Data Driven Methods for Energy Reduction in Large Buildings

Avisek Naug
Institute for Software Integrated Systems
Vanderbilt University
Nashville,TN, USA
avisek.naug@vanderbilt.edu

Gautam Biswas
Institute for Software Integrated Systems
Vanderbilt University
Nashville,TN, USA
gautam.biswas@vanderbilt.edu

Abstract—Modeling of HVAC components and energy flows for energy prediction purposes can be computationally expensive in large commercial buildings. More recently, the increased availability of building operational data has made it possible to develop data-driven methods for predicting and reducing energy use for these buildings. In this paper, we present such an approach, where we combine unsupervised and supervised learning algorithms to develop a robust method for energy reduction for large buildings operating under different environmental conditions. We compare our method against other energy prediction models that have been discussed in the literature using (1) a benchmark data set and (2) a real data set obtained from a building on the Vanderbilt University campus. A Stochastic Gradient Descent method is then applied to tune the controlled variable *i.e.*, the AHU discharge temperature set point so that energy consumption is "minimized".

Index Terms—Data driven modeling, HVAC, Hierarchical Clustering, AdaBoost, Random Forests, Stochastic Gradient Descent.

I. INTRODUCTION

Large commercial buildings typically consume large amounts of energy [1] and the HVAC systems of the buildings are usually the most significant contributors to the energy footprint of these buildings. Given the recent moves toward sustainability and reduced fossil fuel use, researchers have begun to study the key components in HVAC energy consumption and to develop methods for energy use reduction [2], [3]. In our research, we have been applying machine learning algorithms to develop data-driven analytic models of energy consumption, and then apply optimization methods to these models to reduce the energy use of the building for different environmental conditions.

Two primary approaches have been developed for modeling energy consumption in buildings. Model-based approaches rely on the laws of physics and building configurations to develop energy and mass balance equations that model the energy consumption. We review some of these commonly used approaches in Section II-A. However, building accurate physics based models is computationally expensive and time-consuming. As an alternative, as more data on building energy consumption profiles have become available, researchers and practitioners have switched to alternate forms of modeling that are derived from data-driven and machine learning methods.

Popular machine learning methods include Support Vector Machines, Boosting methods, and auto-regressive neural network models. Data-driven methods are reviewed in section II-B.

In this paper, we develop a new data driven technique for modeling energy consumption in large buildings, and then apply a stochastic optimization method to reduce the energy footprint of the building under different environmental conditions. We combine unsupervised learning methods to characterize the different environmental conditions under which buildings operate, and then develop data-driven models of energy consumption for each set of environmental conditions. We then apply optimization methods to the derived analytic models to find operating points that reduce the energy consumption of the building. We demonstrate the effectiveness of our approach by comparing our model's prediction performance with other methods that have been discussed in the literature using a benchmark building energy dataset. As a second experiment, we apply our approach to a building for which we acquired energy consumption data from the Vanderbilt Plant Operations.

The rest of the paper is organized as follows. Section II reviews the existing state of the art in energy prediction and optimization methods. In Section III, we describe the configuration of the building that we have used for our experimental studies. Section IV describes the work flow for energy optimization, and the components of the work flow. Section V discusses the results of our experimental studies, and a comparison of our results to some of the other methods discussed in the literature. Finally, Section VI presents the summary and conclusions of the paper.

II. LITERATURE REVIEW

Parameterized analytic models of building energy consumption can be used to "optimize" energy consumption by tweaking control parameter settings that reduce energy use. Typically, the inputs associated with such models include environmental conditions and the thermostat settings in different locations of the building. The output variable is the building energy consumption, and control parameters include the settings for air handling, heating and cooling units of the building. Once we have an analytic model of the system, the control variables can be adjusted during operation to optimize energy consumption. There are two common approaches to

building models of a system: (1) Physics Based and (2) Data Driven. Representative approaches in either category are discussed below.

A. *Physics Based Modeling for Optimizing Energy Consumption*

These methods apply physical laws of thermodynamics, heat and mass flow to model building energy consumption. Data collected from the system may be used to estimate the system parameter values in the model. Control algorithms are then employed to dynamically set values to some of the system parameters and reduce energy consumption without sacrificing comfort levels in the building. In our case, we might have a model where energy use is modeled using thermodynamic equations. We use an optimization algorithm to tune the discharge temperature variable such that the energy consumption for the process is reduced.

Ghiaus *et al.* [4] used a scaled laboratory model of an HVAC system to model the interactions between different subsystems using a set of thermodynamic and mass balance equations. They used a controlled static relay to feed power to the system and noted the difference in temperature between the input and output air of the coils inside the model. Using this data, they estimated the parameters of the thermodynamic equations. Once these model parameters were estimated, the temperature set points were tuned for the current set of environmental conditions and temperature set points using an optimization algorithm to minimize the energy consumption. They validated their approach under a controlled setup, but inferring parameters of a real dynamic model with uncertainties, noise and disturbances in the measurements can be very difficult.

Another common approach for energy optimization in buildings is to use energy simulation software like EnergyPlus, DOE-2 and TRNSYS [5], [6]. These software packages can be integrated with sensor data from different locations of the building to continuously track building loads and develop a virtual model of the building energy consumption. These methods use preset schedules to optimize electrical loads within the building. However, [7] shows that these method may incur more energy cost during peak energy consumption hours.

