Applying Cluster Analysis to Support Failure Management Policy Selection in Asset Management: A Hydropower Plant Case Study

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

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

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

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

Maximizing the realization of value from physical assets through asset management is a contemporary approach to support the achievement of organizational goals. As consequence, organizations have placed maintenance as a strategic function to deliver business outcomes. The development of appropriate failure management policies for failure prevention is essential but also a challenge for maintenance planning. In this context, this paper proposes a method to support the failure management policy selection in asset management based on the exploratory cluster analysis technique. The proposed method comprised three sections: acquisition of the physical asset performance data, cluster analysis, and selection of the failure management policies. Criticality aspects supported the definition of the criteria and scales for evaluating the performance of physical assets and the composition of the dataset. Then, the distance measure and agglomeration schedule were defined for the application of the cluster analysis. From the set of formed clusters of physical assets, it was assigned an overall failure management policy to each cluster based on their internal homogeneity assessment and the context of the organization. The proposed method is demonstrated through a case study in a maintenance management context of a Brazilian hydroelectric power plant. The results obtained show that the method can support organizations in the selection of appropriate failure management policies according to determined groups of physical assets.

Keywords: Asset management, Maintenance, Failure management, Cluster analysis, Machine learning, Hydropower plant.

1. Introduction

Exposure to the global economy and international competition constantly require organizations the pursuit of better performance. Maximizing the realization of value from physical assets through asset management is a contemporary approach that supports this achievement.

The evolution of maintenance practices over the years and the recent rise of asset management have placed maintenance in a strategic function in the organizations. According to GFMAN (2016), maintenance is one of the key levers to deliver business outcomes. Nevertheless, at the same time, maintenance planning has become an increasingly difficult task as the complexity of equipment and systems evolves constantly (Silva et al. 2019)). Therefore, organizations are now more interested and dependent on the appropriate failure management policies for failure prevention.

As these decisions take into account different aspects and performance of the physical assets, the decision makers should not rely on simple heuristics. Instead, these decisions are expected to be supported by systematic approaches that incorporate data analysis. According to Fávero and Belfiore (2019), whenever treated and analyzed, data are transformed into information that is recognized and applied to the decision-making process.

Reliability-Centered Maintenance (RCM) has been used to help organizations formulate failure management policies for the physical asset in almost every industrialized country in the world (SAE International (2009)). However, its application is often time-consuming in modern engineering systems and requires prioritization in an extensive physical asset portfolio as it is an individual application. Thus, it is suitable and relevant that organizations investigate how modern techniques such as machine learning can be incorporated in solving maintenance management challenges.

In this context, this paper proposes a method to support the failure management policy selection in asset management based on the cluster analysis. It intends to use

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_344-cd

this exploratory machine learning technique as assistance for this challenging decision-making in maintenance planning. Then, it is demonstrated through a case study application in a hydroelectric power plant.

This paper is organized as follows: Section 2 presents a brief discussion on failure management policy selection in asset management. Section 3 describes the cluster analysis technique. Section 4 presents the proposed method to support the failure management policy selection based on the cluster analysis. Section 5 applies the method to a hydropower plant case. Finally, Section 6 presents the authors' conclusions about the proposed method and case study.

2. Failure Management Policy Selection

Failure management is a coordinated activity of an organization that deals with the recognition, prevention, and reaction to failures (Schneider et al. (2019)). It has a strong interface with maintenance management as its policies contribute not only to repairing the failures but to their prediction and prevention.

By definition, a policy formally expresses the top management's intentions and directions of an organization (ISO (2015)). Accordingly, failure management policies for physical assets can support maintenance planning in the elaboration of maintenance tasks and plans aligned with what is expected by the organization. The main failure management policies are presented in Table 1 based on the standard SAE JA1011 (SAE International (2009)).

Table 1. Failure management policies.

