Applying an Unsupervised Machine Learning Method for Defining Maintenance Significant Items

Miguel Angelo de Carvalho Michalski

Department of Mechatronics and Mechanical Systems, University of São Paulo, Avenida Professor Mello de Moraes 2231, São Paulo, SP, 05508-030, Brazil. E-mail: michalski@usp.br

Arthur Henrique de Andrade Melani

Department of Mechatronics and Mechanical Systems, University of São Paulo, Avenida Professor Mello de Moraes 2231, São Paulo, SP, 05508-030, Brazil. E-mail: melani@usp.br

Renan Favarão da Silva

Department of Mechatronics and Mechanical Systems, University of São Paulo, Avenida Professor Mello de Moraes 2231, São Paulo, SP, 05508-030, Brazil. E-mail: renanfavarao@usp.br

Gilberto Francisco Martha de Souza

Department of Mechatronics and Mechanical Systems, University of São Paulo, Avenida Professor Mello de Moraes 2231, São Paulo, SP, 05508-030, Brazil. E-mail: gfmsouza@usp.br

The selection of Maintenance Significant Items (MSI) is a very important phase in the implementation of Reliability-Centered Maintenance (RCM) in any organization, being essentially a screening phase in which the number of items for analysis can be reduced and prioritized. Despite its importance, there are currently few studies that present systematic and structured methods for MSI identification. There are two phases to identify these items in a physical asset portfolio: firstly, based on the system study and analysis, the criteria and scales are established; secondly, the criteria evaluation for each item is performed. In the latter phase, generally, a Multicriteria Decision-Making (MCDM) method is used to rank the MSI. However, their intrinsic subjective evaluation can lead to bias results. To prevent this issue and simplify the process, this work proposes the use of an unsupervised method based on Principal Component Analysis (PCA). With this method application, not only the MSI of a system are defined, but also the importance of the criteria selected in the first phase is assessed based on the variability of the scores associated with each item. To demonstrate the method, it is implemented considering data from a hydroelectric power plant, and the results are compared to those obtained from a more traditional approach. It is noted that the proposed method points to a robust MSI selection, consistent with the analyzed system.

Keywords: Maintenance significant items, MSI, machine learning, principal component analysis, PCA, multicriteria analysis, multicriteria decision-making method, MCDM, reliability-centered maintenance, RCM, hydroelectric power plant.

1. Introduction

At the dawn of the 21st Century, the new organizational structure of maintenance started to focus on increasing the availability, reliability, and maintainability of equipment and systems. Ten years later, in the 2010s, based on more sophisticated computational tools, systems integration, and interdisciplinary work, asset management principles began to be included in maintenance practices focusing on monitoring, reliability, ease of maintenance, risk analysis, systems integration, and interdisciplinary work. Kardec and Nascif (2019)

The good practices introduced from the inclusion of such principles made it clear that it had become economically advantageous to extract the maximum from the assets already installed in industrial parks around the world, in a way associated with the need to achieve sustainability and environmental preservation goals. Therefore, reliability has become an even more relevant topic and the reduction in the number of shutdowns, leading to a longer interval between scheduled maintenances, has become a necessity in practically all types of industries.

Given such needs, it is natural that Reliability-Centered Maintenance (RCM) became the most logical strategy to be adopted by maintainers and maintenance engineers. Rausand (1998)

Being a methodology focused on the system's functions, according to Deepak Prabhakar and Jagathy Raj (2014), RCM can be understood as the process used to determine what must be done to ensure that any component of a system continues to perform the functions for which it was designed, in the existing operating conditions and circumstances. Since this methodology allows the elaboration of maintenance plans with the best techniques and the definition of the best type of maintenance for each situation, it leads to the maximization of the system's availability and the minimization of production losses.

Considered one of the most important phases in the implementation of RCM, the selection of Maintenance Significant Items (MSI) is essentially a screening phase in

Proceedings of the 31st European Safety and Reliability Conference
Edited by Bruno Castanier, Marko Cepin, David Bigaud, and Christophe Berenguer

ISBN: 978-981-18-2016-8; doi:10.3850/978-981-18-2016-8_345-cd

which the number of items for analysis can be reduced. Precisely because it defines a control volume for the application of Prognosis and Health Management (PHM), reducing the number of components to be evaluated, the number of studies that address this subject in a structured, systematic, and convenient operational manner has been increasing. Some recent examples found in the literature are Caminada Netto et al. (2020), Santos et al. (2019), Silva et al. (2019), Yuan et al. (2019), Gupta (2018), Melani et al. (2018), and Tang et. al. (2017).

