events which affect the load and are not included in the model, such as the shut-down of a factory due to a vacation period, the sale of power to other companies, major sports and political events and unusually severe weather conditions.

13) The technique proposed here is very much dependent on a good knowledge of $\underline{A}$, $\underline{b}$, $\underline{d}$, Q, and $\underline{x}_p$, the model parameters which we assume are approximately constant over a period of about one week, say. The determination of these parameters is treated in the following chapter.

The slow time variations of the model parameters are such that identification techniques (Ref. 16) will be able to detect them and adjust the estimation-prediction algorithms accordingly. Any errors in these parameters will hopefully be small enough so that the corrective feedback estimator will be insensitive to them.

Finally, something should be said about the possible disadvantages of this approach.

The primary objection would be to the relative complexity of the technique. This is however the price we have to pay for the treatment of weather and uncertain effects. Actually the filtering and prediction process is relatively simple to implement on a small computer. The more complex parameter identification technique, to be discussed later on, needs to be done much less frequently, e.g. once a week, so that we could carry out this operation at a time when the main computer is not too busy.

III.0 IDENTIFICATION OF LOAD MODEL PARAMETERS

III.1 Preliminary Discussion:

The aim of this chapter is to present a number of techniques for the identification of the parameters of the load model from real data. Although most of the results are well known some adaptation to our specific model are made.

At this level modelling arguments are normally no longer made, rather the model is presumed valid, and we simply desire to find some set of parameters such that model response and actual response are close in some sense. There is however some feedback from this step about the validity of the model back to the modelling step. Thus we may decide from the identification step that additional effects must be incorporated into our model.

In addition to evaluating a valid set of model parameters, identification techniques are used for adaptive reestimation of these parameters with time. In particular in our case, although the model attempts to describe in considerable detail load behaviour, it is valid only for a certain period of time to be determined by experimentation (approximately one week). This is of course due to the slow seasonal variations in power consumption which are not considered in this model.

III.2 System Identification:

A great deal of literature has been written on this subject (Ref. 16, 17), but the basic idea is to minimize some criterion of the difference between model and actual response by appropriately choosing the model parameters. One of the major restrictions in the field is that the system be time invariant at least over the range of data observations. This can however be modified to include so-called "slowly time varying" systems. The theory has furthermore specialized to linear systems, with some specific exceptions, in view of the inherent complexity as well as lack of understanding of non-linear systems. The treatment of model uncertainty has received considerable attention in the literature (Ref. 18, 19), having concentrated totally on probabilistic uncertainty.

In this study we concentrate primarily on an identification technique which can be denoted under specific conditions as Maximum Likelihood, or under looser conditions as Least Squares.

This approach is to the author's knowledge the most powerful one in the sense that given a certain amount of observed data it will provide us with the best parameter estimate (Ref. 19). In addition it can be shown under certain conditions that as more data is considered the parameter estimates are progressively better (Ref. 20). Furthermore we can consider with this technique the identification of models with input uncertainty, necessary in this study, as well as providing us with a quantitative approach for determining the order of the system.

III.3 Definition of Load Model Parameter Identification Problem:

The identification of the load model parameters can be defined in terms of the stochastic load model, which is here rewritten for ease of reference,

$$z(t) = y_p(t) + y(t) \tag{3.1}$$

where,

$$y_p(t) = \underline{\phi}^T(t)\, \underline{x}_p \tag{3.2}$$

$$\underline{\phi}^T(t) = [1, \sin \omega_o t, \sin 2\omega_o t, \ldots, \sin n_p\omega_o t, \cos \omega_o t, \ldots, \cos n_p\omega_o t] \tag{3.3}$$

$$\omega_o = 2\pi/24 \tag{3.4}$$

$$\underline{x}_p = \text{constant vector} \tag{3.5}$$

Also,

$$y(t) = \sum_{i=1}^{n} a_i y(t-i) + \sum_{j=0}^{m} b_j u(t-j) + c(t) \tag{3.6}$$

where c is a coloured process with a constant mean, and a stationary autocorrelation, $R_c(\tau)$, such that,

$$R_c(\tau) = 0 \quad ; \quad |\tau| \geq n \tag{3.7}$$

As shown in section II.3.2 this uncertain process can also be described in terms of a moving average with equal first and second statistics, that is,

$$c(t) = \sum_{i=1}^{n} d_i w(t-i) \tag{3.8}$$

where the d_i coefficients are constant and w is an uncorrelated or white process with variance Q.

The equivalency of these two processes implies that,

$$R_c(\tau) = \sum_{i=1}^{n-\tau} d_{i+\tau} Q \, d_i \quad ; \quad \forall \; \tau = 0, 1, \ldots, n-1 \quad (3.9)$$

The moving average noise model is useful in expressing the total model in state space form which is in turn useful in the estimation-prediction scheme discussed in the previous chapter. The autocorrelation function model for the system noise, c, is in turn useful in the definition of the identification of the entire system.

At this point we could define model identification as the estimation of the parameters $\underline{x}_p$, a_i, i=1, 2,..., n, b_j, j=1, 2,..., m, and $R_c(\tau)$, τ=0, 1,..., n-1.

From equation 3.9 we could get, given $R_c(\tau)$, τ =0, 1,..., n-1, a set of values of Q and d_i, although as can be seen not uniquely. These parameters together with $\underline{x}_p$, a_i and b_j, $\forall$ i and j, are then sufficient to apply the estimation-prediction algorithms discussed in the previous chapter. The non-uniqueness of Q and d_i is however unimportant <u>if what we wish to do with the model is estimate and predict</u>. This is easily seen in the case of Q and d_i. Referring back to estimation equations 2.40 and 2.45 of the previous chapter, we see that the only place where these parameters appear is in the term $\underline{d}Q\underline{d}^T$ so that for this purpose all we need to know are the terms $d_i Q^{1/2}$, i=1, 2,..., n, which are determinable from equation 3.9.

Following the same line of thought we can arrive at a definition of identification which is more appropriate for the purpose we have in mind.

It can be shown that if the Kalman filter is allowed to run a sufficient length of time, a steady state will be reached (Ref. 11). This means that both the filter gain, $\underline{g}$, and the error covariance, $\underline{P}$, will be constant. The filter in steady state is given by,

$$\underline{\hat{x}}(t+1) = \underline{A}\ \underline{\hat{x}}(t) + \underline{b}\ u(t) + \underline{g}\ e(t+1) \qquad (3.10)$$

$$e(t+1) = z(t+1) - \hat{z}(t+1) \qquad (3.11)$$

$$\hat{z}(t+1) = \underline{\phi}^T(t+1)\ \underline{x}_p + \underline{c}^T[\underline{A}\ \underline{\hat{x}}(t)+\underline{b}u(t)] + b_o u(t+1) \qquad (3.12)$$

The terms e(t) are called the filter residuals and can be shown to be a white process (Ref. 21) with,

$$E\{e(t)\} = 0 \qquad (3.13)$$

and,

$$E\{e(t)e(\tau)\} = \begin{cases} R_e & ;\ t=\tau \\ 0 & ;\ \forall\ t\neq\tau \end{cases} \qquad (3.14)$$

where,

$$R_e = \underline{c}^T\ [\ \underline{A}\ \underline{P}\ \underline{A}^T + \underline{d}\ Q\ \underline{d}^T\ ]\ \underline{c} \qquad (3.15)$$

The prediction and predicted error variance equations are given by,

$$\underline{\hat{x}}(t+n+1/t) = \underline{A}\ \underline{\hat{x}}(t+n/t) + \underline{b}\ u(t+n) \qquad (3.16)$$

$$\hat{z}(t+n/t) = \underline{\phi}^T(t+n)\ \underline{x}_p + \underline{c}^T\underline{\hat{x}}(t+n/t) + b_o u(t+n) \qquad (3.17)$$

with initial condition $\underline{\hat{x}}(t/t)=\underline{\hat{x}}(t)$ and,

$$\Sigma_z(t+n/t) = E\{\ [z(t+n)-\hat{z}(t+n/t)]^2\} \qquad (3.18)$$

$$= \underline{c}^T\ \underline{\Sigma}(t+n/t)\ \underline{c} \qquad (3.19)$$

with,

$$\underline{\Sigma}(t+n+1/t) = \underline{A}\ \underline{\Sigma}(t+n/t)\ \underline{A}^T + \underline{d}\ Q\ \underline{d}^T \qquad (3.20)$$

with initial condition $\underline{\Sigma}(t/t)=\underline{P}$, where $\underline{P}$ is the steady state error covariance given by the steady state solution of equation 2.41.

We will show that it is possible to identify the parameters of the steady filter, that is,

$$\begin{aligned} &\underline{x}_p \\ &a_i,\ i= 1, 2, ,..., n \\ &b_j,\ j= 0, 1, ..., m \\ &g_i,\ i= 1, 2, ..., n \end{aligned} \qquad (3.21)$$

and the variance of the residuals R_e. This set of parameters is in turn sufficient to define the desired filter and load predictor as seen from equations 3.10, 3.11 and 3.12. In order now to determine the prediction error variance we also need to know the term $\underline{d}Q\underline{d}^T$, however it will be shown that knowledge of the steady state gain, $\underline{g}$, and the residual variance R_e is also sufficient to determine $R_c(\tau), \tau = 0, 1, ...,$ n-1, which as shown before suffices to determine the desired $\underline{d}Q\underline{d}^T$.

The definition of identification for our purposes is thus to determine the parameters in 3.21 plus R_e. We can also include in our definition of identification the system dimensions n, m and n_p.

III.4 <u>Least Squares Identification</u>:

As discussed earlier the fundamental approach behind the identification of the parameters of a system is to minimize some criterion of the difference between the observed output and the model output.

Since our model is probabilistic we could consider $\hat{z}(t+1)$, the estimate of $z(t+1)$ given $z(i)$, $i=t, t-1, \ldots$, as a possible model output. A reasonable criterion which would describe the difference between the actual and model behaviour is then,

$$J = 1/N \sum_{t=1}^{N} [z(t)-\hat{z}(t)]^2 \qquad (3.22)$$

where N is some large number over which period the system parameters are assumed stationary.

We can recognize 3.22 as the averaged sum of the squares of the filter residuals. Assuming ergodicity it follows that if N is large then,

$$J \simeq R_e \qquad (3.23)$$

where R_e is the variance of the residuals. This will be true if $\hat{z}(t/t-1)$ is generated by the <u>true</u> model parameters, but will yield a larger value of J if the model parameters deviate from their true value. Under general assumptions we can't give a rigorous proof that minimizing J with respect to the system parameters as defined in the previous section will yield the true values of the parameters. This is however a very reasonable engineering criterion which under more rigorous conditions, as will be shown in the next sections, yields a consistent estimate of the system parameters, that is as $N\to\infty$, the parameter estimates converge to their true values.

We also know that the residuals of the Kalman filter, $e(t)=z(t)-\hat{z}(t/t-1)$, are a white process. This gives us an additional test in deciding whether the model is a valid one, that is one can calculate the autocorrelation function of e(t) and check for whiteness (uncorrelation).

III.4.1 Statement of Least Squares Identification Problem:

Minimize the function,

$$J = 1/N \sum_{t=1}^{N} e^2(t) \tag{3.24}$$

where,

$$e(t) = z(t) - \hat{z}(t) \tag{3.25}$$

subject to,

$$\hat{z}(t) = \underline{\phi}^T(t)\underline{x}_p+\underline{c}^T[\underline{A}\ \hat{\underline{x}}(t-1)+\underline{b}u(t-1)]+b_o u(t) \tag{3.26}$$

where,

$$\hat{\underline{x}}(t) = \underline{A}\ \hat{\underline{x}}(t-1) + \underline{b}\ u(t-1) + \underline{g}e(t) \tag{3.27}$$

with respect to the parameter vector $\underline{\theta}$,

$$\underline{\theta} = \begin{bmatrix} \underline{x}_p \\ \underline{a} \\ \underline{b} \\ \underline{g} \end{bmatrix} \tag{3.28}$$

where $\underline{x}_p$ is a $2n_p+1$ vector as defined in equation 2.6 and,

$$\underline{a} = [a_1, a_2, \ldots, a_n]^T \tag{3.29}$$

$$\underline{b} = [b_0, b_1, \ldots, b_n]^T \tag{3.30}$$

$$\underline{g} = [g_1, g_2, \ldots, g_n]^T \tag{3.31}$$

The minimum of J, J*, is such that,

$$J^* \simeq R_e = E\{e^2(t)\} \tag{3.32}$$

Furthermore if the system dimensions n and n_p are uncertain these must be included in the optimization. More on this will be discussed in section III.9.

This is an optimization problem constrained by equations 3.25 through 3.27 with no general closed form solution. The most obvious approach to its solution is an iterative one where one proceeds as follows,

i) Assume some value of $\underline{\theta}$, $\underline{\theta}^*$.

ii) Run the filter equations 3.26 and 3.27 based on the observed data z(t) to yield $\hat{z}(t)$ and hence e(t).

iii) Evaluate the new J and $\partial J/\partial\underline{\theta}$. Calculate a new $\underline{\theta}$ along the direction of decreasing J.

III.4.2 Solution of Least Squares Identification Problem - Autoregressive-Moving-Average (ARMA) Model of Filter:

Before discussing the solution of the problem described in III.4.1, we wish to express the estimation equations 3.25 through 3.27 in autoregressive-moving-average form (ARMA). This is a more convenient form for identification purposes.

We know that the ARMA model,

$$y(t) = \sum_{i=1}^{n} a_i y(t-i) + \sum_{j=0}^{n} b_j u(t-j) + \sum_{j=1}^{n} d_j w(t-j) \tag{3.33}$$

can be put in state space form as,

$$\underline{x}(t+1) = \underline{A}\ \underline{x}(t) + \underline{b}\ u(t) + \underline{d}\ w(t) \tag{3.34}$$

$$y(t+1) = \underline{c}^T \underline{x}(t+1) + b_0 u(t+1) \tag{3.35}$$

and vice versa, where $\underline{A}$, $\underline{b}$, $\underline{d}$ and $\underline{c}$ are as defined in equations 2.21 through 2.24.

Now consider the steady state filter,

$$\hat{\underline{x}}(t+1) = \underline{A}\ \hat{\underline{x}}(t) + \underline{b}\ u(t) + \underline{g}\ e(t+1) \tag{3.36}$$

$$\hat{y}(t+1) \equiv \underline{c}^T[\underline{A}\ \hat{\underline{x}}(t) + \underline{b}\ u(t)] + b_o u(t+1) \tag{3.37}$$

where,

$$e(t+1) = z(t+1) - \hat{z}(t+1) \tag{3.38}$$

$$= \underline{\phi}^T(t+1)\ \underline{x}_p + y(t+1) - \underline{\phi}^T(t+1)\ \underline{x}_p - \hat{y}(t+1) \tag{3.39}$$

$$= y(t+1) - \hat{y}(t+1) \tag{3.40}$$

Thus if we can express 3.36 and 3.37 in the same form as 3.34 and 3.35, we can by the equivalence of the equations obtain an ARMA model for $\hat{y}(t+1)$. Defining,

$$\underline{s}(t) \equiv \underline{A}\ \hat{\underline{x}}(t-1) + \underline{b}\ u(t-1) \tag{3.41}$$

then,

$$\underline{s}(t+1) = \underline{A}\ \hat{\underline{x}}(t) + \underline{b}\ u(t) \tag{3.42}$$

$$= \underline{A}[\underline{A}\ \hat{\underline{x}}(t-1)+\underline{b}\ u(t-1)+\underline{g}e(t)]+\underline{b}u(t) \tag{3.43}$$

$$= \underline{A}[\underline{s}(t) + \underline{g}\ e(t)] + \underline{b}\ u(t) \tag{3.44}$$

$$= \underline{A}\ \underline{s}(t) + \underline{A}\ \underline{g}\ e(t) + \underline{b}\ u(t) \tag{3.45}$$

while from 3.37,

$$\hat{y}(t+1) = \underline{c}^T\underline{s}(t+1)+b_o u(t+1) \tag{3.46}$$

Defining,

$$\tilde{\underline{g}} \equiv \underline{A}\ \underline{g} \tag{3.47}$$

$$\equiv \begin{pmatrix} \tilde{g}_1 \\ \tilde{g}_2 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{g}_n \end{pmatrix} \tag{3.48}$$

and since,

$$\underline{A} = \begin{bmatrix} a_1 & 1 & 0 & . & . & . & 0 \\ a_2 & 0 & 1 & 0 & . & . & 0 \\ . & . & & . & & & . \\ . & . & & & . & & . \\ . & . & & & & . & \\ . & . & & & & & 1 \\ . & . & & & & & \\ a_n & 0 & . & . & . & . & 0 \end{bmatrix} \tag{3.49}$$

it follows that,

$$\underline{\tilde{g}} = \begin{bmatrix} a_1 g_1 + g_2 \\ a_2 g_1 + g_3 \\ . \\ . \\ . \\ a_{n-1} g_1 + g_n \\ a_n g_1 \end{bmatrix} \tag{3.50}$$

We then get the ARMA model for y(t),

$$\hat{y}(t) = \sum_{i=1}^{n} a_i \hat{y}(t-i) + \sum_{j=0}^{n} b_j u(t-j) + \sum_{j=1}^{n} \tilde{g}_j e(t-j) \tag{3.51}$$

with,

$$\hat{z}(t) = \underline{\phi}^T(t)\, \underline{x}_p + \hat{y}(t) \tag{3.52}$$

Equations 3.51 and 3.52 are thus an alternative way of describing the filter equations previously written in state space form.

A second useful way of writing 3.51 is as follows: Subtracting 3.51 from 3.33 and using 3.40 we get,

$$e(t) = \sum_{j=1}^{n} a_j e(t-j) - \sum_{j=1}^{n} \tilde{g}_j e(t-j) + \sum_{i=1}^{n} d_i w(t-i) \tag{3.53}$$

$$= \sum_{j=1}^{n} [a_j - \tilde{g}_j] e(t-j) + \sum_{i=1}^{n} d_i w(t-i) \tag{3.54}$$

But since,

$$\sum_{i=1}^{n} d_i w(t-i) = y(t) - \sum_{i=1}^{n} a_i y(t-i) - \sum_{j=0}^{n} b_j u(t-j) \qquad (3.55)$$

then,

$$e(t) = \sum_{j=1}^{n} [a_j - \tilde{g}_j] e(t-j) + [y(t) - \sum_{i=1}^{n} a_i y(t-i) - \sum_{j=0}^{n} b_j u(t-j)] \qquad (3.56)$$

Now it can easily be shown from the definition of $\underline{g}$ in equation 2.39 that g(1)=1, so that together with 3.50 it follows that,

$$e(t) = \sum_{j=1}^{n-1} g_{j+1} e(t-j) + [y(t) - \sum_{i=1}^{n} a_i y(t-i) - \sum_{j=0}^{n} b_j u(t-j) \qquad (3.57)$$

Since y(t) is not observed in our case but rather,

$$z(t) = \underline{\phi}^T(t)\ \underline{x}_p + y(t) \qquad (3.58)$$

then substituting for y in 3.57 we get,

$$e(t) = \sum_{j=1}^{n-1} g_{j+1} e(t-j) + [z(t) - \sum_{i=1}^{n} a_i z(t-i) - \sum_{j=0}^{n} b_j u(t-j) + \sum_{i=1}^{n} \{a_i \underline{\phi}^T(t-i)\ \underline{x}_p\} - \underline{\phi}^T(t)\ \underline{x}_p] \qquad (3.59)$$

This equation can be used to recursively update the residuals of the filter given the observed data, z and u, and the system parameters, $\underline{\theta}$.

We can now redefine the identification problem as the minimization of,

$$J = 1/N \sum_{t=1}^{N} e^2(t) \qquad (3.60)$$

with respect to the model parameters, $\underline{\theta}$, and subject to the relations 3.59.

The solution to this problem <u>without</u> the term y_p, that is with $\underline{x}_p = \underline{0}$, has been obtained by Aström (Ref. 18) and Kashyap (Ref. 19). Furthermore they show that under Gaussian assumptions for the system uncertainty there exists a unique minimum for the above optimization problem which converges to its true value as N, the amount of data, goes to infinity.

In our case, we can argue that a minimum clearly exists because letting $\underline{x}_p$ equal its true value results in the Aström problem which has a minimum. In addition letting $\underline{x}_p$ diverge clearly would result from 3.59 in a diverging e(t) and hence J.

What we can't prove (or haven't been able to) is that there exists a unique minimum, $\underline{\theta}^*$, which would approach the true value of $\underline{\theta}$ as $N \to \infty$. In a later section we will describe a method of separating the problems of estimating $\underline{x}_p$ and of estimating $\underline{a}$, $\underline{b}$, $\underline{g}$ and R_e which would remove this doubt. At this stage we can however present a set of necessary conditions which will yield the solution of the optimization problem defined above. Since the criterion chosen is a reasonable one from a practical point of view, finding its minimum should yield a valid model.

<u>Necessary Conditions for Parameter Estimation</u> -

For ease of notation we now define,

$$\underline{g} \equiv [g_2, g_3, \ldots, g_n]^T \qquad (3.61)$$

$$\underline{e}(t) \equiv [e(t-1),\ e(t-2),\ \ldots,\ e(t-n+1)]^T \qquad (3.62)$$

$$\underline{a} \equiv [a_1,\ a_2,\ \ldots,\ a_n]^T \qquad (3.63)$$

$$\underline{z}(t) \equiv [z(t-1),\ z(t-2),\ \ldots,\ z(t-n)]^T \qquad (3.64)$$

$$\underline{b} \equiv [b_0,\ b_1,\ b_2,\ \ldots\ b_n]^T \qquad (3.65)$$

$$\underline{u}(t) \equiv [u(t),\ u(t-1),\ \ldots,\ u(t-n)]^T \qquad (3.66)$$

$$\underline{\psi}(t) \equiv [\underline{\phi}(t-1),\ \underline{\phi}(t-2),\ \ldots,\ \underline{\phi}(t-n)]^T \qquad (3.67)$$

so that 3.59 becomes,

$$e(t) = \underline{g}^T\underline{e}(t)+[z(t)-\underline{a}^T\underline{z}(t)-\underline{b}^T\underline{u}(t)-\underline{x}_p^T\underline{\phi}(t)+\underline{a}^T\underline{\psi}(t)\underline{x}_p] \qquad (3.68)$$

$$\equiv \underline{g}^T\underline{e}(t) + f(\underline{\theta},\ t) \qquad (3.69)$$

where f depends only on $\underline{\theta}$ and observed data, and t=n+1, n+2, ..., N.

Since we are dealing with a constrained optimization problem we introduce a set of Lagrange multipliers, $\lambda(t)$, t=n+1, n+2, ..., N, and consider the minimization of the function H, where,

$$H = 1/N \sum_{t=1}^{N} e^2(t) + 1/N \sum_{t=n+1}^{N} \lambda(t)[e(t)-\underline{g}^T\underline{e}(t)-f(\underline{\theta},\ t)] \qquad (3.70)$$

The necessary conditions are,

$$\text{i)}\ \partial H/\partial\lambda(t) = 0 \quad ; \quad t=n+1,\ n+2,\ \ldots,\ N \qquad (3.71)$$

yielding,

$$e(t) = \underline{g}^T\underline{e}(t) + f(\underline{\theta},\ t) \quad ; \quad t=n+1,\ n+2,\ \ldots,\ N \qquad (3.72)$$

$$\text{ii)}\ \partial H/\partial e(t) = 0 \quad ; \quad t=n+1,\ n+2,\ \ldots,\ N \qquad (3.73)$$

yielding,

$$\lambda(t) = \sum_{i=1}^{n-1} g_{i+1}\lambda(t+i) - 2e(t); \quad t=n+1,\ n+2,\ \ldots,\ N \qquad (3.74)$$

with initial conditions,

$$\lambda(N+i) = 0 \quad ; \quad i=1,\ 2,\ \ldots,\ n \qquad (3.75)$$

We can now use 3.72 and 3.74 to calculate for any value of $\underline{\theta}$ the gradient of J with respect to $\underline{\theta}$. This gradient together with the value of J is sufficient to make use of iterative solutions to the minimization problem here discussed. The solution is of the form,

$$\underline{\theta}_{k+1} = \underline{\theta}_k + \underline{H}_k \, \partial J/\partial \underline{\theta}_k \qquad (3.76)$$

where $\underline{H}_k$ depends on the type of iterative procedure. In this study we make use of the Fletcher-Powell technique (Ref. 22) which requires the value of J and $\partial J/\partial \underline{\theta}$ for its implementation.

Now, the gradient of J subject to the constraint of 3.68 is equal to the unconstrained gradient of H, thus,

$$\partial J/\partial \underline{\theta} = \partial H/\partial \underline{\theta} \qquad (3.77)$$

yielding,

$$\frac{\partial J}{\partial \underline{g}} = -\frac{1}{N} \sum_{t=n+1}^{N} \lambda(t)\, \underline{e}(t) \qquad (3.78)$$

$$\frac{\partial J}{\partial \underline{a}} = \frac{1}{N} \sum_{t=n+1}^{N} \lambda(t)[\underline{z}(t)-\underline{\psi}(t)\underline{x}_p] \qquad (3.79)$$

$$\frac{\partial J}{\partial \underline{b}} = \frac{1}{N} \sum_{t=n+1}^{N} \lambda(t)\underline{u}(t) \qquad (3.80)$$

$$\frac{\partial J}{\partial \underline{x}_p} = \frac{1}{N} \sum_{t=n+1}^{N} \lambda(t)[\underline{\phi}(t)-\underline{\psi}^T(t)\underline{a}] \qquad (3.81)$$

Thus from 3.72 and 3.74 we can get values of e(t) and λ(t) for all t, running first 3.72 forward followed by 3.74 backwards in time. This then allows us to determine for any $\underline{\theta}$ the gradient of J as given by 3.78 through 3.81, which together with the value of J can be used in the Fletcher-Powell algorithms for the iterative minimization of J.

It should be noted that e(t) need not be generated via equation 3.72; it may be computationally more convenient to use the state space model equations 3.36 through 3.38. The reason we have used the ARMA model here is for ease of calculation of the gradient of J.

Similarly the equation for λ(t) can be put in state space form as follows,

$$\underline{p}(t-1) = \begin{bmatrix} g_2 & 1 & 0 & \dots & 0 \\ & & & & \cdot \\ g_3 & 0 & 1 & 0 & \dots \\ \cdot & & & \cdot & \cdot \\ \cdot & & & \cdot & 0 \\ \cdot & & & & \cdot \\ g_{n-1} & & & & 1 \\ g_n & 0 & \dots & \dots & 0 \end{bmatrix} \underline{p}(t) - \begin{bmatrix} g_2 \\ g_3 \\ \cdot \\ \cdot \\ \cdot \\ g_n \end{bmatrix} e(t) \qquad (3.82)$$

$$\lambda(t) = p_1(t) - 2e(t) \qquad (3.83)$$

with initial condition,

$$\underline{p}(N) = \underline{0} \qquad (3.84)$$

III.5 Evaluation of $\underline{d}Q\underline{d}^T$ from $\underline{g}$ and R_e:

In the past section we discuss the identification of the system parameters $\underline{a}$, $\underline{b}$, $\underline{g}$ and R_e. We know these to be sufficient to use the estimation and prediction algorithms discussed earlier. R_e is the variance of the residuals or the one-step prediction error variance, however we wish to determine the multi-step prediction error variance. To do this we need to know the additional term $\underline{d}Q\underline{d}^T$ or equivalently $d_i Q^{1/2}$, i=1, 2, ..., n.

We know that,

$$y(t) = \sum_{i=1}^{n} a_i y(t-i) + \sum_{j=0}^{n} b_j u(t-j) + \sum_{j=1}^{n} d_j w(t-j) \qquad (3.85)$$

and from 3.51,

$$\hat{y}(t) = \sum_{i=1}^{n} a_i \hat{y}(t-i) + \sum_{j=0}^{n} b_j u(t-j) + \sum_{j=1}^{n} \tilde{g}_j e(t-j) \qquad (3.86)$$

Thus,

$$y(t) = \hat{y}(t) + y(t) - \hat{y}(t) \qquad (3.87)$$

$$= \hat{y}(t) + e(t) \qquad (3.88)$$

Substituting 3.86 into 3.88 yields,

$$y(t) = \sum_{i=1}^{n} a_i \hat{y}(t-i) + \sum_{j=0}^{n} b_j u(t-j) + \sum_{j=1}^{n} \tilde{g}_j e(t-j) + e(t) \qquad (3.89)$$

or,

$$y(t) = \sum_{i=1}^{n} a_i [y(t-i)-e(t-i)] + \sum_{j=0}^{n} b_j u(t-j) + \sum_{j=1}^{n} \tilde{g}_j e(t-j) + e(t) \qquad (3.90)$$

$$y(t) = \sum_{i=1}^{n} a_i y(t-i) + \sum_{j=0}^{n} b_j u(t-j) + \sum_{j=1}^{n-1} g_{j+1} e(t-j) + e(t) \qquad (3.91)$$

Now 3.91 is equivalent to 3.85 and we must then have since $g_1=1$,

$$\sum_{j=1}^{n} d_j w(t-j) \approx e(t) + \sum_{j=1}^{n-1} g_{j+1} e(t-j) = \sum_{j=0}^{n-1} g_{j+1} e(t-j) \qquad (3.92)$$

at least within first and second order statistics. Since both w and e are white processes we must have, equating their variances,

$$\sum_{j=1}^{n-\tau} d_j d_{j+\tau} Q = \sum_{j=1}^{n-\tau} g_j g_{j+\tau} R_e \quad ; \quad \tau = 0,\ 1,\ 2,\ \ldots,\ n-1 \tag{3.93}$$

The n equations above can readily be solved for $d_i Q^{1/2}$, $i=1, 2, \ldots, n$, in terms of the known g_i, $i=1, \ldots, n$, and R_e.

III.6 Identification by Component Separation:

In this section we consider the possibility of solving the identification problem in two separate steps, first the identification of the y model parameters, the residual load, and second that of $\underline{x}_p$ by the linear estimation procedure discussed in section II.3.4. Two approaches will be discussed:

i) Data prefiltering approach.

ii) Iterative approach.

III.6.1 Data Prefiltering Approach:

The idea here is to prefilter the observed data, z and u, so that the effect of the periodic component becomes negligible. The resulting filtered data is shown to satisfy a system whose uncertain parameters are the same as those of the residual load y, $\underline{\alpha}$. These parameters are identified in the same manner as described in III.4 with the exception that the non-linear search is now restricted to a much lower number of parameters. The search for $\underline{x}_p$ can then be made through the linear estimation technique of section II.8 which is a much less complex and time consuming operation.

Furthermore each identification problem is such that the corresponding search schemes converge to the true value as discussed earlier.

For ease of notation define the delay operator 'd' as follows,

$$d\,x(t) \equiv x(t-1) \qquad (3.94)$$

$$d^2x(t) \equiv x(t-2) \qquad (3.95)$$

and so on.

Using this notation we define,

$$A(d) = \sum_{i=1}^{n} a_i d^i \qquad (3.96)$$

$$B(d) = \sum_{i=0}^{n} b_i d^i \qquad (3.97)$$

$$D(d) = \sum_{i=0}^{n} d_i d^i \qquad (3.98)$$

$$\tilde{G}(d) = \sum_{i=1}^{n} \tilde{g}_i d^i \qquad (3.99)$$

With this notation, the ARMA load model becomes,

$$z(t) = y(t) + y_p(t) \qquad (3.100)$$

$$y(t) = A(d)y(t) + B(d)u(t) + D(d)w(t) \qquad (3.101)$$

Data Prefiltering for Separation of y_p

Since y_p is a periodic described by a finwte sum of sinusoids with maximum frequency $n_p\omega_o$, $\omega_o=2\pi/24$, it follows after some algebra that,

$$B_p(d)\, y_p(t) = 0 \quad ; \quad \forall\, t > 2n_p+1 \qquad (3.102)$$

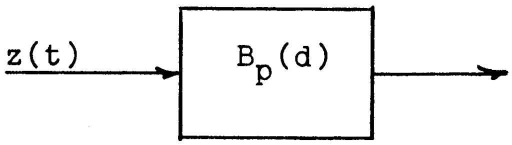
where,

$$B_p(d) = (1-d) \prod_{k=1}^{n_p} (1-2d \cos k\omega_o+d^2) \qquad (3.103)$$

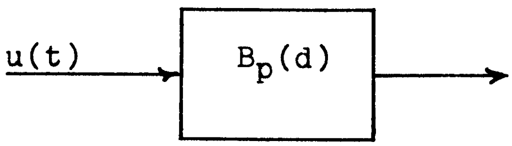
This then implies from 3.100 that,

$$B_p(d)\, z(t) = B_p(d)\, y(t) \quad ; \quad \forall\, t > 2n_p+1 \qquad (3.104)$$

Data Prefilter



$$z_F(t) = B_p(d)\ z(t) \tag{3.105}$$



$$u_F(t) = B_p(d)\ u(t) \tag{3.106}$$

Using the above definitions of z_F and u_F, the prefiltered data, then by multiplying both sides of 3.101 by $B_p(d)$, it follows that,

$$z_F(t) = A(d)\ z_F(t) + B(d)\ u_F(t) + B_p(d)\ D(d)\ w(t) \tag{3.107}$$

Thus the filtered data satisfies a similar equation as y, u and w. The only difference is in the term operating on w(t), the system noise. Part of that operator is $B_p(d)$ which is known. It then follows that the unknown parameters in 3.107 are the same as in the equation relating y, u and w.

Consider now the Kalman filter generating the best linear estimate of y(t) given all past values of z. In operator form this is,

$$\hat{y}(t) = A(d)\ \hat{y}(t) + B(d)\ u(t) + \tilde{G}(d)\ e(t) \tag{3.108}$$

where,

$$e(t) = z(t) - \hat{z}(t) = y(t) - \hat{y}(t) \tag{3.109}$$

Now since,

$$\hat{z}(t) = y_p(t) + \hat{y}(t) \tag{3.110}$$

then,

$$B_p(d)\ \hat{z}(t) = B_p(d)\ \hat{y}(t) \tag{3.111}$$

But since multiplying by B_p is a linear operation then,

$$\hat{z}_F(t) \equiv B_p(d)\ \hat{z}(t) \tag{3.112}$$

is the best linear estimate of z_F, given all past z.

Multiplying both sides of 3.108 by B_p then yields,

$$\hat{z}_F(t) = A(d)\ \hat{z}_F(t) + B(d)\ u_F(t) + \tilde{G}(d)\ e_F(t) \tag{3.113}$$

where,

$$e_F(t) = z_F(t) - \hat{z}_F(t) \tag{3.114}$$

$$= B_p(d)\ e(t) \tag{3.115}$$

We can now proceed as in III.4 and obtain an iterative equation for e(t) as,

$$B_p(d)\ e(t) = e_F(t) \tag{3.116}$$

$$e_F(t)=[A(d)-\tilde{G}(d)]e_F(t)+[z_F(t)-A(d)z_F(t)-B(d)u_F(t)] \tag{3.117}$$

or,

$$e_F(t) = G(d)e_F(t) + z_F(t) - A(d)z_F(t) - B(d)u_F(t) \tag{3.118}$$

where,

$$G(d) = \sum_{i=1}^{n-1} g_{i+1}d^i \tag{3.119}$$

The identification of the unknown parameters $\underline{\alpha}$ = [$\underline{g}$, $\underline{a}$, $\underline{b}$, R_e] is now defined as the minimization of,

$$J = 1/N \sum_{t=1}^{N} e^2(t) \tag{3.120}$$

as before, with respect to $\underline{\alpha}$ which <u>does</u> <u>not</u> include $\underline{x}_p$, and subject to 3.116 and 3.118. Now 3.117 is an assymptotically stable system so it does not matter what the initial conditions are since their effect eventually decays to zero. Equation 3.116 has however all its poles on the unit circle so that in order to prevent e(t) from having a sinusoidal behaviour (e(t) should be a white process) we will force the initial state in 3.116 to be zero.

The solution of this problem is done as before via a Fletcher-Powell type of search. It must be noticed that the search for $\underline{x}_p$ is bypassed here, however the evaluation of e(t) and, as will be seen, of the Lagrange multipliers, $\lambda(t)$, is more complicated. More on the merits and drawbacks of this approach will be discussed later on.

