Short-Term Prediction of Energy Consumption in Buildings Based on Artificial Neural Networks and Nonlinear Time Series Analysis

S. Karatasou and M. Santamouris

Group Building of Environmental Physics, University of Athens, Building 5, Physics University Campus, 15784 Athens, Greece

ABSTRACT

This paper introduces a new approach for the prediction of hourly energy consumption in buildings. The proposed method uses nonlinear timeseries analysis techniques for the reconstruction of energy consumption timeseries and the estimation of the dynamic invariants, and artificial neural networks as a nonlinear modeling tool.
Among the several neural network modeling factors that affect time-series prediction, the most important are the window-size and the sampling lags for the data. Relevant theoretical results related to the reconstruction of a dynamical system are analyzed and the relationship between a correct embedding dimension and network performance is investigated.
The problem is examined initially for the univariate case and is extended to include additional calendar parameters, in the process of estimating the optimum model.
Different network topologies are considered, as well as existing approaches for solving multi-step ahead prediction problems. The performance of short-term predictors is also examined with regard to prediction horizon.
The performance of the predictors is evaluated using measured data from real scale buildings, showing promising results for the development of accurate prediction tools.

KEYWORDS

INTRODUCTION

To predict building energy consumption a large number of building software tools are available, making feasible to model a building for thermal evaluation and study it's exact thermal behavior. Building thermal models which have been widely used in a variety of buildings and for a range of applications, in practice diversify on many factors: the modeling methodology, the physical laws, parameters and data that they encase, the integration of HVAC, passive solar, photovoltaic systems. Thus, depending on the application, these models vary on complexity and can be simple and easy to use, or more sophisticated and time-consuming to set-up and run (ASHRAE, 2001).
In general, for the majority of applications, most of the appropriate software tools are time consuming and computationally heavy, especially when transient numerical methods are used. A large number of assumptions often need to be made when the quantitatively measurement of factors like infiltration or the estimation of parameters like occupancy is not possible. Also, parameters like the cost, the level of expertise and the exhaustive information needed to be collected could be prohibitive for a massive implementation.

Furthermore, almost all energy consumption predictive schemes are based on the prior prediction of weather data. As many weather variables are considered such as dry bulb temperature, relative humidity, solar radiation and cloudiness conditions, the most common practice is to use weather forecasts issued by meteorological centres, yet the direct link with such a centre make the procedure even more complicated.

Artificial Neural Networks (ANN) can provide an alternative approach, as they are widely accepted as a very promising technology offering a new way to solve complex problems. ANNs ability in mapping complex non-linear relationships, have succeeded in several problems such as planning, control, analysis and design. The literature has demonstrate their superior capability over conventional methods, their main advantage being the high potential to model non-linear processes, such as utility loads or individual buildings energy consumption.

As far as it concerns energy modelling for the building sector, many studies have been reported on the use of neural networks. These can be divided mainly into two groups: models to estimate building energy use (Ansett and Kreider, 1993, Kreider and Haberl, 1994, Mackay 1994, Ohlsson et al., 1994, Haberl and Thamilseran, 1996, Dodier and Henze 1996, Ben-Nakhi and Mahmoud 2004, Karatasou et al.,2006) and algorithms for a wide range of HVAC applications, such as design, operation and fault detection (Mistry and Nair, 1993, Curtiss et al. 1993, Curtiss et al., 1994, Kawashima et al., 1996, Ben-Nakhi and Mahmoud, 2002).

There are several important issues for the design of ANN, the dimension of the window size for the input representation of the past data being among the most important. In the absence of systematic approach to neural network modeling, several different approaches have been proposed to treat the aforementioned issues.

Short-term predictors of energy consumption in buildings have the potential to increase the efficiency of energy conservation techniques, such as ice-storage systems (Kawashima et al., 1996) or night ventilation techniques that have a time-shifted influence on the energy consumption in buildings.

In this work, a new approach to predict hourly energy consumption in buildings is examined, using the algorithm of average mutual information (Fraser and Swinney, 1996) and false nearest neighbors (Kennel et al., 1992) to identify the lag of past values and the order of the model respectively. These algorithms stem from the embedding theorem of Takens (1981) and the advances in non-linear dynamics and chaos time series analysis techniques. Thus, a simple one-step predictor, based only on historical data is derived and then used iteratively to extend prediction horizon to 24 hours.

