Applied Acoustics, Volume 147, April 2019, Pages 77-86
DOI:10.1016/j.apacoust.2017.12.003

Reciprocating compressor prognostics of an instantaneous failure mode utilising temperature only measurements

Panagiotis Loukopoulos [a], George Zolkiewski [b], Ian Bennett [b], Suresh Sampath [a], Pericles Pilidis [a], Fang Duan [c], Tariq Sattar [c], and David Mba [d]

[a] School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, United Kingdom
Email addresses: loukopoulospn@gmail.com, s.sampath@cranfield.ac.uk, p.pilidis@cranfield.ac.uk
[b] Shell Global Solutions, Royal Dutch Shell, Rijswijk, Netherlands
Email addresses: george.zolkiewski@hotmail.com, Ian.Bennett@shell.com
[c] School of Engineering, London South Bank University, London, United Kingdom
Email addresses: duanf@lsbu.ac.uk, sattartp@lsbu.ac.uk
[d] Faculty of Technology, De Montfort University, Leicester, United Kingdom
Email address: profdavidmba@gmail.com
Corresponding author: loukopoulospn@gmail.com

Abstract

Reciprocating compressors are critical components in the oil and gas sector, though their maintenance cost is known to be relatively high. Compressor valves are the weakest component, being the most frequent failure mode, accounting for almost half the maintenance cost. One of the major targets in industry is minimisation of downtime and cost, while maximising availability and safety of a machine, with maintenance considered a key aspect in achieving this objective. The concept of Condition Based Maintenance and Prognostics and Health Management (CBM/PHM) which is founded on the diagnostics and prognostics principles, is a step towards this direction as it offers a proactive means for scheduling maintenance. Despite the fact that diagnostics is an established area for reciprocating compressors, to date there is limited information in the open literature regarding prognostics, especially given the nature of failures can be instantaneous. This work presents an analysis of prognostic performance of several methods (multiple linear regression, polynomial regression, K-Nearest Neighbours Regression (KNNR)), in relation to their accuracy and variability, using actual temperature only valve failure data, an instantaneous failure mode, from an operating industrial compressor. Furthermore, a variation for Remaining Useful Life (RUL) estimation based on KNNR, along with an ensemble technique merging the results of all aforementioned methods are proposed. Prior to analysis, principal components analysis and statistical process control were employed to create T^2 and Q metrics, which were proposed to be used as health indicators reflecting degradation process of the valve failure mode and are proposed to be used for direct RUL estimation for the first time. Results demonstrated that even when RUL is relatively short due to instantaneous nature of failure mode, it is feasible to perform good RUL estimates using the proposed techniques.

Keywords: *reciprocating compressor, valve, prognostics, remaining useful life, multiple linear regression, polynomial regression, K-nearest neighbours, instantaneous failure, principal components analysis, statistical process control*

1. Introduction

Reciprocating compressors are essential components in the oil and gas sector, being a key element in refining industry as one of the most frequently used type of equipment, demanding high reliability and availability [1,2]. They are employed extensively, being powerful, flexible, efficient, and dependable in many compression applications. Despite their popularity, their maintenance cost can be several times greater than that of other compressor types [3], since the number of moving parts is higher [4], thus they are expected to experience more failures. Bloch and Heinz [1] note that valves are the most common failing part (36%), making them the weakest component, accounting for almost half the maintenance cost [4].

Published by Elsevier. This is the Author Accepted Manuscript issued with:
Creative Commons Attribution Non-Commercial No Derivatives License (CC:BY:NC:ND 4.0).
The final published version (version of record) is available online at DOI:10.1016/j.apacoust.2017.12.003.
Please refer to any applicable publisher terms of use.

Valves are a vital part of the reciprocating compressor as they play a significant role on its performance from both efficiency and reliability perspectives [2]. Their smooth operation is integral since they regulate the gas flow for compression. Valves suffer numerous hardships during their operation as they may come in contact with liquids, foreign particles or debris, corrosive gases or materials depending on application [2]. Furthermore, pulsations, tension, compression and impact created either by the compressor or the valve motion itself can affect proper valve function [2].

