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

D. Fay and J.V. Ringwood

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 been widely examined in the literature. The main aim of this paper is to present a novel technique for minimizing this degradation. In addition, a supplementary technique is proposed to model weather forecast errors to reflect current accuracy.**

The proposed technique combines the forecasts of several load forecasting models. This approach allows the parameters of the load forecasting models to be estimated using actual weather, thus avoiding introducing noise (i.e. weather forecast error) into the training input set. The effect of the weather forecast error is then minimised during the combination stage.

Index Terms-- **Load forecasting, weather forecast errors, model combination, data fusion.**

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 which in which load forecasts play a central role.

As illustrated above, STLF is an important area and this is reflected in the literature by the many techniques that have been applied, including neural networks [Hippert 1], fuzzy logic [Mastorocostas 2] and statistical techniques [H. Chen 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 [tamimi 4, chen3]. However, in order to use weather information for future load forecasts, weather forecasts must be utilised and these have associated weather forecast errors. Although system dependent, weather forecast errors can be significant [5] and have been attributed as the cause of 17% [6] to 60% [7] of load forecast errors.

Load forecasting models are usually trained using actual past weather readings as opposed to past weather forecasts [8]. This is based on the assumption that to use the latter essentially adds forecast noise to the training data. 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 [8]. 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. [6] 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].

This paper proposes combining several models (called sub-models), or *model fusion*, as a technique for minimising the effect of weather forecast errors in load forecasting models. The concept of model fusion is well known in the general field of forecasting and was pioneered mainly in [12]. Fused forecasts are theoretically more accurate than any of the individual model forecasts [13,14] as different models are often better at modelling different aspects of an underlying process and thus combining the models appropriately gives 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 [13]. 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. 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.

The authors wish to thank Eirgrid, the Irish national grid operator, for their assistance in this research.

Dr. Damien Fay is with the department of Mathematics, NUI, Galway, Ireland, (Damien.fay@nuigalway.ie). Professor John Ringwood is with the department of Electronic Engineering, NUI, Maynooth, Ireland. (john.ringwood@eeng.nuim.ie).

II. Data set details

The range and time-scale of the available electrical demand data is given in Table 1.

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)

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 2). The readings and forecasts are for dry bulb temperature, cloud cover, wind speed and wind direction.

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

The data is subdivided into three sets in order to train and test the load forecasting models (Table 3). The training set is used to calculate model parameters, the validation set is used to aid in model structure determination and the novelty set is used to evaluate model performance.

TABLE III
DIVISION OF DATA SET

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

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.

III. Modeling weather forecast errors.

Due to the sparseness of weather forecast data (Table 2) 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, this may be a good idea. This is because the quality of weather forecasts is changing over time due to improved forecasting techniques and climate change [15]. Previous approaches in STLF have modelled the weather forecast error simply as a Gaussian random variable [16, 17]. However, as seen in Figure 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. Typically some form of aggregate weather variables are normally used in STLF models (e.g. average daily temperature). The error in an aggregate weather variable will have a non-zero mean (Figure 1) and a Gaussian approximation would underestimate this.

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 (Figure 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* (Figure 1) which represent the arrival of a weather front,
- *A level error*, $\tilde{\mu}$, which is the average of the weather forecast error between turning points,
- *A shape error*, $\tilde{\sigma}$, which is the standard deviation of the weather forecast error between turning points, and
- *A random error*, which accounts for the remaining error if $\tilde{\mu}$ and $\tilde{\sigma}$ are removed.

