International Journal of Forecasting 19 (2003) 57–70
www.elsevier.com/locate/ijforecast

Using weather ensemble predictions in electricity demand forecasting

James W. Taylor[a,*], Roberto Buizza[b]

[a]Saïd Business School, University of Oxford, Park End Street, Oxford OX1 1HP, UK
[b]European Centre for Medium-range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK

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.

Keywords: Energy forecasting; Weather ensemble predictions; Forecasting accuracy; Prediction intervals

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, Buizza, Palmer, & Petroliagis, 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 (Adams, Allen, & Morzuch, 1991). A measure of risk is also beneficial for those trading electricity.

In this paper, our analysis is based on daily electricity demand data. We use the demand forecast-

*Corresponding author. Tel.: +44-1865-288-927; fax: +44-1865-288-805.
E-mail addresses: james.taylor@sbs.ox.ac.uk (J.W. Taylor), r.buizza@ecmwf.int (R. Buizza).

0169-2070/02/$ – see front matter
PII: S0169-2070(01)00123-6

ing methodology of the National Grid (NG) as a basis for our analysis. NG 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. Since the re-structuring of the industry in 1990, the NG demand forecasts have also been an important influence on the price and dynamics of the electricity market. Accurate demand forecasts are required by utilities who need to predict their customers' demand, and by those wishing to trade electricity as a commodity on financial markets.

Weather ensemble predictions are described in Section 2. Section 3 presents the method and variables currently used by NG. 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 variance 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 one single model-integration with best estimates for the initial conditions (see Fig. 1).

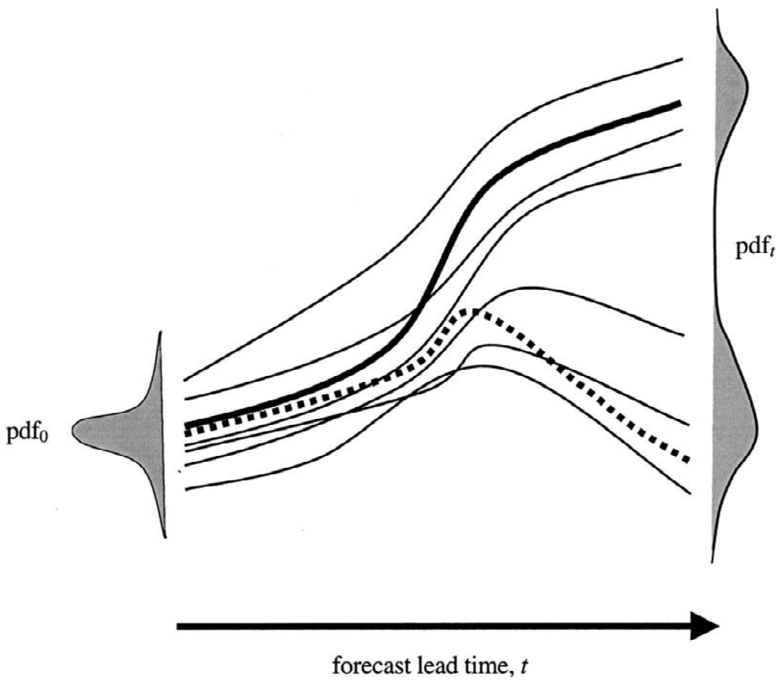

Fig. 1. Schematic of ensemble prediction. The initial probability density function, pdf$_0$, represents the initial uncertainties. From the best estimate of the initial state, a single point forecast (bold solid curve) is produced. This point forecast fails to predict correctly the future state (dash 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.

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 distribution of historical 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. The frequency distribution of the different realisations, which are known as ensemble members, provides an estimate of the pdf. The initial conditions are not sampled as in a statistical simulation because this is not practical for the complex, high-dimensional weather prediction model. Instead, they are designed to sample directions of maximum possible growth (Molteni et al., 1996; Palmer et al., 1993; Buizza et al., 1998).

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 prediction with medium-range ensemble prediction (Toth & Kalnay, 1993; Tracton & Kalnay, 1993; Palmer et al., 1993). The number of ensemble members is limited by the necessity to produce weather forecasts in a reasonable amount of time with the available computer power. In December 1996, after different system configurations had been considered, a 51-member system was installed at ECMWF (Buizza et al., 1998). The 51 members consist of one forecast started from the unperturbed, best estimate of the atmosphere initial state plus 50 others generated by varying the initial conditions. Stochastic physics was introduced into the system in October 1998 (Buizza, Miller, & Palmer, 1999). This aims to simulate model uncertainties due to random model error in the parameterised physical processes.