In order to decrease downtime and cost, while increasing availability and safety of a reciprocating compressor, efficient maintenance is essential [1,2] since failures of such equipment can cause from production loss to human casualties [3,5]. Condition Based Maintenance (CBM) [6–10] is a policy founded on the diagnostics principle and has been increasingly popular over the years, advocating that maintenance should be undertaken only when actually needed depending on unit's health state; it is an effective tool that moves towards this direction [1], with diagnostics being an established area for valve failures [3,11–15]. The equipment of interest is mounted with sensors collecting Condition Monitoring (CM) measurements which are analysed for diagnostics purposes – determine whether the equipment is at healthy or faulty state, and in case of a fault identify the responsible failure mode – and suggest actions to be taken accordingly.

An advancement of CBM is Prognostics and Health Management (PHM) [6–9,16–19] which has been gaining traction during recent years and is founded on prognostics principle [6–10,16–20]. It estimates the time to failure, known as Remaining Useful Life (RUL), after a fault has occurred, enabling the user to schedule maintenance in advance. PHM's proactive nature can assist optimising maintenance by avoiding any unnecessary action. Since PHM can be employed after a fault has been detected, diagnostics is required and thus its coupling with CBM would be unavoidable, leading to CBM/PHM [6]. To the authors' knowledge, there is limited information about prognostics on reciprocating compressors in the open literature. Consequently, the purpose of this work is the comparison of several prognostics methods in order to identify most suitable ones based on accuracy and variability, while at the same time proposing new techniques.

Prognostics methodologies can be divided into two groups [6,7,9,16–21]:

i. *Data-driven. They model the degradation process using historical information, and are suitable when there is limited physical understanding of system under study. They struggle in cases for which they have not been trained like novel events, while their accuracy depends on amount and quality of available data.*
ii. *Physics based. They create a mathematical representation of system's or failure's physical aspect. They are computationally expensive and tend to be application specific though they can outperform data-driven.*

Similarly, there are two ways for calculating RUL [7,21,22]:

i. *Direct estimation. Relationship between information and RUL is modelled. It requires knowledge of historical and current information, with data being the input and RUL being the output. It is useful in cases lacking a failure threshold.*
ii. *Indirect estimation. Relationship between information and a Health Indicator (HI) reflecting machine's health status is modelled. In some cases the HI can be modelled as function of time. HI is extrapolated until a failure threshold is reached and RUL is estimated as the difference between current and failure time. It requires knowledge of historical, current, and future information.*

This paper concentrates on data-driven prognostics methods and employs direct RUL estimation due to availability of CM measurements accompanied by historical failures. The techniques implemented were:

i. *Multiple Linear Regression (MLR) and Polynomial Regression (PR) which belong to trend extrapolation, one of the simplest methods and most commonly used one in industry [17,23–25]*
ii. *K-Nearest Neighbours Regression (KNNR) which belongs to similarity-based prognostics, an emerging trend with great potential [26,27]. Moreover, a RUL estimation variation based on KNNR was proposed.*
iii. *An ensemble method averaging each of the aforementioned algorithms' output was proposed.*

These methods were benchmarked using non-uniformly sampled historical valve failure temperature data from an industrial reciprocating compressor retrieved from a server, rather than raw sensor measurements commonly used. The use of actual information addressed a major prognostics challenge: limited works utilising real-life data [7,16–21], demonstrating PHM's applicability and benefits in industry, while at the same time implementing PHM to a failure mode that is instantaneous in contrast to slowly time varying ones usually examined.

Principal Components Analysis (PCA) and Statistical Process Control (SPC) were employed to fuse the temperature measurements and create Hotelling T^2 and Q residuals metrics which are proposed to be used for the first time to reflect degradation process of a reciprocating compressor valve, and are used for RUL estimation. PCA/SPC has already found limited application in reciprocating compressors as a diagnostics tool. Ahmed et al. [28] used experimental raw sensor data, extracted features, fused them with PCA and performed detection of various faults via SPC. They further enhanced their methodology in [29] by extracting more features and utilising contribution plot of Q metric to identify features associated with faults that can assist identification. Prognostics algorithms were benchmarked while utilising these metrics as inputs.

