TECHNICAL MEMORANDUM

312

Using weather ensemble predictions in electricity demand forecasting

J.W. Taylor and R. Buizza

Research Department

September 2000

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

ABSTRACT

Weather forecasts are an important input to many electricity demand forecasting models. This study investigates the use of weather ensemble predictions in electricity demand forecasting for lead times from 1 to 10 days-ahead. A weather ensemble prediction consists of 51 scenarios for a weather variable. We use these scenarios to produce 51 scenarios for the weather-related component of electricity demand. The results show that the average of the demand scenarios is a more accurate demand forecast than that produced using traditional weather forecasts. We use the distribution of the demand scenarios to estimate the demand forecast uncertainty. This compares favourably with estimates produced using univariate volatility forecasting methods.

1. INTRODUCTION

Weather variables are used to model electricity demand. Demand forecasts are produced by substituting a forecast for each weather variable in the model. Traditionally, single point weather forecasts have been used. In this paper, we consider a new type of forecast, called weather ensemble predictions. An ensemble prediction consists of 51 different members. Each member is a different scenario for the future value of the weather variable. The ensemble, therefore, conveys the degree of uncertainty in the weather variable.

We use the 51 weather ensemble members to produce 51 scenarios for electricity demand at lead times from 1 to 10 days-ahead. Meteorologists sometimes find that the mean of the 51 ensemble members for a weather variable is a more accurate forecast of the variable than a traditional single point forecast (*Leith*, 1974; *Molteni et al.*, 1996). In view of this, we consider the use of the average of the 51 demand scenarios as a point forecast of demand. We use the distribution of the electricity demand scenarios as an input to estimating the uncertainty in demand forecasts. It is important to assess the uncertainty in order to manage the system load efficiently. A measure of risk is also beneficial for those trading electricity.

In this paper, we use the electricity demand forecasting methodology of the National Grid Company (NGC) as a basis for our analysis. NGC is responsible for the transmission of electricity in England and Wales. The company's demand forecasts have always been a crucial input to operational planning where the generation output is scheduled to meet customer demand. However, since the re-structuring of the industry in 1990, and the introduction of the daily electricity power pool, the NGC demand forecasts have also been used to set the price of electricity in the pool. With the anticipated new structure of the industry, accurate demand forecasting will also be required by utilities who will need to predict their customers' demand, and by those wishing to trade electricity on financial markets.

Weather ensemble predictions are described in Section 2. Section 3 briefly reviews electricity demand forecasting before presenting the method and variables currently used by NGC. Section 4 considers how weather ensemble predictions can be used to improve the accuracy of demand forecasts. Sections 5 and 6 investigate the potential for using weather ensemble predictions to assess the uncertainty in demand forecasts. The estimation of demand forecast error standard deviation is considered in Section 5, and demand prediction intervals are the focus of Section 6. The final section provides a summary and conclusion.

2. ENSEMBLE WEATHER PREDICTIONS

The weather is a chaotic system. Small errors in the initial conditions of a forecast grow rapidly, and affect predictability. Furthermore, predictability is limited by model errors due to the approximate simulation of atmospheric processes in a state-of-the-art numerical model. These two sources of uncertainty limit the accuracy of single point forecasts, which are generated by running the model at high resolution with best estimates for the initial conditions (see Figure 1).

Generally speaking, a complete description of the weather prediction problem can be stated in terms of the time evolution of an appropriate probability density function (pdf) in the atmosphere's phase space. An estimate of the pdf provides forecasters with an objective way to understand the uncertainty in single point predictions. Ensemble prediction aims to derive a more sophisticated estimate of the pdf than that provided by a univariate extrapolation of the empirical distribution of historic errors. Ensemble prediction systems generate multiple realisations of numerical predictions by using a range of different initial conditions in the numerical model of the atmosphere, which is run at a slightly lower resolution than for the single point forecast. The frequency distribution of the different realisations, which are known as ensemble members, provides an estimate of the pdf.

Since December 1992, both the US National Center for Environmental Predictions (NCEP, previously NMC) and the European Centre for Medium-range Weather Forecasts (ECMWF) have integrated their deterministic high-resolution prediction with medium-range ensemble prediction (*Toth and Kalnay*, 1993; *Tracton and Kalnay*, 1993; *Palmer et al.*, 1993). These developments followed the theoretical and experimental work of, among others, *Epstein* (1969), *Gleeson* (1970), *Fleming* (1971a, 1971b) and *Leith* (1974).

Routine real-time execution of the ECMWF ensemble prediction system started in December 1992 with a 31-member configuration (*Palmer et al.*, 1993, *Molteni et al.*, 1996). The number of ensemble members is constrained by the running time of the atmospheric model. A major upgrade to a 51-member system took place in December 1996 (*Buizza et al.*, 1998; *Buizza*, 1998). A further upgrade followed in October 1998, with the introduction of stochastic physics into the system (*Buizza et al.*, 1999). This aims to simulate model uncertainties due to random model error in the parameterised physical processes. Ensemble forecasts are produced routinely every day for lead times from 12 hours-ahead to 10 days-ahead. Ensemble forecasts of many different weather variables are archived every 12 hours, and are thus available for midday and midnight. The ECMWF disseminates ensemble forecasts to the National Meteorological Centers of its European member states, as part of an operational suite of weather products.

In this study, we have used ensemble predictions generated by the ECMWF from 1 November 1998 till 30 April 2000. We limited our study to this period because the introduction of stochastic physics in October 1998 substantially improved the characteristics of the ensemble predictions of surface variables.

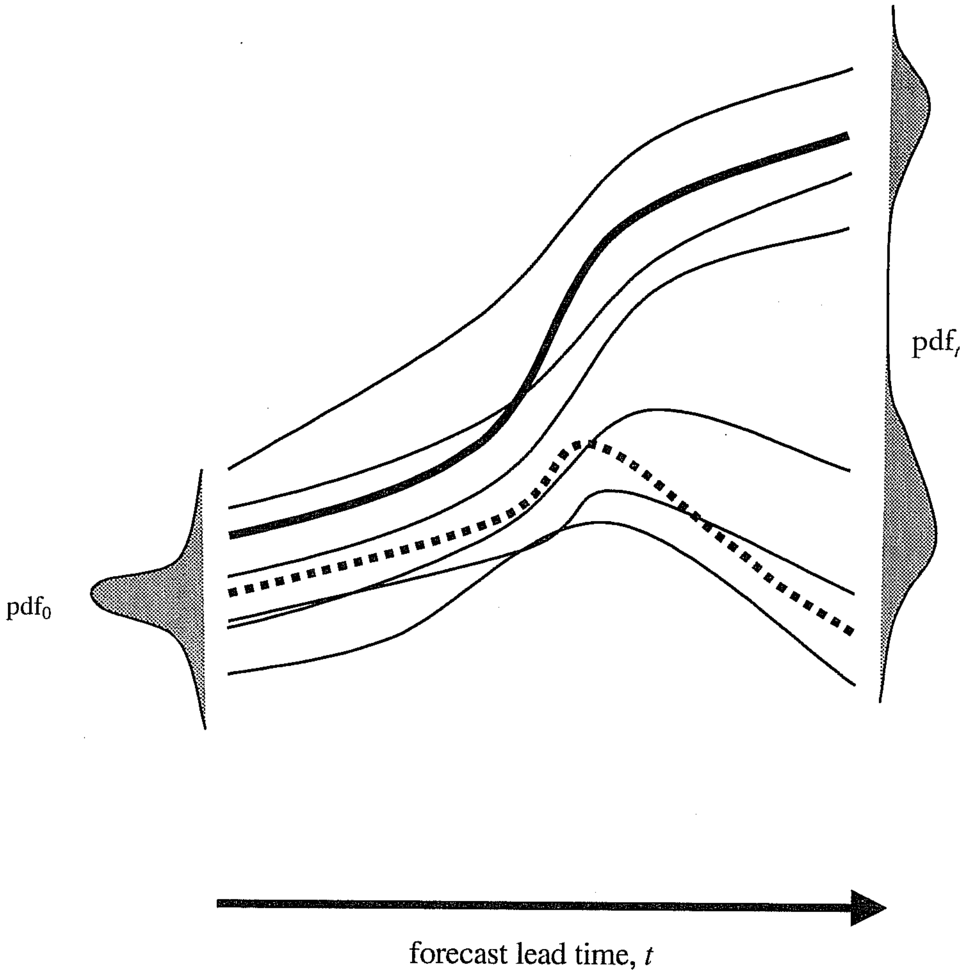

Figure 1: Schematic of ensemble prediction. The initial probability density function, pdf_0, represents the initial uncertainties. From the best estimate of the initial state, a single point forecast (bold solid curve) is produced. This point forecast fails to predict correctly the future state (dash curve). An ensemble of perturbed forecasts (thin solid curves) starting from perturbed initial conditions, designed to sample the initial uncertainties, can be used to estimate the probability of future states. In this example, the estimated probability density function, pdf_t is bimodal. The figure shows that two of the perturbed forecasts almost correctly predicted the future state. Therefore, at time 0, the ensemble system would have given a non-zero probability of the future state.