Huang *et al.* [8] have used a R-C network modeling approach to capture the thermodynamic interactions between different HVAC components of an airport terminal. Model Predictive Control is used to minimize the energy consumption over a finite time horizon by varying the temperature setting within a pre-specified range. They achieved a theoretical 5-18% energy savings using this approach.

Our project involves HVAC systems that use a combination of Air Handling Units(AHU), Variable Refrigerant Flow (VRF) systems, Steam based Heaters and a network of air, chilled water and steam flow systems. To adapt a model-based approach, we have to construct first principles models of the individual components and compose them into a an integrated model of energy flows and consumption in the building. Constructing a system-level model that can simulate the system energy flows and consumption in a sufficiently accurate manner can be a difficult and resource intensive task. The complexities and costs in building sufficiently accurate models for large, complex buildings, naturally motivates data driven models that are reviewed next.

B. *Data Driven Approach for optimizing Energy Consumption*

Since building accurate models of energy flow in buildings can be computationally expensive, an alternative approach may be to derive data-driven models of energy flow and consumption in a building. Progress in machine learning algorithms have made it easier to construct models that make accurate predictions of building dynamics and energy use. Given the model, control variables in the model can be tuned using methods, such as gradient descent, and evolutionary optimization algorithms (e.g., Genetic Algorithm, Differential Evolution) to optimize the outcome variables of interest.

Artificial Neural Nets (ANNs) and Support Vector Regression have been successfully employed to model heating and cooling loads and thermal conditions inside the buildings. Researchers (e.g., [9]–[14]) use weather data like outdoor dry bulb temperature, relative humidity, wind speed, direct and diffuse solar radiation and indoor temperature settings as input and cooling, heating and electrical load data as the output to train these models. The control variables in the model, *viz.*, the indoor temperature set points are then tuned with pre-specified constraints to optimize energy consumption. Novel approaches, such as [15] use Kernel Ridge regression with a K-Nearest Neighbor (K-NN) implementation for efficient prediction of energy consumption in buildings. [16] used Regression Trees for the same purpose.

Data-driven approaches provide a promising alternative to physics-based energy flow modeling, especially when complexity of the system precludes developing sufficiently accurate models. Data-driven models are often black box models that employ abstractions to accommodate complex interactions and enhance the practical utility of the model for prediction. But the models may not have a direct physical interpretation. It should also be kept in mind that such models are susceptible to sample bias. Therefore, it is important to ensure the data represents a sufficiently large spectrum of operating scenarios.

III. DESCRIPTION OF THE SYSTEM

Large buildings typically have a central Building Automation System (BAS) linked to distributed HVAC systems for climate control of a number of interconnected spaces to maintain occupant comfort. The BAS is used by a building personnel to monitor the building conditions and perform supervisory control.

The Alumni Hall at Vanderbilt University is a three-storied building with a collection of small office spaces, large halls, which also serve as meeting spaces, classrooms, cafeteria on the first floor and a gym in the basement. Since it is an old building, its HVAC system has been retrofitted with Variable Refrigerant Flow (VRF) Units. Figure 1 shows the main components of the HVAC system: the Air Handling unit
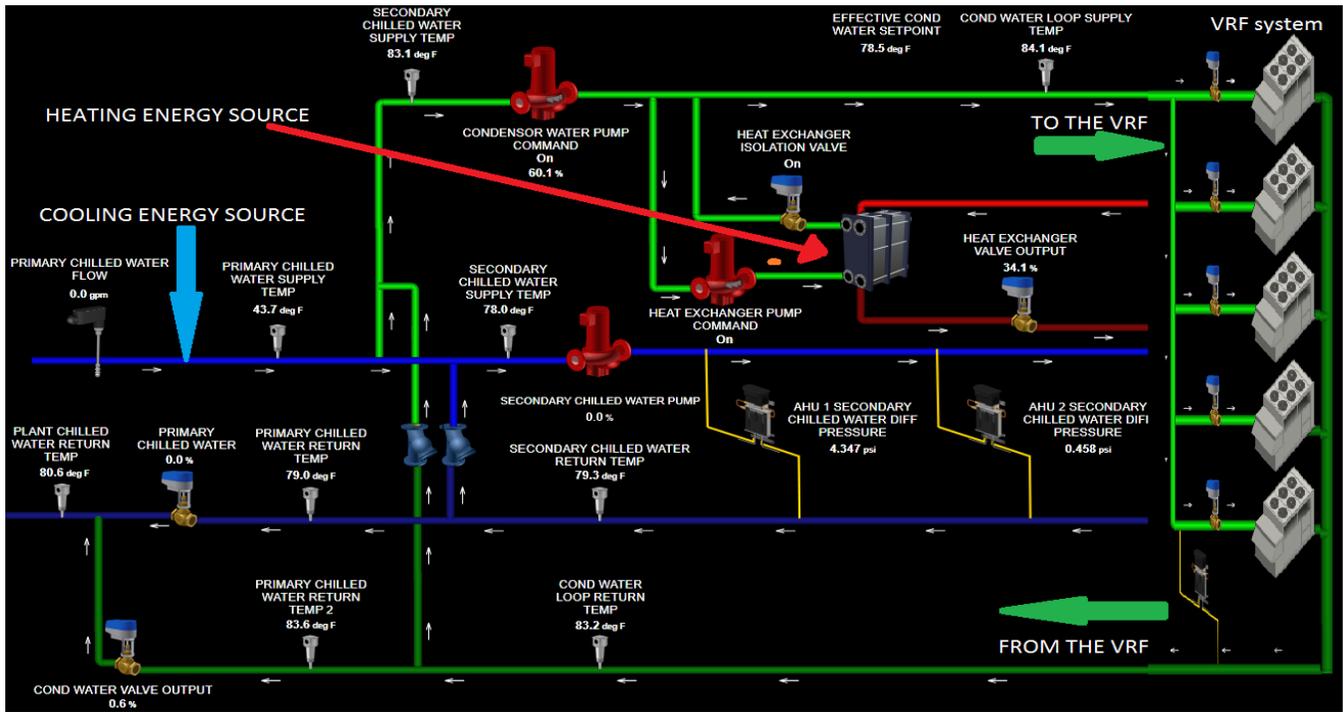
Fig. 1. Schematic of the Energy Flow in the HVAC system at the Alumni Hall at Vanderbilt University

