18th Brazilian Congress of Thermal Sciences and Engineering
November 16–20, 2020 (Online)

ENC-2020-0356
BIG DATA CLUSTERING MODEL FOR THE IDENTIFICATION OF A THERMAL POWER PLANT OPERATING PATTERNS

Jéssica Duarte
Lara Werncke Vieira
Augusto Delavald Marques
Paulo Smith Schneider
Federal University of Rio Grande do Sul, Department of Mechanical Engineering, Porto Alegre, RS, Brazil
jessica.jd.duarte@gmail.com
lara.vieira@ufrgs.br
augustod.marques@gmail.com
pss@mecanica.ufrgs.br

Guilherme Lacerda Batista de Oliveira
Energy of Portugal - EDP
guilherme.oliveira@edpenergiapecem.com.br

Abstract. *Thermal power plant operation depends on the knowledge of a wide range of complex and cross dependent parameters. Information is usually captured through Distributed Control Systems (DCS) which allows to access up to date data but also long periods of recorded operation. Large and available data sets are decisive for plant operation, but they must be properly used and interpreted to achieve effectiveness. The purpose of the present paper is to present an identification of operational patterns from historical data from an actual thermal power plant based on unsupervised machine learning methods. The proposed methodology is applied to a long term data series from the 360 MW Brazilian coal-fired Pecem power plant, for 40 selected parameters, concerning its steam generator and associated mills. Dataset size and redundancy is treated by the Principal Component Analysis (PCA) approach, which defines a lower dimensional space, proper for clustering while preserving most of its variance. The K-means clustering method identifies operating point groups according to their degree of similarity. The appropriate cluster number is defined by means of the average silhouette coefficient, which measures the clusters consistency. Cluster parameter values and distribution are evaluated to verify result consistency. Assessments with the 40 initial parameters and with a subset of 29 parameters from the steam generator and mills system are presented. The latter's generate more useful and physically relevant results, being described globally by a 2 clusters analysis or, for refined observations, by a 10 clusters analysis. The different patterns encountered facilitate an understanding of the parameters arrangement and resulting performance, enabling the identification of higher efficiency operation conditions and supporting practices to improve the plants operation.*

Keywords: *Power plant operation, Operation patterns, Operation parameters, K-means clustering, PCA.*

1. INTRODUCTION

The future of energy generation is one of today's most important and complex challenges. Our society is dependent on electric and thermal energy, from our daily needs into the most energy-intensive industries. Different scenarios on the future of energy were developped by IEA (2018), and an increase in energy demand is expected to grow by more than 25% until 2040. Thermal power plants are nowadays the main source of global energy generation and it's not yet to be counted out of the global power mix predictions, demanding research on an efficient operation.

The comprehension of its optimal operation could improve energy consumption and reduce harmful emissions, but that is not an easy task due to the complexity of industrial scale power plants. Data flow is accessed by a Distributed Control System (DCS), which records long periods of operation and generates an important amount of data. It must be properly interpreted to result in effective information. Different machine learning methods have been being used to infer from this data and investigate such complex processes.

Unsupervised machine learning methods are vastly applied for pattern identification. Kuriak *et al.* applied k-means clustering method to generate control signatures for real-time optimization of the combustion process (Kusiak and Song, 2006), which was expanded to assign different weights to different variables while considering coupling effects (Song and Kusiak, 2007), being able to lead to operation control settings to optimize boiler efficiency. Fuzzy c-mean (FCM) clustering method is applied in different works in order to find patterns on different power plant operations. Hou *et al.* applies it for performance improvement of ultra-supercritical power plants by real-time running data (Hou *et al.*, 2019), which results demonstrate the high precision of identified model. The optimization of a desulfurization system operation (Liu *et al.*, 2020) was also analysed by the FCM method, with the application of k-means clustering on determining the patterns centroids, which was able determining the optimal parameter settings for operation guidance. Xiaoying *et al.* also applied FCM to establish a real-time predictive control of oxygen content in coal-fired power plants, combining it with a subspace method, building a combustion process model that real-time simulation verified the achievement of acceptable accuracy (Xiaoying *et al.*, 2016). Wang *et al.* also modelled thermal power units by k-means clustering method with Spark-based FP-growth algorithm (Wang and Jia, 2019), to mine optimization targets values to improve its economical index.

This paper aims to recognize the different patterns that may occur on a thermal power plant operation, based on historical data with different operation and steam generator parameters, by means of unsupervised machine learning methods due to their ability to handle unlabeled data. The K-means clustering technique is applied to identify groups of datapoints within the data, with the support of Principal Component Analysis (PCA) to reduce the database dimensionality and of the silhouette coefficient to determine the appropriate number of clusters. Since unsupervised machine learning methods do not have delineated validation techniques (Igual and Seguí, 2017), the quality validation of the results is based on a subjective evaluation that requires prior comprehension of the operation processes, rising an interpretation of its usefulness and physical relevance in order to define the final clusters arrangements. The definitive cluster configuration is presented to enrich the comprehension of the operation patterns.

