

UPCommons

Portal del coneixement obert de la UPC

<http://upcommons.upc.edu/e-prints>

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works

Aquesta és una còpia de la versió author's final draft de l'article *Data Analytics for Performance Evaluation Under Uncertainties Applied to an Industrial Refrigeration Plant* publicat a *IEEE access*.

URL d'aquest document a UPCommons E-prints:
<http://hdl.handle.net/2117/165740>

Article publicat / *Published paper:*

Cirera, J.; Carino, J. A., Zurita, D. and Ortega, J.A. "Data Analytics for Performance Evaluation Under Uncertainties Applied to an Industrial Refrigeration Plant," in *IEEE Access*, vol. 7, pp. 64127-64135, 2019. doi: 10.1109/ACCESS.2019.2917079

Received April 15, 2019, accepted May 10, 2019, date of publication May 15, 2019, date of current version May 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2917079

Data Analytics for Performance Evaluation Under Uncertainties Applied to an Industrial Refrigeration Plant

JOSEP CIRERA^{ID}, (Student Member, IEEE), JESUS A. CARINO^{ID}, (Member, IEEE), DANIEL ZURITA^{ID}, (Member, IEEE), AND JUAN A. ORTEGA^{ID}, (Member, IEEE)

MCIA Research Center, Technical University of Catalonia (UPC), 08034 Terrassa, Spain

Corresponding author: Josep Cirera (josep.cirera@upc.edu)

This work was supported in part by the Department of Enterprise and Knowledge of the Generalitat de Catalunya under Grant 2017DI007 and Grant n° SGR 2017.

ABSTRACT Artificial intelligence has bounced into industrial applications contributing several advantages to the field and have led to the possibility to open new ways to solve many actual problems. In this paper, a data-driven performance evaluation methodology is presented and applied to an industrial refrigeration system. The strategy takes advantage of the Multivariate Kernel Density Estimation technique and Self-Organizing Maps to develop a robust method, which is able to determine a near-optimal performance map, taking into account the system uncertainties and the multiple signals involved in the process. A normality model is used to detect and filter non-representative operating samples to subsequently develop a reliable performance map. The performance map allows comparing the plant assessment under the same operating conditions and permits to identify the potential system improvement capabilities. To ensure that the resulting evaluation is trustworthy, a robustness strategy is developed to identify either possible new operation conditions or abnormal situations in order to avoid uncertain assessments. Furthermore, the proposed approach is tested with real industrial plant data to validate the suitability of the method.

INDEX TERMS Artificial intelligence, compression refrigeration, self-organizing maps, uncertainty.

I. INTRODUCTION

Artificial intelligence (AI) is an emerging topic due to its capability of solving complex problems in various scientific areas. The theoretical advances in this field have led its benefits to be tested in real industrial systems, where the implementation of AI in several manufacturing tasks is enabling to tackle issues such as fault detection and diagnosis (FDD) [1], process modelling [2], operation optimization [3] or performance evaluation [4] with more success than ever before. In fact, many currently complex problems in industry can be addressed by taking advantage of AI techniques to overcome the current methods limitations.

Particularly, the performance evaluation issue is a difficult topic to approach [6] due to the vast quantity of variables involved in the industrial environment, especially when the process is composed by several and distributed machines. An example of the aforementioned topic is the refrigeration

process, in which a proper and accurate performance assessment represents a real challenge due to the multiple operation combinations of loads and weather conditions along with the distributed control influences [7].

Industrial refrigeration processes are widely used in several applications to maintain the assets under the optimal conditions, although these systems consume a considerable amount of electricity [8]. The performance assessment applied to these processes is useful in order to compare different control strategies or to quantify the system improvement capabilities. Typically, taking into account the energy efficiency, the refrigeration systems are evaluated using the coefficient of performance (COP) [9]. Such coefficient, does not consider various factors or variables that could also affect or compromise the plant performance and, therefore, it provides a misleading result in order to assess the performance under different operation conditions. An improvement or deterioration in COP index is not able to differentiate between external factors such as wet bulb temperature and improper plant operation, making the coefficient inadequate to compare different

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

operation strategies or detect possible improvement capabilities or malfunctions. Hereby, various authors developed different strategies to analyze refrigeration systems performance, one of them is the exergy analysis. Some physical based methodologies such as Fan et al. [10] in 2014, assess each component separately analyzing the exergy efficiency and the exergy loss ratio. Fang et al. [11] in 2016, take the ideal exergy destruction as the ideal operation level and propose a new evaluation index-improvement potential. Belman-Flores et al. [12] in 2018 proposed a new approach to analyze the exergy performance using artificial neural networks, also applied to each system component. Gill and Singh [13] in 2018, evaluate the exergy destruction and the COP of the components to compare two different refrigerants.

With the recent introduction of the industry 4.0 framework, access to historical databases of refrigeration plants has become more accessible, therefore, data-driven approaches have recently emerged as an alternative for performance evaluation. Li and Ju [7] in 2017, apply a hierarchical cluster method to analyze the operating performance and by the analysis on the COPs of each chiller in each cluster identify the potential energy savings. Wang et al. [14] in 2017, use the operation data to approximate the best performance line of the overall chillers and use it as evaluation benchmark. Wang et al. [15] in 2018 use a dedicated set of coefficients for each of the different cooling conditions to supplement the load-power history data and include various temperature variables.

As proven in the aforementioned works, the current state of the art is mainly approaching this field using two different techniques, the physical based methods which are not effective handling complex processes with massive datasets [16], and data-driven techniques which cannot cope with uncertainty issues [17].

In this work, a data-driven methodology to evaluate the performance of a refrigeration plant is presented. Regarding to the contributions, the aforementioned challenges of data-driven approaches are addressed by proposing a method able to deal with the process signals variability, thus obtaining an assessment with similar conditions and avoiding misleading comparisons with ideal operation or fixed conditions, and applying an uncertainty detection stage to provide robustness and the capability to detect new behaviors to the methodology. To ensure that the performance evaluation is realistic and to notify the operator, each evaluated measurement of the plant must be sufficiently representative by the previously obtained evaluation grid, otherwise, the proposed methodology is able to inform about the novelty or abnormal plant operation.

