Neural Network Load Forecasting with Weather Ensemble Predictions

James W. Taylor and Roberto Buizza
IEEE Trans. on Power Systems, 2002, Vol. 17, pp. 626-632.

***Abstract*--In recent years, a large literature has evolved on the use of artificial neural networks (NNs) for electric load forecasting. NNs are particularly appealing because of their ability to model an unspecified non-linear relationship between load and weather variables. Weather forecasts are a key input when the NN is used for forecasting. This study investigates the use of weather ensemble predictions in the application of NNs to load forecasting for lead times from 1 to 10 days ahead. A weather ensemble prediction consists of multiple scenarios for a weather variable. We use these scenarios to produce multiple scenarios for load. The results show that the average of the load scenarios is a more accurate load forecast than that produced using traditional weather forecasts. We use the load scenarios to estimate the uncertainty in the NN load forecast. This compares favourably with estimates based solely on historical load forecast errors.**

***Index Terms*-- Load forecasting; neural networks; weather ensemble predictions.**

I. Introduction

ACCURATE load 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. Over the last decade, a great deal of attention has been devoted to the use of artificial neural networks (NNs) to model load [1]. Weather variables are an important input to these models for short- to medium-term forecasting. A load forecast is produced by substituting a forecast for each weather variable in the NN model. Traditionally, single weather point forecasts have been used. A weather ensemble prediction is a new type of weather forecast. It consists of multiple scenarios for the future value of a weather variable. The scenarios are known as ensemble members, and in this paper each ensemble prediction consists of 51 members. The ensemble, therefore, conveys the degree of uncertainty in the weather variable. In [2], we found that there was benefit in using ensemble predictions in linear regression load forecasting models. This paper considers the use of these new weather forecasts in the non-linear modelling environment of NNs.

We use the 51 weather ensemble members to produce 51 scenarios for load from a NN for lead times from 1 to 10 days ahead. Meteorologists sometimes find that the mean of the ensemble members for a weather variable is a more accurate forecast of the variable than a traditional single point forecast [3,4]. In view of this, we consider the use of the average of the 51 load scenarios as a point forecast for load. 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. Since NN load models are non-linear functions of weather variables, the traditional procedure of inserting single weather point forecasts 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 load scenarios is appealing because it is equivalent to taking the expectation of an estimate of the load probability density function.

We use the distribution of the load scenarios as an input to estimating the uncertainty in the load forecasts. It is important to assess the uncertainty in order to manage the system load efficiently. A measure of risk is also beneficial when trading electricity. The standard practice in NN load forecasting research is to ignore the impact of weather forecast accuracy; actual weather is predominantly used to evaluate NN models [1]. However, weather forecast error can seriously impact load forecast accuracy [5]. In [6], it is demonstrated that weather uncertainty information can be used to produce improved load predictions and prediction intervals. This is also shown by our study, which uses weather ensembles to provide the weather uncertainty information.

In this paper, our analysis is based on daily load data for England and Wales. The variables used to model load are those used at the National Grid (NG), which is responsible for the transmission of electricity in England and Wales. Weather ensemble predictions are described in Section II. Section III presents the NN and input variables used in this study. Section IV considers how weather ensemble predictions can be used to improve the accuracy of the NN load forecasts. Sections V and VI investigate the potential for using weather ensemble predictions to assess the uncertainty in the load forecasts. The estimation of load forecast error variance is considered in Section V, and load prediction intervals are the focus of Section VI. Section VII provides a summary and conclusions.

II. Weather Ensemble 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

J.W. Taylor is with Saïd Business School, University of Oxford, Park End Street, Oxford, OX1 1HP, UK (e-mail: james.taylor@sbs.ox.ac.uk).

R. Buizza is with European Centre for Medium-range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK (e-mail: r.buizza@ecmwf.int).

numerical model. These two sources of uncertainty limit the accuracy of single point forecasts, generated by running the model once with best estimates for the initial conditions.

The weather prediction problem can be described 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 gauge the uncertainty in single point predictions. Ensemble prediction aims to derive a more sophisticated estimate of the pdf than that provided by the distribution of past forecast 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 model. Instead, they are designed to sample directions of maximum possible growth [4, 7, 8].