Policy	Maintenance task guideline
On-condition	A periodic or continuous task used to detect a potential failure
Scheduled Discard	A task that entails replacing an item at or before a specific age limit regardless of its condition at the time
Scheduled Restoration	A task that restores the capability of an item at or before a specific age limit, regardless of its condition at the time, to a level that provides the acceptable probability of survival to the end of another specific interval
Failure Finding	A scheduled task used to determine whether a specific hidden failure has occurred
One Time Change	Any action taken to change the physical configuration of an asset or system, such as redesign and modification
Run-to-failure	No task attempts to anticipate or prevent a specific failure mode to occur

On-condition, Scheduled discard, Scheduled restoration, and Failure finding policies are aligned with scheduled maintenance tasks. However, any scheduled task is only worth doing if it reduces, avoids, eliminates, or minimizes the consequences of the failure mode to an extent that justifies the direct and indirect costs of doing the task (SAE International (2011)). On the Other hand, One Time Change and Run-to-failure policies are associated with non-periodic maintenance tasks.

The selection of appropriate failure management policies for the portfolio of physical assets is a challenging activity in organizations. Different failure management policies can be essentially more cost-effective than others. Moreover, physical assets have different failure modes and failure consequences such as safety, environmental, operational, or economic. Therefore, organizations should be supported with methods and tools to make better decisions in maintenance planning.

3. Cluster Analysis

The cluster analysis is a multivariate exploratory technique used to study similar behavior between the observations in a dataset regarding certain metric or binary variables and the possible existence of homogeneous clusters of observation (Fávero and Belfiore (2019)). It belongs to the class of unsupervised machine learning as its techniques are employed to detect or observe structure and patterns in the unlabeled dataset (Martey et al. (2017)).

Its main objective is to allocate observations to a relatively small number of clusters that are internally homogeneous and heterogeneous between themselves. In other words, the observation of a certain group must be relatively similar to one another regarding the variables inserted in the analysis, and significantly different from the observations found in other groups (Fávero and Belfiore (2019)).

The hierarchical approach allows to sort and allocate observations to groups and, subsequently, analyze and decide the ideal number of groups for their application. On the other hand, in the nonhierarchical approach, the number of groups is informed *a priori* as a known input in the analysis, together with the dataset (Fávero and Belfiore (2019)). Therefore, according to the context of the application, the user needs to decide which of these approaches is more suitable for its objectives.

The cluster analysis uses a distance or similarity measure as the reference for the observations to be considered less or much closer (Fávero and Belfiore (2019)). Among the distance measures, there are metrics such as Euclidean distance, Manhattan distance, Minkowski distance, Mahalanobis distance, and Squared Euclidean distance (Martey et al. (2017)). The most common distance used in cluster analysis for metric variables is the Euclidian distance that can be calculated between any two observations *a* and *b* by the Eq. (1):

$$d_{ab} = \sqrt{\sum_{j=1}^{k} (X_{ja} - X_{jb})^2} \tag{1}$$

where k is the number of metric variables X in the observed dataset.

It is worth mentioning that if the variables in the dataset do not have the same unit of measures, a data standardization procedure such as Z-scores shall be previously carried out before the distance calculation. The new standardized variable ZX_{ji} for each variable j and observation i is presented in Eq. (2):

$$ZX_{ji} = \frac{X_{ji} - \bar{X}_j}{s_j} \quad (2)$$

where $\bar{X}_j$ and s_j represent, respectively, the mean and the standard deviation of variable X_j.

Without the standardization, the intensity of the distances between the observations may be arbitrarily influenced by the variables that will possibly present greater magnitude in their values, to the detriment of the others (Fávero and Belfiore (2019)).

After deciding the distance measure for the application of the cluster analysis, the user shall choose the clustering method, also known as agglomeration schedule. This is as important as defining the distance or similarity measures and should be made based on the objectives of the user for the analysis (Johnson and Wichern (2007)). Among the hierarchical agglomeration schedules, the most commonly used are these three linkage methods: single-linkage (nearest neighbor), complete-linkage (furthest-neighbor), and average-linkage (between-groups) (Fávero and Belfiore (2019)).

