2016

Experiment Design and Training Data Quality of Inverse Model for Short-term Building Energy Forecasting

Liang Zhang
Drexel University, Department of Civil, Architectural and Environmental Engineering, lz356@drexel.edu

Jin Wen
Drexel University, Department of Civil, Architectural and Environmental Engineering, jw325@drexel.edu

Can Cui
Arizona State University, School of Computing, Informatics, Decision Systems Engineering, ccan1@asu.edu

Xiwang Li
Harvard University, Graduate School of Design, xiwang_li@gsd.harvard.edu

Teresa Wu
Arizona State University, School of Computing, Informatics, Decision Systems Engineering, teresa.wu@asu.edu

Experiment Design and Training Data Quality of Inverse Model for Short-term Building Energy Forecasting

Liang ZHANG[1]*, Jin WEN[1], Can CUI[2], Xiwang LI[3], and Teresa WU[2]

[1]Drexel University, Department of Civil, Architectural and Environmental Engineering,
Philadelphia, Pennsylvania, USA
E-mail: lz356@drexel.edu
E-mail: jw325@drexel.edu

[2]Arizona State University, School of Computing, Informatics, Decision Systems Engineering,
Phoenix, Arizona, USA
E-mail: ccan1@asu.edu
E-mail: teresa.wu@asu.edu

[3]Harvard University, Graduate School of Design,
Cambridge, Massachusetts, USA
E-mail: xiwang_li@gsd.harvard.edu

*Corresponding author

ABSTRACT

For data-driven building energy forecasting modeling, the quality of training data strongly affects a model's accuracy and cost-effectiveness. In order to obtain high-quality training data within a short time period, experiment design, active learning, or excitation is becoming increasingly important, especially for nonlinear systems such as building energy systems. Experiment design and system excitation have been widely studied and applied in fields such as robotics and automobile industry for their model development. But these methods have hardly been applied for building energy modeling. This paper presents an overall discussion on the topic of applying system excitation for developing building energy forecasting models. For gray-box and white-box models, a model's physical representations and theories can be applied to guide their training data collections. However, for black-box (pure-data-driven) models, the training data's quality is sensitive to the model structure, leading to a fact that there is no universal theory for data training. The focus of black-box modeling has traditionally been on how to represent a data set well. The impact of how such a data set represents the real system and how the quality of a training data set affect the performances of black-box models have not been well studied. In this paper, the system excitation method, which is used in system identification area, is used to excite zone temperature set-points to generate training data. These training data from system excitation are then used to train a variety of black-box building energy forecasting models. The models' performances (accuracy and extendibility) are compared among different model structures. For the same model structure, its performances are also compared between when it is trained using typical building operational data and when it is trained using exited training data. Results show that the black-box models trained by normal operation data achieve better performance than that trained by excited training data but have worse model extendibility; Training data obtained from excitation will help to improve performances of system identification models.

1. INTRODUCTION

Buildings are responsible for 40% of global energy use and contribute towards 30% of the total CO_2 emissions (Ahmad et al., 2016). It is significant to develop better building control and operation strategies to improve building energy efficiency and save energy.

Short-term whole building energy forecasting modeling is a core component of building energy control, operation strategy determining and building-to-grid integration. Inverse modeling (data-driven) approach outperforms forward modeling (physics-based) approach in terms of real-time controllability, simplicity, and engineering practicability. However, the performances of data-driven models are strongly affected by the training data. Because typical buildings often are operated with limited set points (one or two zone set points, for example) and operation modes, operational data collected from typical building operations are not rich enough which result in the poor extendibility often observed in data-driven models (Li et al., 2016).

In recent years, there are many studies applying varieties of inverse modeling strategy to short-term building energy forecasting. Some of them also focus on improving the accuracy as well as simplifying models to make them suitable for on-line control and optimization. Dong et al. (2005) presented support vector machines (SVM), a neural network algorithm, to forecast building energy consumption in the tropical region, which examines the feasibility and applicability of SVM in building load forecasting area. Yang et al. (2005) evaluated the performance of adaptive artificial neural network (ANN) models that are capable of adapting themselves to unexpected pattern changes in the incoming data, and therefore can be used for the real-time on-line building energy prediction. Newsham and Birt developed an ARIMAX model to forecast the power demand of the building in which a measure of building occupancy was a significant independent variable and increased the model accuracy (Newsham and Birt, 2010).

Although plenty of existing studies applied different approaches to model and forecast building energy performance, there is still a lack of systematic analysis about the relationship between data quality and model accuracy (Li and Wen, 2014). Inverse modeling approach is based on data. So data quality directly influences model quality. What's more, the complexity of building energy modeling, with many inputs and nonlinearity require sufficient data to reflect the building dynamics. Good experiment design yields data with rich information, which is the basis of a successful modeling process. Operational data often lacks rich information and may have undesirable properties which make it hard to build or train models. As a result, active input data design or experiment design should be used to ensure sufficient data quality and quantity to reflect the model structure and characteristics.

