Forecasting weekly electricity consumption

A case study

J. V. Ringwood, P. C. Austin and W. Monteith

This paper describes the application of time-series modelling techniques to electricity consumption data for a particular power board. Modelling is performed on total consumption, the data being available on a weekly basis with exact measurements for approximately the past 11 years. Both unforced and forced models are considered. An initial data analysis is performed to ascertain the influence of temperature and rainfall inputs on the model, and later on, a spectral analysis is used to investigate the frequency components present in the time-series data. A significant component of the determination of time-series models is the selection of an appropriate model order. Both low and high order models are evaluated, and their properties compared. For the unforced case, both AR (autoregressive) and ARMA (autoregressive moving average) models are considered. For the forced case, these model structures are extended to include ARX and ARMAX models which have one or more exogenous inputs. Such models are further extended by considering the possibility of predicting the inputs to the models, when a forecasting approach is required. Simulation results are provided for all cases together with a measure of the prediction accuracy. Comparisons are made for the various model structures, as well as models based on short and long data records and models which are driven with an external noise sequence or merely released from appropriate initial conditions.

Keywords: Forecasting; Time series; Electricity consumption

Modelling of electricity consumption using the time-series approach has received considerable exposure in the literature. A significant part of the reason for the growing interest in energy modelling is the constant rise in fuel prices. Energy utilities need to make their operation as efficient as possible, in order to offer electricity at affordable prices. To meet this objective, it is important to be able to tailor supply to demand as well as possible, and some method of consumption forecasting is necessary in order to predict future supply requirements.

This paper presents a case study on the total retail weekly electrical energy supplied by one particular power board. For reasons of confidentiality, the power board is not named, and all data have been detrended of zero and first-order components. A variety of model structures are considered, and models are identified under a variety of conditions, for example, different length data records, different model orders and different driving functions. For the particular situation being considered, the only inputs for which data were available were rainfall and temperature.

The MATLAB package [3] (with the *Identification Toolbox* [2]) was used to evaluate the models and MATLAB macros written to perform the predictions for the various model types.

The paper is organized as follows. First the (detrended) data are presented and cross-correlated with prospective inputs to determine the significance of the inputs. Then unforced models for electricity consumption are identified, with significant attention paid to model order selection, length of data record and the determination of suitable initial conditions for

The authors are with the Department of Production Technology, Massey University, Palmerston North, New Zealand.

Final manuscript received 10 August 1992.

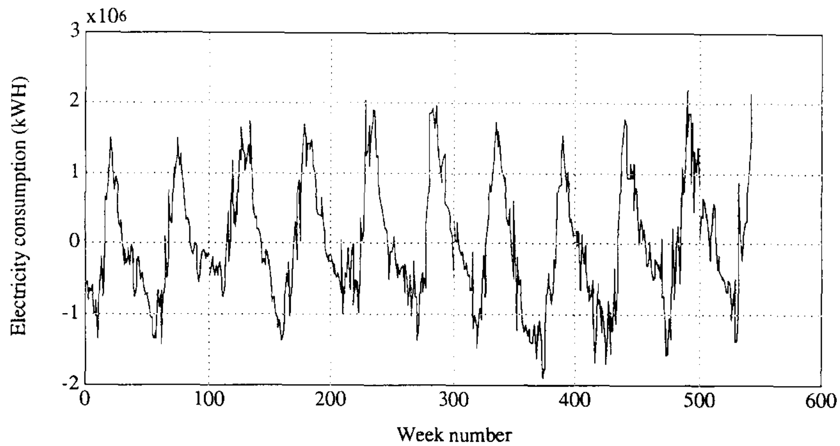

Figure 1. Complete detrended data record.

the models. In the next section, forced models are considered and then we attempt to produce (AR) models for the inputs, allowing the forced models to be used for forecasting. In the next section there is an in-depth discussion of the results and an overall comparison of the models and techniques used. Finally, we draw conclusions from the analysis and make some recommendations based on the experiences with this particular case study.

Data preparation

Weekly consumption data are available from 1980 to 1991, a total of 543 points. The data are detrended to remove constant and first order components. This helps to ensure that the identification algorithms can produce parameter estimates which are unbiased. The complete detrended data record is shown in Figure 1. Note the cyclic variations in the data. These occur not only because of annual temperature profile variations, but also because of the cyclical consumption patterns of seasonal industries. The first 491 data points will be used for system identification and the remaining 52 (corresponding to a year's duration) used for validation purposes. The objective for the current exercise is to produce a projection or forecast of future weekly consumption a year in advance. If inputs to the system are considered, then scenario testing is possible, ie the effect of certain (perhaps known) variations in some input variable on the electricity consumption may be examined. If it is required to do forecasting with a model which has deterministic inputs, then future values for the inputs must be available. Where future values are not available, it may be possible to estimate them, by generating a model for the input. Such a situation is examined in the section on prediction of temperature inputs.

The two inputs which are available for the current analysis are average temperature and rainfall. Since average temperature is available on a daily basis, a daily heating degree day (UK Department of Energy [5]) (HDD) figure may be evaluated and accumulated to provide a weekly HDD figure. The base temperature used in the HDD calculations is 18 °C. Obviously, HDD variations have a significant influence on the consumption pattern and this is verified by the cross-correlation function shown in Figure 2.

Rainfall is considered by some to psychologically influence people to turn on heaters, but the cross-correlation function in Figure 3 does not indicate that this is the case for the current situation.

