Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 54-1 (2021) 31–36

Module Failure Feature Detection by Cluster Analysis for Fleets of Civil Aircraft Engines ⋆

Hang Zhou * **Alexandra Brintrup** ** **Ajith Kumar Parlikad** ***

* *Institute for Manufacturing, Department of Engineering, University of Cambridge, Cambridge, UK (e-mail: hang.zhou@eng.cam.ac.uk)*
** *Institute for Manufacturing, Department of Engineering, University of Cambridge, Cambridge, UK (e-mail: ab702@cam.ac.uk)*
*** *Institute for Manufacturing, Department of Engineering, University of Cambridge, Cambridge, UK (e-mail: aknp2@cam.ac.uk)*

Abstract: During the operation of airplanes in the aviation industry, the stochastic environment in which the airplanes are being operated creates high complexity in planning the fleet overhauls for the aircraft engines. In order to perform efficient and cost-effective overhauls, the maintenance approach is to exchange the modules that caused the removal of engines from airplanes. Typical civil aviation engine consists of eight modules, in order to avoid disturbance for engine end users (airlines), all the modules should be available and well planed in any period of time window. This paper aims to provide a module planning solution based on the fleet operational history of aircraft engines, in order to cluster the engine performances into featured zones for overhaul module demand estimation.

Keywords: cluster analysis, unsupervised learning, aerospace engineering, maintenance

1. INTRODUCTION

The interest in the topic of fleet management has been increasing in recent years, especially in the transportation sector, e.g., airplanes (Burke et al. (2010)), trains (Lu and Schnieder (2015)) and automobiles (Nair and Miller-Hooks (2011)). Fleet management can be defined in a simple form as: the managing fleets of equipment to meet customer requests as they evolve over time, the equipment has to serve customers who typically want to move from one location to the next (Powell and Topaloglu (2005)). One significant improvement for fleet management compared to considering each member of the fleet independently is the increase of efficiency in task planning (Sohoni et al. (2011)), as well as saving costs on the fleet life-cycle. This is particularly beneficial in fleet maintenance planning (Sheng and Prescott (2019)), spare parts procurement (Van Horenbeek et al. (2013)) and warehouse management (Accorsi et al. (2017)). Cutting-edge technologies and algorithms are proposed in the improvement of fleet operational research in recent years. Including multi-agent deep reinforcement learning on managing fleet of taxi sharing services Lin et al. (2018), cloud-based IoT management for prefabrication transportation in public housing production background Xu et al. (2019), the IoT-based architecture is further discussed in its applications on multi-sensor based predictive maintenance for a fleet of buses on keeping peak performance of such vehicles Killeen et al. (2019). In order to perform successful and effective fleet management, the identification of similarities in performance characteristics for member assets within fleets is vital as it enables clustering of similar assets within the fleet and develop asset management strategies appropriate to each cluster of assets.

Fleet planning is especially important in the aviation industry. A typical design of civil aircraft engines contains eight main modules: fan/low pressure compressor (LPC), intermediate pressure compressor (IPC), high pressure compressor (HPC), combustor (CBT), high pressure turbine (HPT), intermediate pressure turbine (IPT), low pressure turbine (LPT), and external gear box (Ackert (2011)), as shown in Figure 1. An engine overhaul involves identification of the modules that are the root causes of the engine removal and replacement of the defective modules for efficient and cost-effective maintenance. Effective planning of such overhauls is beneficial for both the engine manufacturer and its airline customers, as it minimizes the disturbance of flight operations. One key influential factor in guaranteeing the effectiveness of this maintenance strategy is sufficient availability of spare modules in any given time window. In order to make accurate estimation of the number and type of spares required, it is important to understand the performance patterns of the fleet of the engines with historical operational data, both at the entire engine level and the module level.

In the aviation industry, the engine life is measured using two types of measurement scales: flight-hours and the flight-cycles. There have been studies focusing on the flight-hour measurement of operation (Kennet (1994); MacMinn and Jones (1989); Painter et al. (2006); Ashby and Byer (2002)) and flight-cycle measurement of opera-

⋆ This research is funded by the Aerospace Technology Institute (ATI) and InnovateUK.