According to Caminada Netto et al. (2020), some relevant aspects can be implied from these works, as follows:

- There are basically two fundamental phases to identify the MSI in a systematic and structured way: the system study and analysis, where the criteria and scales are established for further evaluation; and the criticality evaluation, where the ranking of the system's MSI is defined;
- Generally, a Multicriteria Decision-Making (MCDM) method is used to rank the most critical items, with Analytical Hierarchy Process (AHP) being one of the most used among these methods.

It is important to note, however, that although decision-making processes concerning maintenance planning have become increasingly critical, including the selection of a system's MSIs, there is still a great deal of epistemic uncertainty inherent to many methods that depend on the knowledge of experts in any of its stages. Bellinello et al. (2020)

Regarding MSI selection, such uncertainty can be present in both phases of the process. The choice of criteria, their importance in the analysis, and the applied scales depend on the knowledge acquired by experts regarding the system. Depending on these choices, the results obtained can be quite biased.

Different approaches can assist in reducing the uncertainties in the process of defining MSI. For example, Silva et al. (2019) had proposed a new framework for the determination of MSI based on the aspects of ISO 55000. Although the standard does not reveal exactly which criteria should be selected, its requirements guide the alignment of physical assets to the organizational objectives. In this work, based on four fundamental critical aspects in asset management, such as safety, environmental impact, performance, and compliance with regulations, nine criteria were chosen to select the MSI of a system. Although the results of this work demonstrate that the proposed framework allows the MSI selection for maintenance planning according to the organizational objectives, the relative importance of the criteria among each other remains open to question.

Thus, the present work proposes the use of an unsupervised method based on Principal Component Analysis (PCA) for MSI selection and ranking. From the application of this method, not only are the MSI of a system defined but the importance of the criteria selected in the first phase is assessed based on the variability of the scores associated with each item.

This new method determines the weight of each criterion previously considered based on the variability of the scores associated with the analyzed components, replacing the use of MCDM methods in the classification process of the most critical items. The objective is to reduce the epistemic uncertainty and possible biases in the MSI selection process since the importance of each criterion will not be attributed subjectively or based on expert knowledge.

The remaining parts of this paper are organized as follows: Section 2 presents a brief discussion regarding PCA fundamentals and its application in the current scenario; Section 3 describes the proposed method; Section 4 presents the application and results of the proposed method to a hydropower plant case; and, finally, Section 5 presents the results discussion and obtained conclusions.

2. Principal Component Factor Analysis

Principal Component Analysis (PCA) is a very useful unsupervised machine learning statistical technique that has found application in many fields, especially the ones where finding patterns in multivariate data are required. PCA can be considered the simplest of the eigenvector-based multivariate analyses and can be used to reveal the internal structure of a dataset in a way that best explains its variation. Sarkar et al. (2014)

Basically, PCA works as follows: for a given dataset of k parameters and n observations, PCA produces a set of principal factors that are, at the same time, a linear combination of the initial parameters and an orthogonal set, i.e., factors are uncorrelated.

These basics assumptions make PCA the most widely used technique in Factor Analysis (FA) among the methods used to determine factors from datasets. The analysis consists of a multivariate technique that seeks to identify a small number of factors that represent the joint behavior of a group of originally interdependent variables. Accordingly, while cluster analysis uses distance or similarity measures to group observations and to build clusters, FA uses correlation coefficients to group variables and generate factors. Fávero (2019)

Mathematically, the extraction of factors from the PCA algorithm, known as Principal Component Factor Analysis (PCFA), occurs as follows: considering the previously mentioned dataset, **X**, with k parameters and n observations, given by Eq. 1:

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix} \qquad (1)$$

the mean of each parameter subset must be calculated as presented in Eq. 2:

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \qquad (2)$$

Where $1 \leq i \leq n$ and $1 \leq j \leq k$.