Based on Figures 2 and 3, HDD18 or average weekly temperature will be the only input to be given further consideration. It may also be noted that some predictions were performed using actual rainfall data; however, no discernible improvement in the prediction was obtained by using this extra input.

Unforced models

In this section, AR and ARMA models for electricity consumption will be identified. The structure of these models is as follows:

$$\text{AR:} \quad A(q)y(t) = e(t) \tag{1}$$

$$\text{ARMA:} \quad A(q)y(t) = C(q)e(t) \tag{2}$$

where

$$A(q) = 1 + a_1 q^{-1} + \ldots + a_{na} q^{-na} \tag{3}$$

$$C(q) = c_0 + c_1 q^{-1} + \ldots + c_{nc} q^{-nc} \tag{4}$$

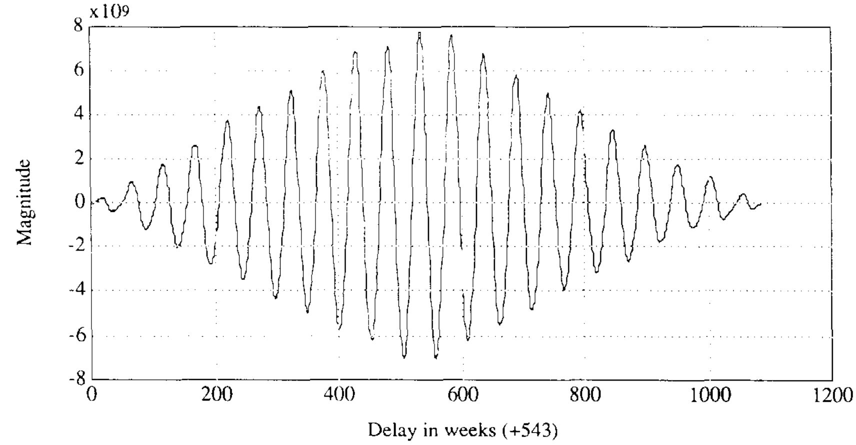

Figure 2. Cross-correlation of HDD18 with consumption.

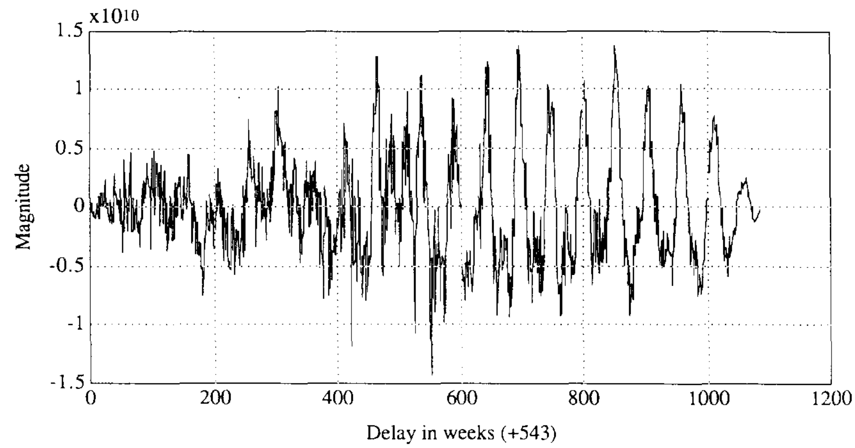

Figure 3. Cross-correlation of rainfall with consumption.

q^{-1} is the delay operator and $e(t)$ is a white noise sequence. In the case of the AR model, a prediction may still be obtained even if no driving noise sequence is used by running the model from non-zero initial conditions. Indeed, for all models, it is important that appropriate initial conditions (corresponding to previous actual consumption) are used, so that the correct starting point is achieved. For the ARMA case, omission of $e(t)$ defaults to the AR model.

To set up the identification problem, the orders of the A and C polynomials (na and nc) must be determined. To assist with this choice, the loss function is plotted for different values of na and nb. The loss function provides a measure of the mean square difference between the model output and the actual consumption for a particular model structure, for zero initial conditions. The loss function is plotted in Figure 4 for the AR term.

The objective is to pick polynomial orders which are sufficient to describe the system. From Figure 4, note how the loss function for the AR term decreases rapidly until order 10 is reached. No significant improvement is apparent until order 24 is reached; na equal to 10 would therefore seem to be one possible choice. Another possible choice is order 53. This has the added significance of corresponding to a year's duration, coinciding with the natural cycle in the data. A popular criterion to use in model order selection is Akaike's information theoretic criterion (AIC) (Ljung [1]), which is order weighted (ie it penalizes higher orders). Selection under this criterion returns an order of 53. Setting $na = 10$ gives a selection of 7 for nc (nc must be $<= na$) and if $na = 53$, then $nc = 17$ is a suitable choice.