Again we define an equivalent unconstrained minimization problem, that is the minimization of,

$$H = 1/N \sum_{t=1}^{N} e^2(t) + 1/N \sum_{t=n}^{N} \lambda(t)[e_F(t)-G(d)e_F(t)-F(\underline{\theta},t)] + 1/N \sum_{t=n_s}^{N} \gamma(t)[e_F(t)-B_p(d)e(t)] \quad (3.121)$$

where,

$$F(\underline{\theta}, t) = z_F(t) - A(d)z_F(t) - B(d)u_F(t) \quad (3.122)$$

and where (t) and (t) are Lagrange Multipliers. Also,

$$n_s = n + 2n_p + 1 \quad (3.123)$$

As before we must have,

$$\text{i) } \partial H/\partial\lambda(t) = 0 \quad (3.124)$$

yielding,

$$e_F(t) = G(d)e_F(t) + F(\underline{\theta}, t) \quad (3.125)$$

$$\text{ii) } \partial H/\partial\gamma(t) = 0 \quad (3.126)$$

yielding,

$$e_F(t) = B_p(d)\, e(t) \quad (3.127)$$

with initial condition

$$e(i) = 0 \;;\; i=1, 2, \ldots, 2n_p+1 \quad (3.128)$$

iii) $\partial H/\partial e_F(t) \;=\; 0$ (3.129)

yielding,

$$\lambda(t) \;=\; G(d^{-1})\lambda(t) - \gamma(t) \qquad (3.130)$$

with terminal conditions,

$$\lambda(N+i) \;=\; -\,\gamma(N+i) \quad ; \quad i=1, 2, \ldots, n_s \qquad (3.131)$$

where d^{-1} is a forward operator such that,

$$d^{-1}x(t) \;=\; x(t+1) \qquad (3.132)$$

iv) $\partial H/\partial e(t) \;=\; 0$ (3.133)

yileding,

$$B_p(d^{-1})\gamma(t) \;=\; 2e(t) \qquad (3.134)$$

with terminal conditions,

$$\gamma(N+i) \;=\; 0 \quad ; \quad i=1, 2, \ldots, n_s \qquad (3.135)$$

The gradient of J with respect to the unknown parameters $\underline{g}$, $\underline{a}$ and $\underline{b}$ becomes,

$$\partial J/\partial \underline{a} = \partial H/\partial \underline{a} \;=\; 1/N \sum_{t=n_s}^{N} \lambda(t)\, \underline{z}_F(t) \qquad (3.136)$$

$$\partial J/\partial \underline{b} = \partial H/\partial \underline{b} \;=\; 1/N \sum_{t=n_s}^{N} \lambda(t)\, \underline{u}_F(t) \qquad (3.137)$$

$$\partial J/\partial \underline{g} = \partial H/\partial \underline{g} \;=\; -1/N \sum_{t=n_s}^{N} \lambda(t)\, \underline{e}_F(t) \qquad (3.138)$$

<u>Discussion on Prefiltering Approach</u>

This approach has the advantage that the set of parameters $\underline{x}_p$ need not be identified as part of the non-linear estimation discussed in III.4. This is of considerable significance if the computer limitations are important as the non-linear scheme is quite demanding. In our model the number of parameters in y_p could be as large as 15, while those in y as large as 9, so that the computing savings may be considerable.

The major drawback is that prefiltering of z and u is such that essentially it requires some high order differencing operating, the equivalent of differentiating in continuous time, on observed data. This results in extremely large values of z_F and u_F, in addition to the magnification of any observation errors. This makes the resulting data untrustworthy.

To get around this problem, instead of defining the prefilter by essentially an all zero system, we can modify it so as to cancel out the differencing effect by adding some poles to the system, that is,

$$A_p(d)\ z_F(t) \ = \ B_p(d)\ z(t) \qquad (3.139)$$

Here $B_p(d)$ is the same as defined before but $A_p(d)$ is such that the resulting system is asymptotically stable as well as being of at least the same order as B_p. The first requirement guarantees that the effect of initial conditions decays to zero and the latter eliminates the differencing effect.

Suppose the input to this filter is y_p and let the output be y_{pF}, then after $2n_p+1$ units of time, $B_p(d)y_p(t)=0$, so that,

$$A_p(d)y_{pF}(t) \ = \ 0 \quad ; \quad t \geqslant 2n_p+1 \qquad (3.140)$$

but since A_p is asymptotically stable then in the limit $y_{pF}(t)=0$.

Before we can use filtered data, z_F and u_F, and be confident that the effect of y_p has died down sufficiently, we then have to wait a few time constants, thus reducing the identification data record. On the other hand making the prefilter too stable is such that the filtered data becomes too small, causing numerical ill-conditioning.

A compromise thus must be made in designing the prefilter poles, primarily by experimentation.

Now we briefly describe how we could use this type of prefilter in the identification approach discussed in this section.

Filtering z we get,

$$A_p(d)\ z_F(t) = B_p(d)\ z(t) \tag{3.141}$$

and in steady state (after a few time constants),

$$z_F(t) = y_F(t) \tag{3.142}$$

where y_F is the output of the prefilter driven by y only, that is,

$$A_p(d)\ y_F(t) = B_p(d)\ y(t) \tag{3.143}$$

Similarly filtering u we get u_F given by,

$$A_p(d)\ u_F(t) = B_p(d)\ u(t) \tag{3.144}$$

We can express the above operationally as,

$$y_F(t) = [\ B_p(d)/A_p(d)\ ]\ y(t) \tag{3.145}$$

and,

$$u_F(t) = [\ B_p(d)/A_p(d)\ ]\ u(t) \tag{3.146}$$

Now applying the prefilter to the estimation equation yields,

$$\hat{z}_F(t) = A(d)\hat{z}_F(t) + B(d)u_F(t) + \tilde{G}(d)e_F(t) \tag{3.147}$$

where,

$$e_F(t) = [\ B_p(d)/A_p(d)\ ]\ e(t) \tag{3.148}$$

Everything now proceeds as in the case without poles except that the relation between e_F and e is slightly more complicated.

III.6.2 Separate Component Identification by an Iterative Approach:

The idea here is to again separate the identification of $\underline{x}_p$ from the rest of the parameters and take advantage of the computationally simpler linear estimation procedure in section II.8 to estimate the latter. In the previous section we discussed a prefiltering technique which does this in one shot and has certain advantages and disadvantages as explained. In this section we skip the need to prefilter the observed data and attempt to iteratively solve the same problem.

We start with an estimate of $\underline{x}_p$, say $\underline{x}_p^0$, and call the corresponding periodic function y_p^0.

This is then used to define a first estimate of the residual term, y^0, as shown below,

$$y^0(t) = z(t) - y_p^0(t) \quad (3.149)$$

$$= y(t) + [y_p(t)-y_p^0(t)] \quad (3.150)$$

$$\equiv y(t) + \Delta y_p^0(t) \quad (3.151)$$

We can consider $\Delta y_p^0(t)$ as a small observation error in y and lump its effect, except for its mean value, into the y model system noise. We then have a model for y^0 as follows,

$$y^0(t) = \sum_{i=1}^{n} a_i y^0(t-i) + \sum_{j=0}^{n} b_j u(t-j) + \sum_{i=1}^{n}[\tilde{d}_i \tilde{w}(t-i)] + m \quad (3.152)$$

where $\tilde{w}$ is a white process with zero mean and m is the mean level introduced by the constant term in Δy_p^0.

Since y^0 is known we can then fit the above model to this data by the least squares technique of section III.4 without $\underline{x}_p$ naturally. This gives us a first set of y^0 model parameters, $\underline{a}^0$, $\underline{b}^0$, $\underline{g}^0$, R_e^0 and m^0.

We can now use linear estimation to reestimate $\underline{x}_p$, $\underline{x}_p^1$, and repeat the procedure a fixed set of times say, or until no further improvement is achieved.

Description of Technique

The basic idea is that the zero mean effect of the error term Δy_p^1 can be approximated by a zero mean moving average as in equation 3.152. The mean value m can be estimated as part of the identification problem.

No guarantee of convergence is given, but to increase our chances of success a good initial estimate of $\underline{x}_p^0$ is important.

Basically what we are doing here is solving the over-all identification problem by searching for a minimum in specific directions, that is constant $\underline{x}_p$ first, followed by constant $\underline{a}$, $\underline{b}$, $\underline{g}$ and so on, whereas the main approach uses the more powerful Fletcher-Powell search.

Evaluation of Initial $\underline{x}_p$

In general y_p is considerably larger than y, particularly if weather conditions do not deviate from normal by much. Furthermore y taken over a large period of time oscillates between positive and negative values. To estimate $\underline{x}_p$ we then make use of a weighted least squares approach which weighs observed data more heavily if weather <u>does not</u> deviate much from normal.

Since,

$$z(t) = \underline{\phi}^T \underline{x}_p + y(t) \tag{3.153}$$

we then define the estimates of $\underline{x}_p$, $\underline{x}_p^0$, as that value of $\underline{x}_p$ which minimizes,

$$J = \sum_{t=1}^{N} [z(t) - \underline{\phi}^T(t)\underline{x}_p]^2 \, W(t) \tag{3.154}$$

where W(t) is a positive weighting factor with the properties discussed above. Its exact value being picked by engineering judgement and trial and error.

Differentiating J with respect to $\underline{x}_p$ and equating to zero yields,

$$\underline{x}_p^0 = \{ \sum_{t=1}^{N} [\underline{\phi}(t)W(t)\underline{\phi}^T(t)]\}^{-1} \sum_{t=1}^{N} \underline{\phi}(t)W(t)z(t) \tag{3.155}$$

III.7 <u>Special Models</u>:

In this section we discuss a much simpler version of the model so far discussed. It is simpler in the sense that it makes the mathematics much more tracktable, in particular the search for a minimum in the identification problem.

The primary assumption is that the system uncertainty c(t), is no longer a correlated process but it is white, that is,

$$c(t) = w(t-1) \tag{3.156}$$

where w is white.

The model structure and advantages remain the same except that a smaller number of parameters are used in fitting the model. In any case this type of model is normally tried before going to more complicated ones.

The noise assumption means that the residual load model is now of the form,

$$y(t) = \sum_{i=1}^{n} a_i y(t-i) + \sum_{j=0}^{n} b_j u(t-j) + w(t-1) \qquad (3.157)$$

while $\hat{y}$, the best linear estimate of y given z (knowing $\underline{x}_p$), is,

$$\hat{y}(t) = \sum_{i=1}^{n} a_i y(t-i) + \sum_{j=0}^{n} b_j u(t-j) \qquad (3.158)$$

so that,

$$e(t) = y(t) - \hat{y}(t) = w(t-1) \qquad (3.159)$$

or,

$$e(t) = y(t) - \sum_{i=1}^{n} a_i y(t-i) - \sum_{j=0}^{n} b_j u(t-j) \qquad (3.160)$$

But from 3.59 we know that,

$$e(t) = \sum_{j=1}^{n-1} g_{j+1} e(t-j) + [y(t) - \sum_{i=1}^{n} a_i y(t-i) - \sum_{j=0}^{n} b_j u(t-j)] \qquad (3.161)$$

which implies that,

$$g_{j+1} = 0 \quad ; \quad \forall\ j = 1, 2, \ldots, n-1 \qquad (3.162)$$

We thus see that the number of parameters we must search for has been drastically reduced.

The function we try to minimize here is still,

$$J = 1/N \sum_{t=1}^{N} e^2(t) \qquad (3.163)$$

$$= 1/N \sum_{t=1}^{N} [y(t) - \underline{a}^T \underline{y}(t) - \underline{b}^T \underline{u}(t)]^2 \qquad (3.164)$$

or in terms of the observed z and $\underline{x}_p$,

$$J = 1/N \sum_{t=1}^{N} [z(t)-\underline{\phi}^T(t)\underline{x}_p-\underline{a}^T\underline{z}(t)-\underline{b}^T\underline{u}(t)+\underline{a}^T\underline{\psi}(t)\underline{x}_p]^2 \quad (3.165)$$

Thus no longer do we have to search for $\underline{g}$ and in addition the minimization is not constrained, a much simpler problem.

III.7.1 Solutions to Special Problem:

The solutions discussed in the previous section can equally be applied to the special problem.

A) Fletcher-Powell

Here the value of J and its gradient can be determined directly from observed data. J is obtained from 3.165 and the gradient as below,

$$\partial J/\partial \underline{x}_p = 1/N \sum_{t=1}^{N} 2e(t)[\underline{\psi}^T(t)\underline{a}-\underline{\phi}(t)] \quad (3.166)$$

$$\partial J/\partial \underline{a} = 1/N \sum_{t=1}^{N} 2e(t)[\underline{\psi}(t)\underline{x}_p-z(t)] \quad (3.167)$$

$$\partial J/\partial \underline{b} = 1/N \sum_{t=1}^{N} 2e(t)[-\underline{u}(t)] \quad (3.168)$$

B) Component Separation

i) Prefiltering

The same results apply. After prefiltering the observed data we have,

$$z_F(t) = A(d)z_F(t) + B(d)u_F(t) + B_p(d)w(t-1) \quad (3.169)$$

which we can use as shown in III.6 to estimate $\underline{a}$, $\underline{b}$ and R_e. Note that we don't have to search over $\underline{g}$ or $\underline{x}_p$, but that we still have a constrained optimization problem due to the term $B_p(d)w$.

ii) Iterative Approach

The technique goes through again without major improvements, unless we make the assumption that the effect of Δy_p^k is negligible. Then we have for the kth iteration,

$$y_k(t) = \underline{a}^T \underline{y}_k(t) + \underline{b}^T \underline{u}(t) + m + w(t-1) \tag{3.170}$$

which can be expressed as,

$$y_k(t) = [\underline{y}_k^T(t),\ \underline{u}^T(t),\ 1] \begin{pmatrix} \underline{a} \\ \underline{b} \\ m \end{pmatrix} + w(t-1) \tag{3.171}$$

$$\equiv \quad \underline{h}_k^T(t) \begin{pmatrix} \underline{a} \\ \underline{b} \\ m \end{pmatrix} + v(t) \tag{3.172}$$

so that the values of $\underline{a}$, $\underline{b}$ and m can be estimated via linear estimation techniques.

If we collect all the k iteration observations into one large vector $\underline{Y}_k$, we have,

$$\underline{Y}_k = \begin{pmatrix} y_k(1) \\ y_k(2) \\ \cdot \\ \cdot \\ \cdot \\ y_k(N) \end{pmatrix} = \begin{pmatrix} \underline{h}_k^T(1) \\ \underline{h}_k^T(2) \\ \cdot \\ \cdot \\ \underline{h}_k^T(N) \end{pmatrix} \begin{pmatrix} \underline{a} \\ \underline{b} \\ m \end{pmatrix} + \begin{pmatrix} v(1) \\ v(2) \\ \cdot \\ \cdot \\ \cdot \\ v(N) \end{pmatrix} \tag{3.173}$$

$$\equiv \quad \underline{H}_k \ \underline{\theta}_k + \underline{v}_k \tag{3.174}$$

The best linear estimate of $\underline{\theta}_k$ is then,

$$\hat{\underline{\theta}}_k = \begin{pmatrix} \hat{\underline{a}} \\ \hat{\underline{b}} \\ \hat{m} \end{pmatrix} = [\ \underline{H}_k^T \ \underline{H}_k\ ]^{-1}\ \underline{H}_k^T \ \underline{Y}_k \tag{3.175}$$

an easy calculation to make.

C) Linearization

Here we take advantage of the simple form of the model to discuss an alternative method of estimating its parameters. Since,

$$z(t) = \underline{\phi}^T(t)\,\underline{x}_p + y(t) \tag{3.176}$$

$$= \underline{\phi}^T(t)\,\underline{x}_p + \underline{a}^T\underline{y}(t) + \underline{b}^T\underline{u}(t) + w(t-1) \tag{3.177}$$

then,

$$z(t) = \underline{\phi}^T(t)\,\underline{x}_p + \underline{a}^T\underline{z}(t) + \underline{b}^T\underline{u}(t) - \underline{a}^T\underline{\psi}(t)\underline{x}_p + w(t-1) \tag{3.178}$$

$$\equiv F(\underline{\theta}, t) + w(t-1) \tag{3.179}$$

Rather than considering the estimation of the parameters assuming these to be totally unknown as we have been doing, let us linearize about some initial guess of $\underline{x}_p$ and $\underline{a}$ and consider a linear estimation problem. Let,

$$\underline{x}_p = \underline{x}_p^0 + \delta\underline{x}_p^0 \tag{3.180}$$

$$\underline{a} = \underline{a}^0 + \delta\underline{a}^0 \tag{3.181}$$

Then ignoring higher order effects we have,

$$\begin{aligned} z(t) = {} & \underline{\phi}^T(t)\underline{x}_p^0 + \underline{\phi}^T(t)\delta\underline{x}_p^0 + \underline{a}^{0T}\underline{z}(t) + \delta\underline{a}^{0T}\underline{z}(t) + \underline{b}^T\underline{u}(t) \\ & -\underline{a}^{0T}\underline{\psi}(t)\underline{x}_p^0 - \underline{a}^{0T}\underline{\psi}(t)\delta\underline{x}_p^0 \delta\underline{a}^{0T}\underline{\psi}(t)\underline{x}_p^0 + w(t-1) \end{aligned} \tag{3.182}$$

so that,

$$\tilde{z}(t) \triangleq z(t) - \underline{\phi}^T(t)\underline{x}_p^0 - [\underline{a}^0]^T\underline{z}(t) + [\underline{a}^0]^T\underline{\psi}(t)\underline{x}_p^0 \tag{3.183}$$

$$= \begin{pmatrix} [\underline{z}(t) - \underline{\psi}(t)\underline{x}_p^0] \\ \underline{u}(t) \\ [\underline{\phi}(t) - \underline{\psi}^T(t)\underline{a}^0] \end{pmatrix}^T \begin{pmatrix} \delta\underline{a}^0 \\ \underline{b}_o \\ \delta\underline{x}_p^0 \end{pmatrix} + w(t-1) \tag{3.184}$$

$$= \underline{h}^T(t)\,\delta\underline{\theta}^0 + \underline{v}(t) \tag{3.185}$$

whose solution by linear estimation techniques as discussed in part B is readily obtained. Having an estimate of $\delta\underline{\theta}^0$

we can then reestimate $\underline{\theta}$ by adding this correction, linearizing again and so on.

This is basically the same result as for part B(ii) except that there we don't assume an initial guess of $\underline{a}$.

A reasonable starting value for the value of $\underline{a}$ could be such that the time constants of the system are within those expected, e.g. a few hours to two days, and that the system be asymptotically stable.

III.8 Maximum Likelihood Interpretation:

So far the criterion used for choosing a set of system parameters has been a least squares criterion,

$$J = 1/N \sum_{t=1}^{N} [z(t)-\hat{z}(t)]^2 \qquad (3.186)$$

Essentially, by minimizing the above we want the variance of the error between observed and model behaviour to be a minimum. Since this is a very reasonable engineering criterion we expect reasonable results.

It is very satisfying though to know that under some additional assumptions this criterion is optimum in a more general form, that is maximum likelihood, (Ref. 14).

This criterion is defined as follows:

Define $\underline{z}_k$ as the vector of all observations up to time k,

$$\underline{z}_k = [z(1), z(2), \ldots, z(k)] \qquad (3.187)$$

Consider the probability density function of $\underline{z}_N$, the whole set of observed data vector, as a function of the unknown system parameters, $\underline{\theta}$, or $\underline{x}_p$, $\underline{a}$, $\underline{b}$, $\underline{d}$ and Q, i.e. $p(\underline{z}_N, \underline{\theta})$.

Then the maximum likelihood estimate (MLE) of $\underline{\theta}$, $\underline{\theta}^*$, is that value of $\underline{\theta}$ which maximizes $p(\underline{z}_N, \underline{\theta})$.

In order to solve this problem we must then define $p(\underline{z}_N, \underline{\theta})$ as a function of $\underline{\theta}$. To do this consider the use of Bayes' rule,

$$p(\underline{z}_N, \underline{\theta}) = p(z(N)/\underline{z}_{N-1}, \underline{\theta})\, p(\underline{z}_{N-1}, \underline{\theta}) \qquad (3.188)$$

which applied N times yields,

$$p(\underline{z}_N, \underline{\theta}) = \prod_{k=1}^{N} p(z(k)/\underline{z}_{k-1}, \underline{\theta}) \qquad (3.189)$$

Now if in our model we assume that the system input uncertainty w is Gaussian then z will also be Gaussian by linearity. This then implies that $p(z(k)/\underline{z}_{k-1})$ will be a normal distribution with mean,

$$\hat{z}(k) = E\,\{z(k) / \underline{z}_{k-1}\} \qquad (3.190)$$

and variance,

$$R_e(k) = E\{ [z(k)-\hat{z}(k)]^2\} \qquad (3.191)$$

Now $\hat{z}(k)$ under Gaussian assumption is the output of the Kalman filter discussed in the earlier sections while $R_e(k)$ is the error variance which for large N approaches a constant value which we will call R_e. Since $p(z(k)/z_{k-1})$ is a Gaussian density function then,

$$p(z(k)/z_{k-1}) = [1/\sqrt{2\pi R_e}]\, \exp\left\{-\frac{1}{2}\,\frac{[z(k)-\hat{z}(k)]^2}{R_e}\right\} \qquad (3.192)$$

Thus, consider the density function $p(\underline{z}_N, \underline{\theta})$,

$$p(\underline{z}_N, \underline{\theta}) = \frac{1}{[2\pi R_e]^{N/2}} \exp\left\{\frac{-1}{2R_e}\sum_{k=1}^{N} [z(k)-\hat{z}(k)]^2\right\} \qquad (3.193)$$

Its maximum and the MLE of $\underline{\theta}$ can be found by maximizing,

$$\ln p(\underline{z}_N, \underline{\theta}) = -\frac{N}{2} \ln(2\pi R_e) - \frac{1}{2R_e} \sum_{k=1}^{N} [z(k)-\hat{z}(k)]^2 \qquad (3.194)$$

Maximizing first with respect to R_e yields,

$$\frac{N}{R_e} - \frac{1}{R_e^2} \sum_{k=1}^{N} [z(k)-\hat{z}(k)]^2 = 0 \qquad (3.195)$$

that is,

$$\hat{R}_e = 1/N \sum_{k=1}^{N} [z(k)-\hat{z}(k)]^2 \qquad (3.196)$$

where $\hat{R}_e$ is the estimate of R_e.

Substituting $\hat{R}_e$ back into 3.194 we reduce the problem to minimizing,

$$J = 1/N \sum_{k=1}^{N} [z(k)-\hat{z}(k)] \qquad (3.197)$$

Under mild assumptions it is shown in references 18, 19 and 20 that the model estimates thus found are sufficient and consistent. Essentially this means that as more data is considered, i.e. $N\to\infty$, one can't do better than this estimate and that this estimate converges to its true value.

Although in practice the Gaussian assumption is not exactly met and our data record is not infinite, we do approximate these conditions, increasing our confidence in the criterion.

III.9 Estimation of Model Order:

The basic idea behind tests to determine the model order is to compare the value of J for different values of the system dimensions and test whether this quantity has decreased "significantly".

Since we are working with a finite data sample, the more parameters we include in the model, the better the fit.

Continuously increasing the number of parameters however results in fitting not the predictable or correlated portion of the model, but the noisy part which cannot be modelled or predicted. Clearly using such a model to predict under different conditions would result in erroneous results.

The principle used in this study is: start with the simplest model; identify the parameters, then predict; next test the one-step prediction errors for whiteness and check the ability of the model to predict within reasonable bounds. If this is approximately satisfied we stop increasing the model's complexity. Real data illustrations of this approach are discussed in section IV.6.2.

If the model structure is exactly as hypothesized, then more rigorous statistical tests for estimating the model order are available (Ref. 18). The basic physical idea is however as described above.

III.10 Adaptive Model Parameter Identification:

In the previous sections we have discussed methods for evaluating the load model parameters from observed data. The general requirement we desired was that a large data sample be used in making this determination. We implicitly assumed that if the sample were sufficiently large, the model input would vary over a wide spectrum and excite all the system modes. This is a necessary condition to be able to identify the system parameters.

We are however limited to the length of the data sample that can be used by the slow time variation of the load behaviour and our requirement that this sample be approximately stationary. For this reason a sample length of 3 weeks was considered for the weekdays, that is 15 days or 360 samples of z and an equal amount of u. Over this period the variations in u are very likely to be considerable and the time variation of the parameters relatively small.

The model thus obtained is considered valid for some time in the future, e.g. one week (later verified by experimentation), and can therefore be used for prediction purposes over that period.

Adaptive reestimation of the model parameters is now done periodically to track their slow variations with time. This can be done once a week at a time when the main computer is not needed for other important tasks.

When this step is carried out the new data is added to the top of the sample throwing out an equal amount of data at the bottom of the sample. This is called a finite memory identifier. Alternatively a longer or infinite memory identifier could be used by using weighting factors which weigh the most recent data most heavily and progressively reduce this weight for less recent data. Although the exact weights and data record should be determined experimentally, depending on the particular power company, the changing nature of the load behaviour with seasons is such that very large samples (i.e. greater than 6 weeks) might not give good results with this model even if weighting is used. The argument is that this model depends considerably on the normal temperature curve for a whole year, that is $\underline{x}_p$ and the y model parameters vary with $\hat{T}$, a variable that changes little over 2 or 3 weeks, but whose change is quite significant for longer samples. Furthermore, more variables then just $\hat{T}$ may be involved in the seasonal variation of load behaviour making the use of large samples unreliable for identification.

III.11 Detection of Anomalous Load Behaviour:

One of the advantages of using a probabilistic load model is that each forecast value has an associated measure of its error in terms of the standard deviation.

On the average, prediction errors will not deviate much from the standard deviation, with larger variations becoming less and less likely. For prediction purposes one could add one or two standard deviations of the prediction error to the actual prediction and expect actual deviations from this forecast to be very unlikely.

A second important use of this model property is that deviations in load behaviour from that expected by the model can be quantitatively analyzed to test for anomalous load behaviour. By anomalous behaviour we loosely mean load patterns deviating excessively from those predicted by the model. This could be caused by a number of reasons, unusually severe weather conditions, factory strikes, school closings, etc. The detection of the effect of these events on the load behaviour is important in order to give advance warning to generating stations in case additional power is needed. In addition this detection is necessary to discontinue updating the state and forecasts with abnormal data, as well as to eliminate this data from the identification record. This must be done as the model we use to forecast and which we periodically reidentify describes normal load behaviour only.

Discontinuing the use of abnormal data to update the state and load forecasts, does not mean however that we should discontinue predicting normal behaviour. We can still do this via the apriori or open loop model given by equations 2.27 through 2.29, 2.32 and 2.33.

The scheme for detecting abnormal load behaviour can take many variations but we will give here one which is intuitively obvious.

Consider the error in the one-step prediction, e(t),

$$e(t) = z(t) - \hat{z}(t) \qquad (3.200)$$

We know this to be a zero mean white process with constant variance R_e, provided z is generated by the assumed model, that is, z is normal.

Under the additional assumption that e is approximately Gaussian the mean and variance are sufficient to describe the probability density of e.

Defining now the standard deviation σ,

$$\sigma = \sqrt{R_e} \qquad (3.201)$$

we can associate probabilities with events |e| exceeding σ, 2σ, 3σ, etc., but in addition since e is white and Gaussian, the variables e(t) and e(τ) are statistically independent for all $t \neq \tau$, so that we can talk abou the probability of more than one residual, e, in succession, exceeding so many standard deviations.

Consider the probability of the following events (for N(0,σ)),

$$Pr\{A_t\} \equiv Pr\{ 2\sigma > |e(t)| > \sigma\} \simeq 0.27 \equiv p_A \qquad (3.202)$$

$$Pr\{B_t\} \equiv Pr\{ 3\sigma \geq |e(t)| \geq 2\sigma\} \simeq 0.04 \equiv p_B \qquad (3.203)$$

$$Pr\{C_t\} \equiv Pr\{ |e(t)| > 3\sigma \} \simeq 0.01 \equiv p_C \qquad (3.204)$$

for all t.

From this we can calculate the probability of combinations of events occuring, for example,

$$Pr\{A_t \text{ and } B_{t+1}\} = p_A p_B \simeq 0.01 \qquad (3.205)$$

$$Pr\{A_t \text{ and } C_{t+1}\} = p_A p_C \simeq 0.003 \qquad (3.206)$$

$$Pr\{A_t \text{ and } C_{t+1} \text{ and } B_{t+2}\} = p_A p_C p_B \approx 0.0001 \qquad (3.207)$$

where the subscript t describes the time at which the event occurs.

We now suggest the following scheme for detecting anomalous load behaviour, described in Fig. 9. A similar scheme can be programmed to detect the return to normal of load behaviour. This can be done by allowing the closed loop predictor to run during the disturbance.. When this subsides, the one-step prediction errors return to normal and a statistical test similar to the above could then be used to detect such an event.

The scheme suggested for the detection of anomalous behaviour is by no means the only or best approach. It has a memory of 3 steps and decides whether the combined probability of one, two or three residuals in succession being excessively high is small. If it is small, it first warns the operator that there is a certain probability of anomalous behaviour. If high residuals persist then the probability of this happening will become smaller than some arbitrarily chosen minimum level, p_{min}, which will determine anomalous behaviour. When a warning is given, no action will be taken unless the operator intervenes, and if the residuals return inside one standard deviation the warning will be removed. If anomalous behaviour does occur, updating will be discontinued, switching then to open loop prediction updating, i.e. using only the observed and predicted inputs, but no outputs.

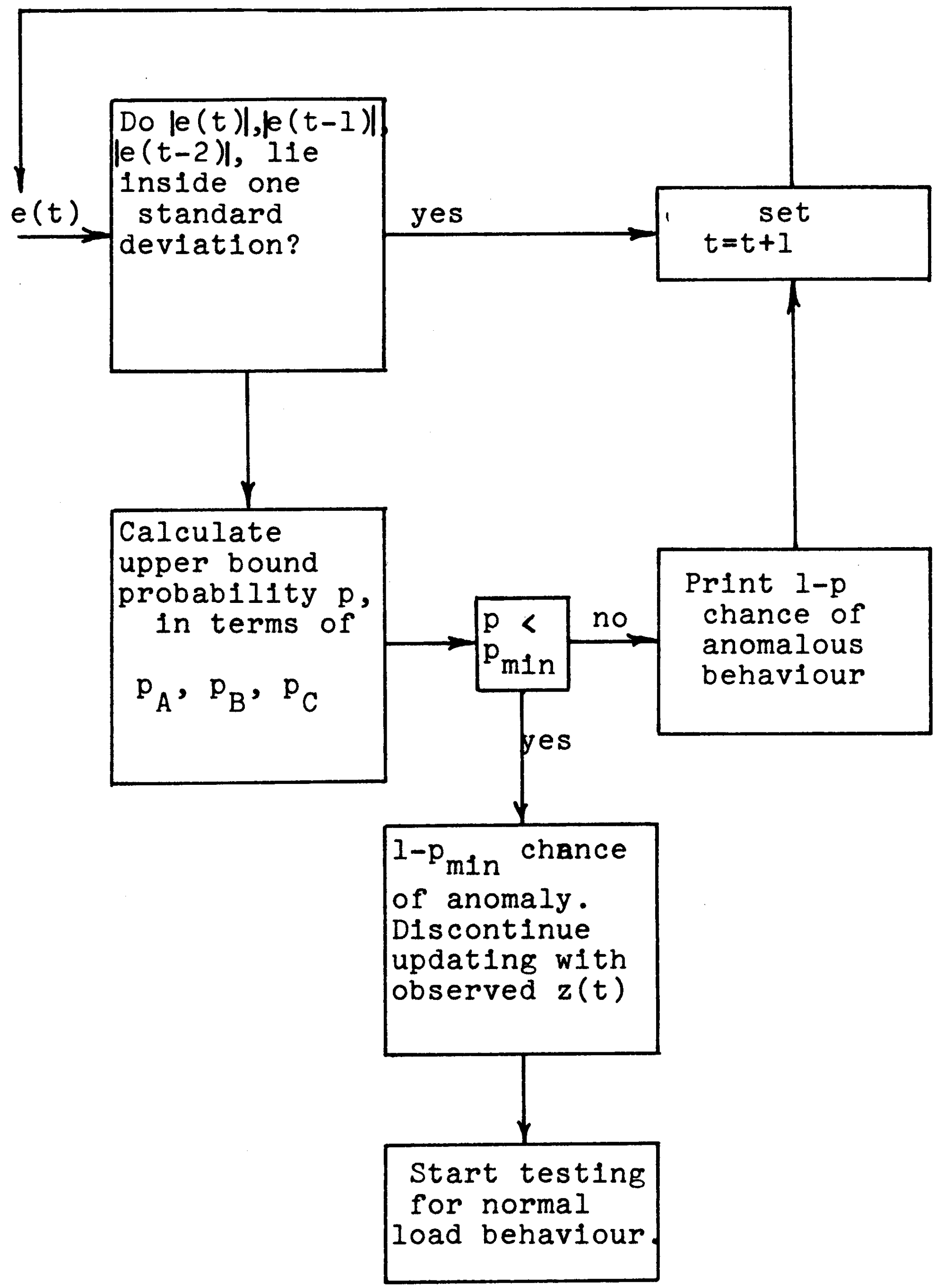


Fig. 9: Scheme for Testing Abnormal Load Behaviour Based on 3 Most Recent Residuals.

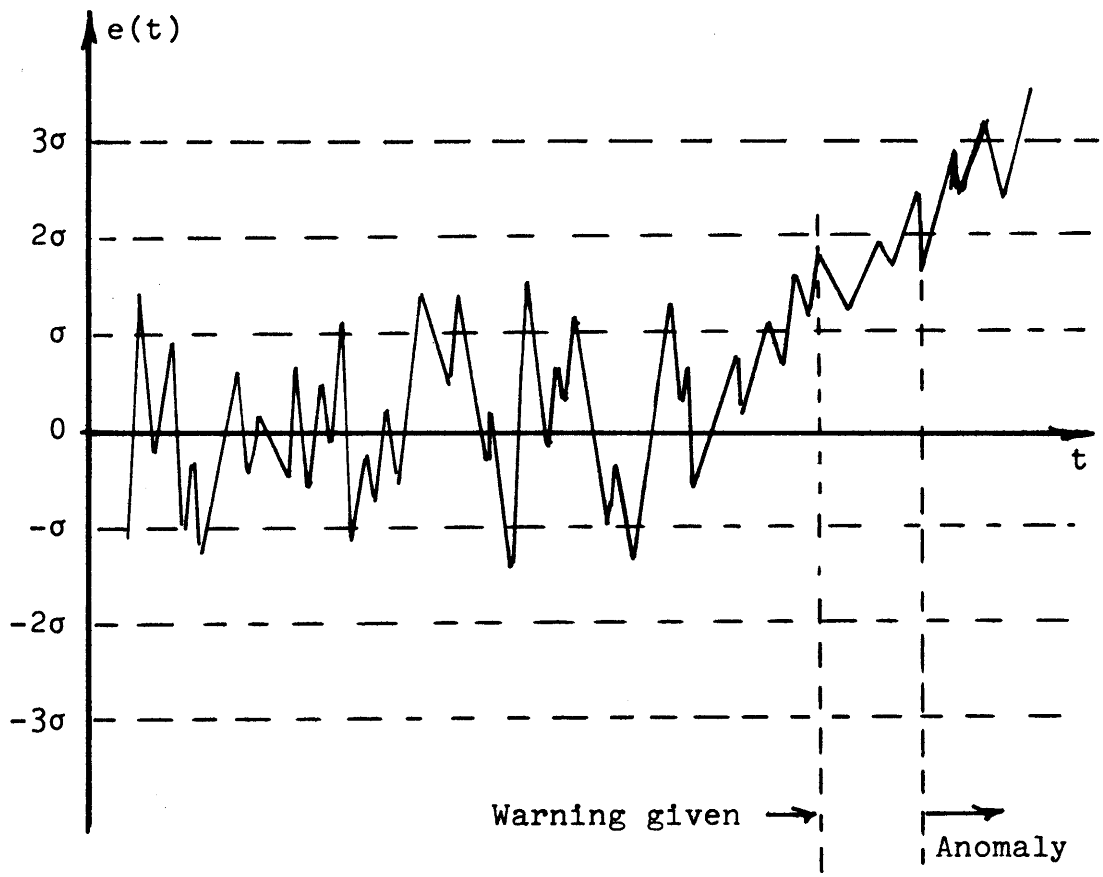


Fig. 10: Use of e(t) for Anomaly Detection.

A possible set of events may occur as shown in Fig. 10.

III.12 Summary of Load Model Identification:

In this chapter we have presented two main identification techniques:

A - Component Separation

B - Simultaneous Identification

A - The former consists of two methods each of which solves the total problem by the separate identification of the y_p and y model parameters. One by a prefiltering scheme which wipes out the effect of y_p, the other by an iterative scheme which alternatively searches over the y_p and y parameters. Although these approaches appear to reduce the computational complexity of the problem, in particular the prefiltering approach, more work is needed both theoretical and experimental before we can confidently use them.

B - The simultaneous identification of the y_p and y parameters makes use of the Fletcher-Powell algorithm for function minimization. This is a powerful technique and the question might be raised as to whether this is necessary to solve our particular problem. Nevertheless, until the previous techniques could be optimized, we decided to make use of it in our load forecasting study.

IV.0 EVALUATION OF LOAD MODELLING & FORECASTING TECHNIQUES

IV.1 Background:

In this chapter we describe the computer simulations and real data experimentation which have been carried out in the testing of the proposed forecasting technique. The various approaches suggested in the previous chapters are evaluated analytically and experimentally.

Specific load behaviour studies have been carried out primarily with load data from the Cleveland Electric Illuminating Company in Cleveland, Ohio. Additional information on more general load behaviour has also been obtained from the American Electric Power Service Corporation, New England Electric Co. and Detroit Edison Company (Ref. 23).

In addition to load data, weather data for the same period was obtained from the United States Weather Bureau at the Cleveland Hopkins Airport.

The load data is in hourly megawatthour values for the years 1968 and 1969, that is measurements of the average power consumed each hour. The technique is however not limited to hourly measurements, and similar studies could be particularly useful in anomaly detection and short term economic dispatch.

The weather data of interest here is dry bulb temperature in degrees Farenheit. Weather bureau measurements at tri-hourly intervals were available for the same period. These 8 daily values together with the maximum and minimum for the day were used to interpolate for the hourly temperatures (Ref. 24).

A normal temperature curve for the Hopkins Airport weather bureau averaged over a ten year period was also obtained. Average hourly values for each month of the year were provided, from which weekly levels were then interpolated.

The first attempt at testing the hypothesized model was a simple weighted least squares fit of a finite sum of sinusoids, that is the periodic component. This was done with real data and the results are discussed in IV.2.

This study was encouraging, and we next proceeded to test the identification techniques previously described. First, identification by component separation. The prefiltering technique was tested with simulated and real data. Linear identification was also tested, used in conjunction with the iterative search for y_p and y. These results are discussed in IV.3.

It appeared at this point that although the above identification procedures were very useful, their numerical refinement in order to use them with confidence were taking us beyond the main thesis objective of load forecasting.

