Journal of the Operational Research Society (1997) 48, 15–24

One day ahead demand forecasting in the utility industries: Two case studies

R Fildes[1], A Randall[2] and P Stubbs[3]

[1]*Lancaster University,* [2]*Independent consultant, Gloucester and* [3]*Mercury Communications, Manchester*

This paper describes two case studies of short-term demand forecasting for the utilities, water and gas, linked to earlier research in similar contexts. In both cases the forecast of demand has important consequences for the operations and control of productive capacity. It is shown that in these two cases extrapolative methods based on the past data history alone are outperformed by more complex multivariate approaches that include information on the effects of weather. The paper concludes with a discussion of how an organization with an important short-term forecasting problem should go about selecting an appropriate forecasting method.

Keywords: forecasting; gas; water; econometrics

Introduction

Short term forecasting in most industries requires the forecaster to process many data series. The forecasts are used to schedule operations and no single series is particularly important to the organization. As a consequence extrapolative forecasting methods have been widely adopted[1]. In contrast, the utilities have a limited number of series relevant to their efficient operations and consequently substantial effort has been expended on forecasting, particularly in electricity[2]. This paper describes and integrates two case studies, based on MSc projects undertaken for British Gas North Western[3] and Thames Water[4] which build on earlier research in the utilities concerned. The objectives of the case studies were similar: to evaluate various alternative methods of forecasting for possible implementation. This research comes to the conclusion that information on relevant explanatory variables, when combined appropriately with the dynamics of the demand process leads to improved levels of forecast accuracy relative to extrapolative techniques. However the benefits depend on both the utility concerned and the accuracy of the weather forecast.

The paper first discusses the need for demand forecasting in the utilities. The next section is concerned with forecasting accuracy and the choices facing an industry forecaster. The paper goes on to consider the data base available in the two case studies. The various models and their accuracy are then described. Finally, conclusions are drawn concerning how these utilities should appraise their forecasting procedures, and the relevance of these results for forecasting research.

Correspondence: R Fildes, Department of Management Science, The Management School, Lancaster University, Lancaster LA1 4YX.

The need for demand forecasting

The privatization of the utility industries in recent years has forced them to reappraise their profit margins. The prices these new privatized companies are able to charge are constrained by government regulatory bodies. To counter the possible erosion of their profit margins, the companies, instead have looked to minimize their costs. On a daily basis, the companies can become more efficient by accurately predicting demand with a consequent reduction in storage and distribution costs.

In the water industry, privatization has forced the water companies to make profits within the confines of their statutory responsibility to provide a constant supply of water to the consumers. The development of the London Ring Main has enabled Thames Water to monitor the supply of water from the London Water Control Centre. The Centre is able to ensure a constant water pressure is maintained by transferring water quickly through the Ring Main. The Ring Main, though, is dependent on the reservoirs around London. Daily forecasts are used to ensure that the amount of water held in the reservoirs and the associated collection and distribution costs are kept to a minimum.

Within British Gas North Western, Grid Control use day ahead demand forecasts to calculate the most efficient intake from the National Transmission System into the regional system. Regional storage can thus be effectively managed and supply and demand smoothly matched and thus more accurate forecasts lead to reduced costs.

The electricity generation industry[5] in England and Wales was fundamentally restructured in 1990/91 from being a single public utility, CEGB (Central Electricity Generating Board). The result was two privatized companies, National Power and PowerGen, and a publicly-owned company Nuclear Electric which retained all the Nuclear capacity. Further competition in the new marketplace also comes from other suppliers, together with possible new entrants. To operate the market, the National Grid Company was set to take over the transmission of electricity and, hence, has the responsibility for secure supply of electricity and the operation of a daily power pool. The 12 regional distribution and supply companies were also privatized to allow them to compete independently from the generating companies.

To minimize the risk arising from large price fluctuations much of the buying of electricity was done on long term contracts although shorter contracts (based on half-hourly forecasts) are expected to become more commonplace as the market stabilizes. Therefore, both the generator companies and the distribution/supply companies need to forecast short-term demand.

The choice of forecasting methods in practice

The daily forecasts of demand in the utility industries described above are important in the day-to-day operations and control of each business. Consequently, in each of the industries there is a *prima facie* case for considering a wide variety of potentially complex forecasting methods, if the result is improved accuracy. There is no 'best' method of forecasting, even when the problem is as tightly defined as in these utilities[6].

Various methods have been proposed for forecasting the demand for gas; Berrisford[7] describes a regression based forecasting application. Lyness[8] examined the effects of a severe winter. Transfer function modelling, a multivariate extension of the univariate Box-Jenkins ARIMA methodology was used by Borgard *et al*.[9], Taylor and Thomas[10], and Piggott[11]. The expertise of Grid Control engineers has been incorporatted into an expert system which selected a forecasting model from a choice that included a Bayesian model, regression and a combined model[12]. The expert system could also make a subsequent adjustment to the chosen model's forecast. However, from the perspective of British Gas NW these various models had not been thoroughly evaluated as to their forecasting performance and therefore this earlier research was seen as offering only limited guidance as to the appropriate methodology to adopt.

Short-term forecasting of water demand has gathered little attention apart from Sterling and Bargiela[13]. The bulk of the published work has been carried out analyzing electricity forecasts. (See Bunn and Farmer[14] for a summary). Engle *et al*.[15,16] have shown that short-term fluctuations in demand have two primary causes: seasonality (within the week, or month) and changes in the weather. Influences such as price and income are likely to have a longer term effect. These studies have conclusively shown that extrapolative methods (which include seasonality but not explicit weather variables) are outperformed by suitably specified explanatory models.