At the time of this study, ensemble forecasts were produced every day for lead times from 12 h ahead to 10 days ahead. The ensemble forecasts were archived every 12 h, and are thus available for midday and midnight. The archived weather variables include both upper level weather variables (typically wind, temperature, humidity and vertical velocity at different heights) and surface variables (e.g. temperature, wind, precipitation, cloud cover). ECMWF disseminates ensemble forecasts to the National Meteorological Centers of its European member states, as part of an operational suite of weather products. In our work we used ensemble predictions generated by ECMWF from 1 November 1998 until 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. We use ensemble predictions for the following three variables: temperature, wind speed and cloud cover.

3. Electricity demand forecasting

3.1. Modelling electricity demand in England and Wales

There is no consensus as to the best approach to electricity demand forecasting. The range of different approaches includes time-varying splines (Harvey & Koopman, 1993), multiple regression models (Ramanathan, Engle, Granger, Vahid-Araghi, & Brace, 1997), judgemental forecasts and artificial neural networks (see Hippert, Pedreira, & Souza, 2001). In this paper, we implement the forecasting process used at NG. 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. The approach taken by NG 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 & Majithia, 2000). Harvey and Koopman (1993) describe a similar approach, which involves fitting a time-varying spline between a number of cardinal points. At NG, 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 Ramanathan et al. (1997), who produced hourly forecasts by using a separate regression model for each hour of the day.

3.2. Modelling midday electricity demand

In this paper, we focus on predicting demand (load) in England and Wales 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 NG, and so there is no need to perform the NG 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 NG and Ramanathan et al. (1997) and produce a model for midday based on demand for previous middays and weather variables.

Fig. 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 5000 MW between typical winter and typical summer demand. Another noticeable seasonal feature occurs within each week where there is a consistent difference of about 6000 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, NG 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. An alternative to this would be to treat the special days as missing observations.

Short to medium-term forecasting models must accommodate the variation in demand due to the seasonal patterns shown in Fig. 2 and due to weather. At NG, 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) for day t is an exponentially smoothed form of TO_t, which is

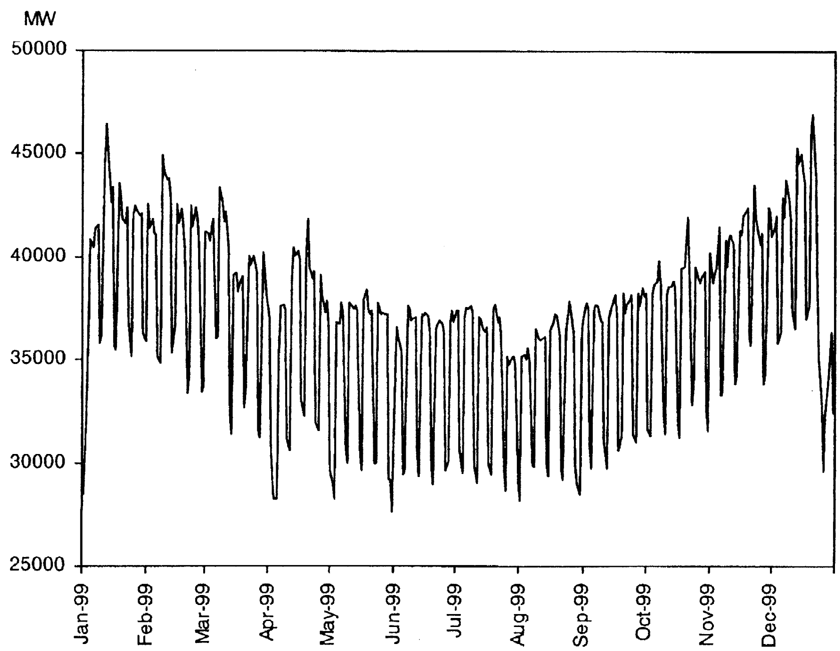

Fig. 2. Demand for electricity at midday in England and Wales in 1999.

the mean of the spot temperature recorded for each of the 4 previous hours

$$TE_t = \frac{1}{2} TO_t + \frac{1}{2} TE_{t-1}. \quad (1)$$

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 non-linear function of wind speed, W_t, and average temperature, TO_t. It aims to describe the draught-induced load variation

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

Effective illumination is a complex function of visibility, number and type of cloud and amount and type of precipitation.

Since NG 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 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 from NG's formulae in expressions (1) and (2).

