On the Influence of Weather Forecast Errors in Short-Term Load Forecasting Models

Damien Fay and John V. Ringwood, *Senior Member, IEEE*

Abstract—Weather information is an important factor in load forecasting models. Typically, load forecasting models are constructed and tested using actual weather readings. However, online operation of load forecasting models requires the use of weather forecasts, with associated weather forecast errors. These weather forecast errors inevitably lead to a degradation in model performance. This is an important factor in load forecasting but has not been widely examined in the literature. The main aim of this paper is to present a novel technique for minimizing the consequences of this degradation. In addition, a supplementary technique is proposed to model weather forecast errors to reflect current accuracy.

The proposed technique utilizes a combination of forecasts from several load forecasting models (sub-models). The parameter estimation may thus be split into two parts: sub-model and combination parameter estimation. It is shown that the lowest PMSE corresponds to training the sub-models with actual weather but training the combiner with forecast weather.

Index Terms—Load forecasting, model combination, neural networks, weather forecast errors.

I. INTRODUCTION

SHORT-TERM load forecasting (STLF) refers to forecasts of electricity demand (or load), on an hourly basis, from one to several days ahead. The amount of excess electricity production (or spinning reserve) required to guarantee supply, in the event of an underestimation, is determined by the accuracy of these forecasts. Conversely, overestimation of the load leads to sub-optimal scheduling (in terms of production costs) of power plants (known as unit commitment). In addition, a deregulated market structure exists in Ireland in which load forecasts play a central role. As indicated above, STLF is an important area and this is reflected in the literature by the many techniques that have been applied, including neural networks [1], fuzzy logic [2] and statistical techniques [3], to mention but a few. In many electricity grid systems, the prevailing weather has a significant effect on the load and it has been found that including weather information can improve a load forecast [3]. However, in order to use weather information for future load forecasts, weather forecasts must be utilized and these have associated weather forecast

errors. Although system dependent, weather forecast errors can be significant [4] and have been attributed as the cause of 17% [5] to 60% [6] of load forecast errors.

Load forecasting models are usually trained using actual past weather readings as opposed to past weather forecasts [7]. This is based on the assumption that to use the latter essentially adds forecast noise to the training data which can lead to biased parameter estimation [8]. Often weather forecasts are unavailable for the entire training period and/or can be subject to increasing accuracy of meteorological models, as mathematical weather models are constantly improved. Therefore, training load models with actual weather can be justified [7]. However, when weather forecast errors not present in the training set are presented, they can have a disproportionate influence on load models [9]. Changing the load model parameters to account for this can be impossible in many conventional models once training is completed. Douglas *et al.* [5] approached this problem by use of a Bayesian framework, but restricted analysis to the use of dynamic linear models. In spite of the importance of weather forecast errors with respect to load forecasting, the literature is sparse [10], [11].

The main contribution of this paper is the combination of several models (called sub-models), or model fusion, as a technique for minimizing the effect of weather forecast errors in load forecasting models. Model fusion is particularly suited to STLF as the sub-models may be trained with actual weather information and the effect of weather forecast errors taken into account when combining the models. While the concept of model fusion is well known in the general field of forecasting and was pioneered mainly in [12], its use to deal with forecasting errors in causal variables is new. For example, a linear combiner is used in [13] to combine ARIMA models, with weights updated iteratively following each additional observation, based on model stability measures. In [14] the short-term load forecasting problem is addressed using a combination of forecasts, where weather inputs are considered using a "factor vector", though autoregressive models are used for the individual forecasts. The weights of the linear combiner are updated at each time step using a forecasting performance measure and errors in weather forecasts are not considered. Similar approaches may be found in [15] and [16].

Fused forecasts are theoretically more accurate than any of the individual model forecasts [17], [18], though there are some results which show little benefit under certain circumstances [19], in particular where an unweighted mean of forecasts is used to model an autoregressive time series. Nevertheless, different models are often better at modeling different aspects of an underlying process and combining the models *appropriately* (i.e., by taking weather forecast errors into account) thus gives

TABLE I
DATA TIME SCALE AND RANGE

Range	29/12/1986-31/03/2000
Time Scale	Hourly
No. of data points	4842 Days (116208 hours)

TABLE II
WEATHER DATA TIME-SCALE AND RANGE

Type	Range	Time scale
Weather readings	29/12/1986 - 31/03/2000	Hourly
Weather forecasts	01/02/2000- 01/03/2000	Hourly

TABLE III
DIVISION OF DATA SET

Set	Training	Validation	Novelty
Range	1987-1997	1998	1999-2000

Fig. 1. Actual and forecast temperature (February 6–15, 2000).

a better forecast. In addition, a single model incorporating all aspects of an underlying process may be more complex and difficult to train than combining individual models [17]. However, it should be noted that a fusion model is not a universal approximator as information may be lost by the sub-models which cannot be recovered by the fusion model.

II. DATA SETS

The range and time-scale of the available electrical demand data is given in Table I. Two categories of historical weather data are available from the Meteorological Office of Ireland (MOI): readings (or actual weather) and forecasts. Both sets of data are for Dublin airport, the closest and most relevant weather station to Dublin (Table II). The readings and forecasts are for dry bulb temperature, cloud cover, wind speed and wind direction.

The data is subdivided into three sets in order to train, validate and test the load forecasting models (Table III). The training set is used to estimate model parameters, the validation set is used to aid in model structure determination and the novelty set is used to evaluate model performance.

Data between Monday and Friday in the months January to March (known as the *late winter working day day-type*) is selected so as to avoid the exceptions associated with weekend, Christmas and changes due to the daylight saving hour.

