2022 Building Performance Analysis Conference and
SimBuild co-organized by ASHRAE and IBPSA-USA
Chicago, IL
September 14-16, 2022

Evaluating the Performance of Chiller Plant Efficiency Using Random Forest Model: A High-Rise Building Case Study

Behzad Salimian Rizi[1,3], Mohammad Heidarinejad[1], Gregory Pavlak[2,3], Vincent Cushing[3] and William Hederman[3]
[1]Illinois Institute of Technology, Chicago, IL, 60616, USA
[2]The Pennsylvania State University, University Park, PA, USA
[3]QCoefficient Inc., Chicago, IL, USA

Abstract

Assessing electricity consumption of chilled-water cooling plants is essential for near-optimal operation and carbon emission reduction. The goal of this study is to develop an efficient chiller sequencing control strategy for different building operating conditions. To that end, this study aims to develop three Random Forest (RF) chiller models for predicting chiller power consumption and two more efficient chiller sequencing control strategies for a 1.3 million ft^2 high-rise commercial office building located in New York City. Chiller cooling load, chiller power consumption, and ambient wet bulb temperature were logged at 15-min intervals in May–September 2019, and used to train RF models for analyzing the two more efficient chiller sequencing strategies. The average value of mean absolute percentage error (MAPE) and root mean squared error (RMSE) for all three RF chiller models are 5.3% and 30 kW, respectively, for the validation dataset, which confirms a good agreement between measured and predicted values. Results of this study provide additional insights on how to accurately predict the total chiller power consumption of cooling plants under different chiller sequencing control strategies.

Introduction

Commercial Building Energy Consumption Survey (CBECS) indicates that chillers provide cooling in more than half of the commercial office building floor spaces in the U.S. (2012 CBECS Survey Data, 2015). In addition, chiller systems are responsible for providing cooling for more than half of the commercial buildings with areas greater than 9,000 m^2 (~100,000 ft^2) (Rizi and Heidarinejad 2022). To address the need of improving energy efficiency of operation of chillers, research studies investigate different chiller sequencing approaches. The recent version of ASHRAE Guideline 36 includes rules for the variable speed centrifugal chiller sequencing strategies, which includes stage up/down part load ratio (SPLR_UP,DN) and the temperature difference between condenser water return and chilled water supply temperature (lift) parameters (“ASHRAE Guideline 36” 2021). One study has evaluated the performance of three different chiller sequencing control strategies in a super tall building from aspects of chiller switch on/off number, electrical peak demand, and overall energy consumption in one week (Sun et al. 2013). Another chiller sequencing strategy utilized the hybrid predictive control to optimize sequencing of chillers using a cooling load prediction as a corrective measure to reduce the unnecessary sequencing action (Liao and Huang 2019). Several researchers deployed multivariate linear regression (MLR) and artificial neural network (ANN) as black-box modeling methods for simulation of a chiller system (Lee et al. 2012; Wei et al. 2014; Labus et al. 2013). There are big challenges to optimize chiller sequencing control strategies because of different parameters including operating parameters, chiller load, cooling tower parameters and weather conditions. However, selecting accurate modelling methods for the chiller sequencing needs to consider the importance of each variable and their contribution to the overall performance of a chiller system, as well as the ease of application and computational time. In this regard, developing empirical models was selected here because of accessibility to required data, superiority in model implementation and prediction accuracy.

The empirical modeling compared to the engineering-based approach is more practical in energy prediction of existing buildings because of accessibility to required data, such as building energy data, environmental data, and occupancy data. Also, results have proven that empirical modeling outperforms engineering-based modeling when the appropriate model is selected and the learning algorithm is chosen properly (Neto and Fiorelli 2008). However, some of the empirical modeling, e.g. Decision Tree and ANN, may introduce significant variations in the output due to small changes made in the input data due to the instability issues (Breiman 1996). As a result, the model accuracy for prediction dramatically decreases. Moreover, ANN models don't identify the relationship between the input and output variables explicitly, while the MLR models can identify the correlation between the input and output variables by regression coefficients. The MLRs models accuracy highly depends on which input variables and their multivariate terms are included. The degree of input variables may also affect the accuracy. Therefore, examining the optimum order and combination of input variables to predict the outputs with the highest accuracy needs some effort. To overcome the instability of the learning algorithm as well as to improve prediction accuracy and identify variable importance a more advanced data mining technique called ensemble learning was introduced in the early 1990s (Hansen and Salamon 1990).

Random Forest (RF) models, as an ensemble learning model partition input variables utilizing decision tree methods to identify importance of variables and to predict output variables with high accuracy. This model also works well when handling outliers and noise in the dataset (Liaw and Wiener 2002). One study showed that RF models could outperform classical linear regression in predicting heating and cooling loads of residential buildings (Tsanas and Xifara 2012). Also, RF, Support Vector Regression (SVR), and Regression Tree (RT) models have been applied for predicting hourly building energy. The comparison of the results confirmed the superiority and feasibility of homogeneous ensemble learning of RF models in building energy prediction (Wang et al. 2018). Moreover, some researchers studied how the RF model can identify important variables. This model was used to identify influential features on the regional energy use intensity (EUI) of residential buildings (Ma and Cheng 2016). Also, the RF model could analyze the importance of operating variables and predict the Coefficient of Performance (COP) of an air-cooled chiller with high accuracy (Yu et al. 2017).