The paper is organized as follows: In Section II theoretical considerations of the COP, Self-Organizing Maps and Multivariate Kernel Density Estimation are presented. The Section III presents the experimental plant description. The Section IV contains the proposed methodology. In Section V the experimental results are exhibited and finally, Section VI describes the study conclusions.

II. THEORETICAL CONSIDERATIONS

A. COP EVALUATION

The COP metric is a classical tool to evaluate refrigeration plants energy efficiency. This performance ratio is expressed as the cooling capacity provided, over the electrical power consumed. To calculate the electrical power, it has to be taken into account the compressors and the condensers [18]. This ratio is typically higher than 1 and greater values represent better performance.

B. SELF-ORGANIZING MAPS

Kohonen [19], in 1990, proposed the Self-Organizing Maps (SOM) neural network, also known as Kohonen map, used to build a topology preserving mapping. The grid of this kind of neural network tries to preserve and allocate its neurons position preserving the topological properties of the input space. The output space, also called mapped space or latent space, is a parameter to be determined. The most common output grid dimensionality is composed of two or three dimensions, which are enough and suitable for most of the applications [20].

The SOM grid is formed by various neurons also called Matching Units (MU). Every *MU* has its own *D*-dimensional weight vector w_{v-j} , where the *v*-th represent the data and *j*-th the neuron. This weight vector is the neuron coordinates in the input space. The assignment of each data point x_{v-i} to one of the grid neurons is the mapping action, the selected neuron is the one whose weight vector is closest to the data point, called the Best Matching Unit (BMU). In the output space, the position vector y_{v-i} is given by the weight vector of the selected BMU. The error function (E_{SOM}) used is shown in (1).

$$E_{SOM} = \sum_j \sum_{i \in s_{y-i}} (w_{v-j} - y_{v-i})^2 \quad (1)$$

where s_{y-i} is the set of data points which have neuron *i* as closest neuron. This error metric represents the average squared distance from the data point to its representative neuron. The objective of this technique is to minimize this error function in order to distribute the neuron grid over the input space preserving its topological properties. This minimization is performed updating the weight vectors w_{v-j} of the neurons and it can be implemented using the classical gradient descend approach:

$$w_{v-j}^{(t+1)} = w_{v-j}^{(t)} - \alpha^{(t)} \left(\nabla E_{SOM}^{(t)} \right)_{v-j} \quad (2)$$

The learning rate is not useful in such algorithm as it does not depend on the output space and does not take into consideration the neighbor neurons. Hereby, the learning rate is substituted with the neighborhood function Nhf_{wn} which depends on the mapped space:

$$Nhf_{wn}^{(t)} = \begin{cases} \alpha^{(t)} & \text{if } i \in N_{wn}^{(t)} \\ 0 & \text{if } i \notin N_{wn}^{(t)} \end{cases} \quad (3)$$

where only the nearest neurons with a certain range of the BMU in the output space are considered, $N_{wn}^{(t)}$. In this way, while executing the training phase the $\alpha(t)$ decrease monotonically and the neighborhood among the neurons in the input and output spaces is preserved.

The algorithm training performance is evaluated using the average quantization error (Qerror) (4). This metric evaluates the average distance between each input data vector with the selected BMU, where N is the number of sample vectors in the input data x_i .

$$Q_{error} = \frac{1}{N} \sum_{i=1}^N \|x_i - BMU_i\| \quad (4)$$

C. MULTIVARIATE KERNEL DENSITY ESTIMATION

The Multivariate Kernel Density Estimation (MVKDE), also referred to as Parzen windows or Parzen-Rosenblatt windows, is a flexible approach to estimate the densities of a given multi-dimensional data distribution [21]. Given a d -dimensional vector $\mathbf{X} = (X_1, \dots, X_d)^T$ where X_1, \dots, X_d are one-dimensional variables, the vector \mathbf{X}_i represents the i -th observation of the d variables: $\mathbf{X}_i = (X_{i1}, \dots, X_{id})$, where $i = 1, \dots, n$, and n correspond to the total number of observations. The variable X_{ij} is the i -th observation of the variable X_j , where $j = 1, \dots, d$. The Probability Density Function (PDF) of \mathbf{X} is, then, given by the joint PDF of the random variables $(X_1, \dots, X_d)^T$:

$$f(\mathbf{X}) = f(X_1, \dots, X_d) \quad (5)$$

Kernel functions are applied to scale distances. For example, in a one-dimensional case where $u = (x - X_i) / h$, the h is the smoothing parameter called bandwidth, and x is the currently analyzed observation. In the multivariate version, the bandwidth can be set individually for each distance $(x - X_i)$, obtaining a d -dimensional bandwidth $\mathbf{h} = (h_1, \dots, h_d)$. There are different approaches to form a multi-dimensional kernel, $K(\mathbf{u}) = K(u_1, \dots, u_d)$, is an example of a multiplicative kernel, $K(\mathbf{u}) = K(u_1) \dots K(u_d)$. Using this approach, the density estimator can be given as Eq. (6).

$$f_{\mathbf{h}}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \left\{ \prod_{j=1}^d h_j^{-1} K\left(\frac{x_j - X_{ij}}{h_j}\right) \right\} \quad (6)$$

The PDF highly depends on the selection of the bandwidth parameter vector [21]. A performing approach is to set the bandwidths through the least squares cross-validation. By this approach, each bandwidth h_j is selected so to minimize the integrated mean square error between the estimated and actual distributions as (7).