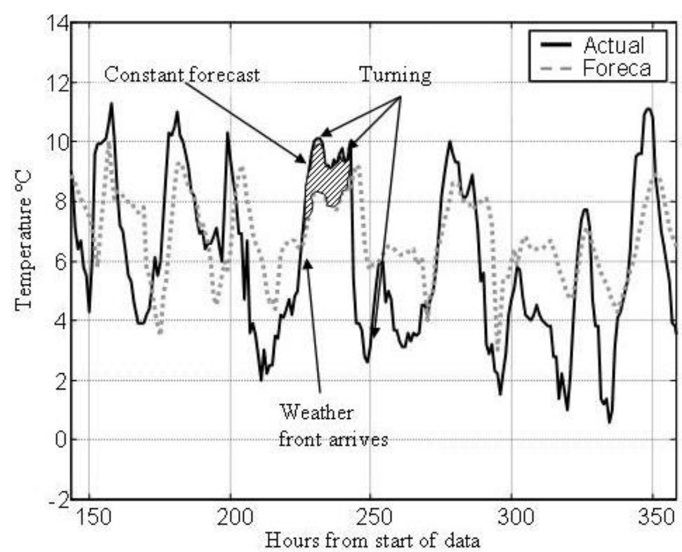

Fig. 1. Actual and forecast temperature (6th to 15th February 2000).

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

$$X(k) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} X(k-1) + \varepsilon(k) \qquad (1)$$

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

$$y(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} X(k) + v(k) \quad (2)$$

where $y(k)$ is the filtered weather variable and $v(k)$ 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 [18]). The turning points are then defined as the maxima and minima within a rolling window of length 5:

$$\vec{y} = \left\{ y(k) / y(k) <> .5 \sum_{i=0}^{11} \frac{y(k-5+i)}{10} \quad , i \neq 5, \forall k \right\} \quad (3)$$

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

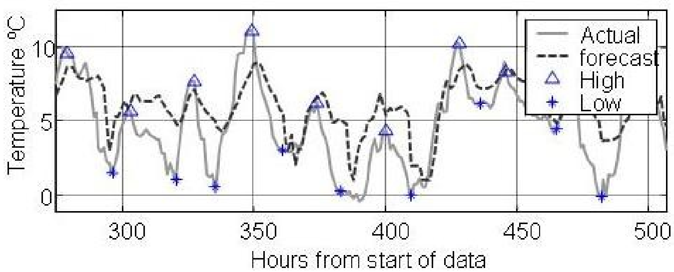

Fig. 2. A sample of the turning points calculated for temperature.

Fig. 3 below shows the histograms, fitted Gaussian distributions and the Sample AutoCorrelation Function (SACF) for the level shape and random error of the temperature forecasts.

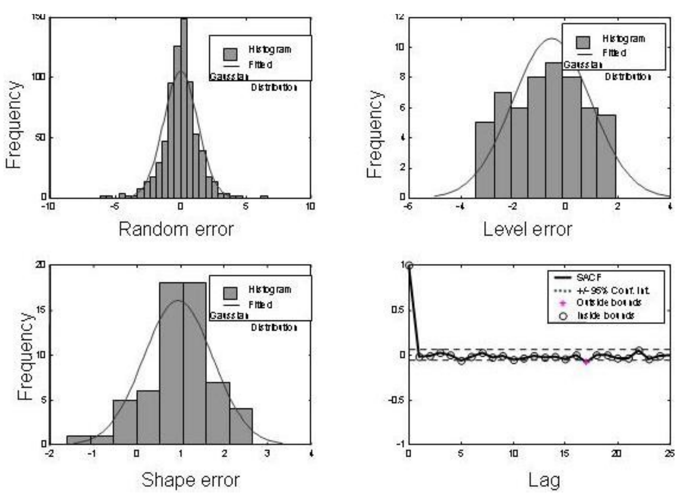

Fig. 3. Distributions and SACF for temperature forecasts.

The shape and level errors of the four weather variables are found to be cross correlated, suggesting that they may be jointly distributed. 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 the weather variables jointly. Fig. 4, below, 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.