In this paper, a literature review of experiment design (including system excitation) and training data quality in short-term building energy forecasting models is presented. Difference of experiment design between system identification and black-box models will be illustrated. Finally, the impact of different training data on different building modeling structures will be studied.

2. LITERATURE REVIEW OF EXPERIMENT DESIGN IN BUILDING ENERGY MODELS

There are many review papers introducing building energy forecasting models (Li and Wen, 2014; Zhao and Magoulès, 2012; Swan and Ugursal, 2009), but few summarized the application of experiment design (including system identification) and the evaluation of data quality in this field. In this section, a general literature review will be given to summarize the existing researches about experiment design and data quality of inverse model for short-term building energy forecasting. Gray-box models (including those frequency-based models from system identification) have relatively fixed model structure and known properties of linearity or nonlinearity, only with parameters to be determined. However, black-box models are more complicated: sometimes most of model information, including properties of linearity and nonlinearity, model structure and its parameters, is based on training data. The experiment design theory and process are quite different between system identification and black-box model. In this section, they will be first discussed respectively and then be compared with each other.

2.1 System Identification

In system identification, the requirement that the data should be informative means that the inputs should contains sufficiently many distinct frequencies. The Fisher information matrix is usually used to identify the model parameters, and guide to design experiment for training data collecting (Ljung, 1998). Using the condition number of the Fisher information matrix can determine if the data are sufficiently rich in frequency for identification of a given model, and also can determine the maximum model orders. Cai et al. (2016) generated an optimal training data set for varying zone temperature set-points that maximizes the accuracy of parameter estimates for an intended building model structure. Optimal zone air temperature set-points were determined to maximize the Fisher information matrix. The required size of the training data could be dramatically reduced to achieve a certain accuracy level, which leads to more cost-effective experiments.

To be more specific, deducted from the theory of maximizing Fisher Information Matrix, we should spend the input power at frequencies where the model frequency response is sensitive to parameter variations. Put more leisurely, if a parameter is of special interest, then vary it and check where the frequency response moves, and put the input power there. In many cases this may give sufficient guidance for good input design.

However, when considering cost-effectiveness, frequency alone is not enough to ensure a good experiment design. The crest factor gives an idea of the compactness of the signal. Signals with an impulsive behavior (having a large crest factor) inject much less power into the system than signals having the same peak value and a small crest factor. The crest factor Cr is defined as:

$$Cr^2 = \frac{max_t u^2(t)}{lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N} u^2(t)} \tag{1}$$

$max_t u^2(t)$ is the peak value of signal u, and $lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N} u^2(t)$ is the average value of signal u. A good signal waveform is consequently one that has a small crest factor. The theoretic lower bound of Cr clearly is 1, which is achieved by binary, symmetric signal: $u(t) = \pm\bar{\mathrm{u}}$. This gives a theoretical advantage for binary signals. Braun et al. (2002) demonstrate the application of minimum crest factor signals to accomplish meaningful "plant-friendly" identification and control in the process industries.

The basic issue for design of input signal is clear: achieve a desired input spectrum for a signal with as small crest factor as possible. Unfortunately, these properties are somewhat in conflict: If it is easy to manipulate a signal's spectrum, it tends to have a high crest factor and vice versa. To resolve the conflict between informative frequency and crest factor, Multi-sines signal and Pseudo-Random Binary signals are most widely used.

Multi-sines signal is defined as:

$$u(t) = \sum_{k=1}^{d} a_k \cos(\omega_k t + \Phi_k) \tag{2}$$

d is number of sinusoid; $a_k$is amplitude; $\omega_k$is frequency; and $\Phi_k$is phase. With d, a_k and ω_k we can thus place the signal power very precisely to desired frequencies. The crest factor is also controllable through choosing the phase Φ_k so that the cosines are "as much out of phase" as possible. Li and Wen (2014) applied Multi-sines signal to excite a system identification model used for short-term building energy forecasting. To guarantee those signals cover the frequency domain around the building, common operation range, Multi-sines signal was used to generate the exciting signal. He decided the parameters of the signal to put power to desired frequency and at the same time minimize the crest factor. Similarly, Privara et al. (2011) presented a comprehensive study of modeling of a large multi-zone office building by physical modeling and RC modeling. In RC modeling part, PRBS signal, Multi-sines signal and MPRS are proposed in black-box subspace identification. Signals are designed to cover required frequency spectrum and meet the demand of switching time.

