Building Energy Data Analysis and Prediction with Recurrent Neural Nets

Lizhong Wu & Mahesan Niranjan

CUED/F-INFENG/TR.124

Cambridge University Engineering Department
Trumpington Street
Cambridge CB2 1PZ
United Kingdom

April 1993

Email: lzw@eng.cam.ac.uk or niranjan@eng.cam.ac.uk

Building Energy Data Analysis and Prediction with Recurrent Neural Nets

Lizhong Wu & Mahesan Niranjan
Cambridge University Engineering Department
Trumpington Street, Cambridge CB2 1PZ, U.K.

Abstract

In this report, we present analysis and prediction of building data using recurrent neural nets. We first explain why a recurrent neural net is chosen by analysing the static and dynamic characteristics of the data, and demonstrating its prediction result. Two techniques are then developed to track the non-stationary state and to catch the long-term memory structure of the data to improve the prediction performance, which cannot be attained with a large recurrent net due to its training difficulty.

1 Introduction

1.1 Task

Building data analysis and prediction is a competition organised by the ASHRAE meeting, Denver, Colorado during December 1, 1992 to April 30, 1993, (Kreider and Haberl, 1993). Two sets of energy and environmental data from real buildings are provided for this task.

1.2 Introduction to this report

In this report, we present the results of analysis and prediction applying recurrent neural nets to data set A. We will first describe the data set and define a validation set and a criterion measure. In Section 2, we analyse the linear and nonlinear, static and dynamic characteristics of the data. We present the results using the linear regression technique and a nonlinear prediction technique using feedforward neural nets, and explain why recurrent neural nets are chosen to model the data. A recurrent neural net model is then given in Section 3. We will describe the training process and demonstrate the prediction result. A recurrent neural net is a general, nonlinear and dynamic model, but a large net is usually required to track the dynamic state of non-stationary, complex signals. However, a large recurrent net will always result in training difficulty. In the following two sections of this report, two techniques are developed to alleviate the above difficulty. In Section 4, we train multi- recurrent nets to cope with the non-stationarity of the data. In Section 5, we analyse the long-term memory structure of the data, present a long-term prediction model and show an improved result. Finally, Section 6 concludes the recurrent neural net prediction model and discusses further modifications and improvements.

1.3 Data preprocessing and error measure

Data set A (as shown in Figure 1) is a time record of hourly energy usage including electricity, chilled water and hot water, for a four-month period in an institutional building. The task is to predict these three data sets for the following two-months. Four environmental data sets of temperature, humidity ratio, solar flux and wind speed, for the same six-month period are also provided as side information.

The given samples are the continuous records from 2am on September 1 1989 to 9am on February 23 1990. 2926 samples dated from September 1 1989 to December 31 1989 are provided as training data. We divided them into a validation set and a training set. Due to the non-stationary nature of the data, a validation set covering samples between $(200, 300)$, $(450, 550)$, $(700, 800), (950, 1050), (1200, 1300), (1450, 1550), (1700, 1800), (1950, 2050), (2200, 2300), (2450, 2550)$ and $(2700, 2800)$ is chosen to keep the probabilistic distribution of the training set and that of the validation set as close as possible. The sizes of the training set, the validation set and the test set are therefore 1826, 1100 and 1282 samples respectively.