Although many attempts have been made to develop chiller model and chiller sequencing strategy in chiller plants, developing high-accurate chiller model when limited data is available is still not well discussed. Also, subject to real world measurement's uncertainty and noise, it is not always known what features will lead to the best prediction model. Therefore, selecting high accurate modelling methods that identify the importance of each input variable and its contribution to the overall performance of a chiller is essential.

This study aims to (i) develop data-driven chiller models, (ii) investigate variables that are importance to chiller electricity consumption predictions, and (iii) predict the power consumption of the chillers for two more efficient chiller sequencing strategies of the cooling plant of a high-rise commercial building located in New York City. The experimental set up of the chillers, RF model development for each chiller, chiller plant actual data, and the results of feature importance technique for each model are discussed in this paper.

Method of study

Description of the chillers

The central water-cooled chiller plant is comprised of four centrifugal electric variable speed chillers with rated capacity of 5,627 kW (1,600 tons) for chillers 1 and 2, and 3,517 kW (1,000 tons) for chillers 3 and 4 to provide chilled water for two air handling units. Chiller 3 is a magnetic-bearing chiller and the building facility managers usually operate it only during mild Spring and Fall days; thus, it has not been considered in this analysis since this study focuses on the warmer June-September weeks of the 2019 cooling season. The logged data during the 2,820 hours of those weeks indicates that chillers 1, 2 and 4 were working individually 27%, 23% and 18% of total working hours, respectively. Also, when working in combination, chillers 2 and 4 and chillers 1 and 4 were the most frequent contribution at 10.5% and 4.5%, respectively.

Random Forest (RF) and model development

RF uses random sets of input variables to build an ensemble of decision trees and is one of the important methods for classification and regression problems. Many studies confirmed that this model outperforms the single decision tree models in terms of overcoming issues of overfitting and model accuracy by aggregating the predictions made by many individual decision trees (Breiman 2001).

The Random Forests can be used to rank the importance of variables in a regression or classification problem. The first step in measuring the variable importance in a data set $D_n = \{(X_i, Y_i)\}_{i=1}^{n}$ is to fit a random forest to the data. During the fitting process the out-of-bag error for each data point is recorded and averaged over the forest (errors on an independent test set can be substituted if bagging is not used during training). To measure the importance of the j-th feature after training, the values of the j-th feature are permuted among the training data and the out-of-bag error is again computed on this perturbed data set. The importance score for the j-th feature is computed by averaging the difference in out-of-bag error before and after the permutation over all trees. The score is normalized by the standard deviation of these differences. Features which produce large values for this score are ranked as more important than features which produce small values. The built-in feature_importances_ attribute in the RF package ranks the importance of the input variables on the predicted variable.

This study used Jupyter Notebook (version 5.7.4) in the Anaconda Navigator platform for developing the RF model. The Sklearn Random Forest Regressor package was called to execute the algorithm. Measured data for each chiller were split into the training and validation datasets by randomly sampling 75% of the dataset for training the model and 25% for testing the model. All the data was standardized between 0 to 1 since features with a wider range can affect the stability of the model during the training process. Therefore, data standardization is a significant step for many machine-learning estimators (Mousavi et al. 2018).

The RF model accuracy was measured by considering two frequently-used prediction accuracy evaluation indices which are: the Root Mean Squared Error (RSME), and the Mean Average Percentage Error (MAPE) given by Equations (1) and (2).

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{t=1}^{n} (x_t - y_t)^2} \tag{1}$$

$$MAPE = \frac{1}{n} \times \sum_{t=1}^{n} \left|\frac{x_t - y_t}{y_t}\right| \times 100\% \tag{2}$$

where n, x_t and y_t represent the sample size, the predicted value and the actual value, respectively. RMSE represents the sample standard deviation of the residuals between predicted and measured data. MAPE is a statistical indicator that explains the accuracy of the prediction by comparing the residual with the measured data and mostly expressed in percentage (Wang et al. 2018). Development and comparison of new staging strategies

This study conducted five steps to compare the existing operating strategy used by the building operators to two new and more efficient chiller sequencing strategies suggested by the RF model as follow:

- **Step 1:** Develop two new chiller sequencing strategies for the chiller plant based on the Chillers 1, 2 and 4 and combinations of them.
- **Step 2:** Apply chiller models for the existing strategy as well as the new strategies 1 and 2 at each time stamp.
- **Step 3:** Predict the power consumption of chillers sequencing strategies based on the chiller models at each time stamp.
- **Step 4:** Sum the chiller power consumptions calculated at each timestamp to determine the total chillers power consumption for all the strategies at each period.
- **Step 5:** Rank the new chiller sequencing strategies by calculating the chiller power consumption reduction compared to the existing sequencing strategy.