Next, for each parameter subset, the parameter's observations must be subtracted from its average, previously obtained in Eq. 2, leading to a new re-centered matrix **B**, given by Eq. 3:

$$\mathbf{B} = \begin{bmatrix} x_{11} - \overline{x}_1 & \cdots & x_{1k} - \overline{x}_k \\ \vdots & \ddots & \vdots \\ x_{n1} - \overline{x}_1 & \cdots & x_{nk} - \overline{x}_k \end{bmatrix} \tag{3}$$

Then, the covariant matrix **S** is obtained from matrix **B**, as given by Eq. 4:

$$\mathbf{S} = \frac{1}{n-1} \mathbf{B}^{\mathrm{T}} \mathbf{B} \tag{4}$$

From the eigendecomposition of the covariant matrix **S**, the loading matrix, **V**, can be obtained, as presented in Eq. 5:

$$\mathbf{S} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^{-1} \tag{5}$$

Where **V** is the square k x k matrix whose j-th column is the eigenvector $\mathbf{v}_j$ of **V** and $\mathbf{\Lambda}$ is the diagonal matrix in which the non-zero elements are the corresponding eigenvalues, ordered highest to lowest ($\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ... \geq \lambda_k$), given by Eq. 6.

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_k \end{bmatrix} \tag{6}$$

The obtained eigenvectors (the principal components) must be ordered by eigenvalues, creating a new data set matrix arranged so that the first vector corresponds to the highest eigenvalue and the last vector corresponds to the lowest eigenvalue. In this way, the principal components are given in order of significance.

Also, the eigenvalues and eigenvectors' elements are respectively normalized considering, respectively, Eq. 7 and Eq. 8, as follows:

$$\lambda'_j = \frac{\lambda_j}{\sum_{j=1}^{k} \lambda_j} \; ; \quad 1 \leq j \leq k \tag{7}$$

$$v'_{ij} = \frac{v_{ij}}{\sum_{i=1}^{n} v_{ij}} \; ; \quad 1 \leq j \leq k \tag{8}$$

The normalized eigenvalues are the factors' loads, proportional to the shared variance of each principal component. On the other hand, the eigenvectors' normalized elements are the factors' scores that correspond to the importance of each original variable in each principal component.

Considering the factor loads, a Cumulative Percentage of Total Variation (CPV) can be obtained thru Eq. (9):

$$\mathrm{CPV} = 100 \cdot \sum_{j=1}^{m} \lambda'_j \tag{9}$$

Where m, in this case, is the maximum number of factor loads necessary for the CPV to be greater than or equal to a pre-established value, which means the percentage of the variance of the original dataset carried by the new dataset. Jolliffe (2002)

Then, the original parameters' weights can be obtained from Eq. 10.

$$\mathbf{p} = \begin{bmatrix} v'_{11} & \cdots & v'_{1m} \\ \vdots & \ddots & \vdots \\ v'_{k1} & \cdots & v'_{km} \end{bmatrix} \begin{bmatrix} \lambda'_1 \\ \vdots \\ \lambda'_m \end{bmatrix} \tag{10}$$

Where **p** is the original parameters' weights vector, with dimension k, whose elements correspond to the weight of each original parameter, p_j ($1 \leq j \leq k$), in the shared variance considering all factors.

Note that the matrix with the normalized eigenvector value has dimensions k x m while the vector with the normalized eigenvalues has dimension m. Thus, although the vector **p** has dimension k, i.e., the number of original parameters, it was obtained considering the number of factors loads, m, necessary to obtain the desired CPV. In other words, the **p** vector already contemplates the dimension reduction obtained with the PCA application.

The value of each p_j must be normalized by the sum of the elements of the **p** vector, giving rise to the normalized vector **p**', so that finally the observations' scores, s_i ($1 \leq i \leq n$) can be obtained considering Eq. 11:

$$\mathbf{s} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} p'_1 \\ \vdots \\ p'_k \end{bmatrix} \tag{11}$$

Where **s** is the vector of observations' scores which, when ordered, allows the ranking of observations in terms of relevance to the total variation of the original dataset.

3. Proposed Method

In this section, the method proposed in this article will be presented. The method aims at ranking the MSI of a system using an unsupervised PCA-based approach. However, for the method to be applied, two premises must be satisfied: criteria and scales have been previously established for further evaluation of the analyzed system's components, and a preliminary study of the system have been prepared to indicate a list of components of interest for the analysis of MSI.

Besides, the preliminary study may have the objective of helping maintainers and specialists to better understand the system and define the scores of each component as a function of the established criteria and scales. This application of the preliminary study is especially necessary when the grades of the criteria are given subjectively, that is, not based on measurements or metric variables.