The rest of the paper is organised as follows. Section 2 contains a literature review of prognostics methods employed. Section 3 analyses HI creation process and overviews prognostics methods. Section 4 describes data acquisition procedure and evaluation metrics used. Section 5 presents results followed by a discussion. Section 6 contains concluding remarks.

2. Prognostics methods literature review

In this work, several prognostics methods were benchmarked on reciprocating compressor valve failure data based on their accuracy and variability. The first two techniques employed (MLR and PR) belong to trend extrapolation family. It is one of the most preferred prognostics methods in industry, being the simplest one, though there are limited published works in literature [17]. Zhao et al. [30] used S-transform, Gaussian pyramid, local binary pattern, PCA and linear discriminant analysis for pre-processing along with MLR for RUL estimation for bearings. Li and Nilkitsaranont [31] employed MLR for prognostics of gas turbine engine during early degradation stage while quadratic regression was used when degradation deteriorated. Alamaniotis et al. [32] applied fuzzy sets and MLR for prognostics of power plant turbine blade. Proposed methodology was superior to simple MLR. MLR has also been used extensively as a benchmarking tool, along with PR. In such works, MLR/PR were used either to compare performance of proposed methodology, usually found inferior [33–35], or to compare performance of several algorithms [36,37]. These works used either experimental [30,33,34,36,37]or simulated [31,35] or actual [32] raw sensor data.

Another method examined was KNNR which lies in the similarity based prognostics group. In similarity based prognostics, a reference data base is created with historical failures which are compared with an ongoing case via distance analysis. Wang et al. [22], used MLR for fusion, curve fitting for smoothing, and segmented failure trajectories. RUL was estimated based on similar reference RULs by measuring distance of ongoing failure trajectory section with historical ones. Zio and Maio [24] segmented and normalised failure signals. During normal operation, RUL was estimated as Mean Time to Failure (MTTF). After fault detection, RUL was calculated as weighted sum of historical RULs based on fuzzy similarity of current segment and reference ones. They further enhanced their methodology in [38] where RUL was calculated continuously and new estimate was compared with previous ones under assumption of stationarity. In case of no significant change healthy state was considered and RUL was

replaced by MTTF. Maio and Zio [25] compared Zio and Maio's technique [24] with Monte Carlo based particle filter where it was shown computationally cheaper. Mosallam et al. [23] implemented symmetrical uncertainty method, PCA and EMD for pre-processing and segmented failure signals. RUL was estimated as most similar historical RUL based on K-nearest neighbour analysis of ongoing segment and reference ones, with discrete Bayesian filter used for uncertainty quantification. They also applied the same methodology in [39], and further enhanced it in [40] by adding GPR in RUL estimation process. Zhang et al. [41] used phase space reconstruction trajectory for pre-processing and segmented failure trajectories. RUL was estimated using weighted average of most similar historical RULs, based on distance analysis of ongoing segment and reference ones. Wang et al. [42] applied MLR for fusion, RVM for offline sparse training, estimated RUL as weighted average of historical RULs based on similarity analysis of ongoing trajectory with reference ones, and quantified uncertainty with uncertainty propagation map. Khelif et al. [43] used MLR for fusion and curve fitting for smoothing. RUL was estimated as weighted sum of most similar historical RULs based on distance analysis of current trajectory and reference ones, with most similar cases being favoured and dissimilar ones being penalised. Li et al. [27] used wavelet packet analysis for pre-processing and applied Zio and Maio's methodology [24] where they compared two membership functions which displayed similar performance. You and Meng [26] segmented historical failures. RUL of current segment was estimated based on weighted RUL of similar historical ones. During similarity analysis, more recent measurements within segment had greater importance. Xue et al. [44] estimated RUL by applying local regression on most similar historical RULs based on fuzzy instance modelling of ongoing failure and reference ones, optimised using evolutionary analysis. Lam et al. [45] applied empirical signal to noise ratio method for pre-processing, PCA for fusion, and kernel regression for smoothing. Similarity of ongoing failure with historical ones was computed using various metrics, while RUL was estimated in several ways according to similarity results. Point estimated RUL via Pearson correlation similarity metric outperformed the rest. These works used either experimental [23,25–27,41,42,44] or simulated [22,24,38–43,45] raw sensor data. Similarity based prognostics has been implemented on turbofan engines [22,39,40,42–45], fission reactor [24,38], crack propagation [25], lithium-ion batteries [23,39], bearings [41], contact resistances of electromagnetic relays [27], and ball grid array solder joints of printed circuit boards [26].