Pseudo-Random Binary Signal (PRBS) is a periodic, deterministic signal with white-noise-like properties. Since the signal is a deterministic, it can be selected in accordance with the interesting part of the frequency scale of variation. It is generated by the difference equation:

$$u(t) = rem(A(q)u(t), 2) = rem(a_1 u(t-1) + \cdots + a_n u(t-n), 2) \tag{3}$$

Here $rem(x, 2)$ is the remainder as x is divided by 2. Like white random binary noise, PRBS has an optimal crest factor. PRBS has advantages over random binary noise: its covariance matrix will have a very special pattern. It can be analytically inverted, which will facilitate certain computations; PRBS has its second order properties secured when evaluated over whole periods. However, the following caution should be mentioned: A binary input will not allow validation against nonlinearities. There is consequently no way to detect that such a non-linearity is present from a binary input. In light of these advantages, it is widely used in experiment design in building models. Ma et al. (2012) used pseudorandom binary sequence to excite Autoregressive exogenous (ARX) models to build a model with inputs of zone temperature set-points of the cooling system and outputs of actual zone temperature and power measurements. A pseudorandom binary sequence (PRBS) is generated as the excitation input to randomly adjust cooling set-point between 21℃ and 25℃. Bacher and Madsen (2011) suggested a procedure for identification of suitable models for the heat dynamics of a building. The controlled heat input is a PRBS, which has white noise properties and no correlation with the other inputs. It is designed to excite the heat dynamics at several ranges of frequencies in which the time constants of the building is expected to be, such that the information embedded in data is optimized for estimation of the heat dynamic properties of the building. Lara et al. (2015) used PRBS signals as the excitation input signal, a linear time invariant (LTI) model to predict return water temperature of a central chiller plant was identified using grey-box technique. Zubiaga et al. (2015) presented the application of a grey box model to evaluate the thermal performance of a reference social housing building, focusing on its potential to evaluate the thermal performance of building passive elements. The controlled heat input (on and off status) was designed as a combination of a Randomly Ordered Logarithmic distributed Binary Sequence and a Pseudorandom Binary Sequence.

The PRBS input, however, is not always well suited for nonlinear problems. Since the PRBS consists of only two levels, the resulting data may not provide sufficient information to identify nonlinear behavior. Additionally, a PRBS signal of too large a magnitude may bias the estimation of the linear kernel. Multi-level pseudo-random sequences (m-level PRS, or MPRS), in contrast, allow the user to highlight nonlinear system behavior while manipulating the harmonic content of the signal to enable unbiased estimation of the linear dynamics in the presence of nonlinearities (Schoukens et al., 1993). Privara et al. (2011) applied MPRS to excite zone temperatures of a large multi-zone office building.

To sum up, Multi-sines, PRBS and MPRS are widely used in the process of experiment design in building energy modeling: Multi-sines can easily place the signal power very precisely to desired frequencies but with poor crest factor and complex design process; PRBS has an ideal crest factor but with poor nonlinearity application and required period length; MPRS has better nonlinearity application but with complex design process.

Input design should be decided in consideration of model structure and characteristics. If there is a high demand for frequency, Multi-sines signal has its advantage over PRBS and MPRS. If cost-effectiveness matters for the input acquisition, PRBS should be considered first. If the model has severe nonlinearity, Multi-sines and MPRS should be used. The comparison among Multi-sines, PRBS and MPRS in terms of frequency, crest factor, nonlinearity, period length and design complexity is shown in Table 1.

Table 1: Comparison among Multi-sines, PRBS and MPRS

	Multi-sines	**PRBS**	**MPRS**
Frequency	Good	Poor	Poor
Crest Factor	Poor	Perfect	Good
Nonlinearity	Medium	Poor	Good
Period Length	Not required	Required	Not Required
Design Complexity	High	Low	High

The sampling intervals are also important in experiment design: very fast sampling leads to numerical problems, model fits in high-frequency bands, and poor returns for extra work; slower sampling leads to data sets that are subsets of the maximal one, and hence is less informative. As the sampling interval increases over the natural time constants of the system, the variance increases drastically. Optimal choices of T for a fixed number of samples will lie in the range of the time constants of the system. These are, however, not exactly known, and overestimating them may lead to very bad results. Ljung (1998) summarized that a sampling frequency that is about ten times the bandwidth of the system should be a good choice in most cases. However, if the model should be used for control purpose, certain other aspects will enter. The sampling interval for which we build the model should be the same as for the control application. For example, in the building energy forecasting models for building-to-grid integration like demand response, the interval for control is about 15 minutes to 1 hour (Li and Wen, 2014 and Ma et al., 2012), which is the same as the sampling intervals. In the application of zone temperature prediction models, the interval for control is about 5 minutes (2011), and the sampling intervals should be the same.

"What inputs to excite" is another important topic. Modelers should choose the outputs and the inputs so that the predicted output becomes sensitive with respect to parameters that are important for the application in question. The input should have validation power to test possible properties of interest in the system. In most of the papers related to building energy forecasting, zone temperature set-points are usually used as the input to be manipulated. Besides, occupancy (Richardson et al., 2008 and Menezes et al., 2012) and ventilation flow rate are also manipulated in some studies to train building energy models. Other operational inputs like supply air temperature and pressure set points, and chilled water temperature set point have been hardly discussed, which could be the potential topic in the future study.