$$IMSE(h_j) = \int \{f_{h_j}(x_j) - f(x_j)\}^2 dx \quad (7)$$

III. PLANT DESCRIPTION

Industrial refrigeration plants are mainly composed by compressors, condensers, expansion valves and evaporators in order to perform the vapor-compression cycle. As previously

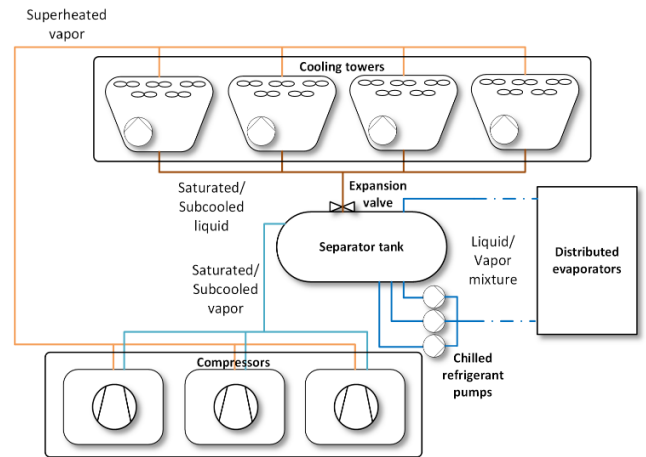


FIGURE 1. Refrigeration plant scheme.

presented in [22], the tested installation, from the company Corporación Alimentaria Guissona, S.A., contains all the aforementioned characteristics to remove the heat from their facilities. The refrigerant (R717), in the state of saturated vapor, is compressed employing three compressors in parallel, two of them are the same model and the third has less capacity. The warmth absorbed by the refrigerant is rejected using four condensers in parallel with its respective water pump and variable speed fans. The used expansion valve reduces the liquid refrigeration pressure, diminishing also the temperature. Finally, the evaporators, where the cold refrigerant is applied to lower the temperature of a closed space, are distributed along the whole facilities. The described plant is shown schematically in Fig.1.

Other components of the plant are the separator tank and the chilled refrigerant pumps. In the separator tank, the refrigerant is stored in both vapor and liquid states, the liquid part is pumped to the distributed evaporators by means of the chilled refrigerant pumps and the vapor part is absorbed by the compressors. Among all these plant components, the ones that consume the most of the energy are the compressors, the condensers water pumps, the condensers fans and the chilled refrigerant pumps [23]. The operation of the aforementioned refrigerant pumps is not modifiable since they are in charge of guaranteeing a minimum differential pressure to go over all the distributed evaporators along the facilities. On the other hand, the operation of the remaining elements can be modulated to improve the whole system performance by taking into account the components, which can be modified to increase the efficiency. Considering the data of the aforementioned components observed during the analyzed period, the condensers, which include the pumps and the fans, consume about the 20% of the overall electrical energy, while the compressors spend about the 80%.

The data employed in this study consists on samples acquired every 5 minutes from 1 year of operation, starting in March 2017 to February 2018, both included, which leads to a total of 367994 samples. Data when the plant was stopped or

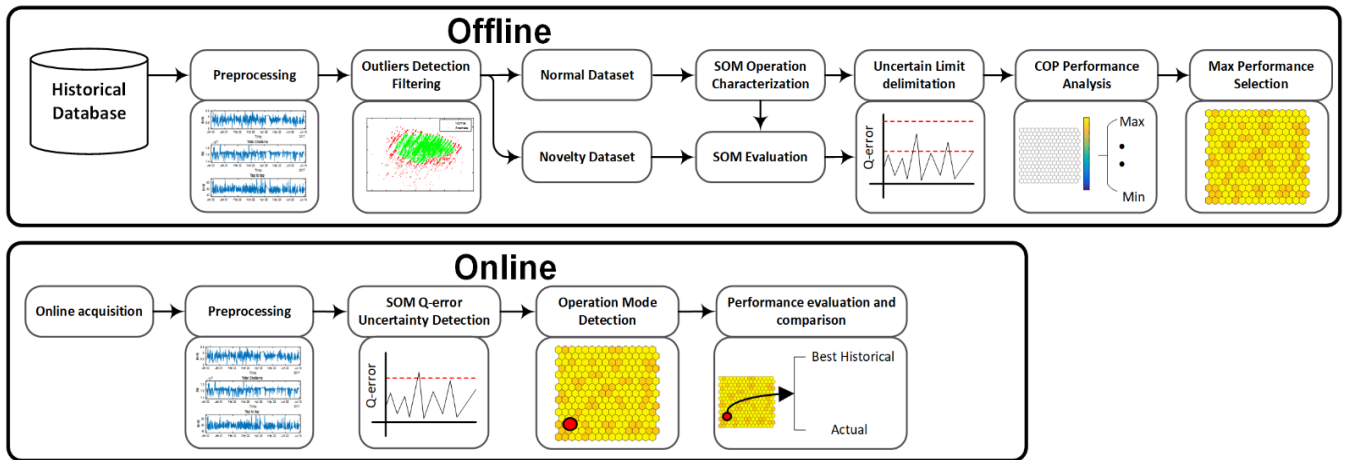


FIGURE 2. Proposed methodology. The offline block is used to create a map containing a near-optimal COP in each zone. The online block maps the new data into the discretized grid. Finally, a comparison can be made and therefore obtain the possible improvement.

out of range values was not taken into account. In comparison with the previous works [22], a reduced set of three relevant variables in order to analyze the plant performance is selected according to the plant experts. The magnitudes studied from the refrigeration plant are the compressors cooling capacity (Q_c) in kW, the suction pressure (Sp) in bar and the wet bulb temperature (T_{wb}) in $^{\circ}C$, obtained according to [24]. Fig.3 shows these signals under a short period of time.

To satisfy the fluctuating cooling demand necessities, the suction pressure is constantly changing, which is a particularity that also limit the plant performance.