Despite its simplicity, KNNR has found limited applications regarding prognostics. Rezgui et al. [46] combined support vector regression with KNNR for diagnostics and prognostics of reverse polarity fault. Hu et al. [47] extracted features and used KNNR, optimised by particle swarm optimisation and k-fold cross validation, for RUL estimation of lithium-ion battery. Zhao et al. [48] extracted features, and used KNNR with Dempster-Shafer belief theory for RUL estimation local oscillator from an analogue circuit of a high frequency receiver. The method outperformed NN, fuzzy NN, and particle filtering. These works used either experimental [47] or simulated [46,48] data. On the other hand, KNNR has found popularity in other fields like forestry [49–51] or traffic forecasting [52–54].

3. Prognostics methods overview

3.1 Health indicator

In data-driven prognostics, data quality is of paramount importance, as it affects RUL estimation accuracy [6,9]. Hence, it is essential data used reflect degradation process adequately. This can be achieved via HIs that can be either features extracted from signals (mean, skewness, kurtosis, etc.), or one-dimensional metrics created by data fusion requiring all useful information to be considered [6,9]. In this work, PCA coupled with SPC were implemented to construct Hotelling T^2 and Q residuals metrics describing compressor's valve degradation, and proposed to be used for the first time as HIs and RUL estimation inputs.

3.1.1 Principal components analysis

PCA is a dimensionality reduction technique that projects a number of correlated variables in a lower space via a linear transformation, while preserving maximum possible variance within original set, creating a new group of uncorrelated, and orthogonal latent variables [55]. Let X be a $n \times p$ data matrix (n: number of measurements, p: number of variables), its PCA transformation is [55]:

$$X = P'T + R, \text{ Equation 1}$$

Where T, the $n \times k$ score matrix, is the projection of X from p-dimensional space to k-dimensional, with $k \leq p$. P, the $p \times k$ component matrix, is the linear mapping of X to T. R is the $n \times p$ reconstruction error matrix. Calculation of principal components can be done with use of singular value decomposition [55]. Selection of appropriate k was done employing Cumulative Percentage of Variance (CPV) [55], where k first components leading to a model capturing a predefined variance percentage are kept. A typical value is 90% [55].

3.1.2 Statistical process control

SPC is used to monitor a process for diagnostics purposes. A univariate process is considered to be healthy when its value lies within some statistical limits decided by the control chart used [56]. For a multivariate process, SPC assumptions of variable independency are inadequate. Hence, Multivariate Statistical Process Control (MSPC) is introduced, where a single control chart is created using information from all variables. A common tool used to facilitate MSPC is PCA by reducing number of monitored variables and decorrelating them. Some good overview works describing the application of PCA and MSPC can be found in [57–60].

After a PCA model has been created, its scores and residuals can be used for SPC. Control charts employed in this work are Hotelling T^2 and Q residuals; these are the most widely employed in relation to PCA/SPC [57–60]. Hotelling metric for score matrix T is [57,60,61]:

$$T^2 = \sum_{i=1}^{k} \frac{t_i}{s_i^2}, \text{ Equation 2}$$

With t_i ith principal component scores, s_i^2 its variance, and control limit [57–60]:

$$T_a^2 = \frac{k(n^2-1)}{n(n-k)} F_a(k, n-k), \text{ Equation 3}$$

With $F_a(k, n-k)$ the $(100-1)a\%$ upper critical point of F distribution with k and $n-k$ numbers of freedom.