We thus next proceeded to the more powerful simultaneous identification of the parameters of the y_p and y models by a Fletcher-Powell minimization scheme. This is discussed in IV.4.

This approach proved very successful so we proceeded to test the model primarily in this manner. First however, the estimation-prediction algorithms of chapter II were tested and programmed for a simulated load. This is discussed in section IV.5.

Section IV.6 discusses the identification of the model parameters with real data. Various models are tried and their forecasting capabilities are compared and evaluated. This is tried for various times of the year thus testing the ability of the model to adapt to seasonal variations. The detection of anomalous behaviour is also tested. In all cases the results are very good.

Section IV.7 discusses the computer requirements of the overall method. Section IV.8 summarizes the results of this chapter and the capabilities of the model based on real data results.

IV.2 Weighted Least Squares Estimation of y_p and y:

The theory behind this study is described in III.6.2 under the heading "Evaluation of initial $\underline{x}_p$", with equation 3.150 giving the estimated value of $\underline{x}_p$ and hence y_p. The difference from actual load is the estimate of y.

$$\underline{x}_p^N = \sum_{t=1}^{N} \{\underline{\phi}(t)W(t)\underline{\phi}^T(t)\}^{-1} \sum_{t=1}^{N} \underline{\phi}(t)W(t)z(t) \qquad (3.150)$$

The dimension of $\underline{x}_p$, $2n_p+1$, found to describe the most significant periodic effects lies between 9 and 15, that is 4 to 7 harmonics of the fundamental frequency $\omega_o=2\pi/24$.

The lower numbers fitted the periodic behaviour faithfully most of the day, but weakened slightly during peak periods, particularly during the twilight hours, around 9 PM, when lighting load suddenly comes on.

The weighting function W was chosen to be a constant for each day of the 12 weekdays considered. The 12th day was assumed to have maximum weight, whereas all earlier days were less heavily weighted. This tried to put less weight on data far in the past. In addition we weighted each day according to the average temperature deviation for that day and the previous one from the normal level. This was a "crude" way of demanding that $\underline{x}_p^N$ should describe only "normal" days, as well as introducing a memory effect into the weight.

The specific relation for W was as follows,

$$W_i = \left\{\frac{1}{1 + C_1 \Delta T_i + C_2 \Delta T_{i-1}}\right\} \beta^i \qquad (4.1)$$

where:

The subscript i stands for the particular day, i=1, ..., 12, to the last day.

β is an arbitrary constant, greater than one, picked in this study as 1.05, although different values were tried. If β is too small then the estimator "remembers" too far back; if β is too large then the memory is too short. The best value can be found by trial and error.

ΔT_i and ΔT_{i-1} stand for the deviations in daily average temperature from the daily normal level for the i^{th} and $(i-1)^{th}$ day.

C_1 and C_2 are arbitrary weighting constants. We picked $C_1=2$ and $C_2=0.4$.

This approach can easily be programmed for daily updating of $\underline{x}_p$. Defining,

$$\underline{H} = \begin{pmatrix} \underline{\phi}^T(1) \\ \underline{\phi}^T(2) \\ \vdots \\ \underline{\phi}^T(24) \end{pmatrix} \quad (4.2)$$

$$\underline{z}_i = \begin{pmatrix} z(1) \\ z(2) \\ \vdots \\ z(24) \end{pmatrix} \quad (4.3)$$

$$\underline{y}_i = \begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(24) \end{pmatrix} \quad (4.4)$$

where the subscript i stands for the particular day, we then have,

$$\underline{z}_i = \underline{H}\ \underline{x}_p + \underline{y}_i \quad (4.5)$$

and the weighted least squares criterion to be minimized is,

$$J = \sum_{i=1}^{N} [\underline{z}_i - \underline{H}\ \underline{x}_p]^T W_i\ [\underline{z}_i - \underline{H}\ \underline{x}_p] \quad (4.6)$$

where N (=12) is the number of days being considered. The estimate of $\underline{x}_p$, $\underline{x}_p^N$, is,

$$\underline{x}_p^N = [\ \sum_{i=1}^{N} W_i\ \underline{H}^T\underline{H}\ ]^{-1} \sum_{i=1}^{N} \underline{H}^T\ W_i\ \underline{z}_i \quad (4.7)$$

$$\underline{x}_p^N = [\underline{H}^T\underline{H}]^{-1}\,\underline{H}^T \sum_{i=1}^{N} W_i \underline{z}_i \Big/ \sum_{i=1}^{N} W_i \qquad (4.8)$$

which can readily be expressed in recursive form as,

$$\underline{x}_p^{N+1} = \underline{x}_p^N + W_{N+1}\,[\underline{x}_p^N - \{\underline{H}^T\underline{H}\}^{-1}\underline{H}^T\,\underline{z}_{N+1}] \Big/ \sum_{i=1}^{N+1} W_i \qquad (4.9)$$

Actually if we define,

$$\hat{\underline{x}}_p^{N+1} \equiv [\underline{H}^T\underline{H}]^{-1}\,\underline{H}^T\underline{z}_{N+1} \qquad (4.10)$$

we can recognize this as the minimum weighted least squares estimate of $\underline{x}_p$ given only the $(N+1)^{th}$ day's data. Thus this approach is on-line in the sense that we don't have to operate on data from previous days to update $\underline{x}_p^N$, but simply use,

$$\underline{x}_p^{N+1} = \underline{x}_p^N + W_{N+1}\,[\underline{x}_p^N - \hat{\underline{x}}_p^{N+1}] \Big/ \sum_{i=1}^{N+1} W_i \qquad (4.11)$$

where $\hat{\underline{x}}_p^{N+1}$ is given by 4.10.

Some of the results are described in Fig. 11 through 14. These clearly indicate the dependence of actual load on temperature. On the relatively hot days indicated in Fig. 11 and 12, the estimate of y, $\hat{y}$, is highly positive, slowly decaying as temperature drops. During the second day, the drop in $\hat{y}$ can be attributed to a sudden temperature drop which weather records indicate were due to a sudden rainstorm. As can be seen during a cool day (Fig. 13), $\hat{y}$ is negative, whereas for a typical day (Fig. 14), $\hat{y}$ is relatively small.

There is a possibility of using this approach to estimate $\underline{y}_p$, and then use the resulting estimate of y as data to identify the parameters of the y model. However no use is made of the assumed structure of the y model in separating

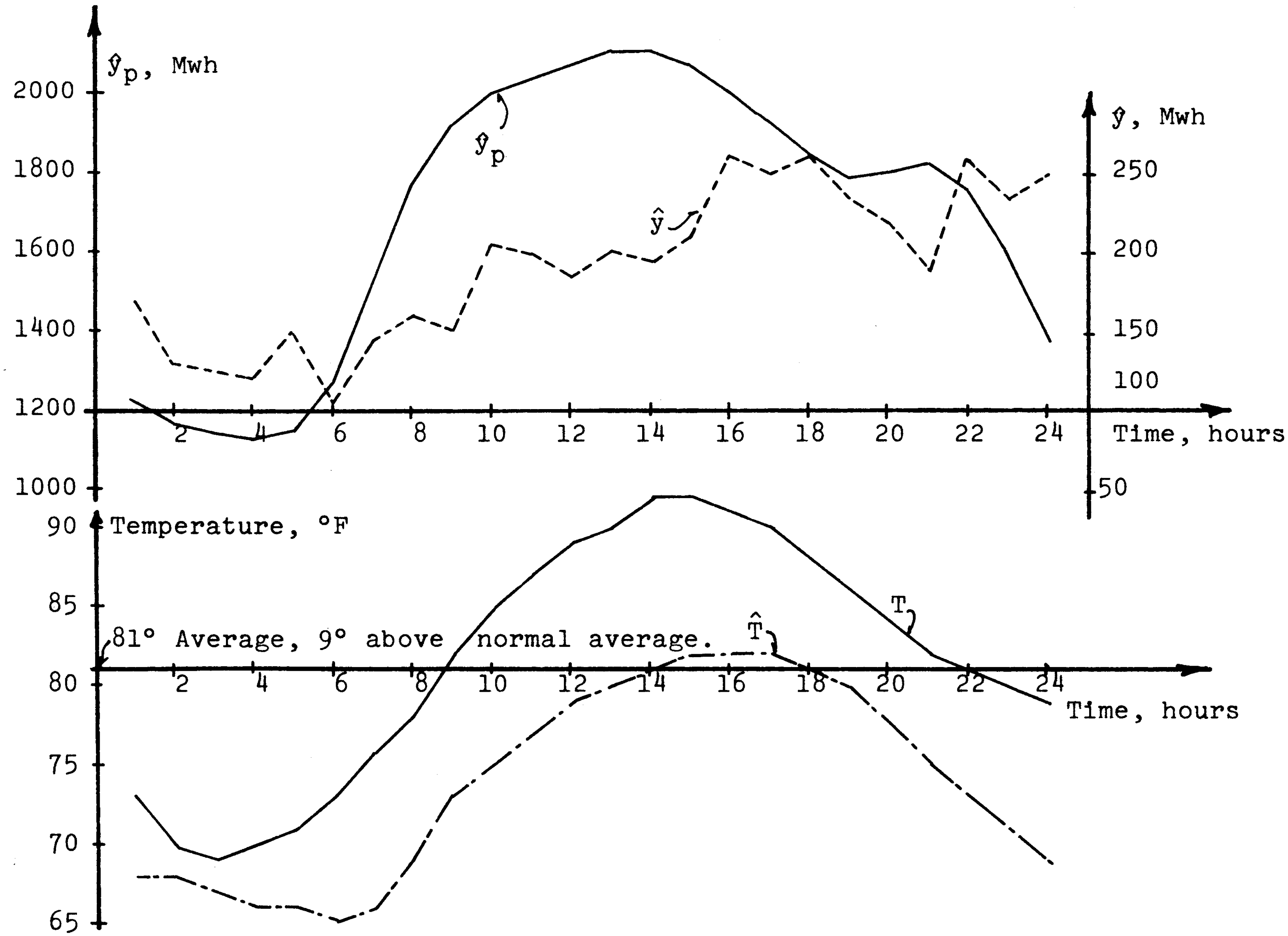


Fig. 11: <u>Weighted Least Squares Estimate of Periodic and Residual Components - July 16th, 1969</u>.

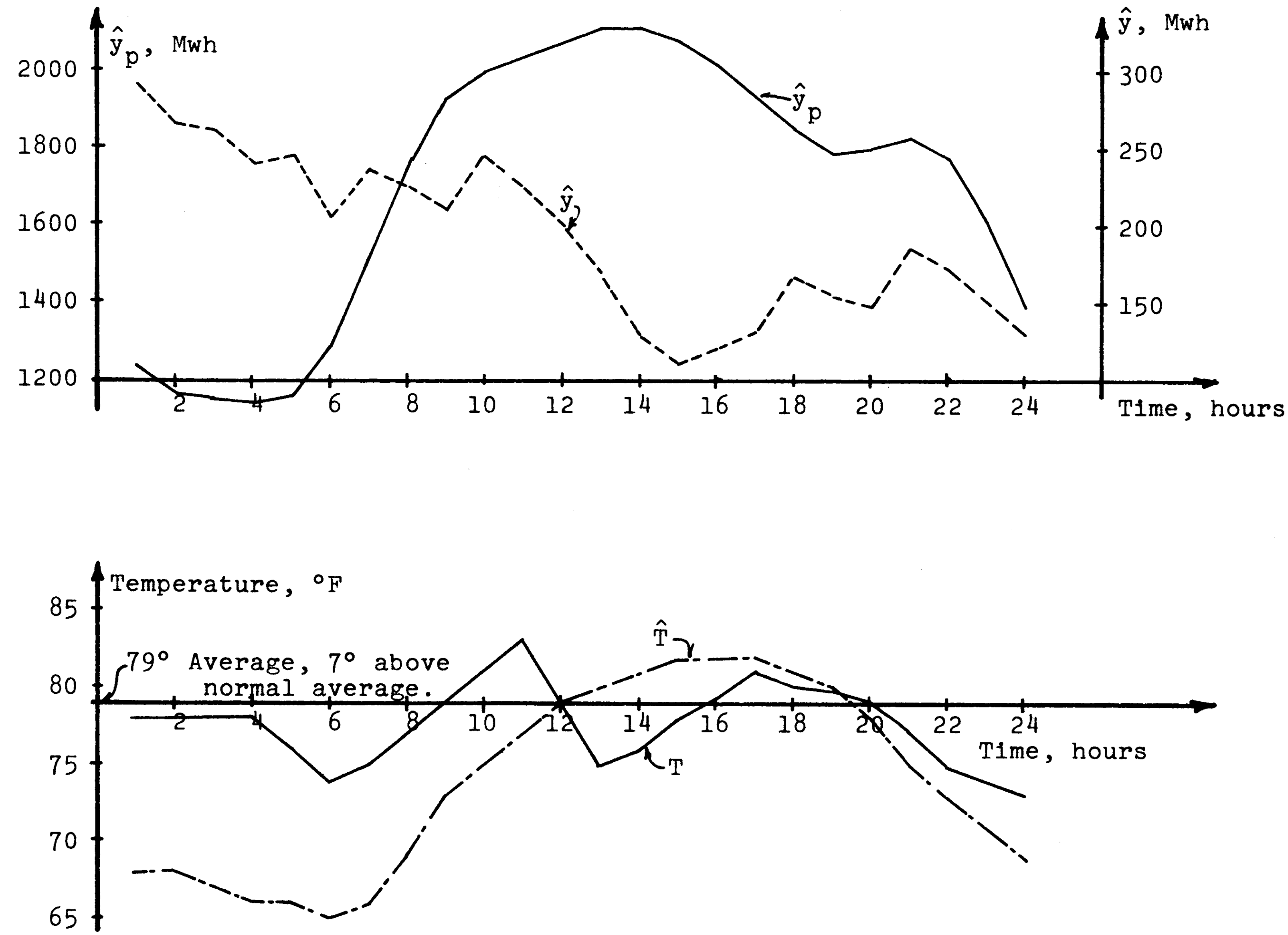


Fig. 12: Weighted Least Squares Estimate of Periodic and Residual Components - July 17th, 1969.

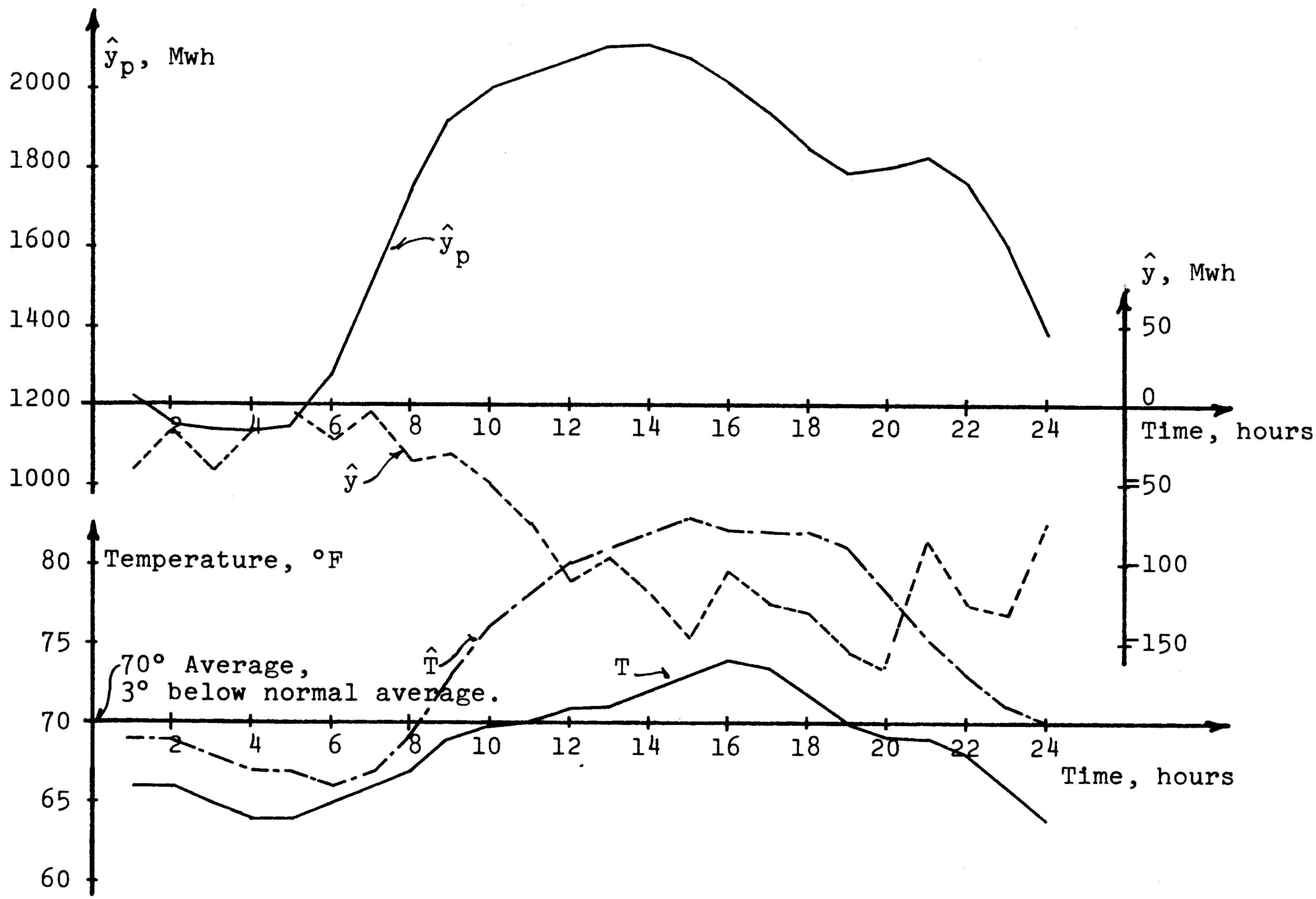


Fig. 13: Weighted Least Squares Estimate of Periodic and Residual Components - July 29th, 1969.

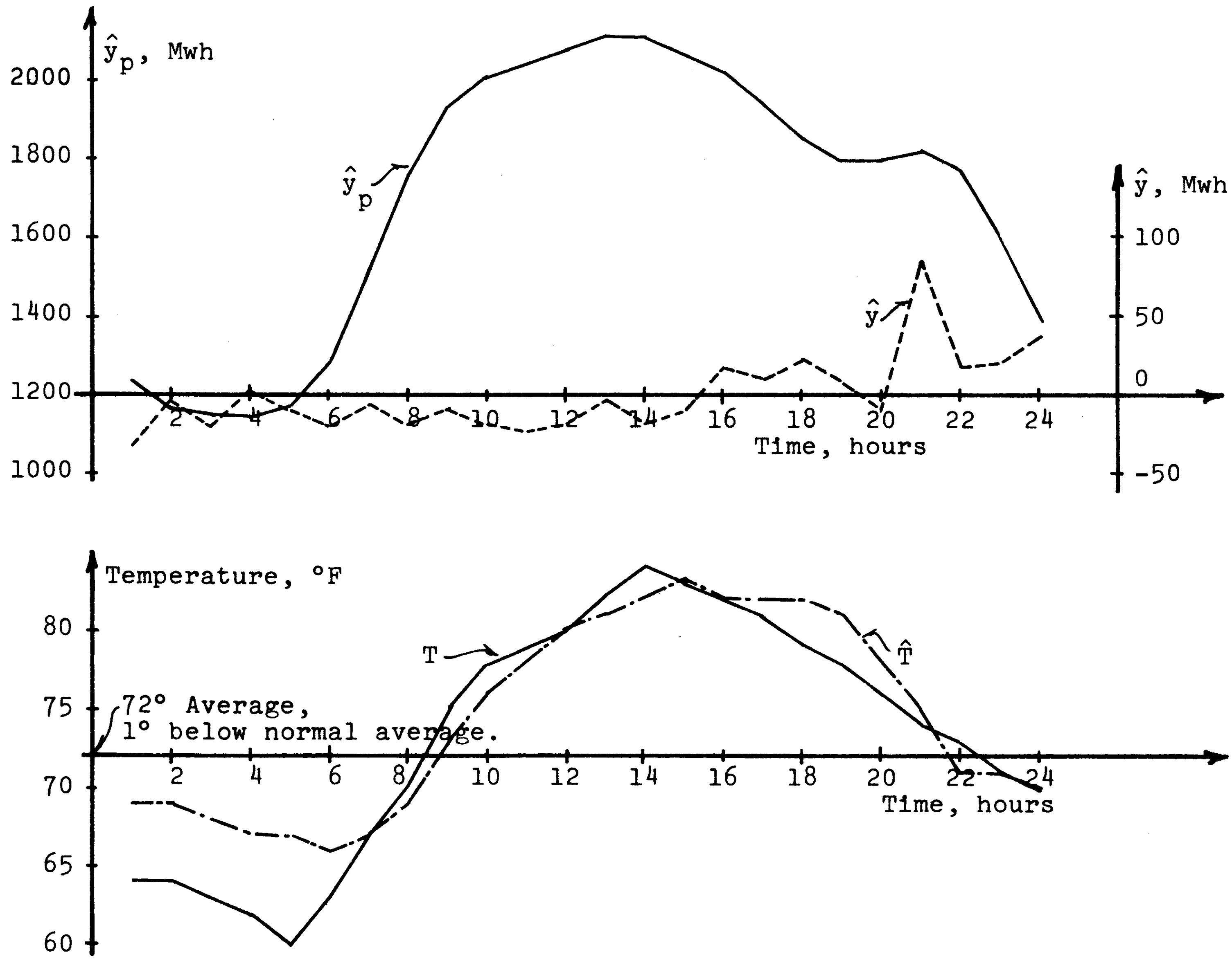


Fig. 14: Weighted Least Squares Estimate of Periodic and Residual Components - July 31st, 1969.

these effects and intuitively we have less confidence in the technique. Nevertheless the computational simplicity of this approach is something to be considered in future studies.

It should finally be noted that the value of the $\underline{x}_p$ estimate here obtained is used in later schemes as an initial estimate.

IV.3 Parameter Identification by Component Separation: Evaluation of Technique:

Considerable time was spent in developing techniques for model identification by component separation. Although insufficient confidence was obtained to apply these techniques to real data some of the results will be discussed, since computational savings may occur using this identification approach.

IV.3.1 Data Prefiltering Tests:

The prefilter was designed as described in the previous chapter, both with and without poles. The first case results in an extremely high sensitivity to observation errors. The prefilter poles were then designed to minimize this sensitivity, and an application of a simulated periodic function of the type hypothesized to such a filter is shown in Fig. 15. We can see that the response is relatively smooth, that is, after a while, no sharp peaks occur in the filtered data. Since theinput to the

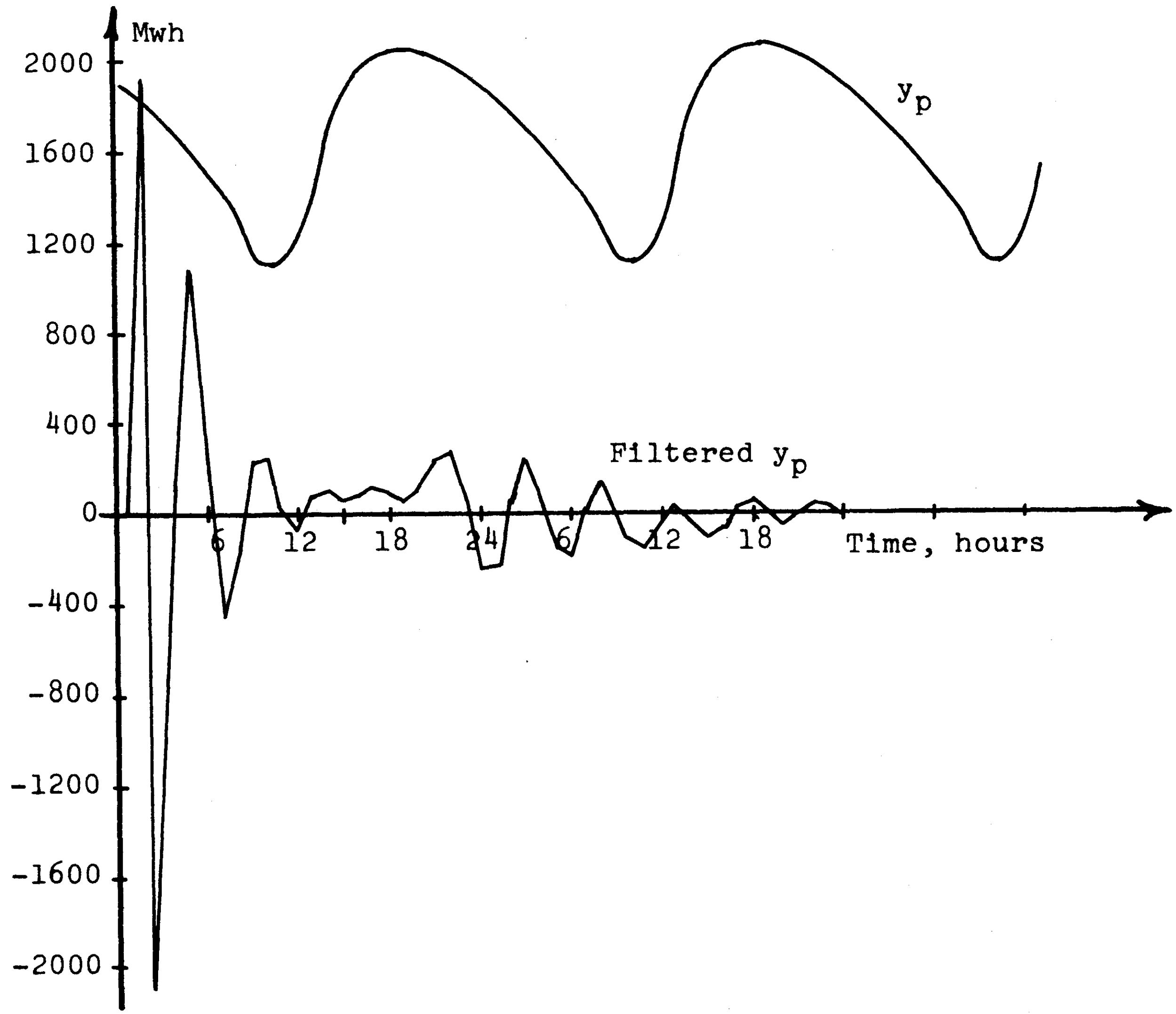


Fig. 15: <u>Test Filtering of Periodic Component (5 harmonics)</u> - After 48 steps filtered y_p less than 100 in magnitude.

to the prefilter is a periodic function made up of a finite sum of sinusoids, then the steady state output should be zero. As can be seen, a definite attenuation is present, but 48 or more data points are needed before we can confidently say that the filter output is close to zero. This naturally cuts down on the amount of data we can use for identification. However, the reduced problem, that is, the identification of the parameters of the filtered data model is much simpler, especially if we assume the system noise in the ARMA model to be white.

In Fig. 16 we show the effect of applying this filter to real load and temperature data. Here we can see one of the main drawbacks in this approach, which is the fact that whereas the unfiltered data shows obvious physical correlation between ΔT and z, the filtered data loses physical significance, making it more difficult to analyze the response from an engineering point of view.

More investigation with simulated data is needed before a concrete evaluation of this approach can be made.

IV.3.2 Iterative Component Separation-Evaluation:

The theory behind this approach is described in III.6.2. The basic idea is to search for the minimum

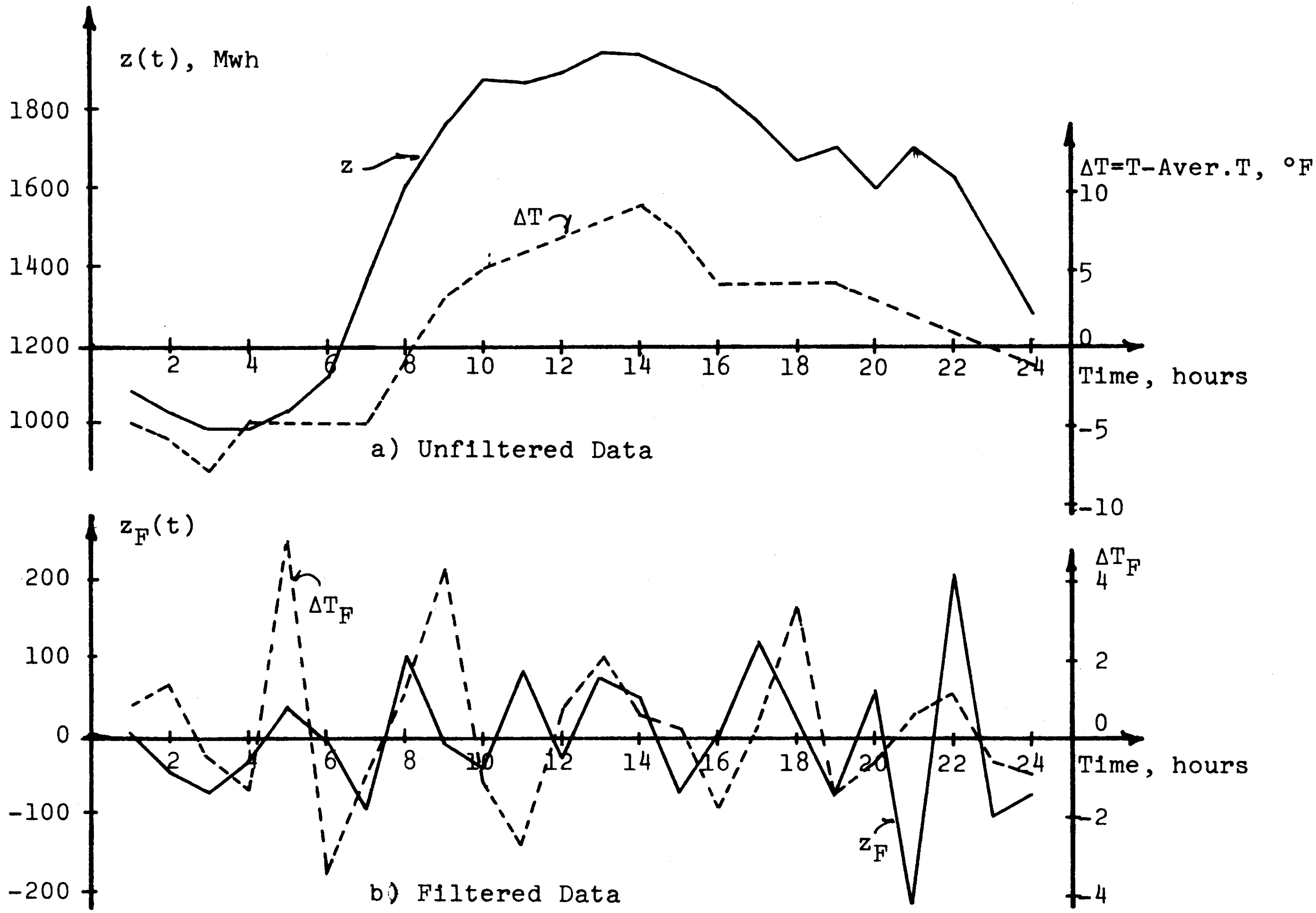


Fig. 16: Result of Prefiltering Load and Temperature Deviation Data July 9th, 1969

of the likelihood function, J, in specific directions. First along $\underline{x}_p^o$ as determined from IV.2, which yields an estimate of the y model parameters, call it $\underline{\theta}^o$, as discussed in III.6.2 and III.7.1. Using $\underline{\theta}^o$ and the results of chapter II on the linear estimation of $\underline{x}_p$, we get a new estimate of $\underline{x}_p$, $\underline{x}_p^1$. The procedure is repeated a number of times or until no further improvement is made.

If we use the special model of III.7, that is with white system noise, then $\underline{x}_p$ can be estimated in <u>one shot</u> given $\underline{\theta}^i$, where i stands for the i^{th} estimate of the y model parameters. The $(i+1)^{th}$ estimate of $\underline{x}_p$, $\underline{x}_p^{i+1}$, is then that $\underline{x}_p$ which minimizes,

$$J = 1/N \sum_{t=1}^{N} [z(t)-\underline{z}^T(t)\underline{a}^i-\underline{u}^T(t)\underline{b}^i-\underline{x}_p^T\underline{\phi}(t)+\underline{x}_p^T\underline{\psi}^T(t)\underline{a}^i]^2 \quad (4.12)$$

This clearly has a minimum given by,

$$\underline{x}_p^{i+1} = \{\sum_{t=1}^{N}[\underline{\phi}(t)-\underline{\psi}^T(t)\underline{a}^i][\underline{\phi}(t)-\underline{\psi}^T(t)\underline{a}^i]^T\}^{-1} \cdot \sum_{t=1}^{N}[z(t)-\underline{z}^T(t)\underline{a}^i-\underline{u}^T(t)\underline{b}^i][\underline{\phi}(t)-\underline{\psi}^T(t)\underline{a}^i] \quad (4.13)$$

This equation is computationally simpler and numerically more accurate than the iterative scheme discussed in chapter II, however it is only valid for the special model.*

The results of III.7.1, part B(ii), are the same as minimizing 4.12 with $\underline{x}_p$ fixed with respect to $\underline{a}$ and $\underline{b}$.

* The iterative approach is not limited to the special problem though.

It is clear at least for the special model, that this iterative approach results in a minimum for J, since for each iteration we find a minimum along the fixed direction. This is so, since fixing $\underline{x}_p$, or $\underline{a}$ and $\underline{b}$, makes J a convex function of the other parameters.

Now since the i+1 iteration is made along the direction defined by the previous minimum then $J^{i+1} \leq J^i$, where i stands for the i^{th} iteration, and since J is bounded below, a minimum must be reached.

The question in using this approach is just how fast can this minimum be reached, that is how many iterations are needed, and whether there are various minima.

In this study we restricted ourselves to a simple test of the iterative approach with simulated data. The reasons are similar. Much work was needed to perfect the iterative approach, which would deviate our attention excessively from the load forecasting problem.

This approach is, however, quite promising in that considerable computationalsavings <u>may</u> be made. As will be discussed later the Fletcher-Powell approach, although more reliable, requires considerable computer time, making long term off-line testing of the model quite expensive.*

The model simulated was of the form,

$$z(t) = 1500+100\sin(2\pi t/24)+100\cos(2\pi t/24)+y(t) \qquad (4.14)$$

* Although the cost of on-line implementation is minimal.

where,

$$y(t) = 1.4y(t-1)-0.49y(t-2)+3u(t)+u(t-1)+w(t-1) \quad (4.15)$$

We started with an initial guess of $\underline{x}_p$, as obtained, for example, from a weighted least squares estimation, given below,

$$\underline{x}_p^o = \begin{pmatrix} 1400 \\ 90 \\ 90 \end{pmatrix}$$

The results of III.7.1, part B(ii), are now applied, or equivalently we minimize J with respect to $\underline{a}$ and $\underline{b}$ with $\underline{x}_p$ fixed at the given value.

It was found by experimentation that $\underline{\hat{\theta}}$ in 3.170 is best solved for by a one shot solution rather than recursively. The latter approach resulting in greater numerical errors. The one shot solution can easily be determined by solving the system of equations,

$$\underline{H}_k^T \underline{H}_k \underline{\theta} = \underline{H}_k^T \underline{Y}_k \quad (4.16)$$

where $\underline{H}_k$ and $\underline{Y}_k$ are as defined in III.7.1. Simulations show this approach to be most accurate.

Solution of 4.16 for our particular example using 96 data points (4 days' data) yields,

$$\begin{aligned} \hat{a}_1 &= 1.69 && (1.4) \\ \hat{a}_2 &= -0.72 && (-0.49) \\ \hat{b}_0 &= 4.73 && (3.0) \\ \hat{b}_1 &= -2.85 && (1.0) \end{aligned} \quad (4.17)$$

with the true values in parentheses.

This first estimate of a is not bad. That of b is worse, but as it turns out this is the least sensitive set of parameters.

The next step would be to substitute the estimated values in 4.17 into 4.13 to obtain a new estimate of $\underline{x}_p$.

This approach was not pursued further as we turned to the more powerful, more general, more reliable and readily available in computer program form, Fletcher-Powell method for minimizing J.

IV.4 Parameter Identification via Fletcher-Powell - Evaluation through Simulation:

IV.4.1 Background:

In the previous chapter we defined the parameter identification problem, and showed that a reasonable solution could be found by solving a minimization problem. We argued that we could save considerable effort by not treating it as a general minimization problem, but taking advantage of the particular structure, to separate the identification of the model parameters in y_p and y. The two approaches proposed, data prefiltering and the iterative approach, are carefully described, but their numerical refinement was not pursued for lack of time and our desire to tackle the main problem of load forecasting.

The Fletcher-Powell approach (Ref. 22) is a set of recursive algorithms for the minimization of a function of many variables given the function value and derivative for any argument. The basic scheme is of the form,

$$\underline{\theta}_{k+1} = \underline{\theta}_k + \underline{G}_k \, \partial J/\partial \underline{\theta}_k \qquad (4.18)$$

where $\underline{\theta}_k$ is the k^{th} iteration estimate of the argument for which the function J is a minimum, $\partial J/\partial \underline{\theta}_k$ is the corresponding gradient and $\underline{G}_k$ is the Fletcher-Powell gain which is chosen according to the algorithm.

Basically $\underline{G}_k$ is such that at first it starts searching in the direction of steepest descent, progressively changing to a Newton-Raphson scheme as it gets closer to the minimum.

This scheme is particularly powerful when dealing with functions which have sharp ravines in given directions, smoother ones in others.

It is difficult to apriori estimate the nature of our function in the general case. However in the special case of white noise only, we see from equation 3.160 that J is a fourth order analytic function of $\underline{\theta}$. J can thus be well approximated by a quadratic near its minima, making the Newton-Raphson type search near the end of the iteration most desirable. Furthermore higher sensitivity is expected along the directions of $\underline{a}$ and $\underline{x}_p$, than along $\underline{b}$, making steep ravines quite possible, a condition for which the Fletcher-Powell method is well suited.