2.2 Black-Box models

Experiment design in system identification models can be improved by using theories and index like Fisher information matrix and crest factor. In black-box modeling, training data decide most of model information, including properties of linearity and nonlinearity, model structure and its parameters. Also, the diversity of inner structure of black-box models make it hard to have a universal criterion or index for experiment design.

In terms of the topic of "what to excite", it is termed as Feature Selection Process (FSP) in black-box modeling. Canonical Correspondence Analysis (CCA) (Härdle and Simar, 2012), Principal Component Analysis (PCA) (Abdi and Williams, 2010), and Entropy-based Feature Selection are used to find out the most important input variables for the modelling techniques. Jurado et al. (2015) proposed a hybrid methodology that combines feature selection based on entropies with soft computing and machine learning approaches, i.e. Fuzzy Inductive Reasoning, Random Forest and Neural Networks. They are also compared with a traditional statistical technique ARIMA (AutoRegressive Integrated Moving Average). Che et al. (2014) proposed a kernel-based SVR combination model by using a novel individual model selection algorithm to get the optimal kernel function of STLF problem in the field of short-term load forecasting. Kusiak et al. (2010) proposed a correlation coefficient matrix and the boosting tree algorithm for variable selection when predicting daily steam load of buildings by a neural network ensemble with five Multi-Layer Perceptrons (MLPs) methods. Fan et al. (2014) used the recursive feature elimination technique, an embedded variable selection method, to select the optimal inputs to the base prediction models developed separately using eight popular predictive algorithms when developing ensemble models for predicting next-day energy consumption and peak power demand, with the aim of improving the prediction accuracy. Li et al. (2010) used an improved PCA, called Kernel Principal Component Analysis (KPCA), before training SVMs to predict building cooling load.

In terms of the topic of "whether training is sufficient", usually what modelers do is to run a benchmark model on the training data sets of different sizes, and evaluate the model performance with some metric, for example, the NRMSE, and find the best size. Because different problems may have different design space and different properties, it is difficult to set a universal threshold for determining the training size. Usually researchers will just assume their selected training data is adequate for prevention of under-training, and they will also add cross validation for prevention of over-training. There is also a rule of thumb about the number of training data in machine learning: roughly 10 times as many examples as there are degrees of freedom in the model. Thinking in terms of dimensionality and number of examples is a convenient shortcut.

Data mining techniques are used in improving training data quality in building energy forecasting. Fan et al. (2014) proposed a data mining based outlier detection method to identify abnormal building operating data, which improve the quality of training data.

The topic of pre-training is widely discussed in many researches, including the field of energy forecasting. Erhan et al. (2010) empirically show the influence of pre-training with respect to architecture depth, model capacity, and number of training examples. The observations so far in this paper confirm that starting the supervised optimization from pre-trained weights rather than from randomly initialized weights consistently yields better performance. Pasa et al. (2015) proposed a novel approach to pre-training sequential neural networks that exploits a simpler, first-order Hidden Markov Model to generate an approximate distribution of the original dataset. The learned distribution is used to generate a smoothed dataset that is used for pre-training to drive the connection weights in a better region of the parameter space, where subsequent fine-tuning on the original dataset can be more effective.

To the best of our knowledge, methods to improve the quality of modeling and training data in black-box modeling are limited to feature selection process, data-mining techniques and pre-training,. The focus in black-box modeling has been on how to find the best model (structure and parameters) to represent an existing data set. There is a lack of investigation on how different training data sets would affect a black-box model's performance, especially when the training data is to represent a nonlinear system. Considering that training data sets obtained from a system identification process, i.e., an excitation process, describe a nonlinear system in a more completed sense, we are interested in here the impacts of using such training data sets on a black-box model's performances, when compared with traditional training data set obtained from typical building operation.

3. MATERIAL AND METHODS

3.1 Research Method

The main goal of this research is, in the field of building energy forecasting modeling, to understand the impacts of using training data obtained from well-designed excitation tests which satisfy the maximum Fisher information matrix as well as the minimum crest factor, on the performance of a black-box model. In this study, the training data obtained by exiting building zone temperature set points using Multi-Sine signals are compared with normal operation data without excitation. The black-box models included in this study are Kriging, Support Vector Regression, Radial Basis Function, Multivariate Adaptive Regression Splines, Artificial Neural Network and Polynomial Regression. A system identification model (SID) using frequency is also included in the comparison study.