Q metric for residual matrix R is [57,58,60]:

$$Q = \sum_{i=1}^{n} (x_i - \hat{x}_i)^2, \text{ Equation 4}$$

With $\hat{x}_i$ reconstructed values of x_i, and control limit [57,58]:

$$Q_a = g x_{h,a}^2, \text{ Equation 5}$$

Where $= \frac{var(R)}{2mean(R)}$, $h = \frac{2(var(R))^2}{var(R)}$, and $x_{h,a}^2$ the $(100-1)a\%$ upper critical point of x^2 distribution with h numbers of freedom.

Metrics created by PCA/SPC were used as HIs for prognostics purpose. Procedure of employing PCA/SPC to create HIs is described in a compact form as follows. In phase I healthy data are centred and scaled to unit variance, and PCA model is created, along with control limits for T^2 and Q. In phase II new data are centred and scaled using healthy means and variances, projected on healthy PCA model calculating their scores and residuals, and their metrics are estimated creating HIs. This process can also be seen schematically in Figure 1.

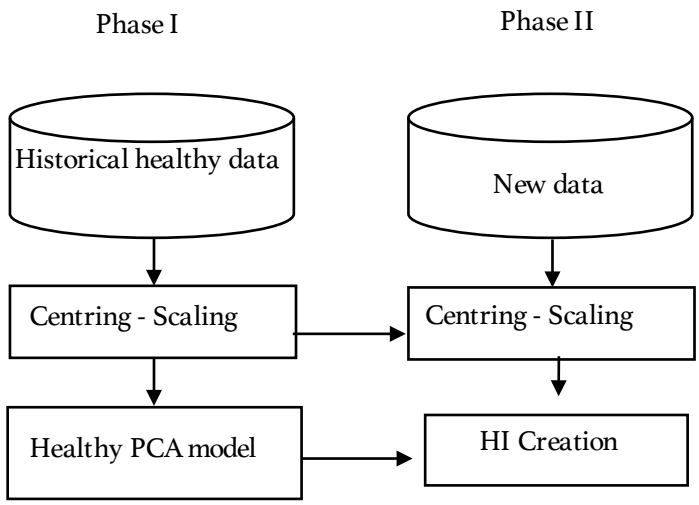


Figure 1 Health indicator creation process

3.2 Prognostics techniques

As already mentioned, there is lack of literature about prognostics on reciprocating compressors. Ergo, several prognostics methods were compared on valve failure data from an operation industrial compressor.

3.2.1 Multiple linear regression

MLR belongs to trend extrapolation family being its simplest representation. Let Y be a $n\times1$ response vector and X a $n\times p$ regressor matrix. MLR is used to predict the dependent variable as linear combination of independent ones [62]:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \text{ Equation 6}$$

With $\beta_0, \beta_1, \dots, \beta_p$ regression coefficients to be estimated, ε the residuals assumed to be uncorrelated and normally distributed, and $i = 1, \dots, n$. Parameters are calculated utilising least squares algorithm [62]:

$$\hat{\beta} = (X'X)^{-1}X'Y, \text{ Equation 7}$$

Fit of model on data can be assessed using coefficient of determination R^2 that measures amount of variability captured [62]:

$$R^2 = 1 - \frac{SS_E}{SS_T}, \text{ Equation 8}$$

with $SS_E = \sum E^2$ [62], and $SS_T = \sum(Y - \bar{Y})^2$ [62]. Another criterion is the adjusted coefficient of determination $R^2_{adjusted}$ [62]:

$$R^2_{adjusted} = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)}, \text{ Equation 9}$$

Both metrics range from zero indicating bad fit to one indicating perfect fit.

In this work MLR was trained using historical failures with HIs being independent variables and RUL dependent one, and was applied for direct RUL estimation.