2. SYSTEM DESCRIPTION

PECEM power plant is located at São Gonçalo do Amarante, in Ceará, and is responsible for 50% of the energy generation complex of the state. The power plant is composed by two independent sub-critical coal-fired power generating units, working in two different electrical regimes of 240MW and 360MW. In this paper, selected parameters covering the mills and the steam generator processes of one generation unit at PECEM power plant are analysed for the recognition of operation patterns. Figure 1 presents a diagram of the steam generator and operating mills for the studied generation unit.

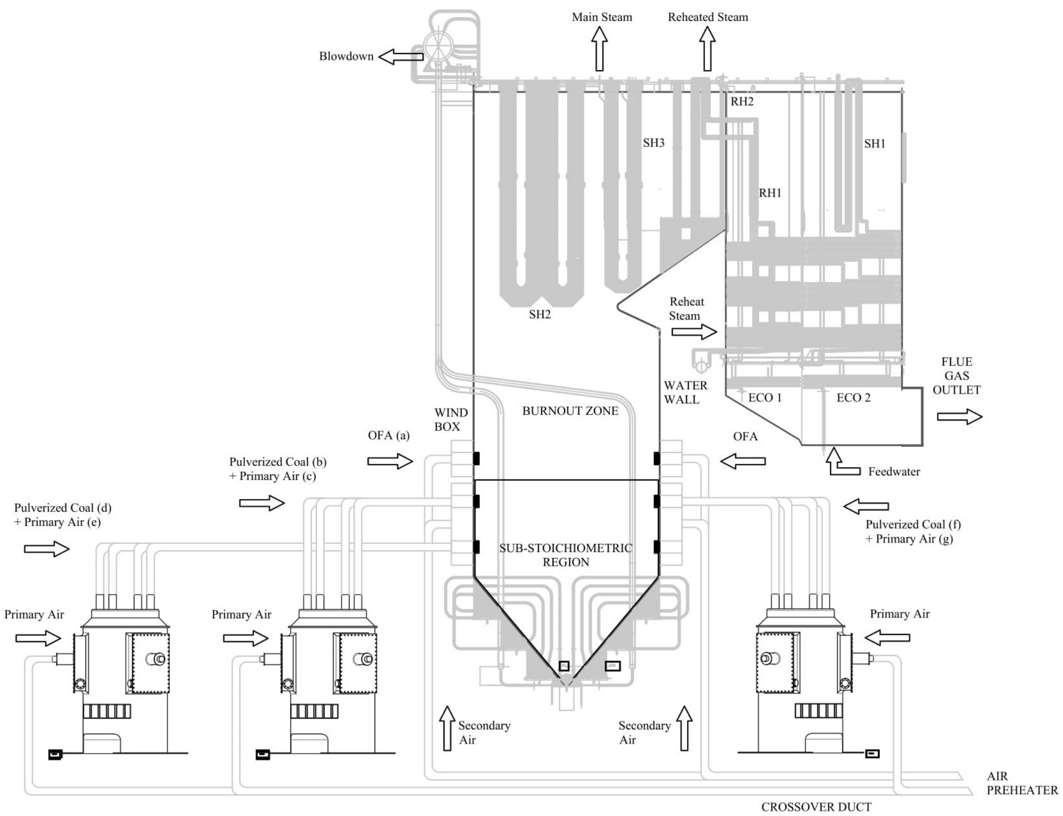


Figure 1: Steam generator and mills flows scheme.

The steam generators operates with two steam flows, the main steam and the reheat steam, absorbing heat released by the fuel combustion and delivering steam at a rated temperature, pressure and capacity to the cicle turbines. There are four independent mills, three of them simultaneously feeding each steam generator with dry pulverised coal on the operation and one remaining stopped as a backup. A preheated air stream is split into two different paths, the primary and secondary air flows. The primary air flow is fed into the mills to dry and transport coal to the steam generator burners, while the secondary air flow is connected to the combustor by the wind box to be directly mixed with the primary air and pulverized coal stream. The coal and air mix burn in the furnace under sub-stoichiometric conditions, completed with more oxygen from the OFA ports at the burnout zone. Feedwater is admitted in countercurrent to the flue gases at the economizers, evaporates at the furnace water walls and is superheated at the superheaters. The main steam and the reheated steam streams feed the cycle turbines, generating the unit electrical power.

3. APPLIED METHODS

3.1 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical technique for data reduction with axis rotation (Aggarwal, 2015), widely used before the application of machine learning methods. The dataset is represented with fewer components, increasing its interpretability is a way to minimize information loss. The original multidimensional space is represented by new orthogonal components, generated by a linear combination of the original ones, in a way to maintain the most of the total data variance. The resulting components are sorted from the highest to the lowest variance representation, and a subset of the first number of components is selected to be the new dimensional space.

3.2 K-means Clustering

Clustering methods are unsupervised machine learning techniques aiming to divide unlabeled data into homogenous subgroups, according to the degree of similarity between them (Kubat, 2017). The k-means clustering method uses the Euclidian distance metric in such a way that each item is only assigned to one cluster, with the nearest cluster centroid. Its implementation requires the specification of the number of clusters beforehand, and finding a reasonable optimal clustering number is essential for the accuracy of the results. A cluster of datapoints from operation parameters is herein considered to represent an operation pattern.