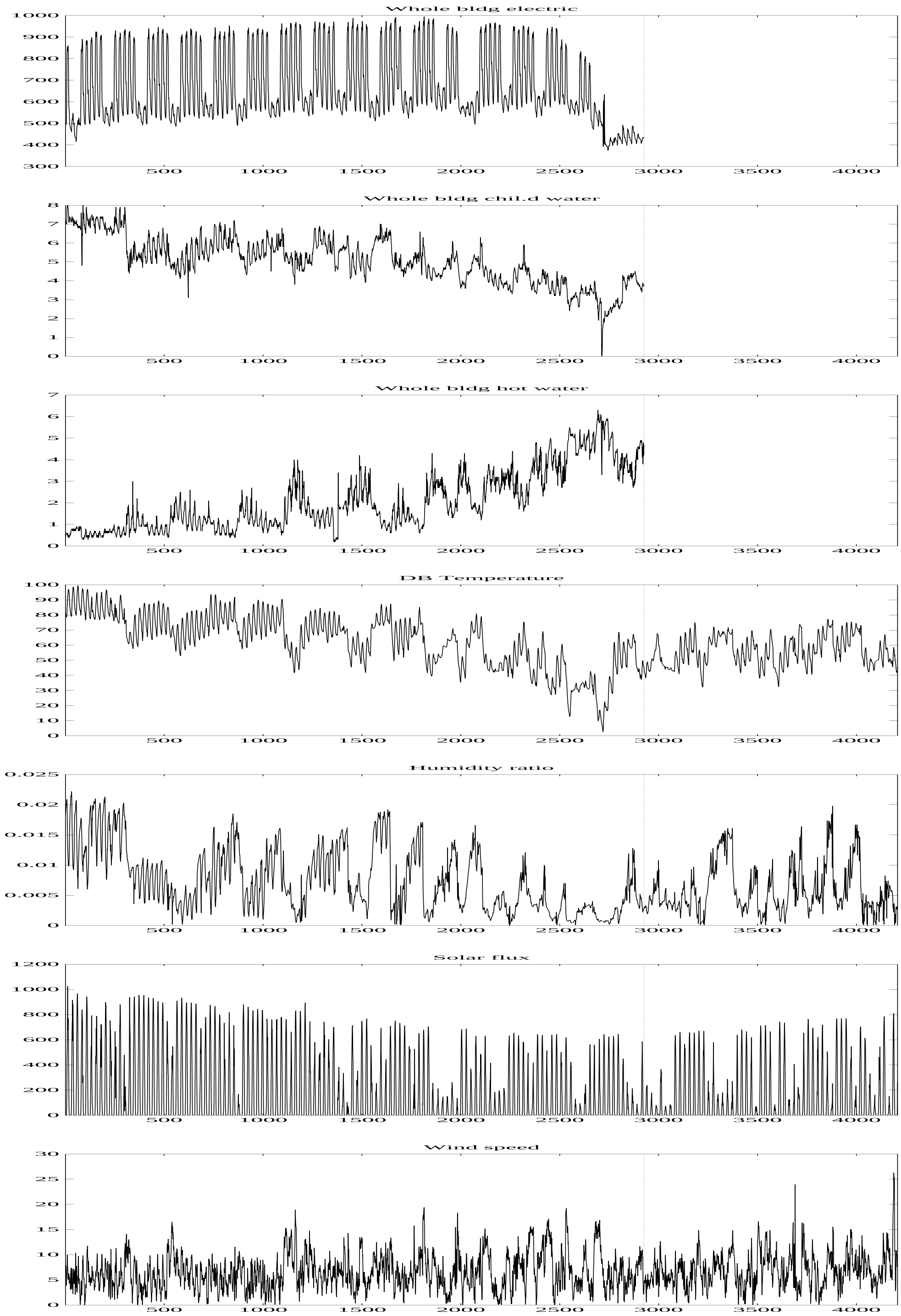

Figure 1: Building energy and environmental data set A. From top to bottom, they are respectively the four-month electricity usage data (kWh/hr), chilled water usage data (millions of Btu/hr), hot water usage data (millions of Btu/hr), the six-month temperature (deg F), humidity ratio (lb water/lb dry air), solar flux(W/sq meter) and wind speed (mi/hr).

Two additional inputs have been considered, one is the time information and the other is the date information. The time input is the hourly record and the date input gives working or non-working information. The non-working information indicates if it is ordinary weekends or long holidays such as Christmas and New Year holidays. We will see from the comparison in the following section that considering these two additional inputs will greatly improve the prediction precision.

Because the data sets are recorded in different measure units, they have respectively been normalised to zero-mean and one-variation before being applied to a predictor.

The coefficient of variance, CV, and the mean bias error, MBE, as described by Equation 1 and Equation 2 are defined as error measures. CV_{Tra} and MBE_{Tra} are defined as error measures over the training set, CV_{Val} and MBE_{Val} are those over the validation set and CV_{TraVal} and MBE_{TraVal} are those over both the training set and the validation set.

$$CV = \frac{1}{\bar{y}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{1}$$

$$MBE = \frac{1}{\bar{y}}\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i) \tag{2}$$

where y_i is an original data value, $\hat{y}_i$ is the prediction of y_i, $\bar{y}$ is the mean value of y_i and N is the number of data in the defined set.

Due to the time-limit, our study concentrates on the electricity usage data only. Our result will extend to the water usage data in future work.

2 Data Analysis

2.1 Static Analysis

We first consider the point-to-point mapping between the electricity data and the environmental data. The mapping relationship may be linear or nonlinear. The cross-plots of Figure 2 show that a nonlinear mapping could be a better selection. This can further be seen from the following comparison.

Using the linear-regression technique, e.g. (Draper and Smith, 1966), the linear correlation between the electricity data and the environmental data can easily be found. Its predictive error is shown in the first row of Table 2. This table also demonstrates the improvement of prediction performance by considering the time information and the date information.

Using the nonlinear prediction technique with a feedforward neural net, as in (Lapedes and Farber, 1987), the prediction error CV decreases by about 11.5% as shown in the second row of Table 2.

The prediction performance of a feedforward neural net relies on the size of the net. Here we set the number of layers to three and determine the number of hidden units by comparing