IV. PROPOSED METHOD

The COP calculation in classical approaches does not consider the influence of conditional variables to the refrigeration plant performance, such as the suction pressure and the wet bulb temperature. Thus, in this work a data-driven performance evaluation methodology is proposed to tackle the aforementioned problem and evaluate the performance of the plant in regard with its operating condition, achieving a more reliable COP potential increment according to the historical plant conditions.

An initial work addressing some of these problems was previously presented [22], nevertheless many deficiencies were detected that led to the development of this improved proposed methodology, including a new reduced set of analyzed variables, a different preprocessing, the introduction of an outlier detection filtering and an uncertainty detection stage. Therefore, the potential savings can be more accurately measured instead of expecting misleading maximum COP gain values which would benefit any posterior control modification by proposing a reliable target. Additionally, an uncertain detection module is also included to increase the robustness of the method as well as the reliability of the evaluation.

The proposed methodology is shown in Fig. 2 and it is composed by two main stages, an offline operation stage where historical data is processed to obtain the maximum

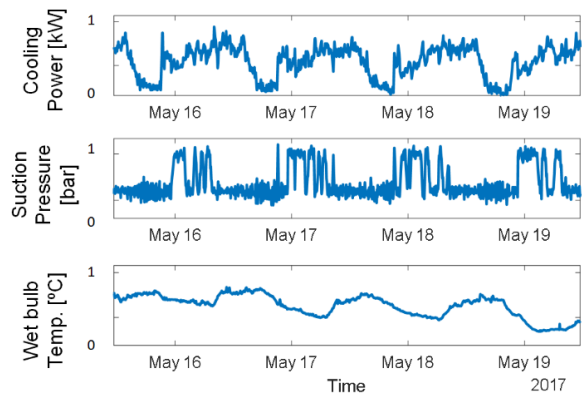


FIGURE 3. One working week of the analyzed variables. Cooling power, suction pressure and the wet bulb temperature.

performance and an online part where the performance is assessed. In the offline part, the method takes advantage of recorded data to identify the maximum COP (COP Max) performance in the different conditional operation modes of the plant. The COP Max measure inform of which can be the maximum performance achievable within the specific conditions based on real operation data. This obtained measure is a near optimal performance ratio as it is based in past operation data and it is robust as avoids non representative samples due to the uncertainty delimitation. Otherwise, in the online stage, the current plant COP (COP Actual) and the COP Max obtained from the previous stage are compared in order to assess the plant performance under the same operation conditions.

A. OFFLINE

First, the available historical data from the database (DB) is analyzed with, at least, the operation measurements of one year. The length of one year is preferred because the operation modes of the refrigeration plant are normally cyclical each year, which means that the performance is very dependent on the outside temperature. As mentioned before, the three

analyzed variables are the cooling power, the suction pressure and the wet bulb temperature.

After the extraction, the DB is filtered to eliminate periods of time where the plant is not working and periods of time where some measurements are incorrectly stored due to registered sensor failures. Additionally, the measurements are scaled from 0 to 1 for generalization purposes in order to apply posterior processing algorithms.

The next step consists on learning the distribution of the data to characterize the plant operation, nevertheless to avoid including outliers that would lead to a misleading characterization of the plant, an outlier filter is first used, in this case the MVKDE. In this particular problem, a multivariate hyperparameter tuning is beneficial for the proper selection of the anomaly boundary, therefore, the MKDE is chosen as it can be optimized for each variable analyzed in comparison with other classical techniques such as One-Class Support Vector Machine (O-C SVM) [25], which hyperparameters are unique regarding the number of the variables analyzed. The selected statistical non parametric anomaly detection technique divides the database in two parts according to the data distribution, the first one which consists on the majority of the data considered normal and a second minority one that consists on novelties or outliers. The distribution of number of samples between both sets is defined by the probability density function and the outlier threshold proposed, which in standard situations lead to around 90% of the data to the normal set and 10% to the novelty set. Nevertheless, these proportions are limited to many circumstances, included the integrity of the database, length of the analyzed period, distribution of the data, etc.

Regarding to the normal set, a SOM is used to characterize the plant operation into the 2D neuron grid of the output layer. A training procedure is employed to adapt the neurons positions to the input space, in order to preserve as much as possible, the original topology with its variance, information and distribution. Once the SOM is trained, all the output grid neurons include information about each process signal considered in the study. Indeed, after the training with the normal set, each MU, represented by each unit of the grid, describes a specific operating area of the plant. In this regard, the plant operating point codification is assessed by associating each sample to its corresponding BMU, which represents the sample operation conditions.

It is important to notice that, since this is an historical-based data-driven approach, the characterization of the plant operation is limited to the scenarios encountered in the analyzed period. Therefore, to provide awareness to the operator a module of uncertainty detection is also included, which would detect if the plant is working under new operating conditions not present in the normal set. For this uncertainty measure, SOM's Qerror is used to label new scenarios according to their value, a high Qerror would imply that the analyzed measurement corresponds to new conditions not previously considered in the training set, and a low Qerror would imply that the data correspond to the known operation conditions.

For this reason, to easily interpret and label the uncertainty of evaluated measurements two thresholds are defined according to the Qerror to obtain three labels: known, uncertain and new.

The first threshold, $Th1$, is obtained by analyzing the Qerrors obtained on the normal set. This threshold represents the first boundary that separates data considered known and data considered uncertain, therefore the uncertain concept is limited to data used in what its considered normal.

$$Th_1 = \max(Qerror(Normal\ Set)) \quad (8)$$

For the second threshold, $Th2$, the novelty set is evaluated by the trained SOM with the normal set, therefore higher Qerrors are obtained which reflect values corresponding to data that have already been considered an outlier or new. Consequently, this threshold (9), where σ represents the standard deviation, with a higher value than $Th1$, explains the limit between uncertain and new. Therefore, data between $Th1$ and $Th2$ is considered uncertain and data with higher Qerror than $Th2$ is considered new. For this second threshold, the standard deviation measure is used since it is commonly used to detect deviations in datasets [26].

$$Th_2 = 3\sigma(Qerror(Novelty\ Set)) \quad (9)$$

Once the thresholds are selected, the trained SOM grid is now associated with a performance metric. Since each MU of the grid delimits a plant operational area, the grid provides the capability to identify a range of COP values for the specific operating condition fixed by the historical registers associated to a certain MU. Therefore, the COP values are calculated for the measurements corresponding to each MU and from this set of COP values the optimal performance is selected to be the representative COP Max value in the historical data under the conditions described by the MU. A graphical description of the selection of the COP Max value in a MU is shown in Fig. 4.