The benefit of using ensemble predictions is illustrated in Fig. 1. pdf_0, represents the initial uncertainties. From the best estimate of the initial state, a single point forecast is produced (bold solid curve). This point forecast fails to predict correctly the future state (dashed curve). The ensemble forecasts (thin solid curves), starting from perturbed initial conditions, can be used to estimate the probability of future states. In this example, the estimated pdf, 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.

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 [7, 9, 10]. 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 [8]. 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 [11]. This aims to simulate model uncertainties due to random model error.

At the time of this study, ensemble forecasts were produced every day for lead times from 12 hours ahead to 10 days ahead. The ensemble forecasts were archived every 12 hours, and are thus available for midday and midnight. The archived weather variables include both upper level variables (typically wind speed, temperature, humidity and vertical velocity at different heights) and surface variables (e.g. temperature, wind speed, precipitation, cloud cover). In our work, we used ensemble predictions for temperature, wind speed and cloud cover generated by ECMWF from 1 November 1998 to 30 June 2000. We did not use earlier predictions because the introduction of stochastic physics in October 1998 substantially improved the ensemble predictions.

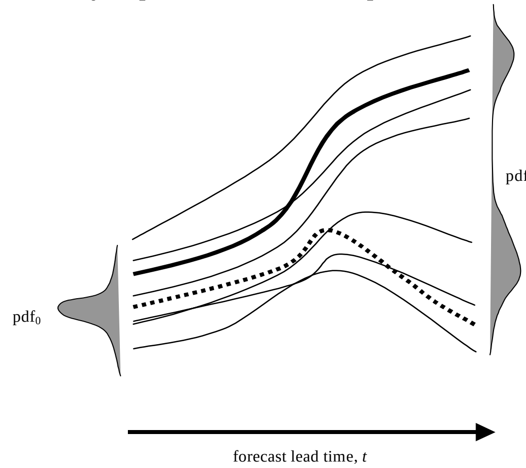

Fig. 1. Schematic of ensemble prediction. Bold solid curve is the single point forecast. Dashed curve is the future state. Thin solid curves are the ensemble of perturbed forecasts.

III. An NN Load Model For England and Wales

A. Load Forecasting

A wide variety of methods have been used for load forecasting. The range of different approaches includes time-varying splines [12], linear regression models [13], profiling heuristics [14] and judgemental forecasts. However, the most significant development in recent years has been the use of NNs, which allow the estimation of possibly non-linear models without the need to specify a precise functional form. Load forecasting is a suitable application for NNs because load is usually an unknown non-linear function of weather variables. Furthermore, there is often a large amount of data available in load modelling, which is a necessity for the effective use of NNs. A useful critical review of the literature on the use of NNs for load forecasting is provided in [1].

The winners of a recent load forecasting competition produced hourly forecasts using separate linear regression models for each hour of the day [13]. In this paper, we follow this general methodology but, instead of linear regression, we use a NN. For simplicity, we focus on predicting load at midday. This is convenient because ensemble predictions are currently available for midday, although in the future they could be produced for any required period of the day. Midday is a particularly important period in many summer months in England and Wales because it is often when peak load occurs. However, it is important to note that our work is not specific to peak load forecasting, and that although we do focus on midday forecasting, the methods that we consider can be used for any period of the day.

Fig. 2 shows a plot of load in England and Wales at midday for each day in 1999. One clear feature 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. In practice, at NG, judgemental methods are often used to forecast load on these days. As in many other studies of electric load (e.g. [6] and [15]), we elected to smooth out these special days, as their inclusion is likely to be unhelpful in our analysis of the relationship between load and weather.

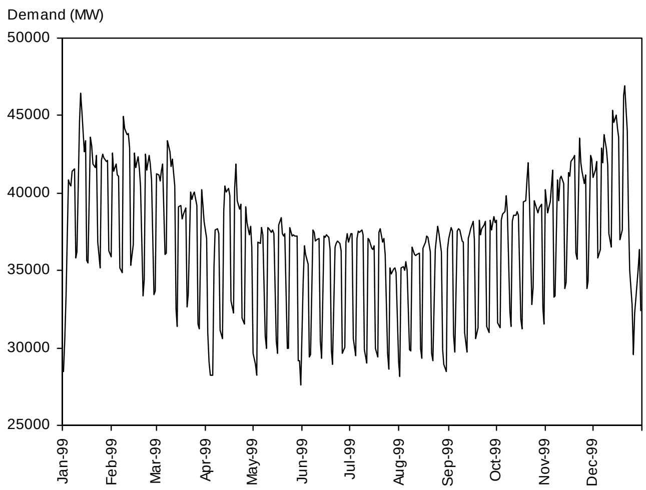

Fig. 2. Load at Midday in England and Wales in 1999.