Results

The measured time series data has been collected during summer 2019 from May 1st to mid-September at a 15-min interval, totaling 15,286 recorded data. The number of measured data were 5,684 for chiller 1, 6,156 for chiller 2 and 5,337 for chiller 4. Three RF chiller models were developed for chiller 1 (model A), chiller 2 (model B) and chiller 4 (model C). This study considered normalized chiller cooling load, normalized wet bulb temperature, day of week, and time of a day as input parameters and chiller power consumption as an output parameter. The importance of input variables on the prediction of chiller power consumption was ranked based on the built-in feature_importances_attribute in the RF package. Results show that for models A, B and C the chiller cooling load and wet bulb temperature are the most important features for determining chiller power consumption. The chiller cooling load has an importance of 67%, 74% and 66% and wet bulb temperature has an importance of 31%, 24% and 31% for models A, B and C, respectively. The importance of other features compared to these features was negligible.

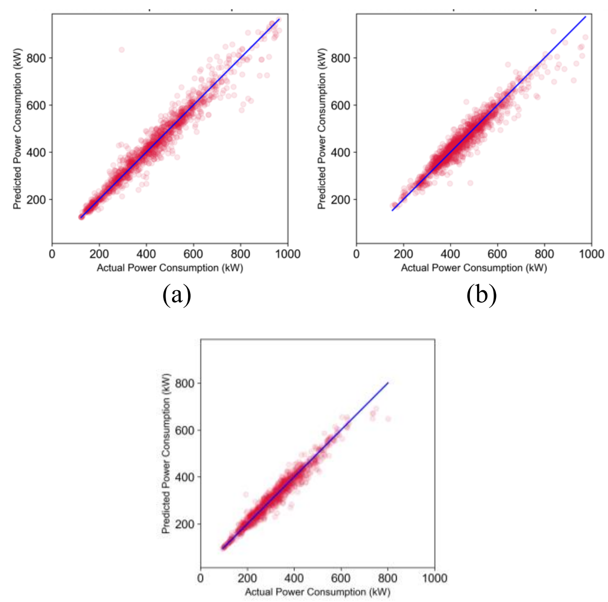
The chiller models A, B and C were applied to the validation dataset to compare the result of predicted data with the measured data. Figure 1(a), Figure 1(b) and Figure 1(c) have depicted the predicted power consumption vs. actual power consumption of the validation dataset using model A, B, and C, respectively.

The blue line represents a perfect regression. The closer the points are to the line, the more accurate the model is. The blue line can also be used to understand if the model is under or over predicting. If points are above the blue line, the model is over predicting; if points are below the blue line, the model is under predicting. Also, to measure the accuracy of the model, the RMSE and MAPE for each model is presented in Table 1. The RMSE for model A, model B and model C are 38 kW, 31.2 kW and 21.9 kW, respectively. Also, the MAPE for model A, model B and model C would be 5.9%, 4.9% and 5.2%. The number of decision trees for the models are 100.

Table 1. Chiller model specification

Chiller model	Number of decision trees	Number of data		Model Evaluation Indices	
		Training	Testing	RMSE (kW)	MAPE (%)
Model A (Chiller 1)	100	4263	1421	38.7	6.1
Model B (Chiller 2)	100	4617	1539	31.0	4.8
Model C (Chiller 4)	100	4003	1334	21.8	5.1

The accuracy of the chiller models would normally decrease when the chiller model is examined by the unseen data due to the overfitting issue. In this study, to evaluate the level of overfitting, and provide an unbiased sense of model effectiveness, two weekday periods of the 2019 cooling season (07/15/2019 to 07/20/2019 and 07/22/2019 to 07/27/2019) in New York City with average wet bulb temperature of 73°F and 68°F were selected as an unseen dataset which were not used to build or tune the models.Figure *2* depicts the wet bulb temperature variation in New York City during cooling season 2019. The studied area is limited by the two-orange dash line.



(a) (b) (c)

Figure 1. Plots of the predicted chiller power consumption against the actual power consumption of validation dataset for (a) model A (b)model B and (c) model C