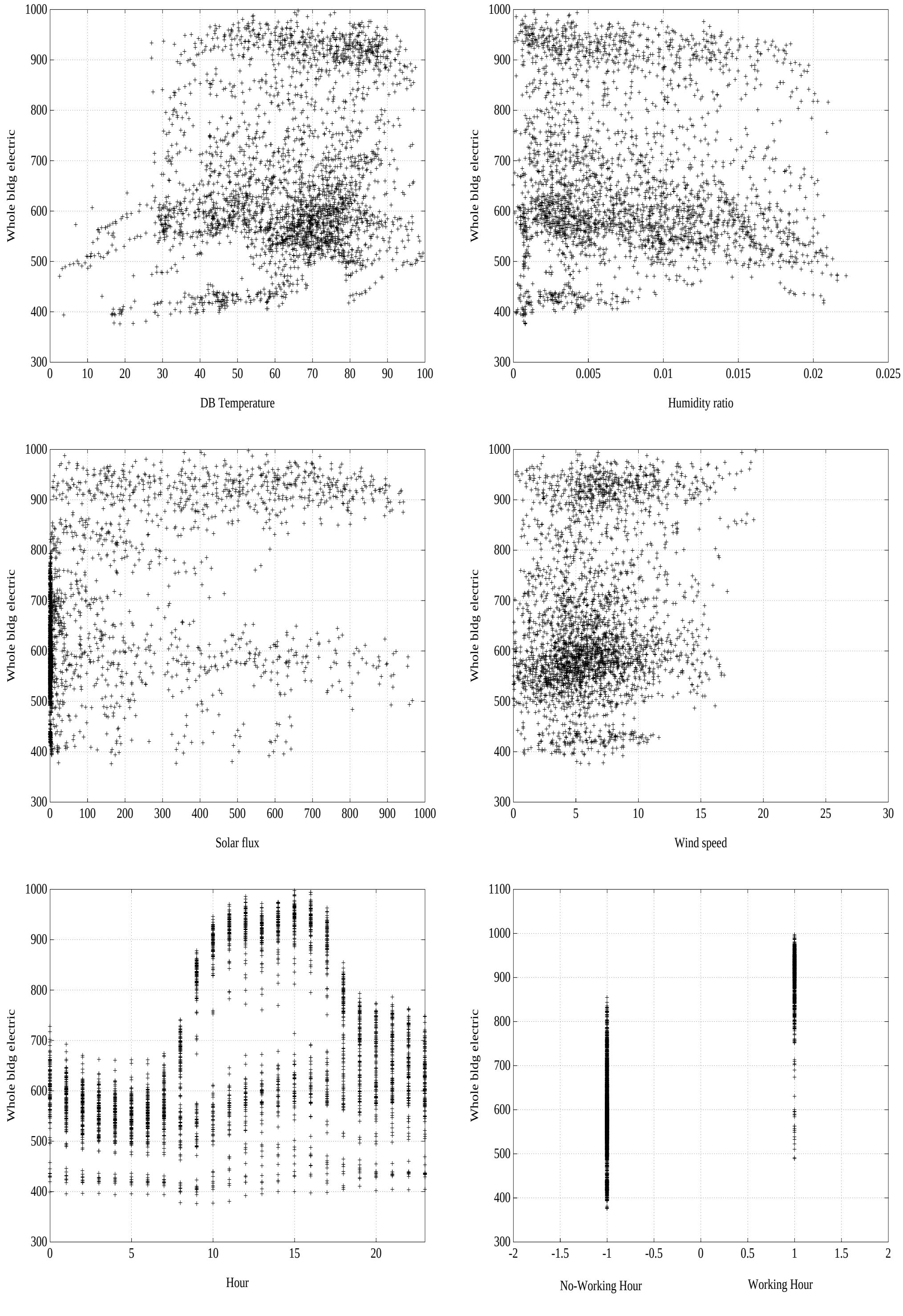

Figure 2: Crossplots between the building electricity usage and each environmental data and the time and date information.

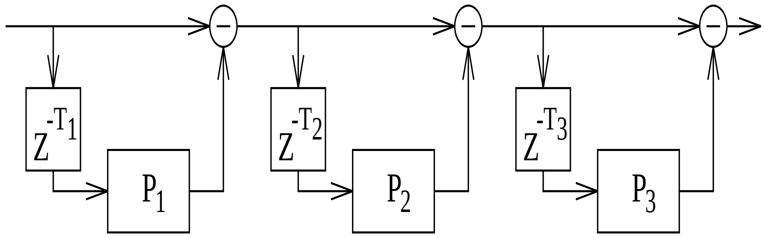

Figure 3: Cascaded short-term and long-term prediction model.

$Predictors$	CV_{Tra}	CV_{Val}	CV_{TraVal}	MBE_{Tra}	MBE_{Val}	MBE_{TraVal}
P_1	0.0407	0.0437	0.0419	0.0069	0.0053	0.0063
$P_1 + P_2$	0.0315	0.0362	0.0334	0.0013	-0.0009	0.0004
$P_1 + P_2 + P_3$	0.0249	0.0313	0.0276	0.0007	-0.0017	-0.0002

Table 1: Predictive error of the cascaded short-term and long-term prediction model to the electricity data.

the prediction error over the validation set. It was found that when the number of hidden units was increased to ten, the prediction error over the validation set did not reduce obviously.

2.2 Dynamic Analysis

The electricity data is not a memoryless signal. We now observe its dynamic characteristics. From the waveform shown in Figure 1, we can see that in addition to the short-term memory, there are two types of long-term information, with periods of 24 (one day) and 168 (one week) respectively.

With a cascaded short-term and long-term prediction model studied in (Wu and Fallside, 1992), one can get the predictive error as shown in Table 1 and Figure 4.

The cascaded short-term and long-term prediction model is formed by three predictors P_1, P_2 and P_3 as shown in Figure 3. Each predictor is a recurrent neural net. The P_1 uses several previous electricity data to predict a current sample. Its predictive residual is then time-delayed by 24 hours and applied to the P_2. The P_2 uses the predictive residual of P_1 to predict the current P_1 predictive residual. In sequence, the P_3 predicts the P_2 predictive residual using the predictive residual of P_2 with 168 hour-delayed.