3.3 Silhouette Coefficient

The Silhouette Coefficient is an internal evaluation of the clustering consistency result and is considered in this paper to determine the more adequate number of clusters for a dataset. This cluster validation criteria considers the intracluster proximity and the intercluster separation. It ranges from -1 to 1, as a high positive value represents a compact cluster. It can be defined as Eq. 1.

$$Silhouette = \frac{b-a}{max(a,b)} \tag{1}$$

Where a is the intracluster distance of a sample in the dataset and b is the nearest cluster distance of a sample to the next cluster, by Euclidian distance metric (Igual and Seguí, 2017).

4. METHODOLOGY

The present work employed a methodology framework composed of 12 steps, that can be observed in Fig. 2.

The first step is the parameter selection. The analysis should contemplate process elements which may be relevant on the characterization of the operation to be studied, considering the available data. Therefore, an understanding of the processes and parameters guides this selection.

Following the parameter selection is the data preprocessing, with the outlier removal and data standardization steps. The data cleaning by the removal of outliers improves the quality of data mining. The data point is identified as an outlier if any of its parameter values is three standard deviations away from the mean. Following the outlier removal, data standardization is the conversion of the different data structures into a common data format. It is required for machine learning applications when features have different ranges.

The PCA first step is the definition of the new components. Those new orthogonal components represent the original multidimensional space, and the cumulative original variance for each component is obtained. The dimensionality reduction step is then the definition of the minimal number of PCA components for the requested cumulative variance, to be used in the analysis.

The K-means Clustering starts with the the clustering step, with the application of the method to the dimensionally reduced data. As its implementation requires the specification of the required number of clusters beforehand, the data is clustered multiple times for different numbers of clusters, in order to select the best fitted number in the next step. The average silhouette coefficient is calculated for each set of the multiple clustering from the previous phase, and an average silhouette coefficient curve is plotted with the objective of evaluating the clustering consistency results. For the

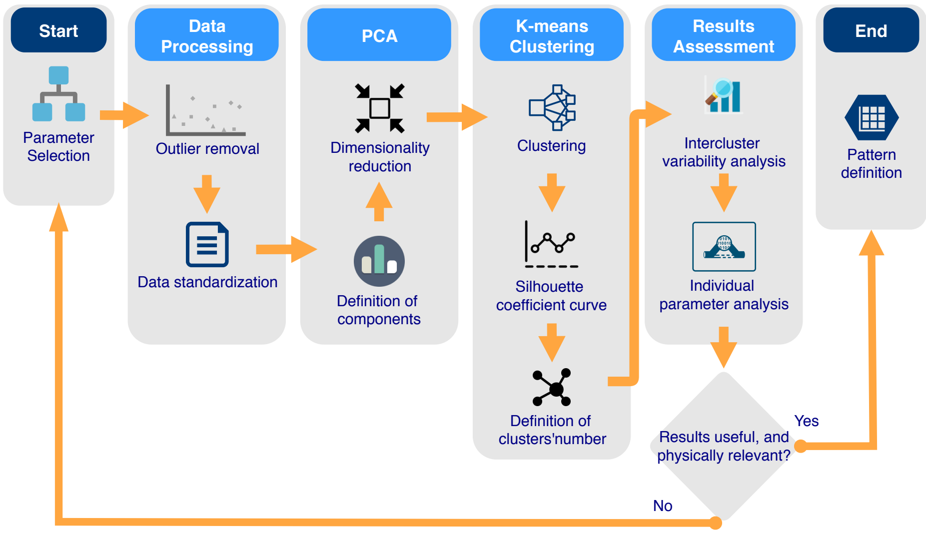


Figure 2: Methodology framework.

definition of the number of clusters step, the higher average silhouette coefficient values are identified on the curve. The corresponding number of clusters is adopted. A low number of clusters will establish well defined global patterns in the operation, but a higher number of clusters will promote a more sensible identification of different patterns in the analysis.

The Results Assessment starts with the the intercluster variability analysis step. The clustering results for the chosen number of PCA components and the chosen number of clusters is analyzed. For each parameter, boxplots for the different clusters are generated and compared. A visual inspection of these plots may indicate the intercluster behavior of the selected parameters and, also, the coordination between different parameters. The following step is the individual parameter analysis, to inspect specific parameter behaviors.

Finally, the usefulness and physical relevance of the different clusters results in order to obtain or not process information is questioned. This step is inherently subjective, since a comprehension of the process is required for such evaluation. If the results are judged as not useful or relevant, the parameters selection step is reconsidered, leading to a new analysis. The parameters that are found not to improve the results discussion or to interfere its interpretation may be excluded from the analysis, as well as different new parameters may be included. On the other hand, if the results observed are physically relevant, it is possible to define the encountered clusters as the power plant operation patterns.

5. CASE STUDY

The presented method is applied to identify operation patterns from 40 selected parameters of one generation unit at PECEM power plant. The selected parameters cover the mills and the steam generator thermodynamic processes, and are presented in Appendix A. The analysis is conducted firstly considering all listed parameters and, following its findings, the parameters selection step is reconsidered and 11 parameters are disconsidered. A second analysis is conducted considering a subset of 29 out of the 40 initial parameters, as a continuation of the application of the methodology. The two parts of the analysis will be labeled Analysis A and Analysis B, respectively.