THE DATA SET

The data set used in this work is the benchmark Proben 1, and comes from the first energy prediction contest, the Great Building Energy Predictor Shootout I, organized by ASHRAE (Haberl and Thamilseran, 1996). It consists of hourly data of building energy use (electricity, hot- and cold-water) of a big building, for which at the time of contest no other details (like type of use, occupancy etc) were available. The total data set covers the period from September 1989 to February 1990, whereas energy consumption data were available only for September-December 1989; the part from January to February 1990 was withheld by the organizers, and used to score the generalization performances.

PREDICTION METHODOLOGY

Building's energy consumption data time series can be seen as a sequence of vectors, depending on time t:

$$\vec{x}(t) \qquad \text{where } t=0,1,\ldots \tag{1}$$

Then, the problem can be stated as finding a function $F : R^d \rightarrow R$ such as to obtain an estimate of $\vec{x}(t+\mathrm{T})$ of the vector $\vec{x}$ at time t+T, given the values of $\vec{x}$ up to time t:

$$\vec{x}(t+\mathrm{T}) = F\big(\vec{x}(t), \vec{x}(t-\tau), \ldots, \vec{x}(t-(d-1)\tau)\big) \tag{2}$$

$$\vec{x}(t+\mathrm{T}) = F\big(\vec{y}(t)\big), \tag{3}$$

where x(t) lies in the d-dimensional time delay space. Dimension d is called the embedding dimension; T is the lag of prediction and normally T=1 so that the next $\vec{x}$ value will be predicted, but can take any value larger than 1. τ is the time delay.

Since F is deterministic, the problem of forecasting the component $\vec{x}(t+\mathrm{T})$ reduces to that of estimating the function F, and the neural network approach of performing prediction is to induce this function in a standard Multilayer Perceptron MLP architecture using a set of samples $\vec{x}(t), \vec{x}(t-\tau), \ldots, \vec{x}(t-(d-1)\tau)$ as inputs and a single output as target value of the network.

In the three-layer perceptron, the neurons are grouped in sequentially connected layers: the input, the output and the hidden layers. Each neuron in the hidden and output layer is activated by a non linear activation function that relies on the weighted sum of its inputs and the neuron parameter, called bias, *b*.

The output of a neuron in the output layer is

$$\hat{y}(k) = \sum_{j=1}^{h} w_j \Psi_j \left[\sum_{i=1}^{n} w_{ji} x_i + b_i \right] + b_j \tag{4}$$

where the h hidden units (processing elements) perform the weighting summation of the inputs x_i and the nonlinear transformation by the sigmoid (log-sigmoid or tan-sigmoid) transfer function $\Psi_j(.)$

In the present study, the *false nearest neighbor* method, proposed by Kennel (1992), is used to determine the minimal sufficient embedding dimension d. To start with, we estimate time delays with the *average mutual information*, a method suggested by Fraser and Swinney (1996) to determine reasonable time delays τ. The time lag can be taken at the first minimum of the mutual information graph. As shown in Figure 1, the first minimum for the data set occurs at τ=12. This is the value of time lag that is used to construct time delay vectors.

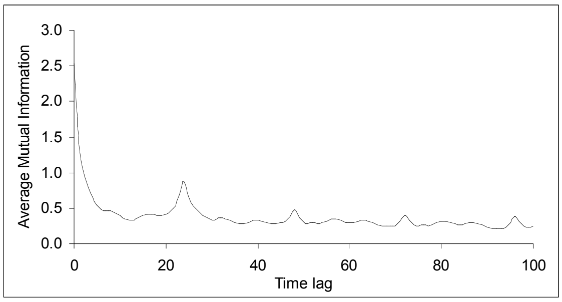

Figure 1: The average mutual information for the energy consumption timeseries.

Next we estimate the embedding dimension using the FNN method. The percentage of false nearest neighbors, as a function of dimension, is shown in Figure 3. The embedding dimension is specified as the embedding where the percentage of FNN first vanishes. The method suggests an embedding dimension of 6, e.g d=6.