As a result, each MU of the grid is now associated with a COP Max, which can lead to a realistic comparison of the COP Actual of the plant and the maximum observed in the historical with similar operation conditions. It is important to emphasize that the previous outlier model by the MVKDE is used to improve the operation characterization of the plant by the SOM and therefore reduce the Qerror of the training set.

B. ONLINE

Once the SOM is properly trained and each MU of the grid is associated with the COP Max, new samples can be evaluated to assess the actual performance and the comparison with the COP Max under the same operation conditions.

Each new sample in the online test is preprocessed with the same measures as the offline test and normalized according to the maximum and minimum values obtained during the training. After the preprocessing, the sample is evaluated by the SOM to identify the BMU on which the COP Max and the COP Actual can be compared as shown in Fig 5. Thus,

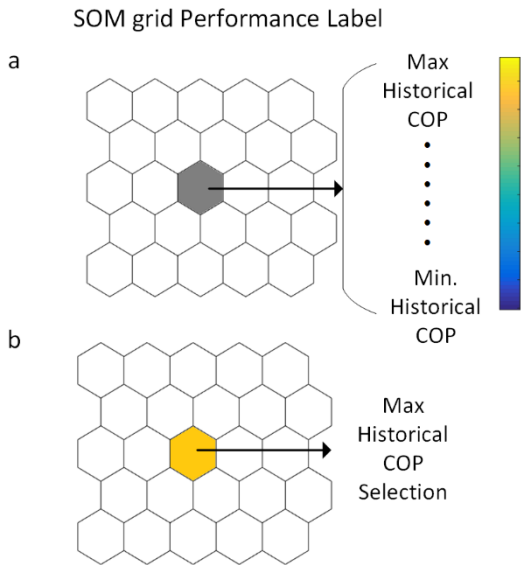


FIGURE 4. COP performance map. (a) MU historical COPs range. (b) COP Max value selection.

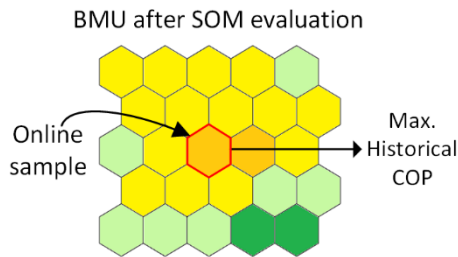


FIGURE 5. Evaluation of a new sample, the best historical performance is selected in each BMU as reference.

the performance evaluation can be made by applying (10):

$$Performance\ Improvement = COP_{Max} - COP_{Actual} \quad (10)$$

Furthermore, to give an index about the uncertainty of the evaluated measurement, the quantization error is compared with the previously defined thresholds and subsequently labeled as known, uncertain and new according to the result.

V. EXPERIMENTAL RESULTS

The proposed methodology is evaluated using data from the described plant in section III. First, the whole dataset is preprocessed as it is explained in section IV, subsequently the data is divided between training and test sets using 3 weeks per month and 1 week per month respectively. Thus, a balanced distribution among both sets is accomplished.

First of all, the data from the training set is selected, after the preprocessing, the MVKDE is employed as an outlier or anomaly filter to divide the training set in the normal set and the novelty set. For the training procedure, the MVKDE with multiplicative function and Gaussian kernel function is used. The MVKDE bandwidths are set through least squares cross-validation. With such configuration, 90% of the data is labeled as normal and 10% as novelty. To ensure the

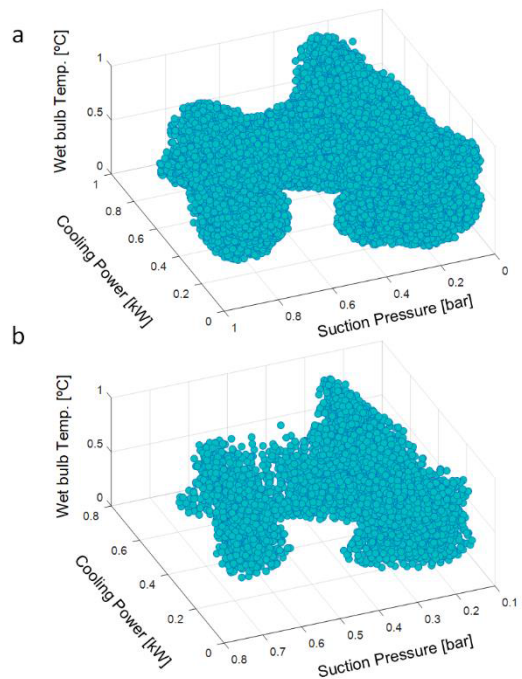


FIGURE 6. Three dimensional space comparison. (a) Data. (b) SOM grid distribution.

capability detecting anomalies of the MVKDE, a validation set previously labeled by a plant expert is used. The method is able to identify the 100% of samples labeled as abnormal, affirming the uncertainty detection effectiveness.

Once the normal set and novelty set are obtained, the SOM training is performed with only the normal set. For the SOM configuration, a rectangular grid type connection is selected with a planar map type, a Gaussian neighborhood function and a 70x70 output grid was used, which means a total of 4900 neurons. Different configurations were tested, nevertheless the aforementioned configuration presented the minimum number of neurons with no-hits without compromising the characterization resolution. Furthermore, a low mean Qerror without overfitting the network to the distribution is achieved, specifically a value of 0.013 is obtained.