For this analysis, we just used a very small size of recurrent neural nets. Each net consisted of three layers with three units in the input layer, two in the hidden layer and one output unit. The result in Table 1 and Figure 4 has shown that a very good prediction accuracy can be obtained by using the contextual memories, both short-term and long-term, of the data. Better prediction performance can be expected with larger nets. However, the cascaded short-term

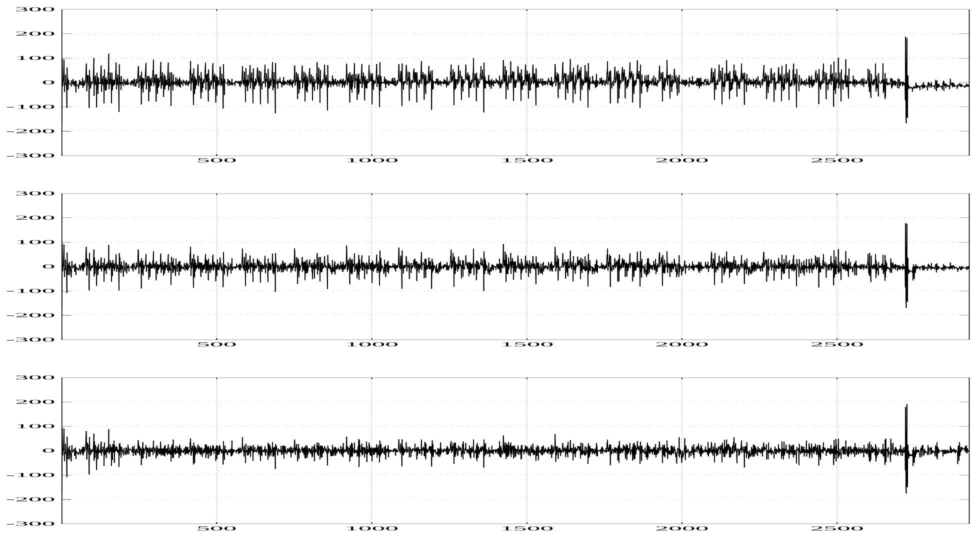

Figure 4: Predictive residual of building electricity usage data with a cascaded short-term and long-term prediction model. From top to bottom, they are respectively: (a) the predictive residual with a one-hour short-term predictor, (b) the predictive residual with a one-day long-term cascaded predictor and (c) the predictive residual with one-day and one-week two long-term cascaded predictors.

and long-term prediction model of Figure 3 cannot directly be applied to the task of building energy data prediction, since the previous samples of data are not available in the test set.

From the above analysis of the static and dynamic characteristics of data, we conclude that the mapping from the environmental data to the electricity data is nonlinear, and that the electricity data is a state-dependent signal and its memory state should be explored to attain good prediction performances.

A recurrent neural net is a general, nonlinear and dynamic system. We will first apply a recurrent neural net to simulating the mapping function between the electricity data and environmental data. The application of recurrent neural nets is then extended to model the long-term memory of electricity data.

3 Prediction with Recurrent Neural Nets

A recurrent neural network is shown in Figure 5. There are six inputs, respectively corresponding to four environmental data, the time information and the date information. The outputs of hidden units are time-delayed and feed-backed to the inputs of hidden units. The time-delayed units are set to one hour. There is only one output, which approximates the electricity data. We use the gradient-based training approach, as in (Wu and Fallside, 1992). The normalised environmental data and the time and date information are applied to the input

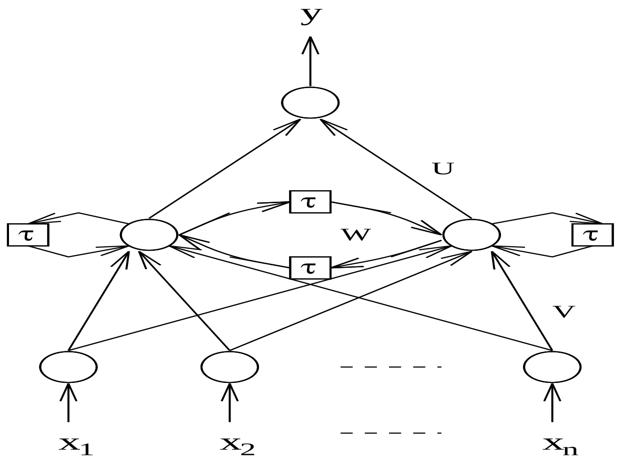

Figure 5: Recurrent neural net model for prediction of building electricity data.

units and fed forwarded from the input layer to the hidden layer, to the output layer. The output is compared to the original electricity data. The difference between the output and the electricity data is then fed back from the output layer to the hidden layer, to the input layer to get the gradient-descent of each connection weight. To avoid under-training or over-training, the number of hidden units and the training process are determined and judged by the change of performance over the validation set

The response of a recurrent neural net depends on its inputs and its states. The current states of a recurrent net are the hidden outputs of the last sample, which reflect the dynamic characteristic of previous inputs. During training, both feedforward and backward processes are carried out through the training and validation sets, but the weights are updated over the training set only.

The gradient-based training approach cannot guarantee a global optimum. Its local property depends on the initialisation of network weights. Usually, the weights are initialised by random values. Different initialisations may lead to greatly different training results. To reduce the effect of initialisation, N recurrent neural nets with different initialisations have been trained. M best nets ($M < N$) are then chosen and averaged to approximate the electricity data.

The prediction of electricity data and its predictive residual are shown in Figure 6, which is obtained by averaging ten nets chosen from twenty nets. All these nets contain only one hidden layer. The number of hidden units varies from six to eight. The predictive error is given in Table 2. Compared to the feedforward neural net, the prediction error CV of the recurrent net decreases by about 42.5%.