As these variations in the clustering method have an impact on the final result of the cluster analysis, the user shall be aware of the implications of each one. For Fávero and Belfiore (2019), the single-linkage agglomeration schedule is advisable in cases in which the observations are relatively far apart as it form clusters considering a minimum of homogeneity. On the other hand, complete-linkage is recommended in cases in which there is no considerable distance between the observation. Finally, the average-linkage merges two groups based on the average distance between all pairs of observations that are in these groups.

From the clustering stages and the distances between the clusters formed, a tree-shaped diagram known as a dendrogram can be developed, as exemplified in Fig. 1.

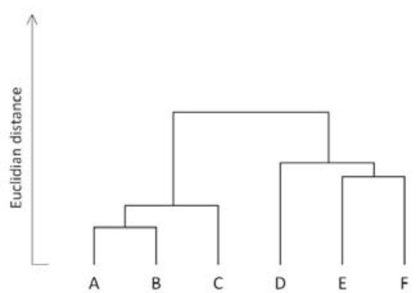

Fig. 1. A simplified dendrogram.

As can be seen in Fig. 1, A and B are the observations that have the smallest Euclidian distance between them and, therefore, grouped first in this dendrogram. This cluster is grouped with observation C, forming a cluster composed of A, B, and C. Then, E and F are the next observations to be grouped together, forming a new cluster that is merged with observation D in the sequence, forming the cluster 'DEF'. Finally, the clusters 'ABC' and 'DEF' are grouped into a single cluster containing all six observations.

Accordingly, the dendrogram summarizes the clustering hierarchical process and explains the allocation of each observation in each cluster (Johnson and Wichern (2007)). It is also useful for recommending the number of clusters by identifying a considerable distance leap in the dendrogram. This criterion considers that a very high leap when grouping two clusters may incorporate observations with characteristics that are not so homogenous. Then, the defined number of clusters should be the number of clusters immediately before the great leap in the dendrogram (Fávero and Belfiore (2019)). For instance, the longest distance in Fig. 1 is when merging of the cluster 'ABC' with the cluster 'DEF', which suggests that the ideal number of clusters is two in this example.

4. Proposed Method

This paper proposes a novel method to support the failure management policy selection in asset management based on cluster analysis. This method includes three sections: Acquisition of physical asset performance data (I), Cluster analysis (II), and Definition of the failure management policies (III). The proposed method is detailed and represented in Fig. 2.

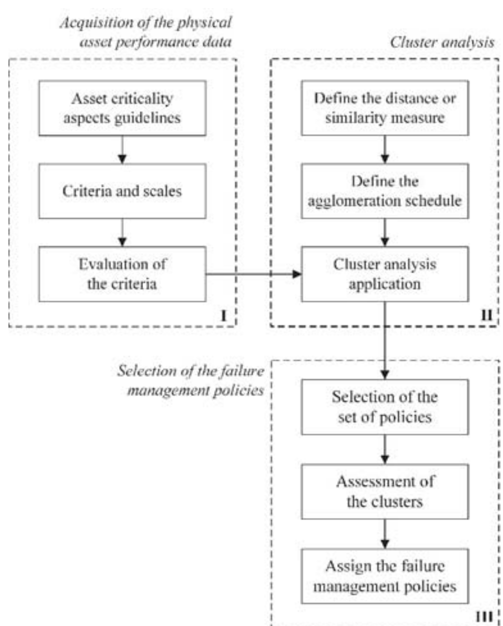

Fig. 2. The proposed support method for the selection of failure management policies.

4.1. *Acquisition of the physical asset performance data*

The proposed method starts with the acquisition of the physical asset performance data to compose the dataset for input in the cluster analysis. As criticality is an important property of any equipment, machinery, and engineering systems influencing the maintenance planning decisions regarding these physical assets, it is pertinent to choose it to guide the definition of the criteria and scales in this method.

