

A Data Fusion Model for Irish Electricity Load Forecasting

Damien Fay[†] John V. Ringwood^{††} Marissa Condon[†] Michael Kelly^{†††}

[†] Dublin City University ^{††} NUI Maynooth ^{†††} Electricity Supply Board

Abstract

Weather inputs are an important factor in load forecasting. The arrival of a weather front or introduction of weather forecast errors lead to a change in the significance of weather as input. A data fusion of individual models is introduced, which uses *a priori* information regarding the significance of weather inputs in these situations, to produce a superior forecast to any individual model.

1. INTRODUCTION

Short term forecasting is required by electricity utilities for unit commitment and allocation of spinning reserve. Errors in short term forecasts cost the Irish Electricity Supply Board a significant amount of money per annum. In accordance with the Electricity Regulation Act of 1999, a deregulated market structure was set up which should lead to increased impetus to reducing forecast error and the associated costs.

Load forecasting models are usually trained using actual past weather readings [1-2] as opposed to past weather forecasts. 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. However, weather forecast errors not present in the training set can have a disproportionate influence on the load in models as illustrated in [1-2]. Changing the load model parameters to account for this can be impossible in many conventional models once training is completed. Douglas *et al.* [4] approached this problem by use of a Bayesian framework, but restricted analysis to the use of dynamic *linear* models. As pointed out by [3], the literature is sparse with respect to load forecasting in the presence of weather forecast error.

An incoming weather front will lead to a shift in temperature and a subsequent shift in the load. In this situation, the influence of temperature on load is increased relative to other factors. Park et al [5] found that by considering these effects, improved results from their model could be obtained.

The three load forecasting methods in [3-5] have the ability to change the parameters of an individual load model given *a priori* information regarding the accuracy of the inputs. However, combining the outputs of several *different* models, or *data fusion*, is an alternative technique that has been applied to many fields [6,7,8] to solve similar problems to those faced in [3-5]. References [7,8] compare several data fusion techniques for rainfall run-off models and find that a fuzzy logic approach is the superior technique in that situation.

The focus of this paper is to present a fusion technique for dealing with *a priori* knowledge regarding weather forecast error and shifts in temperature. The data fusion algorithm of McCabe [9] was chosen as the fusion parameters can be easily changed to cater for perceived shifts in the relative significance of the various data streams used to form the composite output.

2. DATA SET DETAILS

A database of electricity demand, actual temperature, wind speed and cloud cover from 1987 to 2000 on an hourly basis is available. Data between Tuesday and Thursday in the months January to March is selected so as to avoid the exceptions associated with weekend, Christmas and changes due to the daylight saving hour.

Weather *forecast* data for temperature wind speed and cloud cover is also available for February 2000 also on an hourly basis.

Four sets of data are used to train and test the models (Table 1). The training set is used to calculate model parameters. In the case of neural network models the validation set is used for early stopping and structure determination. For the linear models the validation set is used to determine the number of inputs to the model. Since the validation set has some influence on model determination a novelty set is used to evaluate model performance using actual weather readings. Finally, the forecast set is used to evaluate model performance using weather *forecasts*.

Table 1 Segmentation of data set.

Set	Training	Validation	Novelty	Forecast
Range	20 th Jan 1987 20 th Mar 1996	21 st Mar 1996 20 th Mar 1998	27 th Mar 1998 30 th Mar 2000	3 rd Feb 2000 26 th Feb 2000
Size (Days)	320	40	60	14

3. FUSION MODEL

3.1 Preliminary AR linear model

It was previously found by these authors [10] 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 1) which is composed of weather, non-linear AR and white noise components [10].