Within British Gas a wide range of models had been tried and claims made for their performance characteristics. Within Thames Water the ability to control the flow of water was new and therefore the need for short-term forecasting was new. Both project briefs therefore required the evaluation of a range of methods. Methods chosen for inclusion were

- Exponential Smoothing (on the grounds of its strong performance in so-called forecasting competitions)[17].
- Box-Jenkins ARIMA methods.
- Transfer function modelling.

An approach that is sometimes adopted when forecasting electricity demand is to make an adjustment to demand based on the weather. There is no unique method for accomplishing this and performance would typically be worse than the more general regression methods discussed here.

Performance measures

It is essential when evaluating a forecasting method to consider, not just standard statistical measures of the method's (within-sample) fit, but also its stimulated performance when forecasting. Fildes and Howell[18] demonstrated that there is little correlation between models that best fit the data and the models that will produce the best results when actually forecasting, arguing that model performance should be compared by using *ex ante* testing. When explanatory variables are included in the model such *ex ante* comparisons are implicitly conditional on the accuracy of forecasts of the explanatory variables. Typically, forecast accuracy deteriorates sharply when these forecasts are used[19,20]. As a consequence, an out-of-sample *ex post* comparison should also be carried out. This establishes whether forecast inaccuracy arises from model mis-specification (and randomness) or from the failure to forecast the explanatory variables correctly.

Within the above framework it is necessary to decide on appropriate measures with which to compare alternative methods. Ideally, where forecasts are used directly as a key input into a decision, the cost consequences of the errors should be measured although these are seldom available. Instead various alternative measures of forecast accuracy are used. Because there are many different criteria which may be used to evaluate the measures[21,22] such as its ease of interpretation in the problem situation, its reliability etc no ideal single measure is available. Different forecasters and the forecast users will have their own preferences based on these alternative criteria. The overall aim when choosing an error measure in a practical application is, therefore, to recognize how the costs of forecast inaccuracy are reflected in the distribution of forecast errors and make the choice accordingly.

Unsurprisingly the two utilities had different preferences. In so far as both problems under study are only concerned with a single data series, the choice of measures for comparing models is not so problematic as in the more typical case where there are many series to evaluate. The measures chosen in Thames Water were

- Mean Absolute Deviation (MAD): to evaluate the average error regardless of sign.
- Root Mean Square Error (RMSE): to place more emphasis on large errors.
  (With only one data series, the error measures can be scale dependent.)

In addition, because these measures do not fully reflect the error distribution according to the standard laid down in Armstrong and Collopy[21], we also discuss the overall shape of the error distribution. In British Gas, it was regarded as important to distinguish between those errors arising from the method under analysis and those due to poor forecasts of the explanatory variables. Two different error measures were already used in British Gas NW to measure the daily performance of their model. Total % Error shows how each daily forecast actually performed (using forecast and estimated data), and Unexplained % Error shows how the model would have performed had actual weather data been known when the forecast was made. These errors are defined as follows for each day forecast:

$$\text{Total \% Error} = \text{Actual Demand} - \text{Forecast Demand (using forecast explanatory variables)} / \text{Actual Demand}$$

$$\text{Unexplained \% Error} = \text{Actual Demand} - \text{Forecast Demand (using actual explanatory variables)} / \text{Actual Demand}$$

and the mean absolute percentage error (MAPE) calculated in both cases. In addition, extreme errors may be regarded as particularly important for utility forecasting because of legally enforceable service requirements. Therefore the percentage of large absolute errors was also calculated for British Gas North Western. As in the case of Thames Water we comment on the shape of the error distribution.

Preliminary data analysis

Thames Water

Short-term fluctuations in water consumption were thought to be primarily caused by the weather, and any restrictions on demand such as hose pipe bans. Usage itself was calculated from the main sources of supply. Reservoirs act as a buffer between supply and demand, so the following equation was defined to evaluate daily demand.

$$\text{Daily demand} = \text{Total daily supply} - \text{Reservoir usage}$$

where

$$\text{Reservoir usage} = \text{Today's reservoir volume at 8 am} - \text{Tomorrow's reservoir volume at 8 am}$$

Possible causes of inaccuracies in the data were

- The supply and reservoir readings were taken at different times during the day. Staff did not consider this to be a major cause of inaccuracy.
- The reservoir information covered 74% (by volume) of the reservoirs in London. It was not possible to predict the remaining 26% of reservoirs due to the unpredictable nature of their operation. Therefore all reservoir usage figures were under-estimates.
- In most circumstances, leaks in the pipe network can reasonably be assumed to be constant and therefore not significant. However, large leakages caused by thaws were significant.
- Supply constraints such as hose pipe bans were not thought to be a problem during the period studied.
- Some limited interpolation was required to estimate missing reservoir data.

Initially data were available for the period spanning 1 November 1990 to 5 April 1992 (522 data points), which was later extended to the 30 June 1992. A graph of the demand data is given in Figure 1. An initial interpretation of the graph provided the following observations

- A decrease in demand can be seen to occur over the Christmas periods.
- Good weather and the thaw (February 1991) are possible causes for observed increases in demand.
- No significant long term trend can be identified from the graph.

A simple analysis of daily seasonality using a multiplicative decomposition approach showed a clear drop in demand on Friday and Saturday. Possible reasons for the pattern are a shutdown of industry on Friday for the weekend and consumer behaviour at weekends that may differ from usual weekday activities.