3. ELECTRICITY DEMAND FORECASTING

In this section, we describe the forecasting process at NGC. We present the modelling approach and the weather variables in some detail, as they form the basis of our analysis in the remainder of the paper.

3.1 Modelling electricity demand in England and Wales

There is no consensus as to the best approach to electricity demand forecasting. The Puget Power Company in Seattle organised a 1 day-ahead forecasting competition involving a range of different approaches including: time-varying splines (*Harvey and Koopman*, 1993), artificial neural networks (*Connor*, 1996), multiple regression models (*Ramanathan et al.*, 1997), judgemental forecasts produced by Puget Power's own personnel, and Box-Jenkins transfer function intervention-noise models. The approach taken by NGC is first to forecast the demand at the 10 or 11 daily turning points and at several strategically positioned fixed points, such as midday and midnight. These turning points and fixed points are collectively known as cardinal points. Forecasts for periods between cardinal points are then obtained by a procedure known as profiling which involves fitting a curve to the cardinal points (see *Taylor and Majithia*, 2000). The cardinal point forecasts are produced by separate regression models which are functions of seasonal and weather variables (*Baker*, 1985). This method has similarities with the method of the overall winners of the Seattle competition, *Ramanathan et al.* (1997), who produced hourly forecasts by using separate regression models for each hour of the day.

3.2 Modelling midday electricity demand

In this paper, we focus on predicting demand at midday. This is convenient because ensemble predictions are currently available for midday, although in the future they certainly could be produced for any required period of the day. Midday is always chosen as a fixed cardinal point by NGC, and so there is no need to perform the NGC profiling heuristic. Midday is a particularly important period in many summer months because it is often when peak demand occurs. We follow the procedure of NGC and *Ramanathan et al.* (1997) and produce a model for midday based on demand for previous middays and weather variables.

Figure 2 shows a plot of electricity demand in England and Wales at midday for each day in 1999. One clear feature of demand is the strong seasonality throughout the year which results in a difference of about 500 MW between typical winter and typical summer demand. Another noticeable seasonal feature occurs within each week where there is a consistent difference of about 600 MW between weekday and weekend demand. There is unusual demand on a number of 'special days', including public holidays, such as 1 January. In practice, NGC forecasts demand on these days using judgemental methods. As in many other studies of electricity demand, we elected to smooth out these special days, as their inclusion is likely to be unhelpful in our analysis of the relationship between demand and weather.

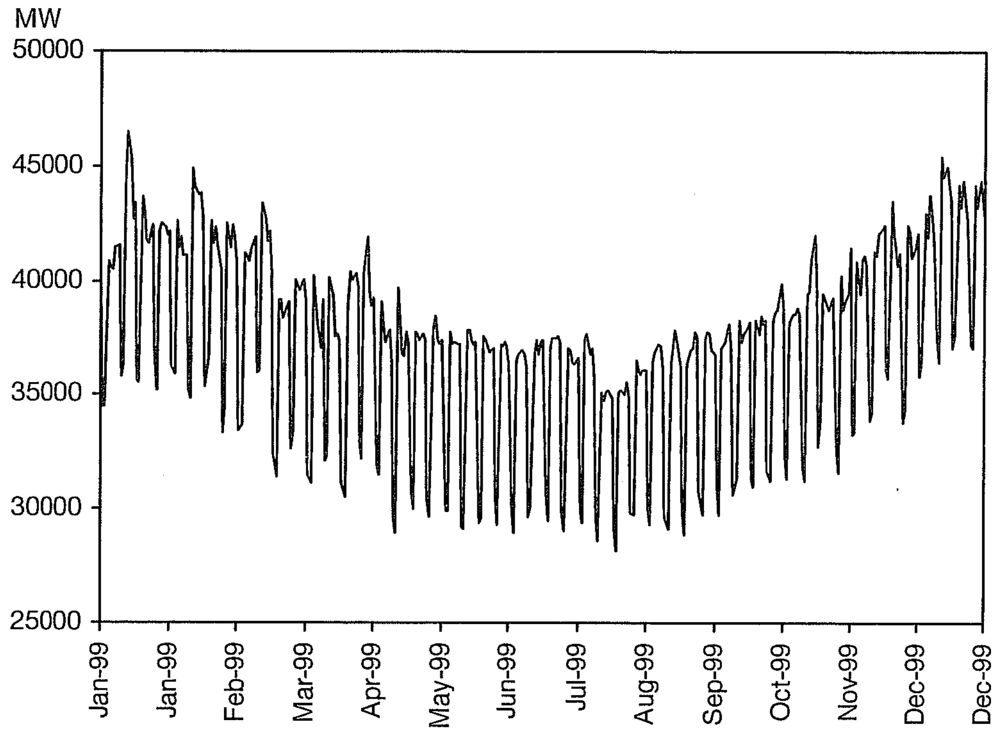

Figure 2: Demand for Electricity at Midday in England and Wales in 1999.

Short to medium-term forecasting models must accommodate the variation in demand due to the seasonal patterns shown in Figure 2 and due to weather. At NGC, demand is modelled using three weather variables: effective temperature, cooling power of the wind and effective illumination. These variables are constructed by transforming standard weather variables in such a way as to enable efficient modelling of weather induced demand variation (*Baker*, 1985). *Effective temperature* (TE_t) is an exponentially smoothed form of TO_t, which is the mean of the spot temperature recorded for each of the four previous hours.

$$TE_t = \tfrac{1}{2}TO_t + \tfrac{1}{2}TE_{t-1}$$

The influence of lagged temperature aims to reflect the delay in response of heating appliances within buildings to changes in external temperature. *Cooling power of the wind* (CP_t) is a nonlinear function of wind speed, W_t, and average temperature, TO_t. It aims to describe the draught-induced load variation.

$$CP_t = \begin{cases} W_t^{\frac{1}{2}}(18.3 - TO_t) & \text{if} \quad TO_t < 18.3 \\ 0 & \text{if} \quad TO_t \geq 18.3 \end{cases} \tag{1}$$

Effective illumination is a complex function of visibility, number and type of cloud and amount and type of precipitation. Since NGC needs to model the demand for the whole of England and Wales, weighted averages are used of weather readings at Birmingham, Bristol, Leeds, Manchester and London. The weighted averages aim to reflect population concentrations in a simple way by using the same weighting for all the locations except London which is given a double weighting.

Since the aim of this paper is to investigate the potential for the use of ensemble predictions in electricity demand forecasting, we opted to use only weather variables for which ensemble predictions were available. Ensemble predictions are available for temperature, wind speed and cloud cover (CC_t) at midday and midnight. In view of this, we replaced effective illumination by cloud cover, and we used spot temperature, instead of average temperature, TO_t, to construct effective temperature and cooling power of the wind. We do not feel that this implies that our modelling was weak because many studies in the literature use only spot temperature to model electricity demand.