This approach also had the marked advantage that it was well tested, documented and programmed. In particular, we made use of an SSP subroutine, DFMFP (Ref. 25), in conjunction with our main identification program, as discussed in Appendix A.

In this section we discuss the identification of a simulated load model using the Fletcher-Powell algorithm in minimizing the function J.

It must finally be reemphasized that this approach may be too pwerful for this particular problem, but until further investigation is carried out on the methods discussed earlier, we will make use of this technique only.

IV.4.2 Simulation Results - Fletcher-Powell:

The model tested was of the form,

$$z(t) = y_p(t) + y(t) \quad (4.19)$$

where,

$$y_p(t) = 1500 + 100 \sin(2\pi t/24) + 100 \cos(2\pi t/24) \quad (4.20)$$

and,

$$y(t) = 1.4y(t-1)-0.49y(t-2)+3u(t)+u(t-1)+w(t-1) \quad (4.21)$$

where w is a white Gaussian process generated by the SSP subroutine Gauss. Its mean value and variance are given by,

$$E\{w(t)\} = 0 \quad (4.22)$$

$$E\{w^2(t)\} = 25 \quad (4.23)$$

The input u is tabulated in Table 3 and is chosen to approximately resemble a typical sequence of temperature deviations as defined in chapter II. As can be seen the actual values of z are in the typical load value neighbourhood.

u	z	u	z
-4	1599	2	1785
-5	1585	1	1787
-3	1562	-1	1772
-3	1535	2	1750
-3	1509	2	1733
-3	1488	2	1710
-3	1453	2	1684
-2	1419	6	1659
-1	1379	10	1652
0	1351	12	1662
1	1337	13	1686
2	1332	14	1724
2	1339	15	1766
3	1360	16	1813
5	1388	14	1856
3	1424	11	1886
3	1458	11	1908
3	1496	11	1928
2	1529	11	1955
0	1565	10	1979
-2	1586	9	1992
-3	1599	8	2004
-4	1598	7	2012
-5	1581	6	2015
-5	1558	5	2011
-7	1527	4	1994
-8	1483	5	1975
-9	1433	5	1947
-8	1382	6	1919
-6	1341	7	1893
-4	1306	8	1873
-2	1287	10	1856
1	1278	12	1851
3	1283	15	1853
5	1304	16	1863
7	1346	17	1884
8	1405	18	1912
10	1471	19	1948
7	1539	20	2003
5	1591	19	2053
4	1623	18	2108
4	1646	17	2148
3	1665	16	2177
3	1690	14	2197
2	1704	11	2205
2	1717	9	2204
2	1744	7	2196
2	1766	5	2172

Table 3: Simulated Inputs and Corresponding Outputs for Testing of Identification Technique.

The initial guess at the values of the system parameters was as follows,

$$
\begin{aligned}
\hat{a}_1 &= 0.9 && (1.4) \\
\hat{a}_2 &= -0.2 && (-0.49) \\
\hat{b}_0 &= 0.8 && (3.0) \\
\hat{b}_1 &= 0.8 && (1.0) \\
\hat{x}_{po} &= 900 && (1500) \\
\hat{x}_{p1} &= 90 && (100) \\
\hat{x}_{p2} &= 90 && (100)
\end{aligned}
\qquad (4.24)
$$

the numbers in parentheses being their true values.

Since w(t) is white, we make use of the identification solution for the special model described in the previous chapter. This solution was programmed and is described in Appendix A. Actually, as section IV.6 will show, the special model yields quite reasonable results, thus, unless further refinements are desired, the special model may prove sufficient.

After about 50 iterations a minimum of J is achieved. The estimated parameters are,

$$
\begin{aligned}
\hat{a}_1 &= 1.32 && (1.40) \\
\hat{a}_2 &= -0.44 && (-0.49) \\
\hat{b}_0 &= 2.1 && (3.0) \\
\hat{b}_1 &= 2.9 && (1.0) \\
\hat{x}_{po} &= 1503 && (1500) \\
\hat{x}_{p1} &= 106 && (100) \\
\hat{x}_{p2} &= 97 && (100) \\
\hat{Q} &= 23 && (25)
\end{aligned}
\qquad (4.25)
$$

with the true values in parentheses.

We can see that $\hat{a}_1$ and $\hat{a}_2$ are within 10% of the true values, and the $\underline{x}_p$ estimates are within 6%, $\hat{b}_0$ is about 30% off, while $\hat{b}_1$ is about 60% in error. The reason for this large error in $\underline{\hat{b}}$ is in the fact that information about $\underline{b}$ can best be extracted by large and fast deviations in u. Since u is a relatively smooth time function, the estimation of $\underline{b}$ becomes difficult. For better estimates a larger data sample should be tried. As will be shown this error will not greatly affect the prediction capability of the identified model.

In all cases the Fletcher-Powell estimates are considerably better than the ones found from the iterative approach discussed in the previous section. In order to more accurately compare the two results, the iterative approach should be tested more thoroughly though.

In the next section we compare the performance in predicting of the model here identified, against the one given the true system parameters.

IV.5 On-Line Forecasting and Updating - Evaluation via Simulation:

IV.5.1 Discussion:

In this section we shall test the capabilities of the forecasting technique via simulations.

A model of the form hypothesized for the load is simulated as in the previous section. The on-line prediction and updating scheme described in chapter II is tested with data from the simulated model, assuming first that the model parameters are exactly known, and second, using the same data, but the parameters as identified by the Fletcher-Powell algorithm. This will give us an indication of how well the identified model predicts, and what the sensitivity of the prediction performance is with respect to parameter errors.

In each case we show typical 24 hour predictions with the associated standard deviation. The effect of closed loop updating is illustrated and discussed. Anomaly detection is tested together with the capability of the scheme to correct itself after the anomaly if allowed to run in closed loop configuration.

The effect of initial state and $\underline{x}_p$ uncertainty is also discussed.

The estimation-prediction algorithms described in chapter II and used throughout are programmed as listed in Appendix B.

IV.5.2 Prediction with Exact Simulated Model:

The model used to simulate data is the same as described in the previous section IV.4, by equations 4.19 through 4.23.

Typical prediction curves are shown for different times together with the actual error and the predicted standard deviation (see Fig. 17, 18). Of particular interest is the error curve between predicted and actual simulated load data and the standard deviation envelope around the error. The standard deviation is smallest at the one-hour prediction, with

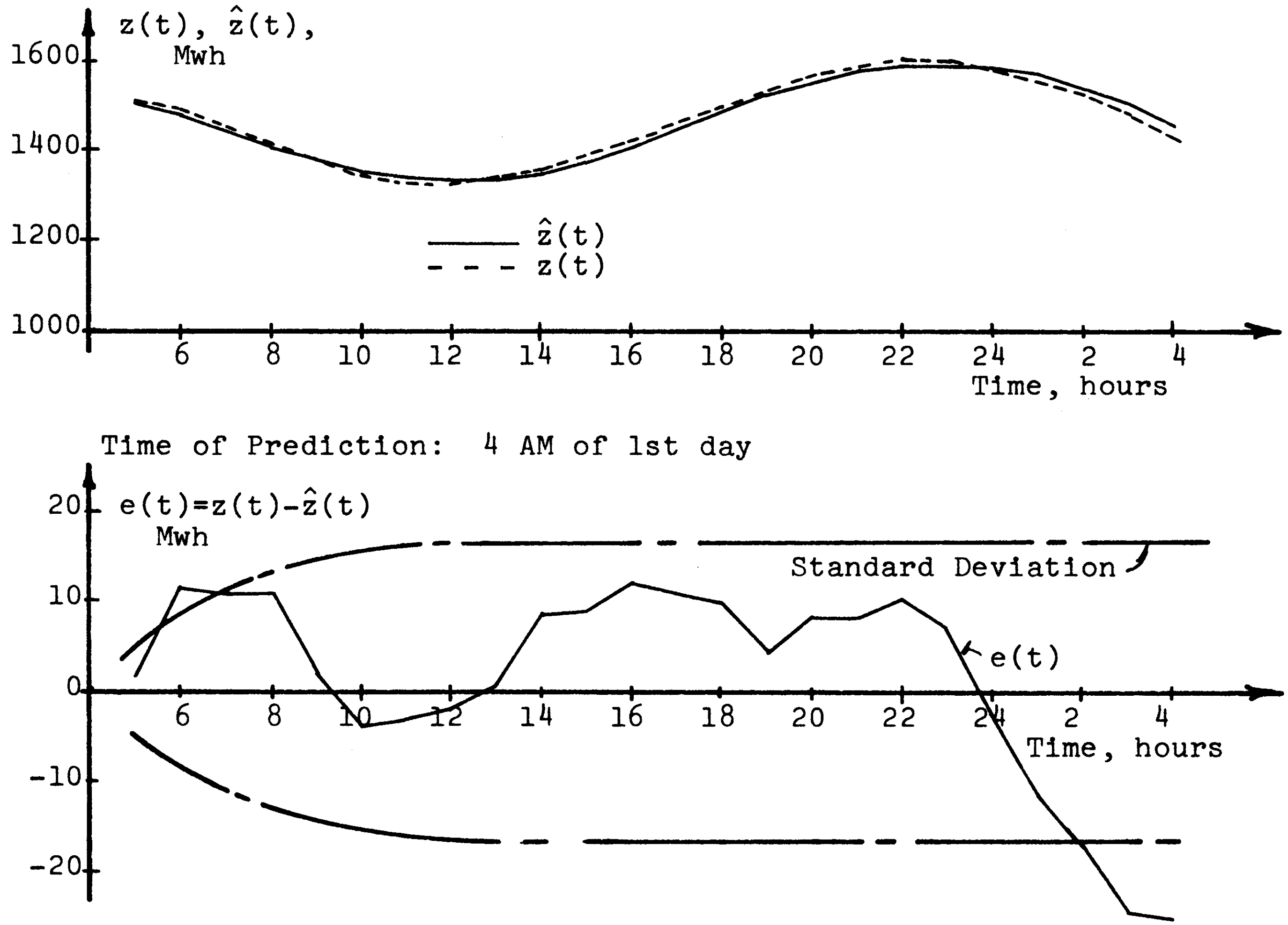


Fig. 17: <u>Prediction of Simulated Load</u> - <u>Assuming Model Parameters Exactly Known</u>,I.

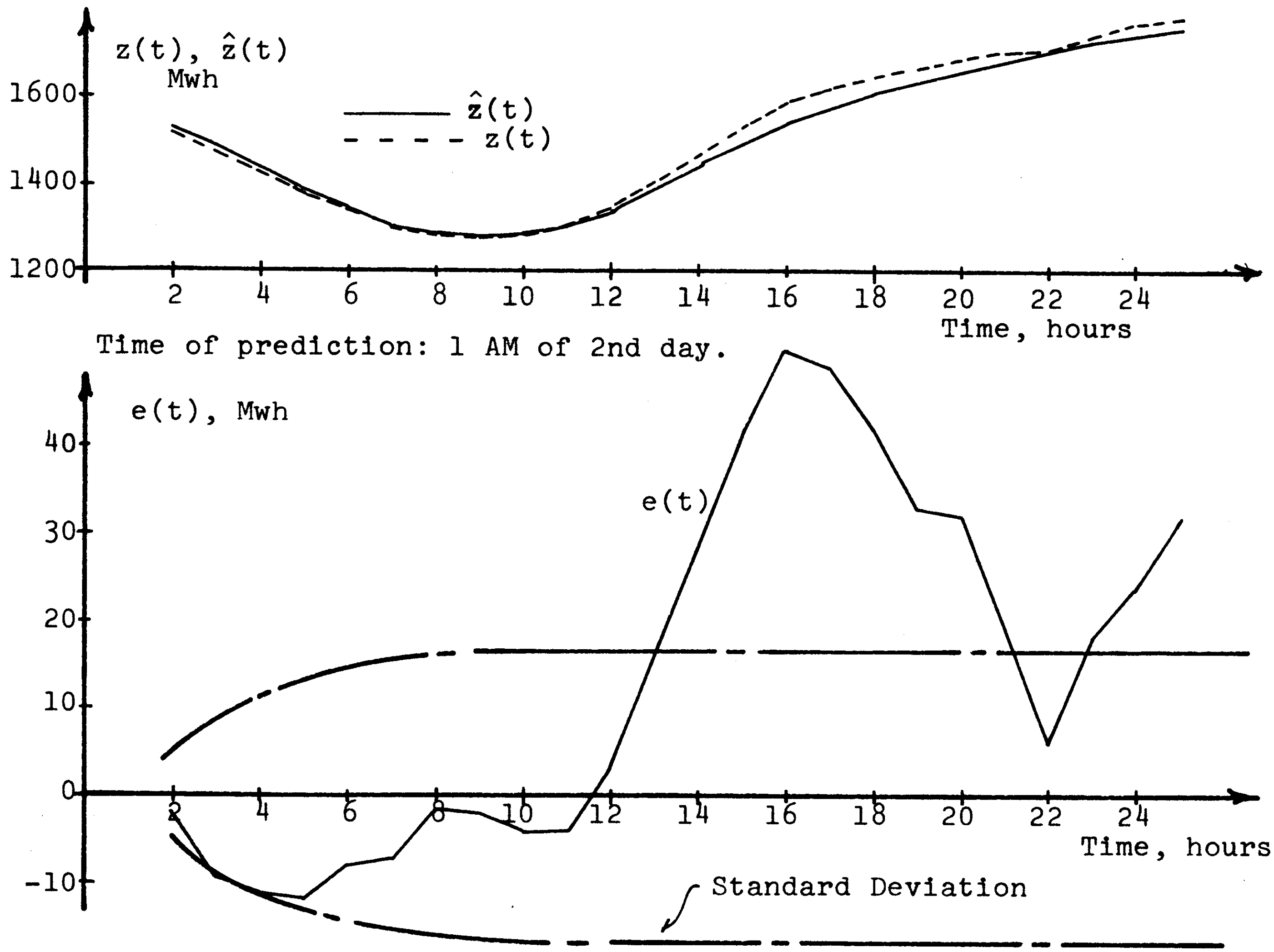


Fig. 18: Prediction of Simulated Load - Assuming Model Parameters Exactly Known, II.

the value of 5. The uncertainty rises as we predict farther into the future, reaching a steady state of 16.8. Figure 18 depicts the probabilistic nature of the model. Although in this case the error exceeds the one standard deviation band considerably, this is entirely possible, although unlikely. Closer analysis shows that the particular noise sequence used in simulating the load data for that particular time period, is unusually large, thus explaining the result. Figure 17 shows a more likely prediction error curve.

In the closed loop configuration, after every hour, a new observation is made which updates the forecast. In the example of Fig. 17 and 18, this means that the standard deviation curve would be shifted to the right by one hour, leaving the prediction uncertainty from the particular hour the same as before.

If we discontinued updating then the standard deviation curve would remain fixed at that point. Eventually we would reach the steady state portion of the curve. In this case the prediction error would have a standard deviation of 16.8 for all prediction times.

In Fig. 19 we have shown the propagation of the 24-hour prediction error with time. As expected, this is not a white process, since a little calculation of the prediction equations show that this error is the result of passing w through the y model without the u input. On the average though the standard deviation is as predicted.

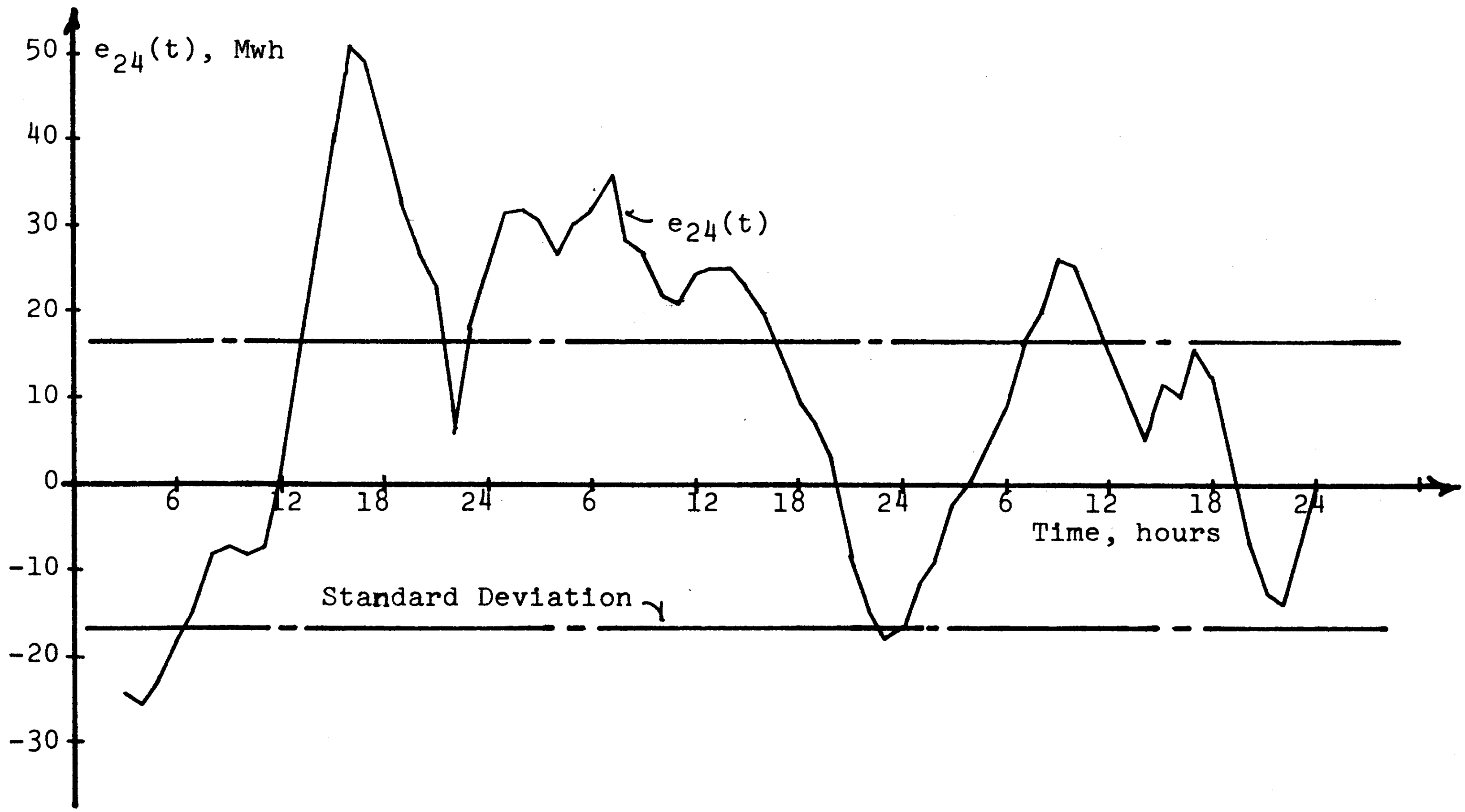


Fig. 19: 24-Step Prediction Error for Simulated Load.

IV.5.3 Anomaly Detection and Self-Adjustment:

As discussed in III.11, anomaly detection is effected by observing the one-hour prediction and testing for statistically unlikely deviations.

Fig. 20 shows a plot vs time of one-step prediction errors in the closed loop configuration. A large disturbance is artificially introduced into the actual data , drastically disturbing the error process. The duration of the disturbance is 5 hours and as shown in the figure, normality returns two hours later. This is so since in our case the state of the model can be exactly determined from the estimation equations in two iterations, given error-free observations.*

The anomaly can quickly be detected from the graph, after one or two steps, depending on the decision level chosen, as discussed in III.11. Of considerable importance here too is the ability of the closed loop scheme to correct itself after the disturbance has died down.

IV.5.4 Prediction with Initial State Uncertainty:

Due to the fact that the y model is asymptotically stable, the effect of errors in the initial guess of the state of this system, $\hat{z}(0)$, is rapidly attenuated with time. In general

* In the more general case of coloured w in the model or observation noise, we can't estimate the state perfectly and the effect of the disturbance would only disappear according to the filter time constant decay.

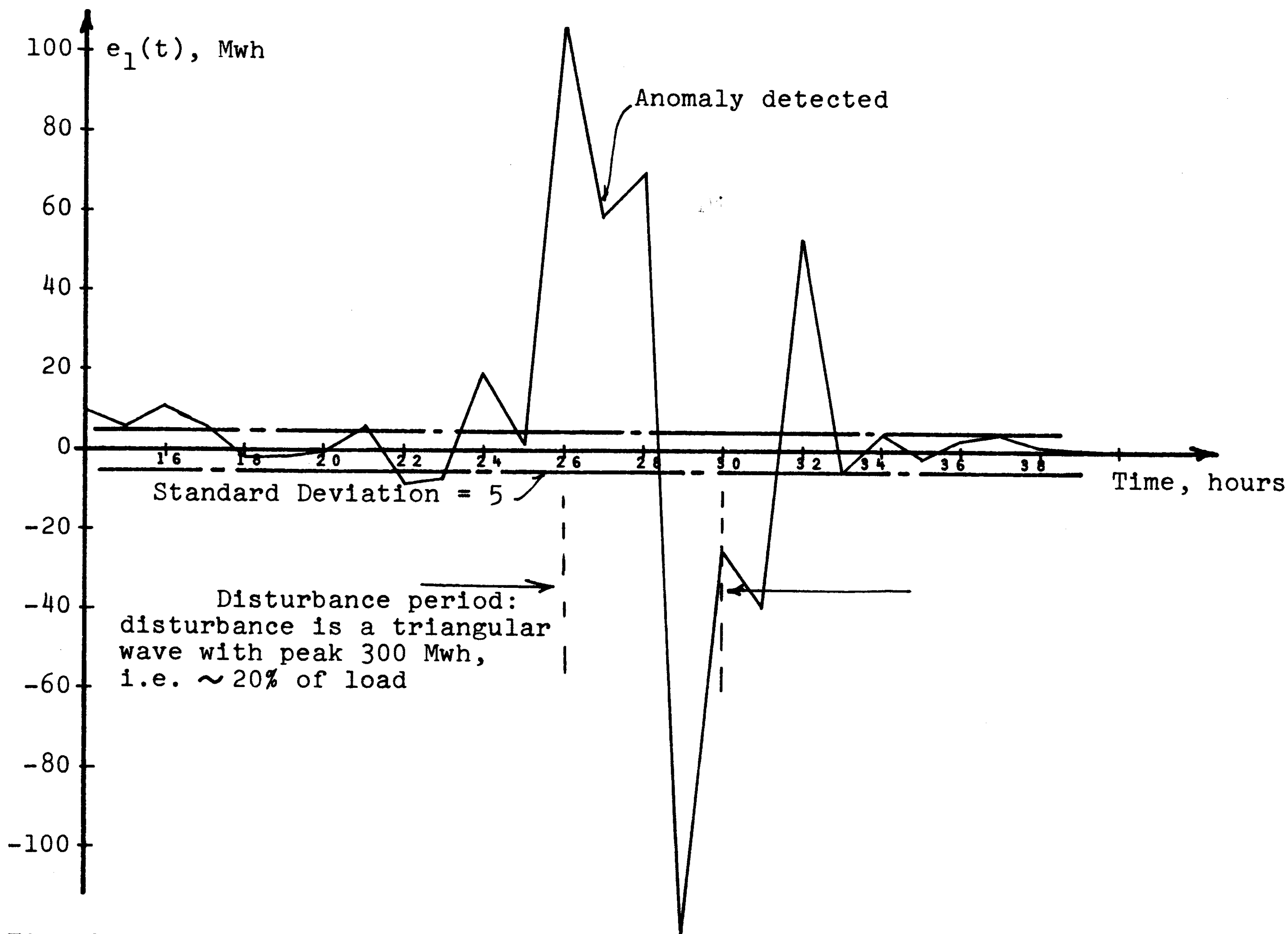


Fig. 20: One-Step Prediction Error - Anomaly Detection - Self Correction; Parameters exactly Known.

this attenuation depends on the time constants of the system. For the special model, as we have simulated here, this effect disappears after two steps (the order of the system). An example of this is seen in the previous section where a disturbance is introduced which essentially creates an error in the system state estimate, therefore disturbing the ability of the model to predict. This disappears two steps after the disturbance ends.

In the case of initial state errors we can introduce a measure of the confidence in this estimate into the covariance equations. This measure is defined by the $\underline{S}(1)$ matrix in our covariance equations of chapter II. Propagation through time together with the value of Q defines the confidence levels in the predictions. At first, if the lack of confidence in the initial state is large, we make $\underline{S}(1)$ correspondingly large. This effect makes the initial uncertainty in the prediction quite large as it should be. As new observations are made the effect of errors in the initial state are attenuated and corrected, so that more confidence in the predictions should be present. This is indeed the case as the effect of $\underline{S}(1)$ in the covariance equations dies down, leaving only that effect introduced by the system noise, Q.

In the special model being considered, the variance of the one-step error starts at 10^9 and at the next step, i.e. after two observations, equals 25=Q. This is as expected, for in our special case two observations are sufficient to exactly determine the state. The one-step prediction uncertainty is thus that introduced by w only.

For the case of coloured noise or when $\underline{x}_p$ is considered uncertain, the confidence levels decay to their steady state much more slowly.

IV.5.5 <u>Prediction with Uncertainty in $\underline{x}_p$ - Linear Estimation of $\underline{x}_p$</u>:

In chapter II we showed that the value of $\underline{x}_p$ could be theoretically determined by linear estimation, i.e. by assuming that $\underline{x}_p$ is part of the state to be estimated.

In this subsection we consider the same simulated model as before. We assume that $\underline{x}_p$ is not exactly known, but that the <u>remaining model parameters are</u>. The estimation-prediction algorithms are run with an initial estimate of $\underline{x}_p$ as given below,

$$\hat{\underline{x}}_p(0) = \begin{pmatrix} 1450 \\ 95 \\ 95 \end{pmatrix} \qquad (4.26)$$

where actually,

$$\underline{x}_p = \begin{pmatrix} 1500 \\ 100 \\ 100 \end{pmatrix} \qquad (4.27)$$

We make the assumption that this initial estimate is subject to some uncertainty, as it should be.

That is,

$$\underline{S}_p(0) = E\{[\underline{x}_p-\hat{\underline{x}}_p(0)][\underline{x}_p-\hat{\underline{x}}_p(0)]^T\} \tag{4.28}$$

$$= \begin{bmatrix} S_{p1} & 0 & 0 \\ 0 & S_{p2} & 0 \\ 0 & 0 & S_{p3} \end{bmatrix} \tag{4.29}$$

where different values of S_{pi} were considered. Consider the case where,

$$S_{p1} = S_{p2} = S_{p3} = 10^6 \tag{4.30}$$

We now present a number of graphs which describe the capability of identifying $\underline{x}_p$ by linear estimation, in addition to testing the prediction scheme when $\underline{x}_p$ is uncertain and is being simultaneously identified. Only simulated data is used here.

From Fig. 21 we see that the uncertainty for the $\underline{x}_p$ parameters shown decays very fast at first, settling down to a very slow asymptotic decay, approximately of the order of 1/n, where n is the number of observations from time zero.

Figure 22 presents the actual propagation of the error in $\underline{x}_p$. Comparison with Fig. 21 shows this error to be compatible with the standard deviation. Near the end of the estimation curve, the error reduces very slowly as more data is observed. This however implies that the ability of the model to predict and model is relatively insensitive to this error as is indicated in Fig. 23, 24 and 25.

In Fig. 23 we see the one-step prediction error propagated with time. At the start this error is quite large, but

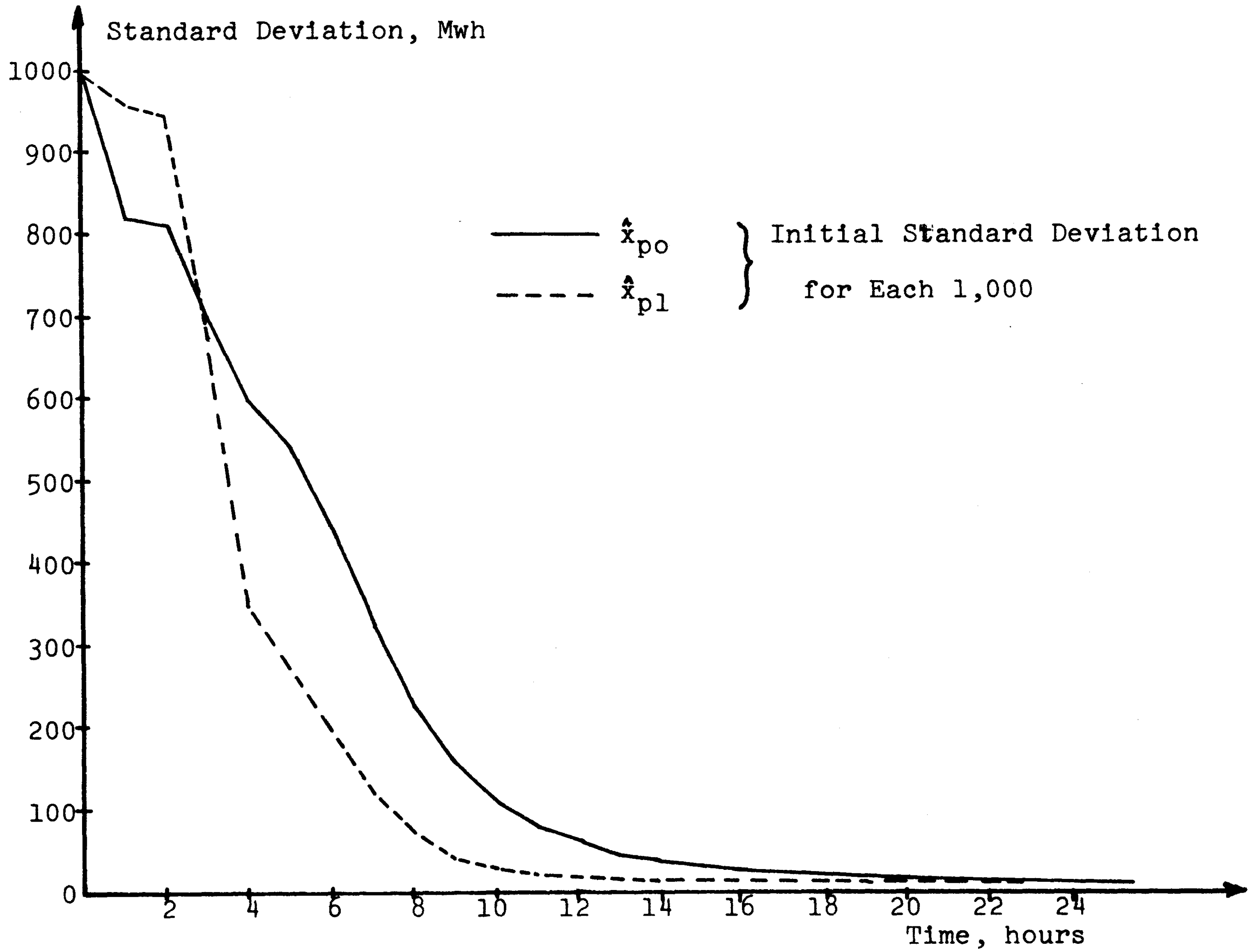


Fig. 21: Linear Estimation of x_p - Propagation of Standard Deviation of $\hat{x}_p$ Components $\hat{x}_{po}$ and $\hat{x}_{pl}$ - Simulated Data.

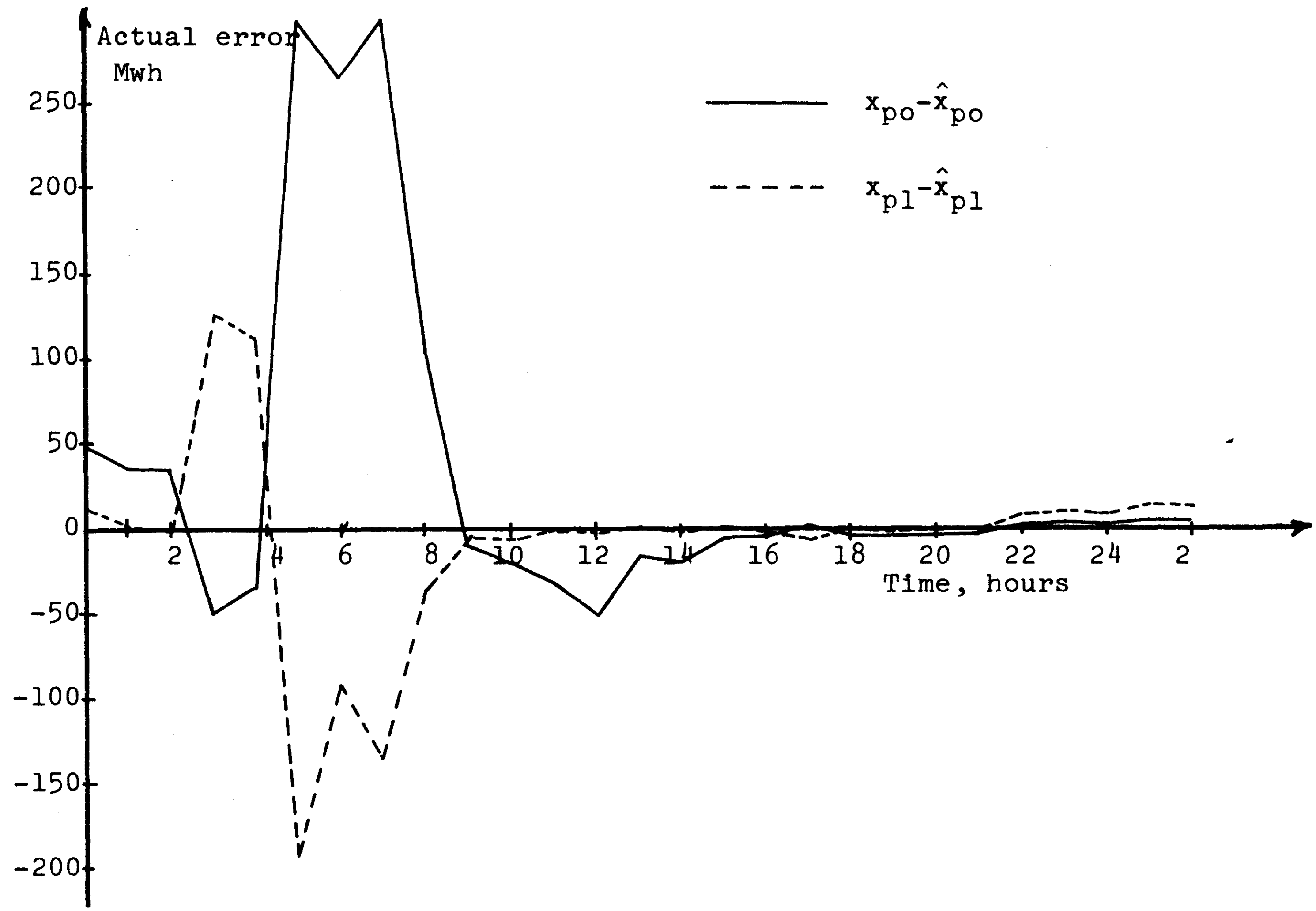


Fig. 22: Actual Error in $\hat{x}_{po}$ and $\hat{x}_{p1}$ from their True Values - Simulated Data.

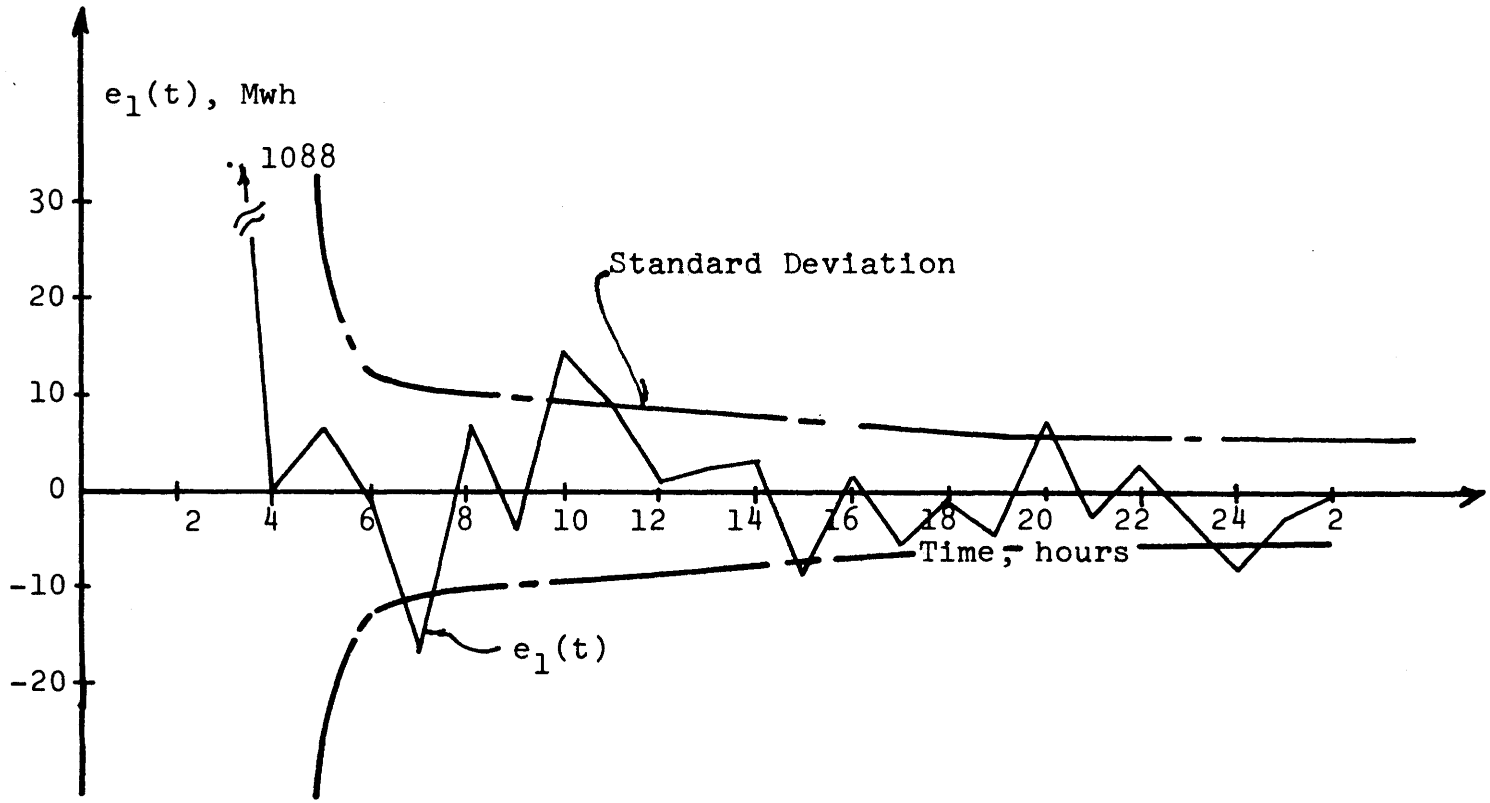


Fig. 23: One-Step Prediction Error with Large Initial uncertainty in x_p - Simulated Data.

always compatible with its standard deviation. With time both the error and the standard deviation decay to much smaller values. In particular the standard deviation appears to reach a steady state of 5 which is the same value as in the case where $\underline{x}_p$ is exactly known. This is why we say that even though $\hat{\underline{x}}_p$ is in error, this error does not greatly affect the model response.