For comparison purposes, both low order and higher order AR models are evaluated. A least squares technique (Soderstrom and Stoica [4]) is used to perform the system identification. The parameter

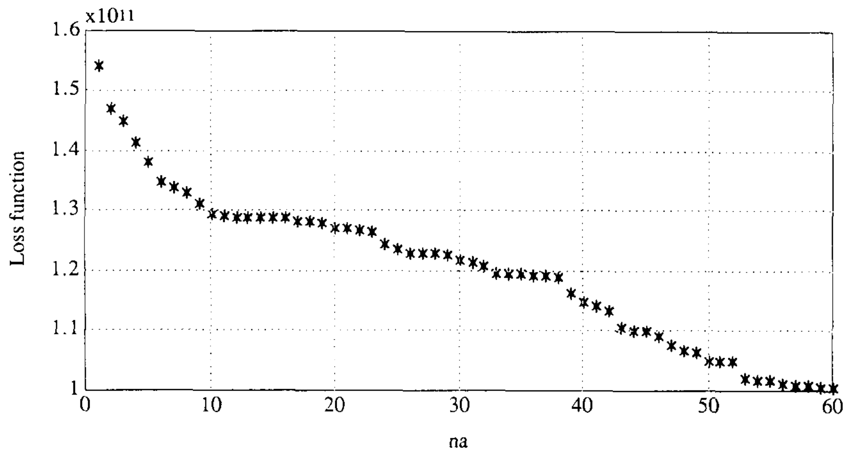

Figure 4. Model order selection for AR term.

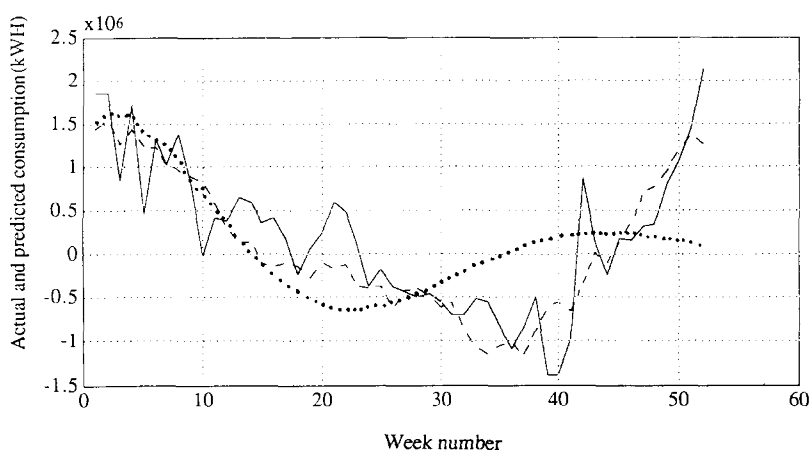

Figure 5. Predictions from initial condition driven AR models. Key: —— Actual, --- order 53 model, ··· order 10 model.

values for the 10th order are as shown in Table 1. The bottom row in Table 1 give the standard deviation of the estimates. Predictions from both 10th and 53rd order AR models are now evaluated and compared against the actual consumption for the validation period of 52 weeks. For systems modelled with AR models,

$$A(q)y(t) = e(t) \tag{5}$$

where the term $e(t)$ represents the modelling error, that is, the difference between the data and the model prediction, $A(q)y(t)$. When, as here, the sequence $e(t)$ is a white noise sequence, it is entirely unpredictable. Consequently, there is no modelling advantage in using a white noise driving sequence, $e(t)$, in conjunction with the AR term $A(q)y(t)$, in making predictions. Accordingly, Figure 5 shows the predictions obtained by releasing the model from initial conditions only.

This means that if enough trials are taken, the average of the responses due to a driving noise

Table 1. Parameter values and standard deviations for the 10th order model.

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
−0.6088	−0.1674	−0.2564	−0.0031	0.0166	0.0545	−0.0077	−0.0174	0.0399	0.1156
0.0452	0.0535	0.0540	0.0553	0.0552	0.0553	0.0560	0.0550	0.0544	0.0463

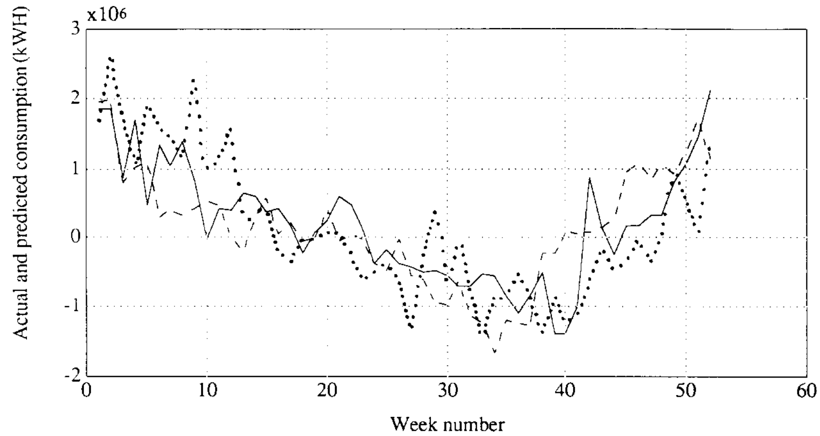

Figure 6. Predictions from ARMA models. Key: —— Actual, --- $nc = 1$, ··· $nc = 20$.

sequence will be the same as that obtained from initial conditions alone. The response obtained from a single trial will deviate from y_{ic}, but the deviation will be purely random, offering no improvement to the prediction.

Some comments on the AR models are pertinent at this point. It is clear from Figure 5 that the 53rd order model produces superior predictions to the 10th order model. This may be attributed to two factors:

(i) The higher order model has the increased complexity required to provide the many inflections in the consumption profile.
(ii) The higher order model is initialized with the full previous year's consumption profile.

The extra computational burden (identification and simulation) associated with the higher order model is not important, since real-time operation is not required.