A common approach to electricity demand forecasting is to predict separately the weather related demand and the non-weather related demand, the 'base load'. For simplicity, in this paper, we follow the two-stage approach of NGC. The first stage aims to identify the weather related component by estimating a regression model similar to the following:

$$demand_t = a_0 + a_1\, TE_t + a_2\, TE_t^2 + a_3\, CP_t + a_4\, CC_t + a_5\, t + a_6\, t^2 + a_7\, t^3 + a_8\, t^4 + a_9\, FRI_t + a_{10}\, SAT_t + a_{11}\, SUN_t + a_{12}\, W1_t + a_{13}\, W2_t + a_{14}\, W3_t + \varepsilon_t \qquad (2)$$

where FRI_t, SAT_t and SUN_t are 0/1 dummy variables for Fridays, Saturdays and Sundays; $W1_t$, $W2_t$ and $W3_t$ are 0/1 dummy variables representing the three summer weeks when a large amount of industry closes; ε_t is an error term; and the a_i are constant parameters. The time polynomial is used to model in a deterministic way the yearly seasonal effect that was evident in Figure 2. We followed NGC in using data from the previous two years to estimate the model, and so a quartic time polynomial was appropriate.

The second stage of the approach involves summing forecasts of the weather related demand and the base load. A forecast for the weather related demand is produced by substituting traditional weather point forecasts in the following expression taken from the estimated regression model in (2):

$$weather\_related\_demand = \hat{a}_1\, TE_t + \hat{a}_2\, TE_t^2 + \hat{a}_3\, CP_t + \hat{a}_4\, CC_t \qquad (3)$$

Forecasts for the base load are produced judgementally. In this study, we predicted base load using a well specified ARMA-regression model of the following form:

$$base\_demand_t = b_0 + b_1\, FRI_t + b_2\, SAT_t + b_3\, SUN_t + b_4 W2_t + b_5 W3_t + \varepsilon_t$$
$$\varepsilon_t = \phi_1\, \varepsilon_{t-1} + \phi_2\, \varepsilon_{t-2} + \theta_1\, u_{t-1} + u_t$$

where u_t is a white noise error term and the b_i, ϕ_i and θ_i are constant parameters.

An alternative to this two-stage approach is to estimate a single well-specified ARMA-regression model for demand, which would include weather variables, dummy variables and ARMA terms. However, the danger is that the autoregression in the weather variables would be modelled by the autoregressive terms and would thus reduce the significance of the weather variables. For example, when we estimated a single well-specified ARMA-regression model for demand, we had to eliminate cooling power of the wind from our model as it was not significant. The result would be that the model fails to identify correctly the weather related component. This could then be a disadvantage when forecasting under conditions of unusual weather, where weather was not conforming to its previous autoregressive structure.

4. USING WEATHER ENSEMBLES FOR DEMAND POINT FORECASTING

4.1 Creating 51 scenarios for weather related electricity demand

A fundamental result in statistics is that the expected value of a non-linear function of random variables is not necessarily the same as the non-linear function of the expected values of the random variables. Let us reconsider the forecast of the weather related demand which was given in expression (3). In view of the definition of cooling power of the wind, given in expression (1), and the presence of the TE_t^2 term in (3), it is clear that the weather related demand is a non-linear function of the fundamental weather variables: temperature, wind speed and cloud cover. The usual approach to forecasting the weather related demand in all electricity demand models simply involves substituting a single high resolution point forecast for each weather variable. Bearing in mind the result regarding the expectation of a non-linear function of random variables, it would be preferable to first construct the probability density function for the weather related electricity demand, and then to calculate the expectation.

Although estimation of the density function of weather related demand is not straightforward, weather ensemble predictions do enable a reasonably sophisticated estimate to be constructed. Since we have 51 ensemble members for temperature, wind speed and cloud cover, we can substitute these 51 weather scenarios into expression (3) to deliver 51 scenarios for weather related demand. The histogram of these 51 demand scenarios is an estimate of the density function. The estimate of the mean is calculated as the mean of the 51 demand scenarios. In Sections 5 and 6, we assess the accuracy of the shape and spread of this estimated distribution. This is less of an issue in this section, as our aim is to estimate the mean of the density function. Meteorologists often find that the mean of the 51 ensemble members for a weather variable is a more accurate forecast of the variable than the single high resolution point forecast. The 51 ensemble members must, therefore, contain information not captured by the single high resolution point forecast. This provides further motivation for forecasting weather related demand using the mean of the 51 demand scenarios.

4.2 Comparison of Forecasting Methods

We used 22 months of daily data from 1 January 1997 to 31 October 1998 to estimate model parameters, and 18 months of daily data from 1 November 1998 to 30 April 2000 to evaluate the different forecasting methods. After eliminating special days from this 18 month period, this gave 500 days for evaluation. We produced forecasts for each day in our evaluation period from lead times of 1 to 10 days-ahead.

After estimating the regression model and the ARMA model for the two-stage approach described in Section 3, we produced forecasts by the usual procedure of substituting traditional single high resolution weather point forecasts in expression (3) for the weather related demand. Using the same models from the two-stage approach, we then produced forecasts using the mean of the 51 scenarios for weather related demand.

In order to establish the limit on demand forecast accuracy that could be achieved with improvements in weather forecast information, we produced demand 'forecasts' using the same two-stage approach with actual observed weather substituted for the weather variables in the weather related demand expression in (3). Clearly this level of weather forecast accuracy is unattainable, as perfect weather forecasts are not achievable.

In order to investigate the benefit of using weather-based methods at different lead times, we produced a further set of benchmark forecasts from the following well-specified model that does not include any of the weather variables:

$$demand_t = c_0 + c_1\, FRI_t + c_2\, SAT_t + c_3\, SUN_t + c_4 W2_t + c_5 W3_t + \varepsilon_t$$
$$\varepsilon_t = \varphi_1\, \varepsilon_{t-1} + \varphi_2\, \varepsilon_{t-2} + \psi_1\, u_{t-1} + u_t$$

where the c_i, φ_i and ψ_1 are constant parameters.

Figure 3 shows the mean absolute percentage error (MAPE) results for the four different methods. We use the MAPE summary measure because it is used extensively in the electricity demand forecasting literature. It is widely accepted that, for 1 day-head forecasting, a weather-based method is preferable to a method that does not use weather information. Indeed, all of the methods entered in the Seattle based 1 day-ahead forecasting competition used temperature as an explanatory variable (*Ramanathan et al.*, 1997). We are not aware of a consensus of opinion regarding lead times up to 10 days-ahead. Our results show that the weather-based methods comfortably dominate the method using no weather variables at all 10 lead times.

It is interesting to note from the MAPE results that, for day-ahead demand forecasting, there is very little difference between the performance of the methods using weather forecasts and that of the benchmark method using actual observed weather. The difference increases steadily with the lead time due to the worsening accuracy of the weather forecasts.

The results show that using weather ensemble predictions, instead of the traditional approach of using single weather point forecasts, led to improvements in accuracy for almost all the 10 lead times. These improvements increased with the lead time, and brought the MAPE results noticeably closer to those of the method using actual observed weather. For lead times of 4 days-ahead or more, the accuracy of the new approach is as good as that of the traditional approach at the previous lead time. This could be described as a gain in accuracy of a day over the traditional approach.

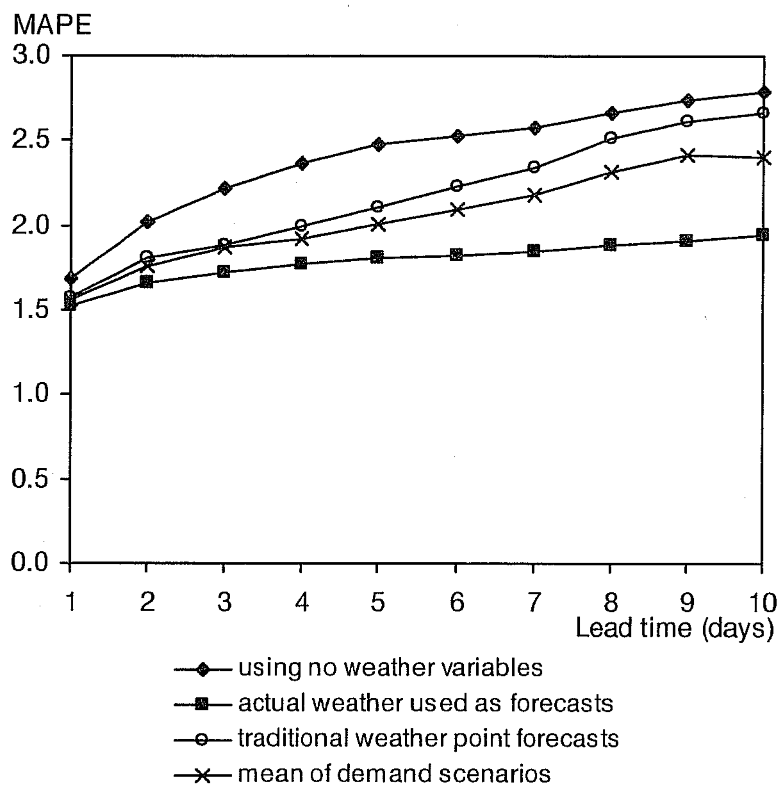

Figure 3: MAPE for electricity demand point forecasts for post-sample period, November 1998 to April 2000.