In the present study, the data is taken at one hour intervals over a 14 month period, from the beggining of September 2018 to the end of October 2019. The analysis was done to the plant's designed power generation, which is the 360MW operating load. The generation unit operates under this condition 33% of the time over the considered period, resulting in 3.357 samples.

5.1 Analysis A

Analysis A considers all 40 parameters of the dataset. The outlier removal removes 16,6% of the original dataset. The dimensionality reduction by PCA is applied to the dataset after standardization. The number of PCA components chosen for this analysis can be defined by the cumulative variance, according to Tab. 1.

The first column in Tab. 1 refers to the cumulative variance of the original data explained according to the number of components. Since the first ten components are able to represent 81.24% of the cumulative variance and reduces the original 40 components dimentionality, it is chosen to be used in the analysis instead of the original 40.

Aiming to understand the most fitting clustering options for the dataset, the data is clustered multiple times, from 2 to 13 clusters, and the results for each are inspected by the average silhouette coefficient. To better understand the dimensionality results and to verify that the definition of 10 components is adequate, this is done for each considered

Table 1: PCA cumulative variance result for Analysis A.

Cumulative Variance	Number of Components
100.00%	40
95.13%	20
90.08%	15
85.47%	12
81.24%	10
76.44%	8
70.95%	6
61.61%	4
45.89%	2

number of components PCA reduction in Tab. 1, from 2 to 40 components, each one presenting a different curve. The average silhouette coefficient is determined for each clustering result, and Fig. 3 presents its results.

Figure 3: The average silhouette coefficient from clustering with different number of clusters for each number of PCA components for analysis A.

It can be observed that the lower the number of components, the higher is the average silhouette coefficient, indicating more consistent clustering results. The data sparcity increases rapidly with the volume of high-dimensional spaces. In other words, with the higher number of components, more difficult it is to cluster the data. On the other hand, as presented in Tab. 1, a low number of components wouldn't represent adequately the original dataset variance. As a consequence, the shape of the curve changes and the results are affected. The defined number of 10 PCA components sufficiently conserves the higher-dimentional average silhouette coefficient curve shape.

Observing Fig. 3, it is also possible to identify that, for 10 components, the maximum average silhouette coefficient of 0.33 is for 3 clusters. Figure 4 present the clustering results with different colors.

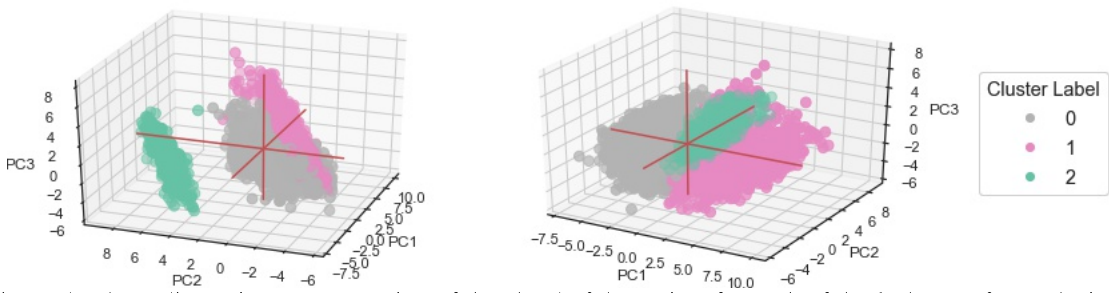


Figure 4: Three-dimension representation of the cloud of datapoints for each of the 3 clusters for analysis A.

It can be noted that the clustering results identified a single separated cluster, represented by the green datapoint cloud at Fig. 4.

The intercluster variability analysis is then performed, examining the clustering results regarding the 40 original selected operation parameters. For each of those parameters, the boxplots for the three different clusters are generated and compared. A visual inspection of the behavior between clusters for all operation parameters found that specifically the parameters from the feedwater stream heat exchangers presented a particular pattern. The boxplots for four of these parameters are presented at Fig. 5.