B. *The Neural Network Design*

In this paper, we use a single hidden layer feedforward network, which is the most widely-used neural network for forecasting [16]. It consists of a set of k inputs, which are connected to each of m units in a single hidden layer, which, in turn, are connected to an output. In regression terminology, the inputs are explanatory variables, x_{it}, and the output is the dependent variable, y_t, which in this study is midday load in England and Wales. The resultant model can be written as

$$f(\boldsymbol{x}_t, \boldsymbol{v}, \boldsymbol{w}) \;=\; g_2\left(\sum_{j=0}^{m} v_j \; g_1\left(\sum_{i=0}^{k} w_{ji} \; x_{it}\right)\right) \tag{1}$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are activation functions, which we chose as sigmoidal and linear respectively, and w_{ji} and v_j are the weights (parameters). We estimated the weights using the following minimisation

$$\min_{\boldsymbol{v},\boldsymbol{w}}\left(\frac{1}{n}\sum_{t=1}^{n}\left(y_t - f(\boldsymbol{x}_t, \boldsymbol{v}, \boldsymbol{w})\right)^2 + \lambda_1 \sum_{j=0}^{m}\sum_{i=0}^{k} w_{ji}^2 + \lambda_2 \sum_{j=0}^{m} v_i^2\right) \tag{2}$$

where n is the number of observations, and λ_1 and λ_2 are regularisation parameters which penalise the complexity of the network and thus avoid overfitting [17, §9.2]. We established suitable values for λ_1 and λ_2 and for the number, m, of units in the hidden layer using a hold out method with a third of the data used for testing [17, §9.8]. This resulted in the same number of hidden units, m, as the number, k, of inputs, which is a rule-of-thumb suggested in [18].

Although in [1] several studies are reviewed, which like ours implement a NN with just one output (e.g. [19, 20]), the use of 24 outputs is also common, where each output corresponds to an hour of the day (e.g. [15, 21]). However, in [1] it is argued that with 24 outputs it is difficult to avoid the number of weights becoming unreasonably large in comparison with the size of the estimation sample. We acknowledge that there are many other effective NN designs in the literature, which we could have implemented in this study. However, as our focus is on improved weather input to the modelling process, we felt that it was important to use a relatively straightforward and uncontroversial NN design. The methods discussed in this paper are relevant to other designs because all neural network load models are likely to be non-linear functions of weather variables. It is this non-linearity that makes the use of weather ensemble predictions particularly attractive.

C. *The Neural Network Inputs*

Our choice of inputs was influenced by the variables that have been used for many years in the linear regression models of NG. Short- to medium-term forecasting models must accommodate the variation in load due to the seasonal patterns shown in Fig. 2 and due to weather. NG forecasters use 0/1 dummy variables for each day of the week and for each of three summer weeks when a large amount of industry closes. In order to capture the autoregressive pattern in load, we included lagged demand variables. We considered lags of 1 to 7 days. A hold out method, with a third of the data used for testing [17, §9.8], indicated that only lags 1, 3 and 5 should be included in the model, and only the dummy variables for Fridays, Saturdays, Sundays and the second week of the industrial closure period.

In addition to these seven variables, we also included as inputs the three weather variables used at NG: 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 load variation [22]. *Effective temperature* (TE_t) for day 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.

$$TE_t = \tfrac{1}{2} TO_t + \tfrac{1}{2} TE_{t-1} \tag{3}$$

The influence of lagged temperature aims to reflect the delay in response of heating appliances within buildings to changes in external temperature. At NG, the non-linear dependence of load on effective temperature is modelled by the inclusion of higher powers of TE_t in their linear regression models. *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.

$$CP_t = \begin{cases} W_t^{\frac{1}{2}}(18.3 - TO_t) & if \quad TO_t < 18.3\ ^0C \\ 0 & if \quad TO_t \geq 18.3\ ^0C \end{cases} \tag{4}$$

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

Since NG forecasters need 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. We used the same weighted averages in this study.