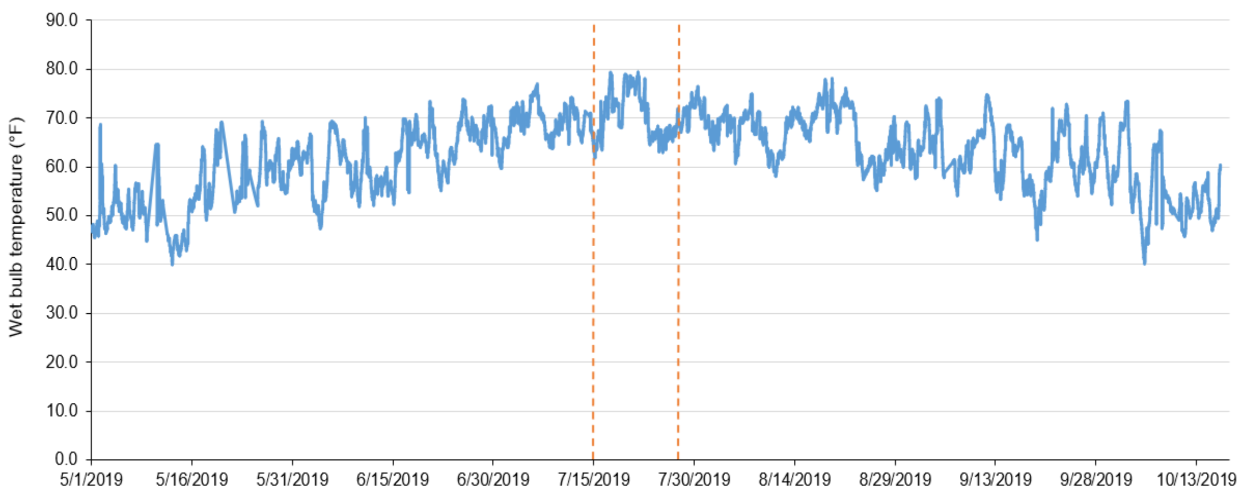
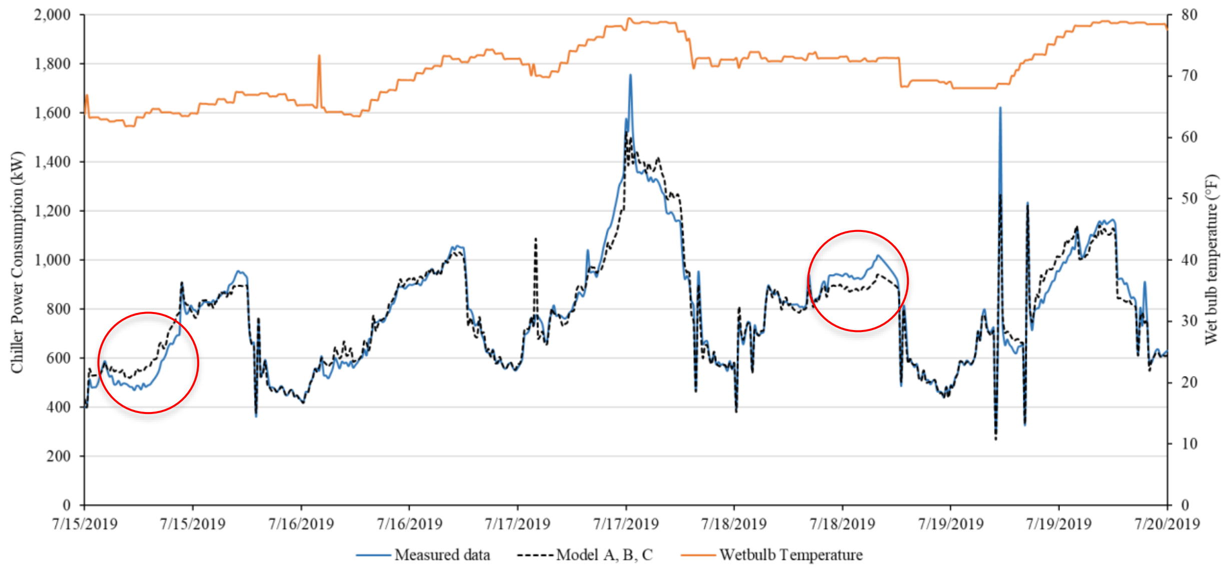
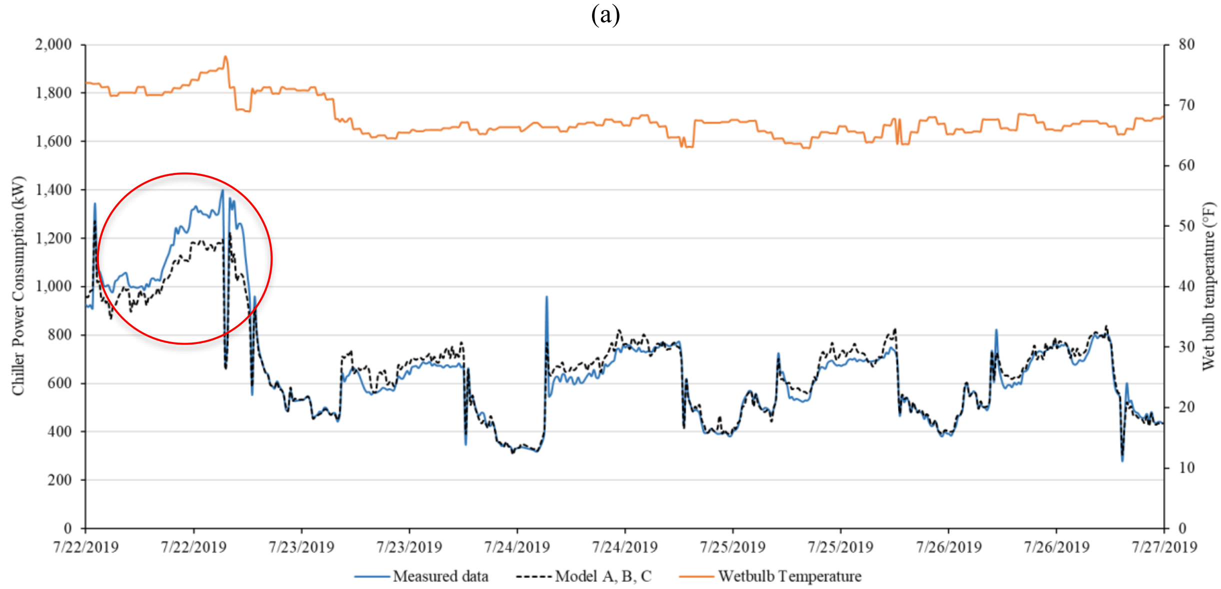