British Gas NW

A forecast sendout, the amount of gas distributed from the National Transmission System (NTS) and Regional storage holders, over the twenty-four hour period starting at 0600 hrs on day t is needed to manage the National

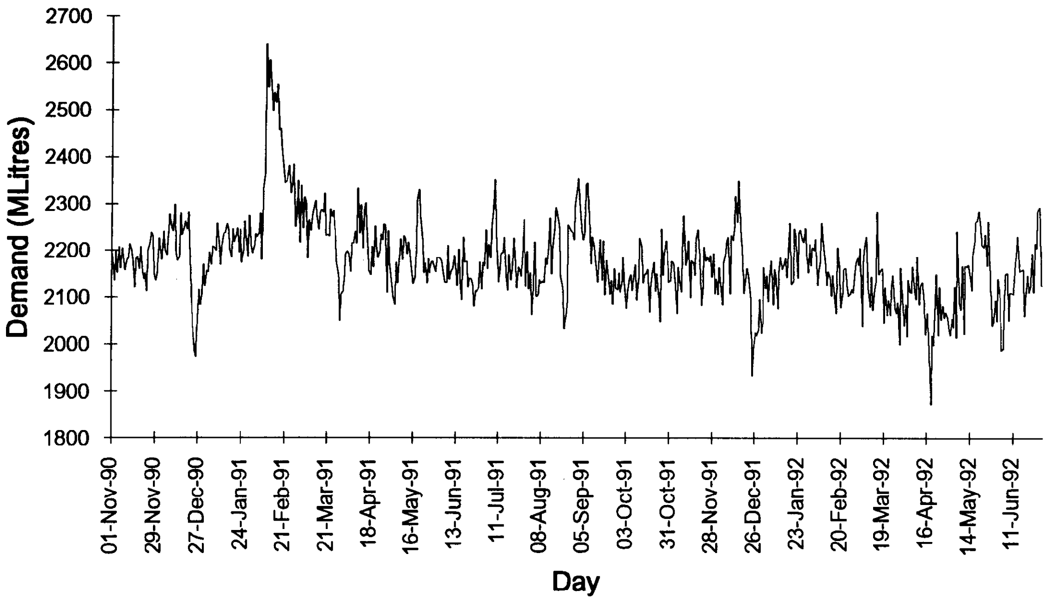

Figure 1 Daily Water Demand in London. Period: 1 November 1990–30 June 1992.

Transmission System. Daily sendout, D_t, is calculated using

$$D_t = \text{Intake from NTS } (I_t) - \text{Start Storage Next Day } (S_{t+1}) + \text{Start Storage Current Day } (S_t)$$

Sendout does not correspond to the amount of gas sold to end-users in the NW, due to such factors as stolen gas, leaks and calorific value assumptions. While sendout is not a completely accurate measure of the amount of gas taken by end-users, there are no identifiable consistent measurement errors. Daily sendout for winter 1990/91 is shown in Figure 2. The day of week and the occurrence of national holidays also affect the level of demand significantly. At weekends and holidays, sendout is less than it would be on weekdays as industrial gas demand decreases, with an analysis of the daily seasonality confirming the 'weekend effect'.

Weather is known to be the major factor affecting demand. Problems were identified with the temperature and wind speed data during the analysis

- In the North West, temperature and wind speed are not constant over the whole Region; consequently, these data

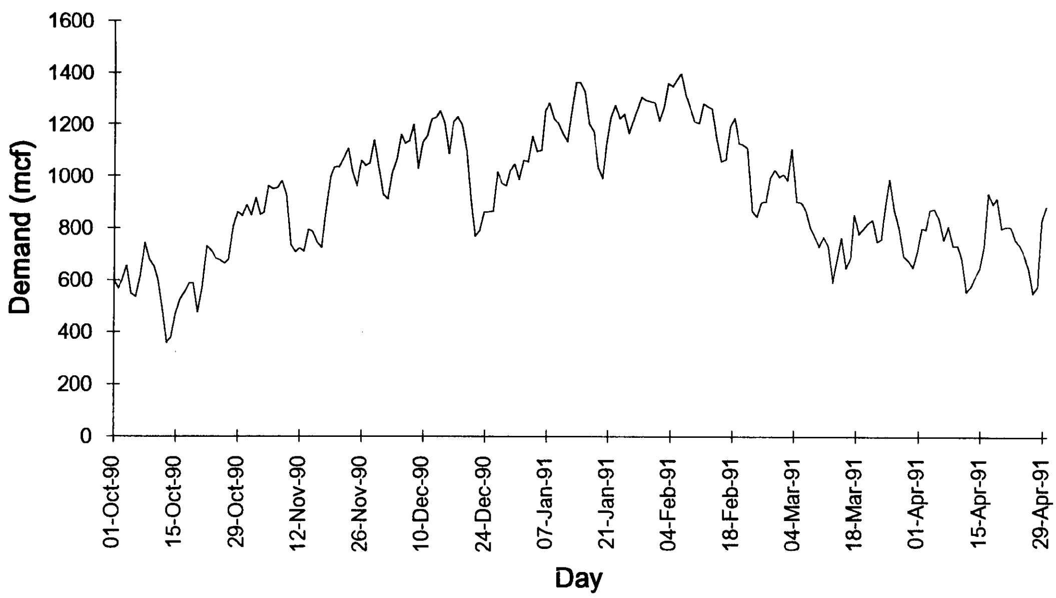

Figure 2 Daily Gas Sendout in the North West. Period: 1 October 1990–30 April 1991.

provide a less useful summary of whether effects throughout the Region when there is a large difference in temperature between different areas.