As the aim of this paper is to investigate the potential for the use of ensemble predictions, we used only weather variables for which ensemble predictions were available. Ensemble predictions are available for spot temperature, wind speed and cloud cover 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 the NG formulae in expressions (3) and (4), respectively. The hold out method indicated that all three weather variables should be included as inputs to the NN model.

One might argue that variables should not be transformed prior to their use as inputs because the NN should be used to identify all non-linearities. However, an important stage of NN modelling is data pre-processing [1]. Since meteorologists and load forecasters have established that expression (4) satisfactorily captures the effect of the cooling power of the wind, it would be unwise to discard this information. Data pre-processing is also performed on weather variables in [23].

IV. Using Weather Ensembles in Load Forecasting

A. Creating 51 Scenarios for Load

When forecasting from non-linear models, such as NNs, it is important to be aware 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 [24]. In addition to the non-linearity in the NN, the definition of cooling power of the wind, given in expression (4), emphasises that our NN load model will be a non-linear function of the fundamental weather variables: temperature, wind speed and cloud cover. The usual approach to load forecasting involves substituting a single point forecast for each weather variable. In view of the result regarding the expectation of a non-linear function, it would be preferable to first construct the load probability density function, and then calculate its expectation.

Weather ensemble predictions enable an estimate to be constructed for the load density function. Since we have 51 ensemble members for temperature, wind speed and cloud cover, we can substitute these into the NN model to deliver 51 scenarios for load. The histogram of these load scenarios is an estimate of the density function. The mean of the load scenarios is an estimate of the mean of the density function. Meteorologists often find that the mean of the weather ensemble members is a more accurate forecast than a single point weather forecast. The collection of ensemble members must, therefore, contain information not captured by the single point forecast. This provides further motivation for forecasting load using the mean of the 51 load scenarios.

B. Comparison of Load Forecasting Methods

We used 22 months of daily data from 1 January 1997 to 31 October 1998 to estimate model parameters. It has been remarked in [1] that many studies implement NNs with far too many parameters in relation to the size of the estimation sample. Our estimation sample of 22 months consisted of 669 daily observations with which to estimate the 121 parameters of our NN. This ratio of sample size to number of parameters is bettered by only one of the studies reviewed in [1]. Design of the NN model, choice of NN inputs and NN parameter estimation were based only on this sample of 669 observations. We used 20 months of daily data from 1 November 1998 to 30 June 2000 to evaluate the resulting forecasts. These 20 months are the months for which we had weather ensemble predictions. We produced forecasts for each day in our evaluation period for lead times of 1 to 10 days ahead. We compared forecasts from the following four methods using the mean absolute percentage error (*MAPE*) summary measure, which is used extensively in the load forecasting literature.

Method 1: *NN using traditional weather point forecasts* - This is the usual procedure of substituting traditional single weather point forecasts in the NN load model.

Method 2: *mean of NN load scenarios* - This is the mean of the 51 load scenarios. This approach is based on the weather ensemble predictions since the 51 scenarios are constructed from the 51 ensemble members.

Method 3: *NN using actual weather as forecasts* - In order to establish the limit on load forecast accuracy that could be achieved with improvements in weather forecast information, we produced load ‘forecasts’ using actual observed weather substituted for the weather variables in the NN load model. Clearly this level of forecast accuracy is unattainable, as perfect weather forecasts are not achievable.

Method 4: *univariate* - 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 univariate model that does not include any of the weather variables:

$$\begin{aligned} demand_t &= b_0 + b_1\, FRI_t + b_2\, SAT_t + b_3\, SUN_t + \varepsilon_t \\ \varepsilon_t &= \phi_1\, \varepsilon_{t-1} + \phi_2\, \varepsilon_{t-2} + \psi_1\, u_{t-1} + u_t \end{aligned} \qquad (5)$$

where FRI_t, SAT_t and SUN_t are day of the week 0/1 dummy variables, and the b_i, ϕ_i and ψ_1 are constant parameters. The model was constructed using the standard Box-Jenkins statistical modelling steps. Comparison of NN predictions with forecasts from a simpler benchmark method is one of the recommendations in [25] for effective NN validation.

Fig. 3 presents the *MAPE* results for the four methods. The figure shows that the weather-based methods comfortably dominate the method using no weather variables beyond a lead time of 1 day. It is interesting to note that, for 1 to 3 day-ahead load 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. As in [5], this shows how weather forecast error can have a significant impact on load forecast accuracy. 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 all 10 lead times. These improvements brought the *MAPE* results noticeably closer to those of the method using actual observed weather, which is an unattainable benchmark. For lead times of 5, 6 and 10 days ahead, the accuracy of the new ensemble based NN approach is as good as that of the traditional NN approach at the previous lead time. This could be described as a gain in accuracy of a day over the traditional approach.