5. USING WEATHER ENSEMBLES FOR ESTIMATING THE SPREAD OF DEMAND FORECAST ERROR

We now turn our attention to estimating the uncertainty in demand forecasts. In Section 6, we consider the estimation of prediction intervals. In this section, we aim to estimate the spread (standard deviation) of the probability distribution of demand forecast error. Since the method using weather ensemble predictions as input produced the most accurate post-sample forecasts in the previous section, we focus on estimation of the spread of the post-sample errors from this method.

The approach that we take is to model the spread in a series of historic post-sample forecast errors. This was also the approach taken by *Engle et al.* (1993). We compare a variety of univariate and multivariate methods. The multivariate methods incorporate weather ensemble information in the estimate of the error spread. We consider a range of lead times, unlike *Engle et al.* who focused only on 1 day-ahead forecasting. For k day-ahead forecasting, we produce k day-ahead variance forecasts in order to evaluate strictly post-sample estimation accuracy.

5.1. Univariate methods for estimating demand forecast error spread

The simplest estimate of future error standard deviation is the standard deviation of previous forecast errors. For each lead time, *k*, we calculated the variance of the *k* day-ahead errors resulting from the estimation period of 22 months, 1 January 1997 to 31 October 1998. These variances were then used as forecasts for each day in the remaining 18 month period. We term this estimator *naïve* because it assumes the variance is constant over time.

A simple approach, which allows the estimator to adapt over time, is to estimate future *k* day-ahead variance as the simple moving average of recent squared *k* day-ahead errors, e_t. This estimator is often used to forecast the volatility in financial returns. We arbitrarily decided to use a 14 day simple moving average. We call this estimator *sma14*.

Another estimator, which is often used in volatility forecasting, is the exponentially weighted moving average (*ewma*) of recent squared errors. We optimised the smoothing parameter, α, separately for all lead times. The ewma day-ahead estimator is calculated as:

$$\hat{\sigma}_t^2 = \alpha e_{t-1}^2 + (1-\alpha)\hat{\sigma}_{t-1}^2$$

An alternative to the *ad hoc* methods described so far is a statistical modelling approach. Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models are widely used to model volatility in finance (see *Engle*, 1982; *Bollerslev*, 1986). GARCH models express the conditional variance as a linear function of lagged squared error terms and lagged conditional variance terms. For example, the 1 step-ahead GARCH(1,1) variance forecast is given by

$$\hat{\sigma}_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \hat{\sigma}_{t-1}^2$$

Using the standard set of GARCH diagnostics and tests, we fitted an appropriate GARCH model to the set of forecast errors from each of the 10 lead times. We are not aware of other studies that have fitted statistical models to post-sample forecast errors from lead times other than 1 step-ahead. An interesting issue arises in fitting statistical models to *k* step-ahead errors. The series of *k* step-ahead errors from an optimal predictor is likely to possess autocorrelation which can be described by a moving average process of order *k*-1 (see p. 130, Granger and Newbold, 1986). This was evident in our forecast errors. In fitting the GARCH models, we controlled for this by fitting ARMA-GARCH models, where the ARMA component was an MA(*k*-1) process for the series of *k* day-ahead errors. In using the GARCH model for prediction, the MA(*k*-1) components play no part as the prediction is for *k* days-ahead.

5.2. Using weather ensemble predictions to estimate demand forecast error spread

If there is a strong degree of uncertainty in the weather variables, then there will be a strong level of uncertainty in the demand forecasts. Therefore, there is strong motivation for using a measure of weather forecast uncertainty to model the uncertainty in the demand forecasts. Weather ensemble predictions provide information regarding the uncertainty in weather, so the issue is how best to use this information to model demand forecast uncertainty. The spread of the 51 demand scenarios, discussed in Section 4, conveys the uncertainty in the weather component of demand. We calculated the standard deviation, $\sigma_{ENS\,t}$, of the 51 scenarios for each day and for each of the 10 lead times. However, this is likely to underestimate the standard deviation of the demand forecast error because it does not accommodate the uncertainty due to the model error and the parameter estimation error associated with expressions (2) and (3). In view of this, for each lead time, we performed a linear bias correction by regressing the absolute value of forecast error on $\sigma_{ENS\,t}$. We used the first 9 (1 November 1998 to 31 July 1999) of the 18 months of ensemble predictions to estimate the bias correction model parameters. The bias corrected estimator is of the form:

$$\hat{\sigma}_t = \hat{a} + \hat{b}\,\sigma_{ENS,t}$$

Explanatory variables can be included in GARCH models. Since there is likely to be useful information in the weather ensemble predictions that is not captured by the univariate time series extrapolation of the GARCH model, we estimated GARCH models with $\sigma^2_{ENS,t}$ as an explanatory variable. In this paper, we refer to these models as *mixed* GARCH models. The day-ahead GARCH variance forecast is then given by

$$\hat{\sigma}_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \hat{\sigma}_{t-1}^2 + \gamma_1 SAT_t + \gamma_2 \sigma_{ENS,t}^2 \qquad (4)$$

We experimented with additional explanatory variables but the dummy variable for Saturdays, SAT_t, was the only one that was significant. The coefficient of $\sigma_{ENS\,t}$ was significant in the model for each of the 10 lead times. As with the bias correction parameters, we used 9 of the 18 months of ensemble predictions to estimate the model parameters. *Engle et al.* (1993) modelled prediction day-ahead error spread in terms of independent and lag variables using the absolute value of the error as dependent variable. However, they did not have access to weather ensemble predictions, and so their choice of regressors differed from those in expression (4).

5.3. Combining standard deviation estimators

If several forecasts are produced for the same period from different information sources, then a combination of the forecasts may be able to synthesise this information to deliver an improved prediction (*Bunn*, 1989). Combining is, therefore, an alternative to the mixed GARCH model for bringing information together from the ensemble predictions and from the history of the spread of the forecast errors.

To select a univariate method for combining, we compared accuracy on the same 9 months of data used to estimate the bias correction parameters (1 November 1998 to 31 July 1999). The sma14 method was overall the most accurate so we included this estimator in the combination. We generated simple average and regression combinations of the sma14 estimator with the bias corrected $\sigma_{ENS\,t}$ estimator. The simple average is the simplest and most widely used combining approach. It is considered to be robust since the weights are not estimated and sum to one. The obvious disadvantage of the simple average combination is that equal weights will be inappropriate when one forecast tends to be superior to the other. *Granger and Ramanathan* (1986) propose the use of unrestricted regression to derive combining weights:

$$\hat{\sigma}_t = w_0 + w_1 \hat{\sigma}_{1t} + w_2 \hat{\sigma}_{2t}$$

Although multicollinearity can be a problem for this approach, it has the advantage of being able to correct for bias in the individual forecasts, unlike the many combining methods that restrict the combining weights to sum to one. The regression parameters were estimated using the 9 months of data from 1 November 1998 to 31 July 1999.

5.4. Comparison of standard deviation estimators

We evaluated post-sample forecasting performance for the 10 standard deviation estimators using the 9 month period from 1 August 1999 to 30 April 2000. After eliminating special days, this period consisted of 252 days. Table I shows the results for the R^2 evaluation diagnostic. (In the remainder of this paper, we present results in tables, as there are too many methods to permit graphical comparison.) The R^2 is the coefficient of determination from a LS regression of the absolute value of the *post-sample* forecast error on the volatility estimator; high values are preferable. The measure is often used to evaluate volatility forecasts in finance (see *Jorion*, 1995, and *Taylor*, 1999). The absolute value of the error acts as a proxy for the unobservable standard deviation. The regression corrects for any bias and the R^2 measures the degree to which the estimator varies with the spread of the error. It is, therefore, a measure of the information content of the estimator. Typically, the R^2 values are very low. For example, Jorion (1995) recorded values of between 1.9% and 5.2% in his work with foreign exchange futures.