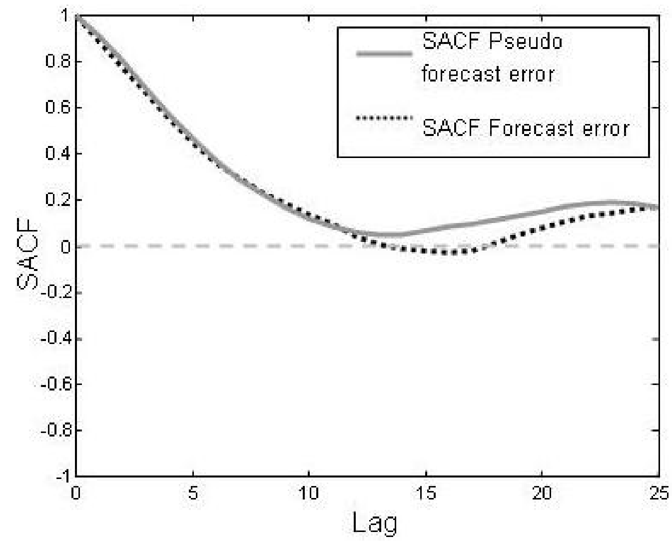

Fig. 4. SACF of forecast and pseudo-forecast temperature errors.

IV. The fusion model

A. Preliminary Auto-Regressive (AR) linear model.

It was previously found by these authors [19] that decomposing load data into 24 parallel series, one for each hour of the day, is advantageous as the parallel series have a degree of independence. 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) leaving a residual, *x*(*j*,*k*), (Figure 5) which is composed of weather, non-linear auto-regressive and white noise components [19].

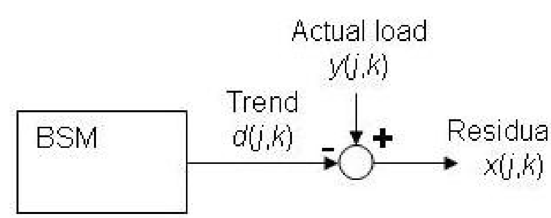

Fig. 5. Preliminary AR linear model overview.

B. Sub-Models.

Three models were chosen which have different types of inputs. These are chosen so that forecast errors can be attributed to particular inputs. A fourth model is included using all the available inputs to capture any non-linear relationships between the inputs and the residual. The models are named after their input types as shown in fig. 4. The fusion technique combines the forecasts of the sub-models, $\hat{x}_1(j,k),...,\hat{x}_4(j,k)$, to give a fused forecast, $\hat{x}_f(j,k)$ of the residual for series *j* on day *k* (Fig. 6).

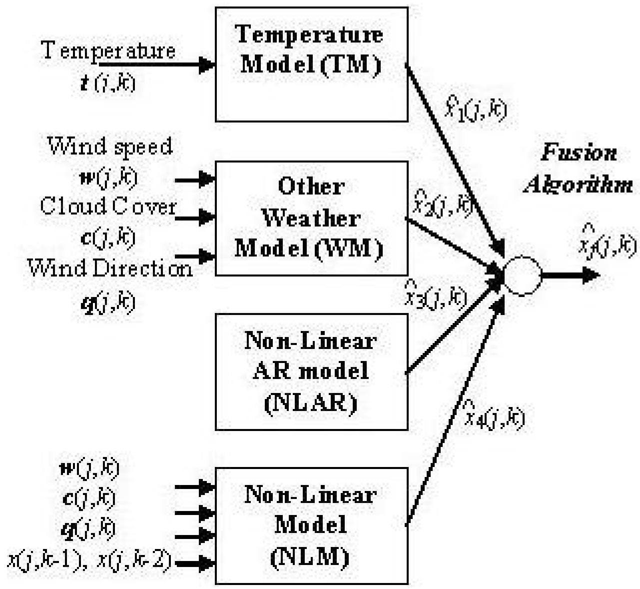

Fig. 6. Data fusion model overview.

The sub-models all use feed forward neural networks, although it should be noted that the choice of modelling 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 non-constant variance. The most likely cause for the non-constant 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 [20]. During training with the back-propagation algorithm the target errors are thus scaled prior to being propagated backwards as:

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

where $\boldsymbol{e}$ is vector of target errors and $\boldsymbol{e}'$ is the adjusted vector.