The possibility of using ARMA models is now examined. Based on the foregoing analysis, it would seem reasonable to retain a 53rd order AR term. Unfortunately, plotting the loss function for a white noise driving sequence does not indicate any outstanding choices for the order of the MA term, and the AIC suggests an order of 1. Due to this uncertainty, a number of ARMA models will be evaluated and their predictions compared. The following orders for the MA term will be examined: 1, 5, 10, 20, 32 and 53, covering the available range. Note that order zero corresponds to an AR model.

The results for $nc = 1$ (dashed line), $nc = 20$ (dotted line), and the actual consumption (solid line) are shown in Figure 6. Since the system is now subject to a stochastic input, these predictions are evaluated as an average of 100 trials with different driving noise sequences. From observation, little benefit seems to be obtained from use of the higher order MA terms. This is verified in the quantitative comparisons in the section analysing the results, where it is confirmed that, of the orders tested, $nc = 1$ gives the best predictions. The ARMA predictions, however, do not offer any improvement over the AR results.

Forced models

In this section, models with deterministic (and possibly stochastic, as well) inputs will be considered. The objective of these models is to include information on certain variables which are known to have a strong influence on electricity consumption. While AR and ARMA models can only forecast the 'predictable' (or regular) variations in consumption, models with deterministic inputs have the capacity to show some 'unpredictable' variations in consumption, where the unpredictability is somehow reflected in the input signals. This property is especially useful in scenario testing, where the effects of global warming (for example) can be analysed in advance. In many cases, of course, future values for these external inputs will not be available. This situation is addressed to some degree in the section on prediction of temperature inputs, where the benefit of attempting to predict the unknown future input values is examined. Two model structures will be used:

$$\text{ARX:} \quad A(q)y(t) = B(q)u(t) + e(t) \tag{6}$$

$$\text{ARMAX:} \quad A(q)y(t) = B(q)u(t) + C(q)e(t) \tag{7}$$

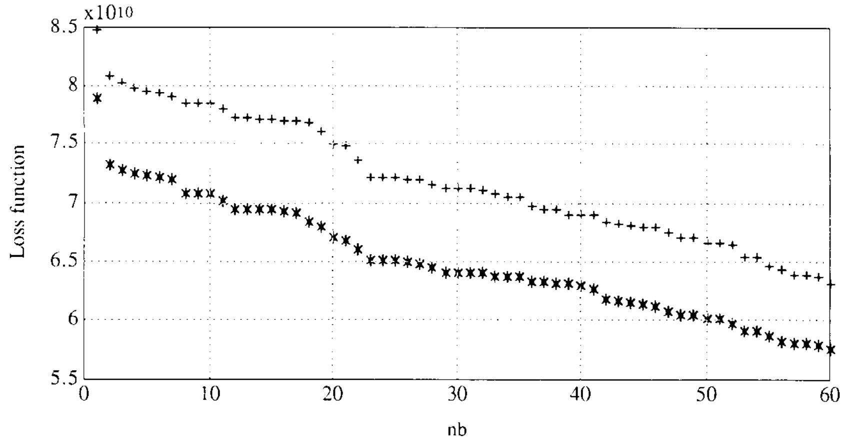

Figure 7. Order selection for $B(q)$ polynomial. Key: * * * HDD input, + + + Average temperature input.

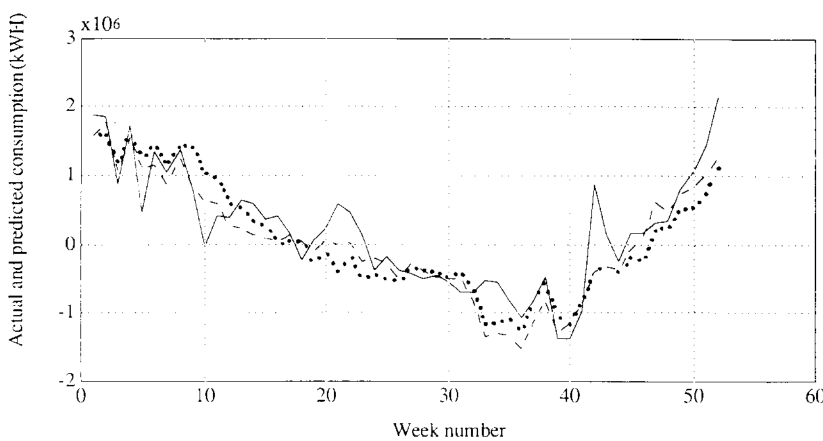

Figure 8. Predictions from HDD ARX models. Key: —— Actual, --- $nb=2$, ··· $nb=23$.

where

$$B(q) = b_0 + b_1 q^{-1} + \ldots + b_{nb} q^{-nb} \tag{8}$$

As with the unforced models, $e(t)$ can be either omitted or retained. From the analysis in the previous section, it is seen that no advantage is to be gained in including $e(t)$ in the ARX model. A non-zero $e(t)$ will be used in the ARMAX model, since setting $e(t)$ equal to zero would result in a default to the ARX model.

The first step in evaluating the ARX and ARMAX models is to select orders for the polynomials $A(q)$, $B(q)$ and $C(q)$. The order of the $A(q)$ polynomial will be retained at 53 for this analysis. This choice was well justified in the section on unforced models, and similar model orders will allow a more meaningful comparison of the performance of the forced models with the unforced models. The order of the C polynomial as determined in that section will also be retained for comparison purposes. The order of the B polynomial still remains to be determined.