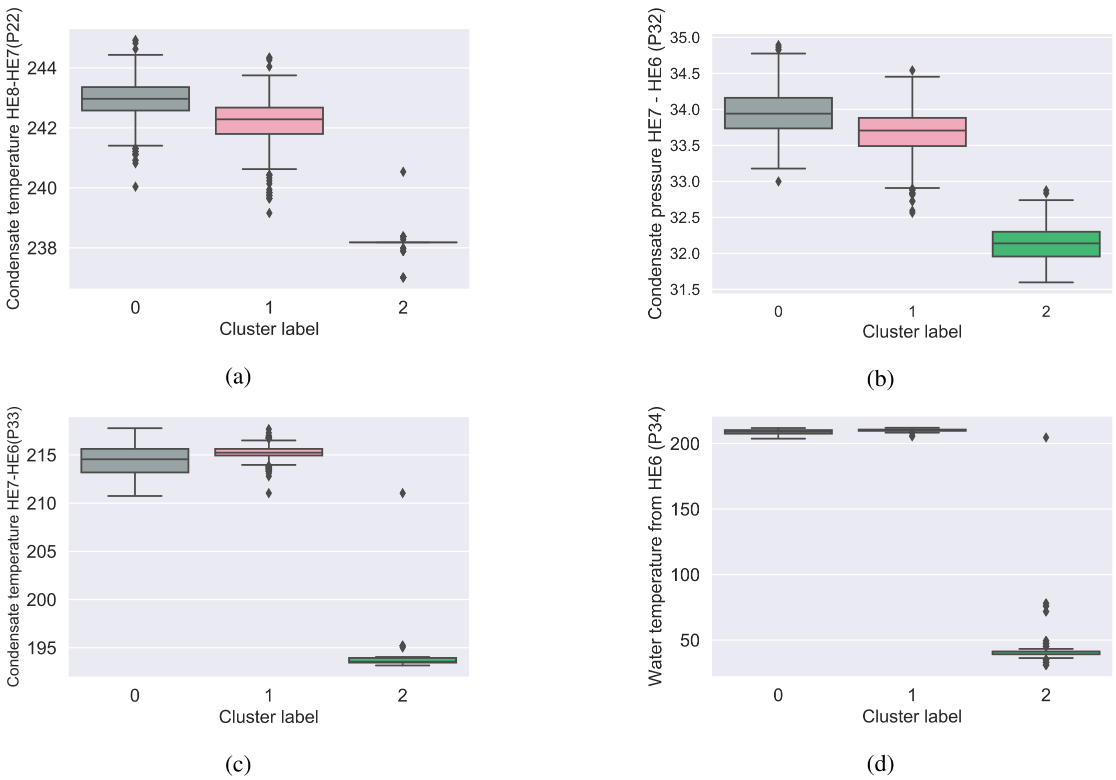


Figure 5: Boxplots for the 3 clusters: (a) condensate temperature from heat exchanger (HE) 8 to HE7 (P22), (b) condensate pressure from HE7 to HE6 (P32), (c) condensate temperature from HE7 to HE6 (P33) and (d) water temperature from HE6 (P34).

The presented parameters are the condensate temperature from heat exchanger (HE) 8 to HE7 (P22), condensate pressure from HE7 to HE6 (P32), condensate temperature from HE7 to HE6 (P33) and water temperature from HE6 (P34). It can be observed in Fig. 5 that cluster 2 identified contrasting values in comparison with cluster 0 and 1.

It is possible to observe that the separated datapoint cloud represented in green at Fig. 4 happens due exclusively to the heat exchangers parameters. These results indicate that the operation of these parameters, from the feedwater flow heat exchangers, present a distinct pattern from the other operation parameters, consequently having an important impact at the cluster results. However, the impact of this heat exchanger system alone do not have as much influence on the global process of the power plant, and are not parameters that the operators have great power to act on.

For this reason, at the methodology step when questioning whether the results are useful and physically relevant, these results are judged as not useful. The analysis is redone, excluding the 11 heat exchangers parameters from the database, as presented in Analysis B.

5.2 Analysis B

Analysis B considers the dataset with 29 parameters. The original dataset was 15% reduced by applying the outlier removal procedure. Next, all the steps of the methodology described in Section 4.were performed for this new dataset. The dimensionality reduction by PCA is applied after the dataset standardization. The cumulative variance obtained by the selected number of components is presented in Tab. 2.

Table 2: PCA cumulative variance result for analysis B.

Cumulative Variance	**Number of Components**
100.00%	29
95.61%	19
90.41%	15
86.67%	13
82.01%	11
76.33%	9
62.35%	5
39.20%	2

For this analysis, the number of components is here defined to be 11. It represents 82,01% of the variance, while it substantially reduces the original data dimensionality. The data is clustered multiple times, from 2 to 13 clusters, which is done for each number of components considered at the PCA reduction. Figure 6 presents the average silhouette coefficient determined for each clustering result.

Figure 6: The average silhouette coefficient from different number of clusters for each number of PCA components for analysis B.

The curve for 11 PCA components is similar to the curves for 13 to 29 PCA components, conserving the higher-dimentional average silhouette coefficient curve pattern. Moreover, the higher average silhouette coefficient of 0.26 is observed occurring for 2 clusters. With this, its possible to understand that the separated datapoint cloud is not present in this analysis. Moreover, a local maximum average silhouette coefficient of 0.19 is observed for 10 clusters, whose analysis may distinguish subtle operation patterns. Figure 7 and 8 present the clustering results with different colors for the 2 and the 10 clustering analysis, respectively.

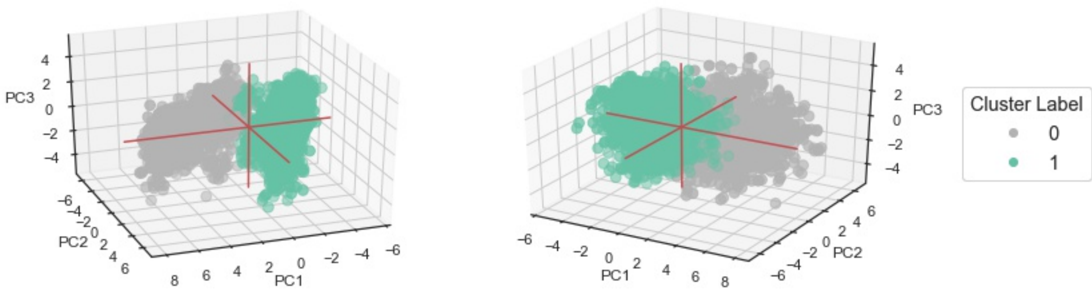


Figure 7: Three-dimension representation of the cloud of datapoints for each of the 2 clusters for analysis B.