In lieu of a real building, a virtual building testbed described in Li wand Wen 2014, which is modeled in EnergyPlus, is used to generate all of the training data sets. The first floor of this reference medium size office building has five zones, and the total floor area is 510 m^2. The window-to-wall ratio of this building's facades is approximately 21.2%, and the windows are equally distributed. The U-factor of these single pane windows is3.4 W/m^2K and the solar heat gain factor is 0.36. The solar absorptivity, transmissivity and reflectivity are 0.06, 0.697 and 0.243, respectively. The location of this building selected for this study is Philadelphia, PA, USA. This building is partitioned into five different air conditioning zones, and an unconditioned attic zone. The five conditioned zone are four perimeter zones and one core zone. The roof insulation has an R-value of 15. The roof is covered in an

asphalt membrane, with a solar absorptivity value of 0.9. The exterior wall has the following construction (from outside layer to inside layer): 2.54 cm of stucco; 20.32 cm concrete masonry units; R-6 continuous insulation; 1.27 cm gypsum wallboard.
When comparing the models, the performance of accuracy is measured using Normalized Root Mean Square Error (NRMSE), where

$$\mathrm{NRMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{N}}/(y_{max}-y_{min}), \tag{4}$$

and y is the true value of the building energy consumption, $\hat{y}$ is the forecast value and N is the number of data sets.
Training error and prediction error will be calculated to show the accuracy of each model. Training error means the error in model building process, illustrating how well the model can be trained to reflect the real building data. Prediction error shows how powerful is the model to predict the future cooling energy by using the given inputs.
Two training datasets, i.e., normal operational data and excited operational data are generated. The normal operational data are generated by allowing the virtual building to be operated like a typical commercial building with very limited zone temperature set point variation. Weather inputs from July 1st to 7th are used to generated this seven day normal operation training data set. The exited training data set is described in Sec. 3.2. Seven days of training data are generated using the virtual testbed for both cases.
Two three-day testing datasets, i.e., similar weather testing and extended testing, are used to examine the model performances. In the similar weather testing, a testing period in which the weather conditions are similar to those during the training period is selected. In the extended testing, a testing period in which the weather conditions are very different from those in the training period is selected to examine how accurate a model is. To be more specific, in this TMY3 weather data in Philadelphia, average temperature of seven-day training data (July 1st to 7th) is 24.7℃ and direct solar radiation is 220.5W/m^2. The average outdoor air temperature of similar weather testing is 25.9℃ and direct solar radiation is 210.3W/m^2 (July13th to 15th). The extended weather testing is under weather condition of 28.0℃ and 277.9 W/m^2 (August 16th to 18th), which is hotter than the situation in training data.
Each model is firstly trained with normal operational data and then with the exited operational data, marked as Model I and II, respectively. Each model is then tested using the two testing datasets. The NRMSEs of each model for each testing period, as well as during the two training periods, are compared to examine the model accuracy and extendibility.

3.2 Excited Training Data Algorithms and Acquisition

The inputs used in all of the models in this study are zone temperature, ventilation rate, occupancy schedule ratio. Model disturbances include: outdoor air temperature, direct solar radiation and diffuse solar radiation. Output of the model is zone heating/cooling energy.
Zone temperature set-points are excited when generating the exited training data. The excitation method is a typical Multi-Sine signal generation process, which is to place the signal power very precisely to desired frequencies first, and then minimizing the crest factor. The excitation signal generation procedure is shown in Figure 1.

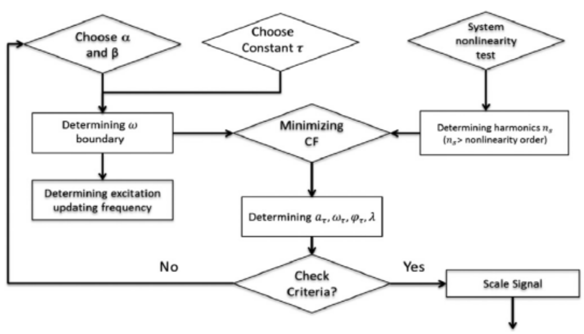

Figure 1: Excitation signal generation procedure (Li and Wen, 2014)

Matlab is used in this study to execute the algorithms to generate excited zone-temperature temperature set-points. A virtual test bed BCVTB connects Matlab to EnergyPlus is used to manipulate the zone temperature set-points in the virtual building testbed in EnergyPlus. According to the theory mentioned in Section 2.1, the sampling interval for which we build the model should be the same as for the control application. In demand response application of building-to-grid operation, the interval of 30 minutes is normally used, thus making the sampling interval of the model to be 30 minutes.

3.3 SID and Black-Box Model Description

Six data-driven models are chosen to test the adaptability of well-trained data from SID model in black-box modeling. They are Kriging, Support Vector Regression, Radial Basis Function, Multivariate Adaptive Regression Splines, Artificial Neural Network and Polynomial Regression.

Kriging (also known as Gaussian process regression) is an interpolation method that assumes the simulation output may be modeled by a Gaussian process. It gives the best linear unbiased prediction of simulation output not yet observed. It generates the prediction in the form of a combination of a global model with local random noise.

Support Vector Regression (SVR) is analogous to support vector classification, which attempts to maximize the distance between two classes of data by selecting two hyperplanes to optimally separate the training data.

Radial Basis Function (RBF) is used to develop interpolation on scattered multivariate data. A RBF is a linear combination of a real-valued radially symmetric function, $\emptyset(x)$, based on distance from the origin.

