Received 14 November 2022, accepted 29 November 2022, date of publication 1 December 2022, date of current version 6 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3225976

RESEARCH ARTICLE

Simulating Floating Head Pressure Control With Artificial Intelligence

ANDREA LOREGGIA[1], FRANCESCO FACCIO[2], DIEGO MUSSOLIN[2], AND MAURIZIO ZAMBONIN[2]
[1]Department of Information Engineering, University of Brescia, 25121 Brescia, Italy
[2]MTA SpA, 35026 Conselve PD, Italy

Corresponding authors: Andrea Loreggia (andrea.loreggia@unibs.it) and Diego Mussolin (dmussolin@mta-it.com)

ABSTRACT In this work, we model with artificial intelligence techniques one of the most used solutions for energy saving in the field of refrigeration. This solution, called floating head pressure control, allows for the optimum pressure condensation depending on the working environment, thus increasing the overall efficiency of the plant. Usually, these mechanisms are controlled by algorithms that are stored in the control memory of the chillers. Under certain environmental conditions, the optimal temperature and condensation pressure fluctuate with the environmental temperature. Thus, the absorbed electricity depends on the working parameters fixed by the control system of the chiller. Results in energy savings in these terms could grow as the external temperature decreases. The problem addressed in this work regards standard floating head pressure control (FHPC) systems that do not combine working environmental factors with the conditions of the machine and historical data because they are implemented adopting static models. We provide evidence about the potential of machine learning models in predicting the velocity of fans in order to adjust their load based on environmental factors and the conditions of the machine. A good setting of fan velocity will result in lower energy consumption. To do that we analyzed and implemented machine learning algorithms to provide instruments that support the operation of the chiller, enabling the floating head pressure control mode in the control system of the machine, during its process. We performed an empirical evaluation on both synthetic and real data to assess the quality of our proposal. Synthetic data are produced by an industrial software that simulates the behavior of chillers, while real-world data is collected from a commercial chiller. The results show that machine learning approaches are able to approximate real data getting errors that are 5 times smaller than the errors committed by the system that is now adopted. All the source code as well as the datasets is available online at http://ieee-dataport.org/10179.

INDEX TERMS Refrigeration, air-conditioning systems, artificial intelligence, energy saving, floating HP, data simulation.

I. INTRODUCTION

Refrigeration and air conditioning account for 15% of global consumption of electricity, which corresponds to 4.5% of global greenhouse gas emissions (GHG) [11]. Cooling systems represent a considerable part of energy consumption in the food industry, warehouses, or large and medium shops (from 30% to 80%). The International Energy Agency (IEA) reports the year 2021 to be one of the warmest years on record. This led to a tremendous space cooling demand, experiencing the highest annual growth among all buildings end uses in 2021 and accounted for nearly 16% of buildings sector final electricity consumption (about 2 000 TWh).[1]

Without any measure to limit this enormous demand for resources, the use of cooling energy could more than double in the next twenty years, while we should limit cooling energy growth between now and 2040 to 19% [11].

Existing solutions for the development of refrigeration systems and air conditioning often privilege the production to the consumption reduction. In this direction, the refrigeration

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

[1]https://www.iea.org/fuels-and-technologies/cooling - October 14, 2022

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

industry is searching for alternatives and new methods to improve sustainability, for instance by replacing the refrigerant used by the plants [14]. Other solutions are focused on the containment of consumption, such as the floating head pressure control (FHPC or floating HP control) technique. Floating HP control is a solution that allows optimizing the operation at any working point, ensuring increased efficiency at partial loads, through the continuous modulation of the fans of an air-condensed chiller. Not all refrigeration systems are equipped with it. Some installations with floating HP control get poor results and save little energy because this solution must be implemented properly in order to maximize energy savings without causing technical problems. Moreover, floating HP control maximizes its benefits based on some environmental conditions (such as the temperature of the environment).

FHPC and other energy-saving techniques are also very interesting in the context of strategies implemented by public authorities and governments. For instance, in the French energy transition, the government designed and developed a set of normative tools that are encouraging more sustainable energy consumption practices, enabling at the same time, the implementation of the European Directive 2012/27/EU[2] for energy efficiency. In this context, the *"Programme fixant les Orientations de la Politique Energétique"*[3] (POPE law), enacted in 2005, created an energy saving certification (Certificats d'Économie d'Énergie - CEE). This is an innovative and ambitious mechanism in terms of the environmental taxation. This system has met its main expansion within the major French energy producers and certification companies. In the context of energy savings with FHPC solution, the POPE law allows companies to benefit from energy-saving vouchers or funds when they certify their products with respect to the standard IND-UT-16.[4]

In developing methods able to correctly drive FHPC solutions, environmental factors have to be taken into account, because they influence the performance of the system. The problem addressed in this work regards standard FHPC systems that do not consider working environmental factors, the conditions of the machine, and historical data to manage the plant. This is due to the fact that these systems are usually implemented with static models. In this regard, embedding machine learning techniques into the control system of the chiller would be beneficial to control the behavior of the plant based on the conditions of the ambient and on historical data. The aim of this study is to provide evidence of the accuracy of machine learning techniques in approximating the floating head pressure process.

We show through empirical analysis that the performance of models based on artificial intelligence (AI) is better than the one of existing frameworks. To do that, we compared the prediction of several AI techniques to the prediction of an industrial simulator, and then with real values generated by a commercial chiller. This is to the best of our knowledge the first study that shows the difference between these approaches and that made the data available for future works. This possibly opens new research lines that can be adopted in the industry and that may increase sustainability. All the resources employed in this work (i.e., source code and datasets) are publicly available online at http://ieee-dataport.org/10179.