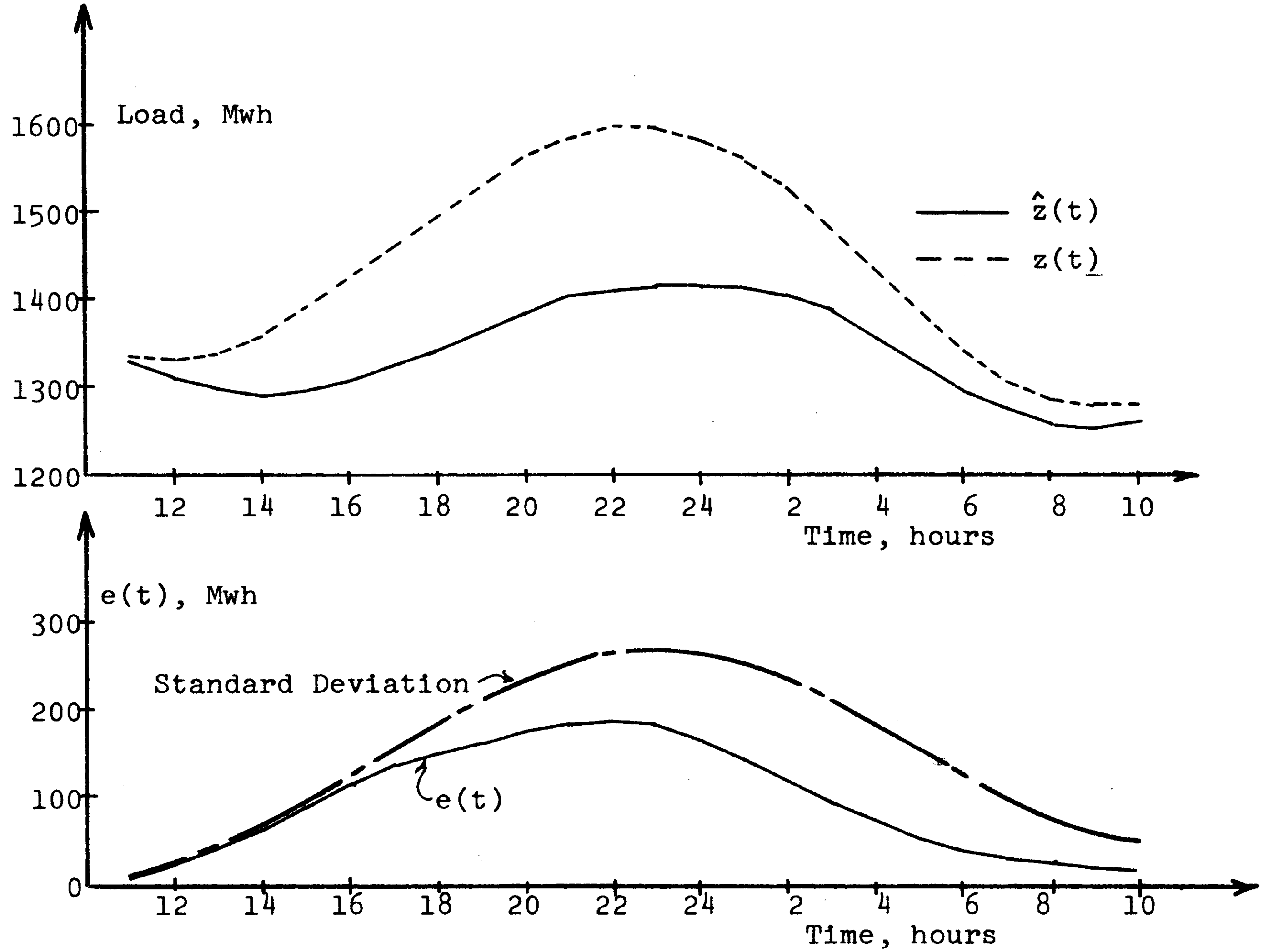
Figure 24 shows a typical prediction curve when the error in $\hat{\underline{x}}_p$ is large enough to significantly affect the ability of the model to predict as accurately as in the case when $\underline{x}_p$ is exactly known. The important point to notice is that even then, the prediction error is compatible with the standard deviation, that is our confidence in the prediction.

Figure 25 is an equivalent case of a prediction curve, except that $\underline{x}_p$ is now near its true value, and we have for practical purposes reached steady state. The standard deviation is much smaller now and the load prediction error curve well within the expected levels.

IV.5.6 Prediction with Identified Model Parameters:

So far we have discussed and evaluated with simulated data the ability to predict load with exact parameter knowledge, detect anomalous behaviour, estimate $\underline{x}_p$, as well as predict load and simultaneously estimate $\underline{x}_p$ when it is uncertain.

Now we discuss the case where the model used for prediction is that found from the identification scheme. This model will have errors not only in $\underline{x}_p$, but also in the parameters of the y model.



Time of Prediction: 10 AM of first day.

Fig. 24: Prediction of Simulated Load, Assuming Model Parameters Uncertain-Before Steady State.

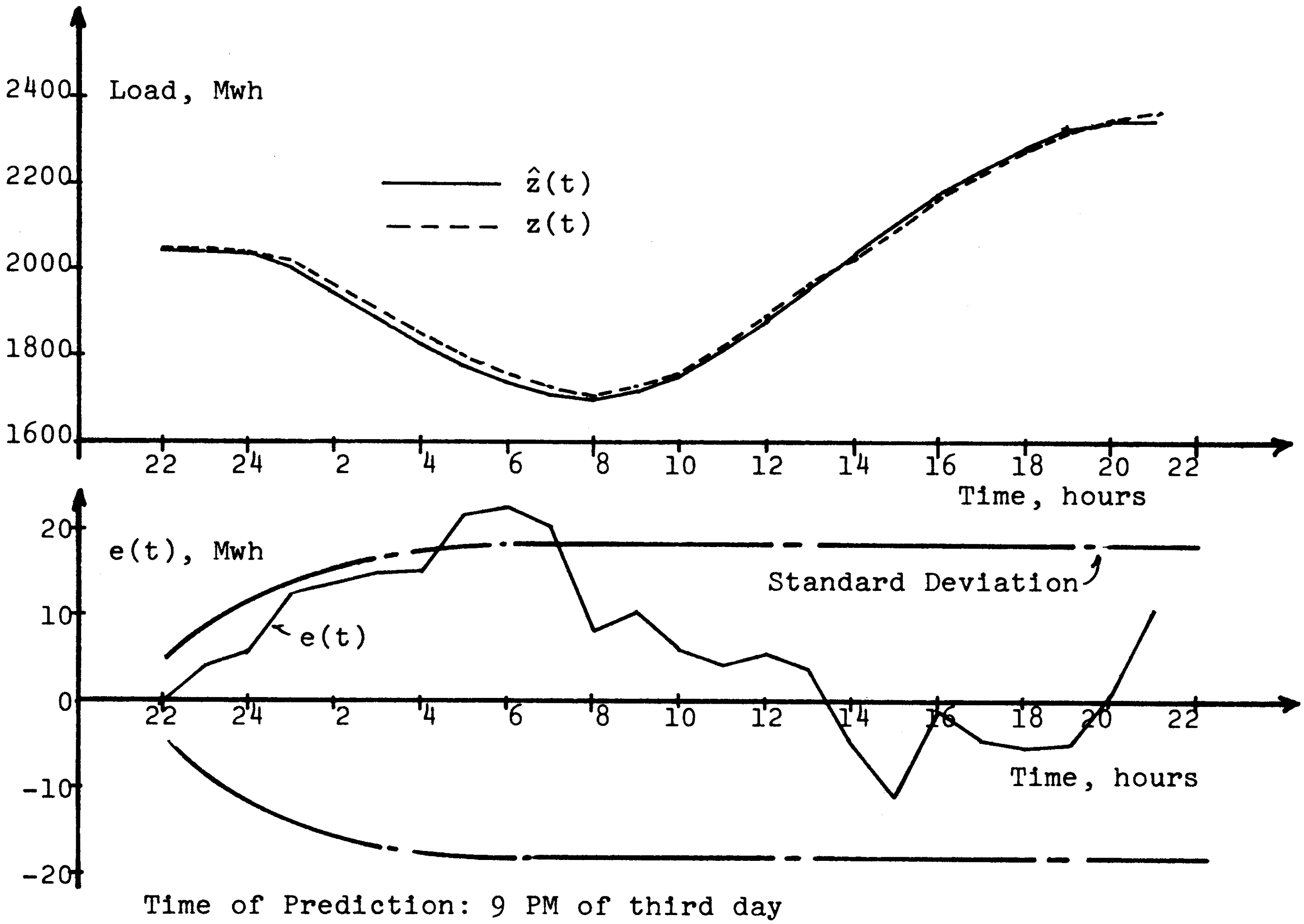


Fig. 25: <u>Prediction of Simulated Load, Assuming Model Parameters Uncertain - in Steady State</u>.

We can proceed exactly as before, but most of the results would be quite similar. Here we simply consider the case where the parameter values obtained in section IV.4.2 by the Fletcher-Powell algorithms are assumed free of uncertainty, and run some prediction tests with them. Uncertainty in $\underline{x}_p$ could be treated as in IV.5.5 by artificially assuming an initial $\underline{x}_p$ uncertainty. Uncertainty in the y model parameters cannot be easily taken into account as these parameters appear non-linearly in the model.

As can be seen from Fig. 26 and 27, the identified model predicts quite well. The prediction error in Fig. 27 is a bit too large, but this can be attributed to two things primarily. First, the identified Q was lower than the true value, thus decreasing the prediction standard deviation, and second, as explained for Fig. 18, the noise sequence for that particular period is unusually although possibly high.

Nevertheless prediction tests made throughout a 4 day simulated data period as exemplified in Fig. 26 and 27 give us sufficient confidence in the ability of the identified model to predict.

Anomaly detection tests based on the one-hour prediction errors were carried out as in the case with the exact parameters yielding very similar results.

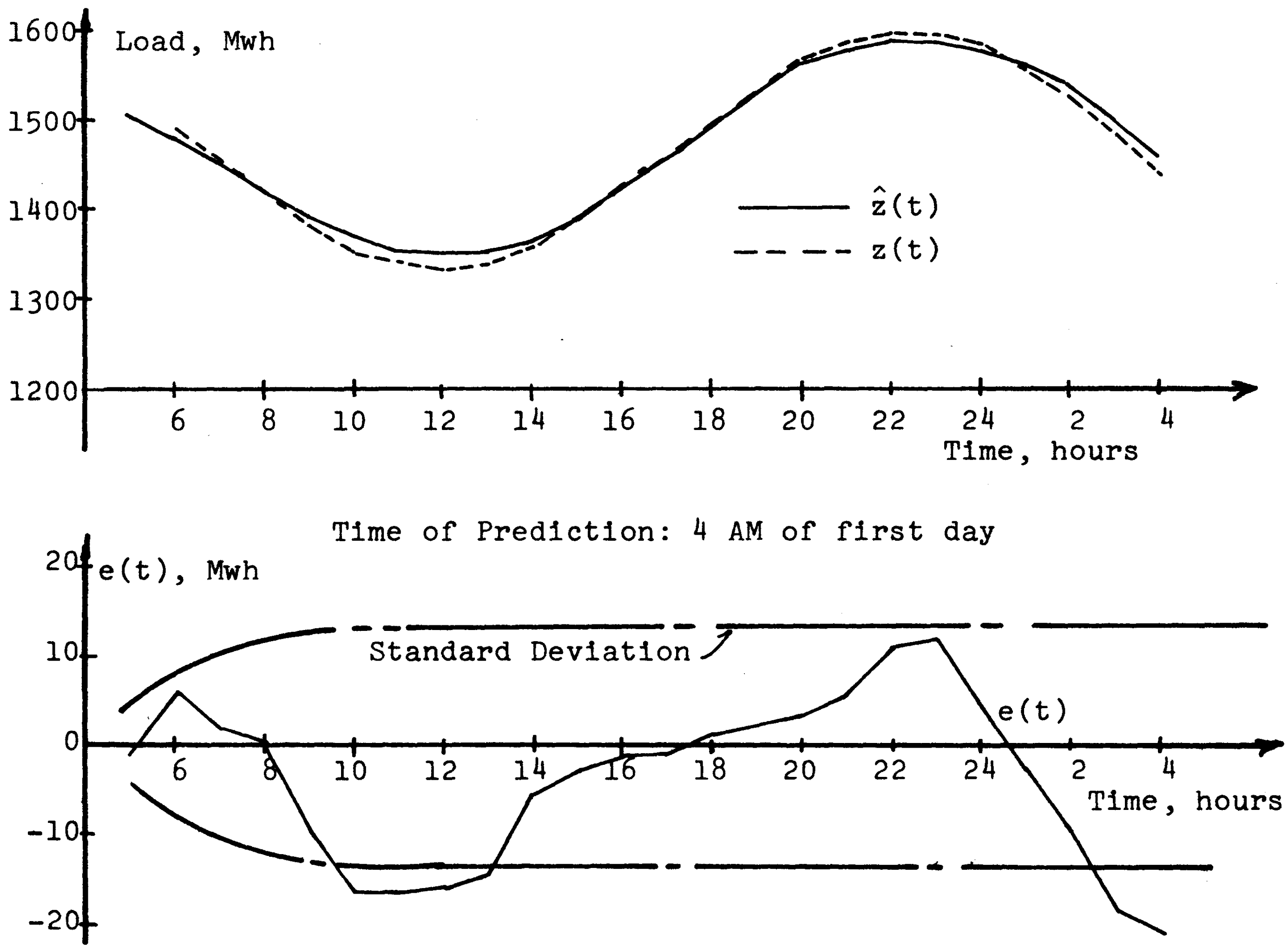


Fig. 26: Prediction of Simulated Load, Using Identified Model Parameters - Case I.

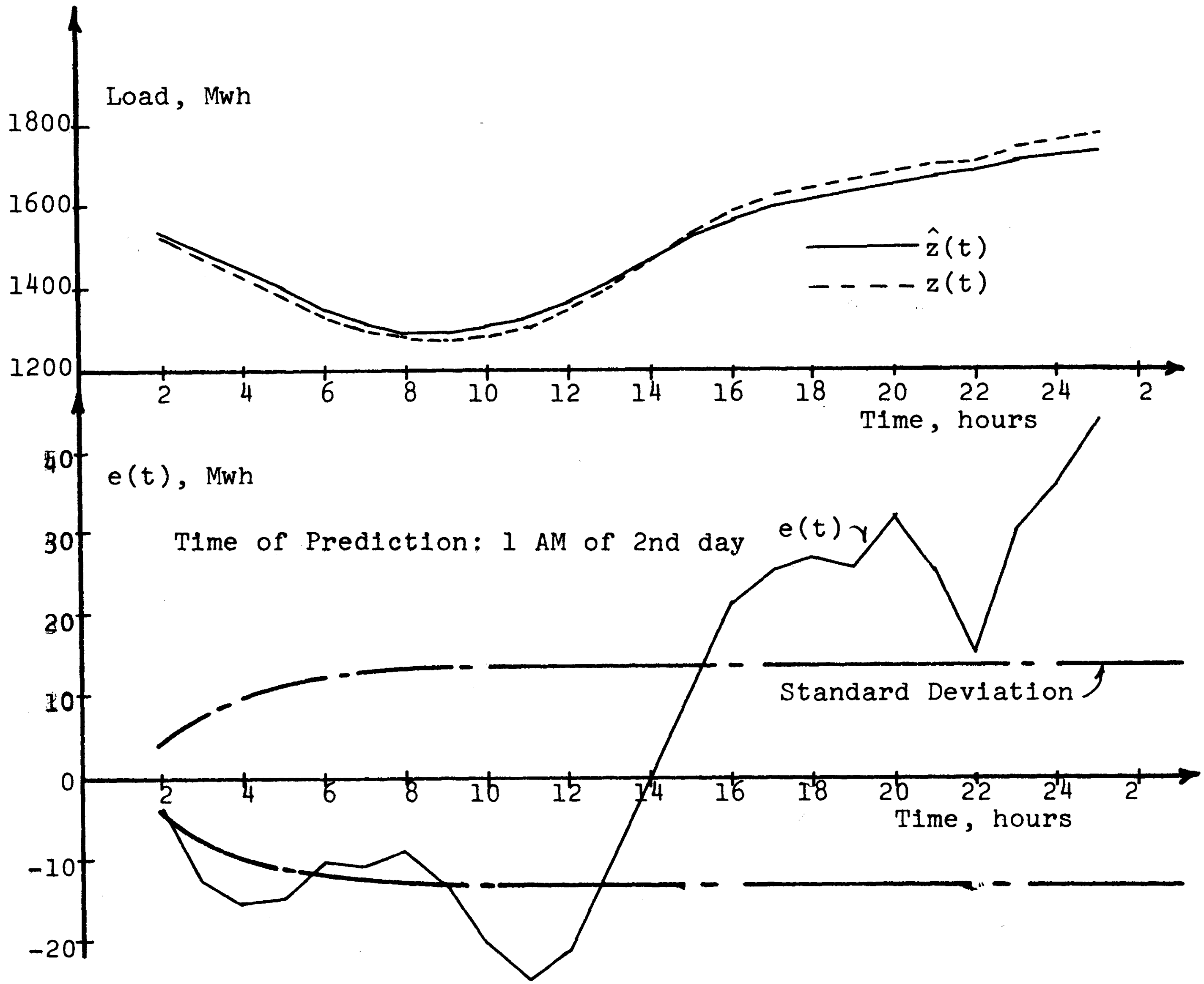


Fig. 27: Prediction of Simulated Load, Using Identified Model Parameters - Case II.

IV.6 Evaluation of Identification & Prediction Techniques with Real Data:

IV.6.1 Preliminary Discussion:

In this section we discuss the tests carried out with real data to verify the hypothesized model's capability to predict load. Since no apriori knowledge of the value of the model parameters is available, the main test as to the validity of the model is whether it can predict load under varying conditions, in compatibility with the associated standard deviation. In other words, we want to predict future loads based on future inputs using a model evaluated from past loads and past inputs.

In addition to the parameter values we are also uncertain as to the dimension of the model. From section IV.2 we can estimate the number of parameters in $\underline{x}_p$, that is $2n_p+1$, however the order of the y model cannot be so easily estimated.

The data considered in this section is still from the Cleveland Electric Illuminating Company, for the year 1969.

Tests (both modelling and prediction) are carried out for the months of July and August, considered warm months, and January, considered a cold month. These tests hopefully show the ability of the proposed model to perform as suggested, and provides a sufficient basis to carry out more detailed and specific studies.

IV.6.2 Estimation of Model Order:

The number of harmonics, n_p, in the component y_p can be estimated from the shape and duration of some of the reoccuring load peaks. More reliable estimates can be obtained from the least squares study described in section IV.2. On this this basis we decided on a value of n_p of at least 5.

The next step was the estimation of the order of the y model. Here we have various dimensions we can play with, the number of a_i parameters, n, the number of b_i parameters, m, and the number of d_i parameters.

In this study we only consider the special model with white noise describing the model uncertainty. We need not thus concern ourselves with the number or values of d_i. This assumption, as will be seen later on, is not critical in getting a good model.

In estimating the values of n and m we tended toward values as small as possible. That is if a low order model performed within reasonable levels, this was preferred. The reasons were manyfold. First, higher order models require more computer time to identify. Second, due to our data record limitations, the identification of many parameters becomes more difficult. Thirdly, the physical process we are trying to model, y, is basically a simple one, which should be describable by a relatively low number of parameters. Finally, if we can describe load behaviour with accuracy with a simple model, then don't complicate life more than necessary.

Keeping in mind the above recommendations, we first fixed n_p at 5 and attempted the identification of various models with different values of n and m. The data considered in this preliminary study was chosen from the month of July, as conditions in this month are probably more stationary than during the spring and fall when load behaviour varies more rapidly over the weeks.

The load data used was provided by the Cleveland Electric Illuminating Company, as described in IV.1. The input data, u, is determined from hourly values of recorded actual and normal temperatures in degrees Farenheit, as explained in chapter II.

We are interested in obtaining 3 models, one for the weekdays and one each for Saturday and Sunday. For the weekday model we considered a data record of 3 weeks from Monday, July 14th, 1969, to Friday, August 1st, 1969, excluding weekends. The record length consisted therefore of 15 days. More on this choice will be discussed later on.

The first attempt was to fit a model with n=m=1 and n_p=5. The behaviour of this model would tell us whether the dynamic behaviour hypothesized is indeed needed. Thus if a simpler static model were sufficient we would expect the identified value of a_1 to be near zero.

Identification using the Fletcher-Powell algorithm as discussed yields the following set of parameter estimates,

$$\hat{a}_1 = 0.99$$

$$\hat{b}_0 = 0.77$$

$$\hat{\underline{x}}_p = \begin{bmatrix} 1931 \\ -279 \\ -128 \\ -16 \\ -13 \\ -15 \\ -362 \\ 85 \\ 23 \\ -49 \\ -18 \end{bmatrix} \qquad (4.31)$$

$$\hat{Q} = 1016$$

The first conclusion to be made here is that there is a very strong support for a dynamical model. The parameter $\hat{a}_1$ is, in this case, the time constant of the model. Since this is equal to 0.99 the memory of this model is quite long.

In Fig. 28 we have shown the result of using the model evaluated above to estimate and predict. Here we have predicted into the first week of August. It can be seen that the standard deviation of the prediction error starts at approximately 32, the square root of $\hat{Q}$=1016. This is a reasonable confidence level, however, due to the large memory of the system, the steady state standard deviation is almost 225, or about 12% of the peak load. This is clearly a very uncertain prediction.

It is interesting to note that the prediction error is well within the standard deviation throughout the 4 days prediction time. In particular, for the first two days the error is very reasonable, deviating only near the end. This

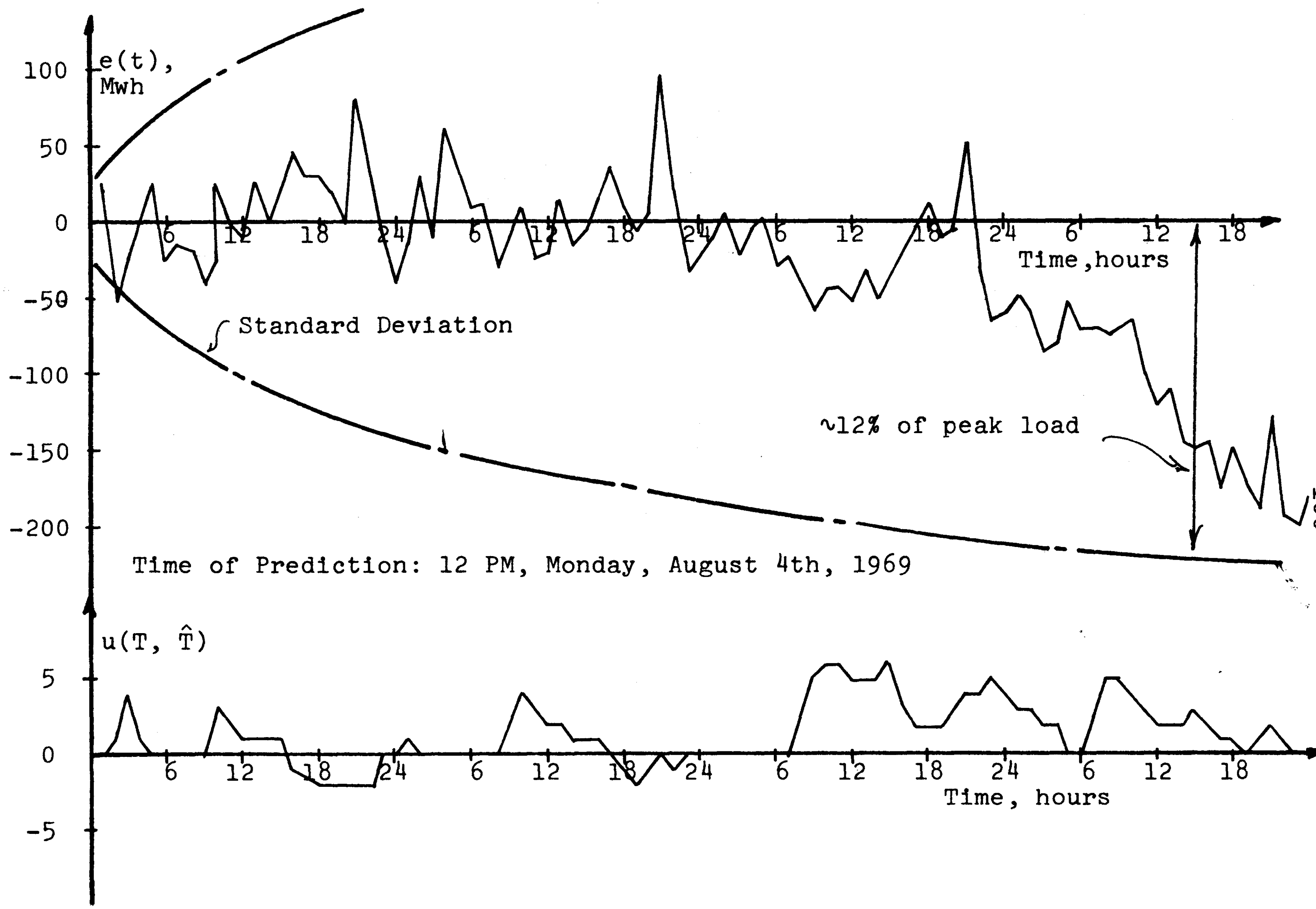


Fig. 28: August Load Forecast and Temperature Deviation - Model $n_p=5$, $n=m=1$.

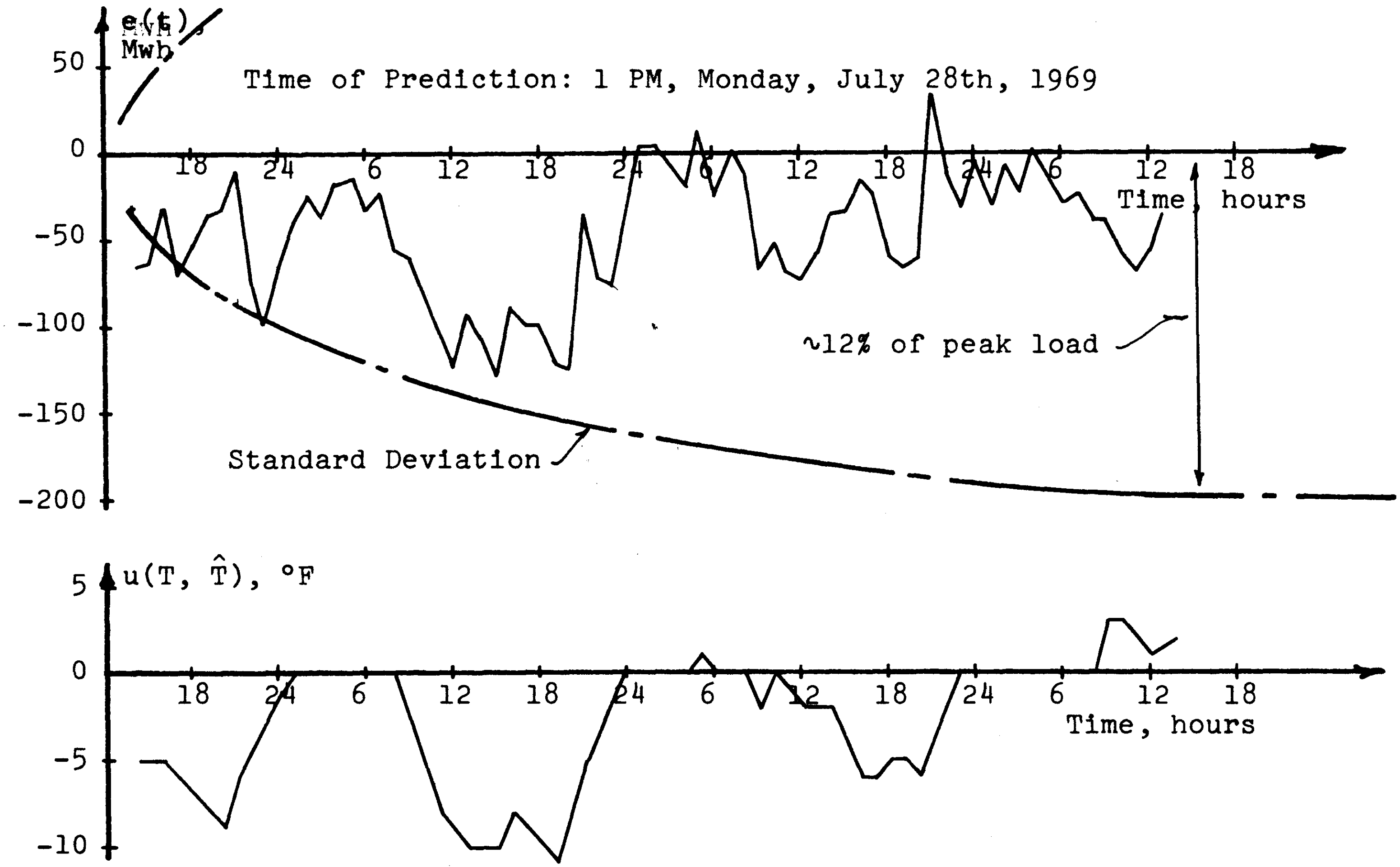


Fig. 29: July Load Forecast Error and Temperature Deviation - Model n_p=5, n=m=1.

seems to imply that for the first two days when the system inputs, u, are relatively small and random, i.e. average days, the prediction is primarily determined by the periodic component. However when the inputs deviate from normal, the errors become much more significant although still compatible with the standard deviation.

To further dramatize this effect we have shown in Fig. 29, again using the same model, the prediction into the last 4 days of the last week in July, where temperature deviations are quite significant. We can see that when u deviates from zero, the prediction error becomes quite significant.

The above results indicate that perhaps a more complex temperature dependent model is needed to eliminate the high sensitivity of prediction errors to large inputs, u. Following our hypothesized model, we increase the order of the y system to n=m=2 and still keep $n_p=5$.

The data record used to identify this model was taken to be the 3 weeks in July starting Monday July 7th. Two of these weeks overlap with those of the previous example so we have a good basis for comparison. Running the identification program resulted in the following model,

$$\begin{aligned} \hat{a}_1 &= 0.65 \\ \hat{a}_2 &= 0.27 \\ \hat{b}_0 &= 4.2 \\ \hat{b}_1 &= -1.4 \end{aligned} \qquad (4.32)$$

$$\hat{\underline{x}}_p = \begin{bmatrix} 1709 \\ -284 \\ -127 \\ -14 \\ -12 \\ -17 \\ -350 \\ 84 \\ 21 \\ -48 \\ -17 \end{bmatrix} \qquad (4.32) \text{ cont'd}$$

$$\hat{Q} = 889$$

We can easily calculate the characteristic roots of this system to be 0.93 and -0.28. Both are less than one in magnitude making the y model asymptotically stable as expected. No oscillatory behaviour is provided by y, a desirable result since we want y_p to take care of this. The larger time constant 0.93, gives an idea of the time it would take the y system response to go to zero. In 24 hours the response would decay to less than 0.1 of the original value. This behaviour is quite within the expected response of the load to temperature effects.

In Fig. 30 we show the result of predicting with the model obtained above. This prediction is made during the week after the three weeks used to evaluate the model, as would be done in real time.

The actual prediction error is plotted vs future time, together with the associated standard deviation. On the same graph we have shown a plot of the prediction error had we not included the predicted value of y in the total load model. This shows the effect on the model behaviour of the temperature

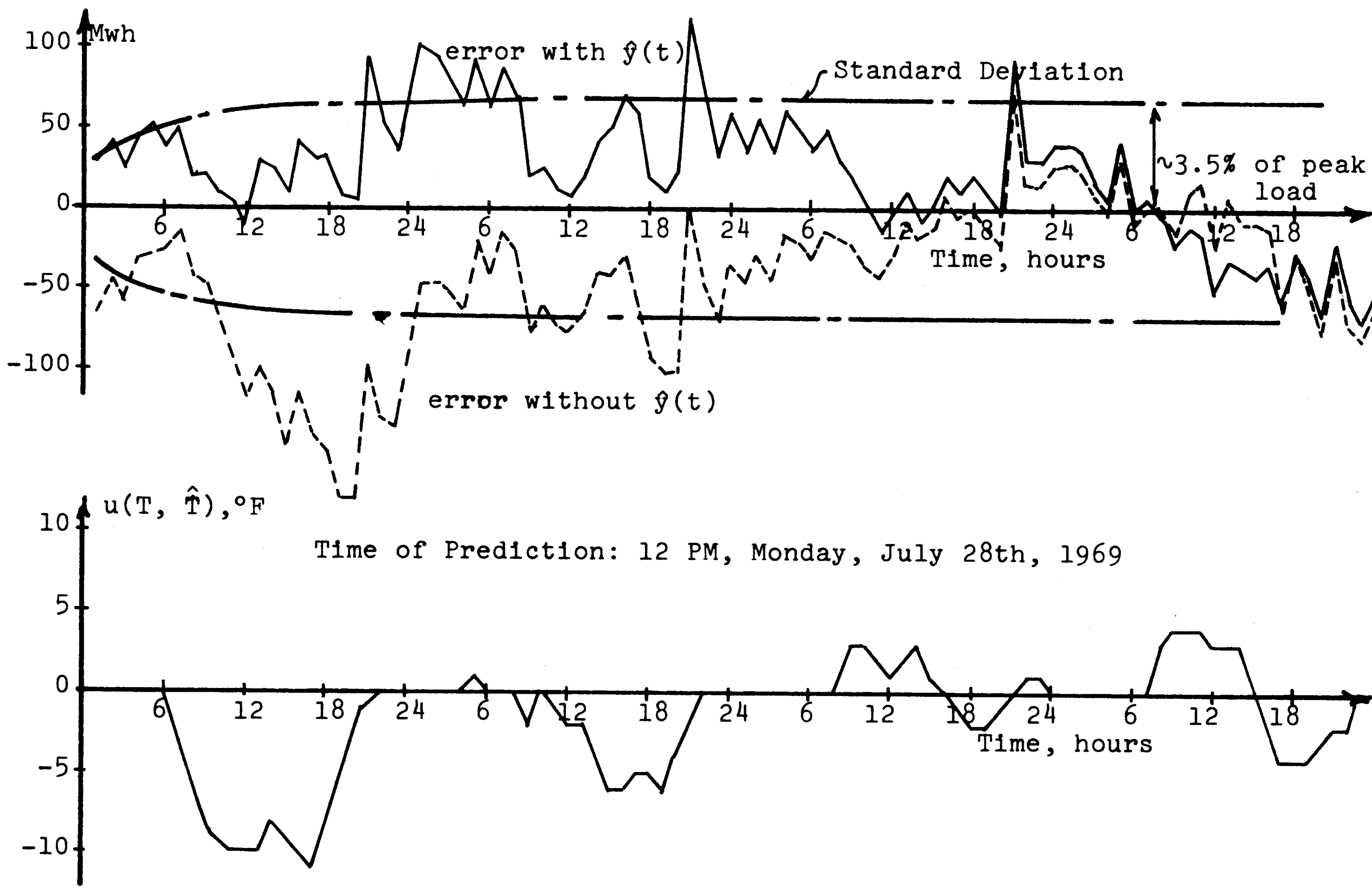


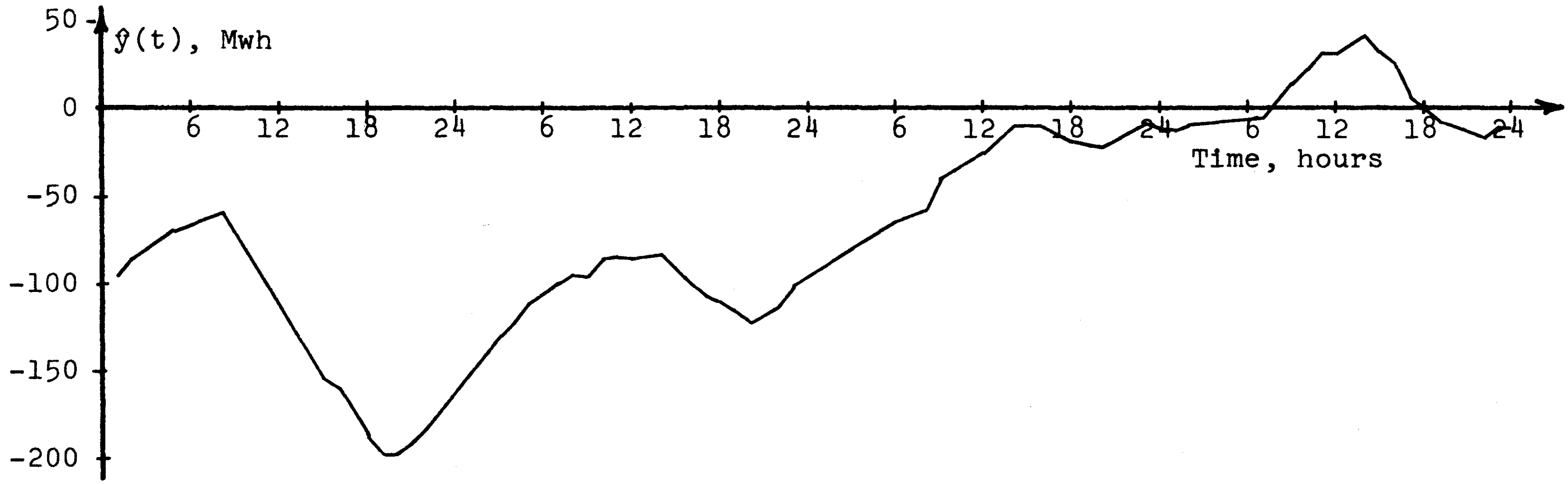
Fig. 30: July Load Forecast Error and Temperature Deviation - Model $n_p=5$, $n=m=2$.

dependent uncertain component, $\hat{y}$. Below we have included the temperature deviation inputs, u, used in this prediction. We notice that when u is large (unusually cool), the effect of $\hat{y}$ is to bring the error curve within the standard deviation, whereas when u is not unusually large, the effect of $\hat{y}$ is less significant, as expected.