	Lead time (days)									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Univariate										
naïve	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
sma14	2.0	1.7	0.9	0.3	0.5	1.1	2.1	4.7	7.1	10.0
ewma	2.7	2.1	1.8	0.0	0.0	0.0	0.0	4.3	0.0	0.0
garch	2.2	4.4	2.5	0.1	0.5	0.9	1.8	2.0	4.3	0.9
Ensemble based										
stdev of scenarios	1.7	5.4	2.0	3.5	4.0	6.4	7.5	11.9	11.6	7.1
bias corrected stdev	1.7	5.4	2.0	3.5	4.0	6.4	7.5	11.9	11.6	7.1
mixed garch	1.7	7.7	2.0	3.5	4.3	6.4	7.5	12.4	12.4	8.8
Combining sma14 and stdev of scenarios										
average	2.9	3.7	1.7	1.8	1.7	6.4	4.6	9.6	11.6	11.0
regression	3.0	6.0	2.3	3.5	3.9	3.2	6.7	12.0	12.1	8.5

Table I: R^2 measure for forecast error standard deviation estimation methods for post-sample period, August 1999 to April 2000.

The R^2 for the naïve estimator was zero for all lead times since it does not vary during the 9 month evaluation period. The sma14 estimator outperformed the other univariate methods beyond 3 days-ahead. The results show that the three weather ensemble based methods comfortably outperformed all the univariate methods at almost all the lead times. The results for the standard deviation of scenarios method and the bias corrected method are identical since the R^2 measures covariation after performing a bias correction on the estimator. The regression method tended to be the better of the two combination methods. Overall, there is little to choose between the regression combination, the bias corrected standard deviation of scenarios method and the mixed GARCH approach. However, the bias corrected standard deviation of scenarios approach is the simplest to compute, and so, on the basis of the R^2 measure, we feel that this is the most attractive method of the ten considered.

Table II shows the post-sample evaluation results using the following mean absolute error (MAE) evaluation diagnostic; low values are preferable:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left\| e_i \right| - \hat{\sigma}_i \right|$$

	Lead time (days)									
	1	2	3	4	5	6	7	8	9	10
Univariate										
naïve	401	401	421	458	456	470	484	498	512	542
sma14	408	405	433	471	463	462	477	480	482	476
ewma	389	386	413	464	443	473	472	481	510	531
garch	399	377	394	431	424	430	444	469	469	498
Ensemble based										
stdev of scenarios	449	457	460	457	451	446	464	485	499	515
bias corrected stdev	361	367	397	432	417	426	430	432	451	479
mixed garch	361	380	397	432	411	426	430	436	455	481
Combining sma14 and stdev of scenarios										
average	376	379	408	443	432	426	445	440	450	459
regression	357	365	396	435	415	437	431	433	450	473

Table II: MAE (megawatts) for forecast error standard deviation estimation methods for post-sample period, August 1999 to April 2000.

Unlike the R^2, the MAE does not correct for bias, and so the results of Table II are a more straightforward reflection of forecasting performance. The most accurate univariate method is GARCH. In view of the R^2 results, it is perhaps surprising to see that sma14 is the poorest of the univariate methods for the first five lead times; there must be a degree of bias in sma14 at these early lead times. In Section 5.2, we anticipated that bias would be a major issue for the standard deviation of scenarios method. Comparing the fifth and sixth rows of results in Table II, we can see that the MAE results for this estimator notably improve with the application of the bias correction.

The bias corrected standard deviation of scenarios method outperforms the univariate methods and it also tends to outperform the mixed GARCH method. The results for the combination methods show that there was little benefit in combining over simply using the bias corrected standard deviation of scenarios approach.

6. USING WEATHER ENSEMBLES FOR ESTIMATING DEMAND PREDICTION INTERVALS

The most common way of conveying the uncertainty in a forecast is a prediction interval. In this section, we consider a number of ways of estimating prediction intervals for electricity demand forecasts. Although 95% and 90% intervals are most common in the research literature, *Granger* (1996) suggests that 50% intervals are also widely used by practitioners. He points out that 50% intervals are more robust to distributional assumptions and are less affected by outliers. He criticises 95% limits for often being embarrassingly wide, and thus not very useful. In order to consider both the tails and the body of the predictive distribution, we focus on estimation of 50% and 90% intervals. More specifically, we evaluate different approaches to estimating the bounds of these intervals: the 5%, 25%, 75% and 95% quantiles. The θ% quantile of the probability distribution of a variable y is the value, $Q(\theta)$, for which $P(y<Q(\theta))=\theta$.

6.1. Estimating quantiles using the standard deviation estimators

The standard deviation estimators, investigated in Section 5, can be used as the basis for estimating the quantiles. A common approach to generating interval limits is to assume that the conditional distribution is Gaussian. For example, the 95% quantile is estimated using $1.645\,\hat{\sigma}_t$. An alternative for estimating the θ% quantile is to multiply each standard deviation estimator by the θ% quantile of the empirical distribution of the corresponding standardised forecast errors, $e_t / \hat{\sigma}_t$. Gaussian and empirical estimators were used by Granger *et al.* (1989), and both were implemented in this study.

6.2. Estimating quantiles using the distribution of the weather-related demand scenarios

The distribution of weather related demand scenarios underlies the better standard deviation estimators in Section 5. In view of this, one might surmise that the quantiles of this distribution would be useful for estimating the quantiles of the predictive distribution. One approach is simply to use the quantiles, $Q_{ENS,t}(\theta)$, of the distribution of scenarios as estimates of the quantiles of the forecast error distribution. It seems likely that this will deliver biased quantile estimates, as the spread of the predictive distribution is likely to be wider than the spread of the scenario distribution. Therefore, we applied a bias correction to the quantile of the distribution of demand scenarios. We used quantile regression to perform bias correction, as proposed by Granger (1989). The forecast error series was used as dependent variable and $Q_{ENS,t}(\theta)$ as regressor. The form of the resultant bias corrected estimator was then:

$$\hat{Q}_t(\theta) = \hat{a} + \hat{b}\, Q_{ENS,t}(\theta)$$

Since quantile regression is not well known, a brief description here is probably useful. *Koenker and Bassett* (1978, 1982) developed the theory of quantile regression for the estimation of the quantiles of a variable y_t. Suppose that the $\theta\%$ quantile of y_t can be written as a linear function of explanatory variables. Let us express these variables as elements of a vector $\boldsymbol{x}_t$.

$$Q_{yt}(\theta \mid \boldsymbol{x}_t) = \boldsymbol{x}_t \beta(\theta) \tag{5}$$

where $\beta(\theta)$ is a vector of parameters dependent on θ. *Koenker and Bassett* (1978) defined the *θth regression quantile* ($0<\theta<1$) as any solution, $\beta(\theta)$, to the *quantile regression* minimisation:

$$\min_{\beta} \left(\sum_{t \mid y_t \geq x_t\beta} \theta \left| y_t - \boldsymbol{x}_t\beta \right| + \sum_{t \mid y_t < x_t\beta} (1-\theta) \left| y_t - \boldsymbol{x}_t\beta \right| \right)$$

Koenker and Bassett (1982) showed that quantile regression delivers parameters that asymptotically approach the parameters, $\beta(\theta)$, in (5) as the number of observations increases.

The common procedure for building an explanatory model for a variable is to look for a relationship between past observations of that variable and past observations of potential explanatory variables. This is not a feasible procedure for building a model for the quantiles of a variable because past observations of the quantiles will not be available, as they are unobservable. The appeal of quantile regression is that past observations of the quantiles are not required. Instead, the variable itself, y_t, is regressed on explanatory variables, $\boldsymbol{x}_t$, to produce a model for the quantile. The software packages STATA (*Stata*, 1993) and SHAZAM (*White*, 1997) perform quantile regression.