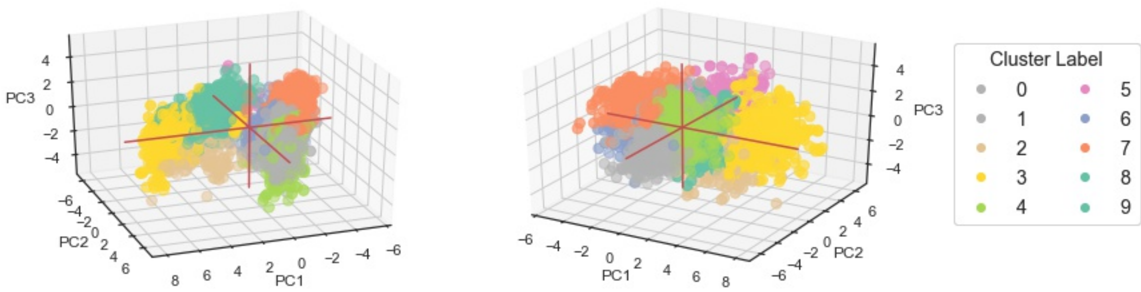


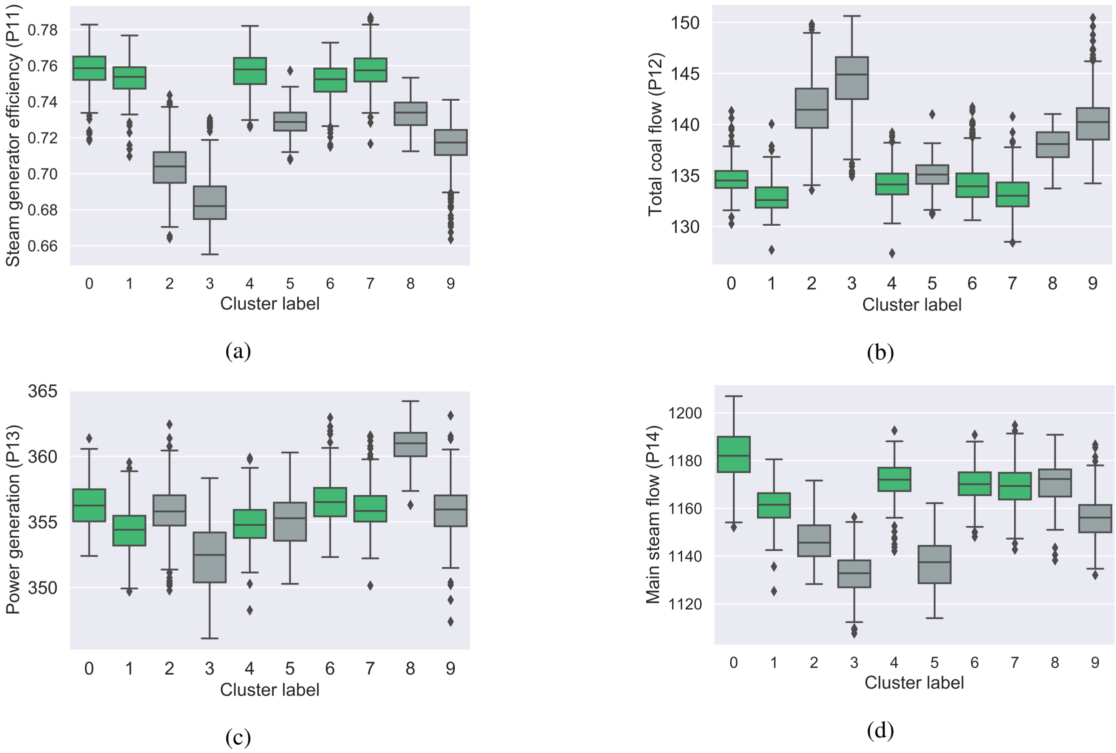
Figure 8: Three-dimension representation of the cloud of datapoints for each of the 10 clusters for analysis B.

Its important to remark that, comparing these figures, the 10 different clusters from the Fig. 8 are subdivisions from the two clusters from Fig. 7. Table 3 identifies these subdivision relations.

Table 3: Subdivision from the 2 cluster analysis to the 10 cluster analysis.

2 clusters	Subdivision in 10 clusters
0	2, 3, 5, 8 and 9
1	0, 1, 4, 6 and 7

Accordingly, the results assessment is performed, examining the clustering results for the 29 considered operation parameters. For each of the 29 parameters, the 10 clusters boxplots are visually inspected. The observations for a selection of the parameters is presented at Fig. 9, for an analysis of the cluster's different operation performances. The boxplots colors represent the correspondancy to the 2 clusters as presented in Tab. 3, as the clusters which are subdivisions of cluster 0 being represented in grey and the clusters which are subdivisions of cluster 1 being represented in green color.