In some conventional prediction techniques, the differential signal $\delta_n = s_n - s_{n-1}$, instead of the original signal s_n itself, is predicted, and then the reconstructed signal $\hat{s}_n$ is estimated by

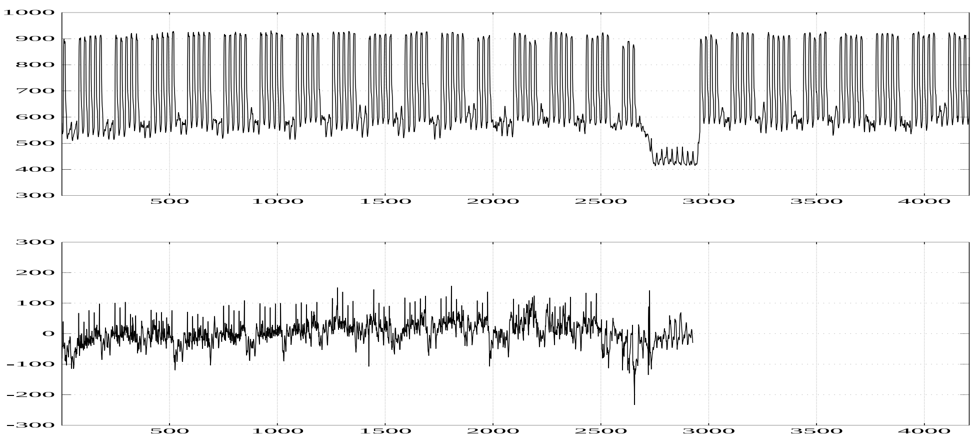

Figure 6: Prediction waveform (top) and predictive residual (bottom) of electricity data with a recurrent neural net.

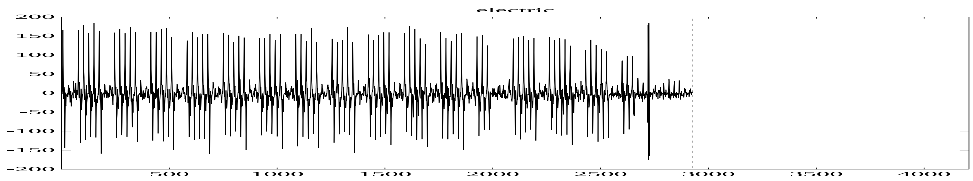

Figure 7: First-order differential signal of electricity data.

$\delta_n + \delta_{n-1}$. The first-order differential signal of electricity data is shown as in Figure 7. It seems that it is more stationary and easier to predict than the original signal, but, actually, we found that the reconstructed signal is far from the original one, because any previous predictive error will be accumulated.

4 Prediction with Multi- Recurrent Nets

In theory, a well-trained recurrent net can automatically track the state of dynamic signals. However, a large net is always needed to learn a transfer function of complex non-stationary signals and, usually, a large recurrent net cannot successfully be trained.

We have simulated the data using a recurrent net as studied in the last section. To deal with the non-stationarity of electricity data, more hidden units have been added to the recurrent net. However, we found that, as the number of hidden units increased up to eight, the prediction performance of the training set and that of the validation set were already saturated. The most probable reason is that a large recurrent net cannot perfectly be trained and becomes

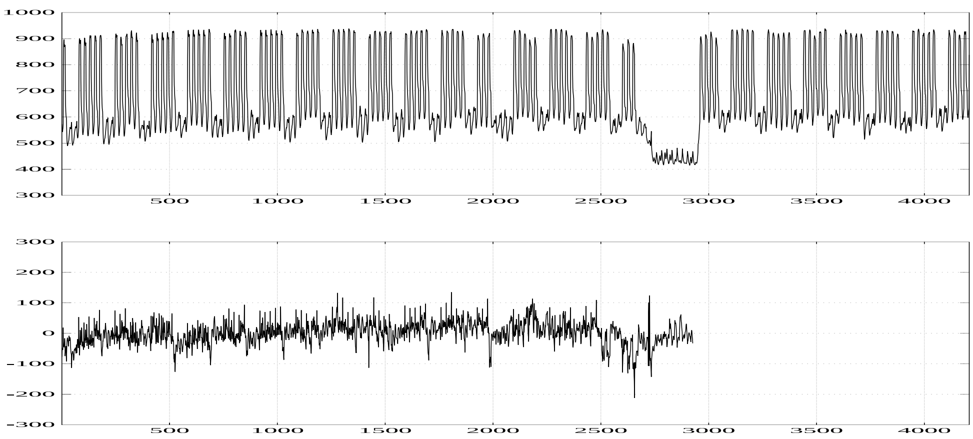

Figure 8: Prediction waveform (top) and predictive residual (bottom) of electricity data with double recurrent neural nets.

more easily trapped in a local minimum. In this section, we will try a modular approach to avoid the training difficulty of large recurrent nets.

The modular approach consists of multi- recurrent nets with the same-architecture. The size of the recurrent net is small so that it can successfully be trained using the gradient-based algorithm. All nets receive the same inputs and the output of each trained recurrent net is expected to cover a certain part of electricity data space. We divide the electricity data into two classes, one class consists of weekdays' data and another consists of weekends' and holidays' data. After a net has been trained using the whole training data, its weights are copied to another net with the same architecture and inputs. One net is further trained using the weekdays' data, and another is further trained using the weekends' and holidays' data. When the training process is switched from one net to another, the outputs of the hidden units are copied from the original net to the new net as current state variables.

The training process is stopped when the prediction error over the validation set does not decrease any more. Because the weights of multi-nets are initialised with the weights of a trained single-net, the prediction performance of multi-nets is at least as good as that of a single-net.

Based on the twenty trained single recurrent nets, which are randomly initialised as mentioned in the last section, twenty sets of double recurrent nets have been produced. After being further trained, ten with best prediction performance are selected. Their average is shown in Figure 8. Compared to the single recurrent net of the last section, the prediction error CV, as shown in Table 2, decreases by about 9.6%.