The rest of the paper is structured as follows: Section II reports related works on the topic analyzed in this work. Section III introduces some basic notions about the refrigeration process, the floating head pressure control, and the adopted machine learning models. Section IV reports the empirical evaluation of the models and the analysis of the collected data. Section V provides a discussion about the results and the proposed method. Section VI concludes our work also reporting some future directions of this research.

II. RELATED WORK

Although air-cooled chillers have been adopted for many years, very few studies have been done to understand how to model and control the condensing temperature and implement these models in industrial production. Some of these studies are focused on designing new refrigerants to be employed in the equipment [14], while others instead carried out simulations with an ammonia compression system [13] where the maximum coefficient of performance (COP) can be reached by adjusting the condenser system based on the wet bulb temperature. Experimental studies showed that COP increases along with the air velocity and the water spray rate of the condenser [12], [19]. Other scholars are focused on the characteristics of the condenser. For instance, [22] showed that pre-cooling the condenser air can influence the COP, while the evaporative air condenser improves the COP even with an increment of the ambient temperature when the compressor power follows the variation of the wet bulb temperature [10].

Ambient temperature and other factors may influence the point of condensing temperature, thus principal ways to adjust this point based on the chiller model and the load ratio can significantly improve the efficiency of the system by achieving maximum COP. Different strategies can be employed to control the compressors and the velocity of the fans in order to reduce electric energy consumption [1], [6], [16], [25]. A different perspective comes from the adoption of bottom-up approaches, they are model driven by the data logged during the operational time of the chiller. This is already supported by some simulation studies (see for instance, [23], [24], [25]), in these studies authors applied simple regression models for the optimization of the whole system without adopting and investigating the performance of more complex machine learning approaches. The performance of a chiller varies accordingly with the ambient, thus different devices from the same series can perform

[2] http://data.europa.eu/eli/dir/2012/27/oj - Last visited October 7 2022

[3] https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000813253/ - Last visited October 7 2022

[4] https://www.legifrance.gouv.fr/loda/article_lc/LEGIARTI000029310643/ - Last visited October 15 2022

completely different because in different scenarios. Tailoring the working behavior to ambient factors can positively affect the performance of the system.

Data-driven models in place or in combination with model-based systems are now considered a necessity of paramount importance to capture the real operating characteristics of chillers [20], [21]: the changes in the operating variables can consistently influence the whole system and possibly reach optimal control. Data-driven systems will define models with better accuracy and with the ability to adjust their performance and operation with the operating factors. For instance, [2], [20] implemented fault detection and diagnosis techniques for chiller systems based on bottom-up approaches.

III. BACKGROUND

In this section, we introduce some basic notions related to the techniques used for refrigeration, the floating head pressure control as well as the machine learning techniques implemented in the empirical evaluation.

A. REFRIGERATION

Refrigeration means transferring heat from an environment that must be kept at a low temperature to another where it can be easily disposed of. Heat spontaneously goes from a body with high temperatures to a body with a lower temperature. But to make the opposite, it is necessary to develop appropriate devices. In this work, we focus on technologies for the development of refrigeration that are based on saturated vapor compression. These systems are based on the reverse steam compression cycle which is one of the cycles used in heat transfer machines from low-temperature systems to high-temperature systems. The cycle is based on the fact that the state transition liquid-vapor can take place in one direction or another with the change of pressure and with the absorption or the release of heat. The key point is adopting a fluid that has the ability to remove heat from the environment to be refrigerated by evaporating at low pressure and temperature while transferring the heat to the external environment and condensing at higher pressure and temperature.

The vapor-compression cycle is represented in Figure 1, the process is carried out by a refrigerant fluid in a circuit refrigerator [17]. The refrigerant flows through an evaporator (step 4 to 1) where the refrigerant comes into contact with the fluid to be cooled (usually air or water), which is at a slightly higher temperature than the refrigerant, then heat is transferred from it to the refrigerant, producing the cooling effect. The refrigerant boils because of the heat it receives and when it leaves the evaporator it is completely vaporized at a low temperature and low pressure. To enable the refrigerating effect continuously, the refrigerant must be brought back to the conditions of a liquid at high pressure. Thus, the fluid flows into a compressor (step 1 to 2) that increases the pressure of the refrigerant gas which also results in increasing its temperature. The refrigerant leaves the compressor as a gas at high temperature and pressure. The refrigerant flows into a condenser (step 2 to 3) where it exchanges heat with another fluid (air or water) that is at a temperature lower than the refrigerant. Heat, therefore, transfers from the refrigerant to the cooling fluid, and as a result, the refrigerant condenses to a liquid.

B. FLOATING HEAD PRESSURE CONTROL

Floating HP control is given by the balance between the power to be evacuated by the entire system and the power that can be evacuated by the condenser. The greater the temperature difference between the air of the environment (e.g., room temperature) and the coolant temperature (condensation temperature), the higher the ability to evacuate. In a low ambient temperature, floating HP control allows for a decrease in the temperature, and consequently the condensation pressure, giving the possibility for the compressor to reduce energy consumption. Increasing the floating HP control efficiency consists of a synthesis of regulating the condensation pressure to a value that allows obtaining the least consumption of couples compressor/condenser.