6.3. Combining quantile estimators

We combined a univariate quantile estimator and the bias corrected quantile of the distribution of weather-related demand scenarios. The sma14 with empirical distribution was chosen as the univariate estimator in order to be consistent with the spread estimator study of Section 5. We used simple average and quantile regression combining which was proposed by Granger *et al.* (1989). The latter involves the quantile regression of the forecast error series on two quantile estimators, $\hat{Q}_{1t}(\theta)$ and $\hat{Q}_{2t}(\theta)$, to deliver an estimator of the form:

$$\hat{Q}_t(\theta) = w_0 + w_1 \hat{Q}_{1t}(\theta) + w_2 \hat{Q}_{2t}(\theta)$$

6.4. Comparison of quantile estimators

The most popular measure of quantile estimator accuracy is the percentage of observations falling below the quantile estimator. For an unbiased estimator of the $\theta\%$ quantile, this will be $\theta\%$. This criterion is used extensively for evaluating quantile estimators and prediction intervals (e.g. *Granger et al.*, 1989; *Taylor and Bunn*, 1999). In this paper, we use this measure as a basis for comparison of the estimators.

Table III compares estimation of the 5% quantiles at the 10 different lead times (1 to 10 days-ahead) for the post-sample period of 9 months, August 1999 to April 2000. The table shows the percentage of post-sample forecast errors falling below the quantile estimators. The asterisks indicate the entries that are significantly different from the ideal value at the 5% significance level. The acceptance region for the hypothesis test is constructed using a Gaussian distribution and the standard error formula for a proportion. The cluster of asterisks in the first six rows of values in Table III highlights the strong tendency for the estimators using the Gaussian assumption to underestimate the width of the forecast error distribution. Four of the six estimators based on standard deviation estimators performed better with the empirical distribution of standardised forecast errors than with the Gaussian distribution. A notable exception is the naïve standard deviation estimator with the Gaussian distribution, which compared favourably with the best of the other estimators. In Section 5.2, we discussed how the spread of the distribution of demand scenarios is likely to underestimate the spread of the forecast error distribution. In view of this, it is not surprising that the results are very poor for the estimator which was constructed simply from the quantiles of the standard deviation of the 51 weather related demand scenarios; the width of the predictive distribution was heavily underestimated. However, the bias correction led to a notable improvement. The combination methods perform reasonably well. It is not easy to see which of the 16 methods is best overall.

	Lead time (days)									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
St dev estimators with Gaussian										
naïve	5.6	5.2	6.7	6.7	5.6	6.0	6.0	4.8	4.8	5.2
sma14	6.7	5.6	6.0	5.6	7.9*	8.7*	8.3*	7.9*	6.7	6.3
ewma	6.0	4.8	5.2	5.2	7.1	6.0	6.0	6.0	5.2	4.8
garch	4.8	6.3	8.7*	9.5*	8.7*	9.1*	10.7*	12.7*	9.9*	12.7*
bias corrected stdev	10.3*	7.9*	9.5*	8.3*	7.9*	9.1*	9.1*	10.3*	10.7*	10.7*
mixed garch	4.0	4.0	7.1	8.3*	8.7*	7.1	12.7*	12.3*	13.5*	14.3*
St dev estimators with empirical										
naïve	4.4	4.4	3.2	4.4	4.0	3.6	4.4	3.2	3.2	2.0*
sma14	5.2	4.4	4.0	5.2	6.3	8.3*	7.9*	6.7	5.2	5.2
ewma	6.3	7.5	4.0	4.8	4.0	3.6	4.0	3.2	3.6	2.4
garch	4.4	3.6	2.8	4.8	4.4	4.4	4.4	3.2	3.6	2.0*
bias corrected stdev	3.6	3.2	2.4	4.4	4.0	2.8	3.2	3.6	5.2	5.6
mixed garch	6.0	2.8	3.2	4.4	4.0	4.8	4.8	4.0	6.3	6.0
Demand scenario quantile										
scenario qu	43.7*	46.0*	27.0*	25.0*	20.6*	20.6*	16.7*	20.6*	17.9*	17.9*
bias corrected scenario qu	4.4	3.6	3.2	4.0	4.4	2.8	3.6	2.8	3.6	4.4
Combining bias corrected scenario quantile and sma14 with empirical										
simple average	4.4	3.6	3.2	4.8	5.6	4.4	5.2	4.4	3.6	4.0
quantile regression	4.4	3.6	5.2	4.0	5.6	6.3	4.0	5.2	4.0	4.4

Table III: Percentage of errors falling below estimates of 5% forecast error quantile for post-sample period, August 1999 to April 2000 indicates significant at 5% level.

Table IV compares post-sample performance for the 25% quantile. For this quantile, the use of an empirical distribution with the standard deviation estimators is clearly preferable to the Gaussian assumption. The fact that this was not so clearly the case for the 5% quantile contradicts Granger's (1996) comment that 50% prediction intervals, and hence 25% and 75% quantiles, tend to be more robust to distributional assumptions than intervals with bounds in the tails of the distribution. Overall, the standard deviation estimators with empirical distribution and the combinations perform relatively well.

We do not report the detailed results for the estimation of the 75% and 95% quantiles as the relative performances of the estimators were broadly similar to the results for the 25% and 5% quantiles, respectively. We chose to report the 5% and 25% results as the lower half of the distribution tends to be more important for electricity demand scheduling purposes; the problems caused by a shortfall in electricity availability tend to be more serious than those resulting from an oversupply of the same size.

	Lead time (days)									
	1	2	3	4	5	6	7	8	9	10
St dev estimators with Gaussian										
naïve	30.6*	31.7*	33.3*	28.6	28.2	27.8	27.0	28.2	28.2	27.8
sma14	29.0	36.1*	34.9*	30.2	31.0*	29.4	30.6*	29.0	27.8	28.2
ewma	30.6*	34.1*	33.7*	28.2	29.4	26.6	27.0	28.6	27.4	28.2
garch	30.2	35.3*	36.9*	32.9*	31.7*	32.5*	30.6*	33.3*	31.7*	31.7*
bias corrected stdev	34.5*	36.5*	35.7*	32.5*	32.1*	31.7*	31.7*	34.5*	33.7*	32.5*
mixed garch	29.4	34.1*	34.9*	31.3*	33.3*	30.6*	35.7*	35.7*	35.3*	36.1*
St dev estimators with empirical										
naïve	29.0	27.4	21.8	24.6	23.4	21.8	23.0	25.0	28.2	27.0
sma14	27.4	30.2	27.8	26.2	26.2	25.0	27.0	26.2	26.2	26.2
ewma	29.4	31.0*	27.8	24.6	23.4	21.8	23.8	25.0	27.4	27.8
garch	27.8	25.4	23.8	24.6	22.6	21.0	23.4	25.0	25.8	27.8
bias corrected stdev	28.6	25.4	25.0	21.4	24.6	21.8	24.6	24.2	25.8	29.0
mixed garch	28.6	26.2	22.2	24.6	25.4	20.2	25.8	23.4	27.8	30.6*
Demand scenario quantile										
scenario qu	52.0*	53.6*	31.3*	37.7*	25.4	32.9*	23.4	30.6*	22.2	27.0
bias corrected scenario qu	29.0	26.6	25.8	22.6	21.4	20.2	21.4	21.8	22.6	24.6
Combining bias corrected scenario quantile and sma14 with empirical										
simple average	27.4	27.4	28.2	22.6	24.2	22.2	23.4	22.2	22.6	26.2
quantile regression	27.0	26.6	27.0	23.0	25.8	23.4	24.2	23.4	21.8	23.0

Table IV: Percentage of errors falling below estimates of 25% forecast error quantile for post-sample period, August 1999 to April 2000. indicates significant at 5% level.

To summarise the relative overall performance for the methods at the different lead times, we calculated chi-squared goodness of fit statistics. For each method, at each lead time, we calculated the statistic for the total number of post-sample forecast errors falling within the following five categories: below the 5% quantile estimator, between the 5% and 25% estimators, between the 25% and 75%, between the 75% and 95%, and above the 95%. Table V shows the resulting chi-squared statistics. The asterisks indicate significance at the 5% level. Unfortunately, we cannot sum the chi-squared statistics across lead times to give a single summary measure for each of the estimators because the chi-squared statistics for the different lead times are not independent. The results show that the use of an empirical distribution is preferable to using a Gaussian distribution. Of the estimators that used an empirical distribution, we would tentatively suggest that the GARCH estimator had the most consistently good performance across the lead times. Perhaps, not surprisingly, the performance of the estimators tends to weaken with the longer lead times. The simple average combination performs the best for two lead times, the quantile regression combination is the best for three lead times, whilst none of the other methods is the best for more than one lead time. This suggests strong potential for combining a univariate estimator with one based on weather ensemble predictions.