We should also emphasize the fact that the steady state standard deviation of the prediction error is now only 75 or about 3.5% of the peak load, considerably better than the first order y model in Fig. 29 (n=m=1).

In Fig. 31 we have shown a plot of the predicted residual load, $\hat{y}$, vs future time for the same period as in Fig. 30. This points out the dependence of y on temperature deviations. It also presents a more explicit quantitative measure of the value of y. When $|u|$ is persistently large as in the first part of the graph, so is $\hat{y}$, reaching a value of 10% of the total peak load. The dynamical behaviour is also clearly evident.

So far the behaviour of this model was quite good, the prediction compatible with the predicted standard deviation, the effect of y very significant in making the prediction valid when u was large, the predicted load errors up to 96 hours in the future and their standard deviation a very reasonable 3.5% of peak load in steady state. The next test consisted of checking the one-step prediction errors for whiteness. As discussed in chapter III, this is not only a check for the model's validity, but its verification would allow the detection of anomalous load behaviour.



Time of Prediction: 12 PM, Monday, July 28th, 1969

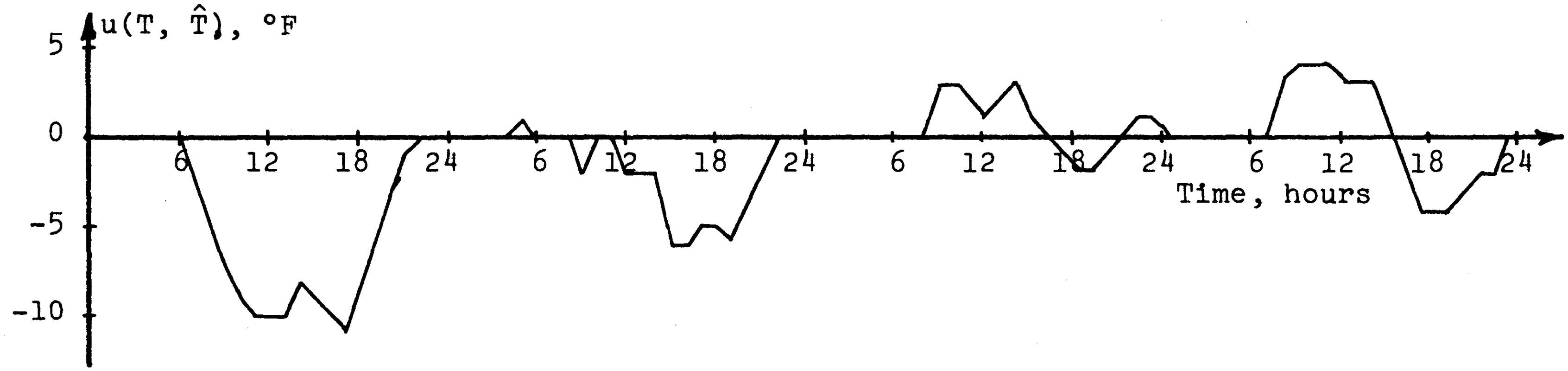


Fig. 31: July Residual Load Forecast and Temperature Deviation - Model $n_p=5$, $n=m=2$.

Since the prediction updating was carried out for the 24 hours of Monday, we have shown in Fig. 32 the one-step prediction errors for each of these 24 hours, and their expected standard deviation. Except for the first 5 or 6 hours, the results are quite "whitish". The initial discrepancy can be attributed to the fact that although the behaviour of the load for Monday is "similar" to that of the other weekdays, during the early hours from Sunday midnight to 6 AM Monday, the load behaviour is quite different, still following the low consumption pattern of the previous Sunday. This fact led us later on to the introduction of a fourth model for Monday, together with one each for Saturday and Sunday, and for the remaining four weekdays.

Now going back to Fig. 30 which shows the load prediction errors, we notice that a periodic peak occurs at approximately 2100 hours, or 9 PM, of each day. This is due to the lighting load which appears as a sharp increase in the daily load curve. The obvious step taken in accordance with our model was to go to a higher n_p value. We then considered for the same data record as before, the model with dimensions $n=m=2$, $n_p=6$.

Identification yields the following set of parameter values,

$$\begin{aligned} \hat{a}_1 &= 0.62 \\ \hat{a}_2 &= 0.30 \\ \hat{b}_0 &= 4.3 \\ \hat{b}_1 &= -1.5 \end{aligned} \qquad (4.33)$$

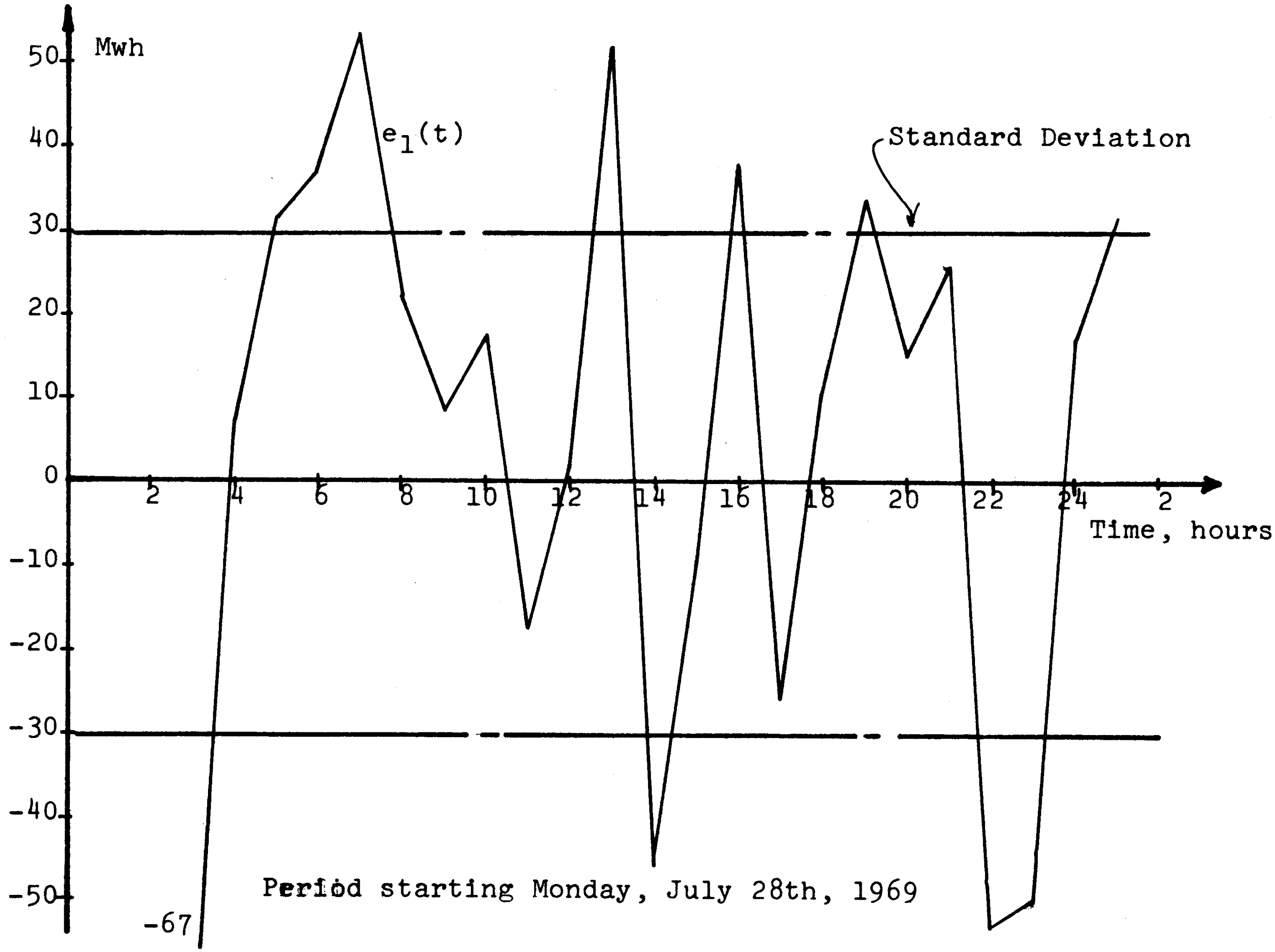


Fig. 32: July One-Step Prediction Error - Model n_p=5, n=m=2.

$$\hat{\underline{x}}_p = \begin{bmatrix} 1716 \\ -284 \\ -127 \\ -13 \\ -12 \\ -17 \\ 10 \\ -350 \\ 84 \\ 21 \\ -48 \\ -18 \\ 3 \end{bmatrix} \qquad (4.33) \text{ cont'd}$$

$$\hat{Q} = 782$$

The system's characteristic roots here are 0.93 and -0.30, almost identical to the model with n=m=2, n_p=5.

Comparison of 4.33 with 4.32 shows the parameters of the y model practically unchanged. The parameters $\hat{\underline{x}}_p$ are unchanged except for $\hat{x}_{po}$ and of course the additional $\hat{x}_{p6}$ and $\hat{x}_{p13}$. The latter are respectively 10 and 3, whereas the change in $\hat{x}_{po}$ is 7.

Not much improvement was expected from these added parameters. The prediction results plotted in Fig. 33 confirm this conclusion. Although the lighting load error peaks are slightly decreased, the overall error stays approximately the same. A perhaps more significant change is that the steady state standard deviation is now only 62, or about 2.5% of the peak load, with the actual errors still compatible with this confidence level.

At this point we were generally satisfied with the performance of the model with n=m=2, n_p=5 or 6, however we investigated the performance of a model with n=3, m=1, n_p=6.

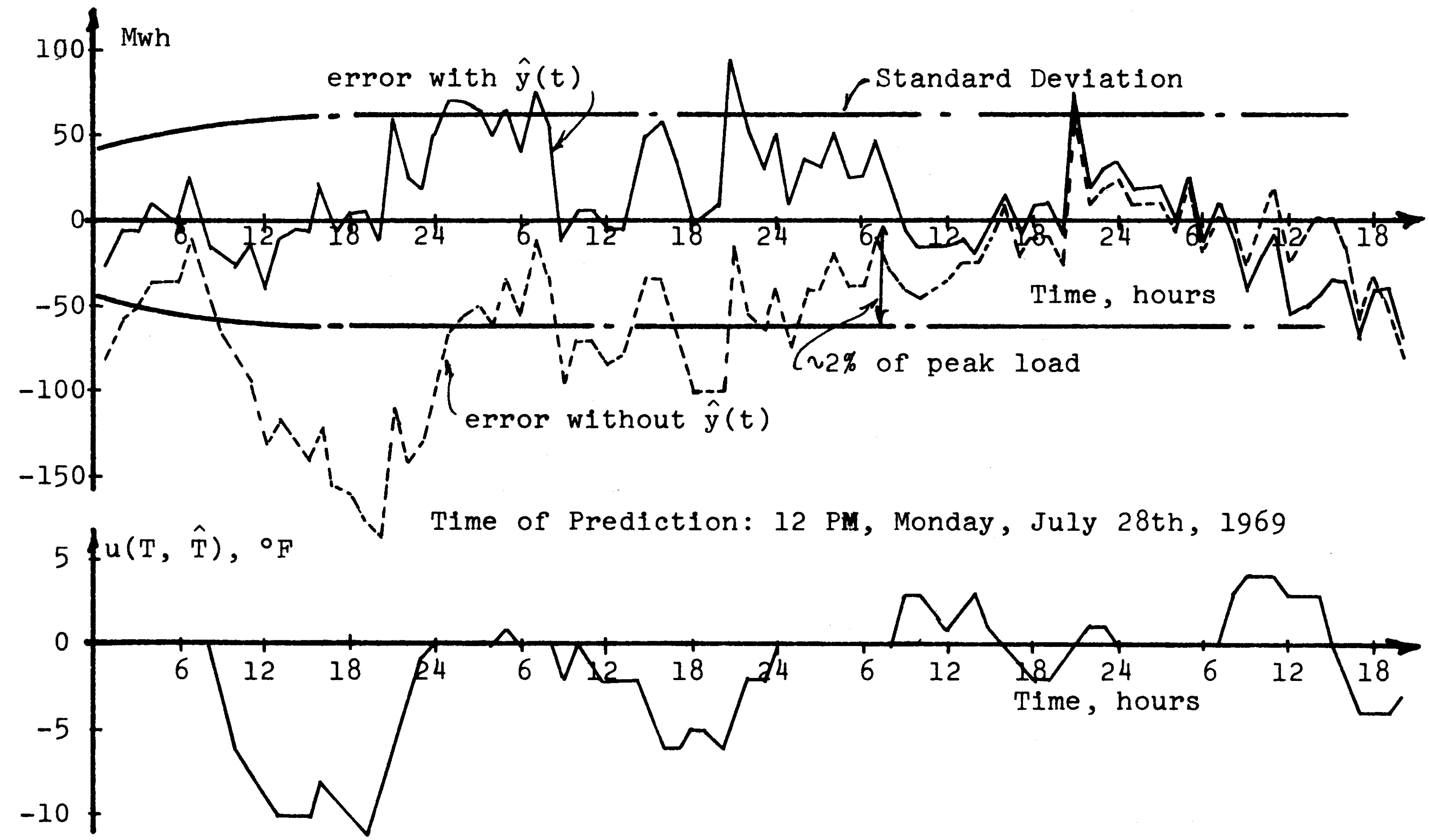


Fig. 33: July Load Forecast Error and Temperature Deviation - Model $n_p=6$, $n=m=2$.

The main change here is that due to n=3, that is increasing the number of system time constants. The fact that m=1 is less significant since the criterion J is relatively insensitive to b_1.

Identification with the same data record again yielded,

$$\hat{a}_1 = 0.46$$

$$\hat{a}_2 = 0.09$$

$$\hat{a}_3 = 0.36$$

$$\hat{b}_0 = 3.6$$

$$\underline{\hat{x}}_p = \begin{pmatrix} 1708 \\ -282 \\ -128 \\ -15 \\ -12 \\ -18 \\ 9 \\ -346 \\ 83 \\ 23 \\ -48 \\ -18 \\ -3 \end{pmatrix} \qquad (4.34)$$

$$\hat{Q} = 700$$

Comparison of 4.34 with 4.33 shows relatively little improvement in the minimum value of J, i.e. 700 against 782. In addition the relatively small value of $\hat{a}_2$ suggests that the previous model may be sufficient. Evaluation of the characteristic roots of the system yields the values of 0.95, and -0.24 ±0.55j. Thus although the dominant time constant has not changed much at 0.95, we now have a pair of complex roots. This is undesirable since it means that the y model is fitting some of the periodic behaviour. Further calculation shows that the complex roots introduce an oscillatory component whose frequency is approximately that which would be provided by a seventh harmonic. This is very reasonable since we still have some small periodic peaks which are not modelled as well.

We attempted to predict with this model as before, yielding consistently poorer results, especially when the system inputs have been large. This indicated that further improvement of the prediction errors would probably not be possible by higher order models, but by introducing other effects as discussed in the Conclusions and Recommendations. This result is consistent with more sophisticated tests for estimating the order of the system. Although the value of J found from the identification scheme is less than for the case with n=2, this does not mean that for other conditions the model fit will be better. In fact it is worse. This could be detected from the small decrease in J (or not significant), when n is increased to 3.

IV.6.3 Further Examples of Modelling and Prediction Capabilities:

In this section we extend the analysis of the above subsection to different seasons of the year.

Preliminary studies showed that load sensitivity to temperature is highest during the warm months.* This fact is illustrated by considering the following two cases:

* For the Cleveland Electric Illuminating Company data. See Fig. 38 and 39.

A) <u>August, 1969</u> -

We first attempted to reidentify the model analyzed in the previous section by eliminating the oldest of the 3 weeks in the data record, and adding the last week of July. The data record was then from Monday, July 14th, to Friday, August 1st, 1969, excluding the weekends. A model with n=m=2 and n_p=5 was attempted. The parameter estimates were,

$$\hat{a}_1 = 0.61$$

$$\hat{a}_2 = 0.23$$

$$\hat{b}_0 = 3.40$$

$$\hat{b}_1 = 0.61$$

$$\underline{\hat{x}}_p = \begin{bmatrix} 1736 \\ -288 \\ -130 \\ -17 \\ -10 \\ -15 \\ -343 \\ 82 \\ 22 \\ -49 \\ -19 \end{bmatrix} \qquad (4.35)$$

$$\hat{Q} = 846$$

Comparison with 4.32, the model parameters based on the previous 3 weeks, shows that the values of 4.35 are quite similar. This is reasonable as load behaviour is not expected to change very rapidly at this time of the year.

This model was used to predict into the first week of August with very similar results to those of Fig. 30. Since this particular week has relatively small temperature deviations, the importance of the y component is less evident than in the

July model previously discussed. Nevertheless the prediction is compatible with the expected error.

We next considered a different model still in the month of August, to emphasize the ability to predict during prolonged heat spells.

Here we implemented the recommendation that a different model be used for Monday due to its early morning low power sonsumption. We thus considered the identification of the weekday model for Tuesday, Wednesday, Thursday and Friday, based on three weeks' observations, that is 12 days' data. The exact dates were Tuesday, August 5th, to Friday, August 22nd, 1969, excluding as explained Mondays, and weekends. The order of the model identified was given by n=m=2, n_p=7 and the resulting parameter values were given by,

$$\hat{a}_1 = 0.52$$

$$\hat{a}_2 = 0.40$$

$$\hat{b}_0 = 3.1$$

$$\hat{b}_1 = -1.0$$

$$\hat{\underline{x}}_p = \begin{pmatrix} 1690 \\ -284 \\ -131 \\ -20 \\ -9 \\ -8 \\ 13 \\ 5 \\ -347 \\ 76 \\ 17 \\ -51 \\ -21 \\ -1 \\ 5 \end{pmatrix} \qquad (4.36)$$

$$\hat{Q} = 437$$

The characteristic roots here are 0.94 and -0.42. We attempted to increase the order of n_p from 6 to 7 in this example to try and fit the sharp periodic lighting load peaks.

In Fig. 34 and 35 we have shown the application of the model to predict into the week following the three used for its identification again as would be done on-line. The results are very satisfactory.

In Fig. 34 we show a 3-day prediction curve at 9 AM of Tuesday, August 26th, 1969. The actual and predicted load are shown together with the prediction error and standard deviation. The latter two are quite compatible with each other and the top graph shows the relatively small error between actual and predicted load.

In Fig. 35 we show the effect of temperature deviations on the predicted value of y. The horizontal line describes the steady state standard deviation of the prediction error and is put there to indicate the importance of the y component. As can be seen at the beginning of the prediction we have very warm weather (in the 90's) making the value of y correspondingly high. The following lower temperature deviations result in a decay of the large value of y due to the previous day's heat wave. In the latter part of the prediction it warms up again creating a new increase in y.

In Fig. 36 and 37 we have shown similar graphs for a 3-day prediction again made later on in the day, that is at 8 PM, after hourly updating throughout the in-between hours.

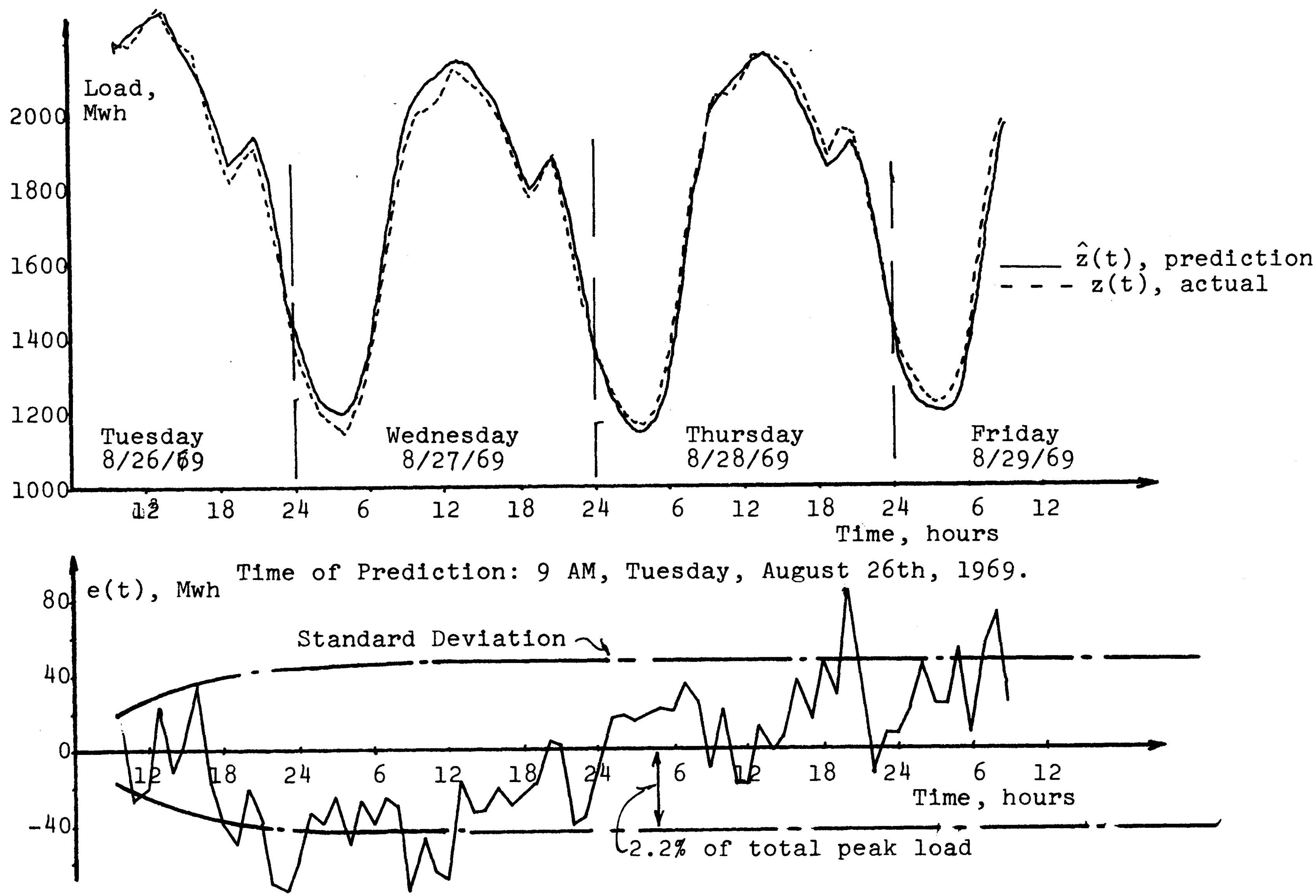


Fig. 34: August Load Forecast and Prediction Error, I - Model $n_p=7$, n=m=2.

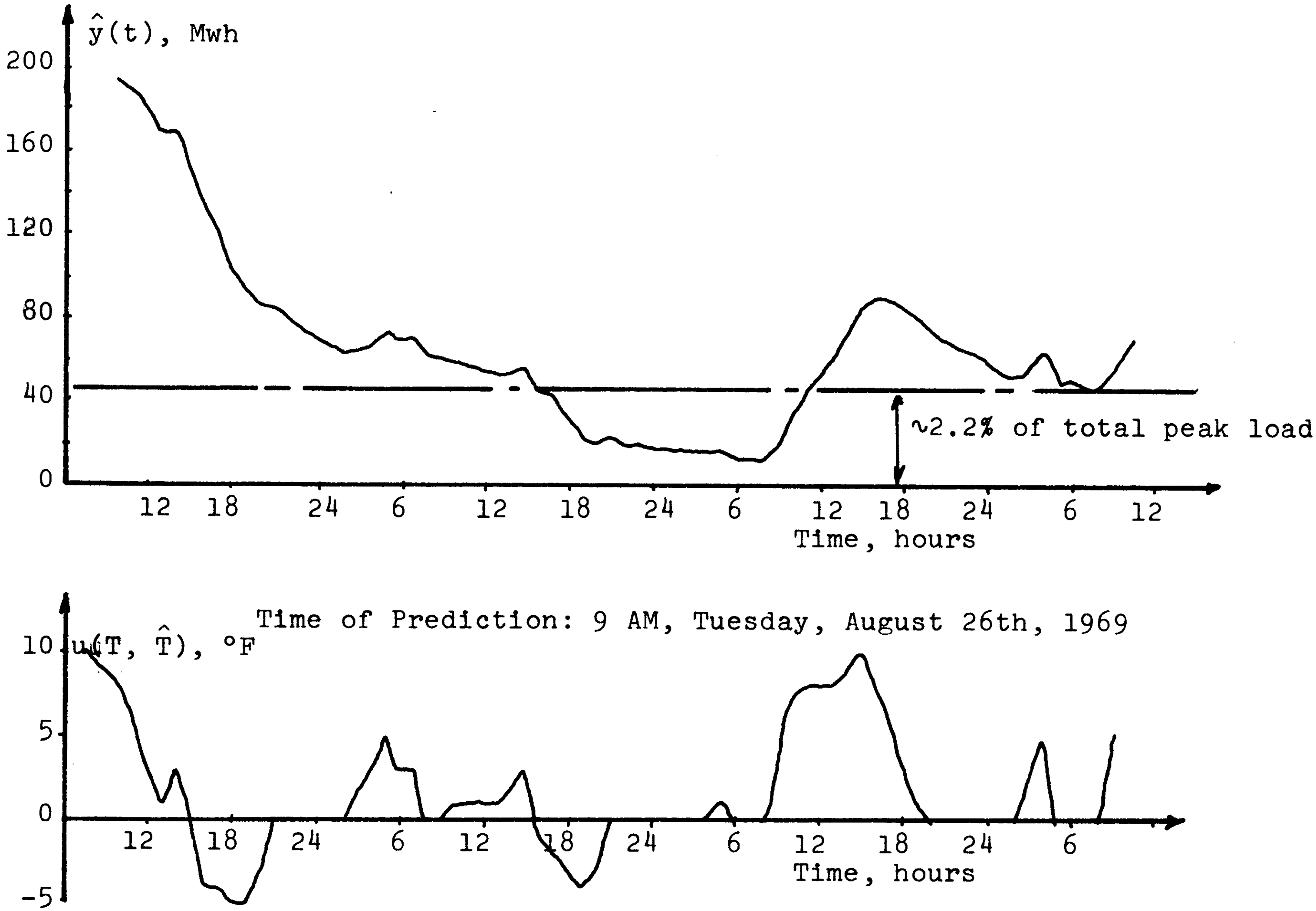


Fig. 35: August Residual Load Forecast and Temperature Deviation, I - Model $n_p=7$, $n=m=2$.

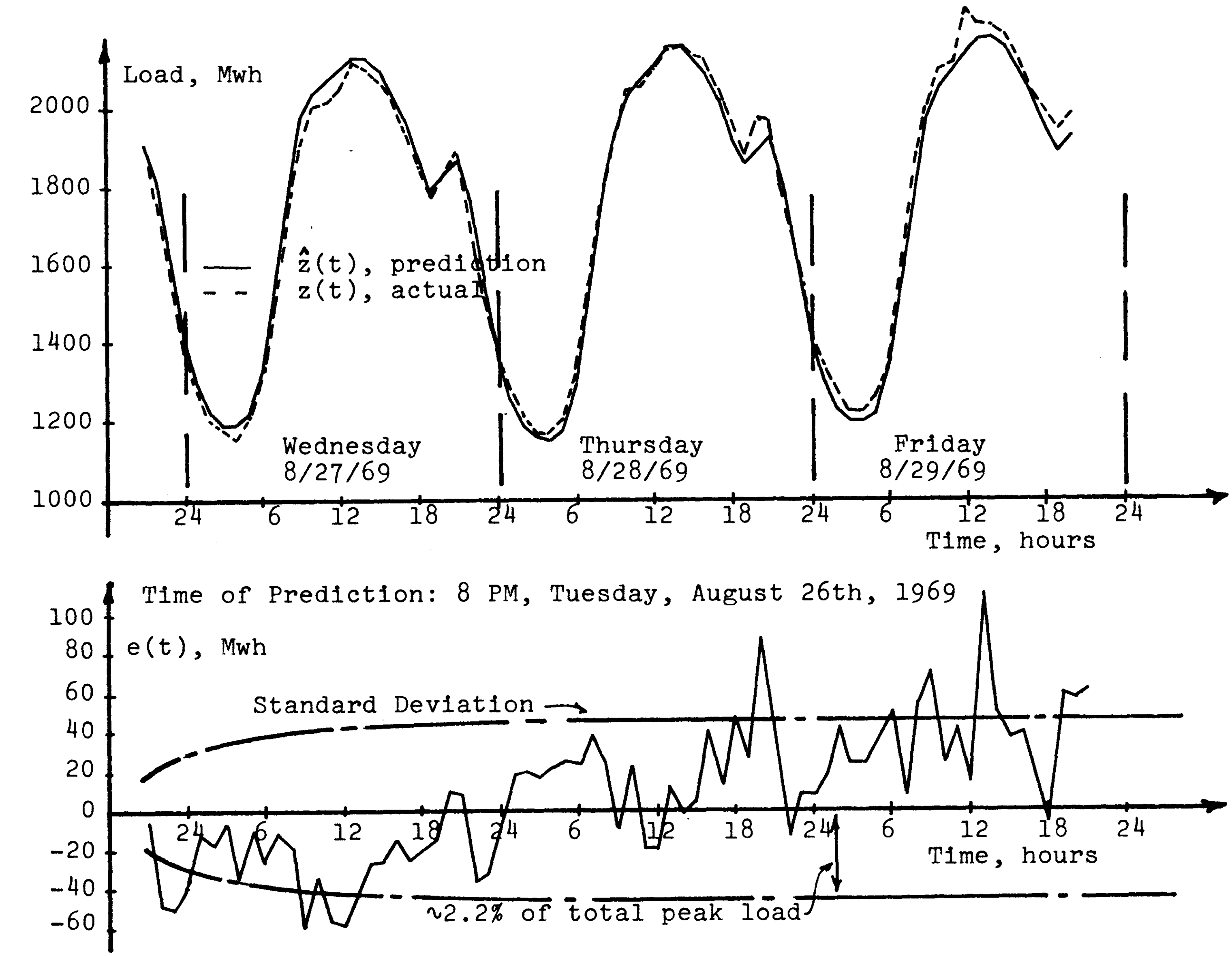
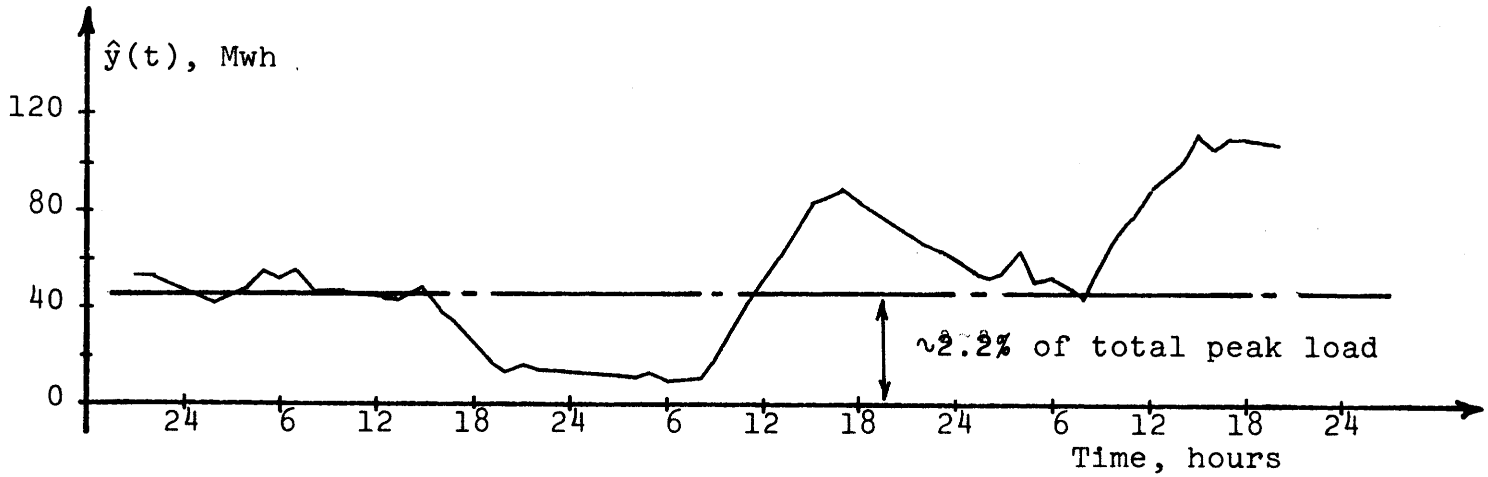


Fig. 36: August Load Forecast and Prediction Error, II - Model n_p=7, n=m=2.



Time of Prediction: 8 PM, Tuesday, August 26th, 1969

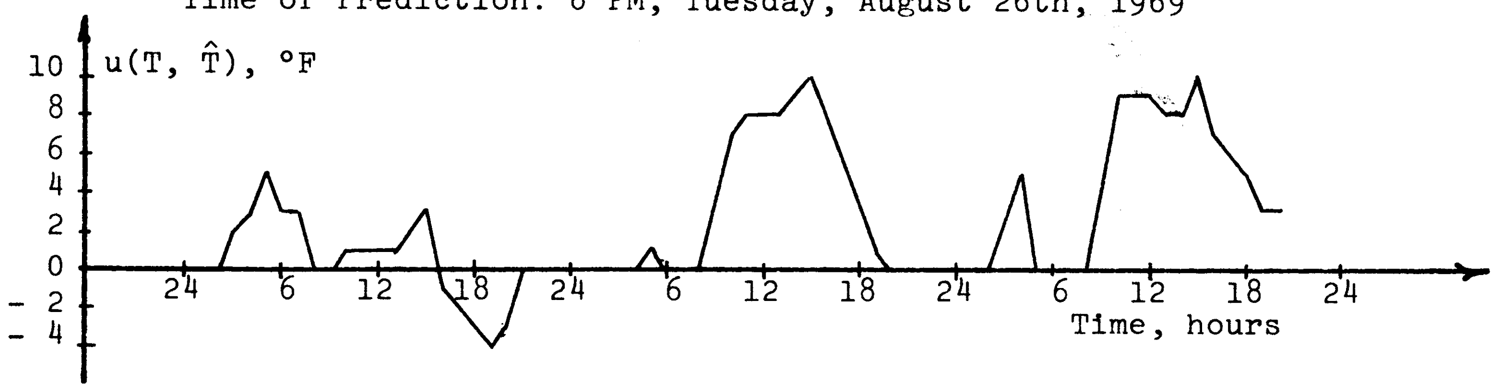


Fig. 37: August Residual Load Forecast and Temperature Deviation, II - Model $n_p=7$, $n=m=2$.

The prediction errors are again compatible with the standard deviation, the largest error occuring at 1 PM of the 3rd day, being approximately 2.5 standard deviations. It must be emphasized that for shorter prediction times the standard deviation is considerably less than in the steady state as seen from Fig. 34 and 36.

The use of n_p=7 instead of 5 or 6, as before, improves the fit of the sharp lighting peaks around 8 PM, however it may have fitted noise as well, resulting in a very low value of $\hat{Q}$. This was responsible for the lower than usual prediction error standard deviation. It is the author's opinion that a lower value of n_p (e.g. 6) with the corresponding higher error standard deviation is a more trustworthy model in the sense that less of the uncertain or random load behaviour is fitted to the deterministic part of the model.

B) <u>January, 1969</u> -

In Fig. 38 and 39 we have illustrated the sensitivity of load to temperature. Fig. 38 shows three Thursdays in July and August relatively close together in the calendar, for a warm, average and cool day. The extremes differ from each other by about 200 Mwh whereas their average temperatures differ by 14°F. In Fig. 39 we show an equivalent case in the month of January. Here the load curves differ by about 100 Mwh while the extreme average temperatures differ by about 32°F. Apparently, in this particular system, electric heat has not been emphasized as much as in other areas of the country (see Fig. 4).