From the criticality aspects, the organization derives the criteria and scales for evaluating the performance of physical assets. As technical knowledge is critical for the criticality evaluation, it is worth mentioning that the organization shall assign appropriate professionals for the task. Besides, a preliminary study of the system may be recommended to support these evaluations.

4.2. *Cluster analysis*

The cluster analysis section (II) begins with two main definitions concerning the cluster analysis settings. First, it needs to be decided the distance or similarity measure to be applied in the analysis. This is directly related to the type of variables in the dataset since distance measure is applied to metric variables and similarity measure to binary variables (Fávero and Belfiore (2019)). Moreover, the approach of how to calculate the chosen measure also needs to be defined (e.g., Euclidian distance).

Then, the agglomeration schedule to be applied in the cluster analysis needs to be decided. Basically, the agglomeration schedules can be classified into two types, hierarchical and nonhierarchical (Fávero and Belfiore (2019)). Each type of agglomeration schedule has different linkage methods to be considered for the application. After these definitions, the cluster analysis is applied with the dataset provided by the evaluation of physical assets across the criteria.

4.3. *Selection of the failure management policies*

Finally, in the Selection of the failure management policies section (III), the proposed framework assigns the failure management policy to each cluster of physical assets provided by cluster analysis. From a set of failure management policies, the user assesses the formed clusters in order to identify the reasons for the internal homogeneity of the physical assets. This information supports the selection of the overall failure management policies for the clusters.

5. Case Study

In this paper, the proposed method is demonstrated through a maintenance case study of a Brazilian hydroelectric power plant composed of four Kaplan turbine generating units with a total installed capacity of around 200 MW. This plant has been undergoing several studies for asset management improvements (Silva et al. (2019), Silva et al. (2020)).

For the acquisition of the physical asset performance data, the ISO 55000 series supported the definition of the criticality aspects due to its recent importance for maintenance and asset management. According to ISO 55000, assets can be critical in safety, environment, or performance and can relate to legal, regulatory, or statutory requirements (ISO (2014)). Thus, the definition of the criteria and scale was based on these four aspects of criticality. From them, nine criteria aligned to the organizational objectives were derived as proposed by Silva et al. (2019) and presented in Table 2.

Table 2. Aspects and criteria of criticality.

Id	Aspect	Criteria
s1	Safety	Safety classification
s2	Safety	History of safety events
e1	Environment	Environmental classification
e2	Environment	History of environmental events
p1	Performance	Reliability
p2	Performance	Maintainability
p3	Performance	Health assessment capacity
p4	Performance	Maintenance compliance
r1	Regulatory	Impact on availability

The scope of this case study only considered one generating unit (GU3) for the demonstration of the proposed method. The hierarchical structure of this generating unit contains eight main systems: Water intake (3.1.), Turbine (3.2.), Electrical connection (3.3.), Generator (3.4.), Water circulation (3.5.), Excitation (3.6.), Speed governor (3.7.), and Suction (3.8.).

The performance evaluation of the main items of these systems across the criteria based on a numerical scale of 1 to 9. Accordingly, 1 represented the best performance in a criterion while 9 is the worst. For better understanding, the performance evaluations of subunits and components of the turbine system (3.2.) of this generating unit are presented in Table 3.

Table 3. Evaluation of the items in the turbine system (3.2).