In practice, FHPC has the following effects:

- increases the cooling capacity of the system;
- reduces of the electrical power consumption;
- increases of the energy efficiency ratio (EER).

These performance gains are significant, particularly during periods of the year when the outside temperature is lower. It makes it possible to generate energy savings from 15% to 25%. Floating HP control is integrated into the control system of the chiller, usually, it is developed through a series of mathematical algorithms that allow the calculation of the optimum condensation pressure depending on the ambient conditions. With the environmental conditions, the temperature and the pressure of condensation fluctuate with the external temperature, resulting in an energy saving that grows with the decrease of external temperature. The purpose of this work is therefore to model algorithms, based on machine learning models, that are able to provide more accurate regulation of head condensation pressure than that given by standard approaches. These are generally of careful experimental analysis made by manufacturers of chillers.

C. MACHINE LEARNING

Machine learning is a branch of Artificial Intelligence (AI) that designs and develops algorithms that learn intrinsic characteristics and possible relationships in the data. Machine learning techniques are usually classified into three main categories. These categories differ in the way information is extracted and used: supervised learning, unsupervised learning, and reinforcement learning [15]. For the purpose of this work, we will focus on the first category, namely supervised learning, which comprehends all those techniques that learn to make a prediction based on the information inferred from a collection of data (called dataset), which describes the studied scenario. A dataset is a matrix with shape $M \times N$, where each column corresponds to a variable (also called "feature") that describes a specific characteristic of the domain, and each row corresponds to a sample. A sample c_i is a tuple $(x_i; y_i)$

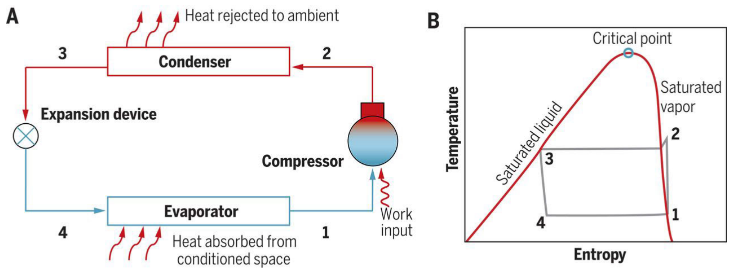

FIGURE 1. Vapor-compression cycle [14].

such that $x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,N})$ is a feature vector that stores a real value for each feature in the domain and y_i is the "label" of the sample (i.e., the real value that we want to predict). The supervised approach assumes that there exists an ideal function $f : X \rightarrow Y$ such that $f(x_i) = y_i$, where X is the space of all possible samples and Y is the set of all possible outputs. Supervised learning tries to find a function $\bar{f} : X \rightarrow Y$ that approximates f as closely as possible, finding the same labels as f for most of the samples. If Y is a finite set of distinct and countable values, then the supervised task is a *classification task*, otherwise when Y is a range of continuous real values the task is said to be a *regression task*. There exist many models designed to tackle with the aforementioned tasks, in this work we adopt the following models: Decision Tree, Random Forest, XGBoost, and AdaBoost.

1) DECISION TREE

Decision tree defines regression or classification models in the form of a tree structure [5]. The training set is splitted into smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision tree is built top-down, starting from a root node. During the training phase, the algorithm partitions the data into subsets that contain instances with homogenous values. Different metrics can be adopted to calculate the homogeneity of a set of sample (e.g., for numerical samples the standard deviation can be adopted). At each iteration, the attribute with the largest homogeneity is chosen to be a decision node. The dataset is splitted based on the values of the selected attribute. This process is run recursively until all data is processed.

2) RANDOM FOREST

The Random Forest algorithm [4] is based on decision trees and applies the bagging technique. This consists in training multiple decision trees on distinct partitions of the dataset by sub-sampling it with re-insertion. The prediction of the whole system is the result of the aggregation of the single predictions from each decision tree in the ensemble. This can be done for instance by using the mean function. Random Forest typically has better generalization performance than a single decision tree, thanks to randomness which helps to contain the overfitting problem, reducing model variance.

3) AdaBoost

Adaptive Boosting (AdaBoost) algorithm is used to define regression or classification models [18]. It is a boosting technique used as an ensemble method in machine learning. The algorithm during the training phase focuses the attention of the models on those instances that are incorrectly classified. The technique adopts weights that are re-assigned to each instance, and in particular higher weights to wrong classified samples.

4) XGBoost

XGBoost is based on Friedman's original Gradient Boosting [8], [9]. It introduces a regularization term to control overfitting, obtaining better performance results. The Gradient Boosting technique creates a final model based on a combination of single models such as Random Forest, but it builds them sequentially by giving more weight to instances with incorrect predictions. Specifically, in each learning cycle, prediction errors are used to calculate the gradient, i.e. the partial derivative of the loss function with respect to the prediction, and build a new tree capable of predicting gradients. Then, the prediction values are updated. After the learning phase, XGBoost derives the final predictions of the target variable by adding the average calculated in the initial step to all the residuals predicted by the trees, multiplied by the learning rate.

D. PERFORMANCE EVALUATION

To evaluate the performance of the models trained for the regression task, we adopted the following metrics: the Root Mean Square Error (RMSE), R^2, and the maximum residual error. Let $\hat{y}_i$ be the predicted value of the i-th sample and y_i is the corresponding observed value, then the two metrics can be defined as follows:

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2}$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$max\_error(y, \hat{y}) = max_{i \in [0,n-1]}(|y_i - \hat{y}_i|)$$

where n is the number of samples and $\bar{y}$ is the average of the observed values of the target variable. RMSE penalizes large errors in the prediction thanks to the quadratic exponent and makes it more sensitive to outliers. RMSE is a measure of error such that in comparing different regression models, the best is the one with the lowest RMSE value. On the contrary, R^2 score, also called the coefficient of determination, represents the proportion of variance of y that has been

FIGURE 2. Chiller model AST 2 140 used for collecting real data.