It was found that this improved the prediction performance of the models in all cases.

The Temperature Model (TM) input, $\boldsymbol{t}(j,k)$, is a vector of the previous 72 hours of temperature from hour j on day k. Similarly the other Weather Model (WM) uses vectors of wind speed, $\boldsymbol{w}(j,k)$, cloud cover, $\boldsymbol{c}(j,k)$, and wind direction, $\boldsymbol{q}(j,k)$ for the previous 72 hours of weather. The Non-Linear Auto-Regressive model (NLAR) uses the previous 2 days of residual, $x(j,k\text{-}1)$ and $x(j,k\text{-}2)$. The Non-Linear Model (NLM) uses all the available inputs.

C. Fusion Algorithm.

The data fusion algorithm described in [21] seeks to minimize the variance of the fused forecast based on the covariance matrix of the sub-model forecasts. 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 $\hat{x}_1(j,k),\ldots,\hat{x}_4(j,k)$ [21]:

$$\hat{x}_f(j,k) = \sum_{i=1}^{4} A_i(j)\hat{x}_i(j,k) \qquad (5)$$

where $A_i(j)$ is the weight applied to the forecast from sub-model i for hour j[2], and is derived from the error covariance matrices of $\hat{x}_1(j,k),\ldots,\hat{x}_4(j,k)$ as:

$$[A_1(j)A_2(j)A_3(j)] = \left[P'_{4,1}(j)P'_{4,2}(j)P'_{4,3}(j)\right]\boldsymbol{P}^{-1} \qquad (6)$$

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

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

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:

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

and

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

where

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

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

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

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

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

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

V. Results.

The overall approach suggested here is that the sub-model parameters are estimated using *actual weather* inputs (thus estimating the sub-model parameters without pseudo-weather forecast errors). The error covariance matrices of the sub-models (10) are then estimated using *pseudo-weather forecasts* as input. The weights, $A_i(j)$, are then estimated using these error covariance matrices. However, the results are here analysed for two cases. The first examines the behaviour of the fusion model without pseudo-weather forecasts and the second examines the behaviour with them:

Case I: The sub-model parameters are estimated using *actual weather* inputs (thus estimating the sub-model parameters without pseudo-weather forecast errors). The error covariance matrices of the sub-models (10) are then estimated using *actual weather* inputs. The weights, $A_i(j)$, are then calculated using these error covariance matrices (as in Section IV part 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 (10) 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 part C). Models are trained and evaluated using pseudo weather forecast inputs.

As an example, the cross-covariance matrix of sub-model forecast errors is shown in Table IV below for the midday series (j=12). The difference between case 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 indicates the degradation of the models due to (pseudo) weather forecast error.

TABLE IV

The cross-covariance matrix of sub-model load forecast errors (Case I⟶Case II)

Sub-model	1	2	3	4
NLAR	**4464** → 4464	**2905** → 3759	**4112** → 4404	**2978** → 3836
TM	**2905** → 3759	**2862** → 4882	**2620** → 3788	**2505** → 3892
WTM	**4112** → 4404	**2620** → 3788	**4155** → 4718	**2939** → 4158
NLM	**2978** → 3836	**2505** → 3892	**2939** → 4158	**2953** → 4537

The corresponding values of $A_1,\ldots A_4$ are shown in Table V below. As can be seen the weights are approximately equal showing that each model has similar forecast accuracy. Note however, how the weights change significantly once pseudo-weather forecasts are introduced to the models.

TABLE V

An example of fusion weights (actual weather inputs)

Sub-model	NLAR	TM	WM	NLM
Case I	-0.59	0.75	0.48	0.35
Case II	0.69	0.21	-0.35	0.45

Fig. 7 below 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.