		Criticality criteria								
Id	**Description**	**s1**	**s2**	**e1**	**e2**	**p1**	**p2**	**p3**	**p4**	**r1**
3.2.1.1.	Filling valve 1	3	1	1	1	1	4	2	1	1
3.2.1.2.	Filling valve 2	3	1	1	1	1	4	2	1	1
3.2.1.3.	Drain valve	3	1	1	1	1	4	2	1	3
3.2.1.4.	Scroll case	7	1	3	1	1	7	9	1	2
3.2.2.	Pre-distributor	7	1	3	1	1	9	9	1	7
3.2.3.	Distributor	7	1	3	1	2	7	9	1	9
3.2.4.	Kaplan rotor	7	1	5	1	1	7	7	1	5
3.2.5.1.	Gasket box	7	1	1	1	2	3	5	1	5
3.2.5.2.	Turbine cover	7	1	1	1	2	3	5	1	5
3.2.5.3.	Drainage system	7	1	1	1	2	3	5	1	7
3.2.6.	Turbine shaft	7	1	1	1	1	7	3	1	7
3.2.7.	Guide turbine bearing	5	1	4	1	2	7	3	1	7
3.2.7.4.	Lubrication system	3	1	2	1	1	3	9	1	9

The application considered a total of 54 main items among all the systems of the generating unit that were evaluated and grouped to compose the dataset for the cluster analysis. Although they were all on the same numerical scale, the evaluations were standardized due to the different aspects of the variables.

Before applying the cluster analysis, it was necessary to define the settings of the application. As the dataset was not composed of binary variables, a distance measure approach was selected. Thus, the greater the differences between the variable values of two observations the smaller the similarity between them (Fávero and Belfiore (2019)). Besides, it was defined the Euclidean distance as the procedure to calculate the distances between the observations according to Eq. (1), previously presented.

Since there was no prior guidance or restriction on the number of desired clusters, it was decided that the cluster analysis would be hierarchical, initiating with all the observations separated. Among the hierarchical agglomeration schedules, it was decided to apply the complete-linkage method as it is advisable in cases in which there is no considerable distance between the observations. In this case study, although the application has nine criteria, the performance evaluation of the subunits and components of the generating unit is not relatively far apart.

With the settings concluded, the cluster analysis was implemented in R language with the support of the 'Cluster' package (Maechler et al. (2019)). The clustering process was summarized in the dendrogram, as presented in Fig. 3, with each item of GU3 represented by its id (e.g., 3.1.2.), as described in Table 4.

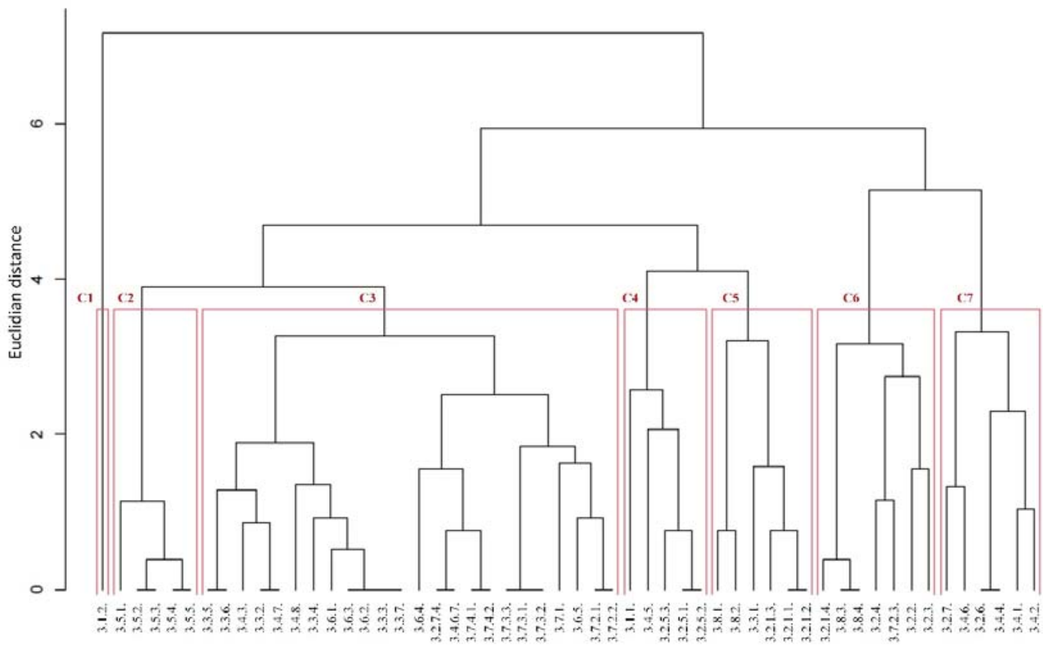

Fig. 3. Dendrogram for the cluster analysis of the GU3.

