Huang, S., Zuo, W., Sohn, M. "A Bayesian Network Model for Predicting Cooling Load of Commercial Buildings," *Building Simulation, 11 (1), pp. 87-101, 2018.*

A Bayesian Network Model for Predicting Cooling Load of Commercial Buildings

Sen Huang [a+] , Wangda Zuo [a,] *, Michael D. Sohn [b]

[a] Department of Civil, Architectural and Environmental Engineering, University of Miami, 1251 Memorial Drive, Coral Gables, FL 33146, U.S.A.
[b] Energy Analysis and Environmental Impacts Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, U.S.A.
[+] Current Address: Electricity Infrastructure and Buildings Division, Pacific Northwest National Laboratory, 902 Battelle Boulevard, WA 99354, U.S.A.

Abstract: Cooling load prediction is indispensable to many building energy saving strategies. In this paper, we proposed a new method for predicting the cooling load of commercial buildings. The proposed approach employs a Bayesian Network model to relate the cooling load to outdoor weather conditions and internal building activities. The proposed method is computationally efficient and implementable for use in real buildings, as it does not involve sophisticated mathematical theories. In this paper, we described the proposed method and demonstrated its use via a case study. In this case study, we considered three candidate models for cooling load prediction and they are the proposed Bayesian Network model, a Support Vector Machine model, and an Artificial Neural Network model. We trained the three models with fourteen different training data datasets, each of which had varying amounts and quality of data that were sampled on-site. The prediction results for a testing week shows that the Bayesian Network model achieves similar accuracy as the Support Vector Machine model but better accuracy than the Artificial Neural Network model. Notable in this comparison is that the training process of the Bayesian Network model is fifty-eight times faster than that of the Artificial Neural Network model. The results also suggest that all three models

will have much larger prediction deviations if the testing data points are not covered by the training dataset for the studied case (The maximum absolute deviation of the predictions that are not covered by the training dataset can be up to 7 times larger than that of the predictions covered by the training dataset). In addition, we also found the uncertainties in the weather forecast significantly affected the accuracy of the cooling load prediction for the studied case and the Support Vector Machine model was more sensitive to those uncertainties than the other two models.

Keywords: Bayesian Network Model, Cooling Load Prediction, Training Dataset, Uncertainties

1. Introduction

In the U.S., building sector accounted for the largest portion of the primary energy consumption in 2010 (U.S. Department of Energy). Furthermore, building energy use is expected to rise by ~31% from 2010 to 2030 (U.S. Department of Energy; The U.S. Energy Information Administration). Thus, even small reductions in building energy used can bring great positive benefits to U.S.'s primary energy use. In fact, many studies have been reported in the literature, which reduce the primary energy use through building energy efficiency measures (Xue et al. 2014; Hughes et al. 2015; Hao et al. 2016; Alajmi 2012; Krati 2016; Corbin et al. 2013; Široky et al. 2011; Ma et al. 2012). Those methods include demand response strategies (Xue et al. 2014; Hughes et al. 2015; Hao et al. 2016), energy audit strategies (Alajmi 2012; Krati 2016), and advanced control strategies (Wetter et al. 2016; Huang, Zuo, et al. 2016a; Ma et al. 2012; Huang et al. 2014; Huang, Zuo, et al. 2016c). To assure the successful application of all these proposed strategies, or for the verification of their implementation (Walter and Sohn 2016), an accurate prediction of building cooling load is necessary (Li et al. 2013). For example, in the demand response strategy proposed by Hao et al. (2016), the predicted cooling load is required to determine the set points for the temperature of each thermal zones. In Krati (2016)'s energy audit study, the cooling load is necessary in predicting the energy saving from different energy saving methods. The predicted cooling load is also a critical input for the model

predictive control strategy proposed by Huang, Zuo, et al. (2016a). This strategy can generate the optimal set points for the future time horizons.

Predicting the building cooling load, however, can be difficult. The challenges come from two aspects: first, building cooling load can be affected by countless factors, including weather, internal activities, and occupant preferences (Kim 2011). Considering all those factors simultaneously requires a lot of detailed information regarding buildings. However, this information may not be assessable or is hard to quantify. Second, the relationship between the factors and the cooling load is a complicated non-linear function which is difficult to be described by the commonly used linear regression (Hou et al. 2006). The complexity of the relationship is mainly due to the highly non-linear nature of the building system. For example, the heat transfer between the ambient environment and the building via radiation is governed by the Stefan-Boltzmann law described by a non-linear equation.

Currently, three broad methods to predict the cooling load have been reported in the literature. In the first method (Eskin et al. 2008; Thevenard et al. 2006), building energy simulation tools such as DOE-2 (Birdsall et al. 1990), EnergyPlus (Crawley et al. 2001), and TRNSYS (Klein et al. 1976) are employed to predict the cooling load based on a physical description of the buildings and surrounding environment. In this physical description, algebraic equations and/or differential equations are usually used to represent the complicated relationships between cooling load and other variables. Due to its explicit nature, this physical description is usually named as a "white-box" model. To achieve an accurate white-box model, detailed specifications of building characteristics, building operation schedules, and occupant behavior, are required.

In the second method, purely data-driven models ("black-box" models) were developed to predict the cooling load according to the pre-defined factors. Those black-box models include Artificial Neural Network models (Kashiwagi et al. 1993; Sakawa et al. 1999; Ben-Nakhi et al. 2004; Kwok et al. 2011;

Leung et al. 2012; Deb et al. 2016) and Support Vector Machine models (Hou et al. 2009; Li et al. 2009a; Li et al. 2009b; Chen et al. 2017; Zhang et al. 2016). Some researchers (Yao et al. 2004; Hou et al. 2006; Li, Ding, Li, et al. 2010; Li, Ding, Lv, et al. 2010) also attempted to achieve a better performance by combining multiple black-box models. For those black-box models, significant amount of training data is usually required to achieve a desired accuracy.

The third method (Braun et al. 2002; Sun, Wang, et al. 2013) is to utilize "gray-box" models that are hybrids of the white-box models and the black-box models. In the gray-box models, thermal network models, which simplify the energy flows in the buildings (such as the heat transfer through the building envelope), are usually employed to calculate the net energy requirement to achieve desired zone temperatures. The values of parameters in the network models are estimated both from rough building descriptions and the optimizations. The optimizations aim to minimize the difference between the outputs of the gray-box model and the training dataset by modulating those parameters. The gray-box models require less building information than the white-box model and fewer training data than the black-box model.

From the large-scale application points of view, black-box models may be highly promising because of the expectation that more building data are becoming available. Additionally, black-box models do not require detailed building information, which may be difficult, sometimes even impossible, to gather, due to the time and/or cost constraints. Second, black-box models are more cost-effective for implementation than the other two types of models. After trained, black-box models can predict the cooling load with very little computational resource demand and fast speed (Li et al. 2013). This feature dramatically lowers the requirement of the hardware in which the prediction models are implemented.

Nevertheless, there are still two problems, which prevent black-box models from being widely adapted in the real-world. Those problems are: first, it is difficult to use those black-box models without relevant background. Black-box models are usually built on sophisticated mathematical theories. For example, in

the Support Vector Machine model proposed by Hou et al. (2009), a dispensable step is to use a set of hyperplanes to classify the training data. However, a hyperplane is not a familiar concept to average practitioners in buildings industry, not to mention for them to regulate the parameters of hyperplane. However, the accuracy of those models is usually sensitive to the parameters (Chapelle et al. 2002). Second, it is lack of quantitative descriptions regarding how the amount of training data affects the performance of different models. Those descriptions can help researchers to identify the best amount of training data, which balances the accuracy and the efforts for preparing the training data.

Besides the two problems mentioned above, there is also another unanswered question: how the uncertainties in the weather prediction affect the results of the cooling load. The weather condition is considered as an important factor for the cooling load (Walter, Price, et al. 2016). In the real implementation, the forecasted weather condition from the weather service providers is usually employed. However, uncertainties in weather forecast are inevitable. The uncertainties may be due to the limitations of the weather forecast models (Gneiting et al. 2005), or the micro-climate effect (Gneiting et al. 2005). Although there are studies aimed to quantify the impacts of the uncertainties in weather data (Sun, Heo, et al. 2013), we did not find the relevant research in the cooling load prediction field. However, to identify how the uncertainties affect the prediction is very important since it can help people to determine an appropriate weather prediction service to pursue.

In this paper, to address the first problem, we developed a new black-box model (Bayesian Network model). The Bayesian Network model is a probability-based graphic model, which is very suitable for non-linear systems. One advantage of the Bayesian Network is that it doesn't require significant efforts to understand, and thus are suitable for large-scale applications. To deal with the second problem, we performed a case study to evaluate the performance of the Bayesian Network model. In this case study, onsite measurement data was used to train and test the Bayesian Network model. To compare the performance of the Bayesian Network model with those of existing black-box models, we also included a Support Vector Machine model

and an Artificial Neural Network model in the case study. We evaluated the three models trained by various training datasets, which had different amounts of data. To handle the last problem, we also quantitatively assessed how the uncertainties in weather forecast affect the cooling load prediction.

This study advances the body of science from four aspects: first, a new Bayesian Network model for cooling load prediction of commercial buildings is proposed. Second, a systematic comparison of the Bayesian Network model with the Support Vector Machine model, and the Artificial Neural Network model, in terms of the prediction accuracy and the training time cost, is performed. Third, insights regarding how the amount of training data affect the cooling load prediction are provided. Lastly, a quantitative assessment on how the uncertainties in weather forecast affect the cooling load prediction is conducted, which is the first one to our best knowledge.

2. Bayesian Network Model

In this section, we present the proposed Bayesian Network model. Firstly, we introduce the theory of the Bayesian Network model is built. Second, we discuss how to develop a Bayesian Network model for building-related applications, generally. Lastly, we discuss how we capitalize on attributes of the Bayesian Network model for predicting the cooling load of a commercial building.

2.1 Theory

Bayesian Network models are probability-based graphic models and have been used in many broad engineering applications (see (Delage et al. 2006; Denoyer et al. 2004; Yu et al. 1999; Kim et al. 2004; Zhang et al. 2011)). In the building industry, we have found only limited uses of them. O'Neill (2014) used a Bayesian Network model to predict building energy performance. In their study, the HVAC and hot water energy consumption in an office building is predicted with the Bayesian Network model. Jensen et al. (2009) used a Bayesian Network model to quantify the effects from the thermal indoor environment on the mental performance of occupancies. Toftum et al. (2009) used a Bayesian Network model to describe the

relationship between acceptable thermal conditions and occupant performance and building energy consumption. Xiao et al. (2016) employed a Bayesian Network model in the fault detection for the air handling unit systems. In the previous study, we also developed a Bayesian Network model for optimizing the condenser water set point (Huang, Malara, et al. 2016) and performed a preliminary study, which aims to extend the model to load prediction purposes (Huang, Zuo, et al. 2016b). While significant, these applications were of limited scope and application of the Bayesian Network models.