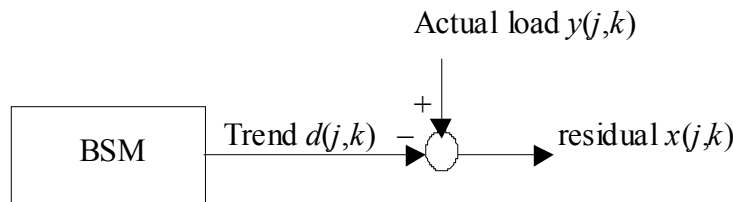
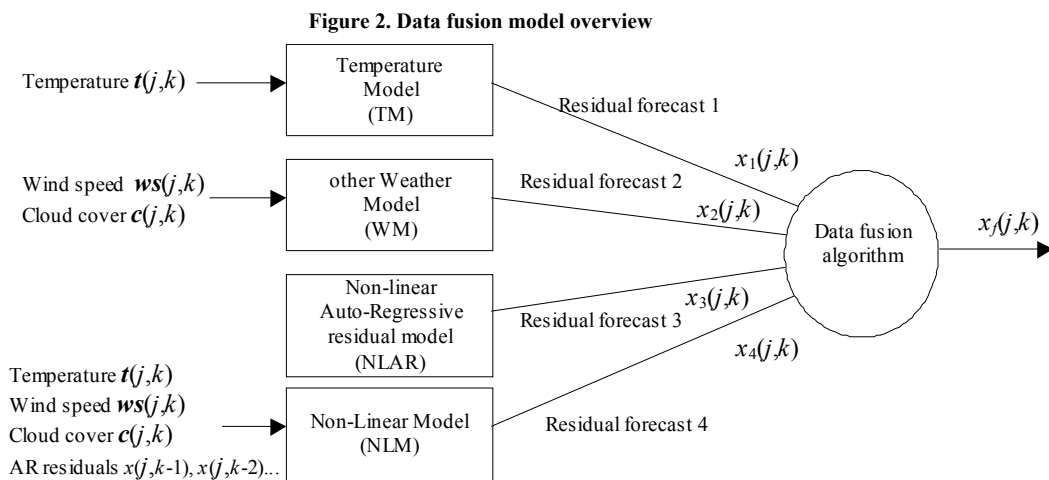


Figure 1. Preliminary AR linear model overview

3.2 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 fusion technique combines the forecasts of the sub-models $x_{1...4}(j,k)$ to give a fused forecast $x_f(j,k)$ of the residual for series j on day k (Figure 2).



The Temperature Model (TM) input, $t(j,k)$, is a vector of the previous 24 hours of temperature from hour j on day k . This vector is transformed using Principal Component Analysis (PCA), to produce a vector of principal components as in [10].

The number of components retained, N , is determined by the model performance over the validation set [10]. The residual $x(j,k)$ is then modeled using a simple linear model of the form

$$x(j,k) = a_1(j)w_1(j,k) + a_2(j)w_2(j,k) + \dots + a_N(j)w_N(j,k) + \varepsilon(j,k) \quad (1)$$

where a_i is the linear coefficient applied to PCA component w_i and $\varepsilon(j,k)$ is a white noise error term. The linear coefficients are calculated by means of least squares as in [10].

The other Weather Model (WM) inputs, $ws(j,k)$ and $c(j,k)$, are vectors containing the previous 24 hours of the wind speed and cloud cover from hour j on day k , respectively. These inputs are transformed via PCA as in the TM and the residual is modeled using a linear model similar to (1).

The Non-Linear AutoRegressive model (NLAR) uses a feed forward neural network trained using the back-propagation algorithm with early stopping using the validation set as in [10]. The network topology is determined from the performance of networks with one to five nodes in both hidden layers over the validation set (refer to [10] for more details). The optimum topology was found to be three nodes in the first hidden layer with two nodes in the second.

The Non-Linear Model (NLM) using all inputs employs a neural network trained the same way as the NLAR. The dimensionality of the weather inputs is reduced using PCA as in the TM and WM (refer to [10] for more details). The optimum topology for the NLM was found to be two nodes in the first hidden and three in the second.