The normal set used to train the SOM and the resulting grid distribution is shown in Fig. 6, the neurons' distribution covers the whole input data space allocating more neurons on denser areas to obtain more resolution.

After the SOM training with the normal set, the first threshold, $Th1$, for the uncertainty analysis is obtained according to (8). Then, the novelty set is evaluated by the trained SOM and the second threshold, $Th2$, is obtained according to (9). The values are 0.06 and 0.13 respectively.

In Fig. 7 the Qerror of the normal set and the novelty set are shown as well as the uncertainty thresholds $Th1$ and $Th2$:

After the uncertainty thresholds are selected, the COP performance analysis is performed in which each neuron of the trained SOM is assigned to a COP Max among the samples in the training set that correspond to each neuron or MU.

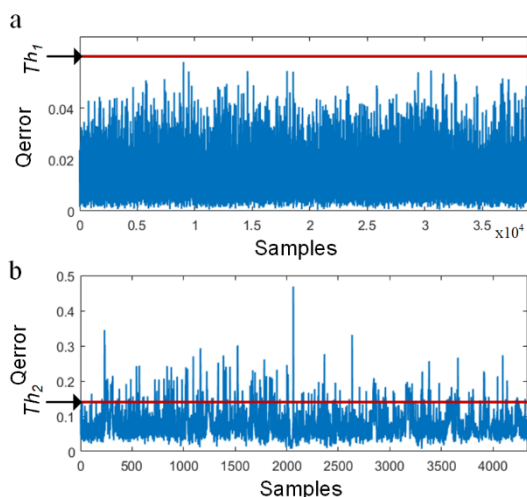


FIGURE 7. Qerror thresholds. (a) Normal dataset. (b) Novelty dataset.

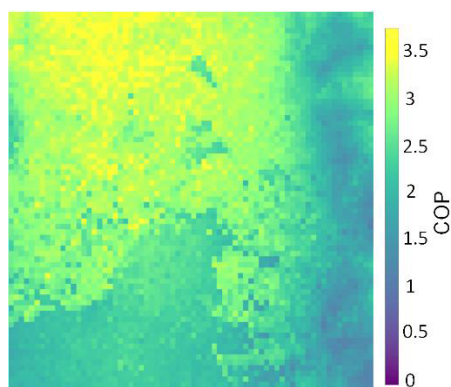


FIGURE 8. Trained SOM grid distribution with the associated COP Max for each BMU.

The anomaly detection in the training phase is used to avoid unrealistic COP Max values in each MU due to the presence of outliers. Once each neuron is assigned to a COP Max, shown in Fig 8, the offline training stage is finished and new samples can be analyzed to obtain a comparison between the COP Actual and the COP Max under similar operation conditions.

Regarding to the test set, a preprocessing is first performed and the normalization considers the maximum and minimum values obtained from the variables on the training set.

Then, to provide a reliable and qualitative value of the performance, each sample of the test set is first labeled, according to the SOM Qerror thresholds previously obtained with the anomaly detection methodology, into known, uncertain and new. The resulting labels are shown in Fig. 9.

As it can be seen, the labeled as known correspond to samples that are similar to the operation distribution of the training set, while the labeled uncertain correspond to samples that started drifting from the known operation distribution. Finally, the new labels correspond to samples that have a considerable distance from the operation conditions. In this case in particular, the lower concentration of uncertain

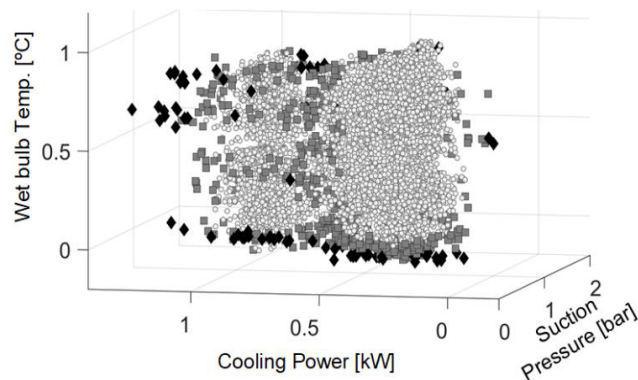


FIGURE 9. Test set labelled in the three dimensional space. \blacklozenge New samples. \blacksquare Uncertain samples. \circ Known samples.

and new samples corresponds to periods of time where the Twb reached the lowest point, such temperatures were not observed in the training set since the meteorological conditions did not coincide explicitly with the test ones. The validation set composed by plant uncertainties and novel operations used to ensure the effectiveness of the MVKDE is also used to test the Qerror thresholds being able to detect all the labeled anomalies.

Afterwards, each sample is evaluated by the SOM to obtain the corresponding BMU and therefore the COP Max with the corresponding operation conditions. The comparison of the classical COP improvement method; the COP Actual and the COP Max is shown in Fig. 10.

An application of a COP Max approach from the literature consists on calculating the COP Max value in regard to specific operation conditions [27]. In this regard, Classical COP curve, only takes into account cooling power and electrical power as it considers that the other variables remain static. With the output map provided by the SOM, it is able to appreciate that more features affect the chiller performance as plant conditions are constantly fluctuating. As it can be seen in Fig. 10, where classical COP is compared with the one achieved by the method and the current COP of the plant, the classical approach presents a lack of adaptation to the real working conditions of the refrigeration plants. Hence, it presents an overoptimistic estimation of the optimal COP since it is only calculated in regard with the current electrical and cooling powers, losing with it the global vision of the process. As it can be seen in the first subplot of the Fig 10 a), the potential improvements are constant independently of the seasonality.