10.1016/j.ifacol.2021.08.154

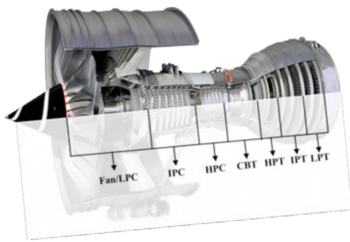

Fig. 1. Civil Aircraft Engine Module Overview

tion (Liao et al. (2018); Wang et al. (2018); Sayah et al. (2020)). However, in the literature there is currently a gap in the evaluation of civil aircraft engine performance by a combination of the time-scale measurements. This paper proposes an approach to evaluate the performance of the engines with a dual time scale measurement in order to provide a more comprehensive insight to services when the fleet of engines are used in a mixture of long-haul, medium-haul and short-haul flights.

Another key challenge is the complexity of engine overhauls associated with a combination of module failures. Additionally, maintenance logbook records are often written in natural language and do not always provide clear indications of the failed modules. In order to obtain a clearer insight of the engine level performance and the module level performance, we propose a solution to bypass the difficulty in the lack of clear informative maintenance records - an unsupervised distribution based clustering approach to divide the engine performances into featured zones purely based on the statistical characteristics of the data points. We then utilise the fraction of informative maintenance records to validate the cluster results, and finally construct a performance map of the hidden engine module failure information.

The paper is structured as: in Section 2 the methodology will be introduced, followed by the results and validation in Sections 3, with conclusion and future work in Section 4 to end this paper.

2. METHODOLOGY

To enable the evaluation of engine level performance and module level performance, in order to ultimately achieve the goal of module demand estimation, the following analysis steps are taken showing in figure 2. The key steps are the definition of dual time scale performance, the fleet survival analysis, the cluster analysis, the validation by information extracted from the maintenance logbook, and the application of the analytical results towards the condition-based estimation of module demands for efficient maintenance.

2.1 Fleet dual time scale performance re-evaluation

One concept this paper proposes is the performance re-evaluation of the civil aircraft engines under dual time-scale measurement. Due to the operation track-record of each individual engine, the final symptoms being diagnosed from the engine which necessitates an overhaul are either contributed more by the hour-measurement or more by the cycle-measurement. A representative failure

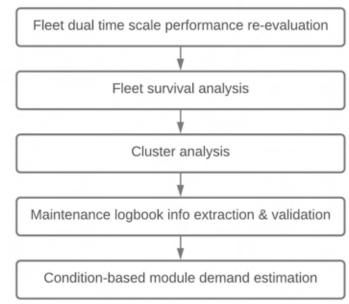

Fig. 2. Flowchart of analysis steps

mechanism where the hour-measurement is critical is the physical degradation of creep for components operating under extreme high temperature (Kassner and Pérez-Prado (2000)). A representative failure mechanism where the cycle-measurement is critical is corrosion fatigue where components are vulnerable with each take-off-landing cycle due to cyclic loading. A typical corrosion fatigue process for aviation components starts by particles striking the surface of the metal material and scratching the anti-corrosive coatings resulting in pits on the material surface. The pit initiation is followed by crack propagation due to fatigue (Zhou et al. (2017, 2018)), for which the crack propagation rate is widely described by the Paris' law (Schütz (1996)).

In order to determine the contribution of the dual-time scale measurement, both the data collected for engine hour-to-failure and cycle-to-failure are normalized, taking the largest value in the dataset as the upper boundary and 0 as the lower boundary. Assume there are n individual engines within the dataset of an engine family, and each of the individual engine goes through an overhaul with a measurement of hour h_i and a measurement of cycle c_i where $i \in [1, n]$. Among all the measurements, the maximum service time measurement by hour is h_{max} and the maximum service time measurement by cycle is c_{max}. For each individual engine with a performance dual measurement $\overrightarrow{Sys_i}(h_i, c_i)$, the normalised dual measurement is $\overrightarrow{Sys_i} = (\frac{h_i}{h_{max}}, \frac{c_i}{c_{max}}), i \in [1, n]$. The performance re-evaluation based on the integration of both time scale measurements is represented as the magnitude of the vector.