Owing to the limited use in the building community, we provide a broad explanation of the Bayesian Network model here. Figure 1 illustrates the graphical structure of a typical Bayesian Network model. A model consists of "nodes" and "arcs". Nodes (e.g., X_a and X_b) represent variables (independent or dependent variables) involved in the studied system. Terminology is such that a node that has impacts on other nodes is called a "parent node" (e.g., X_a and X_f), and a node that is impacted by other nodes is named as "child node" (e.g., X_b and X_d). Of course, A node can be both a parent node and a child node (e.g., X_b and X_c). The arcs indicate the dependent relationships between the nodes. To demonstrate how a Bayesian Network model works, the node X_c will serve as an example in the following section.

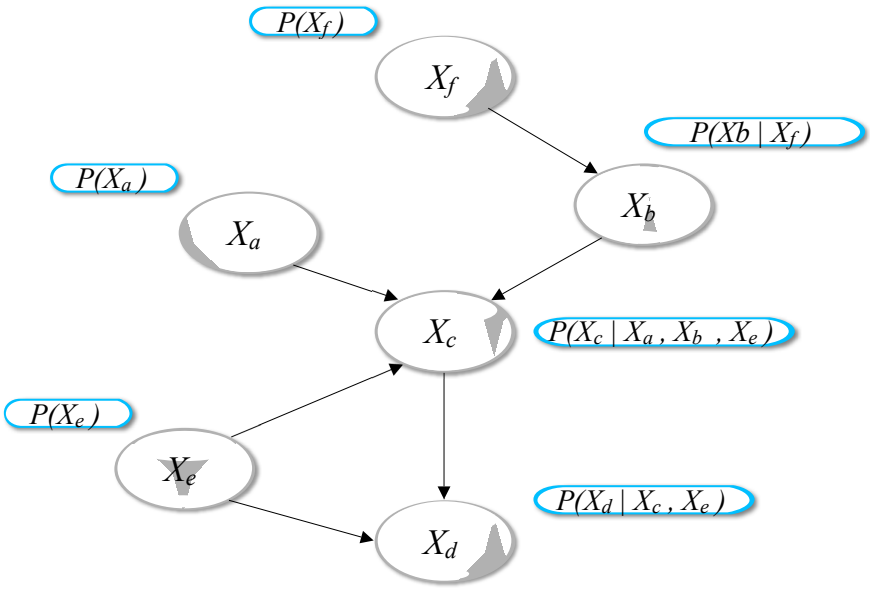


Figure 1 the structure of a typical Bayesian Network model

In Figure 1, the node X_{c} has three parent nodes: X_{a}, X_{b}, and X_{e}. Thus, the node X_{c} is a function of three parent nodes:

$$X_{\mathrm{c}} = f(X_{\mathrm{a}}, X_{\mathrm{b}}, X_{\mathrm{e}}). \tag{1}$$

The relationship between parent nodes and child notes can be expressed as an "exchange" or "transfer" table. For example, if we assume the values of X_{a}, X_{b}, and X_{e} are limited in its ranges: $[x_{\mathrm{a,L}}, x_{\mathrm{a,H}}]$, $[x_{\mathrm{b,L}}, x_{\mathrm{b,H}}]$, and $[x_{\mathrm{e,L}}, x_{\mathrm{e,H}}]$, respectively. Then, the ranges of X_{a}, X_{b}, and X_{e} can be split into smaller sections shown in Table 1:

Table 1 sections of ranges in the Bayesian Network model

Range	**Number of sections**	**Sections**
$[x_{\mathrm{a,L}}, x_{\mathrm{a,H}}]$	m	$A_1 = [x_{\mathrm{a,L}}, x_{\mathrm{a,1}})$ … $A_m = [x_{\mathrm{a,m-1}}, x_{\mathrm{a,H}}]$
$[x_{\mathrm{b,L}}, x_{\mathrm{b,H}}]$	n	$B_1 = [x_{\mathrm{b,L}}, x_{\mathrm{b,1}})$ … $B_n = [x_{\mathrm{b,n-1}}, x_{\mathrm{b,H}}]$
$[x_{\mathrm{e,L}}, x_{\mathrm{e,H}}]$	o	$E_1 = [x_{\mathrm{e,L}}, x_{\mathrm{e,1}})$ … $E_o = [x_{\mathrm{e,o-1}}, x_{\mathrm{e,H}}]$

Additionally, since the parent and child nodes are conditional, we specify the conditions as "conditional possibilities." For example, $X_{\mathrm{c}} = x_{\mathrm{c,i}}$, when the values of X_{a}, X_{b}, and X_{e} are within the set $\{X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l\}$. And its conditional possibility is computed as

$$P\left(\frac{X_{\mathrm{c}} = x_{\mathrm{c,i}}}{(X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l)}\right) = P(X_{\mathrm{c}} = x_{\mathrm{c,i}} \cap X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l) / P(X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l), \tag{2}$$

where $P(X_{\mathrm{c}} = x_{\mathrm{c,i}} \cap X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l)$ is the possibility that $X_{\mathrm{c}} = x_{\mathrm{c,i}}$ and the values of X_{a}, X_{b}, and X_{e} are within the set $\{X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l\}$, $P(X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l)$ is the possibility that the values of X_{a}, X_{b}, and X_{e} are within the set $\{X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l\}$.

The above "possibilities" or probabilities can be calculated for all parent-child relations. For example, if we could express the conditional relationship as Equation (3), assuming that we have a training dataset that is significantly large in terms of size. With a significantly large training dataset, the probabilities is close to the frequencies of that certain values.

$$P(X_{\mathrm{c}} = x_{\mathrm{c,i}} \cap X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l) = \frac{num(X_{\mathrm{c}} = x_{\mathrm{c,i}} \cap X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l)}{num_{tot}}, \tag{3}$$

$$P(X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l) = \frac{num(X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l)}{num_{tot}}, \tag{4}$$

where $num(X_{\mathrm{c}} = x_{\mathrm{c,i}} \cap X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l)$ is the number of training data points in which the values of X_{c}, X_{a}, X_{b}, and X_{e} are within the set $\{X_{\mathrm{c}} = x_{\mathrm{c,i}} \cap X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l\}$, $num(X_{\mathrm{a}} \epsilon\, A_j \cap X_{\mathrm{b}} \epsilon\, B_k \cap X_{\mathrm{e}} \epsilon\, E_l)$ is the number of training data points in which the values of X_{a}, X_{b},

and X_{e} are within the set $\{X_{\mathrm{a}} \epsilon A_j \cap X_{\mathrm{b}} \epsilon B_k \cap X_{\mathrm{e}} \epsilon E_l\}$, and num_{tot} is the number of total training data points.

Equation (2) can be simplified as:

$$P\left(\frac{X_{\mathrm{c}}= x_{\mathrm{c,i}}}{(X_{\mathrm{a}} \epsilon A_j \cap X_{\mathrm{b}} \epsilon B_k \cap X_{\mathrm{e}} \epsilon E_l)}\right)=\frac{num(X_{\mathrm{c}}= x_{\mathrm{c,i}} \cap X_{\mathrm{a}} \epsilon A_j \cap X_{\mathrm{b}} \epsilon B_k \cap X_{\mathrm{e}} \epsilon E_l)}{num(X_{\mathrm{a}} \epsilon A_j \cap X_{\mathrm{b}} \epsilon B_k \cap X_{\mathrm{e}} \epsilon E_l)}. \quad (5)$$

Then, the expectation of X_{c}, when the values of X_{a}, X_{b}, and X_{e} are within the set $\{X_{\mathrm{a}} \epsilon A_j \cap X_{\mathrm{b}} \epsilon B_k \cap X_{\mathrm{e}} \epsilon E_l\}$, can be calculated by

$$E(\frac{X_{\mathrm{c}}= x_{\mathrm{c,i}}}{(X_{\mathrm{a}} \epsilon A_j \cap X_{\mathrm{b}} \epsilon B_k \cap X_{\mathrm{e}} \epsilon E_l)}) = \sum_i^p x_{\mathrm{c,i}} P(X_{\mathrm{c}} = x_{\mathrm{c,i}} / (X_{\mathrm{a}} \epsilon A_j \cap X_{\mathrm{b}} \epsilon B_k \cap X_{\mathrm{e}} \epsilon E_l)), \quad (6)$$

where $x_{\mathrm{c},1}, \ldots, x_{\mathrm{c,p}}$ are the observed values of X_{c}.

If we assume that the value of X_{c}, when the values of X_{a}, X_{b}, and X_{e} are within the set $\{X_{\mathrm{a}} \epsilon A_j \cap X_{\mathrm{b}} \epsilon B_k \cap X_{\mathrm{e}} \epsilon E_l\}$, is equal to its expectation. Thus,

$$X_{\mathrm{c}} = f(X_{\mathrm{a}}, X_{\mathrm{b}}, X_{\mathrm{e}}) \cong E(\frac{X_{\mathrm{c}} = x_{\mathrm{c,i}}}{(X_{\mathrm{a}} \epsilon A_j \cap X_{\mathrm{b}} \epsilon B_k \cap X_{\mathrm{e}} \epsilon E_l)}) \quad (7)$$

Based on the above analysis, the value of X_{c} for the given values of X_{a}, X_{b}, and X_{e} can be determined according to equation (5) and (6).

It is possible that the training dataset may not cover the full range for the parent nodes, which means $num(X_{\mathrm{c}} = x_{\mathrm{c,i}} \cap X_{\mathrm{a}} \epsilon A_j \cap X_{\mathrm{b}} \epsilon B_k \cap X_{\mathrm{e}} \epsilon E_l)$ and/or $num(X_{\mathrm{a}} \epsilon A_j \cap X_{\mathrm{b}} \epsilon B_k \cap X_{\mathrm{e}} \epsilon E_l)$ equal to 0. In that case, equation (5) becomes invalid. To address this issue, the linear interpolation and the nearest

extrapolation methods are applied in the continuously prediction with the Bayesian Network model. For example, one may use the Bayesian Network model to predict X_c for the three successive time steps: t_1, t_2, and t_3. If the equation (5) becomes invalid only at t_2, we can estimate the prediction for t_2 by the linear interpolation method:

$$X_c(t_2) = \frac{X_c(t_3) - X_c(t_1)}{t_3 - t_1}(t_2 - t_3) + X_c(t_3)\,, \tag{8}$$

If the equation (5) becomes invalid only at t_3, we can estimate the prediction for t_3 by the nearest extrapolation method:

$$X_c(t_3) = X_c(t_2)\,, \tag{9}$$

The nearest extrapolation method assumes that the value of the studied child node changes very little by the small change in the values of the parent nodes. We shall admit both the linear interpolation method and the nearest extrapolation method may lead to inaccurate prediction especially when the length of the extrapolation period is large.