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 NG. The first stage aims to identify the weather-related component by estimating a regression model similar to the following:

$$\begin{aligned} \text{demand}_t = a_0 &+ a_1 TE_t + a_2 TE_t^2 + a_3 CP_t + a_4 CC_t \\ &+ a_5 t + a_6 t^2 + a_7 t^3 + a_8 t^4 + a_9 FRI_t \\ &+ a_{10} SAT_t + a_{11} SUN_t + a_{12} W1_t \\ &+ a_{13} W2_t + a_{14} W3_t + \varepsilon_t \end{aligned} \quad (3)$$

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 Fig. 2. We followed NG in using data from the previous 2 years to estimate the model, and so a quartic time polynomial was appropriate.

The second stage of the NG 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 (3):

$$\begin{aligned} \text{weather-related\_demand} = \hat{a}_1 TE_t &+ \hat{a}_2 TE_t^2 + \hat{a}_3 CP_t \\ &+ \hat{a}_4 CC_t \end{aligned} \quad (4)$$

Forecasts for the base load are produced judgementally by NG. Since we do not have the expertise to produce judgemental forecasts, we used the simple alternative of a univariate ARMA-regression model. Using the usual diagnostic tests, we derived the following model:

$$\begin{aligned} \text{base\_demand}_t = b_0 &+ b_1 FRI_t + b_2 SAT_t + b_3 SUN_t \\ &+ b_4 W2_t + b_5 W3_t + \varepsilon_t \\ \varepsilon_t = \phi_1 \varepsilon_{t-1} &+ \phi_2 \varepsilon_{t-2} + \theta_1 u_{t-1} + u_t \end{aligned} \quad (5)$$

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

4. Using weather ensembles for demand point forecasting

4.1. Creating 51 scenarios for weather-related electricity demand

A standard 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. This is an important issue when forecasting from non-linear models (Lin & Granger, 1994). Let us reconsider the forecast of the weather-related demand, which was given in expression (4). In view of the definition of cooling power of the wind, given in expression (2), and the presence of the TE_t^2 term in (4), 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 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 (4) 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 variance and shape 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 point forecast. The collection of 51 ensemble members must, therefore, contain information not captured by the single 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, this 18-month period gave 500 days for evaluation. We produced forecasts for each day in our evaluation period for lead times of 1 to 10 days ahead. We compared four different sets of forecasts using the mean absolute percentage error (MAPE) summary measure, which is used extensively in the electricity demand forecasting literature.

Method 1: *traditional weather point forecasts*. After estimating the models in expressions (3) and (5) for the two-stage approach described in Section 3, we produced forecasts by the usual procedure of substituting traditional single point weather forecasts in expression (4) for the weather-related demand.

Method 2: *mean of scenarios*. Using the same models from the two-stage approach, we produced forecasts using the mean of the 51 scenarios for weather-related demand. This approach is based on the weather ensemble predictions since the 51 scenarios are constructed from the 51 ensemble members.

Method 3: *actual weather used as forecasts*. 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 two-stage approach with actual observed weather substituted for the weather variables in the weather-related demand expression in (4). Clearly this level of forecast accuracy is unattainable, as perfect weather forecasts are not achievable.

Method 4: *pure ARMA*. 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:

$$\begin{aligned}\text{demand}_t &= c_0 + c_1 FRI_t + c_2 SAT_t + c_3 SUN_t \\ &\quad + c_4 W2_t + c_5 W3_t + \varepsilon_t \\ \varepsilon_t &= \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \psi_1 u_{t-1} + u_t\end{aligned}$$

where the c_i, φ_i and ψ_1 are constant parameters.

Fig. 3 shows MAPE results for the four different methods. It is widely accepted that, for 1-day-ahead forecasting, a weather-based method is preferable to a method that does not use weather information.