(AHU), the steam generation unit, and the VRF subsystems. The piping through which the chilled water flows and connects to the different parts of the system are also shown.

The Air Handling Unit within the building is only used to discharge neutral air at 50% relative humidity and a pre-set temperature of 68^oF. The subsystems in the AHU: the cooling, preheat and reheat coils are its main energy consumers. They are regulated by a control set point to humidify and dehumidify the air. The primary heating and cooling energy source in the building is the chilled water, which is supplied by a central plant on the Vanderbilt campus. The hot water, provided by a steam generator, is used to warm up and add energy to the chilled water on cold days using a heat exchanger. Both units are placed inside the building. The current reset schedule for the HVAC operates based only on the ambient humidity, which is not implemented in an energy efficient manner, according to the plant operators. The air has to be cooled down to remove moisture and reheated to a certain level before being released inside the building. When the AHU releases dehumidified air at a higher temperature but regions in the building need to cool them down to lower temperatures, it results in unnecessary energy expenditure. This inefficiency in the system motivated us to look for methods that provide more energy efficient control in large buildings.

Our approach is designed to adjust the discharge temperature set point of the dehumidified air taking into account the environmental conditions (outside temperature, outside humidity, and solar irradiance). To relate to the model that we want to develop, the environmental variables are the input data to our model, and the chilled water and hot water energy

consumption are considered as the output or outcome variable that we wish to minimize, and the AHU discharge temperature is the control variable that we will manipulate to achieve the "minimum" energy use.

IV. PROPOSED APPROACH

Our goal in this work was to adopt a data-driven approach to build a parameterized predictive model of the energy consumption of the building as a function of the AHU temperature set point and environmental variables, which are the input variables. More formally, if the building Energy consumption E can be represented as a function $f(.)$ of the set of inputs that include $\mathbf{X}$, the primary environmental conditions that affect building energy consumption, and θ), a set of control variables of the system. We develop an optimization schedule to perturb the control variables θ under some constraints to get an optimum energy consumption. In other words, we solve the following two problems: Find the function $f(.)$ which predicts the Energy Consumption E according to the relation

$$E = f(\mathbf{X}, \theta)$$

and then find

$$\theta = argmin_\theta f(\mathbf{X}, \theta)$$

In our work, $\mathbf{X}$ is three-dimensional, and made up of the environmental variables: (1) the Outside Air temperature, (2) the Outside Air Relative Humidity and (3) the Solar Insolation. The control variable, the set point θ, is the temperature set point of the Air Handling Unit that will be manipulated to "minimize" energy consumption.
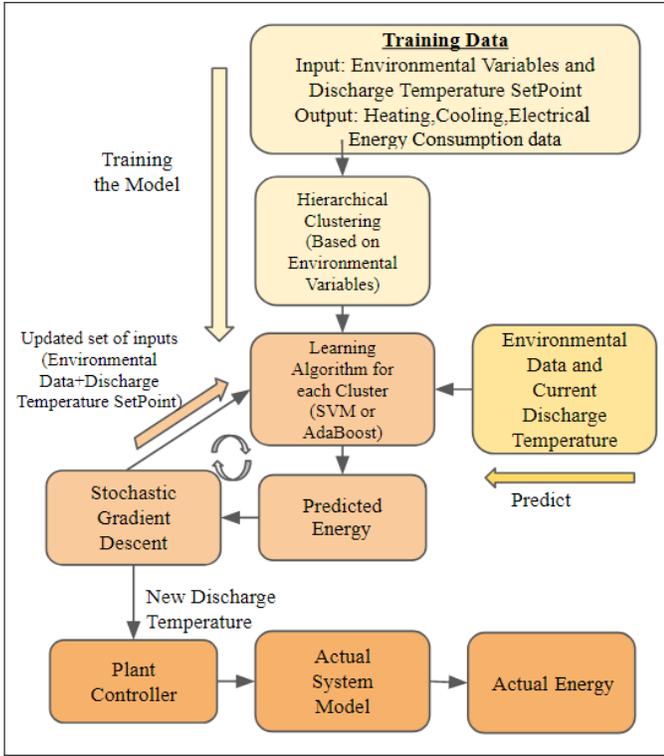
Fig. 2. Schematic of the proposed Framework