(a) (b) (c) (d)

Figure 9: Boxplots for the 10 clusters: (a) steam generator efficiency (P11), (b) total coal flow (P12), (c) power generation (P13) and (d) main steam flow (P14).

The presented parameters are the steam generator efficiency (P11), power generation (P13), total coal flow (P12) and main steam flow (P14). Cluster 3 presents the lower steam generator efficiency (P11), which is coherent due to its high total coal flow consumption (P12) and low power generation (P13). Clusters 2 and 9 are also characterized by lower efficiency, while clusters 0, 4 and 7 present the highest steam-generator efficiency values (P11). The main steam flow (P14) curve is similar to the steam generator efficiency (P11) one, which is coherent since the main steam flow transports energy from the steam generator to the turbines and is definitive to the energy convertion.

Since these clustering results have promoted an adequate analysis of the studied power plant operation, at the methodology when questioning whether the results are useful and physically relevant, these results are judged as useful. The 10 clusters may be understood as the encountered operation patterns and its analysis is developed further.

6. RESULTS AND DISCUSSION

The time the power plant has operated at each pattern (cluster) over the considered period is obtained. Table 4 presents the percentage of the time operated at each of the 10 operation patterns from the encountered clusters.

Table 4: Percentage of the time operated at each of the 10 encountered clusters.

Cluster	0	1	2	3	4	5	6	7	8	9
Occurance	8.41%	12.77%	7.88%	10.87%	7.84%	3.80%	9.78%	19.59%	2.67%	16.39%

It is possible to observe that there are patterns at which the power plant operates for short periods of time, by up to 10% of the considered condition, and others that operate for relatively longer periods, such as cluster 7. Additionally, observing its boxplot at Fig. 9a , the operation pattern for cluster 7 is one with the highest steam-generator efficiency average. The operation ranges for each analysed parameter at the condition for the pattern from cluster 7 is obtained, considering the first and third quartiles as the upper and lower range limits. Thus, tab. 5 presents the range limits for all considered 29 parameters for this cluster, whose parameters details are presented at Appendix A.

Table 5: Lower and upper limit for each parameter operation ranges for cluster 7 pattern.

	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**
Lower	102.20	66.62	0.80	1.96	15.94	75.89	2.00	344.50	13.82	32.95
Upper	105.39	69.54	0.80	2.37	17.30	76.42	3.00	349.59	14.44	33.56
	P11	**P12**	**P13**	**P14**	**P15**	**P16**	**P17**	**P18**	**P23**	**P24**
Lower	0.75	131.97	355.03	1163.71	167.33	536.57	36.03	325.59	1129.30	198.12
Upper	0.76	134.31	356.98	1174.90	167.82	538.48	36.30	327.73	1144.77	198.93
	P25	**P28**	**P29**	**P35**	**P36**	**P37**	**P38**	**P39**	**P40**	
Lower	270.41	31.93	537.30	357.98	76.32	293.28	22.64	341.85	333.00	
Upper	271.97	32.19	541.85	358.31	78.64	298.55	23.86	348.99	337.30	

Conditions with higher steam generator efficiency are desirable for a profitable operation. The operational ranges described in Tab. 5 provide the preferable arrangements to assure the higher steam generator efficiencies, supporting practices to improve plant operation. These conditions were already attained for 19.59% of the analysed period. Further analysis should evaluate the reasons why this condition is not kept and support practices to improve the power plant operation. The power plant steam generator efficiency has a potential of establishing in increased values if such conditions are targeted.

7. CONCLUSIONS

This paper applies unsupervised machine learning methods to recognize different operation patterns at a 360 MW thermal power plant based on one year historical data. The methodology based at the principal component analysis and K-means clustering methods was implemented to the 40 initial selected operation parameters. The analysis identified that the impact of 11 initial parameters on the clustering phase were restraining the interpretation of the results. The methodology is then implemented to a reduced subset of 29 parameters, whose conclusions were more fitted to the power plant operation. It was identified a partition in 2 clusters and a subdivision in 10 clusters. The 2 clusters identified general differences in the operation, while the analysis of the 10 different clusters exposes the patterns that alternate along the operation. It reveals the behavior of the analysed parameters and enables the identification of the conditions for a high steam generator efficiency operation. The defined operation ranges for each parameter for such condition is obtained. The understanding of these configurations and its resulting performance on the plant may support practices to improve its operation and to have an efficient decision making in the plant.

8. ACKNOWLEDGEMENTS

Authors acknowledge Energy of Portugal EDP for the financial and technical support to this project; Duarte acknowledges the financial support from CNPq 154147/2020-6 for her undergraduate scholarship; Vieira acknowledges the INCT-GD and the financial support from CAPES 23038.000776/2017-54 for her PhD grant; Marques acknowledges the financial support from CNPq 132422/2020-4 for his MSc grant; Smith Schneider acknowledges CNPq for his research grant (PQ 305357/2013-1).

9. RESPONSIBILITY NOTICE

The authors are solely responsible for the printed material included in this paper.

10. REFERENCES

Aggarwal, C.C., 2015. *Data Mining*. Springer International Publishing. doi:10.1007/978-3-319-14142-8. URL `https://doi.org/10.1007/978-3-319-14142-8`.