The relatively long data record used permits the data to be disaggregated according to hour-of-day and day type and provides sufficient data to obtain reliable statistics on training, validation and novelty sets. The potential issues of data "currency" and nonstationarity associated with a relatively long dataset are dealt with by:
- the use of a Kalman filter to determine the underlying seasonal and IRW components; and
- the use of a heterogeneity transform (4) to deal with increasing variance.

III. MODELING WEATHER FORECAST ERRORS

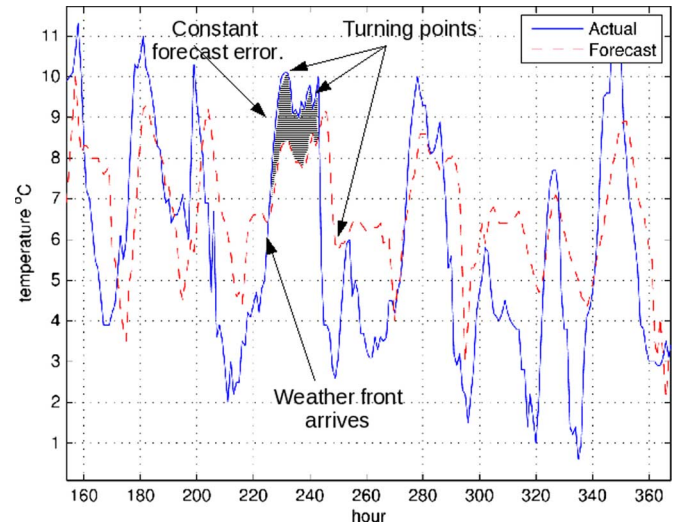Due to the sparseness of weather forecast data available to us (Table II) it is necessary to model the weather forecast error to produce pseudo-weather forecasts for the entire data set. Indeed, even given a long database of weather forecasts, modeling the error is desirable. This is because the quality of weather forecasts is changing over time due to improved forecasting techniques and climate change [20]. Previous approaches in STLF have modeled the weather forecast error simply as an IID Gaussian random variable [21], [22]. However, as seen in Fig. 1 this is not an accurate representation of the statistics of the weather forecast errors in Ireland. Rather, the forecast error displays serial correlation, i.e., it is either above or below the actual for prolonged periods (i.e., the errors are not identically distributed, Fig. 1). Typically some form of aggregate weather variables are normally used in STLF models (e.g., average daily temperature). Serial correlation can introduce large deviations in the sample mean and so an IID noise process is not a good model for weather forecast errors.

The weather in Ireland is dominated by Atlantic weather systems. When a weather system or front reaches Ireland there is a shift in the level of the temperature and other weather variables (Fig. 1) (a similar situation is noted in [3]). This shift is also a factor that the Irish Meteorological Office must forecast. The weather forecast error is thus assumed to have the following structure:
- turning points (Fig. 1) which represent the arrival of a weather front;
- a level error, $\widetilde{\mu}$, which is the average of the weather forecast error between turning points;
- a shape error, $\widetilde{\sigma}$, which is the standard deviation of the weather forecast error between turning points; and
- a random error, which accounts for the remaining error if $\widetilde{\mu}$ and $\widetilde{\sigma}$ are removed.

In order to detect the turning points the following simple algorithm was found to suffice. The weather variable is first smoothed by means of a state space model based on an integrated random walk:

$$\mathbf{X}(k) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \mathbf{X}(k-1) + \boldsymbol{\epsilon}(k) \qquad (1)$$
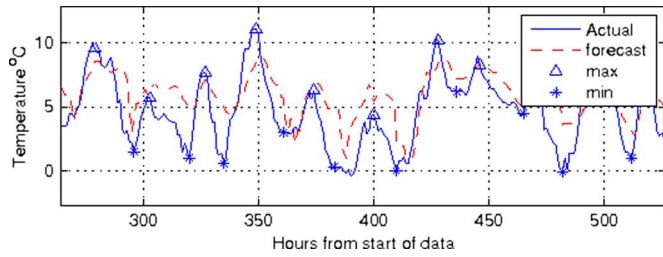
Fig. 2. Sample of the turning points calculated for temperature.

where $\mathbf{X}(k)$ is the state vector at time k and $\boldsymbol{\epsilon}(k)$ is the process noise. The temperature is then extracted from the state vector by means of the measurement equation:

$$y(k) = [1 \quad 0]\mathbf{X}(k) + \boldsymbol{\nu}(k) \tag{2}$$

where $y(k)$ is the filtered weather variable and $\nu(k)$ is the measurement noise. The state vector is estimated using the Kalman filter (Note: the *a-posteriori* state vector estimate is used in (2) as a smoothed version of the original is desired [23]). The turning points are then defined as the maxima and minima within a rolling window of length 5:

$$\mathbf{y} = \left\{ y(k) : y(k) <> 0.5 \sum_{i=0}^{11} \frac{y(k-5+i)}{10}, i \neq 5, \forall k \right\} \tag{3}$$

where $\mathbf{y}$ is the set of turning points and $<>$ denotes greater or less than. A sample of the turning points detected by this algorithm is shown in Fig. 2.

Fig. 3 shows the histograms, fitted Gaussian distributions and the Sample AutoCorrelation Function (SACF) for the level shape and random error of the temperature forecasts. The shape and level errors of the four weather variables are found to be cross correlated, suggesting that they may not be independent. In order to generate pseudo-weather forecast errors, the turning points in the actual weather variables are first identified. Then, a multivariate Gaussian pseudo-random number generator is used to generate the random errors for each of the weather variables jointly. Fig. 4 shows the SACF of the temperature forecast errors and the pseudo-temperature forecast errors. As can be seen, the SACF for both are similar, showing that the pseudo-forecast errors have captured the auto-correlation evident in the temperature forecast errors. A similar situation was found with the other weather variables.