Two possible (and mutually exclusive) inputs are available – heating degree day (HDD) data (accumulated over a week) or weekly average temperature. The effect of using both inputs will be examined.

Figure 7 shows the variation in loss function for an ARX model with increasing nb for the HDD (*) and the average temperature (+) inputs. Both plots are similar in profile, with the loss function being significantly lower for the HDD input. This gives an initial indication that HDD is the superior input to use. Choice of nb, however, is not quite as straightforward.

Two possibilities which will be examined are $nb = 2$ and $nb = 23$. Unfortunately, application of AIC returns a value of $nb = 60$ which is unusable, since nb must be less than na. Figure 8 shows the predictions from the HDD ARX models. The result from the model with the 2nd order $B(q)$ polynomial is given by the dashed line and that from the 23rd order $B(q)$ given by the dotted line. The solid line indicates the

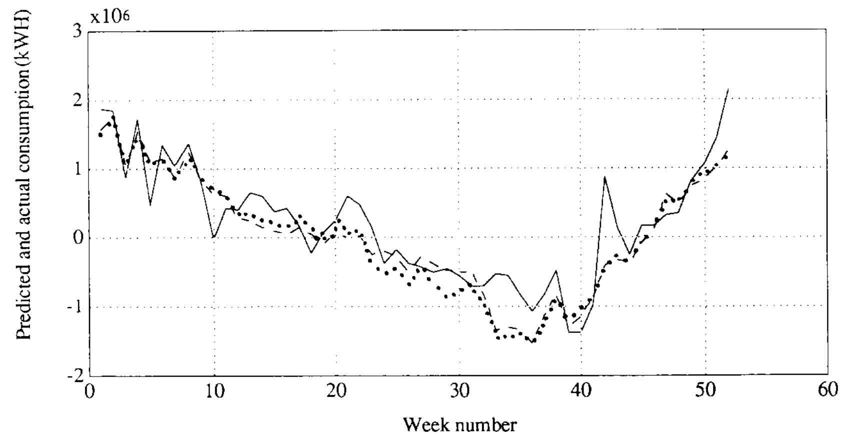

Figure 9. Comparative results from HDD and average temperature inputs. Key: —— Actual, --- HDD input, ··· Average temperature input.

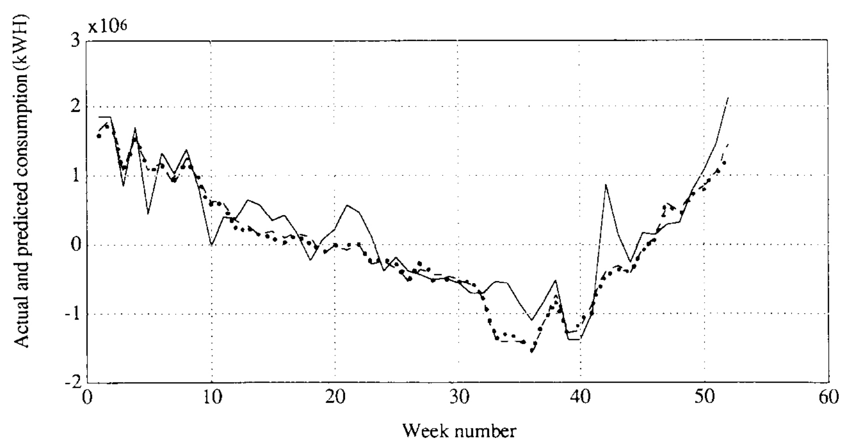

Figure 10. Comparative results for ARMAX and ARX models. Key: —— Actual, --- ARMAX model, ··· ARX model.

actual consumption for that period. From observation, it would appear that the 2nd order $B(q)$ polynomial performs slightly better than the 23rd order one (see the section on analysis of results). This, combined with the issue of increased complexity, are grounds for discarding the higher order choice at this stage.

Note that a similar result was obtained by comparing low and high order predictions from the average temperature input. Figure 9 compares the predictions for HDD and average temperature ARX models with 2nd order $B(q)$ polynomials. The HDD model performs slightly better than the average temperature model (again see section on analysis of results), confirming the initial indications from Figure 7.

Finally an ARMAX model is considered. From previous model order analysis in this and the previous section, the model orders will be chosen as: $na = 53$, $nb = 2$ and $nc = 1$. The model parameters are identified using a prediction error method, which is superior when stochastic models are used. The predictions from this model are shown in Figure 10 (dashed line) with the predictions from the HDD ARX model (with $nb = 2$) (dotted line) and the actual consumption (solid line) shown for comparative purposes. From the plot, it is seen that there is little difference between the predictions from the ARX and ARMAX models, the ARMAX being slightly better. This will be confirmed by the quantitative comparison in the results section. Note that the ARMAX prediction was evaluated as the average over 100 trials using different random noise inputs.