2.2 Procedure for Developing Bayesian Network Model

The typical procedure for developing the Bayesian Network model consists of four steps and the following parts detail each step.

Step 1: To determine the parent nodes for the studied child nodes.

Selecting the parent nodes requires a careful balance between accuracy and accessibility: including more parent nodes tends to give better prediction results. However, more parent nodes also require more efforts in preparing the training dataset. For example, to predict the cooling load for buildings, ideally we should

have the detailed information regarding the building operation, such as the number of the occupancies, and intensive meteorological data, such as the dry bulb temperature, the precipitation, the solar radiation intensity, and how thick the cloudy is. However, many of the above information is not necessarily accessible. For example, the number of the occupancies is hard to collect due to the concern that the privacy may be violated.

Step 2: To prepare and process the training data

Based on the identified parent nodes in Step 1, we must express the relationship between the parent and child notes. We can do so empirically, by collecting data for both parent nodes and the corresponding child nodes. By doing so, we would develop an empirically-based Bayesian Network model. The relationship could also be generated through the use of physics-based models, but we do not do so here, in the application that follows. To make sure the training dataset contains sufficient information, the points should evenly distribute in each range of parent nodes. After the training dataset is ready, we should determine how to split this dataset. There are different ways to perform the split, such as evenly splitting or setting the split intervals so that each split section has the same or close amount of data points. After the split is completed, we can calculate the conditional probabilities according to equation (5), or directly calculate the values of child nodes from equations (8) or (9).

Step 3: To calculate expectations

After we obtain conditional probabilities from Step 2, we can calculate the expectations of the studied child nodes, under different section combinations of the parent nodes, according to equation (6). The calculated expectations combined with the corresponding section combinations of parent nodes form another dataset (named as "output dataset").

Step 4: To generate a lookup table.

To facilitate the implementation of the Bayesian Network model, we can convert the output dataset from Step 3 into a multiple-dimensional lookup table. This table has $n_p + 1$ columns (n_p is the number of the

parent nodes): the first n_p columns contain the split sections for each parent node while the last column contains the values of expectations. Then the lookup table can be directly implemented in either software or hardware.

In this study, we employ Python (Python Software Foundation), which is a script language, to automatize the above procedure. Theoretically, the above procedure can be used to develop the Bayesian Network model to describe any relationships, which is not limited in the building industry. However, whether the desired prediction by the Bayesian Network model can be achieved depends on different settings in the Bayesian Network model, for example, how to select appropriate parent nodes and how to determine the structure of the Bayesian Network model. The general methods to obtain the best settings, however, are beyond the scope of this paper.

2.3 Bayesian Network Model for Cooling Load Prediction

We now process with the application of the theory that discussed in the above section to express the relationship between the cooling load and predefined factors. We do so, in order to have a model that is readily computing, in near real time to facilitate the implementation in the real world, as mentioned in the introduction section. According to section 2.2, the first step to develop the Bayesian network model for cooling load prediction is to determine the parent nodes. The cooling load can be affected by many factors. Generally speaking, those factors can be divided into two categories: the weather condition and the building internal activities. Depending on the type of building, the weather condition and internal activities affect the total cooling load in different ways. For example, in data centers, the cooling load is dominated by the heat gain from the IT (information technology) equipment, and the impact of the weather condition is usually negligible. However, for buildings with a constant and high outdoor air intake (such as the semiconductor manufacturing facilities), the cooling load is mainly due to the processing of the outdoor air. In that case, the cooling load is mainly determined by the weather condition. For a commercial building,

such as an office building, the cooling load is usually affected both by the weather condition and internal activities (ASHRAE 2012).

In this study, we focus on the cooling load prediction for the commercial buildings. Thus, we have to identify the parent nodes that represent the weather condition and internal activities, respectively. As mentioned in the section 3.2, the selection of the weather condition and internal activities should be made based on the balance between the prediction accuracy and the data accessibility. With that in minds, we first consider the parent nodes for the weather condition. As discussed above, there are many meteorological data can be used to describe the weather condition. However, some of the data require special efforts to obtain. For example, to determine how thick the cloud is, we usually rely on the human observation, which is not feasible for the cooling load prediction. To facilitate the large-scale application, we select the outdoor dry bulb temperature and the outdoor wet bulb temperature as the representatives of the weather condition. Those two temperature can describe the thermodynamics patterns of the outdoor air, which is one of the major driven forces for the heat transfer through envelop. More importantly, those two temperature are readily available in either local weather station, or the weather forecast stations. We shall acknowledge that only considering the two temperature may cause the accuracy issue if the solar radiation accounts for the significant portion of the heat gain by the buildings. The same philosophy for selecting the parent nodes is also applied to the internal activities and we select two variables: the hour index and the day category number. This is mainly because they require very little effort to collect. The hour index is the index of the hour in one day: it starts with 0 representing 12 AM. The hour index is used to reflect the changes of the internal activities over the hour-level period. The day category numbers are used to reflect the internal activities of the day-level period. Table 2 shows three category numbers and their descriptions. We expect that days in the same category have a similar internal activity pattern.

Table 2 the category of days

Day Category Number	Day Category name	Description
1	Working Day	Normal working day of the week when no event* occurs. For typical office buildings, the working days are from Monday to Friday
2	Holiday	The non-working days when no event occurs. For typical office buildings, the working days are from Saturday to Sunday
3	Event Day	The days when events occur.

* In this context, "event" means the activity during which the number of occupancies or occupancy schedules are significantly different from the normal working day or Holiday. An example of event in the universities will be the commencement.

Based on the above analysis, a Bayesian Network model shown in Figure 2 is built for the cooling load prediction. This Bayesian Network model has four parent nodes and one child node. There are also four arcs connect the child node with each parent node.

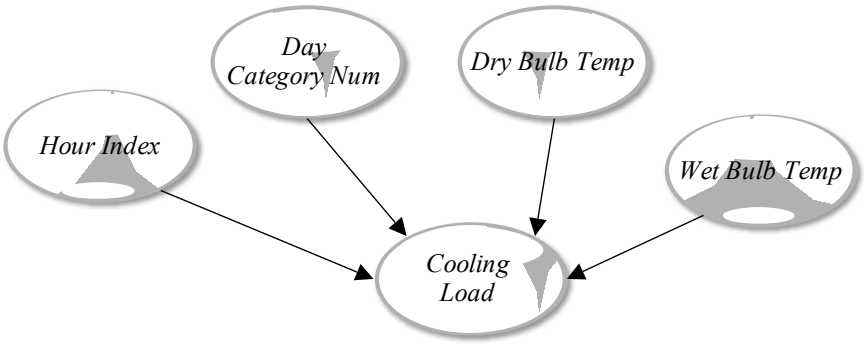


Figure 2 the structure of Bayesian Network model for the cooling load prediction

3. Case Study

In this section, we will detail how to use the proposed Bayesian Network model to predict the cooling load for a real university campus. To better evaluate the performance of the Bayesian Network model, we also employ two other models from the literature, in the cooling load prediction. To assess the impact of the quantity of training data on the accuracy of the cooling load prediction, we generated multiple training datasets with different amounts of data. Then, we trained the three models with those datasets and predicted the cooling load for the testing set. Finally, we studied the sensitivity of the cooling load prediction to the uncertainties in the weather condition forecast.

3.1 Case Description

The studied case is a university campus located in Annapolis, Maryland, U.S. The campus consists of 10 buildings and lies in a subtropical climate zone, which is hot and humid in summers and cool in winters. The university has three academic semesters: the spring semester is from early Januaries to the mid of Mays, the summer semester is from the mid of Mays to the mid of Augusts, and the fall semester is from the mid of Augusts to the end of the calendar year.

Figure 3 shows the onsite measured data from the campus for two periods: 09/08–11/02/2014 and 04/27–09/20/2015. The cooling load was gathered by the supervisor controller of the central chiller plant that served the entire campus. The two periods cover the summer semester and the fall semester, which constitute a typical cooling season for this campus. The cooling load decreased from around 4,000 ton to around 500 ton from September to October, 2014. From February to May, 2015, the cooling load increased from around 500 ton to around 7,000 ton, and then decreased to 4,000 ton in September, 2015.

Besides the cooling load, the hourly outdoor dry bulb temperature and the outdoor wet bulb temperature were obtained from a weather station located in the campus. The day category number was determined according to the academic calendar, which was available on the university website.