Once the premises are satisfied, a dataset table can be assembled, as presented in Table 1, in which the lines refer to the analyzed piece of equipment (P_i, $1 \le i \le n$) and the columns to the evaluated criteria (C_j, $1 \le j \le k$). This table is populated with the grades (G_{ij}) given for each item according to each criterion.

Table 1. Criteria evaluation template table.

	Criteria				
Piece of Equipment	$\mathbf{C_1}$	$\mathbf{C_2}$	$\mathbf{C_3}$	**...**	$\mathbf{C_k}$
$\mathbf{P_1}$	G_{11}	G_{12}	G_{13}	...	G_{1k}
$\mathbf{P_2}$	G_{21}	G_{22}	G_{23}	...	G_{2k}
$\mathbf{P_3}$	G_{31}	G_{32}	G_{33}	...	G_{3k}
...	...	...	...	...	...
$\mathbf{P_n}$	G_{n1}	G_{n2}	G_{n3}	...	G_{nk}

The criteria evaluation table assembling is the first step (I) of the proposed method. The next steps of the method are as shown in Fig. 1.

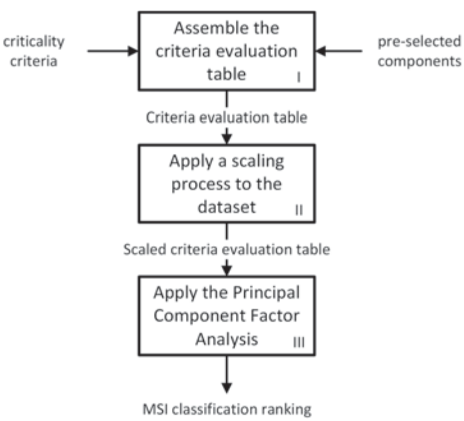

Fig. 1. Proposed method framework

As can be seen from the method framework, the second step (II) is to apply a scaling process to the data set. According to Mujica et al. (2010), since different criteria, whether measurable variables or subjective criteria, have different magnitudes and scales, it is important to apply a scaling process to the original dataset before going further with any analysis.

The choice of the scaling criterion will depend on the application and the nature of the data analyzed. Then, the third step (III) of the method is to apply the PCFA as presented in Section 2.

4. Case Study and Results

The case study presented in this work can be considered as an outcome of the study presented by Silva et al. (2019). The authors considered in the current work the same system and database from this previous study so that the results of the proposed method could be compared with the results of a more traditional MSI selection approach.

Considering a hydroelectric power plant with an installed capacity of 198 MW composed of 4 generating units with Kaplan turbines (GU1 to GU4), each generating unit is divided into eight sub-systems, as presented in Table 2. Silva et al. (2019)

Table 2. Generating units' sub-systems.

ID	Sub-System
x.1.	Adduction
x.2.	Turbine
x.3.	Electrical connection
x.4.	Generator
x.5.	Shaft
x.6.	Excitation system
x.7.	Speed governor
x.8.	Draft tube

Nine evaluation criteria were considered, being two criteria associated with safety aspects: safety classification (s1) and history of safety events (s2); two criteria associated with environmental aspects: environmental classification (e1) and history of environmental events (e2); four criteria related to performance aspects: reliability (p1), maintainability (p2), health assessment capacity (p3), and maintenance compliance (p4); and one criterion related to the regulatory aspect: impact on availability (r1).

Based on the selection of the sub-systems and the evaluation criteria, a numerical classification based on a scale from 1 to 9 for each criterion was developed. The scores considered in this classification were normalized in such a way that the sum of all observations (grade given to each component or sub-system) for each criterion was equal to 1. That is, the 214 scores for each criterion were added and divided by the total value obtained from this sum.

Table 3 presents an example of the given scores considering the speed governor sub-system of the Generating Unit 1 (GU1) evaluated in the case study. The components considered in this case, according to their ID, are the following: Gate mechanism (1.7.1), Kaplan head (1.7.2.1), Head bushing (1.7.2.2), Kaplan mechanism (1.7.2.3), Electronic regulator (1.7.3.1), Distributor control (1.7.3.2), Turbine control (1.7.3.3), Oil cooling system (1.7.4.1), and Oil pressurization system (1.7.4.2). Silva et al. (2019)

Table 3. Scores for the speed regulating sub-system of the GU1.