3.3 Fusion algorithm

The data fusion algorithm described in [9] 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 $x_{1...4}(j,k)$ [9]:

$$x_f(j,k) = \sum_{i=1}^4 A_i(j)x_i(j,k) \quad (2)$$

where $A_i(j,k)$ is the weight applied to forecast i and is derived from the error covariance matrices of $x_{1...4}(j)$ as:

$$[A_1(j) A_2(j) A_3(j)] = [P'_{4,1}(j) P'_{4,2}(j) P'_{4,3}(j)] \begin{bmatrix} P'_{1,1}(j) & P'_{1,2}(j) & P'_{1,3}(j) \\ P'_{2,1}(j) & P'_{2,2}(j) & P'_{2,3}(j) \\ P'_{3,1}(j) & P'_{3,2}(j) & P'_{3,3}(j) \end{bmatrix}^{-1} \quad (3)$$

where

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

with

$$P_{i,n}(j) = E[(x(j) - x_i(j))(x(j) - x_n(j))] = \frac{1}{M} \sum_{k=1}^M (x(j,k) - x_i(j,k))(x(j,k) - x_n(j,k)) \quad (5)$$

where $E[\cdot]$ denotes the expectation operator and M is the number of samples. The $P'_{i,n}(j)$ are calculated from:

$$P'_{i,n}(j) = P_{i,n}(j) - P_{4,n}(j) - P_{i,4}(j) + P_{4,4}(j) \quad i \neq 4, n \neq 4 \quad (6)$$

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

$$A_4(j) = 1 - \sum_{i=1}^3 A_i(j) \quad (7)$$

Finally the fused load $y_f(i,j)$ is estimated by reintroducing the trend:

$$y_f(j,k) = d(j,k) + x_f(j,k) \quad (8)$$

4. COVARIANCE ESTIMATION

4.1 Calculating the covariance matrix

The sample covariance matrix was calculated on a combination of the training and validation sets using (5). It is found that the covariance matrix doesn't change considerably over time and so is not tracked. A normalised sample covariance matrix* is shown (for the 11pm load series as an example) in Table 2.

Table 2 Normalised sample covariance matrix for 11pm load series.

Model	TM	WM	NLAR	NLM
TM	0.6763	0.6240	0.6201	0.6605
WM	0.6240	0.9548	0.9015	0.7711
NLAR	0.6201	0.9015	1.0000	0.8398
NLM	0.6605	0.7711	0.8398	0.9397

As can be seen, the off-diagonal elements of the matrix are significant showing strong cross-correlation between sub-models.

4.2 Covariance matrix for the forecast set

When forecast data is applied to the inputs of the sub-models the covariance matrix will change as weather forecast errors affect the estimated load. The sample covariance matrix is calculated using (5) from the forecast set. A separate set for calculating the covariance matrix would be superior, but due to the small size of the forecast set this is not possible.

4.3 Covariance matrix for a shift in temperature

A shift in temperature indicates a corresponding reaction in the load. Since the NLAR and WM model have no temperature input their forecasts will be less accurate than before. Thus their expected errors will be larger, or the expected errors of the TM and NLM will be relatively smaller. To investigate if this could be taken advantage of a simple formula was applied. The shift is identified by a large difference between the TM forecast and the NLAR forecast. If the difference was above a certain threshold β then the entries in the covariance matrix relating to the TM are reduced by a factor α :

$$\begin{aligned}
 &\text{IF } x_1(j, k) - x_3(j, k) > \beta \text{ THEN} \\
 &P_{1,1}(j) = P_{1,1}(j) / \alpha^2 \\
 &P_{1,n}(j) = P_{1,n}(j) / \alpha \quad n = 2,3,4
 \end{aligned} \tag{9}$$

5. RESULTS

The value for α is tuned over the training set and the results show that as alpha is increased from 1 (i.e. no effect) to 5 the MAPE in both the training and forecast sets falls until α reaches 2.4. Adjusting the covariance matrix for a shift in temperature is shown to be consistent over both sets (Figure 3).

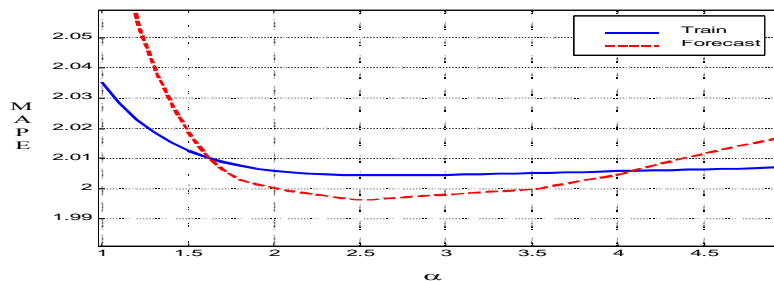


Figure 3. MAPE as a function of α

The performance of the sub-models and the fusion model (with and without adjusting the covariance matrix for a shift in temperature) are shown in Table 3. The results are compared using the standard load forecasting measure, the Mean

* Normalised for confidentiality reasons.