TABLE 1. Description of the variables involved in the simulation.

Variable	Description
T_{ENV}	Temperature of the environment
T_{IN}	Temperature of the water coming into the chiller
T_{OUT}	Temperature of the water coming out of the chiller
β	Empirical coefficient for the condenser heat exchange capacity
C_{cool}	Cooling capacity
P_{Tot}	Total absorbed power
P_{Fan}	Total absorbed power by fan
EER	Energy Efficiency Ratio
RPM	Revolutions per minute

described by the independent variables of the model. This metric provides an indication of the goodness of fit and thus it is a measure of the likelihood that samples never seen by the model are predicted correctly. The best value is 1 and it occurs when it is possible to predict exactly what the value of the target variable will be, knowing the values of the independent variables. The value of R^2 can be negative as the model can be arbitrarily worse than a constant model. Therefore, if we compare two regression models on the same dataset, the model with the greatest R^2 score is the best model because it has the highest predictive power. The maximum residual error reports the worst-case error between the predicted value and the true value. In this case, if we compare two regression models on the same dataset, the model with the smallest maximum residual error is the one that is getting the best performance.

IV. EMPIRICAL EVALUATION

We conducted an empirical evaluation on both synthetic and real-world data. The aim of this evaluation is to provide evidence that AI techniques can better predict some working factors of chillers during working operations and that they may be beneficial in terms of reduction of energy consumption. In the next sections, we will describe the datasets and the results of the adopted models.

A. DESCRIPTION OF EXPERIMENTAL CHILLER AND DATA ACQUISITION

As a first step, we built two different datasets, one contains synthetic data and the other contains real-world data. Both datasets have the same structure, which is represented by the list of variables reported in Table 1. The synthetic dataset has been produced by an industrial simulator that is able to generate a set of samples representing the working parameters of a device during its operation. The simulator is not publicly available. The formalization of the domain used by the simulator is now the same adopted in the control system of the chillers. Thus, the simulator provides a good starting point to collect data that describes the studied domain. After a preprocessing step that avoids the use of unnecessary features, the final dataset results in 68175 samples. A quantitative description of the values for each variable in the dataset is reported in Table 3, where β represents an empirical coefficient adopted to represent the heat exchange capacity of the condenser. In particular, $\beta = Q_c/MTD$ where Q_c is the heat of the condenser and $MTD = T_c - T_{ENV}$ is the maximum temperature difference, with T_c the condenser temperature and T_{ENV} is the temperature of the environment.

TABLE 2. Technical specification of chiller model AST2 140.(1) Data referred to nominal conditions, external ambient temperature 35 Celsius and evaporator water temperature IN/OUT 12/7 Celsius.

Model	**AST2 140**	
Cooling capacity (1)	368.4	[kW]
Total absorbed power (1)	140	[kW]
EER	2,58	
SEER	4,23	
Condenser coils		
Number of Coils	6	
Fans		
Number of Fans	6	
Total airflow	109800	[m³/h]
Nominal power (each)	1,65	[kW]
Dimensions and weights		
Width	2191	[mm]
Length mm	3465	[mm]
Height mm	2424	[mm]
Installed weight	2875	[kg]

The dataset that contains real data is made from samples collected during the testing of a commercial chiller model AST2 140. The device employed is a water chiller with air-cooled units designed for outdoor installation. It has two refrigerant circuits with hermetic scroll compressors in tandem configuration (2 for each circuit). The technical specifications of the device are reported in Table 2.

Model AST 2 140, in the nominal working conditions, is able to provide a total power of 508.4 kW (total power = cooling capacity + total absorbed power). The device is depicted in Figure 2. The chiller was tested in early February 2021 during the winter period, when low environmental temperatures allowed for testing the chiller under favorable conditions for floating HP. During the tests, several parameters were recorded. All the parameters, in particular the environmental temperature, were acquired in the environmental conditions. The final dataset is made of 40 samples, but it has been reduced to 32 for missing or incorrect values. A quantitative description of the values for each variable in the dataset is reported in Table 4. The reduced size of the dataset depends on the fact that collecting data from real-world scenarios is not an easy task as this process is time-consuming and very expensive. Collecting real data requires having commercial chillers available for testing in environmental conditions and thus in a specific period of the year or