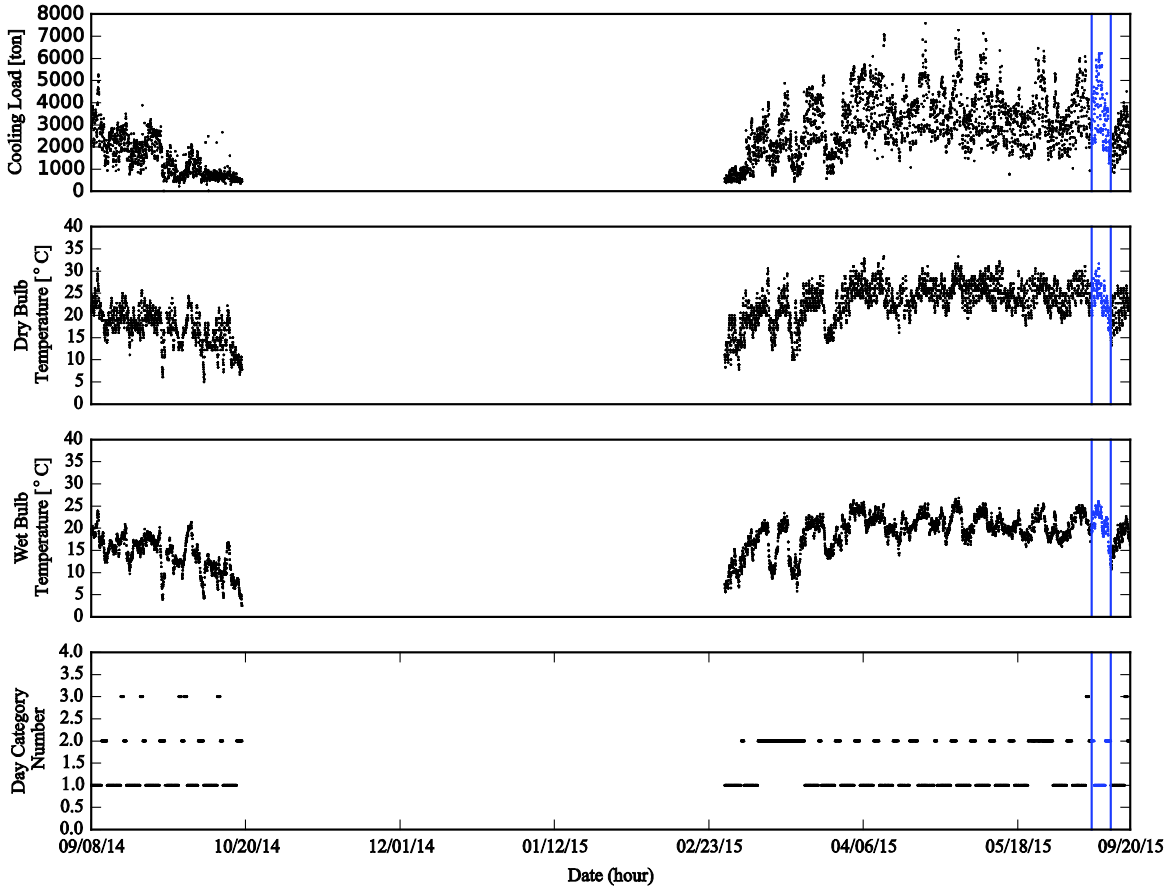


Figure 3 onsite measured data for the studied case (white part: traning data; blue part: testing data

In this study, the measured data was divided into two parts for different purposes: one part (09/09–11/02/2014 & 04/27-09/06/2015 & 09/14-20/2015, black part in Figure 3) for training and the other part (09/07–13/2015, 1 week, blue part in Figure 3) for testing. The ratio of the points in the testing dataset to that in the training dataset is around 0.04.

3.2 Prediction Settings

We used the Bayesian Network model described in Section 2.3 to predict the cooling load for the studied campus. As discussed in Section 2.2, it is necessary to discretize the training dataset into groups. In the Bayesian Network model, two of the parent nodes (day category number and the hour index) are already discrete. For the other two parent nodes (outdoor dry bulb temperature and outdoor wet bulb temperate), we discretized the temperature into 2 degree increments (2^{o}C) that spanned the full range of these temperature data: for the outdoor dry bulb temperature, the discrete sections are $[0, 2)$,…, $[40, \infty)$; for the outdoor wet bulb temperature, the split sections are $[0, 2)$, ..., $[30, \infty)$. We chose 2^{o}C increment because it is the best according to our sensitivity analysis (see details in Huang, Zuo, et al. 2016b).

Besides the Bayesian Network model, we also employed two other models: a Support Vector Machine model and an Artificial Neural Network model. As mentioned in the introduction section, both the Support Vector Machine model and the Artificial Neural Network model have been used for predicting the cooling load in the literature (Hou et al. 2009; Li et al. 2009a; Li et al. 2009b; Kashiwagi et al. 1993; Sakawa et al. 1999; Ben-Nakhi et al. 2004; Kwok et al. 2011; Leung et al. 2012). Table 3 shows the information used by these models to predict:

Table 3 the settings of the Supply Vector Machine model and Artificial Neural Network model

Models	Inputs	Key Settings
Support Vector Machine*	Outdoor dry bulb temperature, Outdoor wet bulb temperature, Hour index, Day category number	Kernel function: Gaussian function, Penalty parameter C of the error term: 1000
Artificial Neural Network**		Number of the hidden layer: 1, Training algorithm: Back-propagation, Maximum iteration number: 100

* implemented with Python package: Scikit-learn (Pedregosa et al. 2011)

** implemented with Python Package: PyBrain (Schaul et al. 2010)

For both the Support Vector Machine and the Artificial Neural Network model, we normalized the training data.

3.2 Evaluation Metrics

To quantitatively evaluate the prediction accuracy, we employ two commonly used variables: the coefficient of determination, denoted R^2, and the root mean squared deviation (RMSD). R^2 is calculated by

$$R^2 = 1 - \frac{\sum_i^{pnum}(\dot{Q}_{p,i} - \dot{Q}_{m,i})^2}{\sum_i^{pnum}(\overline{\dot{Q}_m} - \dot{Q}_{m,i})^2}, \tag{10}$$

where $\dot{Q}_{p,i}$ and $\dot{Q}_{m,i}$ are the ith predicted and measured cooling loads, $pnum$ is the prediction number, and $\overline{\dot{Q}_m}$ is the mean value of $\dot{Q}_{m,i}$. Basically, the more closely R^2 approaches 1, the better the prediction accuracy is.

The RMSD is calculated by

$$RMSD = \sqrt{\frac{\sum_i^{pnum}(\dot{Q}_{p,i}/\dot{Q}_{m,i} - 1)^2}{pnum}} \times 100. \tag{11}$$

The closer RMSD is to 0, the better the prediction results are (Reddy 2011).

3.4 Relationship between Cooling Load Prediction and Amount of Training Data

Typically, the amount of training data used in the calibration face impacts the fidelity of the resulting. In this section, we investigated the relationship between the cooling load prediction and the amount of the data used in the model training. To do so, we will first generate various training datasets and each dataset contains different amount of data. Then we will use each training dataset to train the three models: Bayesian Network model, Support Vector Machine model. After that, we will use the trained model to predict the cooling load for the testing set and compare the prediction results with the measurement. The following section elaborates the above steps.

As a first step, we split the training data by weeks and indexed the weeks by time. As shown in Figure 4, the training data was divided into twenty-eight weeks. The indexes of the weeks, which are just before and after the testing week, are twenty-seven and twenty-eight, respectively. We then assigned the twenty-eight weeks into fourteen groups shown both in Figure 4 and Table 4. For the second step, we trained the Bayesian Network model, the Support Vector Machine model, and the Artificial Neural Network model with the data from 14 groups shown in the Table 4.

Figure 4 index of the training data

Table 4 the groups of the training datasets

Group ID	The Week Index	Number of Data Points
1	26,27	276
2	24,25,26,27	554
3	22,23,24,25,26,27	872
4	20,21,22,23,24,25,26,27	1,188
5	18,19,20,21,22,23,24,25,26,27	1,463
6	16,17,18,19,20,21,22,23,24,25,26,27	1,756
7	14,15,16,17,18,19,20,21,22,23,24,25,26,27	2,079
8	12,13, 14,15,16,17,18,19,20,21,22,23,24,25,26,27	2,413
9	10,11,12,13, 14,15,16,17,18,19,20,21,22,23,24,25,26,27	2,743
10	8,9,10,11,12,13, 14,15,16,17,18,19,20,21,22,23,24,25,26,27	3,077
11	6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26, 27	3,413
12	4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25, 26,27	3,735
13	2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24 ,25,26,27	4,071

14	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23, 24,25,26,27,28	4,379

Lastly, we used the trained models to predict the cooling load for the testing set. Figure 5 shows the RMSDs of three models with different training datasets. For the Bayesian Network model, the RMSD drops from around 25.0% to 14.6% when the training dataset changes from Group One (two weeks) to Group Fourteen (twenty-eight weeks). However, we also notice that the RMSD does not monotonously decrease. For example, when the training dataset group changes from Group Four (eight weeks) to Group Five (ten weeks), the RMSD increases from around 23.0% to 29.4%. For the Support Vector Machine model, when the amount of the training data is less than 12 weeks, its RMSD keeps in a narrow range from 32.1%-34.4%. When the amount of the training data changes from 12 weeks to 16 weeks, its RMSD significantly reduces from around 34.4% to 14.6%. When the amount of the training data further increases, the RMSD of the Support Vector Machine model keeps in a narrow range from 14.5%-16.2%. For the Artificial Neural Network model, the change of its RMSD is not as obvious as that of the rest two models: when the amount of the training data is less than 14 weeks, we can observe some of oscillations (around $\pm 3\%$) in its RMSD. When the amount of the training data is larger than 22 weeks, its RMSD is almost constant.

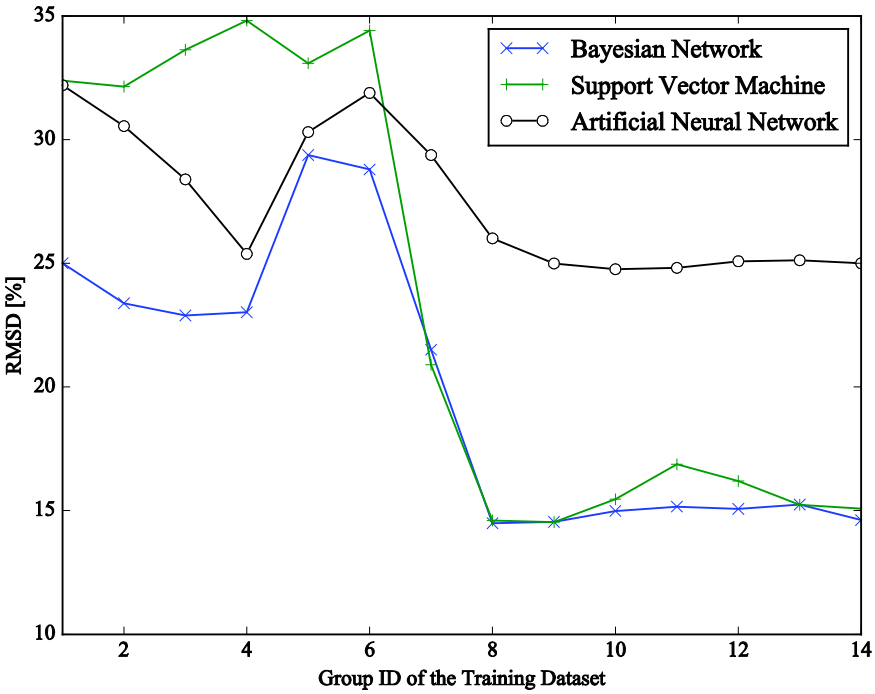
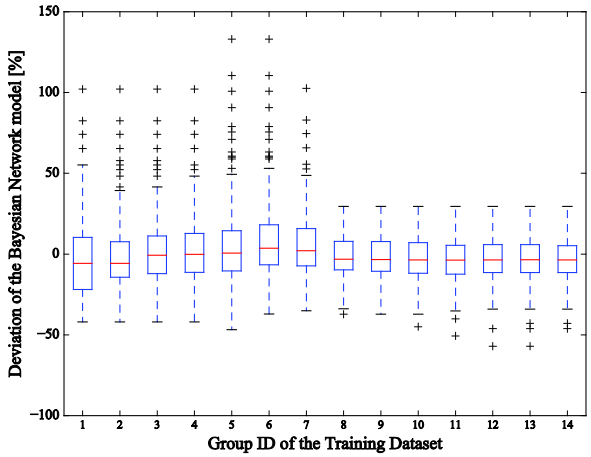