This hierarchical approach of the cluster analysis also allowed to decide the appropriate number of clusters in the selection of failure management policies. As can be seen in Fig. 3, the 54 pieces of equipment were grouped into seven clusters (C1 to C7). This decision was supported by the identification of a great leap between C2 and C3. Therefore, it was decided to maintain total the number of clusters before this clustering stage, resulting in the seven groups of items.

Through the id of each item, it is possible to identify which cluster it was allocated to and which other items belong to the same cluster. Additionally, Table 4 lists each of the 54 items with their respective ids of the hierarchical structure of the plant.

The six failure management policies, as previously presented in Table 1, were selected to compose the set of possible policies for the assignments. Then, each cluster was assessed to identify the reasons for the internal homogeneity of its items. For that, the original input dataset was divided according to the seven clusters, allowing an analysis to identify similarities in the performance evaluation across the criteria. Based on this assessment and the operational context of the organization, an overall failure management policy was selected for each cluster, as presented in Table 5.

As can be seen in Table 5, cluster 1 (C1) was the most apart from all other items, forming a group with only one item (3.1.2. Water intake grilles). This may indicate that this particular asset has an outlier behavior across the evaluation of the criteria. It was confirmed by assessing the performance evaluation and operational context of the plant. Thus, it was selected the 'One Time Change' policy in order to modify the water intake grilles maintenance routine. For instance, installing an automatic grille cleaner.

Table 4. List of ids and descriptions of the GU3 items.

Id and description	Id and description
3.1.1. Water intake gates	3.4.6. Combined generator bearing (CGB)
3.1.2. Water intake grilles	3.4.6.7. CGB lubrication system
3.2.1.1. Filling valve 1	3.4.7. Generator protection system
3.2.1.2. Filling valve 2	3.4.8. Radiators
3.2.1.3. Drain valve	3.5.1. Pipes
3.2.1.4. Scroll case	3.5.2. Self-cleaning filter
3.2.2. Pre-distributor	3.5.3. Cyclone filter
3.2.3. Distributor	3.5.4. W11 refrigeration valve
3.2.4. Kaplan rotor	3.5.5. Motorized refrigeration valve 8 "
3.2.5.1. Gasket box	3.6.1. Automatic voltage regulator
3.2.5.2. Turbine cover	3.6.2. Controlled rectifiers
3.2.5.3. Drainage system	3.6.3. Trigger and control system
3.2.6. Turbine shaft	3.6.4. Excitation transformer
3.2.7. Guide turbine bearing (GTB)	3.6.5. Collecting rings
3.2.7.4. GTB lubrication system	3.7.1. Distributor mechanism
3.3.1. Auxiliary service transformer	3.7.2.1. Kaplan head
3.3.2. Shielded rigid bars	3.7.2.2. Head Bushing
3.3.3. Circuit breaker	3.7.2.3. Kaplan Mechanism
3.3.4. High-voltage switch	3.7.3.1. Electronic regulator
3.3.5. High voltage current transformers	3.7.3.2. Distributor control
3.3.6. High voltage potential transformer	3.7.3.3. Turbine control
3.3.7. Low disconnect switch	3.7.4.1. Oil cooling system
3.4.1. Stator	3.7.4.2. Oil pressurization system
3.4.2. Generator rotor	3.8.1. Filling valve
3.4.3. Braking and lifting system	3.8.2. Drain valve
3.4.4. Generator shaft	3.8.3. Draft tube
3.4.5. Connecting elements	3.8.4. Chamber ring