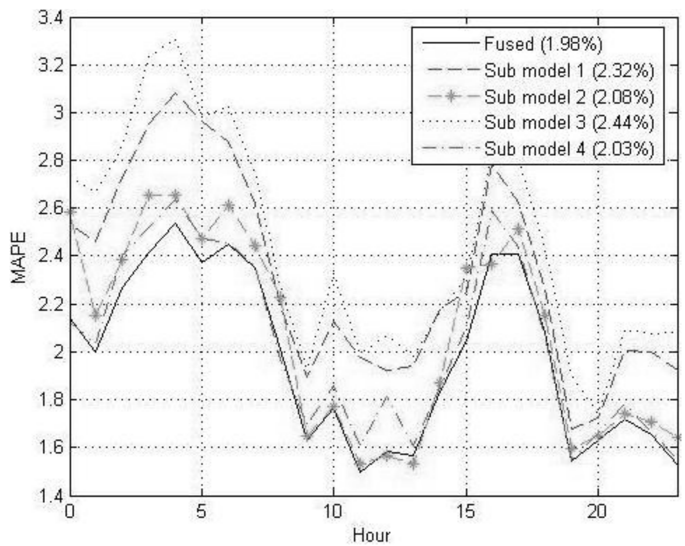

Fig. 7. MAPE as a function of hour of the day for fusion and sub-models (notes: novelty set, actual weather used).

Table VI below, summarise's the results in the training, validation and novelty data sets.

TABLE VI

model performance using actual weather inputs.

Model	Training Set		Validation Set		Novelty Set	
Model	MAPE	Sample error variance	MAPE	Sample error variance	MAPE	Sample error variance
NLAR	2.30	3455	2.17	4911	2.32	6513
TM	2.08	2773	1.95	4084	2.08	5255
WM	2.35	3537	2.21	5045	2.44	7152
NLM	2.06	2782	1.92	3971	2.03	5167
Fusion	**1.98**	**2534**	**1.86**	**3764**	**1.98**	**4837**

Fig. 8 below shows the Mean Absolute Percentage Error (MAPE) for the sub-models and the fusion model using pseudo-forecast weather inputs in the novelty set. The effect of weather forecast errors are now accounted for by calculating the error covariance matrices of the sub-models over the training set *with pseudo-weather forecast inputs*. As can be seen the fusion model again performs best for each hour of the day.

* The MAPE is the standard error measure in the field of STLF as it allows comparison between systems.

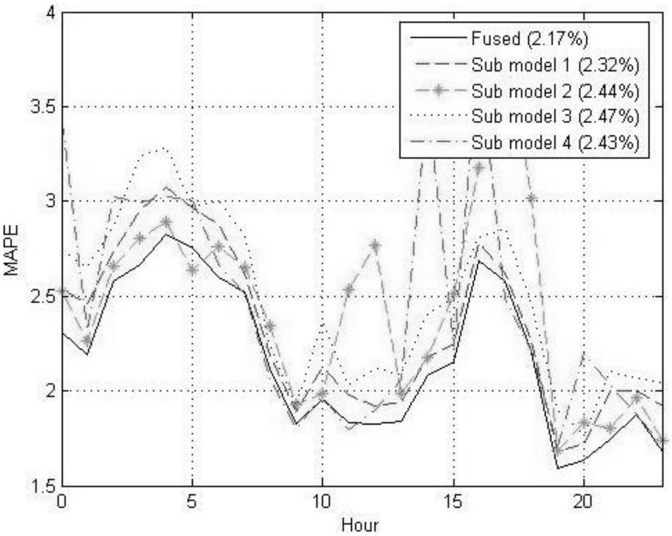

Fig. 8. MAPE as a function of hour of the day for fusion and sub-models (notes: novelty set, pseudo-weather forecasts used).

TABLE VII
SUMMARY OF THE MAPE'S OF THE MODELS USING PSEUDO WEATHER FORECAST INPUTS.