Figure 5 cooling Load Prediction Results with differen training datasets

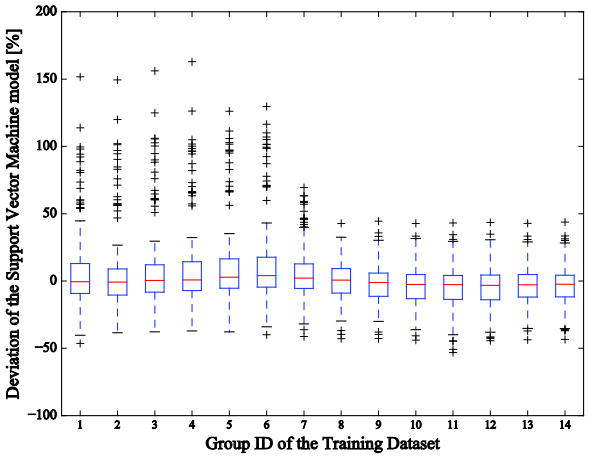
There are two questions associated with the results shown in Figure 5. The first question is why the RMSDs of the Bayesian Network model and the Support Vector Machine model drop significantly when the group ID changes from six to eight. The second question is why the RMSDs for all the three models become almost constant when the group ID is larger than eight.

To examine the first question, we explored in more detail the distribution of the deviations in the three models predictions from the measured cooling load. The deviation is defined as:

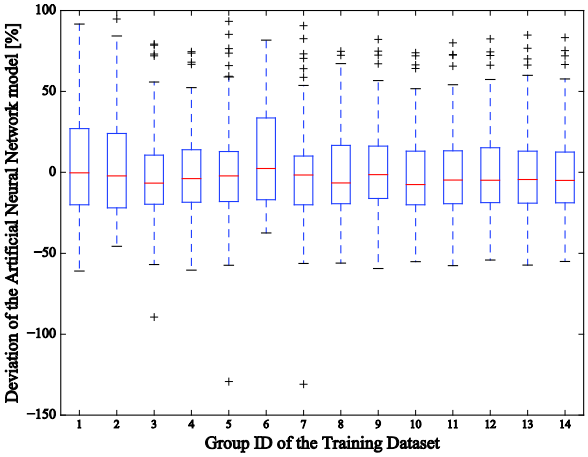
$$DV_i = (\dot{Q}_{p,i}/\dot{Q}_{m,i} - 1) \times 100. \quad (12)$$



(a) Bayesian Network model



(b) Support Vector Machine model



(c) Artificial Neural Network model

Figure 6 the distribution of deviations in the prediction by different models

Figure 6 shows the distribution of the deviations. For the Bayesian Network model, when the group ID is less than eight, there exists a significant number of outliers (the data points that lie an abnormal distance from other points). In statistics, an outlier is viewed as being too far from the central values to be reasonable. When the group ID is larger than or equal eight, the number of outliers reduces dramatically. Since the deviations of outliers are much higher than that of other predicted data points, reducing the number of outliers can contribute a lot to the decrease of the RMSD. Because of this, the RMSD of the Bayesian

Network model drops significantly when the group ID changes from six to eight. In addition, the interquartile range (the difference between the upper and lower quartiles) of the deviation distribution for the Bayesian Network model is slightly reduced when the group ID changes from one to eight. This means the prediction of the Bayesian Network model is becoming more and more accurate in general. For the Support Vector Machine model, the distributions of deviations are very similar to those of the Bayesian Network model. The major difference between the predictions of the Bayesian Network model and those of the Support Vector Machine model is that there are more outliers in the deviation distributions of the latter than that of the former. For the Artificial Neural Network model, the impact of the group ID on its deviation distributions is less obvious than that of the other two models. When group ID changes from one to fourteen, there is not a clear tendency that the interquartile range or the number of outliers is reducing or increasing. Based on the above analysis, we can see that outliers in the prediction results by the Bayesian Network model and the Support Vector Machine model are likely to be caused by the lack of certain training data points. For the Artificial Neural Network model, the outliers in the prediction results are seems to be caused by its incapability to catch the change of the cooling load.

To answer the second question, we firstly check how many testing data points are covered by the training dataset. Here we consider the testing data point is covered by the training dataset if it meets the following condition:

$$\exists\, i \in \{1,..,n\},\ \left|T_{db,train}[i] - T_{db,test}\right| < 1 \cap \left|T_{wb,train}[i] - T_{wb,test}\right| < 1 \cap DC_{train}[i] = DC_{test} \cap HI_{train}[i] = HI_{test}, \tag{13}$$

where n is the number of the data points in the training dataset, $T_{db,train}[i]$, $T_{wb,train}[i]$, $DC_{train}[i]$, and $HI_{train}[i]$ are the dry bulb temperature, the wet bulb temperature, the day category number, and the hour index for the ith data point in the training dataset. $T_{db,test}$, $T_{wb,test}$, DC_{test}, and HI_{test} are the dry bulb temperature, the wet bulb temperature, the day category number, and the hour index for the testing data point. Equation (13) assumes that with very small changes (1°C) in the dry bulb temperature and the wet

bulb temperature, the change in the cooling load, when the day category number and the hour index are constant, is negligible. Thus, if there are data points in the training dataset, which has the same day category number and the same hour index as well as a closer dry bulb temperature and a closer wet bulb temperature as the testing data point, we actually have the information regarding the cooling load for the testing data point in the training dataset. Thus, we consider the testing data point is "covered".

We calculated the percentage of the testing data points that are covered by different training datasets (named as cover-percentage) and Figure 7 shows the results. It is clear that the cover-percentage increases dramatically when the group ID increases from one to eight. However, when the group ID is larger than eight, the cover-percentage almost keeps constant. We then took a close look at the prediction results of the three models when data in the Group One is used for training. Figure 8 shows the deviation distributions of the testing data points covered/not covered by Group One. We can see, for all the three models, the deviation distribution of the testing data points that are not covered by Group One is much worse than that of the testing data points covered by Group One. The maximum absolute deviation of the predictions that are not covered by Group One is up to 7 times larger than that of the predictions covered by Group One. This result suggests that all the three models are lack of ability to extrapolate the training dataset. For the studied case, the cover-percentage basically determines their prediction accuracy. Thus, the RMSDs for all the three models become almost constant when the group ID is larger than eight.

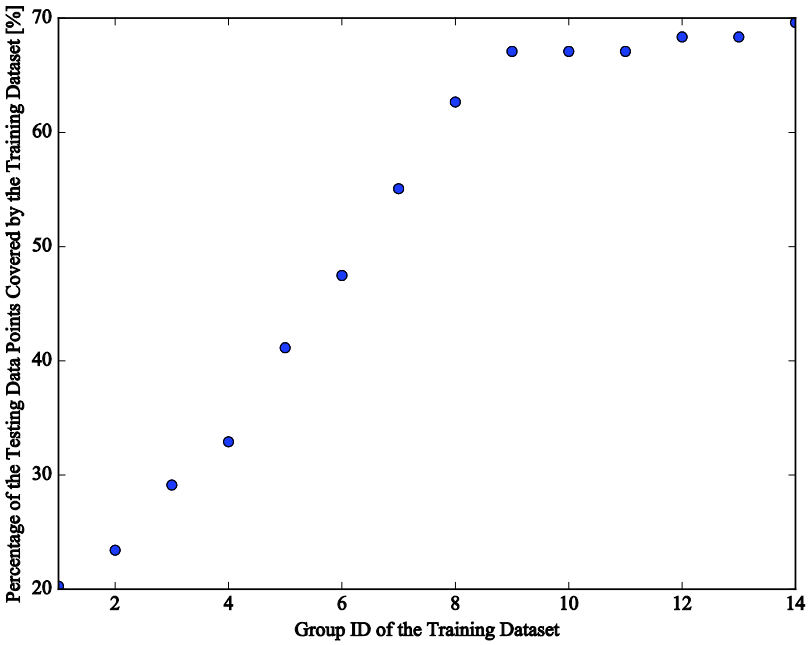


Figure 7 percentage of the testing data points covered by the training dataset

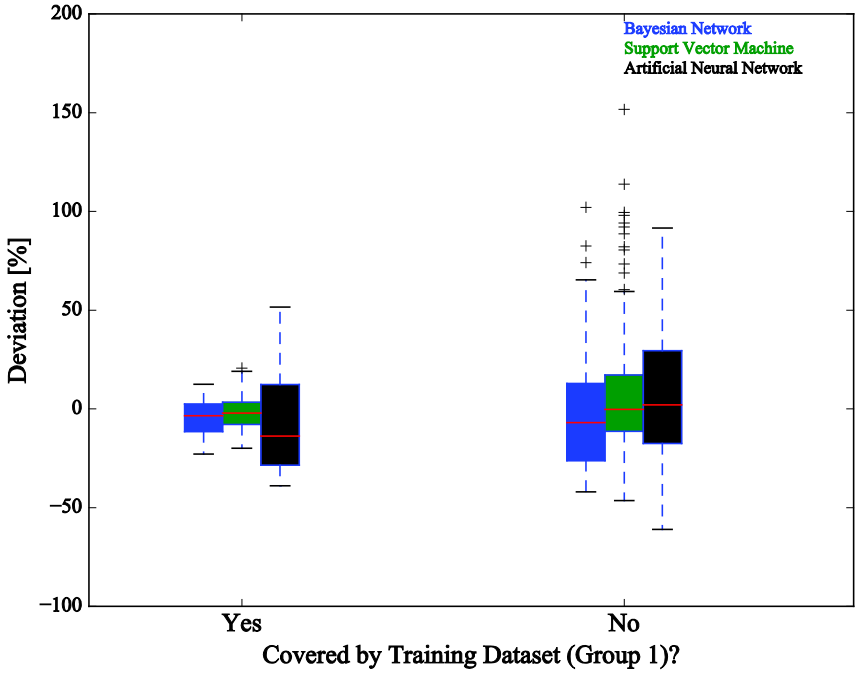


Figure 8 the distribution of the deviations in the three models when training dataset is Group One