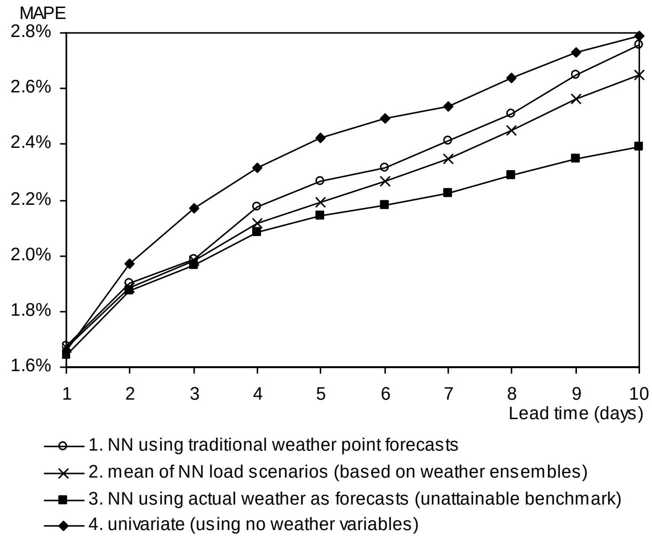

Fig. 3. *MAPE* for load point forecasts for post-sample period 1 November 1998 to 30 June 2000.

V. Using Weather Ensembles to Estimate the Load Forecast Error Variance

The estimation of the variance of the probability distribution of load forecast error is not a trivial task, as the forecast error variance is likely to vary over time due to weather and seasonal effects [5, 6]. The approach that we took was to model the variance in a series of historical post-sample forecast errors. This is similar to the approach taken in [26], where the absolute magnitude of the errors is modelled. Since the method using weather ensemble predictions as input produced the most accurate post-sample forecasts in the previous section, we focused on estimation of the variance of the forecast errors from this method. We considered lead times from 1 to 10 days ahead. We used the first 10 months (1 November 1998 to 31 August 1999) of post-sample errors from our earlier analysis of point forecasting to estimate model parameters, and the remaining 10 months (1 September 1999 to 30 June 2000) of post-sample errors to evaluate the resulting variance estimates. We implemented the following three variance estimation methods.

Method 1: *naïve* - This method produces simple benchmark variance estimates. For each lead time, h, we calculated the variance of the h day-ahead errors in the estimation period of 10 months. For example, we estimated the future variance of the 5 day-ahead forecast errors using the variance of the 5 day-ahead errors from the previous 10 months.

Method 2: *exponential smoothing* - We used an exponentially weighted moving average of past squared errors, e_t^2, to allow the variance estimate to adapt over time. We optimised the smoothing parameter, α, separately for each lead time. This method is used in financial volatility forecasting. This estimator is constructed as:

$$\hat{\sigma}_t^2 = \alpha e_{t-1}^2 + (1-\alpha)\hat{\sigma}_{t-1}^2 \qquad (6)$$

Method 3: *rescaled variance of NN load scenarios* - The level of uncertainty in the load 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 load forecast uncertainty. The variance of the 51 load scenarios, discussed in Section IV, conveys the uncertainty in the load due to weather uncertainty. 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. However, the variance of the 51 scenarios will substantially underestimate the load forecast error variance because it does not accommodate the uncertainty due to the NN model residual error and parameter estimation error. This was confirmed by our empirical analysis. In view of this, for each lead time, we rescaled the estimator by regressing the squared forecast error on $\sigma_{ENS,t}^2$ using just the first 10 months of post-sample data. This results in an estimator of the form $\hat{a} + \hat{b}\,\sigma_{ENS,t}^2$, where $\hat{a}$ and $\hat{b}$ are constant parameters.

Fig. 4 shows the R^2, from the regression of the squared post-sample forecast errors on the variance estimates for the 10-month post-sample evaluation period. Higher values of the R^2 are better. This measure is widely used in volatility forecast evaluation in finance. Typically, the R^2 values are low, with values less than 10% being the norm [27]. The R^2 for the *naïve* estimator was zero for all lead times, as it does not vary during the 10-month evaluation period. *Exponential smoothing* is the best for the first three lead times, but beyond that, it is comfortably outperformed by the *rescaled variance of NN load scenarios*.