Absolute Percentage Error (MAPE) and the Mean Squared Error (MSE) over the training, novelty and weather forecast data sets.

Table 3 MAPE & normalised MSE*

Model	TM	WM	NLAR	NLM	Fused/ $\alpha=1$	Fused / $\alpha=2.4$
Training MAPE	1.93	2.61	2.43	2.06	1.99	1.97
Novelty MAPE	1.65	2.21	2.19	1.90	1.74	1.68
Forecast MAPE	2.10	6.28	2.03	2.27	2.10	1.98
Training MSE	.98	1.82	1.67	1.15	1	.96
Novelty MSE	1.29	1.92	2.10	1.70	1.34	1.27
Forecast MSE	2.41	30.75	2.19	2.46	2.39	1.94

It is interesting to note that the TM is superior to the other sub-models except in the weather forecast set. When weather forecasts are fed into the sub-models all degrade in performance except the NLAR which becomes the best sub-model. The NLAR model does not degrade, since it is not dependent on weather information. The fusion model achieves a comparable performance to the best sub-model in all data sets, but is seen to achieve better consistency across the different data sets.

When the value of α is set to 2.4, the fusion model performance is superior in all data sets except the in the novelty set where the TM MAPE is superior. The MSE for the fusion model is, however, superior and this is due to the fact that the fusion model is trained to minimize the MSE and not the MAPE. As can be seen in Figure 4, the performance of the fusion model for certain hours of the day is preferable to the performance of the other models and thus it may not be desirable to apply the fusion model for all parallel series.

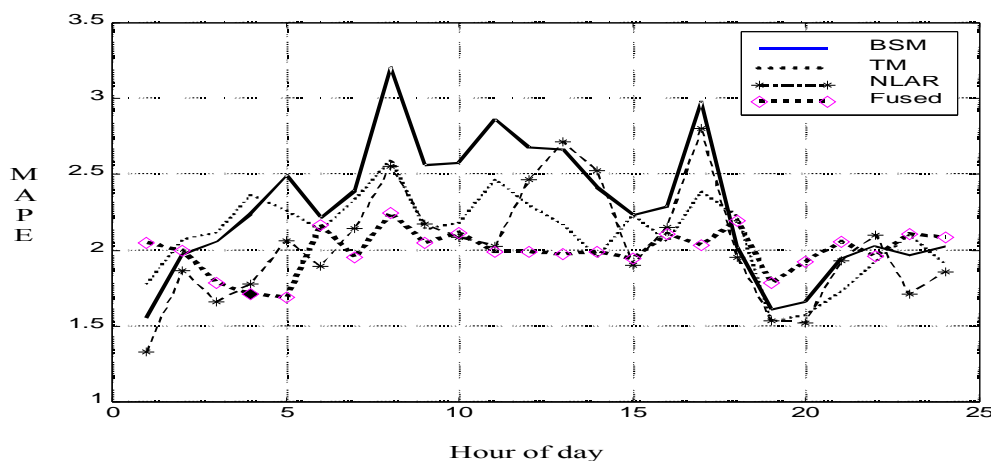


Figure 4. MAPE as a function of hour of day over the forecast set*

Figure 5 shows a sample forecast made using weather forecast data with a shift in temperature occurring at 17:00 and 18:00.