Item ID	Criteria								
	s1	s2	e1	e2	p1	p2	p3	p4	r1
1.7.1	3	1	1	1	1	7	9	1	7
1.7.2.1	3	1	1	1	1	5	9	1	7
1.7.2.2	3	1	1	1	1	5	9	1	7
1.7.2.3	7	1	5	1	1	7	9	9	7
1.7.3.1	2	1	1	1	1	5	7	1	9
1.7.3.2	2	1	1	1	1	5	7	1	9
1.7.3.3	2	1	1	1	1	5	7	1	9
1.7.4.1	3	1	2	1	1	3	9	1	7
1.7.4.2	3	1	2	1	2	3	9	1	7

Note that Table 3 presents just a small sample of the method's complete input table, resulting from the first stage of the framework presented in Fig. 1.

The second stage of the proposed method is the application of a scheduling process to the criteria evaluation scores. In this work, the min-max normalization was chosen. Considering a vector of values **x**, the vector of normalized values **x'** is given by Eq. 12, with min(**x**) being the minimum value of the vector **x** and max(**x**) the maximum value of the same vector.

$$\mathbf{x}' = \frac{\mathbf{x} - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})} \tag{12}$$

Thus, for each criterion, this normalization approach is applied. If the denominator of Eq. 12 is null, then $x_i = 0.5$ for $1 \leq i \leq n$. Table 4 presents the normalized criteria scores for the same components presented in Table 3.

Table 4. Normalized scores criteria for the speed regulating sub-system of the GU1.

Item ID	Normalized Scores								
	s1	s2	e1	e2	p1	p2	p3	p4	r1
1.7.1	0.14	0.50	0.00	0.50	0.00	0.75	1.00	0.00	0.75
1.7.2.1	0.14	0.50	0.00	0.50	0.00	0.50	1.00	0.00	0.75
1.7.2.2	0.14	0.50	0.00	0.50	0.00	0.50	1.00	0.00	0.75
1.7.2.3	0.71	0.50	1.00	0.50	0.00	0.75	1.00	1.00	0.75
1.7.3.1	0.00	0.50	0.00	0.50	0.00	0.50	0.71	0.00	1.00
1.7.3.2	0.00	0.50	0.00	0.50	0.00	0.50	0.71	0.00	1.00
1.7.3.3	0.00	0.50	0.00	0.50	0.00	0.50	0.71	0.00	1.00
1.7.4.1	0.14	0.50	0.25	0.50	0.00	0.25	1.00	0.00	0.75
1.7.4.2	0.14	0.50	0.25	0.50	0.20	0.25	1.00	0.00	0.75

Next, the PCA is applied to the normalized scores values table. The obtained original parameters' weights are presented in Table 5, considering a CPV of 95%.

Table 5. Criteria's normalized weights.

s1	s2	e1	e2	p1	p2	p3	p4	r1
0.156	0.000	0.250	0.000	0.002	0.179	0.065	0.084	0.264

Note that the weight of the criteria 'history of safety events' (s2) and 'history of environmental events' (e2) are equal to zero, i.e., these criteria have no influence on the result of the MSI selection. Also, as the criterion 'reliability' (p1) has a much lower value than those of the other criteria with non-zero values, it will have a very small influence on the result.

The main reason for the negligible influence of such criteria is the lack of historical data for the analyzed system. Without this historical information, the grades attributed to such criteria end up having a very low variability, holding practically no influence on the final result.

For comparison purposes only, considering the work of Silva et al. (2019) in which the AHP is used to determine the MSI with expert judgment to obtain priority scales, the weight of each criterion is given by a priority vector, shown in Table 6.

Table 6. AHP priority vector. Silva et al. (2019)

s1	s2	e1	e2	p1	p2	p3	p4	r1
0.359	0.146	0.161	0.063	0.053	0.053	0.018	0.030	0.116

It is clear that the weights of the criteria considering the two approaches are quite different, with Pearson's correlation coefficient equals 0.33 in this case.

Now, considering Eq. 11 and being the matrix **X** the normalized scores and **p** the criteria's normalized weights, the given vector **s** provides the scores of all the 214 components considered in this work. Table 7 presents this result for the GU1 speed governor sub-system, as an example.

Table 7. GU1 speed governor sub-system ranking.