	Lead time (days)									
	1	2	3	4	5	6	7	8	9	10
St dev estimators with Gaussian										
naïve	10.3*	10.4*	12.2*	7.1	2.6	2.7	2.2	6.5	6.5	7.8
sma14	14.8*	25.1*	25.2*	16.7*	21.9*	20.7*	16.0*	22.3*	15.1*	24.3*
ewma	8.9	18.1*	19.5*	6.3	6.3	2.2	3.6	8.7	6.8	16.1*
garch	10.9*	20.1*	32.3*	27.7*	16.0*	24.2*	34.0*	60.3*	42.3*	75.8*
bias corrected stdev	28.5*	24.5*	31.9*	20.5*	16.4*	17.7*	28.3*	48.9*	85.0*	89.0*
mixed garch	12.8*	19.1*	20.3*	14.6*	20.2*	10.7*	89.9*	91.6*	132.6*	172.1*
St dev estimators with empirical										
naïve	4.9	9.4	4.6	1.1	1.1	1.8	3.7	5.9	11.0*	18.1*
sma14	1.3	8.1	7.3	0.8	2.5	13.4*	11.6*	9.9*	11.5*	23.6*
ewma	8.6	10.4*	4.8	2.0	1.1	1.8	10.0*	4.7	9.2	26.3*
garch	3.1	7.9	4.1	0.7	0.9	3.1	4.3	5.9	4.6	20.7*
bias corrected stdev	7.2	6.9	5.6	2.1	1.1	3.1	4.9	7.0	10.5*	31.0*
mixed garch	3.8	10.6*	3.5	1.1	0.8	4.4	8.3	11.6*	34.4*	41.5*
Demand scenario quantile										
scenario qu	1122.2*	1165.8*	1334.4*	639.2*	1536.0*	676.6*	1240.3*	831.7*	1167.4*	1465.5*
bias corrected scenario qu	4.5	7.5	3.8	2.8	1.7	5.3	2.5	3.3	2.8	34.7*
Combining bias corrected scenario quantile and sma14 with empirical										
simple average	2.2	5.4	6.7	1.0	0.8	1.8	5.4	3.2	4.0	20.8*
quantile regression	1.8	9.9*	12.2*	3.0	0.2	2.6	0.9	0.6	2.9	51.7*

Table V. Chi-squared statistics summarising overall estimator bias for 5%, 25%, 75% and 95% quantiles for the post-sample period, August 1999 to April 2000. *indicates significant at 5% level.

The percentage of errors falling below a quantile estimator evaluates only the bias; we should also consider the variability of the estimation error. For example, consider estimation of the 5% quantile of the 1 day-ahead predictive distribution using the following two estimators: naïve with empirical and GARCH with empirical. The first column of results in Table III shows that 4.4% of the errors fell below both of these estimators. Since the ideal is 5%, both estimators are a little low on average; they possess a degree of bias. Although the bias is equal for the two estimators, one would surmise that the GARCH based estimator would vary in accordance with the varying spread of the distribution better than the naïve estimator, which by construction does not vary at all. It would be useful if we could evaluate this variability characteristic.

In the context of evaluating volatility forecasts, the LS regression R^2 measure has the appeal that the forecasts are effectively corrected for bias, so that the R^2 then reflects variation about the bias. In view of this, there is strong appeal for using a quantile regression R^2 measure to evaluate quantile estimators (*Taylor*, 1999). By performing a quantile regression of the *post-sample* forecast error series, e_t, on the quantile estimator, $\hat{Q}_t(\theta)$, we correct for bias in the estimator before assessing accuracy. The package STATA (*Stata*, 1993) provides a pseudo-R^2 which is analogous to the R^2 in LS regression.

$$\text{pseudo-}R^2 = 1 - \frac{\text{sum of weighted deviations about debiased quantile}}{\text{sum of weighted deviations about raw quantile}}$$

$$= 1 - \frac{\sum_{t \mid e_t \geq \hat{\alpha} + \hat{\beta}\hat{Q}_t(\theta)} \theta \left| e_t - \hat{\alpha} - \hat{\beta}\hat{Q}_t(\theta) \right| + \sum_{t \mid y_t < \hat{\alpha} + \hat{\beta}\hat{Q}_t(\theta)} (1-\theta) \left| e_t - \hat{\alpha} - \hat{\beta}\hat{Q}_t(\theta) \right|}{\sum_{t \mid e_t \geq \hat{Q}(\theta)} \theta \left| e_t - \hat{Q}(\theta) \right| + \sum_{t \mid e_t < \hat{Q}(\theta)} (1-\theta) \left| e_t - \hat{Q}(\theta) \right|}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the parameters derived when e_t is regressed on $\hat{Q}_t(\theta)$. The *raw quantile*, $\hat{Q}(\theta)$, is a time-invariant estimator given by the *θ%* quantile regression with e_t as dependent variable, and the inclusion of an intercept but no regressors.

Tables VI and VII show the pseudo-R^2 for estimation of the 5% and the 25% quantiles respectively; high values of the pseudo-R^2 are preferable. Let us first reconsider the comparison of the performance of the naïve with empirical estimator and the GARCH with empirical estimator for the 5% quantile of the 1 day-ahead predictive distribution. Earlier we noted that the two were equally biased. The first column of results in Table VI shows that the pseudo-R^2 for the naïve estimator is zero, but for the GARCH based estimator, it is 12.7%. These results reflect the fact that the naïve estimator does not vary at all, whilst the GARCH estimator tends to vary with the unobservable quantile to some degree. Note that each of the quantile estimators which are based on a standard deviation estimator from Section 5 have the same pseudo-R^2 when a Gaussian distribution is used as to when an empirical distribution is used. This is because the quantile regression used to estimate the numerator of the pseudo-R^2 performs bias correction and so any constant multiplicative factor used to convert the standard deviation estimator to a quantile estimator ceases to be relevant in calculating the pseudo-R^2.

Table VI suggests that the estimators based on bias corrected standard deviation of the scenarios and those based on the mixed GARCH standard deviation tend to have the highest pseudo-R^2. Many of the others perform well at the early lead times but disappointingly for the longer horizons. Interestingly, the results were very poor for the two estimators based on the quantiles of the distribution of weather related demand scenarios. Although the bias results in Table III were quite respectable for the bias corrected version of this 5% quantile estimator, Table VI indicates that there is really very little covariation between the estimator and the unobservable error quantile. Table VII shows that the relative performance of the estimators for the 25% quantiles was broadly similar to those for the 5% quantiles. One notable difference is that the estimators based on the ewma standard deviation estimator were comfortably the best for the early lead times. The pseudo-R^2 results for the 75% and 95% estimators were broadly similar to the results for the 25% and 5% quantiles, respectively.

Overall, there is no clear single best method according to the estimation bias results in Table V and the variability results in Tables VI and VII. The two combinations and the mixed GARCH with empirical distribution approach perform well. However, in view of its relative simplicity, we feel that there is strong appeal to the estimator formed by using the bias corrected standard deviation of the weather related demand scenarios with an empirical distribution.

	Lead time (days)									
	1	2	3	4	5	6	7	8	9	10
St dev estimators with Gaussian or empirical										
naïve	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
sma14	11.2	8.7	4.0	5.3	3.5	1.7	1.1	0.1	0.4	2.6
ewma	11.0	9.7	6.6	6.0	0.0	0.0	0.1	3.5	0.2	1.2
garch	12.7	6.4	5.8	2.2	6.0	0.7	0.8	0.0	0.3	0.3
bias corrected stdev	5.0	5.9	3.7	3.1	5.9	10.7	7.6	8.3	8.1	3.7
mixed garch	11.1	10.1	3.2	4.1	7.5	0.4	8.3	8.6	6.9	4.6
Demand scenario quantile										
scenario qu	0.2	3.3	0.5	0.2	0.8	0.5	1.1	0.7	0.6	0.7
bias corrected scenario qu	0.2	3.3	0.5	0.2	0.8	0.5	1.1	0.7	0.6	0.7
Combining bias corrected scenario quantile and sma14 with empirical										
simple average	10.5	9.9	5.0	5.3	2.0	2.0	1.9	0.2	0.7	1.8
quantile regression	7.0	4.8	3.9	2.7	0.3	1.6	1.2	0.6	0.6	0.5

Table VI: Pseudo R^2 for estimators of 5% forecast error quantile for post-sample period, August 1999 to April 2000.