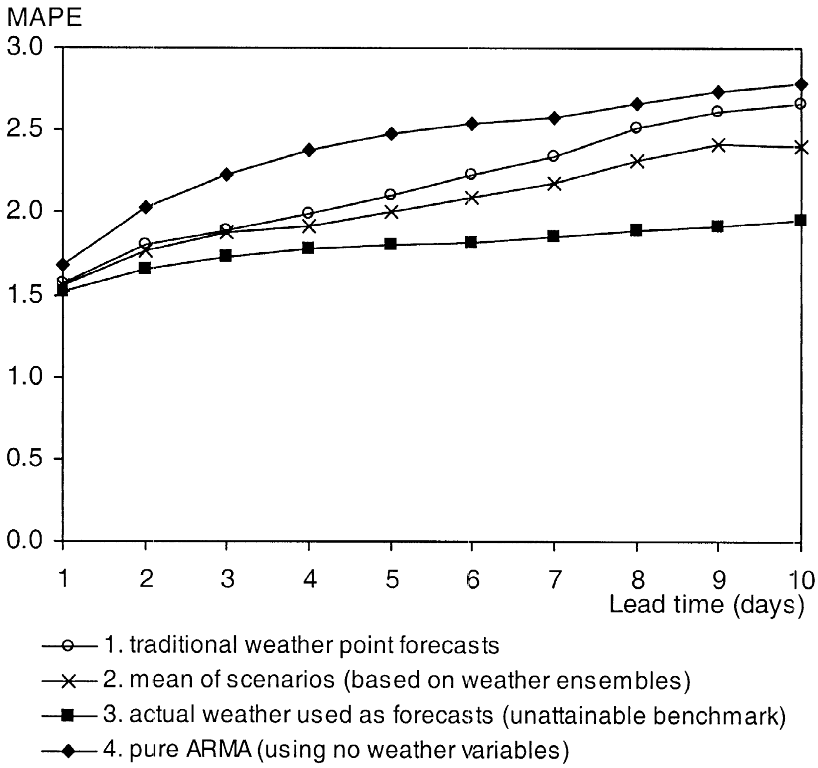

Fig. 3. MAPE for electricity demand point forecasts for post-sample period, 1 November 1998 to 30 April 2000.

Indeed, all of the methods entered in a recent 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 1-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, which is an unattainable benchmark. For lead times of 4 days ahead or more, the accuracy of the new ensemble-based 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.

5. Using weather ensembles to estimate the demand forecast error variance

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 variance of the probability distribution of demand forecast error. This is not a trivial task, as the forecast error variance is likely to vary over time due to weather and seasonal effects. 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 variance of the forecast errors from this method. The approach that we take is to model the variance in a series of historical post-sample forecast errors. Our modelling of the uncertainty focuses on the error series and does not affect the point forecasts. A similar approach is taken by Engle, Granger, Ramanathan, and Vahid-Araghi (1993) who model the magnitude of forecast errors. We consider lead times of 1 to 10 days ahead, unlike Engle et al. who focus only on 1-day-ahead forecasting. We use the first 9 months (1 November 1998 to 31 July 1999) of post-sample errors from our earlier analysis of point forecasting to estimate model parameters, and the remaining 9 months (1 August to 30 April 2000) of post-sample errors to evaluate the resulting variance forecasts.

5.1. Methods for estimating demand forecast error variance

In this section, we present seven methods for estimating the error variance. Methods 2 and 3 are used for forecasting volatility in financial data, and methods 4 to 7 are designed to incorporate weather ensemble information in the estimate of the error variance.

Method 1: *naive*. For each lead time, k, we calculated the variance of the k-day-ahead errors in the estimation period of 9 months.

Method 2: *ewma*. An exponentially weighted moving average of recent squared errors allows the estimate to adapt over time. We implemented this method and optimised the smoothing parameter separately for each lead time.

Method 3: *garch*. An alternative to the ad hoc methods described so far is the GARCH statistical modelling approach (see Engle, 1982; Bollerslev, 1986). In addition to lagged squared error terms and lagged conditional variance terms, exogenous explanatory variables can be included in GARCH models. We experimented with simple univariate explanatory variables. However, the only one that was significant for any of the models was the dummy variable for Saturdays, SAT_t. The 1-day-ahead GARCH(1, 1) variance forecast is given by

$$\hat{\sigma}_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \hat{\sigma}_{t-1}^2 + \gamma_1 SAT_t.$$

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 Granger & Newbold, 1986, p. 130). This was evident in our forecast errors. We controlled for this by fitting the GARCH model to the residuals of an MA($k-1$) model fitted to the 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.

Method 4: *scenario variance*. The level of uncertainty in the demand forecasts depends to an extent on the uncertainty in the weather forecasts. This motivates the use of a measure of weather forecast uncertainty in the modelling of demand forecast uncertainty. The variance of the 51 demand scenarios, discussed in Section 4.1, conveys the uncertainty in the weather component of demand. For each day in our post-sample period, we calculated the variance, $\sigma_{ENS,t}^2$, of the 51 scenarios for each of the 10 lead times, and used this as an estimate of the demand forecast error variance.

Method 5: *recalibrated scenario variance*. The variance of the 51 scenarios is likely to underestimate the demand forecast error variance because it does not accommodate the uncertainty due to the model error and the parameter estimation error associated with expressions (3) and (4). In view of this, for each lead time, we performed a linear bias correction by regressing the squared forecast error on $\sigma_{ENS,t}^2$. The 'recalibrated' estimator is of the form:

$$\hat{\sigma}_t^2 = \hat{a} + \hat{b}\sigma_{ENS,t}^2.$$

Method 6: *mixed garch*. 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 $\sigma_{ENS,t}^2$ as an additional potential explanatory variable. This new variable was significant only in the models for lead times 2, 5, 8, 9 and 10. For the other lead times, the *mixed garch* model was identical to the *garch* model of Method 3.