IV. FUSION MODEL

A. Preliminary Auto-Regressive (AR) Linear Model

It was previously found by these authors [24] that decomposing load data into 24 parallel series, one for each hour of the day, is advantageous, as the parallel series are not interdependent. The parallel series for hour j on day k, $y(j,k)$, has a low frequency trend, $d(j,k)$, which is first removed using a Basic Structural Model (BSM) [25], via an integrated random walk, leaving a residual, $x(j,k)$ (Fig. 5), which is composed of
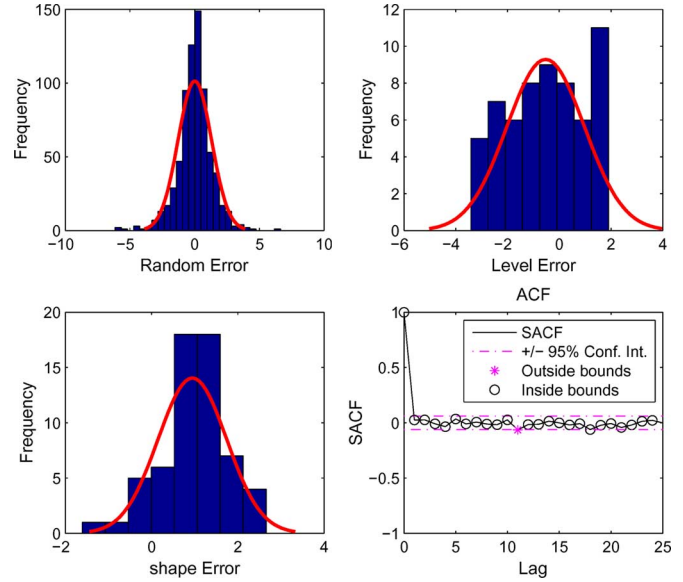

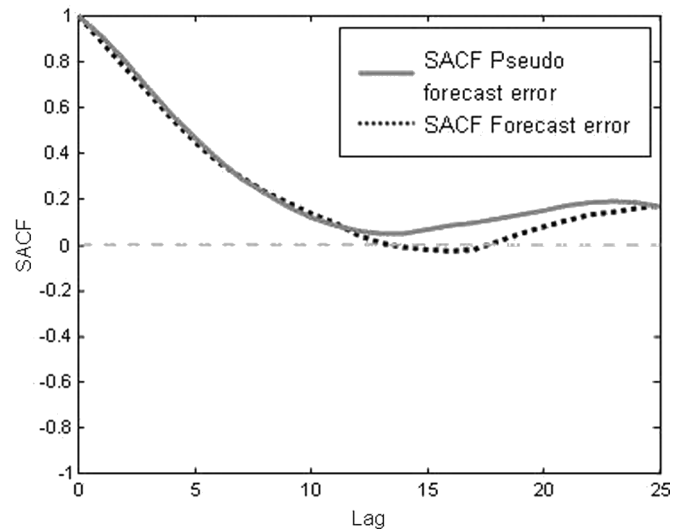Fig. 3. Distributions and SACF for temperature forecasts.

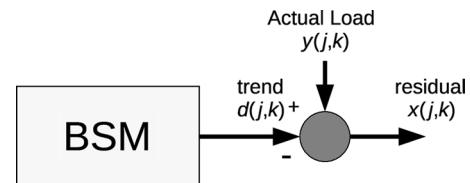Fig. 4. SACF of forecast and pseudo-forecast temperature errors.

Fig. 5. Preliminary AR linear model overview.

weather, nonlinear auto-regressive and white noise components [24].

B. Sub-Models

Three sub-models were chosen which have different types of inputs. These are chosen so that forecast errors can be attributed to particular inputs. A fourth sub-model is included using all the available inputs to capture any nonlinear relationships between the inputs and the residual. The sub-models are named after their
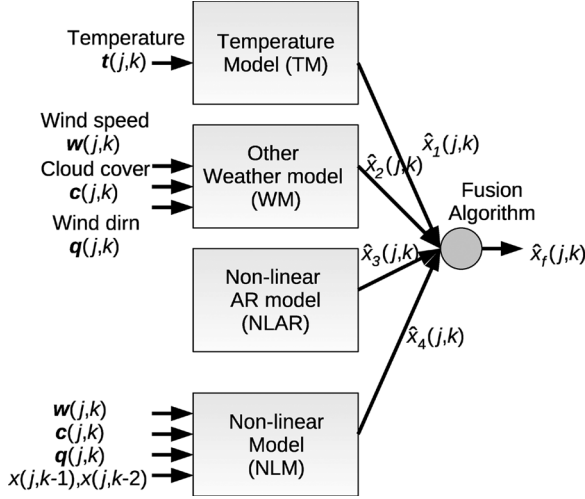
Fig. 6. Data fusion model overview.

input types as shown in Fig. 4. The fusion technique combines the forecasts of the sub-models, $\hat{x}_1(j,k),\ldots,\hat{x}_4(j,k)$, to give a fused forecast, $\hat{x}_f(j,k)$ of the residual for series j on day k (Fig. 6). It should be noted that all of the sub-models are trained on actual, rather than forecast data. The consideration of forecast errors takes place in the fusion algorithm, as detailed in Section IV-C.