Prediction of temperature inputs

From the sections on unforced and forced models, it was seen that notable improvement in the quality of the predictions is possible when appropriate input signals are used. However, the difficulty of providing future values of input signals makes ARX and

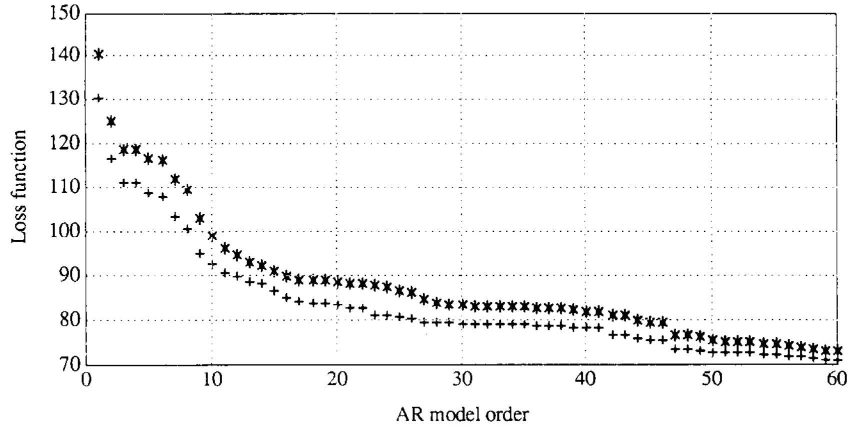

Figure 11. Model order selection for HDD and average temperature models. Key: * * * HDD, + + + Average temperature.

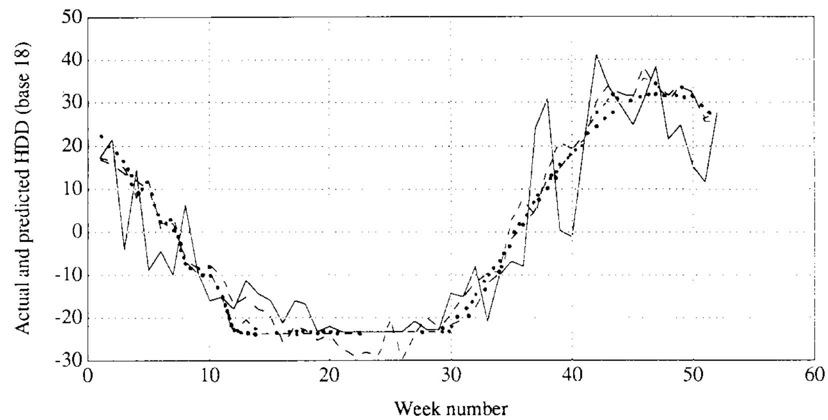

Figure 12. Prediction of detrended HDD input. Key: —— Actual, - - - HDD model, ··· AT27 model, -·- AT52 model.

ARMAX models difficult to use for forecasting purposes. In an effort to retain the effectiveness of these models and extend their use to the forecasting situation, this section will examine if any benefit is obtained from using models for the inputs (ie predicting the inputs). This analysis will be restricted to temperature inputs, since this is the limit of data availability.

From the section on forced models, it may be concluded that the most effective temperature input is HDD. However, due to the truncations inherent in HDD data (no negative values are allowed), it may be more beneficial to predict average temperature and then convert to HDD values. Therefore, AR models for both HDD and average temperature will be determined. Figure 11 shows the variation in loss function for various orders of HDD (*) and average temperature (+, scaled up by a factor of 35 for comparison) models.

Application of the AIC returns orders of 27 and 52 for the average temperature and HDD models respectively. However, considering that the loss function profiles are very similar and the use of different model orders would make a direct comparison difficult, both model orders will be chosen as 52.

Figure 12 shows the actual (solid line) and predicted HDD from the HDD model (dashed line) and the 27th (dotted line) and 52nd (mixed line) order average temperature models. Detrended HDD is evaluated from the average temperature models by predicting detrended average temperature, retrending the prediction, conversion to HDD (with a suitable base temperature) and finally detrending the HDD data.

One difficulty, however, exists with the conversion of weekly average temperature values to HDD figures. HDD is evaluated on a daily basis, with the weekly value being the cumulative sum of the seven daily

Table 2. The mean square fit for different temperature models.

Prediction	AT model (27)	AT model (52)	HDD model	HDD model (T)
MSF	9.2738	9.0235	9.1147	8.7757

values. Daily HDD can be approximately evaluated as:

$$HDD_{\text{base temp}} = \text{average temp} - \text{base temp}$$

$$\text{if } HDD_{\text{base temp}} < 0, \quad HDD_{\text{base temp}} = 0 \qquad (9)$$

Note the non-linear relationship between HDD and average temperature. It is easy to see that if HDD is accumulated over a week and an attempt is made to evaluate weekly HDD using the above relationship, where the average temperature value in the above equation is a weekly one, then a discrepancy will exist between the two weekly HDD values, due to the non-linearity.

One point which arises is that while many discontinuities may exist with daily HDD data (a significant number of zero elements), it is less likely that weekly HDD data will contain very many zeros, ie it is not too likely that daily HDD will be zero for a full week. This may mean that weekly HDD may be reasonably straightforward to model, with no recourse required to an average temperature (AT) model.

Note from Figure 12 that the prediction from the HDD model is not restricted in its minimum value. An improved prediction may be obtained by retrending the sequence, truncating the lower values at zero and subsequently retrending the truncated sequence. The mean square fit (MSF) for each prediction is compared in Table 2. The mean square fit is evaluated as:

$$MSF = \text{norm}(yp - y)/\sqrt{\text{length }(y)} \qquad (10)$$

where yp is the predicted value of the variable y.