- There is significant bias in the forecasts of temperature and wind speed. Forecasts are likely to be more pessimistic (namely forecast temperatures colder and wind speeds higher) than actual.

At the time when the forecast is made, not only must the next day's climate be forecast, but only a part of the current day's demand is known. Predictions made at 1500 h for the current day are known as 'estimates', and predictions for the next day are known as 'forecasts'.

Modelling to improve forecast accuracy

Thames Water

The project brief given by Thames Water requested an exhaustive search of forecasting techniques concentrating on simpler approaches. A naive random walk model, tomorrow's demand forecast = today's demand, was defined as a basis for comparison.

The techniques of exponential smoothing, univariate Box-Jenkins modelling and dynamic regression were tried in order of increasing complexity. Transfer function modelling was not considered because Thames Water perceived the approach as too complex.

After evaluating the naive model, a range of exponential smoothing models including seasonality and trend were tried (as defined by Gardner[24]). The seasonality was identified as daily.

Performance of three models proved to be similar: (i) linear trend and multiplicative seasonality; (ii) no trend and multiplicative seasonality and (iii) no trend and additive seasonality. The first model including trend was unnecessarily complex. The differences between additive and multiplicative seasonality are typically slight when there is no trend in the data.

The choice between the remaining models was based on intuition about the nature of water demand. A change of demand due to a bank holiday, for instance, was considered to have an effect in addition to the regular day-of- the-week seasonality. The model based on no trend and additive seasonality was considered an adequate representation of the data.

The development of a univariate Box-Jenkins model requires the transformation of data if stationarity does not exist in both the mean and variance. No such transformation was required for the Thames Water data. The best model was found to have the form of an ARIMA(2, 0, 0)*(1, 0, 0)[7]: a model with two non-seasonal autoregressive parameters and one seasonal autoregressive parameter. (A constant term was also included in the model). Using this as a basis, a regression model was developed incorporating various weather variables and the lags derived from the ARIMA formulation.

Weather plays a significant role in determining water demand over the summer period. It affects the choice, for instance, of when people go on holiday. The affect of weather in winter is less significant. To identify this period, during the summer, a dummy variable was defined to represent the gardening season (G). The gardening season was defined to lie between 1 April and 14 September.

Weather data were available for minimum temperature (MnT), maximum temperature (MxT), wind (W), rainfall from 9 : 00 to 21 : 00 (R09), rainfall from 21 : 00 to 9 : 00 (R21), and sunshine (Sn). A dummy variable was used to identify the gardening season. A thaw variable was defined to represent the cold spell in February 1991. The variable represents the accumulated degrees of frost

$$\text{cumfz}_t = \min(\text{cumfz}_{t-1} + (\text{MnT}_t + \text{MxT}_t)/2, 0)$$

If the average temperature, $(\text{MnT}_t + \text{MxT}_t)2$, is below freezing cumfz_t decreases. If average temperature is above zero, but the sum of cumfz_{t-1} and average temperature is less than zero, then cumfz_t equals the sum of cumfz_{t-1} and average temperature. If the sum of the average temperature and cumfz_{t-1} is greater than zero than cumfz_t equals zero.

The bank holidays were split up into different variables: Christmas (Xmas); New Year (NY); Easter (E); Summer bank holidays (BH).

After careful analysis of seasonality and autocorrelation, the final regression model was determined as shown in Table 1. The correction for first order autocorrelation was estimated by the Cochran-Orcutt method (using Forecast Master Plus[25]). Autocorrelation still existed for lag 14.

The inclusion of lagged demand implies that the other variables explain changes in the level of demand from the previous periods. The larger coefficient for the change in sunshine hours out of gardening season, rather than in the season, shows that the appearance of the sun has more effect outside the gardening season. Those working within the industry suggested a possible explanation is that the appearance of the sun in Summer is less of a surprise and therefore leads to a more limited change of behaviour. The 'accumulated degrees of frost' variable is such that the higher the value, the lower the demand, therefore a negative coefficient implies a higher demand. (The sum of the various lagged coefficients are negative with the alternating sign explained by the correlation inbuilt into the variable's definition.)

Various diagnostic tests of the model were carried out for evidence of mis- specification and failure of the regression assumptions.

To check how stable the coefficients were, the simple Box-Jenkins demand model was evaluated using the Chow Test (also known as the analysis of covariance test)[26]. The test works by measuring the equality of regression coeffi-

Table 1 The explanatory variables used to predict water demand (D_t)