The sub-models all use feed forward neural networks, although it should be noted that the choice of modeling technique is not central to this paper. Initially, the traditional back-propagation algorithm using Levenberg-Marquadt with cross validation was used to train the networks. Each of the networks has two hidden layers and a single output. To determine a suitable structure for the network (i.e., the number of nodes in each layer), different network structures were trained (ranging from a 1×1 to a 7×7 network) and their prediction mean squared errors (PMSE) compared over the validation set. The best structure was then selected for further evaluation. Given these initial models, the residuals where then examined for homogeneity of variance and it was concluded that the time series possessed nonconstant variance. The most likely cause for the nonconstant variance lies in the considerable growth experienced in Irish electricity demand over the period of the data set. With the increase in electricity demand a corresponding increase in forecasting error (and thus variance) would be expected. The standard approach in this case is to presume that the variance is proportional to the level of the time series squared, specifically $y^2(j,k)$, and then to scale the errors using weighted least squares (see [26, Section 8.3]). During training with the back-propagation algorithm the target errors are thus scaled prior to being propagated backwards as

$$\mathbf{e}' = \begin{bmatrix} \frac{1}{|y(j,1)|} & 0 & \cdots & 0 \\ 0 & \frac{1}{|y(j,2)|} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{|y(j,N)|} \end{bmatrix} \mathbf{e}, \quad j = 1,\ldots,24 \tag{4}$$

where $\mathbf{e}$ is vector of target errors and $\mathbf{e}'$ is the adjusted vector. It was found that this improved the prediction performance of the models in all cases. The Temperature Model (TM) input, $\mathbf{t}(j,k)$, is a vector of the current and previous 71 h of temperature from hour j on day k. Similarly the other Weather Model (WM) uses vectors of wind speed, $\mathbf{w}(j,k)$, cloud cover, $\mathbf{c}(j,k)$, and wind direction, $\mathbf{q}(j,k)$ containing current and previous 71 h of weather. The Non-Linear Auto-Regressive model (NLAR) uses the previous two days of residual, $x(j,k-1)$ and $x(j,k-2)$. The Non-Linear Model (NLM) uses all the available inputs.

C. Fusion Algorithm

The data fusion algorithm described in [27] seeks to minimize the variance of the fused forecast based on the covariance matrix of the sub-model forecasts. In this way, the load forecast errors, due to weather forecast errors, are taken into account. The cross-covariance of the forecasts is considered and the distribution of the forecast error noise is not restricted to Gaussian but merely required to be unbiased. A combined forecast, $x_f(j,k)$, of the load is created using a weighted average of the individual forecasts, $\hat{x}_1(j,k),\ldots,\hat{x}_4(j,k)$ [27]:

$$x_f(j,k) = \sum_{i=1}^{4} A_i(j)x_i(j,k) \tag{5}$$

where $A_i(j)$ is the weight applied to the forecast from sub-model i for hour j,[1] and is derived from the error covariance matrices of $\hat{x}_1(j,k),\ldots,\hat{x}_4(j,k)$ as

$$[A_1(j) \quad A_2(j) \quad A_3(j)] = [P'_{4,1}(j) \quad P'_{4,2}(j) \quad P'_{4,3}(j)] \, \mathbf{P}^{-1} \tag{6}$$

where $P'_{4,1}(j)$, $P'_{4,2}(j)$, $P'_{4,3}(j)$ and $\mathbf{P}$ are auxiliary variables derived from the sample error covariance of $\hat{x}_1(j,k),\ldots,\hat{x}_4(j,k)$:

$$P_{i,n}(j) = \frac{1}{M}\sum_{k=1}^{M}(x(j,k)-\hat{x}_i(j,k))(x(j,k)-\hat{x}_n(j,k)) \tag{7}$$

where $P_{i,n}(j)$ is the error covariance of sub-model i with sub-model n for hour j, and M is the number of samples used. The auxiliary variables are then defined as

$$P'_{4,i}(j) = P_{4,4}(j) - P_{4,i}(j) \quad i \neq 4 \tag{8}$$

and

$$\mathbf{P} = \begin{bmatrix} P'_{1,1}(j) & P'_{1,2}(j) & P'_{1,3}(j) \\ P'_{2,1}(j) & P'_{2,2}(j) & P'_{2,3}(j) \\ P'_{3,1}(j) & P'_{3,2}(j) & P'_{3,3}(j) \end{bmatrix} \tag{9}$$

[1]Although it is assumed that the variance is proportional to $y^2(j,k)$, an adjustment for heteroskedasticity is not necessary here as multiplying $P_{i,n}(j)$ by a scaling factor will not change the weights.

TABLE IV
CROSS-COVARIANCE MATRIX OF SUB-MODEL LOAD
FORECAST ERRORS (CASE I→CASE II)

Sub-model	1	2	3	4
NLAR	3834→3834	2403→2943	3893→3848	2920→3283
TM	2403→2943	2542→3911	2472→3129	2386→3171
WTM	3893→3848	2473→3129	4344→4265	2816→3285
NLM	2920→3283	2386→3171	2816→3285	3026→3444

where

$$P'_{i,n}(j) = P_{i,n}(j) - P_{4,n}(j) - P_{i,4}(j) + P_{4,4}(j) \quad i \neq 4, n \neq 4. \tag{10}$$

The final weight A_4 is determined using the constraint that $x_f(j)$ is unbiased:

$$A_4(j) = 1 - \sum_{i=1}^{3} A_i(j). \tag{11}$$

Finally the fused load forecast, $\hat{y}_f(j, k)$, is estimated by reintroducing the trend:

$$\hat{y}_f(j, k) = \hat{d}(j, k) + \hat{x}_f(j, k). \tag{12}$$

V. RESULTS

The results are here analyzed for three cases. The first examines the behavior of the fusion model without pseudo-weather forecasts and the second and third examines the behavior with them. Case II is the most relevant case with the others included for comparison:

Case I: The sub-model parameters are estimated using actual weather inputs. The error covariance matrices of the sub-models (7) are then estimated using actual weather inputs. The weights, $A_i(j)$, are then calculated using these error covariance matrices (as in Section IV-C).

Case II: The sub-model parameters are estimated using actual weather inputs (as in Case I). The error covariance matrices of the sub-models (7) are then estimated using pseudo-weather forecast inputs (unlike Case I). The weights, $A_i(j)$, are then calculated using these (new) error covariance matrices (as in Section IV-C). Models are trained and evaluated using pseudo-weather forecast inputs.