Method 7: *combination*. Combining is an alternative to the *mixed garch* model for synthesising information from the ensemble predictions and the past forecast error variance. We calculated the simple average of the *recalibrated scenario variance* estimator and the *garch* estimator. We chose the *garch* estimator simply because it is the most sophisticated of the univariate methods.

5.2. Comparison of variance estimators

Table 1 reports the coefficient of determination, R^2, from the regression of the squared post-sample forecast errors on the variance forecasts for the 9-month post-sample evaluation period. This measure is widely used in volatility forecast evaluation. The regression corrects for any bias and the R^2 measures the degree to which the estimator varies with the changing variance of the errors. It is, therefore, a measure of the informational content of the estimator. Typically, the R^2 values are low, with values less than 10% being the norm (Andersen & Bollerslev, 1998). The entries in bold in each column of Table 1 indicate the best performing method for each lead time. 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 results for the *scenario variance* method and the *recalibrated scenario variance* method are identical because the R^2 measures covariation after performing a bias correction on the estimator. Although the *garch* method performed well at the early lead times and at the 10-day horizon, overall, the best results were recorded with the ensemble-based methods and the *combination*.

Table 2 shows the root mean squared error (RMSE) post-sample evaluation results

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_i \left(e_i^2 - \hat{\sigma}_i^2\right)^2}$$

where e_i is the load forecast error and n is the number of observations in the 9-month post-sample evaluation period. Unlike the R^2, the RMSE does not correct for bias, and so the results of Table 2 are a more straightforward reflection of forecasting performance. In Section 5.1, we suggested that bias would be a major issue for the *scenario variance* method. Comparing methods 4 and 5 in Table 2, we can see that the RMSE results for this estimator notably improve with the recalibration. The bold entries in the table indicate that the *combination* is generally the best method up to 5 days ahead and that the *recalibrated scenario variance* method is the best beyond 5 days ahead.

In summary, Tables 1 and 2 show that there is benefit in using weather ensemble information in constructing demand forecast error variance estimates. In view of its strong performance using both evaluation measures and its relative simplicity, we would recommend the *recalibrated scenario variance* method.

6. Using weather ensembles to estimate demand prediction intervals

Prediction intervals are widely used to convey the uncertainty in a forecast. In this section, we consider

Table 1
R^2 percentage measure for forecast error variance estimation methods for post-sample period, 1 August 1999 to 30 April 2000

	Lead time (days)									
	1	2	3	4	5	6	7	8	9	10
Univariate										
1. naïve	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2. ewma	2.1	1.4	2.1	0.0	0.9	0.0	0.0	5.3	0.0	1.8
3. garch	2.6	2.1	**3.2**	0.1	0.8	0.5	3.7	2.4	0.6	**7.2**
Ensemble based										
4. scenario variance	1.2	3.5	0.6	**1.9**	3.0	**5.4**	**6.0**	8.5	6.7	4.1
5. recalibrated scenario variance	1.2	3.5	0.6	**1.9**	3.0	**5.4**	**6.0**	8.5	6.7	4.1
6. mixed garch	2.6	**7.1**	**3.2**	0.1	**4.1**	0.5	3.7	**9.4**	**8.2**	6.1
Combination										
7. average of 3 and 5	**3.0**	4.2	0.9	**1.9**	2.6	4.7	**6.0**	8.8	6.6	6.8

Table 2
RMSE/1000 for forecast error variance estimation methods for post-sample period, 1 August 1999 to 30 April 2000

	Lead time (days)									
	1	2	3	4	5	6	7	8	9	10
Univariate										
1. naïve	1032	958	1090	1198	1340	1315	1401	1624	1800	1880
2. ewma	1025	950	1079	1197	1350	1311	1412	1614	1801	1888
3. garch	1031	954	1088	1217	1358	1334	1397	1690	1891	1857
Ensemble based										
4. scenario variance	1192	1140	1276	1376	1507	1477	1572	1820	2003	2084
5. recalibrated scenario variance	1028	947	1103	1216	1323	**1281**	**1358**	**1556**	**1774**	1874
6. mixed garch	1031	**918**	1088	1217	1359	1334	1397	1682	1902	1994
Combination										
7. average of 3 and 5	**1017**	933	**1078**	**1185**	**1322**	1286	1359	1588	1780	**1838**

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 $P(y < Q(\theta)) = \theta$. As in Section 5, we use 9 months of post-sample errors from our earlier analysis of demand point forecasting to estimate method parameters, and the remaining 9 months of post-sample errors to evaluate the estimators.