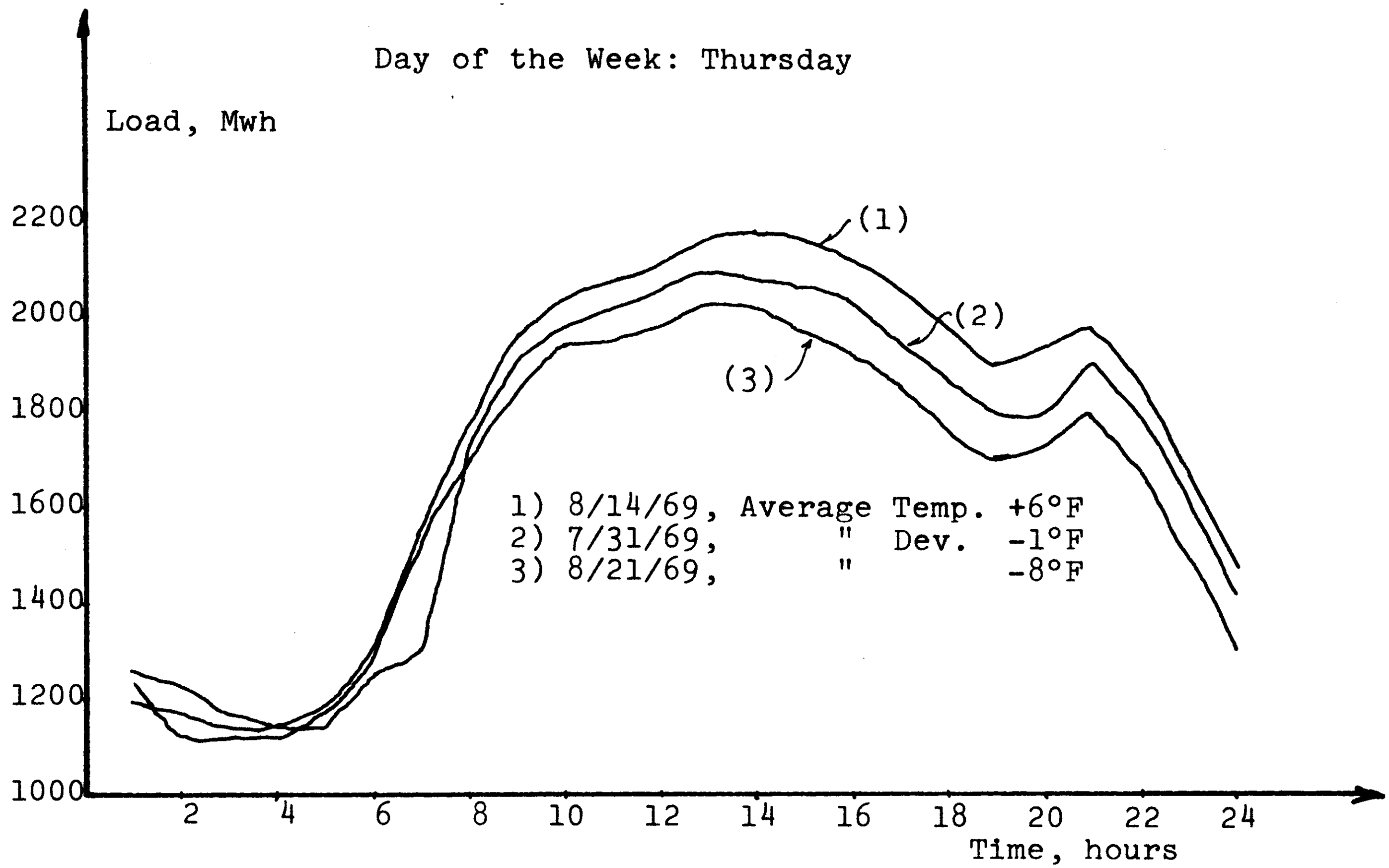


Fig. 38: Effect of Temperature on Summer Load (Cleveland Electric Illum. Co.).

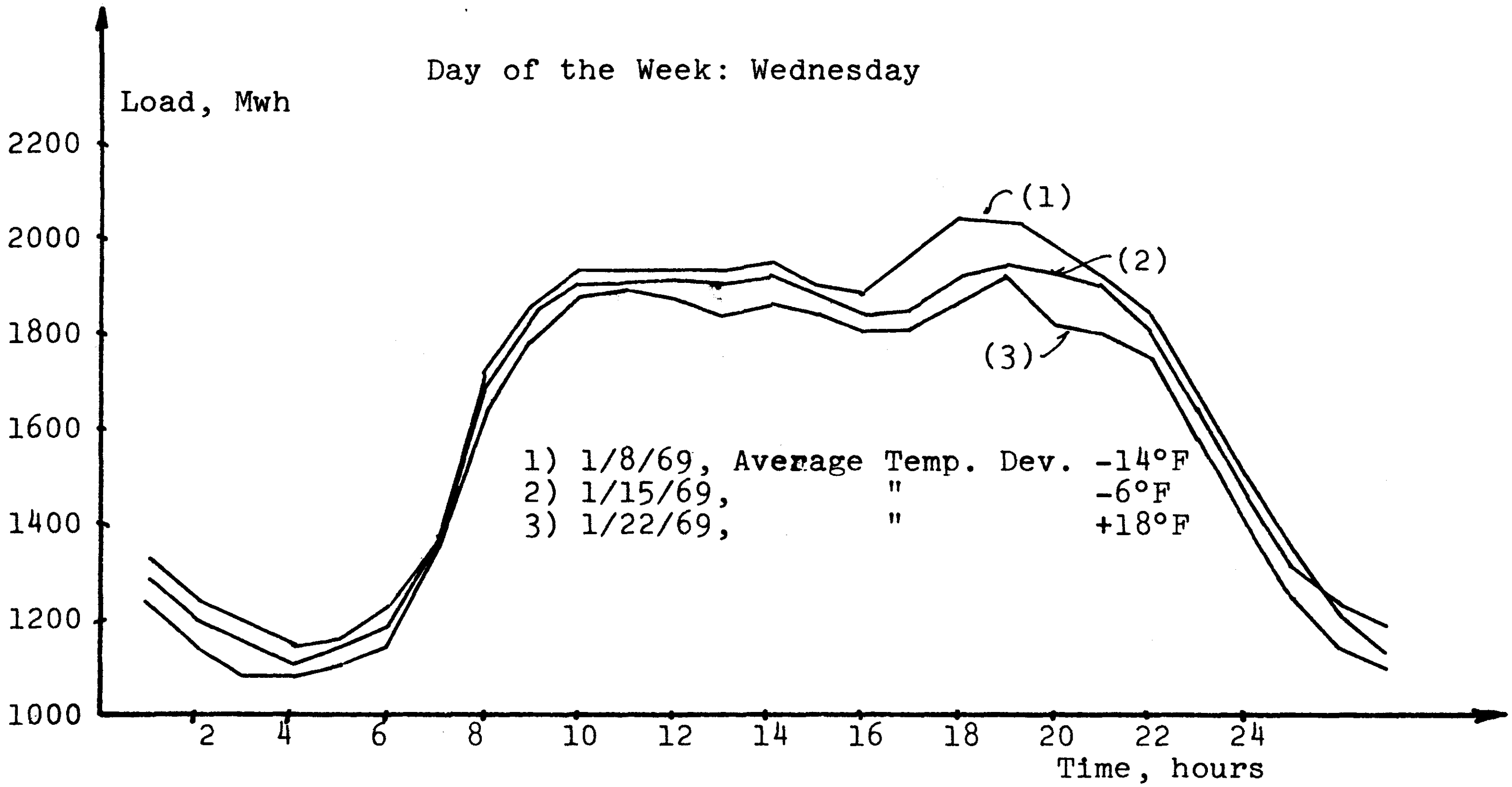


Fig. 39: Effect of Temperature on Winter Load (Cleveland Electric Ill. Co.).

We thus expected the effect of temperature on the winter load model to be relatively minor.

We considered first a model structure with n=m=2 and n_p=5. The data record considered was from Tuesday, January 7th to Friday, January 24th, 1969, that is 12 days excluding Mondays and weekends. This period contained both extremes of cold and warm temperatures for that time of the year, so that sufficient information was available to identify the model. The identified parameters were,

$$\hat{a}_1 = 0.37$$

$$\hat{a}_2 = 0.11$$

$$\hat{b}_0 = 0.68$$

$$\hat{b}_1 = 0.76$$

$$\underline{\hat{x}}_p = \begin{bmatrix} 1653 \\ -321 \\ -167 \\ 25 \\ 2 \\ -25 \\ -241 \\ 65 \\ -4 \\ -35 \\ 15 \end{bmatrix} \qquad (4.37)$$

$$\hat{Q} = 830$$

The characteristic roots are now 0.56 and -0.19, indicating considerably less memory than during the summer time.

The model was used to predict 4 days ahead and update the prediction hourly during the Monday after the last week used in the identification of the parameters, again as would be done on-line.

In Fig. 40 we show the predicted value of y with the standard deviation of the total predicted load error denoted by the horizontal line. In the lower part of the figure we have shown the corresponding inputs which vary from very cold at first to very warm. The temperature sensitive component, y, however remains close to the standard deviation of the error indicating,as expected, the relative load insensitivity to temperature during the cold months.

In Fig. 41 we have shown the actual and predicted load for the same case on the top with the actual prediction error and predicted standard deviation in the bottom. The error is clearly compatible with its level of confidence, i.e. 2% of the peak load approximately.

We also considered the case with n=m=1 and n_p=5. Using the same data record, here we come up with,

$$\hat{a}_1 = 0.43$$
$$\hat{b}_0 = 1.57 \qquad (4.38)$$
$$\hat{Q} = 839$$

and $\underline{x}_p$ identical to the previous model with n=m=2. Prediction with this model yields almost identical results to those of Fig. 40 and 41. This is again expected since the increase in $\hat{Q}$ in going from a second to a first order model is not significant.

IV.6.4 Weekend Models:

Weekend Models were briefly investigated. For this we considered a four week period and the corresponding data for each Saturday and Sunday. The data record length was thus

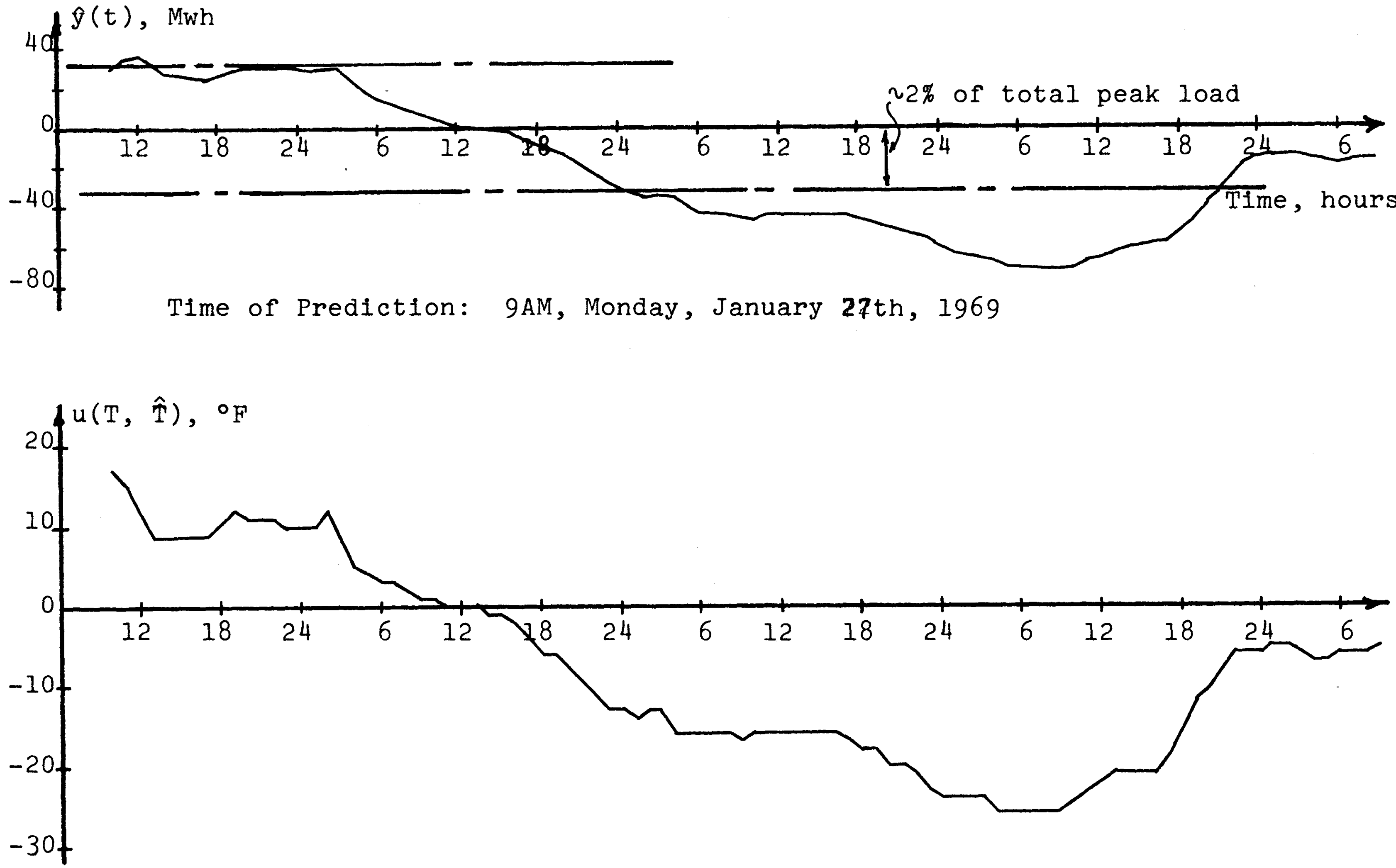


Fig. 40: January Residual Load Forecast and Temperature Deviation - Model $n_p=5$, $n=m=2$.

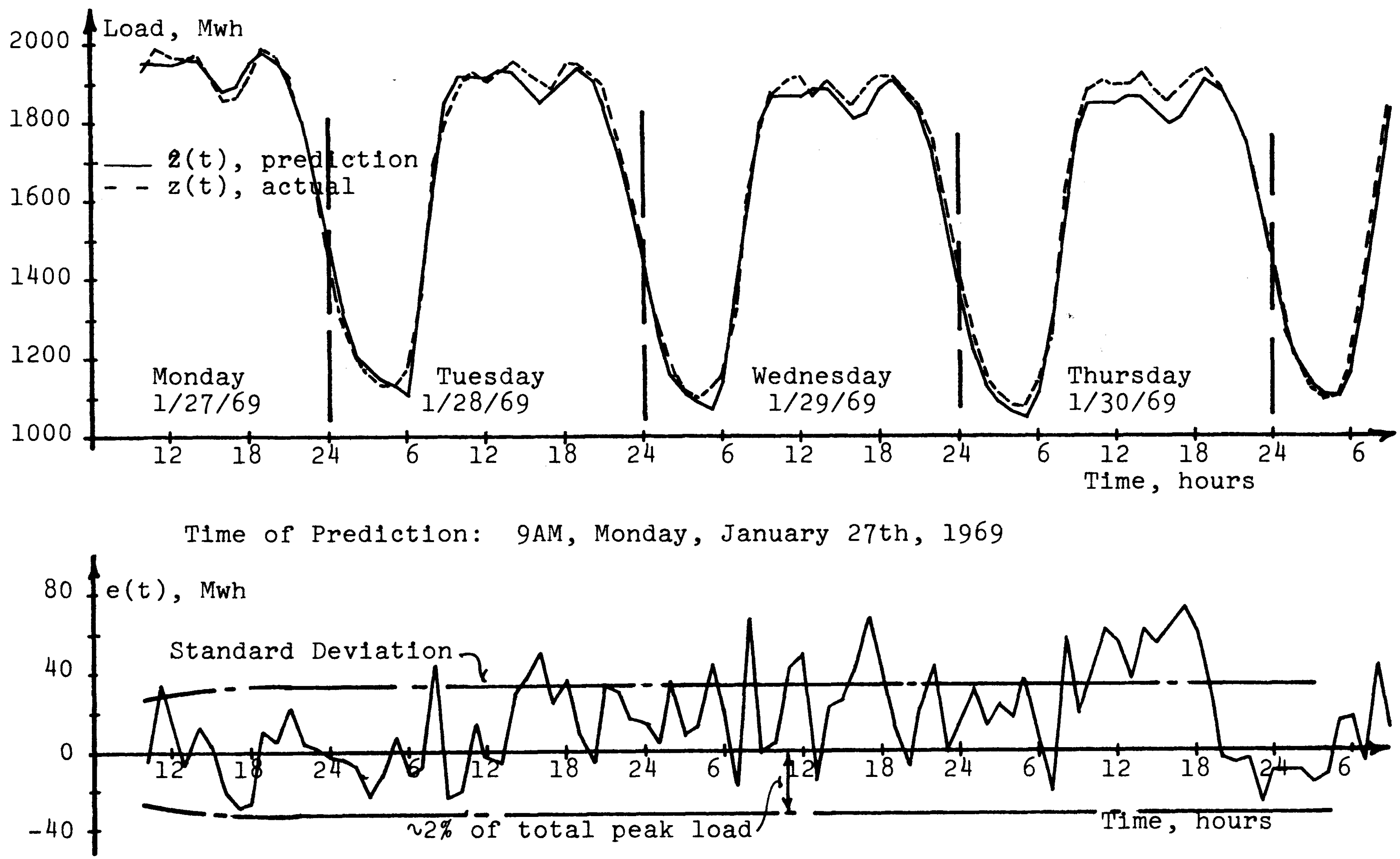


Fig. 41: January Load Forecast and Prediction Error - Model $n_p=5$, $n=m=2$.

four days for each model. Model parameters were evaluated for each case with the identification program yielding as expected a drastically different periodic model with only a slighter variation in the y model.

We considered four weekends in July, 1969, to fit a model with $n_p=5$, $n=m=2$. The results for the Saturday model were,

$$\begin{aligned} \hat{a}_1 &= 0.48 \\ \hat{a}_2 &= 0.40 \\ \hat{b}_0 &= 2.36 \\ \hat{b}_1 &= -1.52 \\ \underline{\hat{x}}_p &= \begin{pmatrix} 1296 \\ -\ 236 \\ -\ 100 \\ -\ 7 \\ -\ 29 \\ -\ 10 \\ -\ 198 \\ 68 \\ 2 \\ -\ 21 \\ -\ 14 \end{pmatrix} \\ \hat{Q} &= 354 \end{aligned} \tag{4.39}$$

while the Sunday model was given by,

$$\begin{aligned} \hat{a}_1 &= 0.71 \\ \hat{a}_2 &= 0.23 \\ \hat{b}_0 &= 1.15 \\ \hat{b}_1 &= -0.24 \end{aligned} \tag{4.40}$$

$$\hat{\underline{x}}_p = \begin{bmatrix} 1115 \\ -199 \\ -56 \\ -22 \\ -31 \\ -6 \\ -1 \\ 72 \\ -15 \\ -5 \\ -8 \end{bmatrix} \qquad (4.40) \text{ cont'd}$$

$$\hat{Q} = 528$$

The characteristic roots for the Saturday and Sunday models are , respectively, [0.93, -0.42] and [0.95, -0.48]. It is interesting to note that the characteristic roots do not change drastically from weekdays to weekends. The main difference lies in the $\underline{b}$ coefficients and the periodic component. This implies that the actual size of the load has little effect on the ability to "remember".

IV.6.5 Anomaly Detection - Real Data:

Anomaly detection is an important characteristic of our model. The theory is discussed in III.11, while a simulated example is presented in IV.5.3. In this section we have used real load and weather data, however since no real anomalies were known to have occured during the period of validity of the identified model, we introduced, artificially, two disturbances which would not be easily detected from the observed load behaviour.

The results are presented in Fig. 42, where the one-step prediction errors are plotted vs time together with the load disturbance. The first anomaly of 100 Mwh lasts only one hour and as seen in the top graph results in two successive

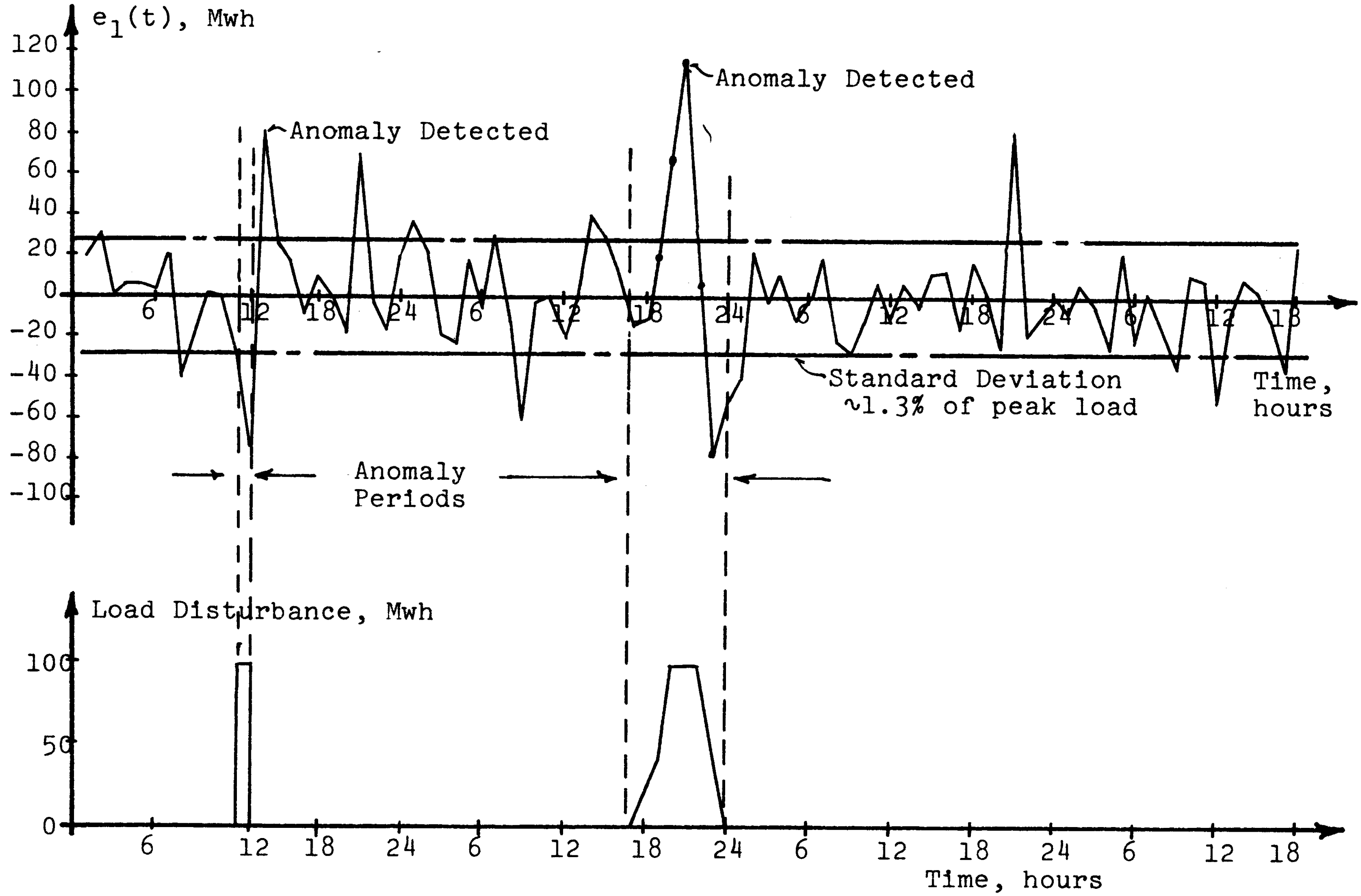


Fig. 42: July One-Step Prediction Error - Anomaly Detection & Self Correction - Model $n_p=6$, $n=m=2$.

unusually large* errors which our scheme detects as an anomaly. The second anomaly has a seven hour duration and appears more gradually so that it is not detected until five observations later.

It is important to notice that the prediction scheme automatically corrects itself after the disturbance has died down, a most significant property.

IV.7 Computer Requirements:

The computer requirements of the proposed identification scheme are the most demanding. Identification of the weekday parameters using 12 days' data used about 128 K of computer space (IBM 360/65) for about 25 to 30 minutes using the double precision option. This program however was not optimized for minimum storage and time requirements. In addition this identification need only be carried out once a week, at a time and day when this computer is not needed for other important tasks.

The estimation-prediction scheme requires minimal storage and computing time. We must only store the system parameters, the state of the system, and the steady state prediction error standard deviations which can be calculated apriori once a week. The input data needed to update the system state is the present and previous input, u, as well as the present load, z. The operations needed are restricted to three matrix multiplications and additions. The prediction requires similar

* Greater then 2 standard deviations.

operations, but forecasted inputs must be fed in to evaluate the forecast load. However the predicted inputs need not be stored in this computer, but merely fed in one at a time. The anomaly detection scheme is a simple program with a 3 hour memory logic.

The estimation-prediction part of the scheme can therefore be programmed on-line in a small computer such as a PDP-8.

IV.8 Summary of Results:

In this chapter we have presented in a somewhat chronological order the experimantal work done with simulated and real data.

Section IV.2 discusses a crude weighted least squares approach of testing the validity of the proposed model. Results indicated that the separation of the load into periodic and temperature sensitive components was very reasonable indeed.

Section IV.3 discusses tests carried out to evaluate the possible separation of the two load components and model each separately. This would reduce the computational requirements of the identification process. Although the theory here was well developed, considerably more experimental work would be required to perfect it , so we proceeded to a more complex but more reliable and readily available Fletcher-Powell approach.

In section IV.4 we discuss the identification problem via the Fletcher-Powell scheme using simulated data. This was very successful, so we next tested in IV.5 the various characteristics of the on line prediction-updating scheme, again with simulated data. Here we also test the ability of the identified model to predict, as is done with real data. Anomaly detection is also tested.

Section IV.6 evaluates the techniques discussed in IV.4 and IV.5 with real data. We started by estimating the best model order. This was done following the guidelines of III.9 and a model with n=m=2, n_p=6 was found to give reasonable results.

We tested the ability of the identified model to predict and presented arguments that indicated the importance of the temperature sensitive component in predicting load, especially for large deviations in temperature. Various models were evaluated for different times of the year. We showed the ability to predict during long hot spells, as well as during the winter time. The ability of the scheme to detect anomalous load behaviour and to correct itself after the disturbance is also shown.

V.0 PRACTICAL IMPLEMENTATION OF PROPOSED LOAD FORECASTING APPROACH

V.1 Preliminary Discussion

In this chapter we discuss a number of guidelines and procedures to be followed when implementing the proposed load forecasting approach in a real system. Naturally only general rules can be given since each particular system has its peculiarities and specific. objectives.

V.2 Recommendations for Implementation

V.2.1 Off-Line Study:

The first recommendation is that an off-line study similar to the one carried out here be made with the particular system's observed data. This would provide an estimate of the model's order, and indicate whether additional complexity is needed for the particular system.

The study should, if possible, be carried out for an entire year's data, including holidays and weekends, simulating on-line operation. This would give us more confidence when implementing the technique on-line. Additional recommendations on the off-line study are described in the next chapter.

V.2.2 Guidelines for Type and Form of Data

A) When carrying out the off-line study the data being used should be screened for abnormal behaviour, since its use would disturb the normal behaviour model.

B) The data used should be such that facilities for their on-line measurements be available. There is no great difficulty in obtaining load data in this form, as most central control units have load telemetering facilities. As far as temperature data is concerned, telemetering from one or more weather stations should provide hourly temperature measurements as well as the forecast temperature profile and its confidence level at that time.

Close contact with the local weather bureau should be made,to determine when forecasts are updated, what their length is, and what their confidence level is.

C) Certain power systems are widely distributed over a large geographical area. For such systems, a possible approach would be to separate the system into two or more regions, each of which has approximately uniform weather effects. Alternatively, data weighting according to the region, could be used to evaluate one average model.

V.2.3 On-Line Implementation

Only after considerable off-line study with observed data does the author recommend the on-line implementation of the technique. After sufficient confidence has been gained in the model and its capabilities by the off-line study then on-line implementation could be carried out.

In this section we discuss as completely as possible at this level the steps necessary to implement the on-line prediction technique. These are as follows:

A) Model Identification and Adaptation

B) On-Line Forecasting and Updating of Forecast

A) Model Identification and Adaptation -

This step is carried out much less frequently than step B. The exact frequency depends on the time of year, and how fast conditions are varying. In our study we considered updating the model once a week, but more or less frequent reestimations may be needed.

The computer requirements of the identification step are quite demanding as discussed in IV.7, however since this needs to be carried out quite infrequently, that is once a week approximately, relatively little computer time is taken up by this step.

If four models are used, i.e. one each for Saturday, Sunday, Monday and the remaining weekdays, then four such identification steps would be carried out. The data record used to identify each model must be kept in storage continuously new data replacing the oldest one as new observations are made. The length of this data record could vary, but as discussed in III.11, it should not exceed six weeks. Note that for prediction purposes we don't have to operate on this entire data record but a much simpler operation is performed. This data is used only to identify the model's parameters which are then used for prediction purposes. The maximum number of days that should be stored is six weeks x 7 days = 42 days.

The identification program needed in this step is described in Appendix A together with the instructions for its use.

B) <u>On-Line Forecasting and Updating of Forecasting</u> -

In this study load forecasts are made once every hour. The length of the forecast is variable and has been shown to be compatible with the model at least up to one week.*

The computer requirements for this step are relatively small, as discussed in IV.7, a small digital computer with A/D lines and converters being sufficient. Alternatively, the scheme could be programmed on a time-sharing basis in a larger computer.

A display console or some type of output mode would be desirable for the operator to observe load predictions when required. A set of controls should be available to the operator to manipulate things like the length of the forecast, personal corrections or interventions, and contingency load predictions. The latter would allow the operator to consider the effect of possible weather variations on the load behaviour, which may be important in unit commitment, power exchanges, and maintenance planning.

If a known load disturbance is expected, operator intervention can be made possible by for example, correcting the y_p component for the time of the disturbance. This correction could be done automatically when the operator feeds in

* This figure may depend on the particular power system and time of year.

the approximate shape and duration of the disturbance. A fitting routine could then be used to alter the $\underline{x}_p$ parameters accordingly. This would allow automatic updating to continue. If the abnormality affects the weather sensitive load, y, such as that due to a known load shedding, or due to a severe loss of power, then automatic updating and prediction should probably be discontinued altogether, and we should revert to the human operator completely.

The output display should also indicate whether anomalous load behaviour has occured as discussed in III.11. This could be indicated as a written message and/or by some visual or audio indicator. The program should be set up, as discussed in III.11, so that when an anomaly is detected, closed loop updating is discontinued, however closed loop updating can continue in order to detect the return of the load to normal levels. This could again be indicated by a written message or by discontinuing the warning signal. Normal load prediction can also continue via the open loop predictor.

To allow for continuous operation throughout the week, the four load models suggested must be stored in the computer, together with the present state estimate of the system. This can be incorporated into the prediction program by having a counter which automatically changes the model parameters at the end of each period, e.g. from a Sunday to a Monday model. The state estimate is kept at its last value. In addition storage for predicted temperatures must be available.

Generally no data from the previous years should be stored, but for some minor exceptions like school vacations, switch to Daylight Saving Time, Christmas lighting, and other regular seasonal effects. This could however be incorporated into the y_p component by the operator as discussed above when the time comes.

In Fig. 43 we have presented in pictorial form the entire operation here described.

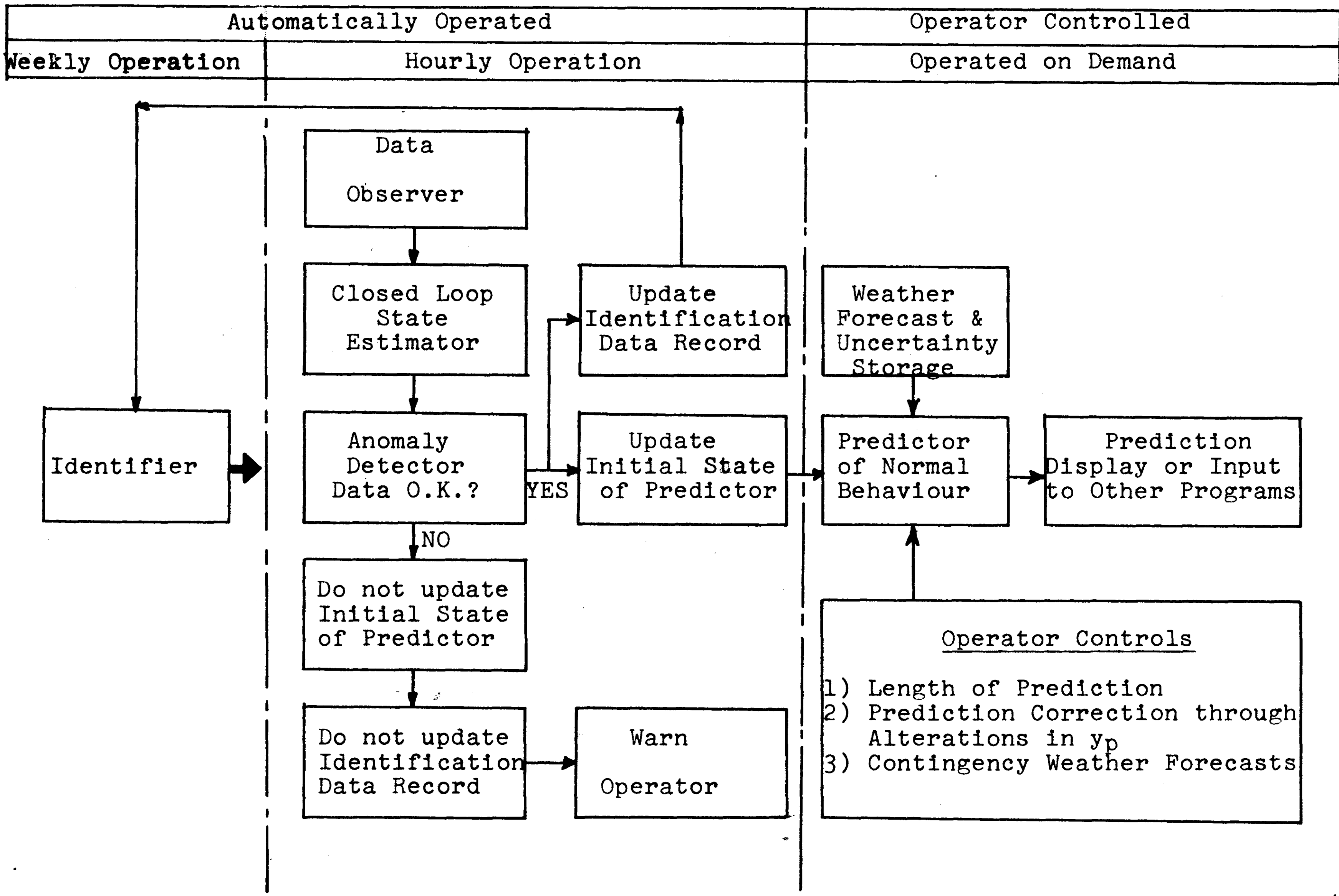


Fig. 43: Block Diagram Implementation of Complete Approach.

VI.0 CONCLUSIONS & RECOMMENDATIONS

VI.1 Conclusions:

A mathematical model for the load of a power system was developed with the main purpose of short term load forecasting. The prediction times investigated ranged from hours to one week. The main characteristics of the model were:

A) Load is modelled as a discrete time process. Hourly intervals were used, but we are not rigidly restricted to this period. This provides more information about load behaviour than models which describe only a small number of load values.

B) A structure is hypothesized, justified physically, and verified by experimentation, which separates load behaviour into periodic and temperature dependent components.

C) The temperature dependent component is described by a non-linear memoryless transformation between actual temperature and normal temperature in cascade with a linear dynamical system.

D) Modelling uncertainty is included in the load model as an additional white noise input to the above mentioned linear dynamical system. This term models all the effects that have not been explicitly described, whether they be due to inherent uncertainty, omitted non-linearities or other weather effects.

E) The model has a number of parameters in both the periodic and weather dependent random components which are later on adjusted via identification techniques to fit the model to the particular load behaviour. These parameters are reidentified approximately once a week.

F) Four models are suggested, one each for Monday, Saturday, Sunday, and the remaining weekdays.

A number of system identification techniques are presented and analyzed. We concluded based on experimentation that the best approach in terms of expediency, but not necessarily the least complex, was one based on the Fletcher-Powell algorithm to minimize the identification criterion.

Linear filtering and prediction algorithms are used in the implementation of the load forecasting scheme. This provides us with a computationally efficient and simple scheme of on-line prediction and updating as new data is observed. In addition it allows us to automatically detect anomalous load behaviour. A number of advantages and disadvantages due to this prediction scheme are discussed in II.9.

A large number of tests are carried out on the proposed schemes, both with simulated and real data, the latter from the Cleveland Electric Illuminating Company.

The real data studies are done for various times of the year to test the model's adaptability to seasonal variations. First, however, some effort is spent in determining a valid model order. For the summer months, a model with $n_p=6$, $n=m=2$, was shown to yield quite good results, while for the winter months, $n_p=5$, $n=m=1$ was sufficient. However these figures will vary from company to company.