Hou, G., Gong, L., Huang, C. and Zhang, J., 2019. "Novel fuzzy modeling and energy-saving predictive control of coordinated control system in 1000 MW ultra-supercritical unit". *ISA Transactions*, Vol. 86, pp. 48–61. doi: 10.1016/j.isatra.2018.10.042. URL `https://doi.org/10.1016/j.isatra.2018.10.042`.

IEA, 2018. *World Energy Outlook 2018 Executive Summary*. International Energy Agency.

Igual, L. and Seguí, S., 2017. *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applicaitons*. Springer, Barcelona.

Kubat, M., 2017. *An Introduction to Machine Learning*. Springer International Publishing. doi:10.1007/978-3-319-63913-0. URL `https://doi.org/10.1007/978-3-319-63913-0`.

Kusiak, A. and Song, Z., 2006. "Combustion efficiency optimization and virtual testing: A data-mining approach". *IEEE Transactions on Industrial Informatics*, Vol. 2, No. 3, pp. 176–184. doi:10.1109/tii.2006.873598. URL https://doi.org/10.1109/tii.2006.873598.

Liu, S., Sun, L., Zhu, S., Li, J., Chen, X. and Zhong, W., 2020. "Operation strategy optimization of desulfurization system based on data mining". *Applied Mathematical Modelling*, Vol. 81, pp. 144–158. doi:10.1016/j.apm.2019.12.004. URL https://doi.org/10.1016/j.apm.2019.12.004.

Song, Z. and Kusiak, A., 2007. "Constraint-based control of boiler efficiency: A data-mining approach". *IEEE Transactions on Industrial Informatics*, Vol. 3, No. 1, pp. 73–83. doi:10.1109/tii.2006.890530. URL https://doi.org/10.1109/tii.2006.890530.

Wang, H. and Jia, L., 2019. "Big data knowledge mining based operation parameters optimization of thermal power". In *2019 IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS)*. IEEE. doi:10.1109/ddcls.2019.8908932. URL https://doi.org/10.1109/ddcls.2019.8908932.

Xiaoying, H., Jingcheng, W., Langwen, Z. and Bohui, W., 2016. "Data-driven modelling and fuzzy multiple-model predictive control of oxygen content in coal-fired power plant". *Transactions of the Institute of Measurement and Control*, Vol. 39, No. 11, pp. 1631–1642. doi:10.1177/0142331216644498. URL https://doi.org/10.1177/0142331216644498.

A APPENDIX A

The extensive operation parameter list presents the 40 selected parameter names and measure unit. The Analysis B column indicate the 11 disconsidered parameters. The last two columns indicate the lower and upper limits for the operating range for cluster 7.

Name	Operation Parameter	Unit	Analysis B	Lower - cluster 7	Upper - cluster 7
P1	Mill A dynamic classifier speed	rpm		102,20	105,39
P2	Secondary air flow	kg/s		66,62	69,54
P3	Stoichiometric ratio	-		0,80	0,80
P4	Average O2 excess	%		1,96	2,37
P5	Secondary air collector pressure	mbar		15,94	17,30
P6	Primary air collector pressure	mbar		75,89	76,42
P7	Average CO furnace output	ppm		2,00	3,00
P8	Average furnace combustion gas temperature	°C		344,50	349,59
P9	Intern consumption A	MW		13,82	14,44
P10	Intern consumption B	MW		32,95	33,56
P11	Steam generator efficiency	-		0,75	0,76
P12	Total coal flow	t/h		131,97	134,31
P13	Power generation	MW		355,03	356,98
P14	Main steam flow	t/h		1163,71	1174,90
P15	Main steam pressure	barg		167,33	167,82
P16	Main steam temperature	°C		536,57	538,48
P17	Steam to be reheated pressure	bar_a		36,03	36,30
P18	Steam to be reheated temperature	°C		325,59	327,73
P19	Steam pressure for steam exchanger HE*8	barg	excl.		
P20	Steam temperature for HE8	°C	excl.		
P21	Condensate pressure from HE8 to HE7	barg	excl.		
P22	Condensate temperature from HE8 to HE7	°C	excl.		
P23	Feedwater flow	t/h		1129,30	1144,77
P24	Feedwater pressure	barg		198,12	198,93
P25	Feedwater temperature	°C		270,41	271,97
P26	HE8 inlet water pressure	barg	excl.		
P27	HE8 inlet water temperature	°C	excl.		
P28	Hot reheated steam pressure	barg		31,93	32,19
P29	Hot reheated steam temperature	°C		537,30	541,85
P30	Steam pressure for HE7	barg	excl.		
P31	Steam temperature for HE7	°C	excl.		
P32	Condensate pressure from HE7 to HE6	barg	excl.		
P33	Condensate temperature HE7 to HE6	°C	excl.		
P34	Water temperature from HE6	°C	excl.		
P35	Drum water temperature	°C		357,98	358,31
P36	Average coal temperature	°C		76,32	78,64
P37	Average mill air temperature	°C		293,28	298,55
P38	Average mill air flow	kg/s		22,64	23,86
P39	Total of primary and secondary air flow	kg/s		341,85	348,99
P40	Average heated air temperature	°C		333,00	337,30

∗ short for Heat Exchanger