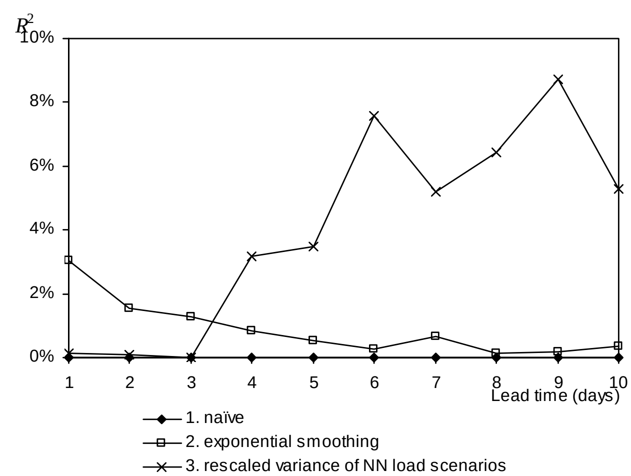

Fig. 4. R^2 percentage measure for forecast error variance estimation methods for post-sample period 1 September 1999 to 30 June 2000.

VI. Using Weather Ensembles to Estimate Load Prediction Intervals

An alternative description of the load forecast error distribution is given by a prediction interval. In order to consider both the tails and the body of the predictive distribution, we focused on estimation of 50% and 90% intervals. More specifically, we evaluated different methods for 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 V, we used 10 months of post-sample errors from our earlier analysis of load point forecasting to estimate parameters, and the remaining 10 months of errors for evaluation.

We constructed quantile estimators using the three variance estimators investigated in Section V with either a Gaussian distribution or the empirical distribution of the corresponding standardised forecast errors, $e_t / \hat{\sigma}_t$ (see [28] and [29]). The use of the empirical distribution was generally more successful than the Gaussian distribution and so, in the remainder of this section, we limit our focus to comparison of the quantile estimators based on the empirical distribution.

The upper tail tends to be the most important part of the load distribution for scheduling purposes; the problems caused by a large shortfall in electricity availability tend to be more serious than those resulting from an oversupply of the same size. Fig. 5 compares estimation of the 95% quantiles at the 10 different lead times for the post-sample period of 10 months. The figure shows the percentage of errors falling below the 95% quantile estimators. For estimation of the 95% quantile, the ideal is 95%. The dashed horizontal lines in Fig. 5 are the bounds of the acceptance region for the test of whether the percentages are significantly different from 95% (at the 5% level). The test uses a Gaussian distribution and the standard error formula for a proportion. Although the *exponential smoothing* based estimator performs well for the early lead times, it fades badly beyond 6 days ahead. The estimator based on the *rescaled variance of NN load scenarios* performs well at the early lead times and comfortably outperforms the other two estimators for the longer horizons.

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%. Fig. 6 shows the resulting chi-squared statistics. Lower values are better. The dashed horizontal line in the figure is the bound of the acceptance region for the 5% significance test on the chi-squared statistic. The chi-squared statistic for the estimator based on the *rescaled variance of NN load scenarios* lies under the statistics for the other two methods for all but two of the 10 lead times indicating an overall superiority of this estimator.

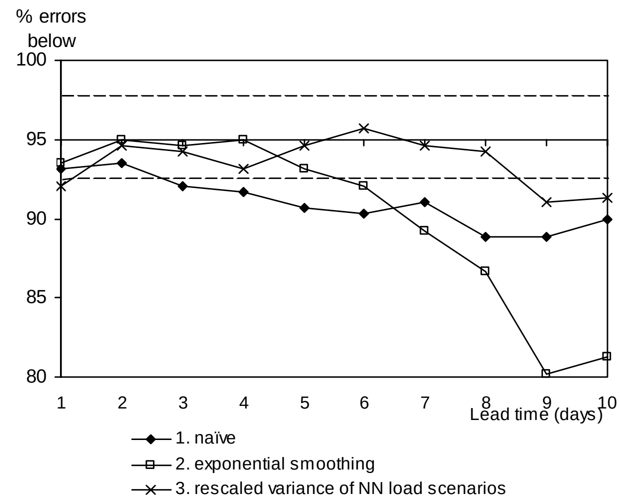

Fig. 5. Percentage of post-sample forecast errors falling below various 95% quantile estimators for the period 1 September 1999 to 30 June 2000. All 3 estimators were based on empirical distribution.