3.2.2 Polynomial regression

PR also belongs to trend extrapolation class. It can be seen as an extension to MLR where predictors are also included in power form. Polynomial order depends on desired power. A second order polynomial for two regressors is [62]:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_{11} x_{i1}^2 + \beta_2 x_{i2} + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \varepsilon, \text{ Equation 10}$$

Estimation of coefficients is done as in 3.2.1.

Depending on polynomial order, number of parameters can be significantly large leading to overfitting. Stepwise regression is the most widely used selection process for including an optimum number of regressors [62]. It is an iterative procedure where terms are included or removed from the model based on a partial F-test. Considering that f_{in} is F-value for including a term and f_{out} for removing one, for a variable to be included it should be $f \geq f_{in}$ and to be excluded $f \leq f_{out}$ [62]. During the initial step, a model is constructed using only most correlated regressor with the dependent variable as it will have highest f value. The process concludes when no variables can be included or excluded [62], leading to polynomial stepwise regression (PSR). Adequacy of model can be examined using R^2 and $R^2_{adjusted}$ metrics. Prognostics application of this method is the same as for MLR.

3.2.3 K-nearest neighbours regression

KNNR is a form of similarity based prognostics, belonging in nonparametric regression family along with similarity based prognostics. It estimates the regression function without making any assumptions about underlying relationship of dependent and independent variables [50,53,63,64] by utilising similarities of current sample to historical points for prediction [54]. KKNR is a distribution free, multivariate method that preserves variable relations and local structure within data, easy to use, fast and computationally cheap [65], but highly affected by amount of historical data available [47].

Let X be a $n \times q$ regressor matrix, Y its $n \times 1$ response vector and u a new sample. Resemblance of new sample's predictors and historical ones is calculated via similarity analysis. Euclidean distance [46,49,52–54,65–68] is the most commonly used similarity metric [47,50,65]:

$$d(u, x_i) = \|u - x_i\| = \sqrt{\sum_{j=1}^{q} \left(u_j - x_{ij}\right)^2}, \text{ Equation 11}$$

with $i = 1, \dots, n$. u's response value is [47,50,65]:

$$y_u = \frac{\sum_{l=1}^{K} w_l y_l}{\sum_{l=1}^{K} w_l}, \text{ Equation 12}$$

K is the number of most similar historical points to current sample according to $d(u, x_i)$, w_l and y_l are the weight and response value of lth neighbour. Hence, the response value is the weighted sum of response values of K closest historical samples based on their predictor similarities. About weighting there is no straightforward formula and can be done in various ways [64]. Formulation used here was:

$$w_l = 1 - d_l / \sum_{l=1}^{K} d_l. \text{ Equation 13}$$

Optimum K can be found via k-fold cross validation [49,63–67]. Historical data are partitioned into k new sets of approximately equal length. For a range of Ks, a model is trained with k-1 sets, leaving one out for validation estimating an error criterion. This is repeated until all subsets are left out once creating k new models. Mean error for each K is calculated and smallest one yields optimum K [47,63,69].

Based on above, a single KNNR model is constructed using historical T^2 and Q measurements and their corresponding RUL values, in order to identify the optimum K. For an ongoing case, its metrics

are presented to the model an RUL is estimated based on similarity analysis. As with similarity based prognostics, the purpose is to identify similar degradation patterns with historical cases, but in this case rather than performing similarity analysis between segments as usually done pointwise similarity was used instead where only latest information was compared with historical samples. Furthermore, RUL was estimated using K most similar samples from all historical data, meaning that one failure might have more than one common points with current sample while another might have none. This method shall be denoted as KNNR 1.

3.2.4 Proposed variation of KNNR based RUL estimation

A variation of RUL estimation process based on KNNR is also proposed. instead of applying KNNR on all historical measurements, it is implemented on each historical case individually. RUL is calculated as the weighted sum of RULs from each separate case based on similarity results. In this variation, instead of using K most similar points from each case only most similar one was used. This variation shall be denoted as KNNR 2.