Figure 2. The wet bulb temperature variation in New York City during the cooling season 2019



(a)



(b)

Figure 3. Total chiller power consumption (kW) during weekday period (a) Period 1, and (b) Period 2

The models A, B and C were applied on the measured data for periods 1and 2 as an unseen dataset. The RMSE and MAPE values for these two periods were calculated using Equations 1 and 2 based on the measured and predicted chiller power consumption at each timestamp. The RMSE for periods 1 and 2 are 55 kW and 58 kW, respectively. Also, MAPE values for periods 1 and 2 are 4.8% and 4.9%, respectively. Therefore, by considering the min, max and average values of measured and predicted chiller power consumption as indicated in Table 3 as well as calculated RMSE and MAPE, there are a good agreement between measured and predicted data in these two periods. Figure 3(a) and Figure 3(b) compare the measured time-series total chiller power consumption with predicted one for the periods 1 and 2 respectively. In these periods as shown in Figure 4, from 07/15/2019 to 07/20/2019 ("Period 1"), chillers 1 and 2 are working 22 and 14 hours individually. In combination mode, chillers 1 and 2, 1 and 4, and 2 and 4 are working about 20, 49 and 14 hours, respectively.

Also, from 07/22/2019 to 07/27/2019 ("Period 2"), chillers 2 and 4 are working 40 and 5 hours individually. Chillers 1 and 2 and chillers 2 and 4 are working about 16 and 44 hours, respectively. in combination. According to the five steps described in the previous section, the results of two developed chiller control strategies were explained as follows.

Step 1. As shown in Figure 4, two new and more efficient chiller sequencing strategies were developed based on a 2019 cooling season graphical analyses of electric input ratio (EIR) vs chiller cooling load for each of the three chillers. The Figure *4*shows the different operating frequencies associated with each chiller scheme mode. Importantly, the graphical analysis identified efficient and inefficient areas of each chiller's historical operation. A 10% dead-band was added to remove additional cost related to the excessive switching off/on of the chillers. Strategy 2 simply lowers the efficient chiller operating ranges adopted in Strategy 1. Note also that Strategies 1 and 2 do not include chiller scheme modes that start with the smaller chiller 4.

Chiller EIR was calculated by dividing chiller power consumption divided by provided chiller cooling load (Ton) at each time stamp and was considered as an index for the chiller efficiency. Table 2 summarizes the thresholds used in the current chiller sequencing strategy, as well as for sequencing strategies 1 and 2.

Step 2. According to the chiller scheme mode at each timestamp, models A, B and C (separately or in combination) were applied.

Step 3. The normalized wet bulb temperature and normalized chiller cooling load at each timestamp were considered as input variables for models A, B and C. The applied model(s) predict the chiller power consumption as an output parameter.

Step 4. To compare the power consumption of chillers in different strategies, the summation of power consumption was calculated over the periods 1 and 2 for each strategy. Table 3 presents the predicted sum, average, min and max value of chiller electricity consumption for each strategy at each period.

Table 2. Threshold of each chiller sequencing control strategy

Chiller sequencing control strategy	Chiller cooling load (ton)	Chiller scheme mode
Current strategy	[400-1100]	Chiller 4
	[400-1400]	Chiller 1/2
	[1300-2300]	Chiller 1/2&4, 4&1/2
	[1700-2700]	Chiller 1&2
Strategy 1	[400-900]	Chiller 4
	[800-1400]	Chiller 1/2
	[1250-1600]	Chiller 1/2&4
	[1450-2700]	Chiller 1&2
Strategy 2	[400-900]	Chiller 4
	[800-1250]	Chiller 1/2
	[1100-1900]	Chiller 1/2&4
	[1700-2700]	Chiller 1&2