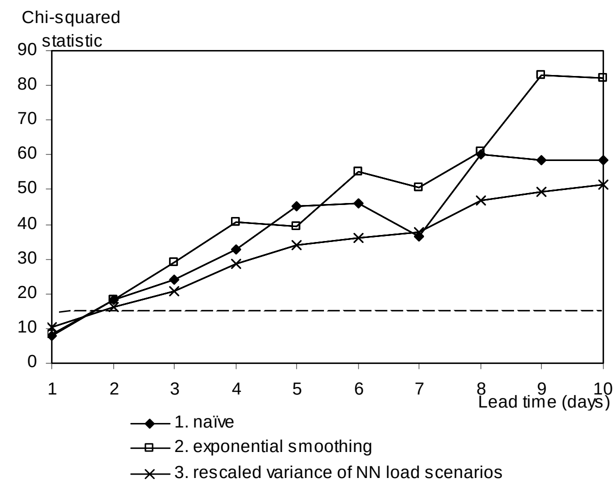

Fig. 6. Chi-squared statistics summarising overall estimator bias for 5%, 25%, 75% and 95% forecast error quantiles for the period 1 September 1999 to 30 June 2000. All 3 estimators were based on empirical distribution.

VII. Summary and Conclusions

We have shown how weather ensemble predictions can be used in NN load forecasting for lead times from 1 to 10 days ahead. We used the 51 ECMWF ensemble members for each weather variable to produce 51 scenarios for load from a NN. For all 10 lead times, the mean of the load scenarios was a more accurate load forecast than that produced by the traditional procedure of substituting a single point forecast for each weather variable in the NN load model. This traditional procedure amounts to approximating the expectation of the NN non-linear function of weather variables by the same non-linear function of the expected values of the weather variables. The mean of the 51 scenarios is appealing because it is equivalent to taking the expectation of an estimate of the load probability density function.

The distribution of the 51 load scenarios provides information regarding the uncertainty in the load forecast. However, since the distribution does not accommodate the NN load model uncertainties, it will tend to underestimate the load forecast uncertainty. In view of this, we rescaled the variance of the load scenarios before using it as an estimator of the load forecast error variance. The resulting estimator compared favourably with benchmark estimators based purely on historical forecast error. Using the same variance estimator as a basis for estimating prediction intervals also compared well with benchmark methods. We, therefore, conclude that there is strong potential for the use of weather ensemble predictions in NN load forecasting.

VIII. Acknowledgements

We are grateful to Shanti Majithia and Chris Rogers of the National Grid for supplying the load data, and to Henrique Hippert for helpful suggestions with the paper. We are also grateful for the useful comments of three anonymous referees.

IX. References

[1] H.S. Hippert, C.E. Pedreira, R.C. Souza, "Neural Networks for Short-Term Load Forecasting: A Review and Evaluation," *IEEE Trans. on Power Systems*, vol. 16, pp. 44-55, 2001.

[2] J.W. Taylor, R. Buizza, "Using Weather Ensemble Predictions in Electricity Demand Forecasting," *International Journal of Forecasting*, forthcoming, 2002.

[3] C.E. Leith, "Theoretical skill of Monte Carlo forecasts," *Monthly Weather Review*, vol. 102, pp. 409-418, 1974.

[4] F. Molteni, R. Buizza, T.N. Palmer and T. Petroliagis, "The new ECMWF ensemble prediction system: Methodology and validation," *Quarterly Journal of the Royal Meteorological Society*, vol. 122, pp. 73-119, 1996.

[5] A.P. Douglas, A.M. Breipohl, F.N. Lee and R. Adapa. "The impact of temperature forecast uncertainty on Bayesian load forecasting," *IEEE Trans. Power Systems*, vol. 13, pp. 1507-1513, 1998.

[6] D.K. Ranaweera, G.G. Karady and R.G. Farmer," Effect of probabilistic inputs in neural network-based electric load forecasting," *IEEE Trans. Neural Networks*, vol. 7, 1528-1532, 1996.