Multivariate Adaptive Regression Splines (MARS) is a form of regression analysis introduced by Friedman (1991). A set of basis functions, defined as constant, hinge function, or the product of two or more hinge functions, are combined in the weighted sum form, to be the approximation of the response function. A MARS model is built with generalized cross validation regularization in a forward/backward iterative process.

Artificial Neural Network (ANN) is a computational model inspired by an animal's central nervous system. It is apt at solving problems with complicated structures. Due to its promising results in numerous fields, ANN has been extensively applied in stochastic simulation modeling. An ANN model typically consists of three separate layers: the input layer, the hidden layer(s), and the output layer. The neurons across different layers are interconnected to transmit and deduce information. In supervised learning, the output unit is trained to simulate the underlying structure of the input signals and response. The trained structure is depicted by several parameters, the weights on each connection, the biases, the number of hidden layers, the transfer functions, and the number of hidden nodes in each hidden layer.

Polynomial Regression (PR) is a variation of linear regression in which an n^{th} order polynomial is modeled to formulate the relationship between the independent variable x and the dependent variable y. PR models have been applied to various engineering domains such as mechanical, medical and industrial

The system identification model in this paper is derived from the research of Li and Wen 2014. Their study proposed a novel methodology to develop building energy estimation models for on-line building control and optimization using a system identification approach. Frequency domain spectral density analysis is implemented in this on-line modeling approach to capture the dynamics of building energy system and forecast the energy consumption with more than 90% accuracy and less than 2 min computational speed.

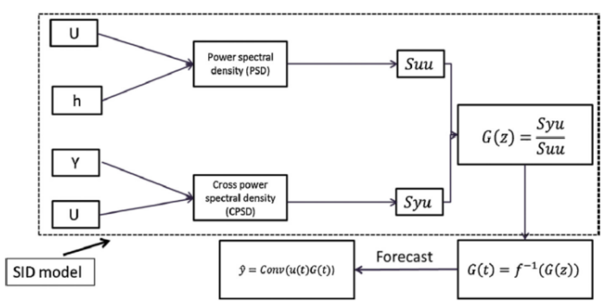

Figure 2: Model structure of system identification method for building energy forecasting (Li and Wen, 2014)

4. Results and Discussion

As mentioned in Section 3, seven models (one SID and six black-box models) and two sets of experiment will be conducted. Two sets of experiments include experiment using normal operation data (non-excited set-points) and well-designed excitation signal in each of the seven models. NRMSE will be calculated in both training period and testing period and it will be used as the criterion of training and prediction accuracy in the following discussion. Table 2 shows the experiment structure as well as the results of the simulation.

Table 2: NRMSE of three testing of two models by seven algorithms

	NRMSE													
Algorithms	**SID[3]**		**Kriging**		**SVR**		**RBF**		**MARS**		**ANN**		**PR**	
Model	I[1]	II[2]	I	II	I	II	I	II	I	II	I	II	I	II
Testing using training data	0.203	0.147	0.000	0.000	0.163	0.180	0.000	0.000	0.039	0.066	0.036	0.083	0.071	0.099
Testing under similar weather	0.153	0.090	0.048	0.098	0.246	0.268	0.231	0.275	0.076	0.171	0.066	0.162	0.080	0.102
Testing under extended weather	0.160	0.089	0.074	0.090	0.250	0.259	0.189	0.199	0.119	0.175	0.092	0.159	0.085	0.093

1 Model I stands for models that trained with normal operation data
2 Model II stands for model that trained with excited operation data
3 SID stands for system identification method by frequency domain spectral density analysis

To compare among seven algorithms, four scenarios are considered: model trained by normal operation data with the best testing result under similar weather testing is Kriging; model trained by normal operation data with the best testing result under extended weather is also Kriging; model trained by excited operation data with the best test result under similar weather testing is SID by frequency domain spectral density analysis; and model trained by excited operation data with the best testing result under extended weather is also SID.
To analyze the simulation results of Model I: by using normal operation training data to train models, most of the black-box models have higher prediction accuracy than SID model (17% better). It performs well when forecasting condition under weather similar to training data, which is better than that under weather much hotter than training data (7% better).
To analyze data of Model II: by using zone-temperate-excited training data to train models, the SID model has higher accuracy than most of the black-box models (48% better). Except RBF model, the rest models perform well under both similar weather and extended weather.
To compare the data in Model I and Model II, we found that training a SID model by using excited training data can raise the accuracy of forecasting (43% better). However, for black-box models, training them by normal operation data can achieve better model accuracy than training them by excited inputs and output (32% worse). As is shown in Figure 3, take ANN as an example, model that trained by normal operation data has much better performance than that trained by excited training data. The fluctuation makes the model trained by excited data unable to capture frequent fluctuation of cooling energy caused by excited zone temperature set-point.