Variable	*Description*	*Coefficient value*	*T-Ratio*
D_{t-1}	Demand (the day before)	0.644	6.58
D_{t-2}	Demand (two days before)	0.154	1.73
D_{t-7}	Demand (seven days before)	0.114	2.29
D_{t-8}	Demand (eight days before)	−0.092	−1.92
M_t	Monday	17.0	2.13
F_t	Friday	−19.7	−2.66
Su_t	Sunday	27.9	3.71
CumFz_{t-1}	Accumulated degrees of frost (the day before)	−25.2	−4.25
CumFz_{t-2}	Accumulated degrees of frost (two days before)	37.1	2.80
CumFz_{t-3}	Accumulated degrees of frost (three days before)	−50.8	−3.29
CumFz_{t-4}	Accumulated degrees of frost (four days before)	55.5	4.21
CumFz_{t-5}	Accumulated degrees of frost (five days before)	−26.0	−4.24
Xmas	Christmas	−95.6	−2.20
Xmas(−1)	Christmas Eve	−99.0	−2.14
Xmas(−2)	23 December	−167	−3.93
BH(+1)	The day after a bank holiday	57.3	2.28
BH(−1)	The day before a bank holiday	−63.2	−2.55
$(\mathrm{G}_t^* \mathrm{Sn}_t')$	Change in the number of sunshine hours in the gardening season	2.49	2.74
$(\mathrm{G}_t^* \mathrm{MxT}_t)$	Change in maximum temperature in the gardening season	4.32	2.84
$(\mathrm{G}_t^* \mathrm{R21}_t)$	Last night's rainfall in the gardening season	−3.18	−2.41
$(\mathrm{G}_t^{-1*} \mathrm{Sn}_t')$	Change in the number of sunshine hours outside gardening season	3.30	3.18
$(\mathrm{G}_t^{-1*} \mathrm{MnT}_t')$	Change in minimum temperature outside the gardening season	3.34	2.13
u_{t-1}	Autocorrelation term	−0.248	−2.25
c	Constant	393	5.40

Estimation Period: 1/11/90–31/9/91. $n = 325$ effective observations.
Average demand = 2210; standard error = 41.6, $R^2 = 79.3$.
N.B. X_t' represents the first difference, $X_t - X_{t-1}$.

cients over two sample periods conditional on the equality of the error variances in the two samples. The test result was significant, indicating instability in the coefficients.

To check whether variable parameters were needed, the coefficients were examined using rolling regression. (A rolling regression is where the parameters are estimated on a moving data window rather than on the full data set). The rolling regression (using *Microfit*[26]) was carried out over the gardening season, 1 April 1991 to 14 September 1991. The size of the window for the regression was 100 data points. The size was chosen for three reasons

- To allow holiday variables to be included in the rolling regression.
- To ensure the coefficients were not affected significantly by individual data points.
- To maximize the importance of the latest data.

An example of how the coefficients vary is shown in Figure 3. The overnight rainfall coefficient, shown in Figure 3, increases towards zero as the end of the gardening season approaches. It would be reasonable to sumise that the coefficient at the beginning of the gardening season started near zero, approached a peak value of − 7 and then went back to zero at the end of the season. The importance of the overnight rain relates to the use of hosepipes.

The coefficients of Monday, Friday and Sunday were also found to vary. This was thought to represent the changes in leisure time behaviour as the season progresses.

To allow the technique to predict water demand for a whole year, the demand data were first adjusted to remove the effects of the thaw, Christmas and bank holidays. The window size was increased to 210 data points so that all the variables were defined during the rolling regression. The results (shown in Table 2) indicate an improvement on earlier models, but at the cost of increased complexity. Note that the models used depend on the actual values of the explanatory variables, the history of local weather forecasts being unavailable except at prohibitive cost.

The performance of the models considered so far is shown in Table 2. Exponential smoothing offers a slightly more appropriate characterization of the data than the univariate Box-Jenkins model identified. The regression models were found to be the best over the gardening season. Outside the gardening season, exponential smooth-

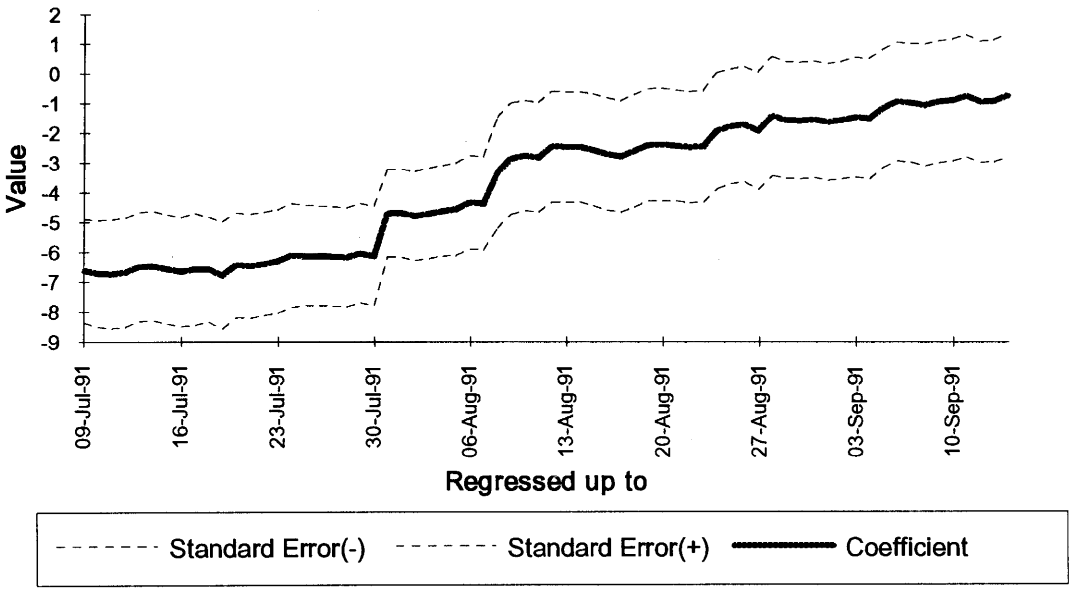

Figure 3 The changing effect of overnight rainfall on Water Demand(Rain21) in a Rolling Regression. Period: 1 April 1991–14 September 1991; Window size = 100.