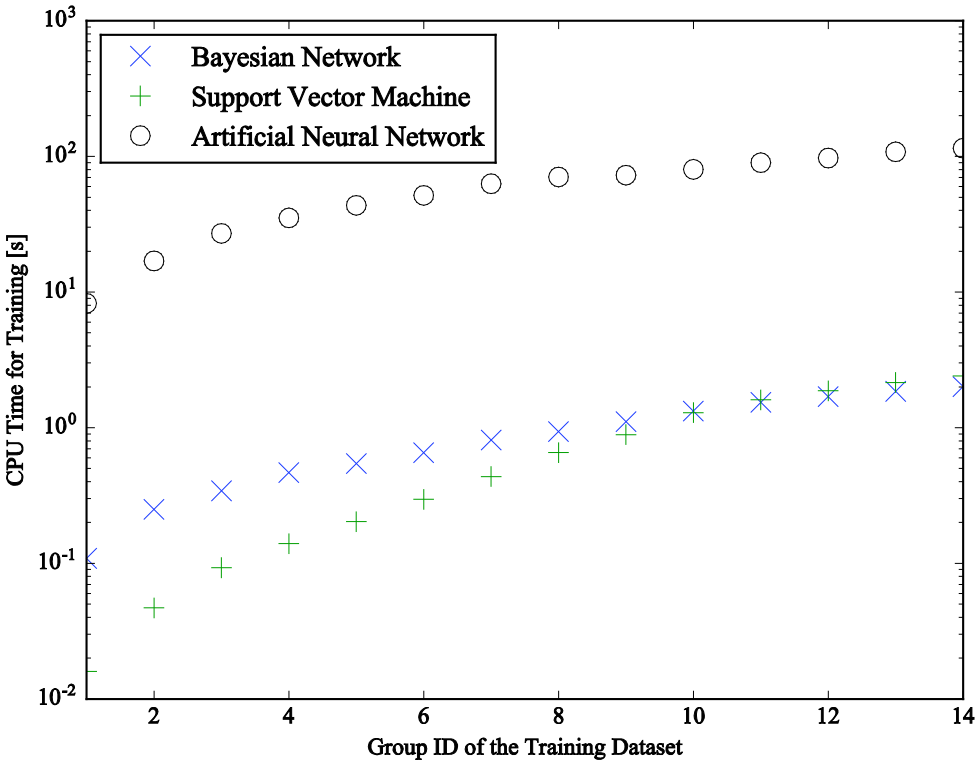


Figure 9 the CPU time for training of different models

In addition to exploring the performance of the Bayesian Network, we also explore the computational benefits of the presented approach. Figure 9 shows the CPU time for the training of three models with different training datasets. The computer we used in this study is a Dell Ultrabook laptop. The CPU and the operation system are Intel Core i7-6600U (2.60GHZ & 2.80 GHz) and Window 7 Enterprise, respectively. In general, for all the three models, the CPU times for training increase significantly when the amount of training data increases. The Support Vector Machine model has the lowest CPU times when the group ID of the training dataset is less than ten. When the group ID of the training dataset is larger than or equal to ten, the Bayesian Network model is the fastest in training. The CPU times of the Artificial Neural Network model are much higher than those of the other models regardless of which training dataset is used (always by at least 10 times). When the Group Fourteen is used as the training dataset, the CPU time for the Artificial Neural Network model is 115 s, which is 48 times higher than the Support Vector Machine model, and 58

times higher than the Bayesian Network model. The reason that the Artificial Neural Network model has very high CPU time is because its training process involves an iteration process to minimize the error functions. In our cause, it is obvious the iteration process doesn't converge very soon.

Lastly, we plotted the predictions of three models against the measured cooling load when the Group Fourteen is used as the training dataset (Figure 10). The Bayesian Network model and the Support Vector Machine model achieve very closer results: their R^2 are both around 0.8. However, the results of the Artificial Neural Network model are much worse than the rest two models and its R^2 is only 0.61. Based on Figure 10, we see that the Bayesian Network model and the Support Vector Machine model tracks the change of cooling load quite well although there are some relatively large deviations in the mid of September 10. For the Artificial Neural Network model, it fails to capture the change of cooling load for September 9 and 10.

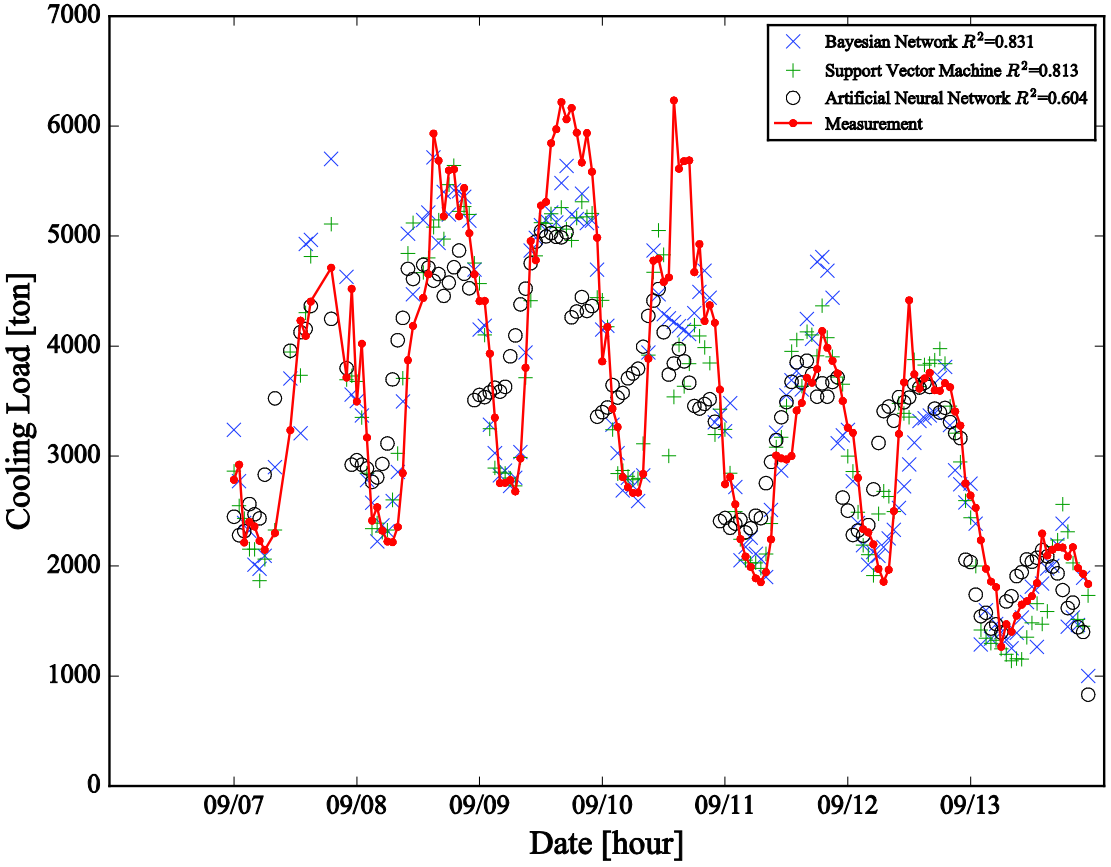


Figure 10 cooling Load Prediction Results with 28 weeks training dataset

Based on the above analysis, we can obtain the following observations:

1) All the three models can't extrapolate the training dataset. For the studied case, the cover-percentage determines the accuracy of the cooling load prediction. If the testing data point is beyond the training dataset, prediction deviations are much larger.

2) For the studied case, the Bayesian Network model and the Support Vector Machine model can catch the trajectory of the cooling load quite well. However, the performance of the Artificial Neural Network model is much worse, with the same training dataset even.

3.5 Relationship between Uncertainties in Weather Forecast and the Cooling Load Prediction

In this section, we will first demonstrate how we mimicked uncertainties in weather. Then we will show how we included the mimicked uncertainties in the inputs for the trained models to predict the cooling load.

To mimic the uncertainties in the weather forecast, we employed the following equation:

$$Pd^* = Pd + Err_{sta} + Err_{ran}, \tag{14}$$

where Pd^* is the forecast with "uncertainties", Pd is the prefect forecast. The Err_{sta} represents static error, which is usually caused by the limitations of the forecast model. The Err_{ran} represents random error, which is usually caused by disturbances, such as the noises in the inputs for the prediction.

We used the measured outdoor dry bulb temperature and outdoor wet bulb temperate shown in Figure 3 as the prefect forecast and applied equation (14) to generate "predictions" of the outdoor dry bulb temperature and the outdoor wet bulb temperate with uncertainties.

For both the two temperature, the random error (unit: ^oC) is computed as:

$$Err_{ran} = random(-0.5, 0.5), \tag{15}$$

where $random()$ is a function that returns a random number between the input range.

For the static error, we considered seven possible values: -2, -1.5, -1, 0, 1, 1.5, 2^{o}C. Figure 11 shows the generated "prediction" of the outdoor dry bulb temperature when the static error is 2^{o}C. We then used the "predictions" with synthetic errors as the input to predict the cooling load for the testing period again.