The second cluster (C2) is composed of items that have a minor impact on the electrical generation that means their failures don't cause unavailability but can impact the operational conditions. Also, as these items have none to low impact on safety and the environment and no monitoring technologies, the 'Scheduled Discard' and 'Scheduled Restoration' were selected as the failure management policies to this cluster. On the other hand, as cluster 3 (C3) has a very critical impact on the generation as well as complete condition monitoring of the items, the 'On-condition' policy was the selected failure management policy for the cluster.

As cluster 4 (C4) is composed of items that critical impact on safety and a major impact on the generation, as appropriate failure management policy shall prevent that these failures don't occur. On the other hand, the items are also extremely reliable with some condition monitoring already in place. Then, the selected policies were 'On-condition' (when applicable) and 'Finding Failure'.

Table 5. Selected failure management policies for the clusters.

#	Item(s)	Similarities	Policy(ies)
C1	3.1.2.	Not applicable	One Time Change
C2	3.5.1.; 3.5.2.; 3.5.3.; 3.5.4.; 3.5.5.	All items belong to the same system; None to low safety and environmental classification; Low repair time and moderate reliability; No monitoring technologies in place; Minor impact on the generation.	Scheduled Discard or Scheduled Restauration
C3	3.3.5.; 3.3.6.; 3.4.3.; 3.3.2.; 3.4.7.; 3.4.8.; 3.3.4.; 3.6.1.; 3.6.3.; 3.6.2.; 3.3.3.; 3.3.7.; 3.6.4.; 3.2.7.4.; 3.4.6.7.; 3.7.4.1.; 3.7.4.2.; 3.7.3.3.; 3.7.3.3.; 3.7.3.2.; 3.7.1.; 3.6.5.; 3.7.2.1.; 3.7.2.2.	Minor impact on safety; Low environmental classification; Extremely reliable items; Low to moderate repair time; Complete condition monitoring; Very critical impact on the generation.	On-condition
C4	3.1.1.; 3.4.5.; 3.2.5.3.; 3.2.5.1.; 3.2.5.2.	Critical impact on safety; Low environmental classification; Extremely reliable items with low repair time; Some condition monitoring; Major impact on the generation.	On-condition and Finding Failure
C5	3.8.1.; 3.8.2.; 3.3.1.; 3.2.1.3.; 3.2.1.1.; 3.2.1.2.	Low safety and environmental classification; Extremely reliable items with low repair time; No or minor impact on the generation.	Run-to-failure
C6	3.2.1.4.; 3.8.3.; 3.8.4.; 3.2.4.; 3.7.2.3.; 3.2.2.;3.2.3.	Critical impact on safety; Minor impact on the environment; Extremely reliable items; High repair time; Complete condition monitoring;	On-condition
C7	3.2.7.; 3.4.6.; 3.2.6.; 3.4.4.; 3.4.1.; 3.4.2.	Low environmental classification; Extremely reliable items; High or very high repair time; Some condition monitoring; Critical impact on the generation.	On-condition or Scheduled Restoration

'Run-to-failure' were selected to the fifth cluster (C5), as its items have low safety and environmental impacts as well as no or minor impact to the generation. Moreover, it also has extremely reliable items with low repair time. Therefore, in an operational campaign, it can operate with no scheduled task. However, in the overhaul stop, all items shall be inspected and restored to as good as new.

Although cluster 6 (C6) has a critical impact on safety and high repair time, its items have complete condition monitoring in place. Thus, the selected failure management selected was on-condition to this cluster. On the other hand, as cluster 7 (C7) has only some condition monitoring in place even if the critical impact on the generation, it was selected the ‘On-condition’ (where applicable) and the ‘Scheduled Restoration’ policies.