Case III: The sub-model parameters and $A_i(j)$ are calculated as in Case I (i.e., based on actual weather). However, in this case the models are evaluated using pseudo-weather forecasts as input (thus a comparison can be drawn with Case II).

As an example, the cross-covariance matrix of sub-model forecast errors is shown in Table IV for the midday series ($j = 12$). The difference between Cases I and II is indicated by an arrow. As can be seen the covariance of sub-models 2 to 4 increases when pseudo-weather forecasts are used. This increase

TABLE V
EXAMPLE OF FUSION WEIGHTS (ACTUAL WEATHER INPUTS)

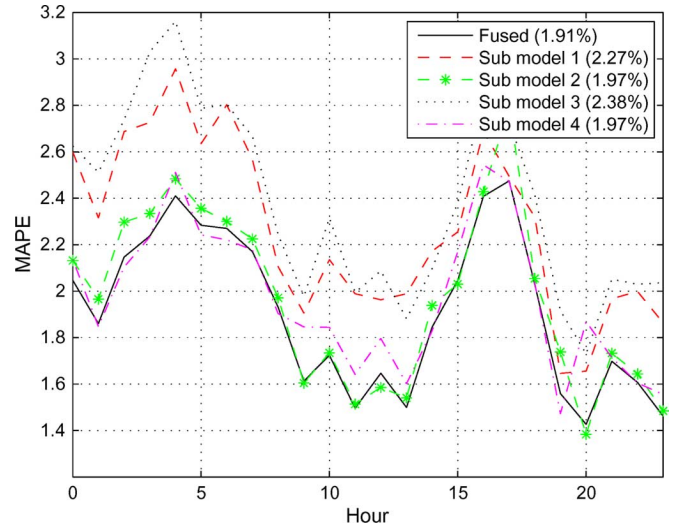| Sub-model	NLAR	TM	WM	NLM
Case I	0.13	0.82	-0.11	0.15
Case II	0.41	-0.31	-0.16	0.43

Fig. 7. MAPE as a function of hour of the day for fusion and sub-models (notes: novelty set, actual weather used).

TABLE VI
MODEL PERFORMANCE USING ACTUAL WEATHER INPUTS

Model	Training set		Validation set		Novelty set	
	MAPE	Sample error variance	MAPE	Sample error variance	MAPE	Sample error variance
NLAR	2.33	3488	2.16	4895	2.27	6311
TM	2.08	2747	1.89	3902	1.97	4810
WM	2.39	3583	2.19	4981	2.38	6938
NLM	2.11	2885	1.91	3951	1.97	5008
Fusion	1.98	2530	1.83	3655	1.91	4584

indicates the degradation of the models due to (pseudo) weather forecast error.

The corresponding values of $A_1, \dots A_4$ are shown in Table V. Note that the weights change significantly in the presence of pseudo-weather forecasts.

Fig. 7. shows the mean absolute percentage error (MAPE) for the sub-models and the fusion model in Case I. As can be seen, the fusion model performs best for each hour of the day.

Table VI, summarizes the results in the training, validation and novelty data sets.

Fig. 8 shows the MAPE for the sub-models and the fusion model using pseudo-forecast weather inputs in the novelty set (Case II). As can be seen the fusion model again performs best for each hour of the day.

Tables VI and VII enumerate the benefit of the fusion model. Table VI is included for reference, corresponding to the case of no weather forecast errors. There is a modest improvement of the fusion model (MAPE = 1.91%) over the best individual model (TM and NLM, with MAPE = 1.97%) for the novelty set. However, in Table VII, a more significant improvement is recorded, with a MAPE of 2.15% for the fusion model (Case
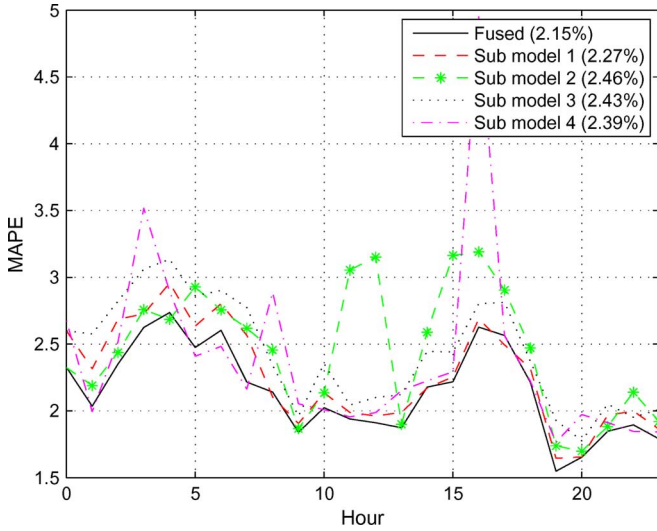
Fig. 8. MAPE as a function of hour of the day for fusion and sub-models (notes: novelty set, pseudo-weather forecasts used).