6.1. Methods for estimating demand forecast error quantiles

The variance estimators, investigated in Section 5, can be used as the basis of quantile estimators. We used either a Gaussian distribution or the empirical distribution of the corresponding standardised forecast errors, $e_t/\hat{\sigma}_t$ (see Granger, White, & Kamstra, 1989).

Method 1: *naïve* variance estimator with Gaussian distribution.

Method 2: *ewma* variance estimator with Gaussian distribution.

Method 3: *garch* variance estimator with Gaussian distribution.

Method 4: *recalibrated scenario variance* variance estimator with Gaussian distribution.

Method 5: *mixed garch* variance estimator with Gaussian distribution.

Method 6: *naïve* variance estimator with empirical distribution.

Method 7: *ewma* variance estimator with empirical distribution.

Method 8: *garch* variance estimator with empirical distribution.

Method 9: *recalibrated scenario variance* variance estimator with empirical distribution.

Method 10: *mixed garch* variance estimator with empirical distribution.

Method 11: *scenario quantile*. We used the quantiles, $Q_{ENS,t}(\theta)$, of the distribution of scenarios as estimates of the quantiles of the forecast error distribution.

Method 12: *recalibrated scenario quantile*. The width of the predictive distribution is likely to be greater than the width of the distribution of scenarios. We used quantile regression to recalibrate the *scenario quantile* estimator with the forecast errors as dependent variable and $Q_{ENS,t}(\theta)$ as regressor (see Granger, 1989). The form of the resultant recalibrated estimator is:

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

Method 13: *combination*. We calculated the average of the *garch*-based estimator with empirical distribution and the *recalibrated scenario quantile*.

6.2. Comparison of quantile estimators

Table 3 compares estimation of the 5% quantiles at the 10 different lead times for the post-sample period of 9 months. The table shows the percentage of post-sample forecast errors falling below the quantile estimators. For an unbiased estimator of the 5% quantile, this will be 5% (see Taylor, 1999). The entries in bold in each column of Table 3 indicate the best performing method for each lead time. 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 results show that the *ewma* variance estimator with Gaussian distribution and the *combination* method perform well, and that the *scenario quantile* method is vastly improved with the quantile regression recalibration.

To summarise the overall relative performance of 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 4 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 indicate that an empirical distribution is preferable to a Gaussian assumption. The *combination* and the *recalibrated scenario quantile* perform consistently well across the 10 lead times.

The percentage of errors falling below a quantile estimator evaluates only bias; we should also consider the variability of the estimation error. For example, the first column of results in Table 3 shows that 4.4% of the 1-day-ahead errors fell below the *naïve* variance estimator with Gaussian distribution. Since the ideal is 5%, the estimator is a little low on average; it possesses a degree of bias. Although the level of bias in this estimator is the second best of the 13 estimators, other estimators should vary in accordance with the varying variance of the distribution better than the *naïve* estimator, which by construction does not vary at all. It would be useful

Table 3
Percentage of errors falling below estimates of 5% forecast error quantile for post-sample period, 1 August 1999 to 30 April 2000

	Lead time (days)									
	1	2	3	4	5	6	7	8	9	10
Variance estimators with Gaussian										
1. naïve	4.4	4.4	3.6	4.0	4.0	**5.2**	4.4	3.2	3.6	3.6
2. ewma	6.0	6.0	**5.2**	**5.2**	6.0	6.0	6.0	**6.0**	3.6	2.8
3. garch	4.0	**5.6**	7.1	8.3*	7.1	7.1	8.3*	9.9*	9.5*	6.0
4. recalibrated scenario variance	4.0	3.6	3.2	4.0	4.4	4.4	3.2	3.6	4.4	4.4
5. mixed garch	4.0	4.0	7.1	8.3*	8.7*	7.1	8.3*	12.3*	15.1*	14.3*
Variance estimators with empirical										
6. naïve	4.4	4.4	3.2	4.4	4.0	3.6	4.4	3.2	3.2	2.0
7. ewma	6.3	7.5	4.4	4.8	6.7	3.6	4.0	3.2	3.2	2.0*
8. garch	6.0	**5.6**	3.2	4.4	6.0	4.8	6.7	3.2	3.2	**4.8**
9. recalibrated scenario variance	3.6	4.4	2.4	4.0	4.4	2.0*	3.2	2.4	**4.8**	4.4
10. mixed garch	6.0	2.8	3.2	4.4	4.0	4.8	6.7	4.0	6.3	6.0
Demand scenario quantile										
11. scenario quantile	43.7*	46.0*	27.0*	25.0*	20.6*	20.6*	16.7*	20.6*	17.9*	17.9*
12. recalibrated scenario quantile	4.4	3.6	3.2	4.0	4.4	2.8	3.6	2.8	3.6	4.4
Combination										
13. average of 8 and 12	**4.8**	4.4	2.8	4.0	**4.8**	3.2	**4.8**	3.2	2.4	4.4

*Indicates significant at 5% level.