$$\mathrm{Perf_i} = \sqrt{(\frac{h_i}{h_{max}})^2 + (\frac{c_i}{c_{max}})^2} \qquad (1)$$

2.2 Fleet survival analysis

The Kaplan-Meier estimator (Kaplan and Meier (1958)) is applied for the survival analysis of a family of aircraft engines:

$$\hat{\mathrm{S}}(\mathrm{Perf_i}) = \prod_{i:0 \leq \mathrm{Perf_i} \leq \mathrm{Perf_n}} (1 - \frac{d_i}{s_i}) \qquad (2)$$

Here d_i represents the total overhaul cases recorded at joint life estimation of $Perf_i$, and s_i is the number of survived engines that have not yet gone through overhaul at the joint life estimation $Perf_i$. Sorting the data points within the dataset according to its survival analysis, each data point contains three types of information: the normalised hour value, the normalised cycle value and the fleet survival rate value. These are represented as $\overrightarrow{Sys_i}' = (\frac{h_i}{h_{max}}, \frac{c_i}{c_{max}}, \hat{S}(\mathrm{Perf_i}))^T$. Thus the dataset for the overhaul data with dual-time scale performance evaluation and the survival analysis is expressed as:

$$\mathbf{SR_{EngineFamily1}} = [\overrightarrow{Sys_1}', \overrightarrow{Sys_2}', \cdots, \overrightarrow{Sys_n}'] = \begin{bmatrix} \frac{h_1}{h_{max}} & \frac{h_2}{h_{max}} & \cdots & \frac{h_n}{h_{max}} \\ \frac{c_1}{c_{max}} & \frac{c_2}{c_{max}} & \cdots & \frac{c_n}{c_{max}} \\ \hat{S}(\mathrm{Perf_1}) & \hat{S}(\mathrm{Perf_2}) & \cdots & \hat{S}(\mathrm{Perf_n}) \end{bmatrix} \quad (3)$$

$\mathbf{SR_{EngineFamily1}}$ is a $3 \times n$ dimensional matrix.

2.3 Cluster Analysis

In order to perform the clustering analysis, the three-dimensional dataset of $\mathbf{SR_{EngineFamily1}}$ is in need of dimensionality reduction. One of the common dimensionality reduction methods is Principal Component Analysis (PCA). The purpose of performing PCA in the dataset is that the clustering algorithm performs better while the dataset maintains the most information and the topological relationships among all the data points. Following the PCA, the clustering of datapoints are trialed with three major clustering approaches: the partition-based approach, the fuzzy-based approach and the distribution-based approach.

PCA The first data processing step, the PCA, is specifically a three-dimensional to two-dimensional transferring problem Jian Yang et al. (2004).

Clustering Methods Three distinctive clustering approaches are trialed in this research, based on three measuring philosophies. First, the K-means algorithm, which is a hard-clustering method where each data point is allocated to one sub-group and one only based on the measurement of distance among the datapoints and their nearest cluster centroids. The distance metrics being used the most frequently are Euclidean Distance (Warren Liao (2005)) and Manhattan Distance (de Amorim and Hennig (2015)). Second, the fuzzy C-means clustering method is an extension of the hard-clustering K-means method. The major difference is the inclusion of the fuzzy-partition matrix (Warren Liao (2005)). Third, the distribution clustering approach, represented by the Gaussian Mixture Model (GMM), allocates each observation in the dataset to the distributions that are most likely to be the same. This method is the main approach applied in this research, while the K-means and the fuzzy C-means are used as references to the determination of the most reasonable and optimized number of clusters without prior knowledge of the datasets.