Use of ensemble learning methods for energy prediction has not been explored much in the past, but we believe it is useful because variation of energy use in buildings can often be difficult to learn using one regression function. [16] applies a popular ensemble method, Random Forests, a well-known adaptive bagging technique for energy prediction. We propose to use AdaBoost [17] employing Regression Trees as base estimators. Even then a single model cannot be sufficient to describe the system because buildings operate in different modes during different times of the day, and across the different seasons. To accommodate these variations, we applied a hierarchical clustering algorithm [18] to identify the different environmental conditions under which the building operated.

We used the three environmental features (outside air temperature, relative humidity, and the solar insolence) to define the data points (total energy consumption at 10 minute intervals) for clustering. Before running the clustering algorithm, we scaled the features to prevent biased cluster formation since solar insolation values typically have higher variance compared to other environmental variables. We considered each cluster to represent different environmental conditions that would entail a different mode of operation for the building. As a result, the energy cost function, f, was derived separately for each cluster of environmental data using the AdaBoost regression technique. We also built Support Vector regression model and compared the results produced by the two methods.

For the optimization step, we assumed the regression model as the cost function and the current weather conditions and the current discharge temperature are used for initializing a Stochastic Gradient Descent(SGD) optimization algorithm [19]. The discharge temperature of the AHU is the primary variable that is adjusted to find a better operating condition while the environmental variables are constrained to vary by small amounts to enable the SGD to explore the search space. Finally, this new discharge temperature was passed to the plant controller of the AHU system. The controller could then automatically adjust the heating and cooling valves of the system to moderate the discharge air temperature. The schematic of the entire process is shown in Figure 2. Details of the algorithms in relation to the application are presented next.

A. Hierarchical Clustering

We employed an agglomerative approach to building clusters. In this paper, we use the Euclidean metric and the *Ward* minimum variance method for minimizing intra cluster variance. The Hierarchical Clustering approach has an advantage over other traditional algorithms, like *K-Means* because additional visual inspection of the generated dendogram can be applied as a heuristic for deciding on the number of clusters to be formed. Once the number of clusters was established, we computed the cluster center for each cluster. It should be noted here that the clusters indicated significant information about the environmental conditions, which is later explained in Table II. During the test phase, a new data point was first assigned to a cluster based on proximity to the cluster center, and then the optimal discharge temperature was computed using the function F associated with that cluster.

B. Support Vector Regression

One of the techniques we used to learn the energy consumption was the Support Vector Regression [20]. The idea is to map a non linear regression function in the input space represented by the 4 dimensional data $\mathbf{X}, \theta$ to a high dimensional space, where it is represented by the hyperplane that minimizes the size of an ϵ tube surrounding the plane that contains all N observations in the training data. Observations lying beyond the tube by an amount $\mathcal{E}_i$ (or $\mathcal{E}_i^*$ if it lies on the other side of the hyperplane) for each observation i are penalized linearly by the constant λ. If ω denotes the hyperplane and b is the intercept, SVR solves the following convex optimization problem given by

$$\mathbf{min}_\omega \frac{||\omega||^2}{2} + \frac{\lambda \sum_{i=1}^{i=N}(\mathcal{E}_i + \mathcal{E}_i^*)}{N} \quad (1)$$

subject to the following constraints,

$$y_i - \omega.\phi(\mathbf{X}, \theta_i) - b \leq \epsilon + \mathcal{E}_i \qquad i = 1, 2 \ldots N \quad (2)$$

$$\omega.\phi(\mathbf{X}, \theta_i) + b - y_i \leq \epsilon + \mathcal{E}_i^* \qquad i = 1, 2 \ldots N \quad (3)$$

$$\mathcal{E}_i^* \geq 0 \qquad i = 1, 2 \ldots N \quad (4)$$

For our problem we used the energy values to be predicted as the set of y_i's. Through grid search, the regularization parameter was found to be 1.5 for almost every cluster.

C. Adaptive Boosting

AdaBoost [17] is a training methodology where a sequence of T learning machines (like regression trees) are trained on samples of the data in a step by step manner over T iterations. In each iteration j, we created the training set for the jth regression tree using N_1 samples selected with replacement from the training set $(\mathbf{X}, \theta)$. It should be kept in mind that $(\mathbf{X}, \theta)$ denotes the data corresponding to a specific cluster that was derived earlier. Any instance i has a probability of selection given by $p_i = \frac{w_i}{\sum_i w_i}$, where the training instances are weighted according to whether the regression algorithm can predict it with sufficient accuracy in the previous iteration. w_i^j is the weight we assigned to the ith instance. This helps us focus more on $(\mathbf{X}, \theta)$s that were difficult to learn from the actual data. We denote the jth regression tree by $\hat{y}^j(\mathbf{X}, \theta)$. Next, we evaluated the prediction error on each data point of the sample denoted by L_i using $\hat{y}^j(\mathbf{X}, \theta)$.