Model	**Training Set**		**Validation Set**		**Novelty Set**	
Model	MAPE	Sample error variance	MAPE	Sample error variance	MAPE	Sample error variance
NLAR	2.30	3455	2.17	4911	2.32	6513
TM	2.53	4431	2.29	6128	2.44	7606
WM	2.43	3749	2.25	5292	2.47	7321
NLM	2.73	5261	2.37	6325	2.43	7583
Fusion	**2.20**	**3167**	**2.06**	**4524**	**2.17**	**5759**

Comparing Tables IV and V, 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. First it should be noted that the errors from the fusion model are correlated to those from the NLAR sub-model and so the assumptions underlying the standard Theil test are violated.

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 resp.; a test statistic may be constructed based on the difference, $u(j)$ and sum, $v(j)$ of their errors [22]:

$$u(j) = e_1(j) - e_2(j) \tag{13}$$

and

$$v(j) = e_1(j) + e_2(j) \tag{14}$$

where $u(j)$ and $v(j)$ are observations of the random variables U and V respectively. As $\mathrm{cov}(U,V) = \sigma_1^2 - \sigma_2^2$ (see [22] for more details) and we wish to show that $\sigma_1^2 - \sigma_2^2 \neq 0$ a null hypothesis may be constructed as:

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

This may be tested [22] using the test statistic:

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

where S_{UV} is the sample cross covariance between U and n is the number of samples used.

Note that in Section IV part B it was assumed that the variance of the errors is proportional to 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).

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. Panel 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.

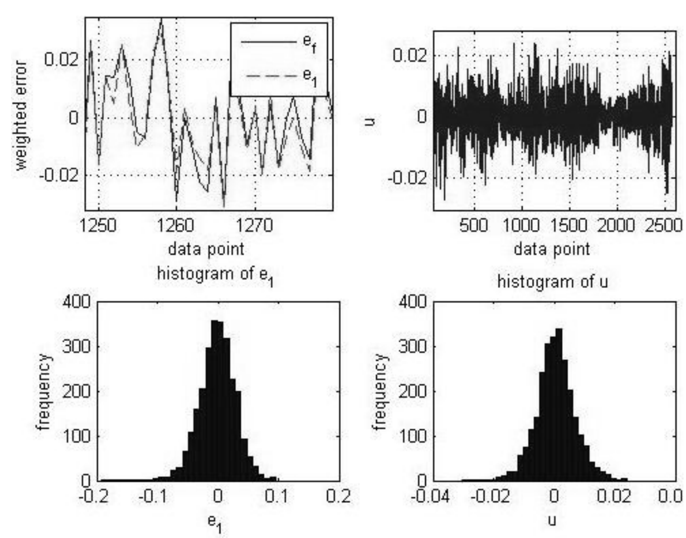

Fig. 9. Plots and histograms of the forecast errors and their differences. (notes: novelty set, pseudo-weather forecasts used).

Table VIII below 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 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.

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

Sample(s)	Test	Hypothesis	Significance	Power	Result (0 –accept, 1-reject).
$e_1(j)$	t-test	H0:mean=0 H1:mean≠0	5%	0.86	0
$e_f(j)$	t-test	H0:mean=0 H1:mean≠0	5%	0.90	0
$e_1(j)$	Ljung-Box test	Series is random	5% (lag 1) 5% … 5% … 5% … 5% (lag5)	0.4 0.6 0.7 0.07 0.01	0 0 0 0 1
$e_f(j)$	Ljung-Box test	Series is random	5% (lag 1) 5% … 5% … 5% … 5% (lag5)	0.72 0.86 0.92 0.05 0.03	0 0 0 0 1
$e_1(j)$	Jarque-Bera test	H0: $e_1(j)$~N H1: $e_1(j)$ skewed	5%	0.0023	1
$e_2(j)$	Jarque-Bera test	H0: $e_1(j)$~N H1: $e_1(j)$ skewed	5%	0.0002	1
$e_1(j)$, $e_2(j)$	Eqn. (15)	H0: cov(U,V) = 0	5%	0.9972	0

Fig. 10 below, 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 hours, 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.