There are different classification methods in signal space. The classification of electricity data into a working date class and a non-working date class is based on intuitive observation of the electricity data waveform. Another classification approach is to group the data into vectors and classify them using a vector quantisation technique. We will study and compare different classification approaches in our future work.

5 Long-term Prediction Modelling

As discussed in Section 2, there exist two long-term memory structures in electricity data. This can also be seen from the regular spikes in the predictive residuals of Figure 6 and Figure 8.

The outputs of hidden units are time-delayed by one interval between two successive samples and fed back in the recurrent net we studied in the last sections. We have simulated a recurrent net with three groups of time-delay, feedback connections. The first group of connections is time-delayed by one hour, the second group by one day and the third group by one week. Because the total number of feedback connections is twice more than the previous net, when the number of hidden number increases, it is found that the recurrent net is more easily trapped into a local minimum and cannot attain any better result than that with a single group of time-delay, feedback connections.

Beside recurrent neural nets, we have also tested the prediction for this building energy data with a time-delay neural net (Waibel et al., 1989). We used a three-layer feedforward neural net with eighteen-input units. Six units correspond to the current inputs, six units are the inputs with one day time-delay and the other six units are the inputs with a one week time-delay. The number of hidden units is varied and determined by the prediction performance from the validation set. It was found that the time-delay neural net could not achieve any better performance than the recurrent neural net model.

In Section 2, a cascaded long-term and short-term prediction model has shown a very good prediction accuracy on the training and validation sets of electricity data. However, this model is not applicable to the test set because the original electricity data is not available during all the segment of test set. In this section, we develop a new long-term prediction model formed by two parallel recurrent neural nets. Instead of using the original data, its inputs receive the estimation of electricity data obtained from the recurrent neural net predictor resulting from the study of the last two sections.

As shown in Figure 9, the RNN_1 is a recurrent neural net similar to that shown in Figure 5. Its inputs consist of four environmental data (x_1, x_2, x_3, x_4), the time information x_5 and the date information x_6. Its output y_1 is an estimation of the electricity data. The RNN_{24} and the RNN_{168} are two recurrent neural nets structured as RNN_1 except that their inputs are the sequentially time-delayed signal, $\bar{Y}_1(t-T) = \{y_1(t-T-N), y_1(t-T-N+1), \ldots,$ $y_1(t-T+N-1)$ and $y_1(t-T+N)\}$. The RNN_{24} is trained to learn the long-term memory structure in the period of 24 hours and the RNN_{168} for that with 168-hour period. Z^{-T_1} and Z^{-T_2} are two time-delayed units with $T_1 = 24 - \frac{N_i-1}{2}$ and $T_2 = 168 - \frac{M_i-1}{2} - T_1$, where N_i

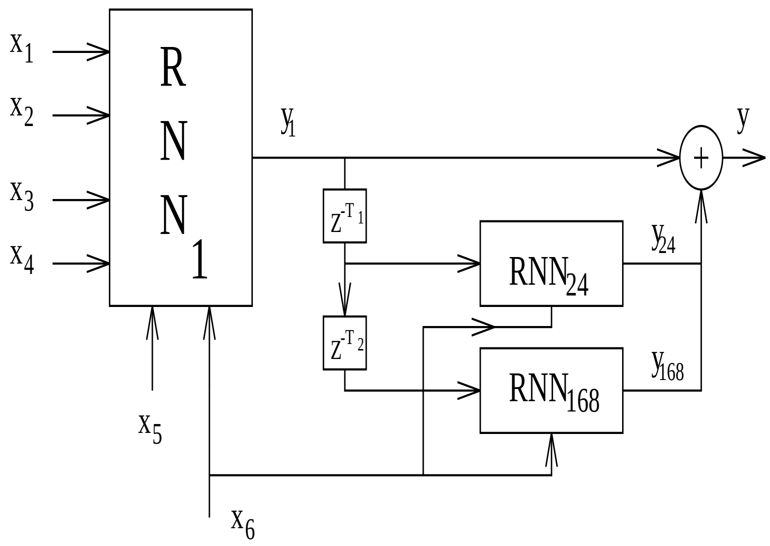

Figure 9: Long-term prediction modelling for building electricity data.

and M_i are respectively the numbers of input units of RNN_{24} and RNN_{168}. The sum of the outputs of RNN_{24} and RNN_{168}, $y_{24} + y_{168}$, is trained to approximate the predictive error of RNN_1 as shown in Figure 8. The RNN_{24} and the RNN_{168} are trained in sequence. First, the RNN_{24} is trained with the predictive error of RNN_1 as a teaching signal, the predictive error of RNN_{24} is then used to train the RNN_{168}.

Both the RNN_{24} and the RNN_{168} are controlled by the date information x_6. As described in Section 2, x_6 is actually a state variable. It consists of two states, working state and non-working state. When using a group of time-delayed signals $\bar{Y}_1(t-T)$ to predict $y_{24}(t)$ or $y_{168}(t)$, $x_6(t-T)$ and $x_6(t)$ should lie in the same state. During prediction, when $x_6(t-T)$ is not equal to $x_6(t)$, both the RNN_{24} and the RNN_{168} are switched off. Their state variables, i.e. the outputs of hidden units in the RNN_{24} and the RNN_{168}, are also re-set to zero.