According to the GMM, with a set of observations $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, and a weighted sum of m clusters. The Gaussian mixture density is defined as (Hedelin and Skoglund (2000)):

$$f_{x|w,\hat{\theta}}(x \mid w, \hat{\theta}) = \sum_{i=1}^{m} w_i f_{x|\hat{\theta}_i}(x \mid \hat{\theta}_i) \quad (4)$$

where w_i is the weight of each observation and the observation densities are:

$$f_{x|\hat{\theta}_i}(x \mid \hat{\theta}_i) = (2\pi)^{-\frac{k}{2}} det(\Sigma_i)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)} \quad (5)$$

Here, μ_i represents mean vectors and Σ_i represents the covariance matrices. The purpose of the GMM clustering is to obtain the parameters of clustered Gaussian distribution parameters $\hat{\theta}$.

In order to obtain the optimised values for the parameters in the GMM, the purpose is to the maximize the log-likelihood function given N independent samples from the identically distributed samples of $\mathbf{x}$ observations (Hedelin and Skoglund (2000)).

$$L(\hat{\theta}) = \sum_{n=1}^{N} ln \sum_{i=1}^{m} w_i f_{x|\hat{\theta}_i}(x_n \mid \hat{\theta}_i) \quad (6)$$

Where the expectation-maximization (EM) algorithm (Dempster et al. (1977); Yang et al. (2012)) is applied.

Cluster Number Determination With limited to no prior knowledge of the datasets, one obstacle for unsupervised clustering analysis is the determination of the most suitable cluster numbers. There are several existing conventional cluster number determination method, including the elbow method (Green et al. (2014)) and the silhouette statistic index (Xu et al. (2012)) for K-means clustering. Applying the fuzzy partition coefficient (FPC) (Trauwaert (1988)) for the fuzzy c-means clustering method. And both the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) for the GMM method. However, it is well known that without pre-defined knowledge and especially when the dataset contains high-complexity information, the conventional clustering number determination can be ineffective (de Amorim and Hennig (2015)). In this paper, the optimized cluster number is determined by the unweighted fuzzy logic decision making approach, by utilizing the values calculated from the fuzzy logic membership (Agrawal et al. (2008)).

$$\mu_i = \begin{cases} 1, & \text{if } \mathcal{F}_\mathrm{i} \leq \mathcal{F}_\mathrm{i}{}^{\max} \\ \frac{\mathcal{F}_i{}^{max} - \mathcal{F}_i}{\mathcal{F}_i{}^{max} - \mathcal{F}_i{}^{min}}, & \text{if } \mathcal{F}_i{}^{max} < \mathcal{F}_i < \mathcal{F}_i{}^{min} \\ 0, & \text{if } \mathcal{F}_i \geq \mathcal{F}_i{}^{max} \end{cases} \quad (7)$$

Here, μ_i is the fuzzy logic membership value of the i^{th} objective function $\mathcal{F}_i$. The normalized membership function for each non-dominated solution is written as:

$$\mu[h] = \frac{\sum_{i=1}^{p} \mu_i[h]}{\sum_{h=1}^{q} \sum_{i=1}^{p} \mu_i[h]} \quad (8)$$

By definition, the optimal number of clusters refers to the minimum normalized fuzzy membership value.

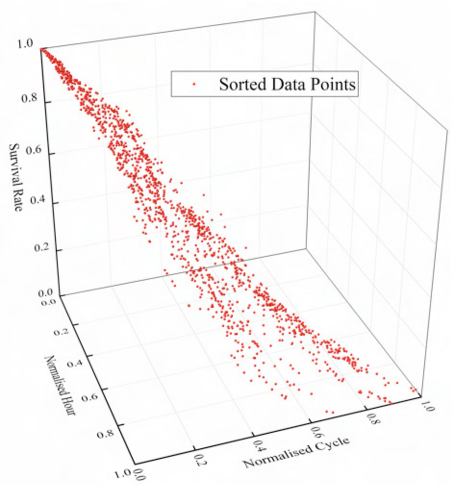

Fig. 3. Sorted overhaul performance profile with survival analysis for engine family

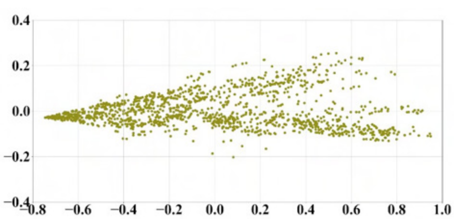

Fig. 4. PCA transformation on the 'Waterfall Model' for engine family 1

3. RESULTS AND VALIDATION

The datasets we use contain fleets of engine families; each family has the same design and thrust, being used by multiple airlines in very different operational environments and regimes. The datasets, taken five rows from the engine family 1 as an example, is shown in Table 1. Note that for confidential reasons, the hour-to-failure and the cycle-to-failure values are scaled by a factor.