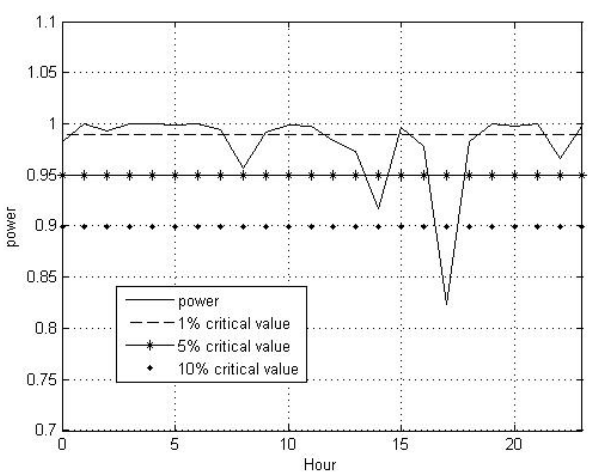

Fig. 10. P-values for each hour of the day. (notes: novelty set, pseudo-weather forecasts used).

VI. CONCLUSION

This paper examined the effect of weather forecast errors in load forecasting models. In Section 3, 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.

A model fusion technique was then proposed for minimising the effect of weather forecast errors. 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 [6] and [7], 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.

VII. REFERENCES

[1] S.H. Hippert, C.E. Pedriera, R.C. Souza, "Neural networks for short-term load forecasting: a review and evaluation", *IEEE Transactions on Power Systems,* 16 (1), pp. 44-55, 2001.

[2] P.A. Mastorocostas, J.B. Theocharis, S.J. Kiartzis, A.G. Bakisrtzis, "A hybrid fuzzy modeling method for short-term load forecasting", *Mathematics and Computers in Simulation*, 51, pp. 221-232, 2000.

[3] H. Chen, Y. Du, J.N. Jiang, "Weather sensitive short-term load forecasting using knowledge-based ARX models", *IEEE Power Engineering Society General Meeting*, Vol 1, pp 190-196, 12-16 June 2005.

[4] M. Tamimi, R. Egbert, "Short term electric load forecasting via fuzzy neural collaboration", *Electric Power Systems Research*, 56, pp. 243-248, 2000.

[5] T. J., Teisberg, , R. F. Weiher, and A. Khotanzad, "The value of national weather service forecasts in scheduling electricity generation.", *Bull. Amer. Meteor. Soc.*, in press, 2005.

[6] A.P. Douglas, A.M. Breipohl, F.N. Lee, R. Adapa, "The impacts of temperature forecast uncertainty on Bayesian load forecasting", *IEEE Transactions on Power Systems,* 13 (4), pp. 1507-1513, 1998.

[7] IEEE committee report, "Problems associated with unit commitment in uncertainty", *IEEE Transactions on Power Apparatus and Systems*, 104 (8), pp. 2072-2078. 1985.

[8] A.D. Papalexopoulos, C.T. Hesterburg, "A regression based approach to short term system load forecasting", *IEEE Transactions on Power Systems,* 5 (4), pp. 1535-1547, 1990.

[9] H. Yoo, R.L. Pimmel, "Short-term load forecasting using a self-supervised adaptive neural network", *IEEE Transactions on Power Systems,* 14 (2), pp. 779-784, 1999.

[10] T. Miyake, J. Murata, K. Hirasawa, "One-day through seven-day-ahead electrical load forecasting in consideration of uncertainties of weather information", *Electrical Engineering in Japan*, 115 (8), pp. 135-142. 1995.