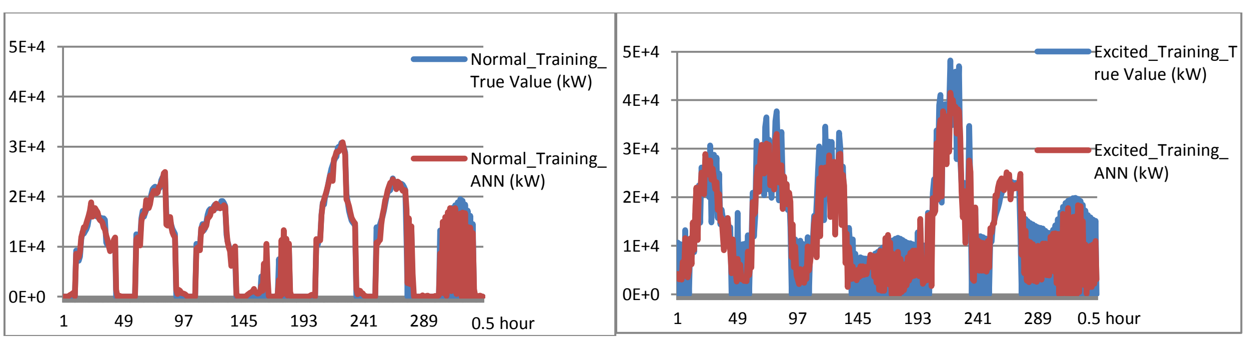

Figure 3. Result of training models excited by normal operation data and excited data in ANN

In the validation period, we could see that model trained by normal operation data predict more accurately than that trained by excited training data, especially when predicting the fluctuation happening in the second day of validation. The second model performs poorly to capture the dynamics. Also, the second model cannot capture the energy consumption pattern during the equipment starting up period.

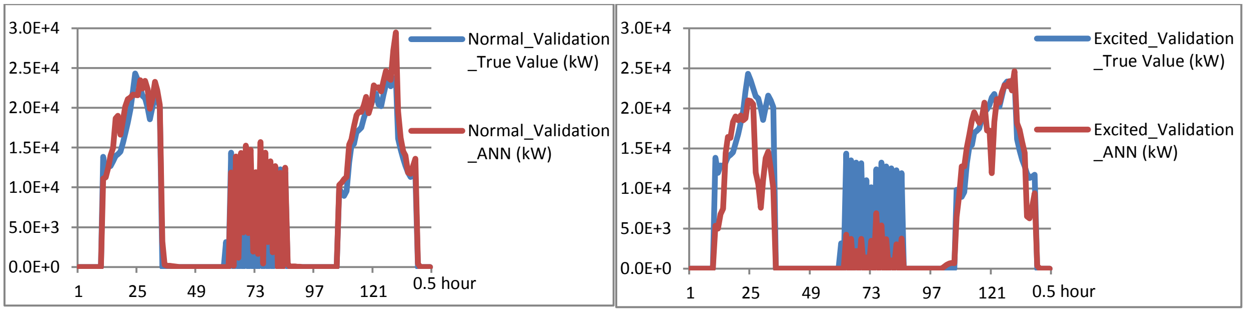

Figure 4. Result of validating models excited by normal operation data and excited data in ANN

To explain this, the experiment design and the excitation theory are used to excite the model to reveal its parameters which are important to the model. However, black-box models do not have parameters with significant meaning. Most black-box models learn only from training data. Excitation will make it hard to capture the normal pattern of the input-output relation, especially in the situation when handling fluctuation of energy and starting up of equipment. Too frequent fluctuation in the excitation input will hinder the black-box model to capture the pattern of equipment starting up period in the real operation. Also frequent on-and-off situation is also hard be detected and well-described in the model trained by exited data.

We also find that in SID models trained by excited data, the difference between NRMSE of similar weather and extended weather is only 0.001. However, this value turns to be 0.024 and 0.018 in black-box models when trained by normal operation data and excited data respectively, which demonstrates the fact that SID models have better extendibility than black-box models. Though the excited training data increase the extendibility of black-box model by 25% but decrease the model accuracy by 32%.

5. CONCLUSIONS

A comparison simulation study using virtual building testbed is conducted to study whether the training data obtained from theory of excitation in SID models can improve the performance of other types of black-box models in the building energy forecasting area. Results show that experiment design, or excitation, is essential to SID modeling: it can increase model accuracy of building energy models. However, for other tested data-driven energy forecasting models, including Kriging, Support Vector Regression, Radial Basis Function, Multivariate Adaptive Regression Splines, Artificial Neural Network and Polynomial Regression, training data obtained from the excitation scheme used in this study can increase the model extendibility. But the overall model accuracy is not improved. In other words, compared with excited operation data generated using the theory of system identification, normal operation data are more suitable for model training in black-box modeling if not considering extendibility.

Future work could further study and compare excitation from different theories, or design specific excitation algorithm that is suitable for black-box models, to increase black-box model accuracy as well as extendibility.

REFERENCES

Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433-459.

Ahmad, M. W., Mourshed, M., Mundow, D., Sisinni, M., & Rezgui, Y. (2016). Building energy metering and environmental monitoring–A state-of-the-art review and directions for future research. Energy and Buildings, 120, 85-102.