Table 1. Example of dataset for engine family

Design Model	Thrust Level	Serial Code	Airline	Overhaul (Hour)	Overhaul (Cycle)
Z	A	1003	AL26	2.2×10^8	1.9×10^7
Z	A	1054	AL28	3.4×10^8	3.1×10^7
Z	A	1079	AL01	6.4×10^8	4.5×10^7
Z	A	1078	AL11	1.2×10^9	9.3×10^7
Z	A	1114	AL15	5.1×10^8	4.3×10^7

The information obtained from this dataset is the time-to-overhaul of each engine from either newly purchased or newly refurbished to its 100% of engine life, the time-to-overhaul for each individual engine being recorded contains both the hour-to-failure and the cycle-to-failure. Applying the fleet dual time scale performance re-evaluation in section 2.1 and the fleet survival analysis in section 2.2, the engine family forms the 'Waterfall Model' as shown in Figure 3. Following the PCA transformation, the datapoints are mapped to a two-dimensional coordinate shown in Figure 4.

In order to determine the optimized cluster number, as stated in section 2.3, all of the three clustering approaches are tested on four different engine families, applying the fuzzy logic decision making method. The results of the determined cluster numbers are shown in table 2 for comparison. Even though the three clustering methods have different measurements justifying the belonging of each datapoint towards determined clusters, fundamentally they are all unsupervised learning processes. With minimum prior knowledge, it is reasonable that the three methods, with a unified cluster number optimization algorithm, should achieve a similar number of cluster numbers without being largely different.

Table 2. Summarised optimal cluster number

Cluster Methods	Engine Family1	Engine Family2	Engine Family3	Engine Family4
$K - means$	5±1	5±1	5±1	5±1
$Fuzzy\ C - means$	4±1	4±1	5±1	6±1
GMM	7±1	5±1	5±1	7±1
Optimal	**6**	**5**	**5**	**6**

The fuzzy logic decision solved two difficulties the conventional determination methods face. First, the cluster performances tend to unify with different clustering method: even though the approaches are fundamentally different, they all reach a certain agreement of the most optimal cluster for a data being collected from each aircraft engine family. Second, the distribution based clustering algorithm, the GMM, by conventional method, is sensitive with the size of the dataset, particularly when the overhaul datasets of the four engine families in this research contains as many as over 1400 records for engine family 1 and as few as 200 records for engine family 4. With the fuzzy logic technique, the GMM clustering provides a stable performance.

Table 3. Example of dataset for engine family

	Cluster (C1)	Cluster (C2)	Cluster (C3)	Cluster (C4)	Cluster (C5)	Cluster (C6)
Sub 1	37.5%	18.6%	8.7%	4.4%	25.8%	11.5%
Sub 2	-	35.6%	-	-	-	-
Sub 3	-	-	41.7%	13.1%	25.8%	-
Sub 4	-	-	12.2%	33.3%	-	9.4%
Sub 5	-	-	29.6%	23.2%	38.7%	18.8%
Sub 6	-	-	-	-	-	46.9%
Sub 7	25.0%	13.6%	-	-	-	-
Sub 8	20.8%	18.6%	-	-	-	-
Sub 9	16.7%	13.6%	7.8%	26.1%	9.7%	13.6%

By definition, the optimal number of clusters refers to the minimum normalized fuzzy membership value . The associated normalized fuzzy membership values are shown in Figure 5.

After the determination of the optimized cluster numbers, the GMM is applied on the datasets, and the cluster results by a visual observation is shown in Figure 6 (taken the engine family 1 as an example for the final cluster results).