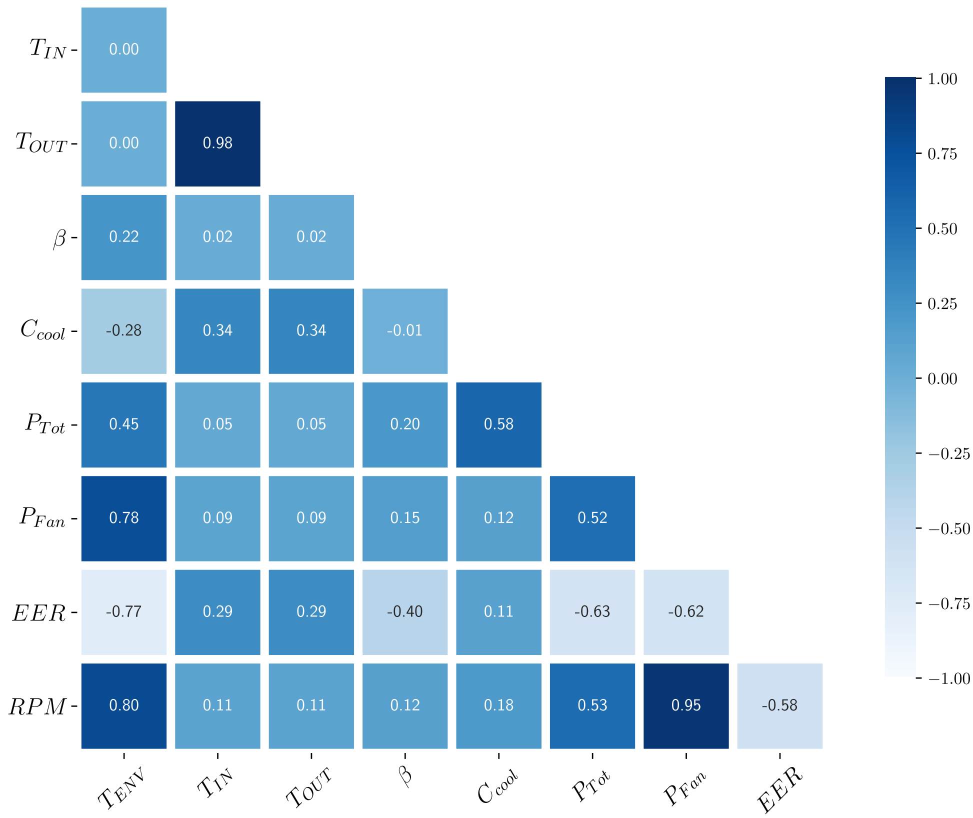

FIGURE 3. Cross-correlation matrix on synthetic data.

TABLE 3. Description of the values in the synthetic dataset that contains 68175 samples.

	Mean	Std	Min	Max
T_{ENV}	17.5545	16.1936	-10.0000	45.0000
T_{IN}	16.0000	4.3205	9.0000	23.0000
T_{OUT}	11.0000	4.3970	3.0000	19.0000
β	13.5767	1.3163	3.9129	13.8900
C_{cool}	160.6168	55.9598	70.8778	324.9042
P_{Tot}	38.3650	18.5571	11.5885	105.6539
P_{Fan}	2.3173	1.4884	0.333749	4.8434
EER	4.7389	1.8133	1.602708	13.0840
RPM	675.9252	215.8686	95.0000	950.0000

TABLE 4. Description of the values in the real-world dataset that contains 32 samples.

	Mean	Std	Min	Max
T_{ENV}	6.5256	2.1328	2.49	9.45
T_{IN}	10.1778	1.7420	7.65	12.67
T_{OUT}	8.6903	1.5923	6.69	10.87
β	8.0263	1.2450	5.88	10.16
C_{cool}	134.2800	43.9773	88.67	238.39
P_{Tot}	29.0422	9.3218	21.76	50.97
P_{Fan}	1.6584	0.9608	0.46	3.83
EER	4.6259	0.4457	3.68	5.42
RPM	13.0313	126.1314	408.00	847.00

having a testing lab that reproduces particular environmental factors. These kinds of operations are very expensive and limited because in the industry the focus is often on the functionality of a plant rather than on its efficiency in terms of energy consumption. We had the chance to collect these data from a commercial chiller that the company made available specifically for this test, but for a limited time and with restricted access to the control system of the plant.

In our experiments, RPM is the target variable: this means that we trained the models to predict the value of RPM based on the values of all the other variables which are used as input to the models.

B. CORRELATION ANALYSIS

To investigate whether some variables may affect or influence other variables, we performed a correlation analysis [7] on the synthetic dataset. This is a statistical analysis that assesses the possible relationships among the features. Among the different existing functions, in this work we employed Pearson's correlation coefficient [3]. Given two vectors $\mathbf{X}$ and $\mathbf{Y}$, the Pearson correlation index can be computed as follows:

$$r_p(\mathbf{X}, \mathbf{Y}) = \frac{\sum\limits_{x_i \in \mathbf{X}, y_i \in \mathbf{Y}} (x_i - \hat{x})(y_i - \hat{y})}{\sqrt{\sum\limits_{x_i \in \mathbf{X}, y_i \in \mathbf{Y}} (x_i - \hat{x})^2 (y_i - \hat{y})^2}}$$

where n is the number of samples, x_i and y_i are the i-th samples. $\hat{x}$ and $\hat{y}$ correspond to the average values of $\mathbf{X}$ and $\mathbf{Y}$ respectively.

The value of r_p is in $[-1, 1]$ and it describes the relationship between two variables. Specifically, a high value, close to +1, indicates a positive correlation, while a low value, close to −1, indicates a negative correlation. A positive correlation exists between two variables when with the increase in the value of one variable we can observe also an increase in the value of the other variable. On the other hand, if a negative correlation exists between two variables, then as the value of one variable increases, the value of the other variable decreases. Furthermore, when the value of the coefficient for two variables is close or equal to 0 there is a poor or no correlation between the two variables, which means that nothing can be inferred for one variable value when we observed an increase or a decrease in the value of the other variable.