Table 2 Comparison of the water demand models: one day ahead forecast error summary statistics

Model	*Out-of-Sample Period*			
	1/10/91–5/4/92: n = 187		*6/4/92–30/6/92: n = 85*	
	MAD	*RMSE*	*MAD*	*RMSE*
Random walk	51	64	56	74
Exponential smoothing	41	53	50	65
Univariate Box-Jenkins	42	56	53	65
Regression	41	51	46	60
Rolling regression	39	51	44	57

ing performed almost as well. These results were also confirmed by further analysis of the out-of-sample error distributions; with distributions which are skewed (as these) examination of a limited range of summary statistics can lead to mis- understanding the results[27]. For example, comparing the performance of Box-Jenkins with Regression for Winter 199192, the difference in MAD (MAPE and RMSE) is solely due to a few observations. (In fact, the Median Absolute Deviation is lower for Box-Jenkins than Regression.) Overall, these results conform to expectations derived from the earlier research cited that weather does have a causal effect on water demand and its inclusion leads to improved forecasting performance.

British Gas NW project

The project brief was

(i) to improve the forecasting methodology then employed by evaluating various possible changes to the existing explanatory model,

(ii) to examine the use of Box-Jenkins methodology,

in the light of the error statistics described in the earlier section.

Prior to the project, British Gas NW had developed a forecasting approach of using seasonal models:

- A winter model operational between the months of October and April inclusive.
- A summer model between May and September inclusive.

British Gas NW regarded it as vital that the winter model was as accurate as possible, since it is on cold days that demand is highest and accurate forecasting is most needed. This project therefore concentrated on developing the winter model.

Little evidence was available from earlier analyses that any out-of-sample comparisons with alternatives had been carried out when the model in operation in British Gas NW was adopted. Because of the emphasis on winter demand, data from the winters 1985–1990 were used to establish the parameter estimates, and the data from winter 199091 were used to test the forecasting performance of the potential models.

The current forecasting method

The starting point for the project was the multiplicative model (*Mult Dual*) developed by British Gas NW to forecast next day demand (sendout) in the winter of 1990/91.

$$\text{Fcst. Demand}_{t+1} = (\text{Est. Demand}_t - \text{Adj}_t) \times (1 + \text{TmpCh}_{t+1}{}^{*}\beta_{\text{Temp}} + \text{WndCh}_{t+1}{}^{*}\beta_{\text{Wind}}) + \text{Adj}_{t+1}$$

where the variables are defined as follows

- Fcst. Demand_{t+1} is the forecast of demand for day $t+1$,
- Est. Demand_t is the estimated demand for day t,
- TempCh_{t+1} which is the change in Effective Temperature (Forecast of EffTemp_{t+1} − Estimate of EffTemp_t),
- WindCh_{t+1} which is the change in average wind speed (Forecast of Average Wind speed for day $t+1$ − Estimate of Average Wind speed for day t), and the parameters are
- β_{Temp} is the temperature sensitivity,
- β_{Wind} is the wind speed sensitivity,
- Adj_t is the adjustment factor for the day of the week, t. This takes into account industry shutdown on Friday, Saturday and Sunday.

Effective temperature is measured as an exponentially weighted average of past temperatures: an attempt to measure the phased adjustment of householders to changes in daily temperature.

$$\text{EffTemp}_t = \alpha\text{Temp}_t + \alpha(1-\alpha)\text{Temp}_{t-1} + \alpha(1-\alpha)^2\text{Temp}_{t-2} + \ldots$$

where Temp_t = Average temperature for day t. Experimentation identified the best value for α as 0.5.

Another feature of the model was dual temperature sensitivity, where temperature sensitivity of end-users depends on the forecast average temperature

$$\beta_{\text{Temp}} = \beta_{\text{Temp1}} \text{ if } \text{Temp}_t >= \text{Split Temperature}$$
$$= \beta_{\text{Temp2}} \text{ if } \text{Temp}_t <= \text{Split Temperature}$$

The split was investigated as the sensitivity of end-users was believed to vary with temperature.

A simplified version of this multiplicative model (*Mult*) was tried without the dual temperature sensitivity. An additive form of the regression model (*Add*), without dual temperature sensitivity, was considered next. The model looked at causes of change in demand:

$$\text{Change in Demand}_{t+1} = \text{TmpCh}_{t+1}{}^{*}\beta_{\text{Temp}} + \text{WndCh}_{t+1}{}^{*}\beta_{\text{Wind}} + \text{Adj}_{t+1} - \text{Adj}_t$$

The model (*Add Dual*) was repeated including the dual temperature sensitivity. Dual temperature sensitivity made little improvement to the performance of the additive model, and many of the error measures deteriorated.

To counteract the autocorrelation noticeable in the residuals of both the additive and multiplicative models a first order autocorrelation term was added. The two models (*Mult AC*, *Add AC*) were repeated with autocorrelation but without the dual temperature sensitivity. The introduction of autocorrelation improved most of the error measures for the multiplicative model. For the additive model, the use of autocorrelation produced the lowest mean absolute total error of all the models considered. For both models, the introduction of autocorrelation slightly improved the mean absolute % total error, but the number of 5% and 10% total errors (based on the forecast weather) increased. In general however the changes were slight.

Alternative models

Base line comparisons were made with exponential smoothing with multiplicative seasonality and no trend. Also ARIMA style models were built with deterministic (dummy) seasonals and two autoregressive parameters. Attention was given to the need to use estimated sendout for the previous period as well as the unusual observations around Christmas. The transfer function class of models (*Transfer*) was next considered; this links the explanatory variables, temperature and wind speed, and their lagged values to the dependent variable (and its lags). The transfer model selected was similar to those derived by Piggott and Borgard *et al*[10,11] but using wind speed instead of solar radiation as an explanatory variable.