The RNN_{24} and the RNN_{168} are defined as three layers with two hidden units. Determined by the prediction performance from the validation set, the numbers of input units of RNN_{24} and RNN_{168} are set to nine. The waveforms of y_{24}, y_{168}, y and its predictive residual are plotted in Figure 10. Compared to the recurrent net model studied in the last section, the prediction error CV, as shown in Table 2, decreases by about 9.8%.

In Figure 9, each of three recurrent nets are trained in sequence. It is possible jointly to re-train and fine-tune the network weights so as to further improve the prediction accuracy.

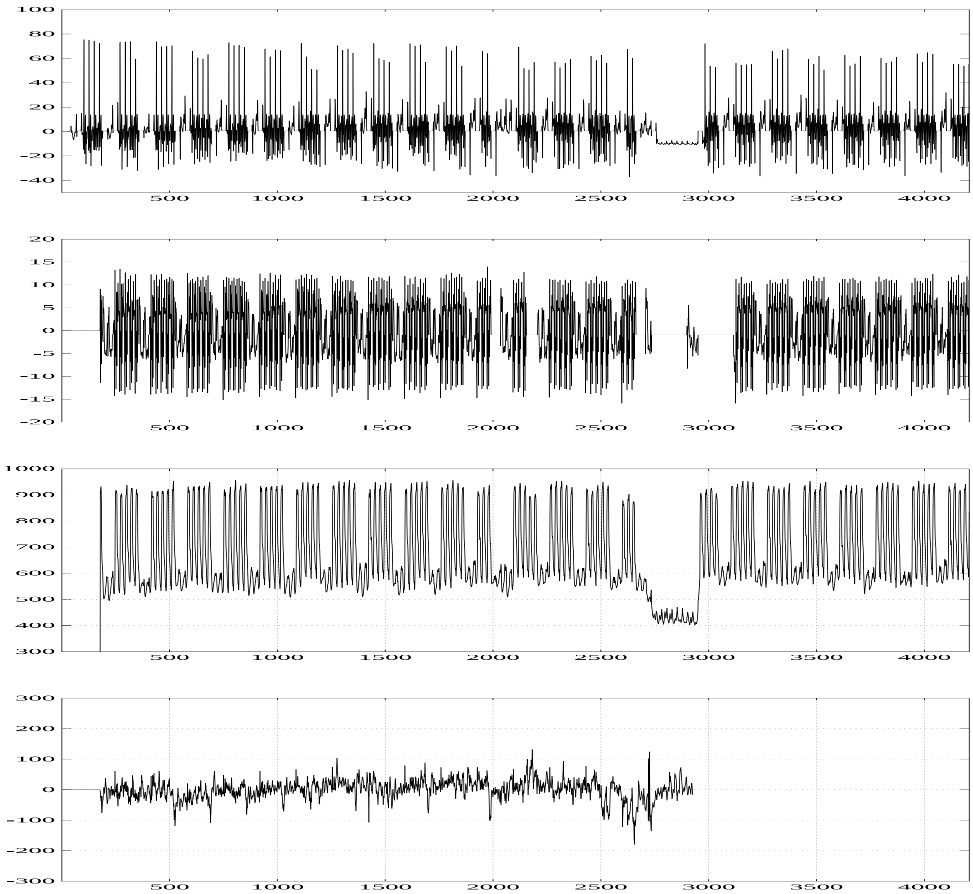

Figure 10: Prediction waveform and predictive residual of electricity data with a long-term prediction model. From top to bottom, they are respectively the prediction waveforms of y_{24}, y_{168}, y and its predictive residual.

$Model$	CV_{Tra}	CV_{Val}	CV_{TraVal}	MBE_{Tra}	MBE_{Val}	MBE_{TraVal}
A	0.2063	0.2019	0.2047	0.0137	0.0256	0.0212
B	0.1205	0.1114	0.1172	0.0604	0.0562	0.0588
C	0.1028	0.1052	0.1037	0.0273	0.0261	0.0269
D	0.0613	0.0570	0.0597	0.0032	0.0034	0.0033
E	0.0540	0.0536	0.0539	0.0008	0.0006	0.0007
F	0.0483	0.0490	0.0486	0.0023	-0.0024	0.0004

Table 2: Comparsion of predictive errors, where model A is a linear regression with four environmental inputs; model B is a model A adding the inputs of the time and the date information; model C is a nonlinear predictor with a feedforward neural net; model D is a nonlinear, dynamic predictor with a single recurrent neural net; model E is a nonlinear, dynamic predictor with double recurrent neural nets and model F is a model E with a long-term prediction model.

6 Concluding Remarks and Discussion

Based on the analysis of the data, a recurrent neural net has been chosen as a model to predict the building energy data. A recurrent net, in theory, can model any kind of data without making any assumption on the data. Its only disadvantage is the training difficulty for nets of large sizes. We have reported a modular approach which avoids training a large recurrent neural net. The main principle of approach is to divide the whole task into several related sub-tasks and attain each sub-task with a rather small recurrent neural net.

Because the electricity data in the test set is not available, we do not know the prediction error over the test set at the moment. Figure 11 shows cross-plots between the prediction of electricity data and the environmental data in the test set. Compared to Figure 2, the cross-plots in the training data set, both the distributions in the training set and the test set are very similar.

In this prediction task, we have used a priori knowledge, the date information, as one of the inputs of the predictor. The date information is identified by two states of working and non-working days. The data set does not provide such a priori knowledge. We have assumed that the Thanksgiving holidays started on Thursday the 23rd of November and ended on Sunday the 26th of November and that the Christmas holidays started from Thursday the 21st of December and ended Monday 1st January 1990.