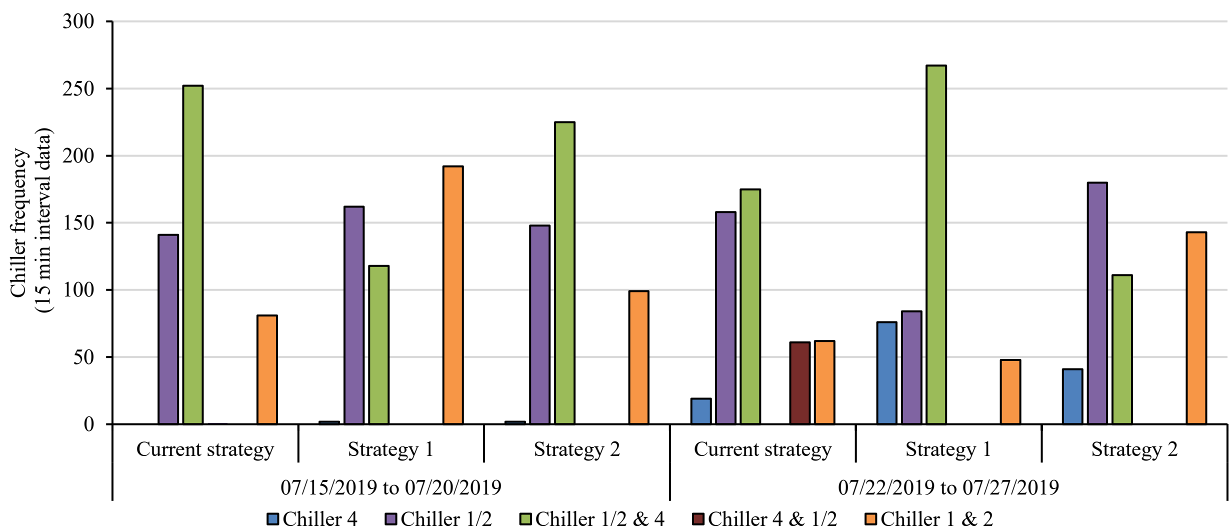
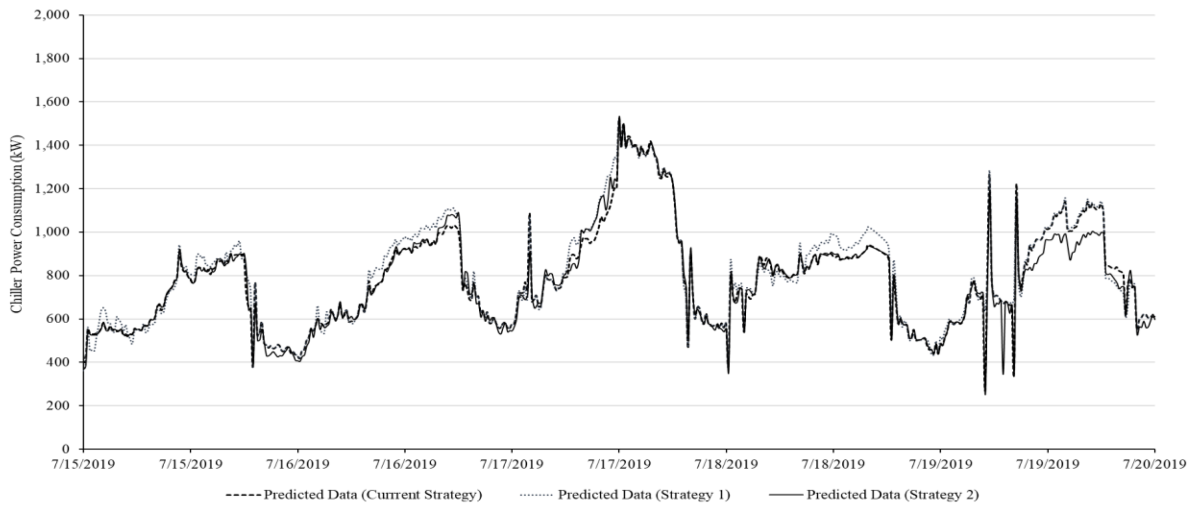


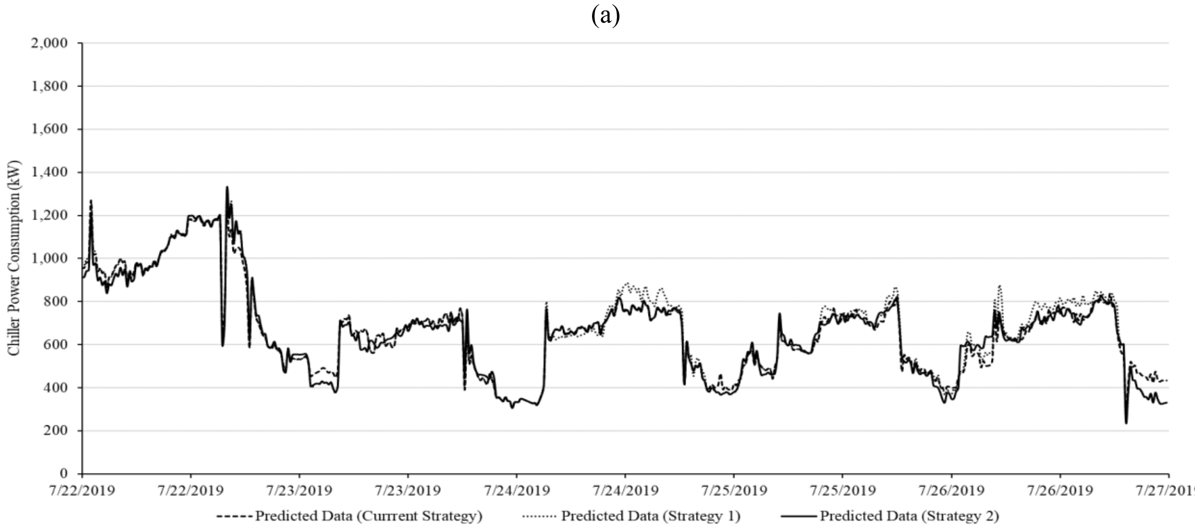
Figure 4. Chiller sequencing strategies specification during Period 1 and Period 2