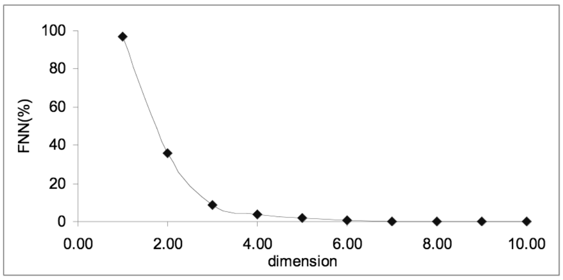

Figure 2: The percentage of global false nearest neighbors.

As neural networks with a single hidden layer of neurons with hyperbolic tangent activation function and a linear output neuron are universal function approximators (Hornik et al., 1989), we don't consider more complex architectures. In this way, the determination of the best model structure reduces to the determination of the appropriate number of inputs and hidden units. We thus use feed forward neural networks with a single hidden layer of tanh units, and a single linear output to predict hourly cooling load, where the number of past inputs is set equal to the embedding dimension d=6. In this way, six energy consumption values were selected at t-1, to t-6.
As energy consumption data present a daily cycle the hour of day is considered as well as an input variable, and coded by means of its sine and cosine values, into a clock representation in which $sh = \sin 2\pi h(t)/24$ and $ch = \cos 2\pi h(t)/24$ represent the hour of the day (where h is the hour of the day ranging from 0 to 23).
Moreover, as the occupancy of the building has a strong effect on the energy use, weekends and holidays were identified and days were classified and encoded as 1 (weekday or working day) and 0 (weekend or holiday).
The data set includes a total of 4208 time steps, where data [1,1296] are available for training, and [2927,4208] for testing. For the test set the energy consumption were withheld by the organizers, and used to score the generalization performances. For consistency reasons, we use this part of the

data, only at the very end, to present fairly comparable results with previous works.
The subset of [1:1296] data patterns selected as the training set and used for training the networks, using Lavenberg-Marquardt algorithm (LM) (Hagan and Menhaj, 1994). LM optimization technique is a more sophisticated method than gradient descent. It is based on Gauss-Newton method, and it is very powerful and fast.
Considering that the notion "n_1:n_2:n_3" denotes a network with n_1 inputs, n_2 hidden neurons and n_3 outputs units, the number of hidden neurons (n_2) was obtained by testing different structure of the network in the range $2 \leq n_2 \leq 2n_1$.
For each number of hidden units, networks are trained q times, where q is the number of its parameters, each time starting from different random initial parameters values. The model that is kept is the one with the minimum performance error, calculated for the test set. In this way the number of hidden neurons was selected equal to 8.
To validate the proposed method, other cases are investigated, keeping in all models the three last inputs (the variables associated with the hour of the day and weather the building is in session or not) and varying the number of delayed input units. Figure 3 shows the relationship between the number of delayed inputs and the Coefficient of Variation CV. It can be observed that the model with 6 delayed input has is the one with the smaller CV, equal to 2.5%.

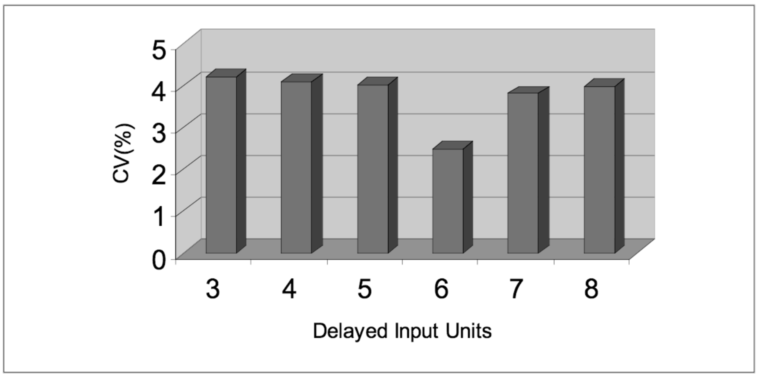

Figure 3: Coefficient of variation vs. delayed input units

The proposed model is then used to get a prediction for one to 24 hours ahead. As weather data are not included on the set of input variables, it can be used iteratively to perform multiple step prediction, by feed back the network outputs as inputs, when required. The CV for the 24-step predictor is 11.08%. A graphical representation of comparison between predicted and measured values is shown in Figure 4.