$$L_i = Loss \left[|\hat{y}_i^j(\mathbf{X}, \theta)_i - y_i| \right] \qquad (5)$$

The average loss for all the instances generated by the regression tree $\hat{y}^j(\mathbf{X})$ is given by the following expectation measure

$$L = \sum_{i=1}^{N_1} L_i * p_i \qquad (6)$$

The average loss helps us generate a measure of confidence for the jth regression function $\hat{y}^j(\mathbf{X}, \theta)$. We have a mathematical measure for our confidence denoted by β_j,

$$\beta_j = \frac{L}{1 - L} \qquad (7)$$

Thus β_j helps our model decide how much importance to attach to the jth regression model when predicting a test instance. Then at the end of iteration j, w_i^j is updated using

$$w_i^{j+1} \leftarrow w_i^j * \beta_j * (1 - L_i) \; \forall i = 1 \ldots N_1 \qquad (8)$$

Thus we update the importance weights of the samples selected at the jth iteration depending on the amount of error. It becomes evident that w_i is assigned lower values if instance i is predicted without much loss. We use the simple linear loss function scaled to the interval $[0, 1]$ by dividing each L_i by the $\max_i L_i$. This was done to ensure β_j is positive in equation 7.
In order to predict the y_i for a test instance $(\mathbf{X}, \theta)_i$, it is passed to all the T regression functions and the cumulative prediction is made according to a weighted median [17] approach. The predictions of all the regression functions are arranged in increasing order of magnitude and the β_js associated with them are also rearranged. Then starting from the index with smallest $\hat{y}_i^j$, the cumulative sum $log\beta_j$ is calculated until it is greater than $\frac{\sum_{j=1}^T log\beta_j}{2}$. The index j for which this is satisfied is taken to be the ensemble prediction. Naturally, if all the regression functions have equal values for β, this would simply indicate a median prediction.

D. Stochastic Gradient Descent

Stochastic Gradient Descent is an approximation to the Gradient Descent optimization algorithm used to obtain the optimum of a cost function. We use the regression model $F((\mathbf{X}, \theta)_i)$ (using SVR or AdaBoost), the current set of environmental variables and temperature set points included in the vector $(\mathbf{X}, \theta)_i$ and iteratively converge to a better temperature set point using a search method. Mathematically SGD updates the parameters sequentially using the following equation,

$$\theta_{\mathbf{i}t} = \theta_{\mathbf{i}t-1} - \alpha \nabla_\theta F((\mathbf{X}, \theta)_i), \qquad (9)$$

where α is the learning rate for the problem and ∇_θ indicates the gradient with respect to the control variable θ. This formula can be applied when the cost function is pseudo convex. Since we trained the classifier over a large dataset, using SGD proved more useful than batch gradient descent since the later makes a pass over all the data before making a single step towards an optimum. Moreover, SGD can converge very fast.

V. RESULTS

We compare the prediction accuracy of our model with other energy prediction models proposed in the literature, in particular the ones that were used in a Energy Predictor Shootout competition.

A. Prediction Accuracies on the Benchmark data set: Great Energy Predictor Shootout (EPS) data set 1992

The EPS data set is used as a benchmark data set for comparing efficiency of algorithms in predicting energy consumption. The data set was initially presented as a part of the first ASHRAE competition on building data analysis in 1992. There are eight variables, recorded over a period of 4 months (Sept-Dec) at 1 hour intervals. Four of the variables are related to the weather: Outside Air Temperature in ^oF, Outside Air Humidity Ratio (wt of water/wt of dry air), solar insolation (irradiance) (watts/ m^2) and Wind Speed (in m.p.h). Machine learning methods were employed to learn regression models from the first four months of data, and then used to predict the three energy variables: chilled water energy (CHW), hot water energy (HW) and whole building electricity consumption (WBE) as a function of the environment variables for the next two months.

In this paper, we compare the prediction accuracy using the Coefficient of Variation Root Mean Square Error (CVRMSE) metric:

$$CV_{RMSE} = \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}}{\bar{y}} \qquad (10)$$

As discussed in section IV we clustered the data to enable the derived energy models to be a function of different environmental modes. The data grouped into clusters that indicated the time of day.

We trained the AdaBoost algorithm on the training data from each cluster using a 10-fold cross validation approach. During the testing phase, the two month data was assigned to the respective clusters and the Heating, Cooling and Whole

Method	WBE CV_{RMSE}	CHW CV_{RMSE}	HW CV_{RMSE}	Average CV_{RMSE}
Team#9*	10.36*	13.02*	15.24*	15.24*
HC+AdaBoost	**13.22**	**10.21**	**24.24**	**15.89**
HC+RF	**13.72**	**10.08**	**24.65**	**16.15**
RF*	11.72*	14.88*	28.13*	18.24*
KR	13.61	12.40	33.01	19.67
Team#6*	11.78*	12.97*	30.63*	18.46*
Team#3*	12.79*	12.78*	30.98*	18.85*
Team#2*	11.89*	13.69*	31.65*	19.08*
Team#7*	13.81*	13.63*	30.57*	19.34*
HC+SVR	**19.09**	**11.84**	**30.23**	**20.39**

HC:Hierarchical Clustering.   RF:Random Forests.
KR:Kernel Regression, reproduced from [15]   *:reproduced from [16].