Table 3. The predicted results of different chiller sequencing strategies for Period 1and Period 2

Time period	Chiller power consumption (kW)	Measured data	Predicted data			Power consumption reduction(kW)
			Current strategy	Alternative 1	Alternative 2	
Period 1	Sum	369,771	369,989	378,020	365,510	(4,479)
	Average	782	784	799	772	
	max	1,757	1,519	1,521	1,529	
	min	295	269	361	250	
Period 2	Sum	319,164	318,689	322,158	316,636	(2,053)
	Average	672	670	678	666	
	max	1,396	1,270	1,316	1,319	
	min	280	305	245	241	



(a)



(b)

Figure 5. The comparison of predicted data for the strategies 1 and 2 with the current strategy for (a)Period 1 and (b)Period 2

Step 5. To evaluate the performance of each strategy, the total chiller power consumption of strategies 1 & 2 during periods 1 and 2 was compared with the actual measured chiller power consumption in terms of power consumption reduction percentage as indicated in Equation 3. The results were summarized in Table 3.

$$Power\ consumption\ reduction\ (\%) = \frac{Strategy\ (i) - Current\ strategy}{Strategy\ (i)} \times 100 \quad for\ i = 1, 2, \dots, 5 \tag{3}$$

where Strategy (i) denotes the summation of chiller power consumption for ith strategy at the specific period.

Figure 5 (a) and Figure 5(b) visualizes these comparisons between strategies 1, 2 and current strategy as a time series for periods 1 and 2, respectively.

Discussion

The conventional chiller sequencing control strategies such as return chilled water temperature-based control and cooling load-based control modeled for high rise commercial buildings require detailed information from the old existing chiller plants which mostly are not accessible. To address these limitations this study introduces a data driven modeling approach as a solution for developing three random forest chiller models, model A, model B and model C to calculate the chillers power consumption. As shown in

Figure 1, the results of predicted chiller power consumption of the validation dataset agree well with the measured chiller power consumption. Also the perfect blue regression lines in Figure 1(a), Figure 1(b), and Figure 1(c) demonstrate how well the power consumption data has been spread around the blue line. Figure 3(a) and Figure 3(b) show how the chillers models A, B and C perform well on the unseen data during period 1 and 2. In some areas of the plots, the deviation between the predicted data and measured data are higher which was indicated by the red circle. These areas represent the state when chillers are in combination mode to satisfy the building cooling load. Therefore, this study entails some limitations in terms of chiller power consumption prediction when chillers are working together. In the future work, we aim to address these limitations by considering the chiller mode operation as an input parameter. Also, the time series wetbulb temperature in periods 1 and 2 follows the trend of timeseries chiller power consumption. The feature importance technique also confirms this fact that the wet bulb temperature with average contribution of 29% is one of the important factors for determining the chiller power consumption.

Figure 5 (a) and Figure 5(b) clearly show that strategy 2 has placed below the current strategy and strategy 1 most of the time during periods 1 and 2, while strategy 1 compares to the current strategy has located above the current strategy. The amount of power consumption reduction compared to current strategy calculated from Equation (3) demonstrates that strategies 2 by reduction of 4,479 kW (1.2%) and 2,053 kW(0.7%) for periods 1 and 2 ,respectively can be introduced as a potential strategy to the chiller plant.

Conclusion

This study considers the Random Forest (RF) method to predict power consumption of water-cooled variable speed centrifugal chillers under two different chiller sequencing strategies in a high-rise commercial office building with building area of about 1,300,000 ft^2 located in New York City.

Chiller cooling load, chiller power consumption, and wet bulb temperature were logged at a 15-min interval from May 2019 to Mid-September 2019. Three RF chiller models are developed for Chiller 1(model A), Chiller 2 (model B) and Chiller 4 (model C) based on collected data. The input parameters for the models are normalized chiller cooling load, normalized wet bulb temperature, day of week, time of a day as input parameters and the output parameter is normalized chiller electricity consumption. This study used feature importance technique to identify the most influential parameters on the chiller power consumption. Among the input features, the chiller cooling load and wet bulb temperature with importance of about 70% and 30% for models A, B and C, respectively, considered as rank 1 and 2 for predicting chiller power consumption. The Root Mean Squared Error (RMSE) and Mean Average Percentage Error (MAPE) for each model suggest that developed models are predicting well.