Other evaluation techniques using data-driven methodologies such as the proliferation strategies [14] do not contemplate the impact of external signals variability on the plant or compressors performance. Also, the deterioration of components, such as compressors, is not considered which leads to an unrealistic performance evaluation if the COP is proliferated. Such data-driven approaches present a higher COP expectation than the classical COP curve due to the nature of oversampling without taking into account plant operation conditions.

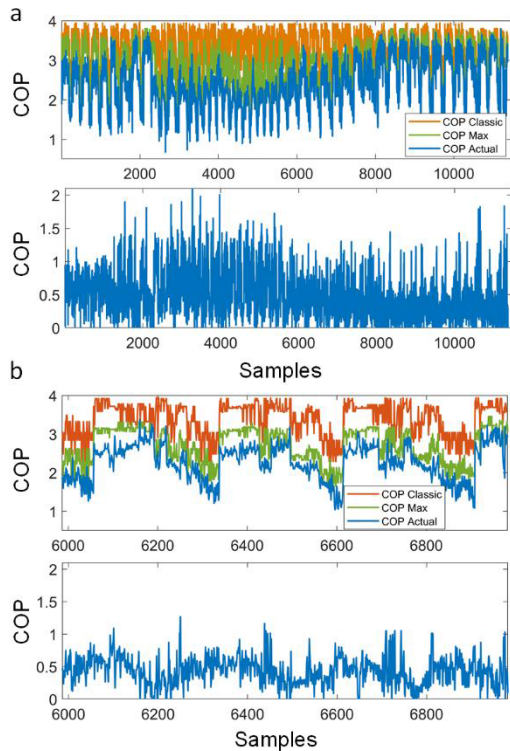


FIGURE 10. First subplot: classical COP, COP Max and COP actual. second subplot: Possible COP optimization in each time step. (a) Whole period. (b) Week detail.

The proposed methodology is able to quantify the possible COP improvement more accurately as the evaluation is performed using various operation conditions and also being robust to uncertainties. Furthermore, the demonstration that the Classical COP curve is overoptimistic since it is not possible to achieve the same performance with the same cooling load due to the affectation of other factors, is not dependent of the parametrization of the technique.

VI. CONCLUSIONS

The proposed performance evaluation methodology takes advantage of the data acquired from the refrigeration process to avoid the physical based approaches deficiencies. The strategy obtains a near-optimal performance map taking into account the process variability and the multiple variables that limit the operation efficiency. Thus, a realistic potential energy savings can be estimated since the map is developed using real plant data in comparison to classical approaches. This COP Max map can be used as a benchmark to assess new data acquired in real time from the refrigeration plant. Furthermore, the methodology is able to discriminate uncertainties from the online analysis in order to provide a robust evaluation. This tool can be used to compare different control strategies, identify abnormal behaviors and quantify the potential operation improvement.

In order to validate the study, data from a real refrigeration plant, during a period of one year, is evaluated. The results demonstrate that the normality model, uncertainty

thresholds, and the development of the discretization characterization technique, fit the plant operation multidimensional space to provide an accurate assessment of the performance. Therefore, the output performance map benchmark obtained from the offline phase contains reliable historical information according to the operation conditions. With the test dataset is observed that the plant has room for improvement varying the control strategy and the non-well represented samples are identified as abnormal situations. Also, if the control strategy is constant, a low performance could indicate machine failures, since not optimal COPs are being obtained in similar conditions.

The abnormal situations match with the conditions less represented in the offline training stage, in the test dataset are principally identified with low wet bulb temperatures since the training set did not contain such similar values.

Further work can be approached incorporating these new behaviors detected to the performance map, thus the knowledge of the benchmark map will be increasing automatically whenever new operation conditions appear. Also, furtherly testing this methodology to provide it of a fault tolerant or fault detection strategies.

ACKNOWLEDGMENT

The authors would like to thank the support of Corporación Alimentaria Guissona S.A. for providing access to their refrigeration plant dataset and their recommendations.

REFERENCES

- [1] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, "Enabling health monitoring approach based on vibration data for accurate prognostics," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 647–656, Jan. 2015.
- [2] T. Vantuch, S. Mišák, T. Jeżowicz, T. Buriánek, and V. Snášel, "The power quality forecasting model for off-grid system supported by multiobjective optimization," *IEEE Trans. Ind. Electron.*, vol. 64, no. 12, pp. 9507–9516, Dec. 2017.
- [3] S. R. Thangavelu, A. Myat, and A. Khambadkone, "Energy optimization methodology of multi-chiller plant in commercial buildings," *Energy*, vol. 123, pp. 64–76, Mar. 2017.
- [4] D. B. Jani, M. Mishra, and P. K. Sahoo, "Application of artificial neural network for predicting performance of solid desiccant cooling systems—A review," *Renew. Sustain. Energy Rev.*, vol. 80, pp. 352–366, Dec. 2017.
- [5] J. Beiranvand and H. Meghdadi, "Analytical performance evaluation of MRC receivers in massive MIMO systems," *IEEE Access*, vol. 6, pp. 53226–53334, 2018.
- [6] Y. Fang, Y. Yuan, L. Li, J. Wu, W. Lin, and Z. Li, "Performance evaluation of visual tracking algorithms on video sequences with quality degradation," *IEEE Access*, vol. 5, pp. 2430–2441, 2017.
- [7] M. Li and Y. Ju, "The analysis of the operating performance of a chiller system based on hierarchal cluster method," *Energy Buildings*, vol. 138, pp. 695–703, Mar. 2017.
- [8] J. Gill and J. Singh, "Performance analysis of vapor compression refrigeration system using an adaptive neuro-fuzzy inference system," *Int. J. Refrig.*, vol. 82, pp. 436–446, Oct. 2017.
- [9] M. R. Braun, P. Walton, and S. B. M. Beck, "Illustration de la relation entre le coefficient de performance et le coefficient de performance du système grâce à un système frigorifique de supermarché au R404," *Int. J. Refrig.*, vol. 70, pp. 225–234, Oct. 2016.
- [10] B. Fan, X. Jin, X. Fang, and Z. Du, "The method of evaluating operation performance of HVAC system based on exergy analysis," *Energy Buildings*, vol. 77, pp. 332–342, Jul. 2014.