TABLE VII
MODEL PERFORMANCE USING PSEUDO-WEATHER FORECAST INPUTS

| Model	Training set MAPE	Training set Sample error variance	Validation set MAPE	Validation set Sample error variance	Novelty set MAPE	Novelty set Sample error variance
NLAR	2.33	3488	2.16	4895	2.27	6311
TM	2.75	4955	2.39	6249	2.46	7629
WM	2.46	3832	2.26	5292	2.43	7190
NLM	2.85	5555	2.43	6657	2.39	7480
Fusion Models						
Case II	2.24	3252	2.06	4590	2.15	5740
Case III	2.57	4336	2.27	5616	2.32	6835
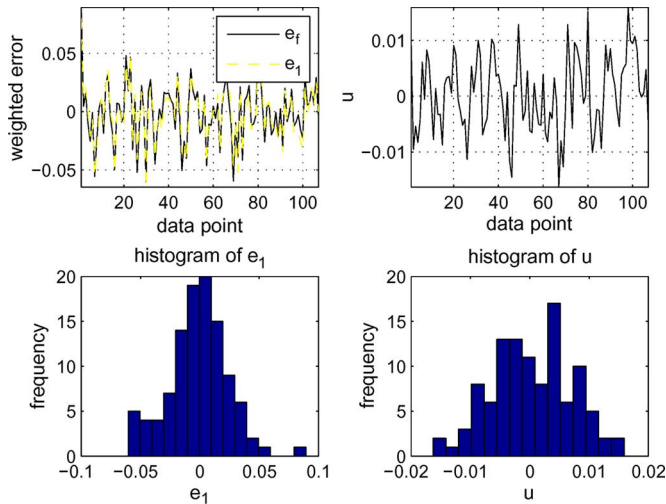
Fig. 9. Plots and histograms of the forecast errors and their differences (notes: novelty set, pseudo-weather forecasts used).

II), compared to 2.27% for the NLAR model, which is now the best individual model. It is interesting to note that had weather forecast errors not been taken into account prior to combining the sub-models, the fused results would have been significantly worse (Table VII, Case III).

Comparing Tables VI and VII, it can be seen that the NLAR models are unaffected by weather forecast errors as they have

no weather inputs. The other sub-models deteriorate with the inclusion of pseudo-weather forecast errors. The fusion model deteriorates with the inclusion of pseudo-weather forecast error but maintains its position as the best model. Next the question must be asked if the difference between the performance of the fusion model and the other models is actually *significant* or due to chance. For this purpose the errors from the NLAR sub-model (the best sub-model) are compared to those from the fusion model.[2] In general, under the assumptions that the forecast errors of two estimators, $e_1(j)$ and $e_2(j)$, are cross-correlated, zero mean and possess constant variance, σ_1^2 and σ_2^2, respectively; a test statistic may be constructed based on the difference, $u(j)$ and sum, $v(j)$ of their errors [28]:

$$u(j) = e_1(j) - e_2(j) \quad (13)$$

and

$$v(j) = e_1(j) + e_2(j) \quad (14)$$

where $u(j)$ and $v(j)$ are observations of the random variables U and V, respectively. As $\text{cov}(U,V) = \sigma_1^2 - \sigma_2^2$ (see [28] for more details) and we wish to show that $\sigma_1^2 - \sigma_2^2 \neq 0$ a null hypothesis may be constructed as

$$H0 : \text{cov}(U,V) = 0. \quad (15)$$

This may be tested [28] using the test statistic:

$$\frac{S_{UV}}{\left(\sum_{j=1}^n u^2(j)v^2(j)/n^2\right)^{1/2}} \overset{asy.}{\sim} N(0,1) \quad (16)$$

where S_{UV} is the sample cross covariance between U and V, and n is the number of samples used.[3] Fig. 9 (panel 1) below shows an example plot of the forecasting errors for the NLAR model, $e_1(j, 12)$, and the fusion model, $e_f(j, 12)$ (note: this is for the mid-day series, $k = 12$). As can be seen there is a high degree of cross-correlation between the forecast errors. Panels 2 and 4 show the histogram of $e_1(j, 12)$ and $u(j, 12)$ which appears to show that they are drawn from a normal distribution. Panel 3 shows a plot of $u(j, 12)$ for completeness. The corresponding plots for $e_f(j, 12)$ and $v(j, 12)$ are similar.

Table VIII gives a summary for the statistics used in ensuring that the assumptions required for (16) hold (as an example the mid-day time series is used). The t-test is used to check that the residuals are zero mean which is confirmed in all cases. The Ljung-Box test is used to test if the residuals are random. It was found that there does exist some serial correlation in the residuals, however this is not evident until later lags. The Jarque-Bera

---

[2]The errors from the fusion and NLAR models are correlated and so the standard Theil test is not appropriate.

[3]Note that in Section IV-B it was assumed that the variance of the errors is proportional to the value of the time series. As the test statistic in (16) is based on the assumption of constant variance, the forecast errors (from the NLAR model and fusion model) are first scaled as in (4) prior to constructing the test statistic in (16)

TABLE VIII
SUMMARY OF FORECAST ERROR STATISTICS ($k = 12$)

Sample	Test	Hypothesis	Significance	Power	Result (0 accept, 1-reject).
$e_1(j)$	t-test	H0:mean=0 H1:mean$\neq$0	5%	0.86	0
$e_f(j)$	t-test	H0:mean=0 H1:mean$\neq$0	5%	0.90	0
$e_1(j)$	Ljung -Box	Series is random	5% (lag1) 5% ... 5% ... 5% ... 5% (lag 5)	0.36 0.61 0.8 0.07 0.02	0 0 0 0 1
$e_f(j)$	Ljung -Box	Series is random	5% (lag1) 5% ... 5% ... 5% ... 5% (lag 5)	0.72 0.87 0.93 0.06 0.04	0 0 0 0 1
$e_1(j)$	Jarque- Bera	Normality	5%	0.0023	1
$e_f(j)$	Jarque- Bera	Normality	5%	0.0002	1
$e_1(j)$, $e_f(j)$	Eqn(15)	H0:cov(U, V) = 0	5%	0.9973	0