Table 4
Chi-squared statistics summarising overall estimator bias for 5%, 25%, 75% and 95% forecast error quantiles for the post-sample period, 1 August 1999 to 30 April 2000

	Lead time (days)									
	1	2	3	4	5	6	7	8	9	10
Variance estimators with Gaussian										
1. naïve	11.7*	9.3	9.0	5.1	4.2	3.3	**0.6**	4.0	4.5	**5.0**
2. ewma	8.9	15.9*	20.3*	6.6	10.6*	**2.2**	3.7	8.2	4.7	4.8
3. garch	12.8*	17.7*	20.3*	14.6*	12.2*	10.7*	14.9*	30.6*	41.8*	25.7*
4. recalibrated scenario variance	12.2*	10.8*	14.7*	4.8	3.0	2.8	3.1	4.0	13.9*	9.7*
5. mixed garch	12.8*	19.1*	20.3*	14.6*	20.2*	10.7*	14.9*	91.6*	153.0*	172.1*
Variance estimators with empirical										
6. naïve	4.9	9.4	4.6	1.1	1.1	1.8	3.7	5.9	11.0*	18.1*
7. ewma	9.3	10.3*	9.7*	**2.0**	3.1	1.8	9.4	6.4	11.0*	19.0*
8. garch	3.8	**1.5**	**3.5**	1.1	1.0	4.4	9.3	7.6	6.3	28.6*
9. recalibrated scenario variance	6.5	7.0	5.0	1.7	**0.3**	5.0	3.9	8.2	19.0*	23.1*
10. mixed garch	**3.8**	10.6*	**3.5**	1.1	0.8	4.4	9.3	**11.6***	31.1*	41.5*
Demand scenario quantile										
11. scenario quantile	1122.2*	1165.8*	1334.4*	639.2*	1536.0*	676.6*	1240.3*	831.7*	1167.4*	1465.5*
12. recalibrated scenario quantile	4.5	7.5	3.8	2.8	1.7	5.3	2.5	3.3	**2.8**	34.7*
Combination										
13. average of 8 and 12	**2.9**	4.9	3.8	1.7	**0.3**	5.6	2.7	2.8	6.0	19.3*

*Indicates significant at 5% level.

if we could evaluate this variability characteristic. The R^2 measure used for evaluating the variance estimators in Section 5 corrects for bias, so that the R^2 then reflects variation about the bias. Similarly, a quantile regression R^2 measure can be used to evaluate quantile estimator prediction variance (Taylor, 1999). The package STATA (Stata, 1993) provides a pseudo-R^2, analogous to the R^2 in LS regression. Table 5 shows this pseudo-R^2 for estimation of the 5% quantiles; high values of the pseudo-R^2 are preferable.

The first column of results in Table 5 shows that the pseudo-R^2 for the *naïve* estimator is zero, but for many of the other estimators it is considerably more. These results reflect the fact that these estimators vary more with the unobservable quantile. The pseudo-R^2 reflects covariation between estimator and unobservable quantile. Consequently, the quantile

Table 5
Pseudo R^2 percentage measure for estimators of 5% forecast error quantile for post-sample period, 1 August 1999 to 30 April 2000

	Lead time (days)									
	1	2	3	4	5	6	7	8	9	10
Variance estimators										
1. & 6. naïve	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2. & 7. ewma	10.9	9.3	**9.9**	**6.7**	5.3	0.0	0.2	2.6	0.0	1.3
3. & 8. garch	11.0	6.3	3.2	4.1	4.3	0.4	5.4	0.3	0.2	**5.2**
4. & 9. recalibrated scenario variance	5.1	5.6	3.2	2.7	4.8	**10.3**	7.5	**8.6**	**8.1**	3.6
5. & 10. 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										
11. scenario variance	0.2	3.3	0.5	0.2	0.8	0.5	1.1	0.7	0.6	0.7
12. recalibrated scenario variance	0.2	3.3	0.5	0.2	0.8	0.5	1.1	0.7	0.6	0.7
Combination										
13. average of 8 and 12	**11.2**	7.6	1.0	0.6	2.0	1.1	3.9	0.5	0.4	3.2

estimators based on the same variance estimator, which differ only by a linear transformation, have the same pseudo-R^2. Table 5 suggests that the estimators based on the *recalibrated scenario variance* estimator and those based on the *mixed garch* variance estimator 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. Based on the chi-squared results in Table 4 and the pseudo-R^2 results for the four different quantiles, we would tentatively conclude that, overall, the methods that perform the best for quantile estimation are the *combination* and *the recalibrated scenario variance* with empirical distribution.