ID	Score	Accumulated Score	Rank
1.7.1.	0.003535	0.73171168	103
1.7.2.1.	0.0030668	0.82713469	133
1.7.2.2.	0.0030668	0.83020149	134
1.7.2.3.	0.0330454	0.0661164	2
1.7.3.1.	0.0030814	0.79633547	123
1.7.3.2.	0.0030814	0.79941685	124
1.7.3.3.	0.0030814	0.80249823	125
1.7.4.1.	0.0044364	0.62356876	75
1.7.4.2.	0.004462	0.58360597	66

In this case, the score value of each item was normalized by the sum of the scores of all items, so that the sum of the normalized scores is equal to 1. Also, the accumulated score column represents the value of the sum of the scores of all items up to the component in question, giving an idea of the position of this component in the general classification table, i.e., the lower the accumulated score, the more critical the item is and vice versa.

Sorting the scores from the highest to the lowest, the MSI are presented in decreasing order of criticality, as shown in Table 8 for the first and last 24 components.

Still, considering as a reference the results of Silva et al. (2019), a column with the classification of the same 48 components considered, obtained with the AHP method, was added to Table 8. In this way, it is possible to compare the Rank PCA with the Rank AHP and notice how in this case study there is some adherence between the results of the two approaches.

Table 8. MSI ranking for the first 24 components.

ID	Component	Rank PCA	Rank AHP
4.7.2.3.	Kaplan Mechanism	1	1
1.7.2.3.	Kaplan Mechanism	2	2
2.7.2.3.	Kaplan Mechanism	3	3
3.7.2.3.	Kaplan Mechanism	4	4
1.2.4.	Runner	5	9
2.2.4.	Runner	6	10
3.2.4.	Runner	7	11
4.2.4.	Runner	8	12
1.1.2.	Adduction grids	9	5
2.1.2.	Adduction grids	10	6
3.1.2.	Adduction grids	11	7
4.1.2.	Adduction grids	12	8
1.2.7.	Turbine guide bearing	13	21
3.2.7.	Turbine guide bearing	14	22
2.2.7.	Turbine guide bearing	15	27
4.2.7.	Turbine guide bearing	16	28
3.2.3.	Wicket gate	17	13
1.2.2.	Stay vane	18	17
2.2.2.	Stay vane	19	18
3.2.2.	Stay vane	20	19
4.2.2.	Stay vane	21	20
1.2.3.	Wicket gate	22	14
2.2.3.	Wicket gate	23	15
4.2.3.	Wicket gate	24	16
…	…	…	…
4.5.3.	Cyclone filter	191	187
1.2.1.3.	Drain valve	192	188
2.2.1.3.	Drain valve	193	189
3.2.1.3.	Drain valve	194	190
4.2.1.3.	Drain valve	195	191
1.8.2.	Drain valve	196	207
2.8.2.	Drain valve	197	208
3.8.2.	Drain valve	198	209
4.8.2.	Drain valve	199	210
1.2.1.1.	Filling valve 1	200	196
1.2.1.2.	Filling valve 2	201	197
2.2.1.1.	Filling valve 1	202	198
2.2.1.2.	Filling valve 2	203	199
3.2.1.1.	Filling valve 1	204	200
3.2.1.2.	Filling valve 2	205	201
4.2.1.1.	Filling valve 1	206	202
4.2.1.2.	Filling valve 2	207	203
1.3.1.	Auxiliary service transformer	208	204
3.3.1.	Auxiliary service transformer	209	205
4.3.1.	Auxiliary service transformer	210	206
1.8.1.	Filling valve	211	211
2.8.1.	Filling valve	212	212
3.8.1.	Filling valve	213	213
4.8.1.	Filling valve	214	214

In fact, it is noted that such adherence is greater at the beginning and at the end of the table, in which there is practically no difference between the classification obtained through the two approaches. Considering the 214 analyzed components, in 12.15% of the cases, there is total adherence between the methods and more than 30% of the analyzed components have a difference between the position of the PCA ranking and the AHP ranking less than or equal to 5.

The consistency of these results can also be attributed to the fact that the team that developed the AHP analysis is experienced in the subject, which suggests a relatively small epistemic uncertainty in the values attributed to the parameter weights. It is to be expected that the results of the proposed method would have a greater divergence with the results of the AHP-based method if the team that performed the analysis were less experienced. In other words, hypothetically, it is expected that the divergence between the results of the PCA-based method and that of MCDM methods will increase due to the uncertainties of the experts regarding the analyzed system.

Classifying the result of the proposed method of the 214 items in three groups, A, B, and C, according to their criticality based on the following criteria, an ABC curve can be obtained, as shown in Fig. 2.