[11] X. Fang, X. Jin, Z. Du, and Y. Wang, "The evaluation of operation performance of HVAC system based on the ideal operation level of system," *Energy Buildings*, vol. 110, pp. 330–344, Jan. 2016.

[12] J. M. Belman-Flores, J. M. Barroso-Maldonado, S. Ledesma, V. Pérez-García, A. Gallegos-Muñoz, and J. A. Alfaro-Ayala, "Exergy assessment of a refrigeration plant using computational intelligence based on hybrid learning methods," *Int. J. Refrig.*, vol. 88, pp. 35–44, Apr. 2018.

[13] J. Gill and J. Singh, "Component-wise exergy and energy analysis of vapor compression refrigeration system using mixture of R134a and LPG as refrigerant," *Heat Mass Transf.*, vol. 54, no. 5, pp. 1367–1380, May 2018.

[14] Y. Wang, X. Jin, and X. Fang, "Rapid evaluation of operation performance of multi-chiller system based on history data analysis," *Energy Buildings*, vol. 134, pp. 162–170, Jan. 2017.

[15] Y. Wang, X. Jin, Z. Du, and X. Zhu, "Evaluation of operation performance of a multi-chiller system using a data-based chiller model," *Energy Buildings*, vol. 172, pp. 1–9, Aug. 2018.

[16] C. Fan, F. Xiao, Z. Li, and J. Wang, "Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review," *Energy Buildings*, vol. 159, pp. 296–308, Jan. 2018.

[17] S. Yin, X. Li, H. Gao, and O. Kaynak, "Data-based techniques focused on modern industry: An overview," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 657–667, Jan. 2015.

[18] F. W. Yu and K. T. Chan, "Modelling of the coefficient of performance of an air-cooled screw chiller with variable speed condenser fans," *Building Environ.*, vol. 41, pp. 407–417, Apr. 2006.

[19] T. Kohonen, "The self-organizing map," *Neurocomputing*, vol. 21, nos. 1–3, pp. 1–6, 1998.

[20] T. Kohonen, "Essentials of the self-organizing map," *Neural Netw.*, vol. 37, pp. 52–65, Jan. 2013.

[21] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Process.*, vol. 99, pp. 215–249, Jun. 2014.

[22] J. Cirera, M. Quiles, J. A. Carino, D. Zurita, and J. A. Ortega, "Data-driven operation performance evaluation of multi-chiller system using self-organizing maps," in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, Feb. 2018, pp. 2099–2104.

[23] X. Wei, G. Xu, and A. Kusiak, "Modeling and optimization of a chiller plant," *Energy*, vol. 73, pp. 898–907, Aug. 2014.

[24] R. Stull, "Wet-bulb temperature from relative humidity and air temperature," *J. Appl. Meteorol. Climatol.*, vol. 50, no. 11, pp. 2267–2269, 2011.

[25] K. Yan, Z. Ji, and W. Shen, "Online fault detection methods for chillers combining extended Kalman filter and recursive one-class SVM," *Neurocomputing*, vol. 228, pp. 205–212, Mar. 2017.

[26] H. Yu, F. Khan, and V. Garaniya, "Risk-based fault detection using self-organizing map," *Rel. Eng. Syst. Saf.*, vol. 139, pp. 82–96, Jul. 2015.

[27] Y. Yao, Z. Lian, Z. Hou, and X. Zhou, "Optimal operation of a large cooling system based on an empirical model," *Appl. Therm. Eng.*, vol. 24, no. 16, pp. 2303–2321, 2004.



JESUS A. CARINO (M'13) received the M.S. degree in electrical engineering from the University of Guanajuato, Mexico, in 2012, and the Ph.D. degree in electronics engineering from the Technical University of Catalonia (UPC), Barcelona, Spain. His research interests include fault diagnosis in electric machines, novelty detection, pattern recognition, artificial intelligence applied to industrial processes monitoring, data analytics, and digital signal processing on FPGAs for applications in mechatronics.



DANIEL ZURITA received the Ph.D. degree in electronics engineering from the Technical University of Catalonia (UPC), in 2017. He is currently a Data Scientist and a Ph.D. Researcher with MCIA-UPC. His research interests include machine learning, data analysis, and industrial process monitoring. He is currently developing machine learning algorithms and smart monitoring solutions in industrial projects in the field of Industry 4.0, and the IIoT.



JOSEP CIRERA was born in Lleida, Catalonia, Spain, in 1988. He received the B.S. degree in electronic engineering from the Technical University of Catalonia (UPC), Barcelona, Spain, in 2012, and the M.S. degree in automation and electronics engineering from UPC, Terrassa, Spain, in 2016, where he is currently pursuing the Ph.D. degree in electronic engineering.

From 2012 to 2017, he was an Automation Engineer with the company Corporación Alimentaria Guissona S.A. (CAGSA), Guissona, Spain. Since 2017, he has been developing his research project at CAGSA in collaboration with UPC. His research interests include performance evaluation, load forecasting, and industrial process optimization methodologies by means of data science algorithms.



JUAN A. ORTEGA (M'94) received the M.S. degree in telecommunication engineering and the Ph.D. degree in electronics from the Technical University of Catalonia (UPC), in 1994 and 1997, respectively. In 1994, he joined the UPC Department of Electronic Engineering, where he is currently an Assistant Professor. Since 2001, he has been with the MCIA Research Center, where he has coordinated and collaborated on multiple national and international research projects on his research interests. He has authored and coauthored over 150 journal and conference papers. His current research activities include industrial process monitoring, fault detection algorithms, machine learning, signal processing, smart sensors, and embedded systems.

...