Finally, it should be noted that in the selection of failure management policies for the clusters, the possibility of investment to adequate the physical asset to monitoring condition was not considered. Therefore, the on-condition policy was only selected when the monitoring condition was in place as evaluated through one of the criticality criteria.

6. Conclusions

Failure management policies are essential for properly translate the intentions and directions of the organization to maintenance management. From them, maintenance planning can elaborate maintenance plans for the physical assets, ensuring that maintenance tasks have been planned in line with the organization's expectations. Nevertheless, the selection of the failure management policies for a portfolio of physical assets is still challenging in organizations.

Creating homogenous clusters and reducing data structurally are some of the main reasons that make researchers choose to work with cluster analysis (Fávero and Belfiore (2019)). This exploratory technique can contribute to maintenance planning in the selection of failure management policies to the portfolio of physical assets as it identifies groups of assets that have internal similarities. Accordingly, the same overall failure management policy can be assigned for all the items in the formed cluster.

Therefore, the present work proposed a novel method to support the failure management policy selection in asset management based on the cluster analysis. It intended to integrate an exploratory machine learning technique in order to provide information for maintenance decision-making. For that, a three-section method was developed and applied with a hydropower plant case study.

As a result, the proposed method showed to support the organizations in the selection of appropriate failure management policies according to determined groups of physical assets. Instead of analyzing the portfolio of physical assets individually, the application of the cluster analysis allows the items to be allocated in a cluster with internal similarities. This is an interesting benefit as it reduces the number of items during the selection of failure management policies. Moreover, the formed cluster may highlight similarities that were not so evident for maintenance management.

It is worth mentioning that as cluster analysis is an exploratory technique of machine learning it does not have a predictive nature for other observations not initially present in the sample (Fávero and Belfiore (2019)). In other words, it means that the inclusion of new items in the analysis can imply a rearrangement of the items in the groups.

Although this case study considered work in the subunits and components of the systems of the generating unit, an opportunity for future works may consider replicating it to the level of failure modes. However, in this suggestion, other aspects shall be appropriate defined to guide the criteria and scales for the performance evaluations.

Finally, the proposed method and results of this work are expected to contribute to asset management research and maintenance practitioners facing the challenge of defining the appropriate failure management policy to prevent failures in a portfolio of physical assets.

Acknowledgement

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, for the development of the present research.

References

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

Global Forum on Maintenance & Asset Management (GFMAN) (2016). *The Maintenance Framework* (1 ed.). GFMAN.

International Organization for Standardization (ISO) (2014). ISO 55000, Asset management – Overview, principles and terminology.

International Organization for Standardization (ISO) (2015). ISO 9000, Quality management systems – Fundamentals and vocabulary.

Johnson, R. A. and D. W. Wichern (2007). *Applied Multivariate Statistical Analysis*. Pearson Education.

Maechler, M., P. Rousseuw, A. Struyf, M. Hubert, K. Hornik (2019). *Cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.0.

Martey, E. N., L. Ahmed, and N. Attoh-Okine (2017). Track geometry big data analysis: A machine learning approach. *IEEE International Conference on Big Data (BIGDATA)*, 3800-3809.

Schneider, M., D. Lucke, and T. Adolf (2019). A cyber-physical failure management system for smart factories. *52nd CIRP Conference on Manufacturing Systems*, 300-305.

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.

Silva, R. F., A. H. A. Melani, M. A. C. Michalski, G. F. M. Souza, S. I. Nabeta, and F. H. Hamaji (2020). Defining maintenance performance indicators for asset management based on ISO 55000 and Balanced Scorecard: A hydropower plant case study. *Proceedings of 30th European Safety and Reliability Conference*, 371-378.

Society of Automotive Engineers International (SAE International) (2009). JA1011, Evaluation Criteria for Reliability-Centered Maintenance (RCM) Processes.

Society of Automotive Engineers International (SAE International) (2011). JA1012, A guide to the Reliability-Centered Maintenance (RCM) Standard.