[11] K. Methaprayoon, W.J. Lee, S. Rasmiddatta, J. Liao, R. Ross, "Multi-Stage Artificial Neural Network Short-term Load Forecasting Engine

with Front-End Weather Forecast", *IEEE Industrial and Commercial Power Systems Technical Conference*, pp1-7, 30-05 April 2006.
[12] J.M. Bates, C.W.J. Granger, "The combination of forecasts", *Operational Research Quarterly*, 20, pp. 451-468, 1969.
[13] A.K. Palit, D. Popovic, "Nonlinear combination of forecasts using artificial neural network, fuzzy logic and neuro-fuzzy approaches", *in: Proceedings, IEEE international conference on fuzzy systems*, 2, pp. 566-571, 2000.
[14] R.M. Salgado, J.J.F. Periera, T. Oshishi, R. Ballini, C.A.M. Lima., F.J. Von Zuben, "A hybrid ensemble model applied to the short-term load forecasting problem", *in: Proceedings, 2006 International Joint Conference on Neural Networks,* pp. 2627 - 2634 , July 16-21, 2006.
[15] S. Parkpoom, G.P. Harrison, J.W. Bialek, "Climate change impacts on electricity demand", *39th International Universities Power Engineering Conference, 2004*, Vol. 3, pp. 1342-1346, 6-8 Sept., 2004.
[16] D. Park, O. Mohammed, A. Azeem, R. Merchant, T. Dinh, "Load curve shaping using neural networks", *in: Proceedings, Second International Forum on Applications of Neural Networks to Power Systems*, *Yokohama, Japan*, April 1993, IEEE, pp. 290-295, 1993.
[17] S.T. Chen, D.C. Yu, A.R. Moghaddamjo, "Weather sensitive short-term load forecasting using non-fully connected artificial neural network", *IEEE Transactions on Power Systems,* 7 (3), pp. 1098-1104, 1992.
[18] A. Gelb, Applied Optimal Estimation, Cambridge: MIT Press, 1974, pp 156 -173.
[19] D. Fay, J.V. Ringwood, M. Condon, M. Kelly, "24-hour electrical load data – a sequential or partitioned time series?", *Neurocomputing*, 55 (3-4), pp. 469-498, 2003.
[20] J.D. Hamilton, Time Series Analysis, Princeton, NJ.: Princeton University Press, 1994, pp. 221.
[21] H. McCabe, "Minimum trace fusion of *N* sensors with arbitrary correlated sensor to sensor errors", *in: Proceedings, IFAC Conference on Distributed Intelligent Systems*, Virginia, USA, August 1991, pp. 229-234. 1991.
[22] Mizrach, B., "Forecast comparison in L2", Departmental Working Papers 199524, Rutgers University, Department of Economics. 1996. Available: ftp://snde.rutgers.edu/Rutgers/wp/1995-24.pdf.

VIII. Biographies

Damien Fay was born in 1973 in Dublin, Ireland. He received an honours degree in Electronic Engineering from the National University of Ireland, Dublin, in 1995. In 1997 he received a Masters of Engineering at Dublin City University. From 1997 to 1998 he worked for Alsthom ltd. UK, researching models for steel rolling processes and differential positioning systems for naval vessels. In 2003 he obtained a PhD in electronic engineering at Dublin City University. Currently he is a lecturer in mathematics in the National University of Ireland, Galway. His main research interests are forecasting techniques, non-linear systems modelling, data fusion techniques and wavelet packet applications in time series analysis.

John Ringwood received the Diploma in Electrical Engineering from Dublin Institute of Technology and the PhD (in Control Systems) from Strathclyde University, Scotland in 1981 and 1985 respectively. He was with the School of Electronic Engineering at Dublin City University 1985 to 2000 and during that time held visiting positions at Massey University and the University of Auckland in New Zealand. He is currently Professor of Electronic Engineering with the National University of Ireland (NUI), Maynooth and Associate Dean (for Engineering) of the Faculty of Science and Engineering. He was Head of the Electronic Engineering Department at NUI Maynooth from 2000 until 2005, developing the Department from a greenfield site. John's research interests include time series modelling, control of wave energy systems and plasma processes, and biomedical engineering. He is a Chartered Engineer and a Fellow of the Institution of Engineers of Ireland.