- A: Items whose cumulative score is less than or equal to 50% must be in group A, i.e., the components with the greatest criticality among the MSI;
- B: Items whose cumulative score is between 50% and 80% must be in group B, i.e., the components with moderate criticality among the MSI;
- C: Items whose cumulative score is greater than 80%, i.e., the less critical components among the MSI.

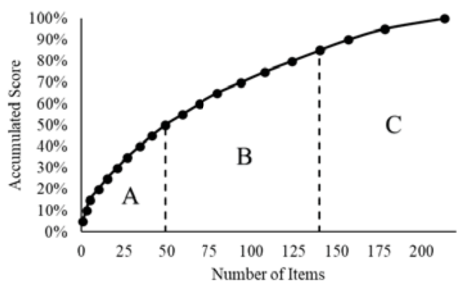

Fig. 2. ABC curve for the MSI classification according to the proposed method.

The vertical dashed lines are used in the graph to separate regions A, B, and C, and to determine the amount of equipment in each region. The first line, in this case, defines the number of items that received an A classification and the second line defines the sum of the items classified as A and B.

Thus, it is noted that 49 items are classified in group A, 91 items in group B, and 74 in group C. In other words, of the analyzed 214 MSI, 49 items (approximately 23%) are considered critical and therefore must be prioritized from the maintenance point of view.

The same classification process was considered for the results by Silva et al. (2019) and, in this way, it is possible to compare the different approaches, as shown in

Fig. 3. In this case, with the application of the AHP to carry out the classification of MSI, 83 items are classified in group A, 86 items in group B, and just 45 in group C. That is, using the AHP according to the proposal by Silva et al. (2019), 34 items are further classified as critical from the maintenance point of view. This difference becomes even clear when observing the flattening of the ABC curve shown in Fig. 3 for the analysis considering the AHP.

It could be said, from the analysis of the ABC curves, that the AHP-based approach generated a more conservative result than the approach proposed in this work, based on the application of the PCA. This is because the approach using AHP classifies a greater number of components, approximately 70% of the total, in regions A and B, while the PCA-based approach classifies 65.5% of the number of items analyzed in the same region. And, specifically for the number of items classified as most critical (region A), this difference is even greater: 39% of the MSI are classified in group A according to the AHP-based method and 23% of the MSI are classified in the same group according to the PCA-based approach.

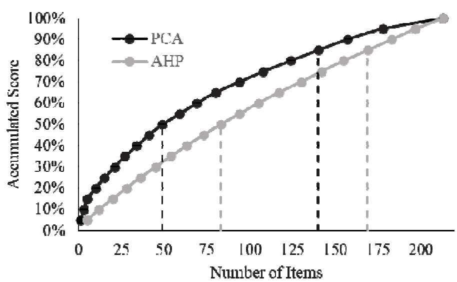

Fig. 3. Comparison of the ABC curves for the MSI classification according to the proposed method and Silva et al. (2019).

5. Conclusions

This work proposed the use of an unsupervised statistical decision-making method in the selection process of MSI of a system. The idea is to reduce the epistemic uncertainties and possible biased results entailed in the classification process of a system MSI with this proposal. A PCA-based technique was applied considering data from a Brazilian hydroelectric power plant and the results obtained were extremely robust and relevant.

The results obtained with the proposed method were compared to the results obtained from a more traditional approach, in which the AHP is used as a method for classifying MSI, as the authors previously presented in Silva et al. (2019).

It is important to emphasize that the purpose of such a comparison is not to point out which method is better or more efficient but to demonstrate the consistency of the approach proposed in this work since the system, its sub-systems and components, as well as the analysis criteria considered in the present work are the same used in Silva et al. (2019). That is, from the same inputs the idea was to observe the changes in the outputs.

Likewise, the results presented in this work are *ad hoc*, i.e., they are related to the current analyzed system. Any change in this system will lead to variations in the responses obtained. However, this issue should not be seen as a weakness of the proposed method, but quite the contrary, it emphasizes that the method ends up being more flexible than an approach based on multicriteria decision-making (MCDM) methods such as AHP.

This is because, if the considered criteria or system changes, a new analysis process must be initiated, as it would happen with any method considered. But the entire phase of evaluation and selection of MSI, which is generally supervised and requires the knowledge of experts, is replaced by an unsupervised process. That is, once the scores of the components considered are given for each criterion analyzed, the process proceeds from there totally autonomous.