Fig. 11. Effect of training data record length on prediction accuracy (start year of dataset shown).
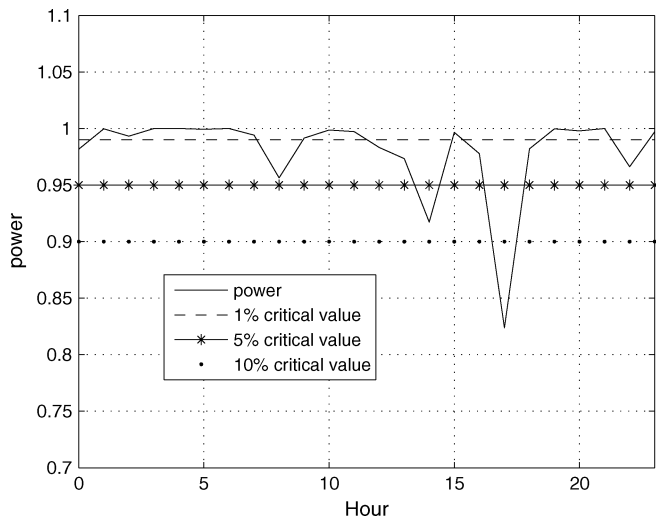

Fig. 10. P-values for each hour of the day (notes: novelty set, pseudo-weather forecasts used).

test is used to test for normality. It is found that the hypothesis of normality is rejected. On further examination this is due to several outliers on the right tail of the distribution. These are caused by the large error which occurs between the transitions from year to year in the late winter working day day-type. Given this limitation the hypothesis (15) is tested.

Fig. 10 shows the p-value for the testing the hypothesis that the variance of the residuals from the two models are statistically different. As there are 24 h, 24 tests are conducted. The results show that the hypothesis is accepted at the 1% confidence level for most of the hours, at the 5% confidence level for all but one of the hours where the p-value is 0.83. Thus empirical evidence would seem to show that the fusion model is indeed a better model than the NLAR model.

A final comment relates to the length of the training set used. There is a trade off between more data (which reduces parameter estimate variance) and irrelevant (older) data which
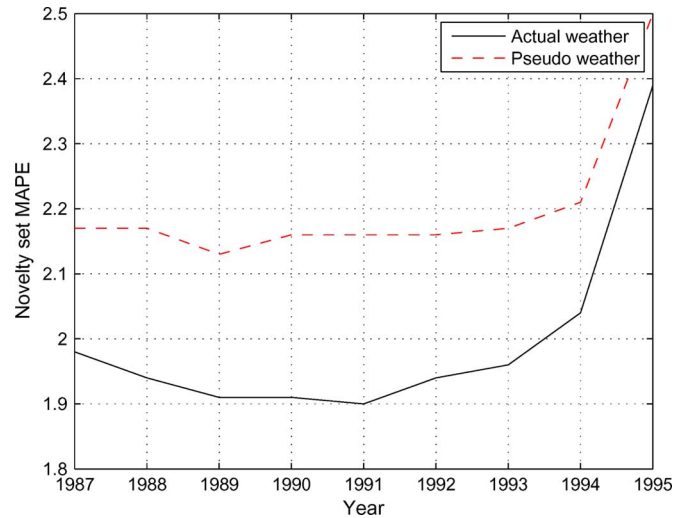
increases the variance. In order to examine the effect of data record length on the result, the fusion model and sub-models were evaluated for training data length of two years (from 1995) to ten years (from 1987). The MAPE achieved in the novelty set for record lengths of two to ten years is shown in Fig. 11, which shows that there is a marked decrease in model performance as the training data record length reduces to three years. There is relatively little variation in MAPE for record lengths greater than four years using pseudo-weather inputs, with some minor local minima in MAPE at six and eight years, using actual and pseudo-weather inputs, respectively. Given the results in Fig. 11, a training data record of seven years, beginning in 1990, was employed, which was seen as a reasonable compromise between actual and pseudo-weather MAPE indicators.

VI. CONCLUSION

This paper examined the effect of weather forecast errors in load forecasting models. In Section III, the distribution of the weather forecast errors was examined and it was found that a Gaussian distribution was not appropriate in this case. Rather, a structure exists which means that the weather forecast error will have a large effect on any aggregate weather variables. The structure of the weather forecast errors was then used to produce pseudo-weather forecast errors from 1986 to 2000 which have the accuracy of current weather forecasts. This is important as, for example, weather forecasts from 1986 are less accurate than current weather forecasts and thus of no relevance in predicting future loads.

It was argued that splitting parameter estimation into two phases; one weather forecast error dependent and the other independent was appropriate and advantageous. A model fusion technique was employed for this task. In general weather forecast error causes approximately 1% deterioration in load forecasts of all models used here. This figure, though important, is not as high as suggested by [5] and [6], for their systems. However, the fusion model was capable of adjusting the weighting of the sub-models to reflect that the weather based sub-models deteriorated relative to the AR model. Finally, the fusion model

was shown to successfully separate the tasks of model training and rejecting weather forecast errors.

ACKNOWLEDGMENT

REFERENCES

[1] H. Hippert, C. Pedreira, and R. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, Feb. 2001.

[2] P. A. Mastorocostas, J. B. Theocharis, S. J. Kiartzis, and A. G. Bakirtzis, "A hybrid fuzzy modeling method for short-term load forecasting," *Math. Comput. Sim.*, vol. 51, no. 3–4, pp. 221–232, 2000.

[3] H. Chen, Y. Du, and J. Jiang, "Weather sensitive short-term load forecasting using knowledge-based ARX models," in *Proc. IEEE Power Eng. Soc. General Meeting*, Jun. 2005, vol. 1, pp. 190–196.

[4] T. J. Teisberg, R. F. Weiher, and A. Khotanzad, "The economic value of temperature forecasts in electricity generation," *Bull. Amer. Meteorol. Soc.*, vol. 86, no. 12, pp. 1765–1771, 2005.

[5] A. Douglas, A. Breipohl, F. Lee, and R. Adapa, "The impacts of temperature forecast uncertainty on Bayesian load forecasting," *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp. 1507–1513, Nov. 1998.