The additional complexity proved to be of no benefit, since the conventional explanatory models outperformed the transfer function methodology. Further, it remained an important issue that the engineers who made the daily forecast understood how the model worked[9]. The additive autocorrelation was understood at British Gas NW, but this comprehensibility would have been lost with the implementation of the transfer model.

The performance of the models is shown in Table 3. The multiplicative models produced results slightly inferior to the corresponding additive models given above. It was decided that the additive autocorrelation model (*Add AC*) was to be recommended for winter 1991/92. The results of the model show a large mean total error and a small mean unexplained error. This is explained by bias in the temperature and wind speed forecasts of winter 1990/91. Bias adjustments were therefore considered for the additive autocorrelation model.

Two methods of calculating the adjustment factors were tried. The first calculated the recent bias in forecasting the explanatory variables but led to an over-adjustment. The second method involved identifying the adjustments by a

Table 3 Comparison of gas demand models: one day ahead forecast error summary statistics (effective sample size = 196)

	Out-of-Sample period winter October 1990–April 1991							
	Total error (%): Forecast weather variables				*Unexplained error (%) Actual weather variables*			
Model	*Mean error*	*MAP-E*	*APE > 5%*	*APE-> 10%*	*Mean error*	*MAP-E*	*APE > 5%*	*APE-> 10%*
Random walk	−0.470	7.953	54.1%	30.6%	—	—	—	—
Exponential smoothing	−0.427	7.048	52.0%	25.5%	—	—	—	—
ARIMA	0.373	6.652	59.7%	18.9%	—	—	—	—
Additive	−1.932	5.393	41.0%	12.2%	−0.104	3.045	19.4%	2.6%
Add Dual	−1.915	5.428	42.3%	12.2%	−0.100	3.002	17.3%	2.6%
Add AC	−2.406	5.358	42.3%	13.3%	−0.096	2.906	16.3%	2.6%
Transfer[a]	−0.728	5.518	42.3%	13.6%	0.132	3.263	19.6%	3.5%
Bias Adj	−0.245	5.013	35.7%	11.7%	−0.096	2.906	16.3%	2.6%

Table 4 Forecasting gas demand: value of parameters in the additive model with autocorrelation (Add AC)

Variable	*Coefficient value*	*t-ratio*	*Summary statistics*
Temperature	−75.6	−53.1	
Wind	8.83	21.2	Average = 870
Friday (Adj)	−44.3	−11.7	$R^2 = 83\%$
Saturday (Adj)	−158	−39.0	Std. Error = 34.5
Sunday (Adj)	−137	−36.3	$n = 968$
Autocorrelation Term	−0.214	na	

Estimation Period: October 1986–April 1991: Bank holidays omitted.

process of trial and error, using winter 1989/90 as the test period. The best adjustment factors were then applied to the data of winter 1990/91. The results of the additive autocorrelation model with adjustments (*Bias Adf*) are shown in Table 4 and this was the model recommended for adoption. The parameters for the proposed model using the data of winters 1986/87 to 1990/91, scaled to disguise their magnitude, are shown in Table 4.

Conclusions

In the forecasting literature there has been some controversy about whether the inclusion of explanatory variables typically improves forecasting accuracy. The British Gas North Western study adds further support to the view that where forecasts of the explanatory variables are reasonably accurate, gains can be made. Short-term weather forecasts fulfil that requirement. The Thames study is more tentative in that forecasts of the weather variables were not available, therefore the analysis only pointed towards the same conclusion. Engle *et al*[15] study of electricity adds further evidence. We therefore conclude that those utilities still relying on extrapolative modelling are likely to be incurring unnecessarily high levels of forecast inaccuracy and correspondingly higher costs.

Whilst the water study again supports the conclusions of Makridakis *et al*.[28] that exponential smoothing typically outperforms the Box-Jenkins methodology, for gas the relationship is in the opposite direction. Arguments also persist concerning the relative virtues of the transfer function approach compared to dynamic regression. The former is now recognized to be a special case of the latter (see for example, Fildes[19]) with particular problems concerning stationarity of variables included in the model. However, this does not necessarily imply poorer performance in practice—the parameter constraint implicit in the transfer function methodology may prove useful. Here the regression approach was better.

The question as to whether variable parameter modelling leads to improved performance remains moot with only limited evidence available[29]. In the Thames study small gains were achieved using the simplest form of this class of model. Both of the studies reported on here split the available data base, for example using the winter period alone. Parameter fluctuations were established throughout the year (and perhaps from year to year). Although the use of time varying parameter models gained only limited support, using different models for distinct sub-sets of the data (which recognizes the shifts in parameters over the seasons) should lead to improved accuracy, for example summer and winter models.

In summary, when selecting a forecasting method for an important application, the methodological issues that need careful attention are: (1) the choice and definition of the explanatory variables and the accuracy with which they are forecast, (2) data correction, to deal with such features as missing observations and holidays, (3) comparative testing of the various alternative forecasting methods using out-of-sample data, (4) appropriate choice of forecast error measures and (5) diagnostic checks, in particular of a model's stability across the range of its potential applications.