To evaluate the level of overfitting of both models, the recorded data for two weekdays' periods of cooling season 2019 (07/15/2019 to 07/20/2019, and 07/22/2019 to 07/27/2019) in New York City was introduced to the models as an unseen dataset. The RMSE and MAPE were calculated during these two periods and the results confirmed that three models are predicted with a_high confidentiality. These models were applied to compare two feasible chiller sequencing strategies during these periods with incorporation of Chillers 1, 2 and 4 and combination of them based on predicted chiller power consumption at each time stamp (15-min interval). The results indicate that chiller sequencing control strategy 2 by reduction of 1.2% and 0.7% of total chiller's power consumption compared to the current strategy in period 1 and 2, respectively can be considered as a potential strategy to the chiller plant. To make confidence of reporting smaller savings estimates for different chiller sequencing strategies using RF models it needs to provide reasonable estimates of prediction uncertainty bands. Also, applying some feature engineering techniques for training the chiller models may improve the chiller's model performance. This will constitute the future research work of the authors.

References

"2012 CBECS Survey Data." 2015. U.S. Energy Information Administration. https://www.eia.gov/consumption/commercial/data/2012/index.php?view=microdata.

"ASHRAE Guideline 36." 2021. ASHRAE.

Breiman, Leo. 1996. "Heuristics of Instability and Stabilization in Model Selection." *The Annals of Statistics* 24 (6). https://doi.org/10.1214/aos/1032181158.

Breiman. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Hansen, L.K., and P. Salamon. 1990. "Neural Network Ensembles." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (10): 993–1001. https://doi.org/10.1109/34.58871.

Lee, Tzong-Shing, Ke-Yang Liao, and Wan-Chen Lu. 2012. "Evaluation of the Suitability of Empirically-Based Models for Predicting Energy Performance of Centrifugal Water Chillers with Variable Chilled Water Flow." *Applied Energy* 93 (May): 583–95. https://doi.org/10.1016/j.apenergy.2011.12.001.

Liao, Yundan, and Gongsheng Huang. 2019. "A Hybrid Predictive Sequencing Control for Multi-Chiller Plant with Considerations of Indoor Environment Control, Energy Conservation and Economical Operation Cost." *Sustainable Cities and Society* 49 (August): 101616. https://doi.org/10.1016/j.scs.2019.101616.

Liaw, Andy, and Matthew Wiener. 2002. "Classification and Regression by RandomForest."

Ma, Jun, and Jack C.P. Cheng. 2016. "Identifying the Influential Features on the Regional Energy Use Intensity of Residential Buildings Based on Random Forests." *Applied Energy* 183 (December): 193–201. https://doi.org/10.1016/j.apenergy.2016.08.096.

Mousavi, Seyed Sajad, Michael Schukat, and Enda Howley. 2018. "Deep Reinforcement Learning: An Overview." In *Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016*, edited by Yaxin Bi, Supriya Kapoor, and Rahul Bhatia, 16:426–40. Lecture Notes in Networks and Systems. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-56991-8_32.

Neto, Alberto Hernandez, and Flávio Augusto Sanzovo Fiorelli. 2008. "Comparison between Detailed Model Simulation and Artificial Neural Network for Forecasting Building Energy Consumption." *Energy and Buildings* 40 (12): 2169–76. https://doi.org/10.1016/j.enbuild.2008.06.013.

Rizi, Behzad Salimian, and Mohammad Heidarinejad. 2022. "Analysis of Hydronic Heating and Cooling Systems in Commercial Buildings Using CBECS Microdata." *Journal of Architectural Engineering* 28 (3): 04022016. https://doi.org/10.1061/(ASCE)AE.1943-5568.0000543.

Sun, Yongjun, Shengwei Wang, and Fu Xiao. 2013. "In Situ Performance Comparison and Evaluation of Three Chiller Sequencing Control Strategies in a Super High-Rise Building." *Energy and Buildings* 61 (June): 333–43. https://doi.org/10.1016/j.enbuild.2013.02.043.

Tsanas, Athanasios, and Angeliki Xifara. 2012. "Accurate Quantitative Estimation of Energy Performance of Residential Buildings Using Statistical Machine Learning Tools." *Energy and Buildings* 49 (June): 560–67. https://doi.org/10.1016/j.enbuild.2012.03.003.

Wang, Zeyu, Yueren Wang, Ruochen Zeng, Ravi S. Srinivasan, and Sherry Ahrentzen. 2018. "Random Forest Based Hourly Building Energy Prediction." *Energy and Buildings* 171 (July): 11–25. https://doi.org/10.1016/j.enbuild.2018.04.008.

Yu, F.W., W.T. Ho, K.T. Chan, and R.K.Y. Sit. 2017. "Critique of Operating Variables Importance on Chiller Energy Performance Using Random Forest." *Energy and Buildings* 139 (March): 653–64. https://doi.org/10.1016/j.enbuild.2017.01.063.