Bacher, P., & Madsen, H. (2011). Identifying suitable models for the heat dynamics of buildings. Energy and Buildings, 43(7), 1511-1522.

Braun, M. W., Ortiz-Mojica, R., & Rivera, D. E. (2002). Application of minimum crest factor multisinusoidal signals for “plant-friendly” identification of nonlinear process systems. Control Engineering Practice, 10(3), 301-313.

Cai, J., Kim, D., Braun, J. E., & Hu, J. (2016). Optimizing Zone Temperature Setpoint Excitation to Minimize Training Data for Data-driven Dynamic Building Models. In American control conference, submitted, Boston, July (pp. 6-8).

Che, J., & Wang, J. (2014). Short-term load forecasting using a kernel-based support vector regression combination model. Applied Energy, 132, 602-609.
Dong, B., Cao, C., & Lee, S. E. (2005). Applying support vector machines to predict building energy consumption in tropical region. Energy and Buildings, 37(5), 545-553.
Erhan, D., Bengio, Y., Courville, A., Manzagol, P. A., Vincent, P., & Bengio, S. (2010). Why does unsupervised pre-training help deep learning?. The Journal of Machine Learning Research, 11, 625-660.
Fan, C., Xiao, F., & Wang, S. (2014). Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Applied Energy, 127, 1-10.
Härdle, W. K., & Simar, L. (2012). Canonical correlation analysis. In Applied Multivariate Statistical Analysis (pp. 385-395). Springer Berlin Heidelberg.
Jurado, S., Nebot, À., Mugica, F., & Avellana, N. (2015). Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques. Energy, 86, 276-291.
Kusiak, A., Li, M., & Zhang, Z. (2010). A data-driven approach for steam load prediction in buildings. Applied Energy, 87(3), 925-933.
Lara, B. G. V., Molina, L. M. C., & Yanes, J. P. M. (2015). Modeling and identification of the cooling dynamics of a tropical island hotel. Energy and Buildings, 92, 19-28.
Li, X., & Wen, J. (2014). Building energy consumption on-line forecasting using physics based system identification. Energy and Buildings, 82, 1-12.
Li, X., & Wen, J. (2014). Review of building energy modeling for control and operation. Renewable and Sustainable Energy Reviews, 37, 517-537.
Li, X., Wen, J., & Bai, E. W. (2016). Developing a whole building cooling energy forecasting model for on-line operation optimization using proactive system identification. Applied Energy, 164, 69-88.
Ljung, L. (1998). System identification (pp. 163-173). Birkhäuser Boston.
Ma, J., Qin, J., Salsbury, T., & Xu, P. (2012). Demand reduction in building energy systems based on economic model predictive control. Chemical Engineering Science, 67(1), 92-100.
Menezes, A. C., Cripps, A., Bouchlaghem, D., & Buswell, R. (2012). Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. Applied Energy, 97, 355-364.
Newsham, G. R., & Birt, B. J. (2010, November). Building-level occupancy data to improve ARIMA-based electricity use forecasts. In Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building (pp. 13-18). ACM.
Pasa, L., Testolin, A., & Sperduti, A. (2015). Neural Networks for Sequential Data: a Pre-training Approach based on Hidden Markov Models. Neurocomputing, 169, 323-333.
Prívara, S., VáHa, Z., Gyalistras, D., Cigler, J., Sagerschnig, C., Morari, M., & Ferkl, L. (2011, September). Modeling and identification of a large multi-zone office building. In Control Applications (CCA), 2011 IEEE International Conference on (pp. 55-60). IEEE.
Richardson, I., Thomson, M., & Infield, D. (2008). A high-resolution domestic building occupancy model for energy demand simulations. Energy and buildings, 40(8), 1560-1566.
Schoukens, J., Guillaume, P., & Pintelon, R. (1993). Perturbation signals for system identification. Prentice Hall, New York.
Terés-Zubiaga, J., Escudero, C., García-Gafaro, C., & Sala, J. M. (2015). Methodology for evaluating the energy renovation effects on the thermal performance of social housing buildings: Monitoring study and grey box model development. Energy and Buildings, 102, 390-405.
Xuemei, L., Lixing, D., Jinhu, L., Gang, X., & Jibin, L. (2010, January). A novel hybrid approach of KPCA and SVM for building cooling load prediction. In Knowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on. IEEE.
Yang, J., Rivard, H., & Zmeureanu, R. (2005). On-line building energy prediction using adaptive artificial neural networks. Energy and buildings, 37(12), 1250-1259.

ACKNOWLEDGEMENT

This research was supported by project: CPS: Synergy: Collaborative Research: GOALI: SMARTER - Smart Manager for Adaptive and Real-Time decisions in building clustERs. It was also supported by U.S. Department of Energy's Building Grid Integration Research and Development Innovators Program: VOLTTRON Compatible Short-term Transactive Load Control Strategy.