[7] T.N. Palmer, F. Molteni, R. Mureau, R. Buizza, P. Chapelet and J. Tribbia, "Ensemble prediction," Proceedings of the ECMWF Seminar on *Validation of Models Over Europe: Vol. I*, 1993. ECMWF, Shinfield Park, Reading, RG2 9AX, UK.

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

[9] Z. Toth and E. Kalnay, "Ensemble forecasting at NMC: The generation of perturbations," *Bulletin of the American Meteorological Society*, vol. 74, pp. 2317-2330, 1993.

[10] M.S. Tracton and E. Kalnay, "Operational ensemble prediction at the National Meteorological Center: Practical aspects," *Weather and Forecasting*, vol. 8, pp. 379-398, 1993.

[11] R. Buizza, M. Miller and T.N. Palmer, "Stochastic simulation of model uncertainties," *Quarterly Journal of the Royal Meteorological Society*, vol. 125, pp. 2887-2908, 1999.

[12] A.C. Harvey and S.J. Koopman, "Forecasting hourly electricity demand using time-varying splines," *Journal of the American Statistical Association*, vol. 88, pp. 1228-1236, 1993.

[13] R. Ramanathan, R. Engle, C.W.J. Granger, F. Vahid-Araghi and C. Brace, "Short-run forecasts of electricity loads and peaks," *International Journal of Forecasting*, vol. 13, pp. 161-174, 1997.

[14] J.W. Taylor and S. Majithia, "Using combined forecasts with changing weights for electricity demand profiling," *Journal of the Operational Research Society*, vol. 51, pp. 72-82, 2000.

[15] T.W.S. Chow and C.-T. Leung, "Neural network based short-term load forecasting using weather compensation," *IEEE Tans. Power Systems*, vol. 11, pp. 1736-1742, 1996.

[16] G. Zhang, B.E. Patuwo and M.Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, pp. 35-62, 1998.

[17] C.M. Bishop, *Neural Networks for Pattern Recognition.* Oxford: Oxford University Press, 1997.

[18] Z. Tang and P.A. Fishwick, "Feedforward neural nets as models for time series forecasting," *ORSA Journal on Computing*, vol. 5, pp. 374-385, 1993.

[19] I. Drezga and S. Rahman, "Short-term load forecasting with local ANN predictors," *IEEE Trans. Power Systems*, vol. 13, pp. 1238-1244, 1998.

[20] J.S. McMenamin and F.A. Monforte, "Short-term energy forecasting with neural networks," *Energy Journal*, vol. 19, pp. 43-61, 1998.

[21] A.G. Bakirtzis, V. Petridis, S.J. Kiartzis, M.C. Alexiadis and A.H. Maissis, "A neural network short term load forecasting model for the Greek power system", *IEEE Trans. Power Systems*, vol. 11, pp. 858-863, 1996,

[22] A.B. Baker, "Load forecasting for scheduling generation on a large interconnected system," in *Comparative Models for Electrical Load Forecasting*, D.W. Bunn and E.D. Farmer, Eds. Chichester: Wiley, 1985, pp 57-67.

[23] A.D. Papalexopoulos, S. Hao and T.-M. Peng, "An implementation of a neural network based load forecasting model for the EMS," *IEEE Trans. Power Systems*, vol. 9, pp. 1956-1962, 1994.

[24] J. Lin and C.W.J. Granger, "Forecasting from non-linear models in practice," *Journal of Forecasting*, vol. 13, pp. 1-9, 1994.

[25] M. Adya and F. Collopy, "How effective are neural networks at forecasting and prediction? A review and evaluation," *Journal of Forecasting*, vol. 17, pp. 481-495, 1998.

[26] R. Engle, C.W.J. Granger, R. Ramanathan, and F. Vahid-Araghi, "Probabilistic methods in forecasting hourly loads and peaks," Electric Power Research Institute, Palo Alto, CA, EPRI TE-101902, 1993.

[27] T.G. Andersen and T. Bollerslev, "Answering the critics: Yes, ARCH models do provide good volatility forecasts," *International Economic Review*, vol. 39, pp. 885-906, 1998.

[28] C.W.J. Granger, H. White, and M. Kamstra, "Interval forecasting: an analysis based upon ARCH-quantile estimators," *Journal of Econometrics*, vol. 40, pp. 87-96, 1989.

[29] J.W. Taylor, "Evaluating volatility and interval forecasts," *Journal of Forecasting*, vol. 18, pp. 111-128, 1999.

X. 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 year 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.