Further studies were performed to test the length of the data record needed in the identification program. Here we are limited as to its length, on the upper side by the slow time variations of load behaviour over the seasons to not more than six weeks (less during spring and fall), and on the lower side by the numerical efficiency of the identification scheme to at least three or four days. For the weekdays' model we picked a period of three weeks, or twelve days, which yielded good results.

We investigated models for the months of July, August, and January, which included unusually cold and warm conditions in both summer and winter. In all cases the prediction error varied from about 1% of peak load for a one hour prediction, to about 2.5% of peak load for a one week forecast, assuming perfect knowledge of the future weather. This was always compatible with the predicted error standard deviations. Weather prediction uncertainty can also be incorporated into the load forecast uncertainty as discussed II.7.2.

The temperature dependent component during summer was found to be very important whenever temperature deviations were large. During periods of relatively small temperature deviations from normal the effect of this component was less significant, as expected.

It was found, for the particular power system considered, that the sensitivity of load to temperature during winter was quite small, so that the model obtained did not depend as heavily on temperature deviations.

Anomaly detection was tested by introducing artificial disturbances into the actual data. These disturbances were of finite duration and less than 5% of peak load, so that their detection with the naked eye could not be easily done. The anomaly detection scheme detects these disturbances, and when they end, the closed loop prediction-updating scheme returns the prediction to normal.

Finally, in chapter V, a number of guidelines and recommendations are given for the implementation of the proposed techniques. Implementation into a real system is certainly recommended, provided a thorough study, like the one described above, of the specific behaviour of the particular system is first carried out.

Interaction between operator and machine is an important characteristic of this approach. The scheme here proposed does not have the replacement of the load forecasting-scheduling operator as its objective. Instead, a complementary action is all that is planned. Thus, the proposed forecastor would automatically predict under normal conditions, warning the operator whenever anomalous load behaviour is detected. This would free the operator, under normal conditions, to perform other important tasks. Further interaction is also possible.

Known disturbances, such as those due to special national or local events, factory strikes, school shutdowns, holidays, Christmas lighting, time shifts, etc., may be incorporated into the prediction by the operator to a fitting routine, given the estimated shape and duration of the disturbance.

Theoretically the operator need not understand the mathematics or structure of the model, however, this ability would prove very valuable in making full use of the combined capabilities of the operator-predictor operation, and it is highly recommended. The basic structure of the model and its operation are relatively simple and do not require knowledge of advanced mathematics to understand.

VI.2 Recommendations:

VI.2.1 Data Recommendations:

A) The possibility of taking load data observations at shorter time intervals should be considered. This may be useful in detecting anomalous load behaviour more quickly.

B) Temperature data was only available for this study in tri-hourly form, so that interpolations had to be made. This could have introduced some errors. Future studies may consider more frequent weather observations.

C) Further studies could be carried out under varying weather conditions and different times of the year.

D) Data from companies that emphasize winter electric heat could be tried to test the model's capability under these conditions.

VI.2.2 Modelling Recommendations :

A) Instead of modelling the periodic component as a sum of sinusoids, we could investigate its modelling by a polynomial time series. This may allow us to describe the sharp load curve peaks without as many parameters as are now required.

B) Certain seasonal variations in the load behaviour could be stored as a correction term to be added to the periodic component when the time comes. Such variations could be due to events like Christmas holidays, time shifts, school holidays, etc. This would improve the model's performance during these periods. Alternatively, we could rely on the operator's personal estimate of such effects which he would incorporate into the model as suggested in chapter V.

C) The no-effect region, (60°,70°), in the nonlinear relation between u,T and $\hat{T}$, may very well vary over the seasons, or may depend on the power company. For example, it is possible that in summer time no heating load will appear even if the temperature falls below 60°F. Alternatively in winter, no more decrease in heating load could occur if the temperature exceeds 50°F, say. A number of possibilities arise, which should be tested by experimentation. In this study the structure chosen was sufficiently valid, so that unless results indicate more complicated models between u,T and $\hat{T}$ may be needed, we don't recommend any changes.

D) Different temperature sensitive models may be needed in some power systems for the working and rest hours. However, again, this study does not indicate that this complexity should be added, as prediction errors are already quite reasonable.

E) In some systems it may be necessary to include other weather effects such as humidity and light intensity. These can be treated as additional inputs to the weather dependent linear system. Again, in this study no need for additional weather effects was found necessary. It should be noted that light intensity may vary drastically over the load area, and may be difficult to forecast. The author's opinion is that unless a significant contribution to the total load is made by light intensity, it should be treated as part of the model uncertainty.

VI.2.3 System Identification Recommendations:

A) The separation algorithms described in chapter III should be investigated more thoroughly to check for possible savings in the computer requirements of the identification scheme.

B) Inclusion of coloured noise in the ARMA temperature dependent model was not found necessary here since the residual errors are essentially white. It is however possible that the coloured noise assumption may yield better results for data from a different system.

Appendix A: SYSTEM IDENTIFICATION PROGRAM

This appendix describes a computer program, written in FORTRAN language, which solves the identification of the special load model (111.7) by the Fletcher-Powell method.

We make use of the double precision subroutine DFMFP which applies the Fletcher-Powell algorithm for the minimization of the identification criterion J.

Equation 3.160 is used to define J while 3.161, 3.162,3.163 define the gradient of J with respect to the unknown parameters.

The program is thus essentially to read load and temperature deviations for the data record being considered, and to define the function J and its gradient for any given value of the unknown system parameters by the above equations. Subroutine DFMFP is then called to minimize J.

```
C
C     MAIN PROGRAM, TO IDENTIFY THE PARAMETERS OF A LINEAR DISCRETE
C     TIME SYSTEM DRIVEN BY A DETERMINISTIC INPUT U, AND WHITE NOISE
C     FLETCHER POWELL METHOD
C
      DOUBLE PRECISION Z,U, ARG,GRAD,H,F,Q
      DIMENSION Z(400),U(400),ARG(26),GRAD(26),H(429)
      EXTERNAL FUNCT
      COMMON NS,MS,NP,NDAY,NWK
      COMMON/DATA/Z,U
C
C     TO DEFINE SYSTEM DIMENSIONS
C
      NS=1
      MS=1
      NP=5
      NP3=2*NP+1
      NV=NS+MS+NP3
      NS1=NS+1
      NM=NS+MS
      NM1=NM+1
C
C     NUMBER OF NON LINEAR ITERATIONS
C
      LIMIT=100
C
C     TO READ NO. OF DAYS DATA BEING USED IN IDENTIFICATION PER WEEK
C
      READ 5,NDAY
C
C     NUMBER OF WEEKS BEING CONSIDERED  IN DATA RECORD
      NWK=3
C
    5 FORMAT(I3)
      NDAS=NWK*NDAY
      NHOURS=NDAS*24
```

```
      ND=NDAY*24
      NDT=ND-NS
      NDT=NDT*NWK
      PRINT 1,NDAS,NHOURS
    1 FORMAT('NUMBER OF DAYS AND DATA POINTS =', I5,',',I5)
C
C     ESTIMATE OF MINIMUM LIKELIHOOD FUCNTION
C
      EST=1000
C
C     TO READ LOAD AND TEMPERATURE DATA
C
    6 READ 157, (Z(I),I=1,NHOURS)
  157 FORMAT(20X,12D5.0)
   15 FORMAT(8D10.2)
      READ 15,(U(I),I=1,NHOURS)
      DO 2 I=1,NHOURS
      PRINT 3, Z(I),U(I)
    3 FORMAT(2D14.5)
    2 CONTINUE
C
C     TO READ INITIAL ESTIMATE OF PARAMETER VALUES
C
C
C     FIRST NS ARE A VECTOR, SECOND MS ARE B VECTOR, THIRD NP3 ARE XP
C
      READ 16, (ARG(I),I=1,NV)
   16 FORMAT(8D10.5)
      PRINT 4, (ARG(I),I=1,NV)
    4 FORMAT(' ARG =',9D13.5/6X,9D13.5)
C
C     TO CALL DFMFP SUBROUTINE
C
      EPS=0.1E-15
      CALL DFMFP(FUNCT,NV,ARG,F,GRAD,EST,EPS,LIMIT,IER,H)
C
```

```
C      TO PRINT OUTPUT
C
C
C      SYSTEM PARAMETERS
C
       PRINT 19,IER
    19 FORMAT('      IER  =',I2)
       PRINT 20,(ARG(I),I=1,NS)
    20 FORMAT(1H0,' A VECTOR =',15F8.4)
       PRINT 21,(ARG(I),I=NS1,NM)
    21 FORMAT(1H0,' B VECTOR =',15F8.3)
       PRINT 22,(ARG(I),I=NM1,NV)
    22 FORMAT(1H0,' XP VECTOR=',15F8.3)
       PRINT 23,F
    23 FORMAT(1H0,'Q ESTIMAT =',D11.4)
       PRINT 24, LIMIT
    24 FORMAT(1H0,'NO OF ITER =',I3)
C
       GO TO 6
     7 CALL EXIT
C
       END
```

```
      SUBROUTINE FUNCT(NV,ARG,VAL,GRAD)
C
C     THIS ROUTINE DEFINES VALUE OF J AND GRADIENT FOR ANY GIVEN PARAME-
C     TER VECTOR VALUE
C
      DOUBLE PRECISION ARG,VAL,GRAD,ZIT,ZITV,UITV,A,B,XP,PHI,PSI,F,GRADS
      DOUBLE PRECISION CK
      DIMENSION ARG(26),ZITV(5),UITV(6),A(5),B(5),XP(15),PHI(15),
     1PSI(5,15),GRADS(26),GRAD(26)
      COMMON NS,MS,NP,NDAY,NWK
      COMMON/ERROR/ZIT,ZITV,UITV
      COMMON/AGO/A,B,XP
C
C     TO DEFINE VALUE OF FUCNTION, VAL
C
      VAL=0.
      NS1=NS+1
      NP3=2*NP+1
      NV=NS+MS+NP3
      CK=1.0/DFLOAT(NWK*(NDAY*24-NS))
      DO 5 I=1,NV
    5 GRAD(I)=0.
      CALL SEPAR(ARG)
      DO 30 IWEEK=1,NWK
      JDAY=(IWEEK-1)*NDAY+1
      DO 10 IT=NS1,24
      CALL ZU(IT,JDAY)
      CALL ERRORS(IT,F,GRADS)
      VAL =VAL+F**2
      CALL SMPY(GRADS,2.0,GRADS,NV,1,0)
   10 CALL GMADD(GRAD,GRADS,GRAD,NV,1)
      IF(NDAY.EQ.1) GO TO 30
      KDAY=JDAY+NDAY-1
      JDAY=JDAY+1
      DO 15 IDAY=JDAY,KDAY
      DO 20 IT=1,24
```

```
      CALL ZU(IT,IDAY)
      CALL ERRORS(IT,F,GRADS)
      VAL=VAL+F**2
      CALL SMPY(GRADS,2.0,GRADS,NV,1,0)
      CALL GMADD(GRAD,GRADS,GRAD,NV,1)
   20 CONTINUE
   15 CONTINUE
   30 CONTINUE
      VAL=CK*VAL
      CALL SMPY(GRAD,CK,GRAD,NV,1,0)
      PRINT 25, VAL
   25 FORMAT(1H0,' FUNCTION =',D13.5)
      PRINT 26,(ARG(I),I=1,NV)
   26 FORMAT(1H ,' ARGUM = ',9D13.5/8X,9D13,5/8X,9D13.5)
      PRINT 27,(GRAD(I),I=1,NV)
   27 FORMAT(' GRAD=',9D13.5/6X,9D13.5/6X,9D13.5)
      RETURN
      END
```

```
      SUBROUTINE ZU(IT,IDAY)
C
C     THIS ROUTINE CALCULATES Y OF T , U OF T AND THE
C     CORRESPONDING VECTORS GIVEN THE TIME OF DAY AND THE SYSTEM DIMENS
C
      DOUBLE PRECISION Z,U,ZIT,ZITV,UITV
      DIMENSION ZITV(5),UITV(6),Z(400),U(400)
      COMMON/DATA/ Z,U
      COMMON/ERROR/ZIT,ZITV,UITV
      COMMON NS,MS,NP,NDAY,NWK
C
C     TO DEFINE ZIT AND UIT
C
      IN=IT+24*(IDAY-1)
      ZIT=Z(IN)
C
C     TO DEFINE ZITV AND UITV
C
      DO 10 I=1,NS
   10 ZITV(I)=Z(IN-I)
      DO 20 I=1,MS
   20 UITV(I)=U(IN-I+1)
      RETURN
      END
```

```
      SUBROUTINE PSISUB(IT,PSI)
C
C     THIS ROUTINE DEFINES PSI MATRIX AS A FUNCTION OF THE HOUR OF DAY
C
      DOUBLE PRECISION PHI,PSI,PI
      DIMENSION PSI(5,15),PHI(15)
      COMMON NS,MS,NP,NDAY,NWK
      NP3=2*NP+1
      DO 15 I=1,NS
      ITI=IT-I
      IF(ITI.LE.0) ITI=24+ITI
      CALL PHISUB(ITI,PHI)
      DO 20 J=1,NP3
   20 PSI(I,J)=PHI(J)
   15 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE PHISUB(IT,PHI)
C
C     THIS DEFINES PHI MATRIX AS FUNCTION OF HOUR OF DAY
C
      DOUBLE PRECISION PHI,PI
      DIMENSION PHI(15)
      COMMON NS,MS,NP,NDAY,NWK
      NP1=NP+1
      NP2=NP+2
      NP3=2*NP+1
      PI=3.14159265359
C
C     TO DEFINE PHI
C
      PHI(1)=1.
      DO 10 I=2,NP1
   10 PHI(I)=DSIN(2.*PI*DFLOAT(IT*(I-1))/24.)
      DO 20 I=NP2,NP3
   20 PHI(I)=DCOS(2.*PI*DFLOAT(IT*(I-NP1))/24.)
      RETURN
      END
```

```
      SUBROUTINE GRADO(F4,UITF,G1,GRADS)
C
C     THIS DEFINES GRADS FROM COMPONENTS
C
      DOUBLE PRECISION F4,G1,UITF,GRADS
      DIMENSION F4(5),G1(15),UITF(6),GRADS(26)
      COMMON NS,MS,NP,NDAY,NWK
      NS1=NS+1
      NM=NS+MS
      NM1=NM+1
      NP3=2*NP+1
      NV=NS+MS+NP3
      DO 10 I=1,NS
   10 GRADS(I)=F4(I)
      DO 15 I=NS1,NM
   15 GRADS(I)=-UITF(I-NS)
      DO 20 I=NM1,NV
   20 GRADS(I)=G1(I-NM)
      RETURN
      END
```

```
      SUBROUTINE SEPAR(ARG)
C
C     THIS SEPARATES ARG INTO A,B,AND XP
C
      DOUBLE PRECISION ARG,A,B,XP
      DIMENSION ARG(26),A(5),B(6),XP(15)
      COMMON NS,MS,NP,NDAY,NWK
      COMMON/AGO/A,B,XP
      NS1=NS+1
      NM=NS+MS
      NM1=NM+1
      NP3=2*NP+1
      NV=NS+MS+NP3
      DO 10 I=1,NS
   10 A(I)=ARG(I)
      DO 15 I=NS1,NM
      J=I-NS
   15 B(J)=ARG(I)
      DO 20 I=NM1,NV
      J=I-NM
   20 XP(J)=ARG(I)
      RETURN
      END
```

```
      SUBROUTINE ERRORS(IT,F,GRADS)
C
C     THIS CALCULATES ERROR,Z(T)-ZHAT(T) FOR ANY GIVEN TIME IT
C
      DOUBLE PRECISION ZIT,ZITV,UITV,A,B,XP,PSI,PHI,F,F1,F2,F3,F4,
     1F5,GRADS,G1,UITF
      DOUBLE PRECISION D,PHIX,PSIX
      DIMENSION D(15,15) ,PHIX(15),PSIX(5,15)
      DIMENSION ZITV(5),UITV(6),PSI(5,15),PHI(15),A(5),B(6),XP(15),
     1F1(1),F2(1),F3(1),F4(5),F5(1),GRADS(26), G1(15),UITF(6)
      COMMON/ERROR/ZIT,ZITV,UITV
      COMMON/AGO/A,B,XP
      COMMON NS,MS,NP,NDAY,NWK
      NP3=2*NP+1
C
C     DEFINE XP VECTOR SCALING MATRIX D
C
      DO 5 I=1,NP3
      DO 5 J=1,NP3
    5 D(I,J)=0.
      D(1,1)=1000.0
      D(2,2)=100.
      D(3,3)=100.
      D(4,4)=10
      D(5,5)=10
      D(6,6)=10
      D(7,7)=100
      D(8,8)=100
      D(9,9)=10
      D(10,10)=10
      D(11,11)=10
C
C     D IS DEFINED HERE FOR NP=5
C
      CALL ARRAY(2,NP3,NP3,15,15,D,D)
      CALL GTPRD(A,ZITV,F1,NS,1,1)
```

```
      CALL GTPRD(B,UITV,F2,MS,1,1)
      CALL PHISUB(IT,PHIX)
      CALL GMPRD( D,PHIX,PHI,NP3,NP3,1)
      CALL GTPRD(XP,PHI,F3,NP3,1,1)
      CALL PSISUB(IT,PSIX)
      CALL ARRAY(2,NS,NP3,5,15,PSIX,PSIX)
      CALL GMPRD(PSIX,D,PSI,NS,NP3,NP3)
      CALL GMPRD(PSI,XP,F4,NS,NP3,1)
      CALL GTPRD(A,F4,F5,NS,1,1)
      F=ZIT-F1(1)-F2(1)-F3(1)+F5(1)
C
C     TO DEFINE GRAD
C
      CALL GMSUB(F4,ZITV,F4,NS,1)
      CALL GTPRD(PSI,A,G1,NS,NP3,1)
      CALL GMSUB(G1,PHI,G1,NP3,1)
      CALL SMPY(F4,F,F4,NS,1,0)
      CALL SMPY(UITV,F,UITF,MS,1,0)
      CALL SMPY(G1,F,G1,NP3,1,0)
      CALL GRADO(F4,UITF,G1,GRADS)
      RETURN
      END
C
C     A NUMBER OF DOUBLE PRECISION MATRIX OPERATING SUBROUTINES
C     HAVE NOT BEEN LISTED HERE SINCE THESE ARE SSP ROUTINES
C     WHOSE ONLY MODIFICATION IS AN ADDITIONAL DOUBLE PRECISION
C     STATEMENT. THESE ARE GMPRD,ARRAY,SMPY,GTPRD,GMADD,AND GMSUB.
C
```

Appendix B: ESTIMATION PREDICTION ALGORITHMS

This program,also written in FORTRAN, solves the estimation-prediction equations used in forecasting load and its standard deviation as well as updating these quantities in a recursive fashion.

The equations programmed are as depicted in Figs. 7 and 8.

The system parameters read by the program,B,D, and X, assume that the model is in state space form as given by equations 2.19 through 2.25 with the total state defined by$[\frac{x}{\underline{x}}p]$. The parameters AS stand for the a_i in the ARMA model. BO stands for b_0 in the same model.

```
C      MAIN PROGRAM - LOAD STATE ESTIMATION AND PREDICTION
C
       DOUBLE PRECISION SQP,YV
       DOUBLE PRECISION Z,U,AS,B,D,S,X,Q,R,BO,ZV,EV,ERR,UV,C,P,G,A
       DOUBLE PRECISION ZI,UI,U1
       DOUBLE PRECISION Y
       DIMENSION YP(96),EYP(96)
       DIMENSION SQP(20),YV(96)
       DIMENSION Z(480),U(480),AS(5),B(25),D(25),S(25,25),X(25),Q(1),R(1)
      1,ZV(96),ERR(96),UV(97),C(20),P(20,20),G(20),A(20,20) ,EV(96)
C
C      TO DEFINE SYSTEM DIMENSIONS
C
       NS=1
       NP=5
       NP1=2*NP+1
       NP2=NP1+1
       N=NP1+NS
       NP3=NP+2
C
C      TO DEFINE SYSTEM PARAMETERS AND INITIAL STATE
C
       Q(1)=839
       R(1)=0.0
    10 FORMAT(8D10.3)
       READ 10,(AS(I),I=1,NS)
       READ 10, (B(I),I=1,N),BO
       READ 10, (D(I),I=1,N)
       READ 10,(X(I),I=1,N)
C
C      DEFINE INITIAL COVARIANCE MATRIX
C
       DO 15 I=1,N
       DO 15 J=1,N
       S(I,J)=0.0
    15 S(I,I)=0.1D+10
```

```
      DO 17 I=1,NP1
   17 S(I,I)=0.0
      CALL ARRAY(2,N,N,25,25,S,S)
C
C     DEFINE A MATRIX
C
      CALL DEFINE(AS,NS,NP,A)
      CALL ARRAY(2,N,N,20,20,A,A)
C
C     READ INPUT DATA
C
      NDATA=120
      READ 11,(Z(I),I=1,NDATA)
   11 FORMAT(20X,12D5.0)
      READ 10,(U(I),I=1,NDATA)
C
C     MAXIMUM PREDICTION TIME=NTP<96, ESTIMATE THUS UP TO MAX NMAX=NDATA
C - NMAX
C
      NTP=96
      NMAX=NDATA-NTP
C
C     TO ESTIMATE,PREDICT NTP AHEAD FROM TIME = 1,NMAX
C
      DO 20 I=2,NMAX
      IT=MOD(I,24)
C
C     DEFINE C MATRIX
C
      CALL CIT(IT,NS,NP,C)
      PRINT 42,IT,(C(K),K=1,N)
   42 FORMAT(1H0,' C(',I2,')=',9D13.5/8X,9D13.5)
C
C     TO DEFINE G AND P, THE GAIN AND COVARIANCE MATRICES
C
      CALL SIGMA(N,1,A,C,Q,D,R,S,P,G)
```

```
      PRINT 40,(G(K),K=1,N)
   40 FORMAT(1HO,' GAIN  =',9D13.5/8X,9D13.5)
      DO 47 JK=1,N
      CALL ARRAY(1,N,N,20,20,P,P)
      SQP(JK)=DSQRT(P(JK,JK))
      CALL ARRAY(2,N,N,20,20,P,P)
   47 CONTINUE
      PRINT 48,(SQP(K),K=1,N)
   48 FORMAT(1HO,' VAR   =',9D13.5/8X,9D13.5)
C
C     UPDATE STATE
C
      ZI=Z(I)
      UI=U(I)
      U1=U(I-1)
      CALL STATES(X,A,B,BO,C,G,N,ZI,U1,UI)
      PRINT 41,(X(K),K=1,N)
   41 FORMAT(1HO,' XP EST=',9D13.5/8X,9D13.5)
      PRINT 50, U(I)
   50 FORMAT(1HO,' TEMP DEVIAT = ', F8.2)
C
C     PREDICT LOAD NTP HOURS AHEAD
C
C     FIRST DEFINE FUTURE INPUT VECTOR UV
C
      NTP1=NTP+1
      DO 25 K=1,NTP1
      J=I+K-1
   25 UV(K)=U(J)
C
      CALL PREDIC(N,NS,NP,IT,NTP,X,P,A,B,BO,D,Q,UV,ZV,EV,YV)
      PRINT 31
   31 FORMAT(1HO,5X,'TIME',5X,'PREDICTION',5X,'ACTUAL',5X,' ERROR VAR
     1',5X,'ACTUAL ERROR',3X,'PRED DTEMP',3X,'PRED RES LOAD',3X,'PER LOA
     1D',5X,'Z-YP')
C
```

```
C      DEFINE ACTUAL PREDICTION ERROR
C
       DO 30 K=1,NTP
       J=I+K
       ERR(K)=Z(J)-ZV(K)
       YP(K)=ZV(K)-YV(K)
       EYP(K)=Z(J)-YP(K)
C
C      TO PRINT OUTPUT DATA
C
       K1=K+1
       L=MOD(J,24)
       PRINT 32,L,ZV(K),Z(J),EV(K),ERR(K)   ,JV(K1),YV(K),YP(K),EYP(K)
    32 FORMAT(6X,I2,6X,F8.1,5X,F8.1,6X,F10.1,5X,F8.1,13X,F5.0,8X,F8.1,5X,
      1F8.1,3X,F8.1)
    30 CONTINUE
    20 CONTINUE
       END
```

```
      SUBROUTINE STATES(X,A,B,BO,C,G,N,ZT,UT,UT1)
C
C     THIS ROUTINE ESTIMATES THE STATE OF A LINEAR SYSTEM DRIVEN BY
C     WHITE NOISE GIVEN THE PREVIOUS STATE ESTIMATE, THE SYSTEM
C     MATRICES A,B, AND C, THE GAIN G, THE SYSTEM DIMENSION N, AND THE
C     PRESENT OUTPUT AND INPUT ZT AND UT. NEW STATE IS X.
C
      DOUBLE PRECISION X,A,B,BO,C,G,ZT,UT,ZT1,X1,AX,ET1,UT1
      DIMENSION X(1),A(1),B(1),C(1),G(1),X1(20),AX(20),ZT1(1)
C
C     TO DEFINE ONE STEP PREDICTION OF X AND ZT
C
      CALL GMPRD(A,X,AX,N,N,1)
      CALL SMPY(B,UT,X1,N,1,0)
      CALL GMADD(AX,X1,X1,N,1)
      CALL GTPRD(C,X1,ZT1,N,1,1)
C
C     TO UPDATE STATE
C
      ET1=ZT-ZT1(1)-BO*UT1
      CALL SMPY(G,ET1,G,N,1,0)
      CALL GMADD(X1,G,X,N,1)
C
      RETURN
      END
```

```
      SUBROUTINE PREDIC(N,NS,NP,IT,NTP,X,P,A,B,BO,D,Q,UV,ZV,EV,YV)
C
C     THIS ROUTIN PREDICTS NTP HOURS IN ADVANCE( MAX. OF 96) AND
C     CALCULATES THE EXPECTED ERROR IN EACH PREDICTION
      DOUBLE PRECISION X,P,A,B,BO,D,Q,UV,ZV,EV,DT,AT,PP,CPC,   BU,AX,C,
     1CX
      DOUBLE PRECISION AP,APA,QD,PC
      DOUBLE PRECISION XP,UVI,YV
      DIMENSION X(1),P(1),A(1),B(1),D(1),UV(1),ZV(1),EV(1),DT(20),AT(20,
     120),PP(20,20),CPC(1)        ,BU(20),AX(20),C(20),CX(1)
      DIMENSION QD(20,20),AP(20,20),APA(20,20),PC(20)
      DIMENSION XP(20),Q(1),YV(1)
      NP2=2*NP+2
C
C     TO GENERATE ESTIMATE OF PREDIVTION ERROR
C
      CALL GMTRA(D,DT,N,1)
      CALL GMTRA(A,AT,N,N)
      CALL GMPRD(D,DT,QD,N,1,N)
      CALL SMPY(QD,Q(1),QD,N,N,0)
      DO 5 I=1,NTP
      IF(I.GT.1) GO TO 10
      CALL GMPRD(A,P,AP,N,N,N)
      GO TO 11
   10 CALL GMPRD(A,PP,AP,N,N,N)
   11 CALL GMPRD(AP,AT,APA,N,N,N)
      CALL GMADD(APA,QD,PP,N,N)
C
C     TO CALCULATE PREDICTED ERROR
C
      J=I+IT
      K=MOD(J,24)
      CALL CIT(K,NS,NP,C)
      CALL GMPRD(PP,C,PC,N,N,1)
      CALL GTPRD(C,PC,CPC,N,1,1)
    5 EV(I)=DSQRT(CPC(1))
```

```
C
C     TO GENERATE PREDICTED OUTPUT
C
      DO 20 I=1,NTP
      UVI=UV(I)
      CALL SMPY(B,UVI,BU,N,1,0)
      IF(I.GT.1) GO TO 21
      CALL GMPRD(A,X,AX,N,N,1)
      GO TO 22
   21 CALL GMPRD(A,XP,AX,N,N,1)
   22 CALL GMADD(AX,BU,XP,N,1)
      J=I+IT
      K=MOD(J,24)
      CALL CIT(K,NS,NP,C)
      CALL GTPRD(C,XP,CX,N,1,1)
      ZV(I)=CX(1)+BO*UV(I+1)
      YV(I)=XP(NP2)+BO*UV(I+1)
   20 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE CIT(IT,NS,NP,C)
C
C     THIS DEFINES C GIVEN TIME OF DAY
C
      DOUBLE PRECISION C,PI
      DIMENSION C(1)
      PI=3.14159265359
      C(1)=1.
      DO 5 I=1,NP
      J=I+1
      C(J)=DSIN(PI*DFLOAT(I*IT)/12.)
      K=NP+I+1
    5 C(K)=DCOS(PI*DFLOAT(I*IT)/12.)
      NP1=2*NP+1
      NP2=NP1+1
      N=NP1+NS
      DO 10 I=NP2,N
   10 C(I)=0.
      C(NP2)=1.0
      RETURN
      END
```

```
      SUBROUTINE DEFINE(AS,NS,NP,A)
      DOUBLE PRECISION A,AS
      DIMENSION A(20,20),AS(1)
C
C     THIS DEFINES A FROM AS PARAMETERS
C
      NP1=2*NP+1
      NP2=NP1+1
      N=NP1+NS
      DO 5 I=1,N
      DO 5 J=1,N
    5 A(I,J)=0.0
      DO 10 I=1,NP1
   10 A(I,I)=1.0
      DO 15 I=NP2,N
      A(I,NP2)=AS(I-NP1)
      IF(I.EQ.N) GO TO 15
      I1=I+1
      A(I,I1)=1.0
   15 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE SIGMA(N,M,A,C,Q,D,R,S,P,G)
C
C     THIS ROUTINE IS VALID FOR ONLY ONE OBSERVSTION , I.E. M=1, HOWEVER
C     IT COULD EASILY BE ADAPTED TO MORE THAN ONE OBSERVATION USING
C     SUBROUTINE MINV AND A FEW OTHER MINOR CHANGES
C
      DOUBLE PRECISION A,C,Q,D,R,S,P,G,SC,CTSC,SUM,SCT,GSCT,AT,DT,AP,APA
     1,DQ,DQD
      DIMENSION A(1),C(1),Q(1),D(1),R(1),S(1),P(1),G(1),SC(20),CTSC(1)
     1,SUM(1),SCT(20),GSCT(20,20),AT(20,20),DT(20),AP(20,20),APA(20,20),
     1DQ(20),DQD(20,20)
C
C     TO REDEFINE P, NEW COVARIANCE MATRIX
C
      CALL GMPRD(S,C,SC,N,N,1)
      CALL GTPRD(C,SC,CTSC,N,1,1)
      CALL GMADD(CTSC,R,SUM,1,1)
      SUM(1)=1./SUM(1)
      CALL SMPY(SC,SUM(1),G,N,1,0)
      CALL GMTRA(SC,SCT,N,1)
      CALL GMPRD(G,SCT,GSCT,N,1,N)
      CALL GMSUB(S,GSCT,P,N,N)
C
C     TO REDEFINE S, NEW ONE STEP PREDICTED COVARIANCE MATRIX
C
      CALL GMTRA(A,AT,N,N)
      CALL GMTRA(D,DT,N,1)
      CALL GMPRD(A,P,AP,N,N,N)
      CALL GMPRD(AP,AT,APA,N,N,N)
      CALL SMPY(D,Q(1),DQ,N,1,0)
      CALL GMPRD(DQ,DT,DQD,N,1,N)
      CALL GMADD(APA,DQD,S,N,N,N)
      RETURN
      END
C
C     A NUMBER OF DOUBLE PRECISION MATRIX OPERATING SUBROUTINES
```

```
C      HAVE NOT BEEN LISTED HERE SINCE THESE ARE SSP ROUTINES
C      WHOSE ONLY MODIFICATION IS AN ADDITIONAL DOUBLE PRECISION
C      STATEMENT. THESE ARE GMPRD,ARRAY,SMPY,GTPRD,GMADD,AND GMSUB.
C
```

REFERENCES

1. Ross, C. W., "Error Adaptive Control Computer for Interconnected Power Systems," IEEE PAS Trans., Vol. PAS 85, No. 7, pp. 742-749, July 1966.

2. Glover, D.J., "Control of Systems with Unknown but Bounded Disturbances - Application to Electric Power Systems," Ph.D. Thesis, Dept. Elect. Eng., M.I.T., Jan. 1971.

3. Matthewman, P. D. and H. Nicholson, "Techniques for Load Prediction in the Electricity Supply Industry," Proc. IEE, Vol. 115, No. 10, October 1968.

4. Gupta, P. C., "Load Forecasting Survey," Systems Control, Palo Alto, California, 1969.

5. Farmer, E. D. and M. J. Potton, "Developing of On-Line Prediction Technique with Results from Trials in the South West Region of the CEGB," Proc. IEE, Vol. 115, No. 10, October 1968.

6. Dryar, H. A., "The Effect of Weather on the System Load," Trans. Amer. Inst. Eng., Vol. 63, pp. 1006-1013, 1944.

7. Davies, M., "The Relation Between Weather and Electricity Demand," Proc. IEE, Vol. 106 C, pp. 27-37, 1958.

8. Heinemann, G. T., D. A. Nordman and E. C. Plant, "The Relationship Between Summer Weather and Summer Loads - A Regression Analysis," IEEE Trans. on Power Apparatus and Systems," Vol. PAS-85, No. 11, November 1966.

9. Christiaanse, W. R., "Short-Term Load Forecasting Using General Exponential Smoothing," presented at the IEEE summer Power Meeting, Los Angeles, Calif., July 12-17, 1970.

10. Toyoda, J. et al., "An Application of State Estimation to Short Term Load Forecasting"- Part I and Part II, presented at the IEEE Winter Power Meeting, New York, N.Y., 1970.

11. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME, J. Basic Eng., Series D, Vol. 82, pp. 35-45, 1960.

12. Atherton, J. R. and B. F. Hastings, "Short Term Load Forecast Study," Talk presented at the Industry-Education Conference, Detroit Edison Co., Aug. 1967.

13. Bioastronautics Data Book, NASA Special Publication, SP-3006, 1965.

REFERENCES (cont'd)

14. Schweppe, F. C., "Uncertain Dynamical Systems," course 6.606 Notes, Dept. Elect. Eng., M.I.T., 1970.

15. Jazwinski, A. H., Stochastic Processes and Filtering Theory, Academic Press, 1970.

16. Galiana, F. D., "A Review of Basic Principles and of Available Techniques in System Identification," Power Systems Eng, Group, Report No. 20, M.I.T., Nov. 1969.

17. Aström, K. J. and P. Eykhoff, "System Identification," IFAC Symposium on Identification and Process-Parameter Estimation, Prague, Czechoslovakia, June 1970.

18. Aström, K. J. and T. Bohlin, "Numerical Identification of Linear Dynamic Systems from Normal Operating Records," Proc. 2nd IFAC Symposium of Self-Adaptive Systems, pp. 76-111, 1965.

19. Kashyap, R. L., "Maximum Likelihood Identification of Stochastic Linear Systems," IEEE Trans. on Automatic Control, Vol. AC-15, No. 1, Feb. 1970.

20. Wald, A., "Assymptotic Properties of the Maximum Likelihood Estimate of an Unknown Parameter of a Discrete Stochastic Process," Amer. Math. Statist., Vol. 19, pp. 40-48, 1948.

21. Kailath, T., "An Innovation Approach to Least Squares Estimation," Part I, IEEE Trans. on Automatic Control, Dec. 1968.

22. Fletcher, R. and M. J. D. Powell, "A Rapidly Convergent Descent Method for Minimization," Computer Journal, Vol. 6, Iss. 2, pp. 163-168, 1963.

23. Personal communication with American Electric Power Service Corp., New England Electric, Detroit Edison Co., and Cleveland Illuminating Co.

24. Cleveland Hopkins International Airport Station, "Local Climatological Data," obtained from U.S. Dept. of Comm., Environmental Sciences Services Admnistration, National Weather Records Center, Federal Building, Asheville, N.C.

25. IBM System/360 Scientific Subroutine Package, Programmer's Manual (360 A-CM-03X), Subroutine DFMFP.

26. Wilkinson, J. H., The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1965.

BIBLIOGRAPHY

Chen, K. C. and P. R. Winters, "Forecasting Peak Demand for an Electric Utility with a Hybrid Exponential Model," Management Science, Vol. 12, No. 12, Aug. 1966.

Farmer, E. D. and M. J. Potton, "A Method of Load Prediction," Proc. of 3rd Intern. IFAC Congress, London, England, 1966.

Gupta, P. C., "A Statistical Approach to Peak Power-Demand Forecasting in Electric Utility Systems," presented at the IEEE Summer Power Meeting & EHV Conference, Loas Angeles, Calif., July 12-17, 1970.

Lathan, J. H. et al., "A Probability Approach to Electric Utility Load Forecasting," IEEE Summer Power Meeting, New Orleans, La., July 10-15, 1966.

Stanton, K. N., "Medium-Range, Weekly and Seasonal Peak Demand Forecasting by Probability Methods," presented at the IEEE Summer Power Meeting & EHV Conference, Los Angeles, Calif., July 12-17, 1970.

Kirchmayer,L.K.; Economic Control of Interconnected Systems John Wiley,][1958.

Kirchmayer,L.K.; Economic Operation of Power Systems, John Wiley, 1958.

BIOGRAPHICAL NOTE

Francisco Daniel Galiana was born in Alicante, Spain, September 23rd. 1944. After having lived in Spain and Brazil, he moved to Canada in 1957. There he went to high school and received the B.Eng. in Honours Electrical Engineering from McGill University in May, 1966. He received an M.I.T Fellowship the same year and in February 1968 and June 1968 respectively obtained the S.M and Eng. degrees in electrical engineering. The title of his masters thesis was " Signal Design Under Bandwidth and Energy Constraints". His interests lie in the area of applied systems and control theory. He is taking up employment at the research center of Brown Boveri & Co. in Baden, Switzerland.