Building Electricity energy estimates were computed. To establish the effectiveness of clustering in improving prediction, we applied a similar method to predict energy consumption using Support Vector Regression and Random Forests. The results were compared with the prediction accuracies of top teams that participated in the ASHRAE competition, the Random Forests approach employed by [16], and the Kernel Regression technique used in [15]. The results are shown in Table I. Our Random Forest and AdaBoost methods generated prediction accuracies that were better than most of the methods. The CV_{RMSE} was better by 2 to 3% than the teams placed 2nd through 5th. The SVR method did not perform well probably because of the skewness in the weather data.

B. Prediction Accuracies on the Alumni Hall data set

In order to see the prediction capabilities of the Hierarchical Clustering based AdaBoost, Support Vector Regression and Random Forests for a second scenario, we applied our algorithms to predict the Heating and Cooling energies of the Vanderbilt Alumni Hall data set. The data was collected for 1 year from October 2016 to September 2017 at 1 hour intervals. However, due to missing data in the Hot Water energy consumption (in BTUs) we used fewer points where all the variables were simultaneously logged. There were two variables for energy prediction: Chilled Water energy consumption (in BTUs) associated with the chilled water line circulating through out the building, acting as a heat sink and the Hot Water/Steam energy consumption, which is the heat source within the building. There were 3 input variables: the Outside Air Temperature, the Outside Air Relative Humidity and the Solar Insolation. The AHU discharge temperature was the control variable.

As described earlier, as a first step, we clustered the data. Table II shows that the time of the day as well as season explains the clusters formed. [15] maps the time of the day to a unit circle to explain the periodicity in energy values and used a kernel regression method to predict energy consumption. We did not use the time variables in our prediction algorithm.

The CV_{RMSE} metric was used to calculate the error in prediction for each of the three methods: AdaBoost, Random Forests and SVR with and without Hierarchical Clustering. During training under a clustered approach, the 12 month test data was grouped into 6 clusters. The data in each cluster was split into two: 75% training and 25% testing. The training involved a 10-fold cross validation. A problem we faced was that the energy function for cluster 3 could not be learned because of the large irregular variation in the data.

As a comparison, to show the effectiveness of the clustering, we used the entire data to learn the energy function, again using 75% of the data for training and 25% for testing. A similar 10-fold cross validation was used to train on the entire train data set in this case.

We tested our regression models and the results are reported in Table III. We ran the above experiment for each method 10 times to get an estimate of the average CV_{RMSE}. We noticed that there is a significant decrease in the CV_{RMSE} error value when we use the clustered approach. This is also supported by performing an unpaired t-test on the corresponding methods where we compare the CV_{RMSE} for 10 samples, one under a clustered approach and one without clustering. The $p-$value was found to be very low ($p = 0.001$) for all the three regression methods supporting the hypothesis that grouping the data by environmental conditions improves accuracy.

Given the strong nonlinearities in the data, AdaBoost, which uses an ensemble of models performs a little better than the Random Forests. This is mostly because some of the scenarios in Alumni Hall show sensor readings where the building conditions are such that because of energy transfer between different sections of the building, the net heating and cooling energy consumption within the building become 0. Sometimes, this scenario occurred multiple times during a day. Another situation is frequently encountered during the late spring and early fall when the outside Relative Humidity remains in a pleasant range during the day. But at night as temperature of the air drops significantly, the capacity of the air to hold moisture decreases and Relative Humidity sensors record a high value immediately causing the Air Handling Unit to switch to climate control mode. This causes both the cooling and heating energy consumption to rise temporarily. In typical building scenarios, the AHUs are mainly guided by temperature requirements but in our system in Alumni Hall the AHU is guided by humidity. Therefore, as stated, there are situations where there were sudden spikes in AHU energy consumption. These are examples of scenarios that are hard to accommodate in a single energy consumption prediction model, but AdaBoost can use a sequence of separate regressors to learn them in a more accurate manner.

We also notice that the accuracies compared to those in the EPS data is lower. There could be several reasons for this. For one, there are other variables, such as thermostat settings that affect energy consumption. We did not have access to the data. Second, the mode of operation for the building was frequently changing based on the controller settings whose schedule was not known to us.

TABLE II
MEAN STATISTICS OF THE CLUSTER CENTERS FOR ALUMNI HALL DATA

Cluster	Outside Air Temperature (F)	Outside Air Relative Humidity(%)	Direct Irradiance ($\frac{W}{m^2}$)
High Temp Med RH (Sunny & Summer)	72.9 (11.8)	59.2 (19.5)	321 (79.8)
High Temp Low RH (Aft & Eve &Summer)	79.9 (8.4)	38.3 (10.5)	77.6 (77.6)
High Temp Med RH (Whole Day &Whole Year)	69.4 (7.2)	61.4 (11.5)	49 (61.9)
Low Temp Low RH (Whole Day & Winter)	42.4 (10.9)	43.6 (13.5)	45 (72.3)
Med Temp High RH (Eve&Night &Summer&Winter)	64.4 (7.4)	91.5 (6.7)	18 (40.1)
Low Temp High RH (Whole Day &Mostly Winter)	64.4 (7.4)	91.5 (6.72)	18 (40.1)