[6] Taskforce, "Problems associated with unit commitment in uncertainty," *IEEE Trans. Power App. Syst.*, vol. PAS-104, no. 8, pp. 2072–2078, Aug. 1985.

[7] A. Papalexopoulos and T. Hesterberg, "A regression-based approach to short-term system load forecasting," *IEEE Trans. Power Syst.*, vol. 5, no. 4, pp. 1535–1547, Nov. 1990.

[8] S. Jo and S. W. Kim, "Consistent normalized least mean square filtering with noisy data matrix," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2112–2123, Jun. 2005.

[9] H. Yoo and R. Pimmel, "Short term load forecasting using a self-supervised adaptive neural network," *IEEE Trans. Power Syst.*, vol. 14, no. 2, pp. 779–784, May 1999.

[10] T. Miyake, J. Murata, and K. Hirasawa, "One-day-through seven-day-ahead electrical load forecasting in consideration of uncertainties of weather information," *Elect. Eng. Jpn.*, vol. 115, no. 8, pp. 22–32, 1995.

[11] K. Methaprayoon, W. Lee, S. Rasmiddatta, J. Liao, and R. Ross, "Multi-stage artificial neural network short-term load forecasting engine with front-end weather forecast," in *Proc. IEEE Ind. and Comm. Power Syst. Tech. Conf.*, 2006, pp. 1–7, 0-0.

[12] J. M. Bates and C. W. J. Granger, The Combination of Forecasts, pp. 391–410, 2001.

[13] H. Zou and Y. Yang, "Combining time series models for forecasting," *Int. J. Forecast.*, vol. 20, pp. 69–84, 2004.

[14] C. Kang, X. Cheng, Q. Xia, Y. Huang, and F. Gao, "Novel approach considering load-relative factors in short-term load forecasting," *Elect. Power Syst. Res.*, vol. 70, pp. 99–107, 2004.

[15] S. Fan, L. Chen, and W. Lee, "Short-term load forecasting using comprehensive combination based on multi- meteorological information," in *Proc. Ind. and Comm. Power Syst. Tech. Conf.*, May 2008, pp. 1–7.

[16] H. Yuansheng and Z. Huijuan, "Grey correlation in the combined weights of power load forecasting application," in *Proc. Int. Conf. Inf. Manage., Innov. Manage. and Ind. Eng.*, Dec. 2008, vol. 1, pp. 292–295.

[17] A. Palit and D. Popovic, "Nonlinear combination of forecasts using artificial neural network, fuzzy logic and neuro-fuzzy approaches," in *Proc. 9th Int. Conf. Fuzzy Syst.*, 2000, vol. 2, pp. 566–571.

[18] R. Salgado, J. Periera, T. Oshishi, R. Ballini, C. Lima, and F. Von Zuben, "Nonlinear combination of forecasts using artificial neural network, fuzzy logic and neuro-fuzzy approaches," in *Proc. Int. Joint Conf. Neural Networks*, Jul. 2006, pp. 2627–2634.

[19] M. Hibon and T. Evgeniou, "To combine or not to combine: Selecting among forecasts and their combinations," *Int. J. Forecast.*, vol. 21, pp. 15–24, 2005.

[20] S. Parkpoom, G. Harrison, and J. Bialek, "Climate change impacts on electricity demand," in *Proc. 39th Int. Univ. Power Eng. Conf.*, Sep. 2004, vol. 3, pp. 1342–1346.

[21] D. Park, O. Mohammed, A. Azeem, R. Merchant, T. Dinh, C. Tong, J. Farah, and C. Drake, "Load curve shaping using neural networks," in *Proc. 2nd Int. Forum App. of Neural Networks to Power Syst.*, 1993, pp. 290–295.

[22] S.-T. Chen, D. Yu, and A. Moghaddamjo, "Weather sensitive short-term load forecasting using nonfully connected artificial neural network," *IEEE Trans. Power Syst.*, vol. 7, no. 3, pp. 1098–1105, Aug. 1992.

[23] A. Gelb, *Applied Optimal Estimation*. Cambridge, MA: MIT Press, 1974.

[24] D. Fay, J. V. Ringwood, M. Condon, and M. Kelly, "24-h electrical load data—A sequential or partitioned time series?," *Neurocomputing*, vol. 55, no. 3–4, pp. 469–498, 2003.

[25] A. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, 4th ed. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[26] J. Hamilton, *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press, 1994.

[27] H. McCabe, "Minimum trace fusion of n sensors with arbitrary correlated sensor to sensor errors," in *Proc. IFAC Conf. Distrib. Intelligent Syst.*, Aug. 1991, pp. 229–234.

[28] B. Mizrach, Forecast Comparison in l2, Dept. Working Paper 199524, Dept. Econ., Rutgers Univ., New Brunswick, NJ, 1996.

Damien Fay received the B.Eng. degree from University College Dublin, Dublin, Ireland, in 1995 and the M.Eng. and Ph.D. degrees from Dublin City University in 1997 and 2003, respectively.

He worked as a Mathematics Lecturer at the National University of Ireland (2003–2007) before joining the NETOS group, Computer Laboratory, University of Cambridge, Cambridge, U.K., in 2007 as a Research Associate. His research interests include network analysis and time series analysis.

John V. Ringwood (SM'97) received the Diploma in electrical engineering from Dublin Institute of Technology, Dublin, Ireland, in 1981 and the Ph.D. degree from Strathclyde University, Glasgow, U.K., in 1985.

He is currently Professor of electronic engineering with the National University of Ireland (NUI), Maynooth, and Associate Dean of the Faculty of Science and Engineering. His research interests include time series modeling, control of wave energy systems and plasma processes, and biomedical engineering.