Regarding the results obtained for the system considered in this work, it is interesting to note its consistency and how there was a relative adherence between these and the results obtained with the AHP-based method for the same system and the same criteria. The coherence of the proposed method would not need to be validated from such a comparison, since the analysis of the obtained results already demonstrates it. But the comparison with an already traditional method and the similarity found in the responses reinforces that the proposed method can be applied in other systems and be quite useful in defining MSI.

Besides, the MSI ranking from the proposed method can be done free from human interference, as long as the criteria used in the evaluations are metric variables, such as the failure rate, mean time to repair, mean time to fail, reliability, maintainability, number of open work orders, among others. That is, if a scale or expert assessment is not necessary to obtain the values of such criteria. Considering the same scenario in the application of an MCDM method (such as AHP) for the ranking of MSI, even with criteria based on metric variables, such application will depend on the evaluation and weighting of criteria by experts.

As future work it is expected to the method be applied in other systems and, if possible, to compare the results obtained with other methods to give more credibility to the method, as well as demonstrate its full usability.

Acknowledgment

The authors thank the financial support of FDTE (Fundação para o Desenvolvimento Tecnológico da Engenharia), FUSP (Fundação de Apoio à Universidade de São Paulo) and EDP Brasil through an ANEEL R&D project.

References

Bellinello, M. M., M. A. C. Michalski, A. H. A. Melani, A. Caminada Netto, C. A. Murad, and G. F. M. Souza (2020). Decision-Making Consistency in Maintenance Management. *Applied Sciences 10*, 1-26.

Caminada Netto, A., A. H. A. Melani, C. A. Murad, M. A. C. Michalski, G. F. M. Souza, and S. I. Nabeta (2020). A Novel Approach to Defining Maintenance Significant Items: A Hydro Generator Case Study. *Energies 20*, 1-20.

Deepak Prabhakar, P. and V. P. Jagathy Raj (2014). CBM, TPM, RCM and A-RCM – A Qualitative Comparison of Maintenance Management Strategies. *International Journal of Management & Business Studies 4*, *3*, 49-56.

Fávero, L. P. and P. Belfiore (2019). *Data Science for Business and Decision Making*, 1st Edition. Academic Press.

Gupta, G. and R. P. Mishra (2018). Identification of Critical Components using ANP for Implementation of Reliability Centered Maintenance. *Procedia CIRP 69*, 905-909.

Jolliffe, I. T. (2002). *Principal Component Analysis*, 2nd Edition. Springer.

Kardec, A. and J. Nascif (2019). *Manutenção – Função Estratégica*, 5th Edition. QualityMark.

Melani, A. H. A., C. A. Murad, A. Caminada Netto, G. F. M. de Souza, S. I. Nabeta (2018). Criticality-based maintenance of a coal-fired power plant. *Energy 147*, 767-781.

Mujica, L. E., J. Rodellar, A. Fernández, A. Güemes (2010). Q-statistic and T^2-statistic PCA-based measures for damage assessment in structures. *Structural Health Monitoring 10*, 539-553.

Rausand, M. (1998). Reliability Centered Maintenance. *Reliability Engineering and System Safety 60*, 121-132.

Santos, T., F. J. G. Silva, S. F. Ramos, R. D. S. G. Campilho, L. P. Ferreira (2019). Asset Priority Setting for Maintenance Management in the Food Industry. *Procedia Manufacturing 38*, 1623-1633.

Sarkar, J., S. Sahab, S. Agrawalb (2014). An Efficient Use of Principal Component Analysis in Workload Characterization - A Study. *AASRI Procedia 8*, 68-74.

Silva, R. F., A. H. A. Melani, M. A. C. Michalski, G. F. M. Souza, and S. I. Nabeta (2019). Defining maintenance significant items based on ISO 55000 and AHP: A hydropower plant case study. *Proceedings of 29th European Safety and Reliability Conference*, 3437-3445.

Tang, Y., Q. Liu, J. Jing, Y. Yang, Z. Zou (2017). A framework for identification of maintenance significant items in reliability centered maintenance. *Energy 118*, 1295-1303.

Yuan, Y., J. Wang, W. ShangGuan, B. Cai, H. Song (2019). Research on Identification of Maintenance Significant Items in Reliability Centered Maintenance for Train Control System. *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2817-2822.