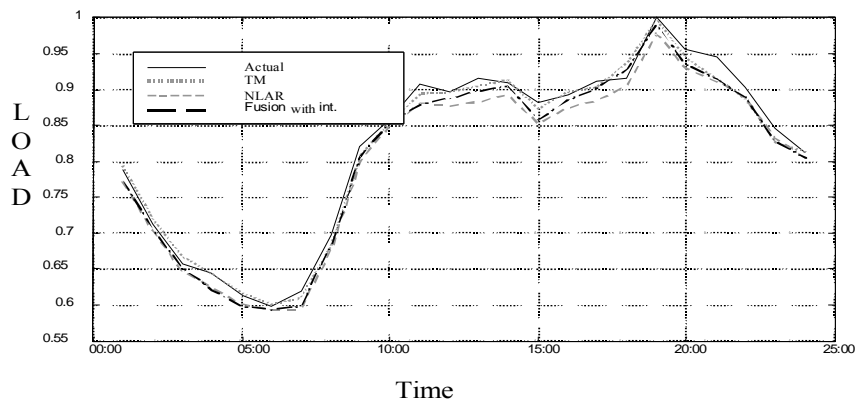


Figure 5. Sample forecast*

& The WM and NLM are excluded for clarity.

* Normalised for confidentiality.

5. CONCLUSION

A fusion model has been introduced which can adapt to *a priori* information via the covariance matrix giving superior results to any individual model. The type of sub-model used is not constrained in any way. The majority of load forecasting models use weather forecasts online. The danger of training models with inputs, some of which will later be forecast, as in the NLM has been shown. The NLM was trained in the absence of weather forecast noise. Its inability to reject this noise is shown in the forecast set. By using the TM and NLAR, the weather forecast error could be de-coupled from the AR error. Thus, the fusion technique can reduce the relative importance of the temperature inputs and maintain performance. Although the algorithm applied for a shift in temperature was possibly too simple and ignored the relative improvement of the NLM, it was found to give considerable improvement and warrants further research.

REFERENCES

1. H. Yoo, R.L. Pimmel, "Short term load forecasting using a self-supervised adaptive neural network", *IEEE Transactions on Power Systems*, Vol. 14, No. 2, May 1999, pp 779-784.
2. P.K.Dash, A.C. Liew, S.Rahman, "A real-time short-term load forecasting system using functional link network", *IEEE Transactions on Power Systems*, Vol. 12, No. 2, May 1997, pp 675-680.
3. T.Miyake, J.Murata, K.Hirasawa, "One day through seven day ahead electrical load forecasting in consideration of uncertainties of weather information", *Elec. Eng. Jpn (USA)*, Vol. 115, no. 8, 15th December 1995, pp 22-32.
4. A.P. Douglas, A.M. Breipol, F.N.Lee, R. Adapa, "The impacts of temperature forecast uncertainty on Bayesian load forecasting", *IEEE Transactions on Power Systems*, Vol. 14, No. 4, November 1998, pp 1507-1513.
5. D.Park, O.Mohammed, R.Merchant, T.Dinh, C.Tong, A.Azeem, J. Farah, C.Drake, "Forecasting abnormal load conditions with neural networks", *Proc. of the Second International Forum on Applications of Neural Networks to Power Systems*, Yokohama, Japan, 19-22 April 1993, pp 73-78.
6. M.P. Perrone, L.N. Cooper, "When networks disagree: Ensemble methods for hybrid neural networks", in *Artificial Neural Networks for Speech and Vision*, R.J. Mammone (eds), Chapman & Hall, 1993.
7. A.Y. Shamseldin, K.M. O'Connor, G.C. Liang, "Methods for combining the outputs of different rainfall-runoff models", *Journal of Hydrology*, 197, 1997, pp 203-229.
8. L.Xiong, A.Y. Shamseldin, K.M. O'Connor, "A non-linear combination of the forecasts of rainfall-runoff models by the first order Takagi-Sugeno fuzzy system", *Journal of Hydrology*, 245, 2001, pp 196-217.
9. H. McCabe, "Minimum trace fusion of N sensors with arbitrary correlated sensor to sensor errors", *Proceedings of IFAC Conference on Distributed Intelligent Systems*, Virginia, USA, August 1991, pp 229-234.
10. D.Fay, J.V.Ringwood, M.Condon, M.Kelly, "Comparison of linear and neural parallel time series models for short term load forecasting in the Republic of Ireland", *35th Universities Power Engineering Conference*, Vol 1 Belfast, UK., 6-8th September 2000, (Not paginated on CD-ROM).