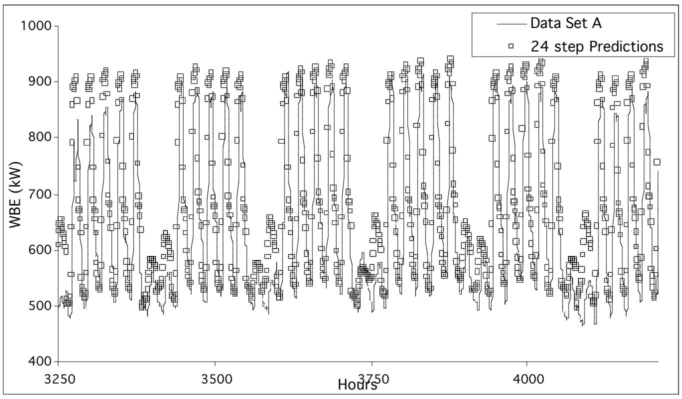

Figure 4: Predicted whole building electricity (WBE) consumption compared with data.

CONCLUSIONS

This paper has proposed a method for predicting hourly energy consumption in buildings. The False Nearest Neighbors method has provided a successful way for selecting the number of delayed input units. The main advantage of the proposed prediction scheme is its design: it used only the measured variable, and thus eliminates the necessity of predicting weather variables as well. Single step predictor is very accurate, but the results show that it can be effective for predictions extended to the next 24 hours

REFERENCES

ASHRAE. Energy estimating and modeling methods. ASHRAE Fundamentals, 2001.

Anstett, M. and Kreider, J.F. (1993). Application of neural networking models to predict energy use. *ASHRAE Transactions* **99**: **1**, 505 –17.

Ben-Nakhi, A. and Mahmoud, M. (2002). Energy conservation in buildings through efficient A/C control using neural networks. *Applied Energy* **73**, 5-23.

Ben-Nakhi, A. and Mahmoud, M. (2004). Cooling load prediction for buildings using general regression neural networks. *Energy Conversion and Management* **45**, 2127-2141.

Curtiss, P.S., Brandemuehl, M.J. and Kreider, J.F. (1994). Energy management in central HVAC plants using neural networks. *ASHRAE Transactions* **100: 1**, 476-493.

Curtiss, P.S., Kreider, J.F. and Brandemuehl, M.J. (1993). Adaptive control of HVAC processes using predictive neural networks. *ASHRAE Transactions* **99: 1**, 496-504.

Dodier, R.H. and Henze, G.P. (1996). Statistical analysis of neural network as applied to building energy prediction. *Proceedings of the ASME ISEC*, San Antonio, TX, 495-506.

Fraser, A. M. and Swinney, H. L. (1996). Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **33**, 1134-1140.

Haberl, J.S. and Thamilseran, S. (1996) The Great Energy Predictor Shootout II: Measuring Retrofit Savings- Overview and Discussion of Results. *ASHRAE Transactions* **102: 2**, 419-435.

Hagan, M.T. and Menhaj, M. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Network* **5**, 989-993.

Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer feed forward networks are universal approximators. *Neural Networks* **2**, 359-366.

Karatasou, S., Santamouris, M., Geros, V. (2006). Modeling and predicting building's energy use with artificial neural networks: Methods and results. *Energy and Buildings* **38**, 949-958

Kawashima, M., Dorgan, C. and Mitchell, J. (1996) Optimizing System Control with Load Prediction by Neural Networks for an Ice-Storage System. *ASHRAE Transactions* **102: 1**, 1169-1178.

Kennel, M. B., Brown, R. and Abarbanel, H. D. I. (1992) Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**, 3403-3411.

Kreider, J.F. and Haberl, J.S. (1994) Predicting hourly building energy use: The great energy predictor shootout- Overview and discussion of results. *ASHRAE Transactions* **100: 2**, 1104-18

MacKay, D. (1994) Bayesian Nonlinear Modeling for the Prediction Competition. *ASHRAE Transactions* **100:2**, 1053–1062.

Mistry, S.I. and Nair, S.S. (1993) Nonlinear HVAC computations using neural networks. *ASHRAE Transactions* **99: 1**, 775-784.

Ohlsson, M., Peterson, C., Pi, H., Rognvaldsson, T. and Soderberg, B. (1994) Predicting System Loads with Artificial Neural Networks. *ASHRAE Transactions* **100(2)**, 1063-1074.

Takens, F. (1981) Detecting strange attractors in turbulence, In D.Rand and L.S. Young (Eds), Dynamical Systems and Turbulence, Warwick, 1980, Springer, Berlin, 1981.