The ARMAX model will now be used to evaluate a prediction of consumption using the best predicted HDD input, which comes from the truncated version of the HDD model output. The consumption prediction is shown in Figure 13 (dashed line), with the actual consumption (solid line) and the prediction from the ARMAX model with actual HDD input (dotted line) shown for comparison. No significant degradation in prediction is observed when the predicted HDD input is used. This result is quantified in the following section, along with the result for an ARMAX model with a predicted average temperature input. This section will attempt to present quantitative comparative results for the various models and also examine the spectral properties of the estimates. In addition, some attention will be devoted to looking at the effects of different length identification sequences and a recursive analysis to examine the consumption data for stationarity.

Table 3 examines the mean square fit (MSF, as given in Equation (10)) for the various predictions obtained. It would appear that the best model, using the MSF as a criterion, is the ARMAX model with actual HDD input (model 12). However, if a forecast

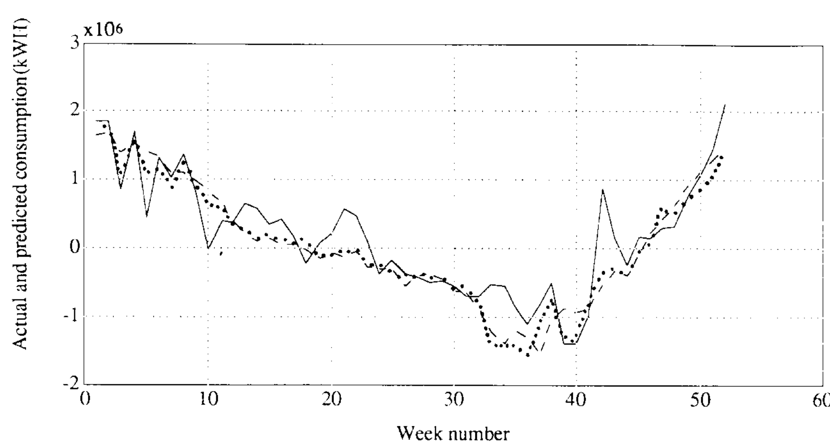

Figure 13. Output of ARMAX model with predicted HDD input. Key: —— Actual consumption, - - - Predicted HDD input, ··· Actual HDD input.

Table 3. The mean square fit for the various predictions obtained.

Model type	*na*	*nb*	*nc*	Input	MSF
(1) AR	10	—	—	—	7.1361*e* + 05
(2) AR	53	—	—	—	4.3271*e* + 05
(3) ARMA	53	—	1	—	4.3912*e* + 05
(4) ARMA	53	—	5	—	4.8173*e* + 05
(5) ARMA	53	—	10	—	4.9043*e* + 05
(6) ARMA	53	—	20	—	8.8201*e* + 05
(7) ARMA	53	—	32	—	5.2562*e* + 05
(8) ARMA	53	—	53	—	6.1490*e* + 05
(9) ARX	53	2	—	Actual HDD	3.8394*e* + 05
(10) ARX	53	23	—	Actual HDD	4.4403*e* + 05
(11) ARX	53	2	—	Actual AT	4.0834*e* + 05
(12) ARMAX	53	2	1	Actual HDD	3.7842*e* + 05
(13) ARMAX	53	2	1	Predicted HDD	4.2771*e* + 05
(14) ARMAX	53	2	1	Actual AT	3.9682*e* + 05
(15) ARMAX	53	2	1	Predicted AT	4.3224*e* + 05

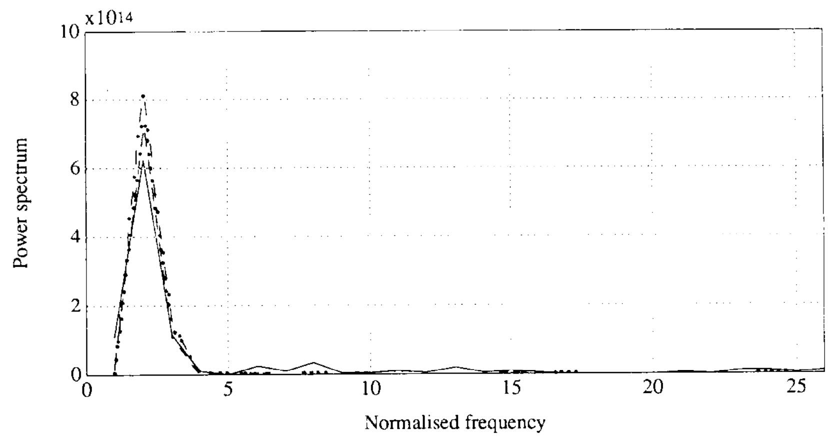

Figure 14. Spectral measures of data and predictions. Key: —— Actual, – – – Model 2, ··· Model 5, –·– Model 9.

is required (ie the future values of the input are not available), then the best option would seem to be the ARMAX with predicted HDD input (model 13), closely followed by the AR model (model 2).

The MSF provides a time domain fit criterion. Further insight into the quality of predictions from the various models may be obtained by taking the analysis into the frequency domain. Figure 14 shows the power spectra for the actual data (solid) and models 2 (dashed), 5 (dotted) and 9 (mixed line). Note the appearance of a dominant low frequency component corresponding to the annual cycle in electricity consumption. It is seen that this component is faithfully reproduced by the models represented in Figure 14.