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 demand model. Since demand is a non-linear function of weather variables, this traditional procedure amounts to approximating the expectation of a non-linear 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 demand forecast uncertainty. In view of this, we recalibrated measures of variance and quantiles taken from the scenario distribution. The resulting variance estimator compared favourably with estimators produced using univariate volatility forecasting methods. Using the same variance 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

We are very grateful to Shanti Majithia and Chris Rogers of the National Grid 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. We are also grateful to three anonymous referees for their helpful comments.

References

Adams, G., Allen, P. G., & Morzuch, B. J. (1991). Probability distributions of short-term electricity peak load forecasts. *International Journal of Forecasting*, *7*, 283–297.

Andersen, T. G., & Bollerslev, T. (1998). Answering the critics: yes ARCH models do provide good volatility forecasts. *International Economic Review*, *39*, 885–906.

Baker, A. B. (1985). Load forecasting for scheduling generation on a large interconnected system. In Bunn, D. W., & Farmer, E. D. (Eds.), *Comparative Models for Electrical Load Forecasting*. Chichester: Wiley, pp. 57–67.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

Buizza, R., Miller, M., & Palmer, T. N. (1999). Stochastic simulation of model uncertainties. *Quarterly Journal of the Royal Meteorological Society*, *125*, 2887–2908.

Buizza, R., Petroliagis, T., Palmer, T. N., 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.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1008.

Engle, R., Granger, C. W. J., Ramanathan, R., & Vahid-Araghi, F. (1993). *Probabilistic Methods in Forecasting Hourly Loads and Peaks*. Palo Alto, CA: Electric Power Research Institute, EPRI TE-101902.

Granger, C. W. J. (1989). Combining forecasts — twenty years later. *Journal of Forecasting*, *8*, 167–173.

Granger, C. W. J. (1996). Can we improve the perceived quality of economic forecasts? *Journal of Applied Econometrics*, *11*, 455–473.

Granger, C. W. J., & Newbold, P. (1986). 2nd edn, Forecasting Economic Time Series, Academic Press, London.

Granger, C. W. J., White, H., & Kamstra, M. (1989). Interval forecasting: an analysis based upon ARCH-quantile estimators. *Journal of Econometrics*, *40*, 87–96.

Harvey, A. C., & Koopman, S. J. (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association*, *88*, 1228–1236.

Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on Power Systems*, *16*, 44–55.

Leith, C. E. (1974). Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review*, *102*, 409–418.

Lin, J., & Granger, C. W. J. (1994). Forecasting from non-linear models in practice. *Journal of Forecasting*, *13*, 1–9.

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

Palmer, T. N., Molteni, F., Mureau, R., Buizza, R., Chapelet, P., & Tribbia, J. (1993). Ensemble prediction. In *Proceedings of the ECMWF Seminar on Validation of Models over Europe*, Vol. I. Shinfield Park, Reading, RG2 9AX, UK: ECMWF.

Ramanathan, R., Engle, R., Granger, C. W. J., 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.* College Station, TX: Stata Corporation.

Taylor, J. W. (1999). Evaluating volatility and interval forecasts. *Journal of Forecasting*, *18*, 111–128.

Taylor, J. W., & 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, M. S., & Kalnay, E. (1993). Operational ensemble prediction at the National Meteorological Center: practical aspects. *Weather and Forecasting*, *8*, 379–398.

Biographies: James W. TAYLOR is a lecturer at Saïd Business School, University of Oxford. His research interests include load forecasting, prediction intervals, combining forecasts, exponential smoothing and volatility forecasting.

Roberto BUIZZA is a principal scientist at the European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK. After 4 years at the Centre for Thermonuclear Research of the Italian Electricity Board (CRTN/ENEL), he joined ECMWF in 1991. His research interests include instability theory, chaos and predictability of the atmospheric flow, diagnostics of deterministic and probabilistic numerical weather forecasting systems, and economic value of weather forecasts.