In order to validate the rationality of the cluster results, a further step is taken, which is to extract the root cause of engine removal from the maintenance logbook. The maintenance logbooks contain descriptions of engine removals in natural language. However, the information that clearly states the failed modules is limited. We applied the natural language processing and recognition, to utilize all the clearly stated removal reasons and validate the

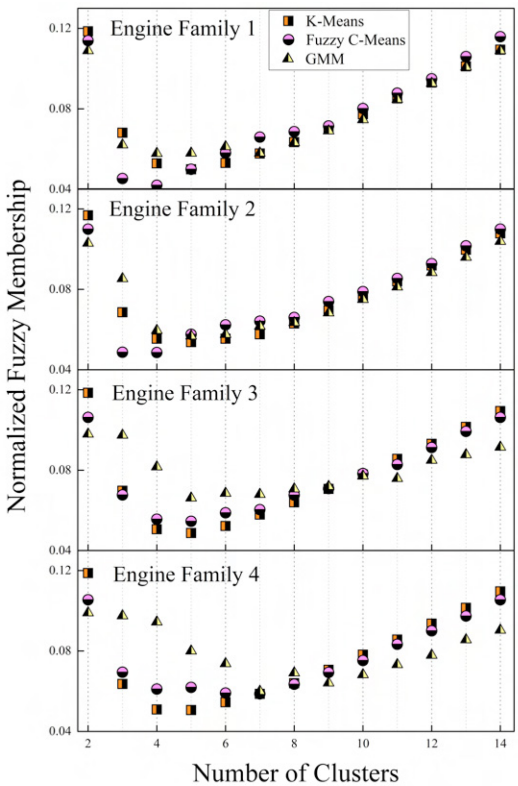

Fig. 5. Normalized fuzzy membership values for four engine families

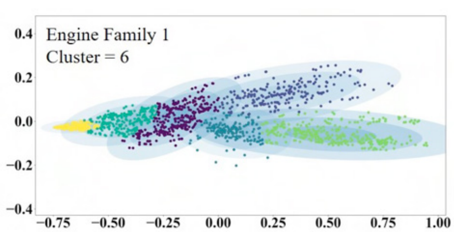

Fig. 6. GMM clustering on engine family 1 with the optimal number of determined clusters

cluster results. Table 3 lists the validation results, combining the extracted knowledge and statistical inference, to conclude the hierarchy of the risky modules in each of the determined clusters. Due to confidential reasons, the exact module names are not provided in the table, but they are the eight modules of the civil aviation engines plus fan case, in total represented by nine sub-systems, or nine 'Sub's. The example shown here is the validation results of the engine family 1. The columns represent the percentage of failed subsystems within each cluster.

The validation results provide the evidence that there is a distinctive featured failure hierarchy within each divided clusters based on the engine performances. Each cluster is distinctive with one leading root cause on one of the modules for engine removal (red in Table 3), which proves the rationality of the unsupervised learning result. It is worth noticing that the clustering results are distribution based clustering analysis, meaning the determination of maintenance records in its allocated cluster is by confidence. The final module demand estimation is thus calculated by considering the confidence of the engine useful life falling into one cluster by the location of the data value in the distribution confidence zones.

4. CONCLUSION AND FUTURE WORK

One important application of the clustering results lies in the estimation of module demands in any given time window for spares planning purposes. With the fleet of engine being used by airlines everyday, the location of the engines in the 'Waterfall Model' is dynamic. The end-of-life engines are refurbished through overhaul and rejoins the fleet at the starting point with 100% remaining useful life. Dynamically the fleet of engines enters one of the six clustered zones with a different probability of failure, bearing a degradation feature presented in the results of this paper. Such properties of the clustering analysis enable the application of this research as an efficient and accurate health monitoring and module demand estimation tool, both at the engine level and at the module level, for the purpose of efficient and cost-effective maintenance planning.

The next step of this research is to describe the three-dimensional 'Waterfall' model and the featured clustering zones using mathematical functions to improve the simplicity and usability of the research results.