Figure 3 displays the cross-correlation matrix for each pair of variables obtained on the synthetic dataset. In particular, the dark blue color depicts positive correlations while the white color depicts negative correlations. Among all the positive correlation, the more significant and of main interest for this work are the one between the variables RPM and T_{ENV} (0.80), RPM and P_{Fan} (0.95), RPM and P_{Tot} (0.53). Among all the negative correlation, the more significant and of main interest for this work are the one between RPM and EER (−0.58), EER and P_{Tot} (−0.63) and EER and P_{Fan} (−0.62). In active floating HP control, with lower working ambient temperatures, fan absorption proportionally affects more than those of compressors with regard to the total power consumption. This once again indicates the goodness of the floating head pressure control solution, when it can be enabled. Correct adjusting of the velocity of the fans can give great advantages, reducing the electrical energy absorption and bringing chillers to performance much higher. The correlation analysis summarized in Figure 3 highlights the correlation among the temperature, the velocity of the fan, and the efficiency energy ratio. In particular, the positive correlation between T_{ENV} and RPM suggests that with the increase in the environmental temperature there will be an increase in the number of revolutions per minute of the fan, this is consistent with the need to lower the temperature of condensation. On the contrary, the negative correlation between RPM and EER suggests that the higher the value for RPM the lower will be the value for EER because the fan is consuming more energy. This is a piece of empirical evidence about the usefulness of floating head pressure control: being able to correctly adjust the value for RPM with the decrease of the environmental temperature will result in a higher EER.

C. MACHINE LEARNING RESULTS

We developed and trained some regressor models for predicting the number of RPM that should be adopted in the chiller based on some other factors. The aim of these models is to predict the velocity of the fan, given the measurements of the environmental features.

The simulations are developed in Python 3.7. We adopted *DecisionTreeRegressor, RandomForestRegressor, and AdaBoostRegressor* by Scikit-learn, *XGBRegressor* by XGBoost.

In order to estimate how good is the generalization, we adopted a 10-fold cross-validation approach: the dataset was randomly divided into ten disjoint sets. One set is left out for testing while the remaining nine are used for training. The process is repeated ten times, the subset that is left out for testing changes at each iteration and we average the performance of the models. In such a way, it is possible to have a more stable performance of the different models.

We train all the models on the synthetic dataset. The average performances are reported in Table 5. They represent the average value of the metrics computed for each of the 10 folds. In each fold, the metrics are computed on the test set with synthetic data and on the whole real-world dataset. We can notice that all the metrics for machine learning models are better on synthetic data rather than real data. This is consistent with the fact that all the models are trained on the synthetic data and thus they basically learn to approximate the simulator. But on real data, an important fact is that all the machine learning models perform better than the simulator, meaning that machine learning can generalize to better approximate the distribution of real values based on environmental factors. Indeed, it can be noticed that the simulator has the worst performance in predicting a real value for fan velocity (i.e., last row in Table 5). Among the machine learning models, all the ensembles report very similar performance, with AdaBoost reporting the best prediction with lower RMSE and higher R^2 on the real data. The Decision tree model instead reports good performance on the training set, but this degrades in the testing phase, suggesting possible overfitting. In Figure 4 and 5, we plot the values predicted by the worst (i.e., decision tree) and the best (i.e., AdaBoost) prediction models along with the real values, and those predicted by the simulator. In both charts, the orange line represents the average prediction of a machine learning model, the green line represents the simulated values and the blue line represents the real values. It can be noticed that the average predictions of the machine learning models (i.e., orange dots) are always very close to the real values, while the values produced by the simulator (i.e., green dots) are far apart. This means that the machine learning models are able to predict values that are very similar to the real ones, while the simulator has a lower accuracy in the prediction. If embedded in the control system of a chiller, this higher capacity of the machine learning models would result in a better setting of the chiller and thus in lower consumption of energy. These results are just another evidence suggesting that methods tailored to data collected from the working environment may help in developing better control systems able to improve the efficiency of the chillers.

V. DISCUSSION

In this work, we have explored a new way to analyze the floating head pressure control technique employing machine learning models. These models analyze available data and, where possible, discover new patterns that can be exploited to improve the performance of the system. As already

TABLE 5. Average performance of the different models in predicting the value for variable RPM. Standard deviation in parentheses. Average performance is computed on the 10-fold test sets for synthetic data and on the whole real-world dataset for real data. For each column, best values in bold.

Model	RMSE		R^2		Max Error	
	Synthetic Data	Real Data	Synthetic Data	Real Data	Synthetic Data	Real Data
Decision Tree	6.1377 (1.5637)	49.2569 (18.0697)	0.9991 (0.0004)	0.8214 (0.0970)	284.9000 (71.1877)	237.3000 (132.4364)
Random Forest	**4.4566 (0.6443)**	27.6135 (3.6913)	**0.9996 (0.0001)**	0.9496 (0.0143)	190.7690 (45.0721)	97.4360 (32.4322)
AdaBoost	27.5563 (0.5298)	**27.3223 (1.5439)**	0.9837 (0.0007)	**0.9514 (0.0055)**	**113.4919 (5.4592)**	**62.6791 (3.9348)**
XGBoost	11.2831 (0.3108)	27.9201 (1.1706)	0.9973 (0.0002)	0.9493 (0.0043)	116.1771 (5.9115)	64.1600 (6.6556)
Simulator	-	124.0034	-	0.0022	-	311.0000