A further analysis was undertaken to examine the effect of using different length identification sequences. An AR model (order 53) was evaluated for eight different cases, ranging from use of (the most recent) two years' identification data (104 points) to nine years' data (468 points). The (time domain) mean square fit was then evaluated for each of the model outputs and is plotted in Figure 15. This would seem to indicate that three years of data is the realistic minimum data length for the current case. Finally, a recursive identification test is performed to examine for drift etc in the model parameters. A forgetting factor of 0.9 was used to allow variation in the parameters. Figure 16 shows the variation in the first four parameters of a 53rd order AR model. No apparent drift in the parameter estimates is observed.

Conclusions

For the case under consideration, time-series analysis has been shown to be a viable and useful tool in the modelling of electrical energy consumption. For the best case observed (model 12), the mean percentage error in the consumption estimates over a year was 2.3% for the retrended data.

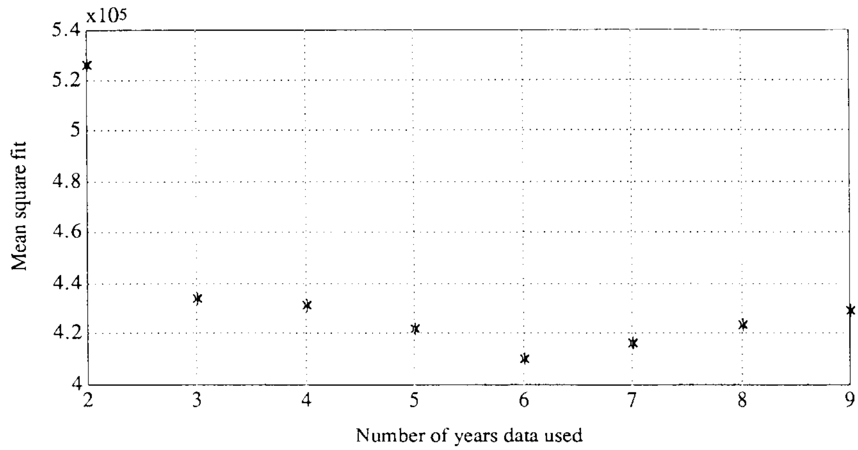

Figure 15. Effect of different length identification sequences.

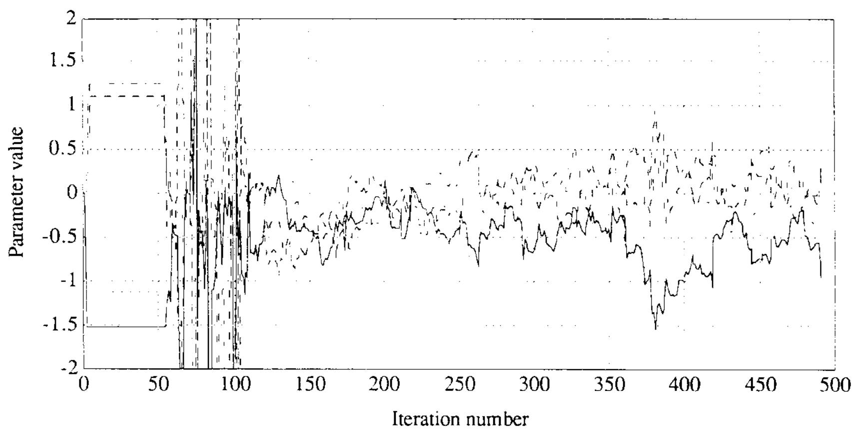

Figure 16. Recursive parameter calculations. Key: —— Parameter 1, – – – Parameter 2, ··· Parameter 3, –·– Parameter 4.

Within the availability of data, a reasonable variety of model types and orders were investigated. The ARMAX model generally proved best, with the AR model doing surprisingly well, considering its simplicity. Both from intuitive (annual cycle) and analytical (AIC) reasoning, an order of 53 was found most suitable for the AR term. In general, little benefit seemed to be obtained in adding a coloured noise model, since the predictions from the AR model were superior to the ARMA case, and although the ARMAX model was slightly better than the ARX one, the difference was not great. The use of the external input (HDD or average temperature) improved the prediction, and in cases where future values of the input are not available (corresponding to the forecasting case), it would still seem to be best to use an ARMAX model with a predicted HDD input. For a well behaved (almost periodic) input such as HDD or average temperature, it would seem to be straightforward to estimate future values of the input. This may not be the case with other inputs eg production data.

It would appear that three years' data is sufficient to estimate a good weekly model. No significant advantage was seen to be gained in the current example of using extra data. However, no disadvantage was noted, either, since no appreciable drift in the model parameters with time was observed. Note that the detrending operation tends to remove the lower frequency components at the start (dc and linear components are removed). The model then concentrates on relatively high frequency components remaining in the detrended data, with the predicted data from the model being subsequently retrended to reinstate the correct low frequency variations.

Overall, the benefits of time-series modelling for this particular example are clear. Good predictions may be obtained even without procuring external input data, with a mean percentage error of only 2.7% for the AR model. This follows a straightforward procedure of detrending, identification, simulation and subsequent retrending, using only consumption data. With appropriate software, this exercise may be completed in approximately five minutes, giving good predictions of weekly consumption a year in advance.

References

1 L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, 1987.

2 L. Ljung, *System Identification Toolbox (for use with MATLAB^{TM})*, The Math Works Inc, 1991.

3 *Matlab^{TM} for Macintosh Computers*, The Math Works Inc, 1991.

4 T. Soderstrom and P. Stoica, *System Identification*, Prentice-Hall, 1989.

5 UK Department of Energy, *Fuel Efficiency Booklet 7: Degree Days*, 1981.