REFERENCES

Accorsi, R., Bortolini, M., Gamberi, M., Manzini, R., and Pilati, F. (2017). Multi-objective warehouse building design to optimize the cycle time, total cost, and carbon footprint. *Int J Adv Manuf Technol*, 92, 839–854. doi: https://doi.org/10.1007/s00170-017-0157-9.

Ackert, S. (2011). Engine maintenance concepts for financiers - elements of turbofan shop maintenance costs.

Agrawal, S., Panigrahi, B.K., and Tiwari, M.K. (2008). Multiobjective particle swarm algorithm with fuzzy clustering for electrical power dispatch. *IEEE Transactions on Evolutionary Computation*, 12(5), 529–541. doi:10.1109/TEVC.2007.913121.

Ashby, M.J. and Byer, R.J. (2002). An approach for conducting a cost benefit analysis of aircraft engine prognostics and health management functions. In *Proceedings, IEEE Aerospace Conference*, volume 6, 6–6. doi:10.1109/AERO.2002.1036124.

Burke, E.K., De Causmaecker, P., De Maere, G., Mulder, J., Paelinck, M., and Vanden Berghe, G. (2010). A multi-objective approach for robust airline scheduling. *Comput. Oper. Res.*, 37(5), 822–832. doi: 10.1016/j.cor.2009.03.026.

de Amorim, R.C. and Hennig, C. (2015). Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Information Sciences*, 324, 126–145. doi:https://doi.org/10.1016/j.ins.2015.06.039.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–38.

Green, R., Staffell, I., and Vasilakos, N. (2014). Divide and conquer? k-means clustering of demand data allows rapid and accurate simulations of the british electricity system. *IEEE Transactions on Engineering Management*, 61(2), 251–260. doi:10.1109/TEM.2013.2284386.

Hedelin, P. and Skoglund, J. (2000). Vector quantization based on gaussian mixture models. *IEEE Transactions on Speech and Audio Processing*, 8(4), 385–401. doi: 10.1109/89.848220.

Jian Yang, Zhang, D., Frangi, A.F., and Jing-yu Yang (2004). Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1), 131–137. doi: 10.1109/TPAMI.2004.1261097.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457–481. doi: 10.1080/01621459.1958.10501452.

Kassner, M.E. and Pérez-Prado, M.T. (2000). Five-power-law creep in single phase metals and alloys. *Progress in Materials Science*, 45(1), 1–102. doi: https://doi.org/10.1016/S0079-6425(99)00006-7.

Kennet, D.M. (1994). A structural model of aircraft engine maintenance. *Journal of Applied Econometrics*, 9(4), 351–368. doi:10.1002/jae.3950090405.

Killeen, P., Ding, B., Kiringa, I., and Yeap, T. (2019). Iot-based predictive maintenance for fleet management. *Procedia Computer Science*, 151, 607–613. doi: https://doi.org/10.1016/j.procs.2019.04.184. The 10th International Conference on Ambient Systems, Networks and Technologies (ANT 2019) / The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40 2019) / Affiliated Workshops.

Liao, Y., Zhang, L., and Liu, C. (2018). Uncertainty prediction of remaining useful life using long short-term memory network based on bootstrap method. In *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 1–8. doi: 10.1109/ICPHM.2018.8448804.

Lin, K., Zhao, R., Xu, Z., and Zhou, J. (2018). Efficient large-scale fleet management via multi-agent deep reinforcement learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*, KDD '18, 1774–1783. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3219819.3219993. URL `https://doi.org/10.1145/3219819.3219993`.

Lu, D. and Schnieder, E. (2015). Performance evaluation of gnss for train localization. *IEEE Transactions on Intelligent Transportation Systems*, 16(2), 1054–1059. doi:10.1109/TITS.2014.2349353.

MacMinn, S.R. and Jones, W.D. (1989). A very high speed switched-reluctance starter-generator for aircraft engine applications. In *Proceedings of the IEEE National Aerospace and Electronics Conference*, 1758–1764 vol.4. doi:10.1109/NAECON.1989.40453.