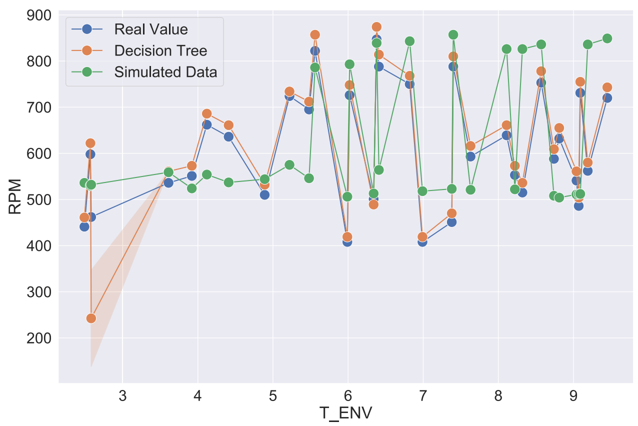

FIGURE 4. Decision tree average predictions compared with real values and the values predicted by the simulator.

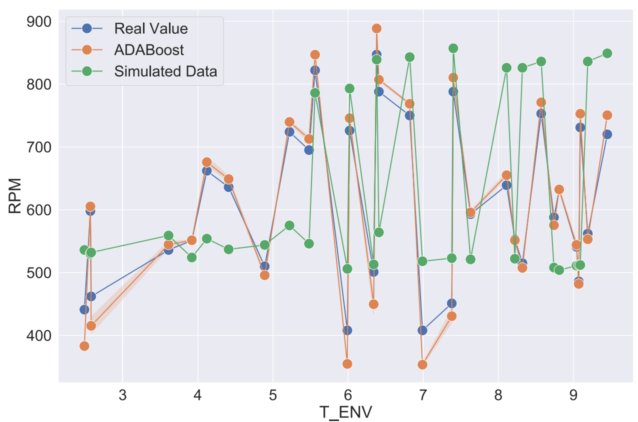

FIGURE 5. AdaBoost average predictions compared with real values and the values predicted by the simulator.

mentioned before, similar plants may perform differently due to the characteristics of the ambient and the deterioration of the components. These nuances can be identified by exploring data collected during the working time. Tailoring the operating behavior to collected data and to ambient factors can positively affect the performance of the system. In our view, these models would not substitute the existing one, but rather a combination of data-driven models and model-based systems can be adopted to benefit from the dynamic nature of the formers while at the same time maintaining safety constraints defined by the latter. This is now considered by the industry a necessity of paramount importance to capture the real operating characteristics of chillers.

Based on the environmental factors and the working conditions of the machine there exist multiple valid values for RPM but only one optimal value provides maximum energy efficiency. Finding this value depends also on the compressor operating envelope. To address this research question, we are developing a more complex model for the prediction that will take into consideration the compressor operating envelope. Validating such a model would require different more structured real data and free unconstrained access to the control system of a chiller. These conditions were not available at the time of the simulation, but we are confident that the positive results of this study will enable us to pursue this line of research in future works.

The next step is to embed these models into the control system of the chiller to automatically adapt to change by learning from new data collected at the functional time of the plant. In this way, the models can continuously improve with the experience and the changes due by the deterioration by adapting the operation of the plant to achieve maximum efficiency and lower energy consumption. These may reveal unexpected correlations or possible new behaviors, for instance, predicting unforeseen anomalies or faults. Applying ML techniques to dig in large amounts of data can help to discover patterns that were not immediately noticeable. This last step will require new and detailed experiments and simulations that would require more effort from the industry in terms of research investments and funds, but that will pay back the efforts with higher sustainability and a higher customization service offered to customers.

VI. CONCLUSION

This study aims to develop a simulator that better approximates a floating head pressure control solution for energy saving in industrial chillers. The work shed more light on the utility of using AI technologies to design and develop systems that can adapt their behavior in response to changes in environmental factors, plant conditions, and historical data.

The positive results of this study will encourage the industry to investigate these models more closely and thus they can be rigorously tested to better understand the effectiveness of these methods.

Among the many research directions, two possible future works regard the possibility of designing, developing, and testing new control logic in order to enable the floating head pressure control equipped with data-driven controls. Data collected by chillers in different scenarios can be used to retrain AI simulators to improve the accuracy of simulations loaded in the control RAM.

ACKNOWLEDGMENT

The authors would like to thank MTA SpA[5] a leading Italian company in the domain of industrial refrigeration for the support provided during the whole research, for instance by providing the synthetic dataset and to make available the commercial chiller used to collect real-world data.

DECLARATION OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] C. Aprea, R. Mastrullo, and C. Renno, "Experimental analysis of the scroll compressor performances varying its speed," *Appl. Thermal Eng.*, vol. 26, no. 10, pp. 983–992, Jul. 2006.

[2] A. Beghi, R. Brignoli, L. Cecchinato, G. Menegazzo, M. Rampazzo, and F. Simmini, "Data-driven fault detection and diagnosis for HVAC water chillers," *Control Eng. Pract.*, vol. 53, pp. 79–91, Aug. 2016.

[3] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*. Berlin, Germany: Springer, 2009, pp. 1–4.

[4] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[5] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. London, U.K.: Routledge, 2017.

[6] F. L. D. Carrir, C. Biserni, D. B. Aguiar, E. D. Dos Santos, and I. C. A. Junior, "Implementation of a floating head pressure condensation control to reduce electrical energy consumption in an industrial refrigeration system," *Appl. Sci.*, vol. 11, no. 24, p. 11923, Dec. 2021.