	Lead time (days)									
	1	2	3	4	5	6	7	8	9	10
St dev estimators with Gaussian or empirical										
naïve	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
sma14	0.6	0.9	1.6	0.7	0.5	0.3	0.5	1.8	2.6	4.0
ewma	1.7	5.5	4.5	4.8	0.0	0.0	0.6	2.1	0.9	3.1
garch	0.7	3.2	4.0	0.5	0.6	0.4	0.2	0.0	1.8	0.5
bias corrected stdev	0.2	1.1	1.2	0.8	1.2	1.0	2.3	3.8	3.6	2.3
mixed garch	0.3	4.4	1.3	2.5	1.4	0.4	3.9	4.7	3.8	3.7
Demand scenario quantile										
scenario qu	1.0	0.1	0.0	0.0	0.9	1.0	0.2	0.1	1.6	0.1
bias corrected scenario qu	1.0	0.1	0.0	0.0	0.9	1.0	0.2	0.1	1.6	0.1
Combining bias corrected scenario quantile and sma14 with empirical										
simple average	0.7	0.9	1.6	0.8	1.3	0.6	0.5	1.5	3.1	4.4
quantile regression	0.9	0.8	1.6	0.2	1.5	0.5	0.3	1.9	3.0	4.1

Table VII: Pseudo R^2 for estimators of 25% forecast error quantile for post-sample period, August 1999 to April 2000.

7. SUMMARY AND CONCLUSIONS

We have investigated the scope for using weather ensemble predictions in electricity demand forecasting for lead times from 1 to 10 days-ahead. We used the 51 ensemble members for each weather variable to produce 51 scenarios for the weather-related component of electricity demand. For almost all 10 lead times, the mean of the demand scenarios was a more accurate demand forecast than that produced by the traditional procedure of substituting a single point forecast for each weather variable in the electricity demand model. Since demand is a nonlinear function of weather variables, this traditional procedure amounts to approximating the expectation of a nonlinear function of random variables by the same non-linear function of the expected values of the random variables. The mean of the 51 scenarios is appealing because it is equivalent to taking the expectation of an estimate of the demand probability density function.

The distribution of the 51 demand scenarios provides information regarding the uncertainty in the demand forecast. However, since the distribution does not accommodate demand model uncertainties, it will tend to underestimate the forecast uncertainty. In view of this, we applied a linear bias correction to inflate measures of standard deviation and quantiles taken from the scenario distribution. The resulting standard deviation estimator compared favourably with estimators produced using univariate volatility forecasting methods. Using the same standard deviation estimator as a basis for estimating prediction intervals also compared well with univariate methods. We, therefore, conclude that there is strong potential for the use of weather ensemble predictions in improving the accuracy and uncertainty assessment of electricity demand forecasts.

Acknowledgements

The authors are very grateful to Shanti Majithia and Chris Rogers of the National Grid Company for supplying data and information regarding the company's approach to demand forecasting, and to Tony Hollingsworth of the European Centre for Medium-range Weather Forecasts for helpful suggestions with the paper.

References

Baker AB. 1985. Load forecasting for scheduling generation on a large interconnected system. In: *Comparative Models for Electrical Load Forecasting*, Bunn DW, Farmer ED (eds), Wiley: Chichester, pp 57-67.

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31:** 307-327.

Buizza R. 1997. Potential forecast skill of ensemble prediction, and spread and skill distributions of the ECMWF Ensemble Prediction System. *Monthly Weather Review* **125:** 99-119.

Buizza R, Miller M, Palmer TN. 1999. Stochastic simulation of model uncertainties. *Quarterly Journal of the Royal Meteorological Society* **125:** 2887-2908.

Buizza R, Petroliagis T, Palmer TN, Barkmeijer J, Hamrud M, Hollingsworth A, Simmons A, Wedi N. 1998. Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society* **124**: 1935-1960.

Bunn DW. 1989. Forecasting with more than one model. *Journal of Forecasting* **8:** 161-166.

Connor JT. 1996. A robust neural network filter for electricity demand prediction. *Journal of Forecasting* **15:** 437-458.

Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50:** 987-1008.

Engle R, Granger CWJ, Ramanathan R, Vahid-Araghi F. 1993. Probabilistic methods in forecasting hourly loads and peaks. Electric Power Research Institute, EPRI TE-101902, Palo Alto, California.

Epstein ES. 1969. Stochastic dynamic prediction. *Tellus*, **21:** 739-759.

Fleming RJ. 1971a. On stochastic dynamic prediction. I: The energetics of uncertainty and the question of closure. *Monthly Weather Review*, **99:** 851-872.

Fleming RJ. 1971b. On stochastic dynamic prediction. II: predictability and utility. *Monthly Weather Review*. **99:** 927-938.

Gleeson TA. 1970. Statistical-dynamical predictions. *Journal of Applied Meteorology* **9:** 333-344.

Granger CWJ. 1989. Combining forecasts - twenty years later. *Journal of Forecasting* **8:** 167-173.

Granger CWJ. 1996. Can we improve the perceived quality of economic forecasts? *Journal of Applied Econometrics* **11:** 455-473.

Granger CWJ, Newbold P. 1986. *Forecasting Economic Time Series*, London: Academic Press (2nd edition).

Granger CWJ, Ramanathan R. 1984. Improved methods of combining forecasts. *Journal of Forecasting* **3:** 197-204.

Granger CWJ, White H, Kamstra M. 1989. Interval forecasting: an analysis based upon ARCH-quantile estimators. *Journal of Econometrics* **40:** 87-96.

Harvey AC, Koopman SJ. 1993. Forecasting hourly electricity demand using time-varying splines, *Journal of the American Statistical Association* **88:** 1228-1236.

Jorion P. 1995. Predicting volatility in the foreign exchange market. *Journal of Finance* **50:** 507-528.

Koenker RW, Bassett GW. 1978. Regression quantiles. *Econometrica* **46:** 33-50.

Koenker RW, Bassett GW. 1982. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50:** 43-62.

Leith CE. 1974. Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review* **102:** 409-418.

Lorenz EN. 1993. *The essence of chaos*, UCL Press.

Molteni F, Buizza R, Palmer TN, Petroliagis T. 1996. The new ECMWF ensemble prediction system: methodology and validation. *Quarterly Journal of the Royal Meteorological Society* **122:** 73-119.

Palmer TN, Molteni F, Mureau R, Buizza R, Chapelet P, Tribbia J. 1993. Ensemble prediction. Proceedings of the ECMWF Seminar on *Validation of models over Europe: vol. I*, ECMWF, Shinfield Park, Reading, RG2 9AX, UK.

Ramanathan R, Engle R, Granger CWJ, Vahid-Araghi F, Brace C. 1997. Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting* **13:** 161-174.

Stata. 1993. STATA 3.1, Stata Corporation, College Station: TX.

Taylor JW. 1999. Evaluating volatility and interval forecasts. *Journal of Forecasting* **18:** 111-128.

Taylor JW, Bunn DW. 1999. A quantile regression approach to generating prediction intervals. *Management Science* **45:** 225-237.

Taylor JW, Majithia S. 2000. Using combined forecasts with changing weights for electricity demand profiling. *Journal of the Operational Research Society* **51:** 72-82.

Toth Z, Kalnay E. 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bulletin of the American Meteorological Society* **74:** 2317-2330.

Tracton MS, Kalnay E. 1993. Operational ensemble prediction at the National Meteorological Center: practical aspects. *Weather and Forecasting* **8:** 379-398.

White K. 1997. *SHAZAM User's Reference Manual Version 8.0*, McGraw-Hill: New York.