Nair, R. and Miller-Hooks, E. (2011). Fleet management for vehicle sharing operations. *Transportation Science*, 45(4), 524–540.

Painter, M.K., Erraguntla, M., Hogg, G.L., and Beachkofski, B. (2006). Using simulation, data mining, and knowledge discovery techniques for optimized aircraft engine fleet management. In *Proceedings of the 2006 Winter Simulation Conference*, 1253–1260. doi: 10.1109/WSC.2006.323221.

Powell, W. and Topaloglu, H. (2005). *Fleet Management.* SIAM. doi: https://doi.org/10.1137/1.9780898718799.ch12.

Sayah, M., Guebli, D., Zerhouni, N., and Masry, Z.A. (2020). Towards distribution clustering-based deep lstm models for rul prediction. In *2020 Prognostics and Health Management Conference (PHM-Besançon)*, 253–256. doi:10.1109/PHM-Besancon49106.2020.00049.

Schütz, W. (1996). A history of fatigue. *Engineering Fracture Mechanics*, 54(2), 263–300. doi: https://doi.org/10.1016/0013-7944(95)00178-6.

Sheng, J. and Prescott, D. (2019). A coloured petri net framework for modelling aircraft fleet maintenance. *Reliability Engineering System Safety*, 189, 67 – 88. doi: https://doi.org/10.1016/j.ress.2019.04.004.

Sohoni, M., Lee, Y.C., and Klabjan, D. (2011). Robust airline scheduling under block-time uncertainty. *Transportation Science*, 45(4), 451–464. doi:10.1287/trsc.1100.0361. URL `https://doi.org/10.1287/trsc.1100.0361`.

Trauwaert, E. (1988). On the meaning of dunn's partition coefficient for fuzzy clusters. *Fuzzy Sets and Systems*, 25(2), 217–242. doi:https://doi.org/10.1016/0165-0114(88)90189-3.

Van Horenbeek, A., Scarf, P.A., Cavalcante, C.A.V., and Pintelon, L. (2013). The effect of maintenance quality on spare parts inventory for a fleet of assets. *IEEE Transactions on Reliability*, 62(3), 596–607. doi: 10.1109/TR.2013.2270409.

Wang, J., Wen, G., Yang, S., and Liu, Y. (2018). Remaining useful life estimation in prognostics using deep bidirectional lstm neural network. In *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, 1037–1042. doi:10.1109/PHM-Chongqing.2018.00184.

Warren Liao, T. (2005). Clustering of time series data—a survey. *Pattern Recognition*, 38(11), 1857–1874. doi: https://doi.org/10.1016/j.patcog.2005.01.025.

Xu, G., Li, M., Luo, L., Chen, C.H., and Huang, G.Q. (2019). Cloud-based fleet management for prefabrication transportation. *Enterprise Information Systems*, 13(1), 87–106. doi:10.1080/17517575.2018.1455109. URL `https://doi.org/10.1080/17517575.2018.1455109`.

Xu, R., Xu, J., and Wunsch, D.C. (2012). A comparison study of validity indices on swarm-intelligence-based clustering. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(4), 1243–1256. doi:10.1109/TSMCB.2012.2188509.

Yang, M.S., Lai, C.Y., and Lin, C.Y. (2012). A robust em clustering algorithm for gaussian mixture models. *Pattern Recognition*, 45(11), 3950–3961. doi: https://doi.org/10.1016/j.patcog.2012.04.031.

Zhou, H., Gnanasambandam, S., Foresta, M., Li, F., Blanc, M.L., Weston, D., and Pan, J. (2018). Life prediction of phosphor bronze reinforcing tape used in underground power cables. *CORROSION*, 74(5), 530–542. doi: 10.5006/2627.

Zhou, H., Gnanasambandam, S., Foresta, M., Weston, D., Li, F., Pan, J., and Blanc, M.L. (2017). Measurement and modeling of pitting depth distribution for phosphor bronze tapes used in underground power transmission cables. *CORROSION*, 73(7), 844–852. doi: 10.5006/2227.