Method	CHW CV_{RMSE} (std)	HW CV_{RMSE} (std)	Average CV_{RMSE} (std)
HC+AdaBoost	22.14 (0.61)	34.79 (0.53)	28.47 (0.57)
HC+Random Forests	28.59 (0.91)	32.88 (0.24)	30.73 (0.57)
HC+SVR	38.01 (0.86)	42.14 (0.92)	40.08 (0.80)
AdaBoost	27.71 (1.15)	43.76 (1.69)	35.74 (1.42)
Random Forests	36.26 (2.01)	43.61 (1.21)	39.93 (1.61)
SVR	54.34 (1.79)	71.1 (4.19)	62.73 (2.99)

HC:Hierarchical Clustering.
SVR:Support Vector Regression.

C. Optimizing Energy Consumption using the model

In this part of the study, we used the Hierarchical Clustering based AdaBoost algorithm since it produced more accurate predictions for the Alumni Hall data set. As we recall from section V-B, one of the inputs to the model was the Temperature set point of the AHU system. We should emphasize here that a better control problem could have been solved if we could access the individual thermostat set points within the rooms. We used the set of control variables that were accessible to us, but our approach can be easily extended to situations where multiple control variables are present. To derive the optimal energy consumption for a given set of environmental conditions, we ran the Stochastic Gradient Descent algorithm.

Our implementation of the optimization algorithm progresses as follows: We note the current values of the four input variables: the outside air temperature, the outside air

TABLE IV
PROGRESS OF OPTIMIZATION ON THE TOTAL COST FUNCTION
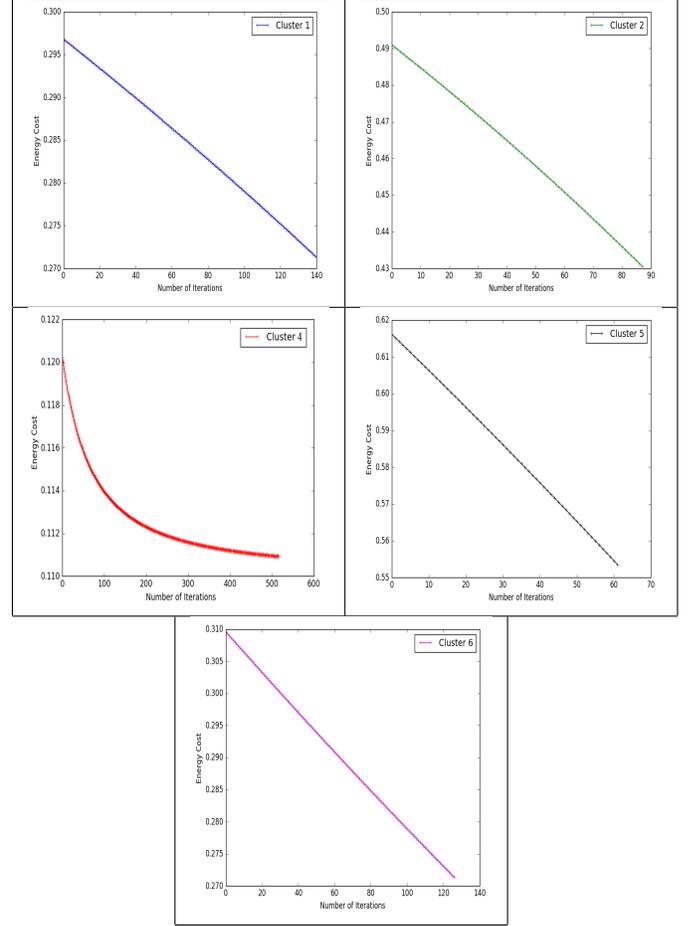
relative humidity, solar insolation value and Air Handling Unit discharge temperature set point. We assign this input to the cluster whose center this data point is closest to. Once we decide the cluster, we consider the summation of the functional form of the heating and cooling load models as a single cost function: the input to each model are the three input variables and the cost is sum of the energy consumed by the chilled water and hot water systems. These current input values and the control variable are used to initialize the Stochastic Gradient Descent Algorithm. We allowed the temperature set point to vary only by $\pm 2^o F$ around the current set point value for the reasons of accuracy, as discussed earlier. We also allowed the other three environmental variables to vary by a small range of $\pm 1^o F$ for outside air temperature, $\pm 2\%$ for outside air relative humidity and $\pm 10 \frac{W}{m^2}$ for incoming solar insolation. This allows Stochastic Gradient Descent to explore the search space in a more effective manner. The stochastic gradient algorithm traverses the search space such that the sum of the hot water energy consumption and cold water energy consumption is minimized. The discharge temperature corresponding to the smallest total energy point is then reported as the output of the optimization algorithm.