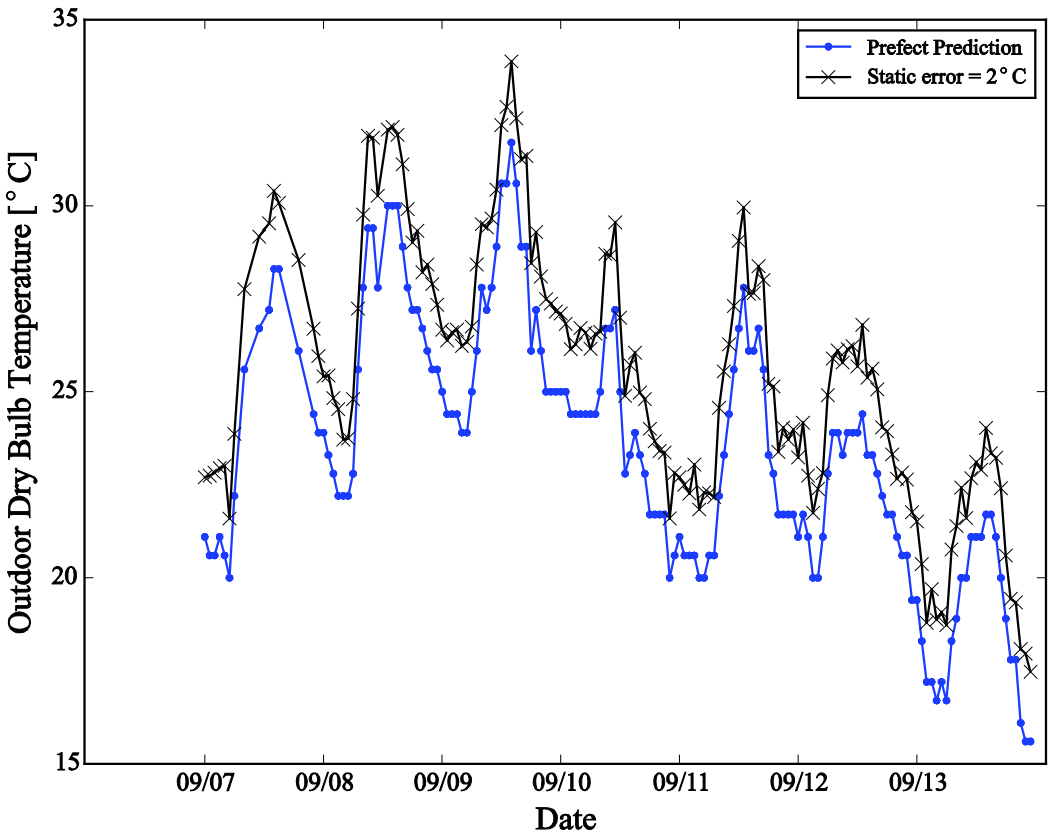


Figure 11 the "prediction" of the outdoor dry bulb temperature with uncertainties

Figure 12 shows the prediction results of the three models. For the Bayesian Network model, when the static error increases from 0 to 2°C, the RMSD increases from 16.7% to 23.6%. If the static error decreases from 0 to -2°C, the RMSD increases from 16.7% to 24.2%. This means the RMSD will increase up to 20.0% when the absolute value of the static error increases by 1°C. For the Support Vector Machine model, the uncertainties in the weather forecast have similar impacts on its accuracy as that on the accuracy of the Bayesian Network model. The RMSD increases up to 30.0% when the absolute value of the static error increases by 1°C. For the Artificial Neural Network model, the impact from the uncertainties is smaller: The RMSD increases up to 16.0% by 1°C increase in the absolute value of the static error.

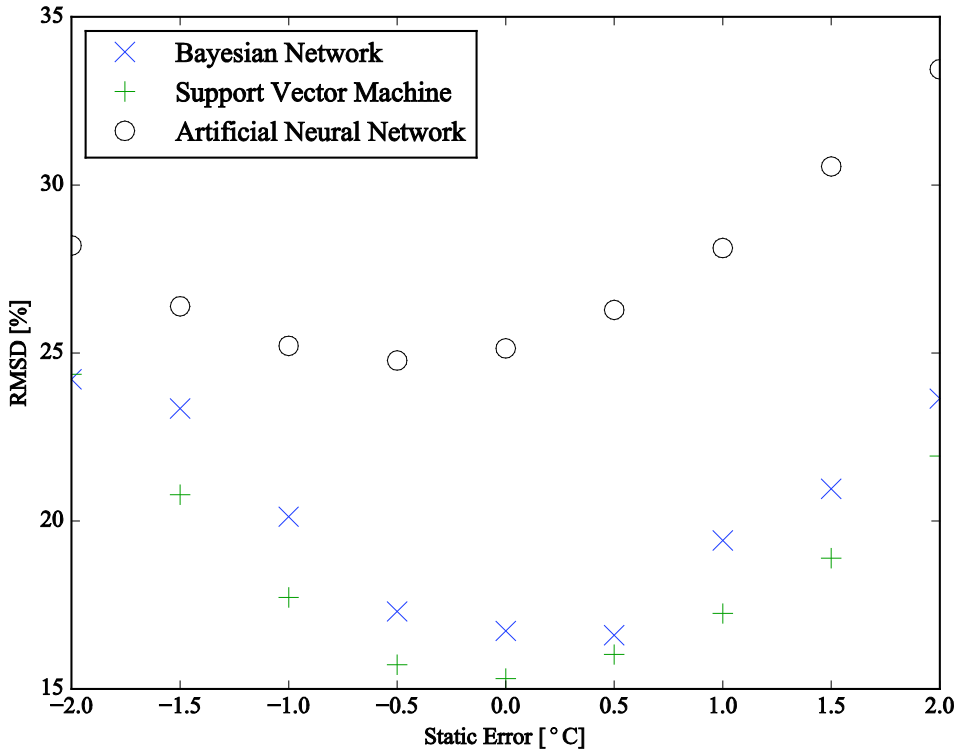


Figure 12 the cooling load predictions when the uncertainities exist in the outdoor dry bulb temperature and outdoor wet bulb temperatature forecast

Based on the above analysis, we can see for all the three models, the uncertainties in the weather forecast have dramatic impacts on their prediction accuracy: the RMSD can increase by 30% when the absolute value of the static error increases only by 1°C. This indicates, for the studied case, the importance of relatively accurate weather forecasts. It also suggests that the accuracy of the forecasting model must be weighed in light of the accuracy of the model inputs, like forecasts of weather. Finally, we also observed that the Support Vector Machine model is slightly more sensitive to the change of the uncertainties.

4. Conclusion

This paper proposes a Bayesian Network model for predicting the cooling load of a commercial building. We show that the proposed Bayesian Network model has the potential of achieving similar or better performance than a Support Vector Machine model or an Artificial Neural Network model. In the case study, the Bayesian Network model used the lowest CPU time for training when the amount of the training data is more than ten weeks. The CPU time cost by the Artificial Neural Network model is higher than that of the Bayesian Network model by up to 5,700%. Moreover, using the Bayesian Network model does not require background in sophisticated mathematical theories. These benefits suggest that the Bayesian Network model is promising for real-world applications.

In this paper, we also explore the relationship between performance of the candidate prediction models and the amount of data available to train the models. We found all the three models can't extrapolate the training dataset. For the studied case, the three models tend to have much larger prediction deviation if the testing data point lies far distance from the training dataset. On the other hand, we also notes that increasing the amount of the training data, but not the percentage of the testing data points that are covered by the training dataset, doesn't benefit the prediction a lot. Based on the above statement, we suggest to increase the percentage of the testing data points that are covered by the training dataset, rather than only the amount of data in the training dataset, if the three models are employed for prediction.

Another insight from this paper is that all the three models don't have the ability to tolerate the uncertainties in the inputs. For the studied case, uncertainties in the weather forecast significantly decrease the accuracy of the cooling load prediction in the studied case. Among the three models, the Support Vector Machine model is more sensitive to uncertainties. The above results reveal that when evaluating the accuracy of the prediction model, the accuracy of the model input, such as weather forecast, should be taken into account.

In this study, the evaluation of the Bayesian Network model focuses on one single case and we selected a relatively short test period. In the future, it will be beneficial to extend the evaluation to longer testing periods, or different types of buildings under different weather conditions when relevant data is available. By doing that, we can see if the Bayesian Network model still works and the above observations can still be applied. It will also interesting to study how to further increase the accuracy of the Bayesian Network model, by identifying better indicators for the internal activities. In this study, we chose the hour index and the day category since they are readily available. However, there are limitations from using them. For instance, the cooling load may be more sensitive to the hour index in working days for a typical office building, which makes the training process more difficult. It is also worth mentioning that this paper focuses on exploring the possibility of applying the Bayesian Network model in the cooling load prediction for the commercial buildings. To effectively achieve this goal, we employ the use of the commonly used settings for the Bayesian Network model, such as discrete inputs, a uniform discretization interval, and a predefined structure. In the future study, it will be important to study if better performance can be achieved by employing continuous inputs, and letting the Bayesian Network model learn the discretization interval and structure from the training dataset.

Acknowledgement

This research was supported by the U.S. National Science Foundation under award number IIS-1633338. This research was also supported by the U.S. Department of Defense under the ESTCP program. The authors thank, Ana Carolina Laurini Malara, Marco Bonvini, Michael Wetter, Mary Ann Piette, Jessica Granderson, Oren Schetrit, Rong Lily Hu and Guanjing Lin for the support provided through the research.

Reference

Alajmi, Ali. 2012. "Energy Audit of An Educational Building in a Hot Summer Climate." *Energy and Buildings* 47 (2012):122-30.

ASHRAE. 2012. "*ASHRAE Handbook—HVAC Systems and Equipment*." Atlanta, GA, USA.

Ben-Nakhi, Abdullatif E., and Mohamed A. Mahmoud. 2004. "Cooling Load Prediction for Buildings using General Regression Neural Networks." *Energy Conversion and Management* 45 (2004):2127-41.

Birdsall, B., W.F. Buhl, K.L. Ellington, A.E. Erdem, and F.C. Winkelmann. 1990. "Overview of the DOE-2 Building Energy Analysis Program, Version 2. 1D." *Technical Report LBL-19735-Rev.1*

Braun, James E., and Nitin Chaturvedia. 2002. "An Inverse Gray-Box Model for Transient Building Load Prediction." *HVAC&R Research* 8 (1):73-99.

Chapelle, Olivier, Vladimir Vapnik, Olivier Bousquet, and Sayan Mukherjee. 2002. "Choosing Multiple Parameters for Support Vector Machines." *Machine Learning* 46 (1-3):131-59.

Chen, Yongbao, Peng Xu, Yiyi Chu, Weilin Li, Yuntao Wu, Lizhou Ni, Yi Bao, and Kun Wang. 2017. "Short-term Electrical Load Forecasting using the Support Vector Regression (SVR) Model to Calculate the Demand Response Baseline for Office Buildings." *Applied Energy* 196 (2017):659–70.

Corbin, Charles D., Gregor P. Henze, and Peter May-Ostendorp. 2013. "A Model Predictive Control Optimization Environment for Real-time Commercial Building Application." *Journal of Building Performance Simulation* 6 (3):159-74.

Crawley, Drury B., Linda K. Lawrie, Frederick C. Winkelmann, W.F. Buhl, Y. Joe Huang, Curtis O. Pedersend, Richard K. Strand, et al. 2001. "EnergyPlus: Creating a New-generation Building Energy Simulation Program." *Energy and Buildings* 33 (4):319-31.

Deb, Chirag, Lee Siew Eang, Junjing Yang, and Mattheos Santamouris. 2016. "Forecasting Diurnal Cooling Energy Load for Institutional Buildings using Artificial Neural Networks." *Energy and Buildings* 121 (2016):284-97.

Delage, Erick, Honglak Lee, and Andrew Y. Ng. 2006. "A Dynamic Bayesian Network Model for Autonomous 3d Reconstruction from a Single Indoor Image." Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA.