With OR projects, implementation of an apparently improved procedure can never be taken for granted. For

British Gas North Western the improvements deriving from the 'best' model described above were sufficient to lead the user engineers to discard their previous approach in favour of the additive autocorrelation model with adjusted weather forecasts. The search for further improvements continues with current interest focusing around the use of neural networks. For Thames Water, in what was a more exploratory study, the accuracy of the forecasts derivable from the simpler models were not thought to offer sufficient improvement in accuracy to merit replacing the judgmental forecasts made by the controllers. However further research has continued into the determinants of water demand, an increasingly important topic with the adoption of water metering and the effects of the drought in Summer 1995.

In conclusion, although there are good reasons to believe the best forecasting model will be a complex, non-linear or time-varying one, it appears that careful treatment of weather, weekends, holidays and other factors are the key to successful forecasting. If simplicity is all important, the users should stick to exponential smoothing or rely on their experience; the loss in accuracy is less than 10% for water and about 25% for gas. If such gains are worth having, the key to achieving the improvements is careful model building and out-of-sample model comparisons.

Acknowledgements—This work was carried out with the help of staff in British Gas North Western and Thames Water. However the opinions and the results contained in this paper are the sole responsibility of the authors.

References

1 Fildes R and Beard C (1992). Forecasting systems for production and inventory control, *J Oper and Prod Mgmt* **12** (5): 4–26.
2 Bunn DW and Farmer ED (1985). *Comparative Models for Electricity Load Forecasting*. Wiley: Chichester.
3 Stubbs PA (1991). Short term demand forecasting at British Gas North Western. MSc Project, Lancaster University.
4 Randall AE (1992). Forecasting water demand in London for Thames Water. MSc Project, Lancaster University.
5 Bunn DW and Larsen ER (1992). Sensitivity of reserve margins to factors influencing behaviour in the electricity market of England and Wales, *Energy Policy* **20**: 420–429.
6 Chatfield C (1988). What is the 'best' method of forecasting, *J App Sta* **15**: 19–38.
7 Berrisford HG (1965). The relation between gas demand and temperature: a study in statistical demand forecasting, *Opl Res Q* **16**: 229–246.
8 Lyness FK (1981). Consistent forecasting of severe winter gas demand, *J Opl Res Soc* **32**: 347–359.
9 Taylor PF and Thomas ME (1982). Short term forecasting: horses for courses, *J Opl Res Soc* **33**: 685–694.
10 Borgard C, George G, Jenkins GM and McLeod C (1982). Analysing a large number of energy time series for a utility company. In *Case Studies in Time Series Analysis Vol 1*. (Jenkins GM and McLeod G, Eds.), GJP Publications: Lancaster, pp. 113–149.
11 Piggott JL (1985). Short-term forecasting at British Gas. In *Comparative Models for Electricity Load Forecasting*. (Bunn DW and Farmer ED, Eds.), Wiley: Chichester, pp. 173–211.
12 Ashouri F (1993). An expert system for predicting gas demand: a case study, *Omega* **21**: 307–317.
13 Sterling MJH and Bargiela A (1985). Adaptive forecasting of daily water demand. In *Comparative Models for Electrical Load Forecasting*. (Bunn DW and Farmer ED, Eds.), Wiley: Chichester, pp. 213–225.
14 Bunn DW and Farmer ED (1985). Review of short-term forecasting models in the electrical power industry. In *Comparative Models for Electrical Load Forecasting*. (Bunn DW and Farmer ED, Eds.), Wiley: Chichester, pp. 13–30.
15 Engle RF, Brown SJ and Stern G (1988). A comparison of adaptive structural forecasting methods for electricity sales, *J Forecasting* **7**: 149–172.
16 Engle RF, Granger CWJ, Ramanathan R, Vahid-Araghi F, and Brace C (1997). Short-run forecasts of electricity loads and peaks. *Int J Forecasting*, forthcoming.
17 Fildes R (1988). Recent developments in time series forecasting, *OR Spektrum* **10**: 195–212.
18 Fildes R and Howell S (1979). On selecting a forecasting model. In *Forecasting*, *TIMS Studies in the Management Sciences*. (Makridakis S and Wheelwright SC, Eds.), North-Holland: Amsterdam, pp. 297–312.
19 Fildes R (1985). Quantitative forecasting—the state of the art: econometric models, *J Opl Res Soc* **36**: 549–580.
20 Ashley R (1988). On the relative worth of recent macroeconomic forecasts, *Int J Forecasting* **4**: 363–376.
21 Armstrong JS and Collopy F (1992). Error measures for generalizing about forecasting methods: empirical comparisons, with discussion, *Int J Forecasting* **8**: 69–80, 99–111.
22 Yokum JT and Armstrong JS (1995). Beyond accuracy: Comparison of criteria used to select forecasting methods, *Int J Forecasting* **11**: 591–597.
23 Chatfield C (1992). A commentary on error measures, *Int J Forecasting* **8**: 100–102.
24 Gardner ES Jr (1985). Exponential smoothing: The state of the art, with discussion, *J Forecasting* **4**: 1–28.
25 Forecast Master Plus (1989). Business Forecast Systems, Belmont, Ma. 02178.
26 Pesaran MH and Pesaran B (1991). *Microfit 3.0 User Manual*, Oxford University Press: Oxford.
27 Fildes R (1992). The evaluation of univariate extrapolative forecasting models, *Int J Forecasting* **8**: 81–98.
28 Makridakis S *et al* (1993). The M-2 competition: A real-time judgementally based forecasting study, with discussion, *Int J Forecasting* **9**: 5–29.
29 Riddington G (1993). Time varying coefficient models and their forecasting performance, *Omega* **21**: 573–583.

Received September 1994
accepted May 1996 after two revisions.