TABLE V
ENERGY SAVINGS IN EACH CLUSTER AVERAGED OVER 20 RUNS

Cluster	Average Percentage Savings (Standard Deviation)
Cluster 1	11.59 (0.45%)
Cluster 2	14.53 (0.55%)
Cluster 4	12.92 (0.36%)
Cluster 5	10.25 (0.21%)
Cluster 6	12.35 (1.12%)

The results of applying the Stochastic Gradient Descent to each of the clusters are shown in Table IV for a set of test cases. The *Y-axis* shows the scaled values of the energy consumption in BTU since the models were trained on normalized data. The *X-axis* shows the number of iterations over which the algorithm ran. Iteration 0 shows the total Chilled Water and Hot Water energy consumption under the current environmental conditions, while the end of iteration shows the total Chilled Water and Hot Water energy consumption after SGD has found an new optimal point for the discharge temperature value. The percentage energy savings in each cluster averaged over 10 different initializations is shown Table V. We see that on average the savings is about 12%.

VI. SUMMARY AND CONCLUSIONS

This paper has successfully applied a data driven approach to modeling and subsequently optimizing the energy consumption in large buildings, where energy flow models are not available, and would be very expensive to develop. Machine learning based models like AdaBoost, Random Forests and Support Vector Regression are applied to energy and environmental data available for buildings to predict the total energy consumption for the building. These models are then treated as an energy cost function with certain control variables as inputs to the function. An optimizing algorithm: Stochastic Gradient Descent is used to obtain a setting for the control variable at which the energy cost is minimum. The new values of the control variables are then recommended as the set points for the HVAC subsystems.

When compared again a standard data set, our approach produces better results compared to other algorithms discussed in the literature. When used with a Vanderbilt building data set, our algorithm predicts a 12% decrease in energy consumption over the current control algorithms that are applied. In future work, we will work to incorporate additional energy related features to further improve the accuracy of prediction modeling and improve the energy savings that our approach provides. We will also extend our approach to combine model and data driven methods for more real time monitoring of building energy consumption.

ACKNOWLEDGMENT

REFERENCES

[1] U. Berardi, "A cross-country comparison of the building energy consumptions and their trends," *Resources, Conservation and Recycling*, vol. 123, no. Supplement C, pp. 230 – 241, 2017.
[2] H. Geller, P. Harrington, A. H. Rosenfeld, S. Tanishima, and F. Unander, "Polices for increasing energy efficiency: Thirty years of experience in oecd countries," *Energy Policy*, vol. 34, no. 5, pp. 556 – 573, 2006, hong Kong Editorial Board meeting presentations.
[3] J.-H. Kamlage and P. Nanz, "Crisis and participation in the european union: Energy policy as a test bed for a new politics of citizen participation," *Global Society*, vol. 31, no. 1, pp. 65–82, 2017.
[4] C. Ghiaus, A. Chicinas, and C. Inard, "Grey-box identification of air-handling unit elements," *Control Engineering Practice*, 2007.
[5] D. B. Crawley, C. O. Pedersen, L. K. Lawrie, and F. C. Winkelmann, "Energyplus: energy simulation program," *ASHRAE journal*, vol. 42, no. 4, p. 49, 2000.
[6] S. Karmacharya, G. Putrus, C. Underwood, and K. Mahkamov, "Thermal modelling of the building and its hvac system using matlab/simulink," in *Environment Friendly Energies and Applications (EFEA), 2012 2nd International Symposium on*. IEEE, 2012, pp. 202–206.
[7] M. Behl, F. Smarra, and R. Mangharam, "Dr-advisor: A data-driven demand response recommender system," *Applied Energy*, vol. 170, pp. 30–46, 2016.
[8] H. Huang, L. Chen, and E. Hu, "Model predictive control for energy-efficient buildings: An airport terminal building study," in *Control & Automation (ICCA), 11th IEEE International Conference on*. IEEE, 2014, pp. 1025–1030.
[9] A. E. Ben-Nakhi and M. A. Mahmoud, "Cooling load prediction for buildings using general regression neural networks," *Energy Conversion and Management*, vol. 45, no. 13, pp. 2127–2141, 2004.
[10] J. Yang, H. Rivard, and R. Zmeureanu, "On-line building energy prediction using adaptive artificial neural networks," *Energy and buildings*, vol. 37, no. 12, pp. 1250–1259, 2005.
[11] B. B. Ekici and U. T. Aksoy, "Prediction of building energy consumption by using artificial neural networks," *Advances in Engineering Software*, vol. 40, no. 5, pp. 356–362, 2009.
[12] H.-x. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586–3592, 2012.
[13] Q. Li, Q. Meng, J. Cai, H. Yoshino, and A. Mochida, "Applying support vector machine to predict hourly cooling load in the building," *Applied Energy*, vol. 86, no. 10, pp. 2249–2256, 2009.
[14] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy," *Applied Energy*, vol. 123, pp. 168–178, 2014.
[15] M. Brown, C. Barrington-Leigh, and Z. Brown, "Kernel regression for real-time building energy analysis," *Journal of Building Performance Simulation*, vol. 5, no. 4, pp. 263–276, 2012.
[16] M. Behl and R. Mangharam, "Evaluation of dr-advisor on the ashrae great energy predictor shootout challenge," 2014.
[17] H. Drucker, "Improving regressors using boosting techniques," in *ICML*, vol. 97, 1997, pp. 107–115.
[18] C. Romesburg, *Cluster analysis for researchers*. Lulu. com, 2004.
[19] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.
[20] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in neural information processing systems*, 1997, pp. 155–161.