Denoyer, Ludovic, and Patrick Gallinari. 2004. "Bayesian Network Model for Semi-structured Document Classification." *Information Processing and Management* 40 (2004):807-27.

Eskin, Nurdil, and Hamdi Türkmen. 2008. "Analysis of Annual Heating and Cooling Energy Requirements for Office Buildings in Different Climates in Turkey." *Energy and Buildings* 40 (5):763–73.
Gneiting, Tilmann, and Adrian E. Raftery. 2005. "Weather Forecasting with Ensemble Methods." *Science* 310 (5746):248-9.
Hao, He, Chad D. Corbin, Karanjit Kalsi, and Robert G. Pratt. 2016. "Transactive Control of Commercial Buildings for Demand Response." *IEEE Transactions on Power Systems* 32 (1):774-83.
Hou, Zhijian, and Zhiwei Lian. 2009. "An Application of Support Vector Machines in Cooling Load Prediction." Paper presented at the International Workshop on Intelligent Systems and Applications, Wuhan, China.
Hou, Zhijian, Zhiwei Lian, Yao Ye, and Xinjian Yuan. 2006. "Cooling-load Prediction by the Combination of Rough Set Theory and an Artificial Neural-network based on Data-fusion Technique." *Applied Energy* 83 (2006):1033-46.
Huang, Sen, Ana Carolina Laurini Malara, Wangda Zuo, and Michael D. Sohn. 2016. "A Bayesian Network Model for the Optimization of a Chiller Plant's Condenser Water Set Point." *Journal of Building Performance Simulation* In Press.
Huang, Sen, and Wangda Zuo. 2014. "Optimization of the Water-cooled Chiller Plant System Operation." Paper presented at the ASHRAE/IBPSA-USA Building Simulation Conference, Atlanta, GA, USA.
Huang, Sen, Wangda Zuo, and Michael D. Sohn. 2016a. "Amelioration of the cooling load based chiller sequencing control." *Applied Energy* 168 (2016):204-15.
Huang, Sen, Wangda Zuo, and Michael D. Sohn. 2016b. "A Bayesian Network Model for Predicting the Cooling Load of Educational Facilities." Paper presented at the ASHRAE/IBPSA-USA Building Simulation Conference, Salt Lake City, UT, USA
Huang, Sen, Wangda Zuo, and Michael D. Sohn. 2016c. "Improved Cooling Tower Control of Legacy Chiller Plants by Optimizing the Condenser Water Set Point." *Building and Environment* 111(2017):33-46.
Hughes, Justin T., Alejandro D. Domínguez-García, and Kameshwar Poolla. 2015. "Virtual Battery Models for Load Flexibility from Commercial Buildings." Paper presented at the 48th Hawaii International Conference on System Sciences, Hawaii, USA
Jensen, Kasper L., Jørn Toftum, and Peter Friis-Hansen. 2009. "A Bayesian Network Approach to the Evaluation of Building Design and its Consequences for Employee Performance and Operational Costs." *Building and Environment* 44 (2009):456– 62.
Kashiwagi, Norihito, and Toshikazu Tobi. 1993. "Heating and Cooling Load Prediction using a Neural Network System." Paper presented at the International Joint Conference on Neural Networks, Nagoya, Japan.
Kim, Sean Hay. 2011. "Development of Robust Building Energy Demand-side Control Strategy under Uncertainty." Georgia Institute of Technology.
Kim, Sunyong, Seiya Imoto, and Satoru Miyano. 2004. "Dynamic Bayesian Network and Nonparametric Regression for Nonlinear Modeling of Gene Networks from Time Series Gene Expression Data." *BioSystems* 75 (2004):57-65.
Klein, S. A., J. A. Duffie, and W. A. Beckman. 1976. "TRNSYS – A Transient Simulation Program." *ASHRAE Transactions* 82 (1):623-33.
Krati, Moncef. 2016. "*Energy Audit of Building Systems: An Engineering Approach*." Second ed. New York: CRC Press.

Kwok, Simon S.K., Richard K.K. Yuen, and Eric W.M. Lee. 2011. "An Intelligent Approach to Assessing the Effect of Building Occupancy on Building Cooling Load Prediction." *Building and Enviroment* 46 (2011):1681-90.

Leung, M.C., Norman C.F. Tse, L.L. Lai, and T.T. Chow. 2012. "The Use of Occupancy Space Electrical Power Demand in Building Cooling Load Prediction." *Energy and Buildings* 55 (2012):151-63.

Li, Qiong, Qinglin Meng, Jiejin Cai, Hiroshi Yoshino, and Akashi Mochida. 2009a. "Applying Support Vector Machine to Predict Hourly Cooling Load in the Building." *Applied Energy* 86 (2009):2249-56.

Li, Qiong, Qinglin Meng, Jiejin Cai, Hiroshi Yoshino, and Akashi Mochida. 2009b. "Predicting Hourly Cooling Load in the Building: A Comparison of Support Vector Machine and Different Artificial Neural Networks." *Energy Conversion and Management* 50 (2009):90-5.

Li, Xuemei, Lixing Ding, Yan Li, Gang Xu, and Jibin Li. 2010. "Hybrid Genetic Algorithm and Support Vector Regression in Cooling Load Prediction." Paper presented at the third International Conference on Knowledge Discovery and Data Mining, Phuket, Thailand.

Li, Xuemei, Lixing Ding, Jinhu Lv, Gang Xu, and Jibin Li. 2010. "A Novel Hybrid Approach of KPCA and SVM for Building Cooling Load Prediction." Paper presented at the third International Conference on Knowledge Discovery and Data Mining, Phuket, Thailand.

Li, Zhengwei, and Gongsheng Huang. 2013. "Re-evaluation of Building Cooling Load Prediction Models for Use in Humid Subtropical Area." *Energy and Buildings* 62 (2013):442-9.

Ma, Yudong, Francesco Borrelli, Brandon Hencey, Brian Coffey, Sorin Bengea, and Philip Haves. 2012. "Model Predictive Control for the Operation of Building Cooling Systems." *IEEE Transactions on Control Systems Technology* 20 (3):796-803.

O'Neill, Zheng. 2014. Development of a Probabilistic Graphical Energy Performance Model for an Office Building. Paper presented at the 2014 ASHRAE Annual Meeting, Seattle, WA, USA, June 28–July 1.

Pedregosa, Fabian, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 2 (2011):2825-30.

Python Software Foundation. "The official home of the Python Programming Language.". https://www.python.org/.

Reddy, T. Agami. 2011. "*Applied Data Analysis and Modeling for Energy Engineers and Scientists"*, Springer US.

Sakawa, M., S. Ushiro, K. Kato, and K. Ohtsuka. 1999. "Cooling Load Prediction in a District Heating and Cooling System through Simplified Robust Filter and Multi-layered Neural Network." Paper presented at the IEEE International Conference on Systems, Man, and Cybernetics, Tokyo, Japan.

Schaul, Tom, Justin Bayer, Yi Sun, Martin Felder, Frank Sehnke, Thomas Ruckstieß, and Jurgen Schmidhuber. 2010. "PyBrain." *Journal of Machine Learning Research* 11 (2010):743-6.

Široky, Jan, Frauke Oldewurtel, Jirˇí Cigler, and Samuel Prívara. 2011. "Experimental Analysis of Model Predictive Control for an Energy Efficient Building Heating System." *Applied Energy* 88 (2011).

Sun, Yongjun, Shengwei Wang, and Fu Xiao. 2013. "Development and Validation of a Simplified Online Cooling Load Prediction Strategy for a Super High-rise Building in Hong Kong." *Energy Conversion and Management* 68 (2013):20-7.
Sun, Yuming, Yeonsook Heo, Huizhi Xie, C.F. Jeff Wu, and Godfried Augenbroe. 2013. "Uncertainty Quantification of Microclimate Variables in Building Energy Models." *Journal of Building Performance Simulation* 7 (1):17-32.
The U.S. Energy Information Administration. 2016. "The State Energy Data System." Accessed March 9. http://www.eia.gov/state/seds/.
Thevenard, D., and K. Haddad. 2006. "Ground Reflectivity in the Context of Building Energy Simulation." *Energy and Buildings* 38 (8):972–80.
Toftum, J., R.V. Andersen, and K.L. Jensen. 2009. "Occupant Performance and Building Energy Consumption with Different Philosophies of Determining Acceptable Thermal Conditions." *Building and Environment* 44 (2009):2009-16.
U.S. Department of Energy. 2014. "Buildings Energy Data Book." Accessed May 15. https://catalog.data.gov/dataset/buildings-energy-data-book.
Walter, Travis, Phillip N. Price, and Michael D. Sohn. 2016. "Uncertainty Estimation Improves Energy Measurement and Verification Procedures." *Applied Energy* 130 (2014):230-6.
Walter, Travis, and Michael D. Sohn. 2016. "A Regression-based Approach to Estimating Retrofit Savings using the Building Performance Database." *Applied Energy* 179 (2016):996-1005.
Wetter, Michael, Marco BonvinI, and Thierry S. Nouidui. 2016. "Equation-based Languages – A New Paradigm for Building Energy Modeling, Simulation and Optimization." *Energy and Buildings* 117 (2016):290-300.
Xiao, Fu, Yang Zhao, Jin Wen, and Shengwei Wang. 2016. "Bayesian Network based FDD Strategy for Variable Air Volume Terminals." *Automation in Construction* 41 (2016):106-18.
Xue, Xue, Shengwei Wang, Yongjun Sun, and Fu Xiao. 2014. "An Interactive Building Power Demand Management Strategy for Facilitating Smart Grid Optimization." *Applied Energy* 116 (2014):297-310.
Yao, Ye, Zhiwei Lian, Shiqing Liu, and Zhijian Hou. 2004. "Hourly Cooling Load Prediction by a Combined Forecasting Model based on Analytic Hierarchy Process." *International Journal of Thermal Sciences* 43 (2004):1107-18.
Yu, David C., Thanh C. NguyeN, and Peter Haddawy. 1999. "Bayesian Network Model for Reliability Assessment of Power Systems ". *IEEE Transactions on Power Systems* 14 (2):426-32.
Zhang, Fan, Chirag Deb, Siew Eang Lee, Junjing Yang, and Kwok Wei Shah. 2016. "Time Series Forecasting for Building Rnergy Consumption using Weighted Support Vector Regression with Differential Evolution Optimization Technique." *Energy and Buildings* 126 (2016):94-103.
Zhang, Lei, and Qiang Ji. 2011. "A Bayesian Network Model for Automatic and Interactive Image Segmentation." *IEEE Transactions On Image Processing* 20 (9):2582-93.