[7] M. Franzese and A. Iuliano, "Correlation analysis," in *Encyclopedia of Bioinformatics and Computational Biology*. Cambridge, MA, USA: Academic Press, 2019, pp. 706–721.

[8] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[9] J. H. Friedman, "Stochastic gradient boosting," *Comput. Statist. Data Anal.*, vol. 38, no. 4, pp. 367–378, 2002.

[10] E. Hajidavalloo and H. Eghtedari, "Performance improvement of air-cooled refrigeration system by using evaporatively cooled air condenser," *Int. J. Refrigeration*, vol. 33, no. 5, pp. 982–988, Aug. 2010.

[11] IEA. (2018). *The Future of Cooling*. Accessed: Oct. 14, 2022. [Online]. Available: https://www.iea.org/reports/the-future-of-cooling

[12] H. Liu, Q. Zhou, Y. Liu, P. Wang, and D. Wang, "Experimental study on cooling performance of air conditioning system with dual independent evaporative condenser," *Int. J. Refrigeration*, vol. 55, pp. 85–92, Jul. 2015.

[13] W. Lu, Z. Meng, Y. Sun, Q. Zhong, and H. Zhu, "Improved energy performance of ammonia recycling system using floating condensing temperature control," *Appl. Thermal Eng.*, vol. 102, pp. 1011–1018, Jun. 2016.

[14] M. O. McLinden, C. J. Seeton, and A. Pearson, "New refrigerants and system configurations for vapor-compression refrigeration," *Science*, vol. 370, no. 6518, pp. 791–796, Nov. 2020.

[15] T. M. Mitchell and T. M. Mitchell, *Machine Learning*, vol. 1. New York, NY, USA: McGraw-Hill, 1997.

[16] C. F. Pinnola, J. V. C. Vargas, C. L. Buiar, and J. C. Ordonez, "Energy consumption reduction in existing HVAC-R systems via a power law controlling kit," *Appl. Thermal Eng.*, vol. 82, pp. 341–350, May 2015.

[17] E. G. Pita and S. Stevenson, *Air Conditioning Principles and Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

[18] R. E. Schapire, "Explaining AdaBoost," in *Empirical Inference*. Berlin, Germany: Springer, 2013, pp. 37–52.

[19] J. Tissot, P. Boulet, F. Trinquet, L. Fournaison, M. Lejeune, and F. Liaudet, "Improved energy performance of a refrigerating machine using water spray upstream of the condenser," *Int. J. Refrigeration*, vol. 38, pp. 93–105, Feb. 2014.

[20] D. A. T. Tran, Y. Chen, H. L. Ao, and H. N. T. Cam, "An enhanced chiller FDD strategy based on the combination of the LSSVR-DE model and EWMA control charts," *Int. J. Refrig.*, vol. 72, pp. 81–96, Dec. 2016.

[21] H. Wang, "Empirical model for evaluating power consumption of centrifugal chillers," *Energy Buildings*, vol. 140, pp. 359–370, Apr. 2017.

[22] J. Yang, K. T. Chan, X. Wu, F. W. Yu, and X. Yang, "An analysis on the energy efficiency of air-cooled chillers with water mist system," *Energy Buildings*, vol. 55, pp. 273–284, Dec. 2012.

[23] F. W. Yu and K. T. Chan, "Optimizing condenser fan control for air-cooled centrifugal chillers," *Int. J. Thermal Sci.*, vol. 47, no. 7, pp. 942–953, Jul. 2008.

[24] F. W. Yu, K. T. Chan, R. K. Y. Sit, and J. Yang, "Energy simulation of sustainable air-cooled chiller system for commercial buildings under climate change," *Energy Buildings*, vol. 64, pp. 162–171, Sep. 2013.

[25] F. W. Yu, W. T. Ho, K. T. Chan, and R. K. Y. Sit, "Logistic regression-based optimal control for air-cooled chiller," *Int. J. Refrig.*, vol. 85, pp. 200–212, Jan. 2018.

ANDREA LOREGGIA received the master's degree *(cum laude)* from the University of Padova, in 2012, and the Ph.D. degree in computer science, in 2016. He is currently an Assistant Professor with the Department of Information Engineering, University of Brescia. His research interest includes artificial intelligence span from knowledge representation to deep learning. His studies are dedicated to designing and providing tools for the development of intelligent agents capable of representing and reasoning with preference and ethical-moral principles. He is a member of the UN/CEFACT group of experts, he actively participates in the dissemination and sustainable development of technology.

FRANCESCO FACCIO received the master's degree in mechanical engineering from the University of Padova, in 2019. He is currently a Mechanical Engineer with the Research and Technology Division, MTA SpA. His work is mainly focused on researching and developing simulation software useful to obtain preliminary data and perform feasibility analysis. An important part of his work is devoted to run and test prototype simulation.

DIEGO MUSSOLIN received the bachelor's degree in computer engineering from the University of Padova, in 2022. He is currently an IT Engineer with MTA SpA. His work his focused on data analysis and the development of software models to optimize business processes. His research interests include data science, big data analysis, and model development in the field of artificial intelligence.

MAURIZIO ZAMBONIN is currently a Mechanical Development Engineer. He is also a Project Manager with MTA SpA and his work is focused on the design of new model of chillers and he is in charge of the mechanical office. His research interests include thermodynamics and mechanical development.

[5]https://www.mta-it.com/ - Last visited October 7, 2022

Open Access funding provided by 'Università degli Studi di Brescia' within the CRUI CARE Agreement