If there were other holidays which happened during the test set, for example, Martin Luther King Day on Monday of the 15th of January in 1990 and Washington's Birthday on Thursday of the 22nd of February in 1990, (referred to "Hints to Exporters, The United States of American, 1989/90" by British Department of Trade and Industry), a different prediction in the test set will result.

We suppose that providing the information on public holidays in prediction is practical

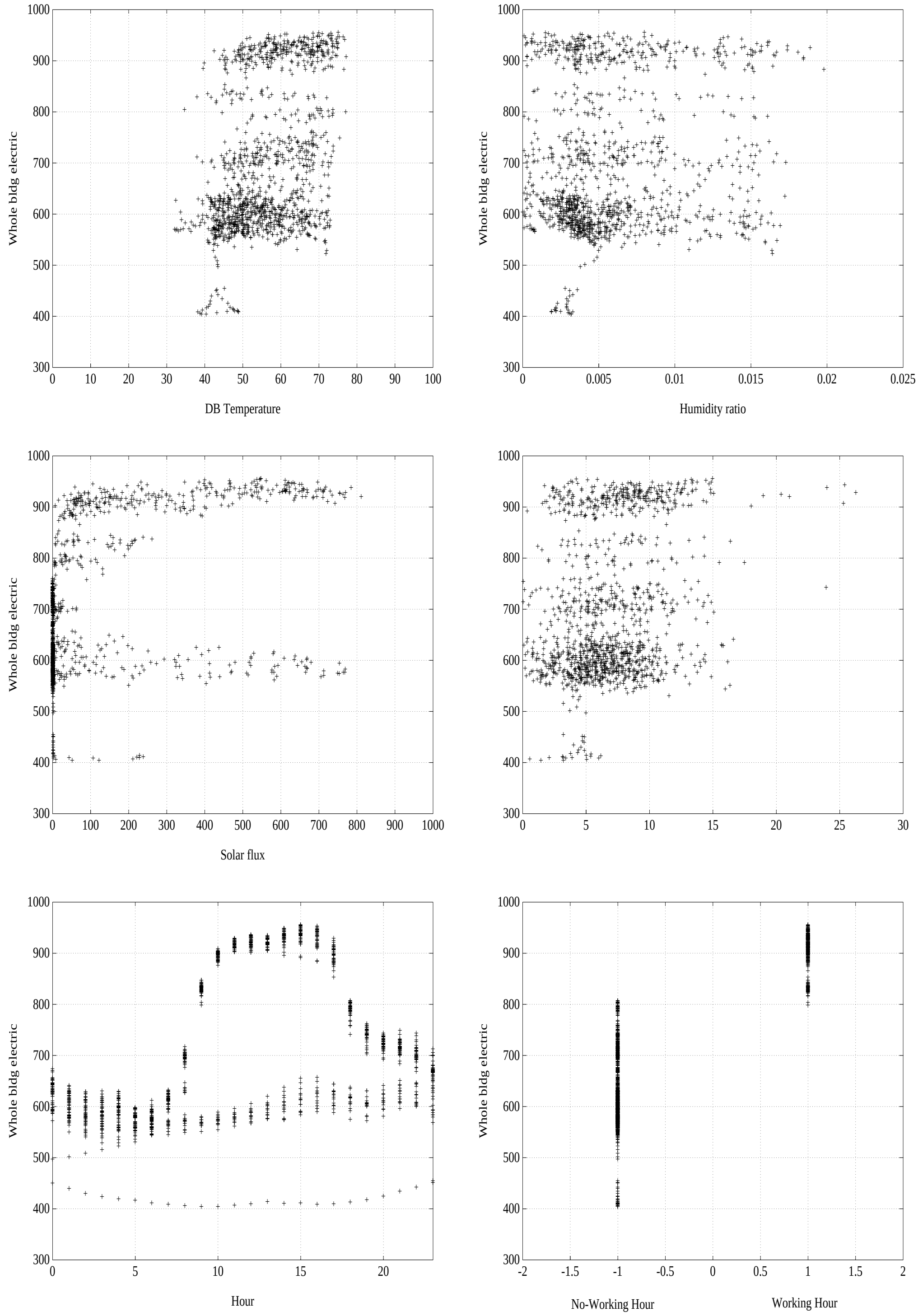

Figure 11: Crossplots between the prediction of electricity data and each environmental data and the time and date information in the test set.

and applicable since which holidays will be observed in a given building should certainly be a priori knowledge.

Acknowledgements

We wish to thank G.Saleh, T.Burrows and S. Harvey for their valuable comments on the manuscript.

References

Draper, N. and Smith, H., editors (1966). *Applied Regression Analysis.* John Wiley & Sons, Inc.

Kreider, J. and Haberl, J. (1993). The great energy prediction shootout. Concept and Summary of the first building data analysis and prediction competition.

Lapedes, A. and Farber, R. (1987). Nonlinear signal processing using neural networks: prediction and system modelling. Technical Report LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, New Mexico 87545.

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., and Lang, K. (1989). Phoneme recognition using time-delay neural networks. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-37:328–339.

Wu, L. and Fallside, F. (1992). Fully vector-quantised neural network-based code-excited nonlinear predictive speech coding. Technical Report CUED/F-INFENG/TR.94, Cambridge University Engineering Department, Trumpington St., Cambridge CB2 1PZ, U.K.