3.2.5 Ensemble technique

In this paper, an ensemble methodology was also tested where output of each prognostics algorithm reviewed in this chapter was combined. Fusion was undertaken via averaging having the following form:

$$Ensemble = (MSR + PR + SOM1 + SOM2 + KNNR1 + KNNR2)/6, \text{ Equation 14}$$

The purpose is to improve prognostics results by combining strengths of multiple techniques, refining their results.

3.3 Prognostics metrics

In order to quantitatively benchmark performance of methods several criteria were used, as there is no universal criterion available yet [36]. The metrics can be separated into two categories: a) accuracy (NMSE, MAPER, CRA) measuring distance between estimated and actual RUL with higher accuracy desired, and b) precision (MAD) measuring error variability with low volatility desired. Let $RUL(t)$ be actual RUL at time t, $t = 1, \dots, N$ number of available samples, $\widehat{RUL}(t)$ be estimated RUL, and $\Delta RUL(t) = RUL(t) - \widehat{RUL}(t)$ be difference of actual and estimated RUL. Employed metrics are:

i. *Normalised Mean Square Error (NMSE) [70]:*

$$NMSE = 1 - \frac{1}{N}\sum_{i=1}^{N}\frac{\Delta RUL(i)^2}{(RUL(i)-\overline{RUL})^2}, \text{ Equation 15}$$

With $\overline{RUL}$ the mean value of RUL.

ii. *Mean Absolute Percentage Error (MAPER) [36]:*

$$MAPER = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{100\Delta RUL(i)}{RUL(i)}\right|, \text{ Equation 16}$$

iii. *Cumulative Relative Accuracy (CRA) [36]:*

$$CRA = \frac{1}{N}\sum_{i=1}^{N} RA(i), \text{ Equation 17}$$

With $RA(i)$ the Relative Accuracy at each time instance [36,71]:

$$\mathrm{RA(i)} = 1 - \frac{|\Delta \mathrm{RUL(i)}|}{\mathrm{RUL(i)}}$$

iv. *Mean absolute deviation [36]:*

$$MAD = \frac{1}{N}\sum_{i=1}^{N}|\Delta RUL(i) - median(\Delta RUL(i))|, \text{ Equation 18}$$

NMSE and CRA range in (-∞, 1] with 1 indicating perfect score, while MAPER and MAD range in [0, ∞) with 0 indicating perfect score.

4. Data acquisition

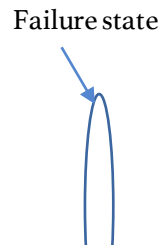
Information employed in this work came from an operational industrial two-stage, four-cylinder, double-acting reciprocating compressor that has been used in various applications (compressing different gasses). The machine is instrumented with sensors collecting both process (temperature, pressure, speed, etc.) and mechanically (bearing vibration, bearing temperature, seal pressure etc.) related measurements, that stream continuously, via internet, to a central location. They are stored, pre-processed, and analysed for CBM purposes. The large volume of data created, considering each sensor's sampling frequency, requires a huge amount of storage. Hence, a ruleset was created deciding which values should be stored, leading to non-uniformly sampled sets. To compensate, linear interpolation is used.

The fault mode under study was a valve failure. A ring valve was the defective component with cause of failure: broken valve plate leading to leakage. There were 13 defective cases available that all took place in the same cylinder within a period of one and a half years. Depending on case, the failing valve was either Head End (HE) or Crank End (CE) discharge valve. In all failures, valves were of same type, model, and manufacturer. Failure was denoted as the point when it was deemed as incapable of performing its intended function.

Historical information of 16 temperature measurements, one for each valve (two suction (HE/CE) and two discharge (HE/CE) per cylinder, four cylinders), was extracted from a server with sampling period $T_s = 1s$ *(*$f_s = 1Hz$*). Each case contained roughly two and a half days' worth of data, depicting both healthy and failing states. Table 1 contains the fault duration of each case (from the moment of detection until moment of failure). The instantaneous nature is evident as failure occurs in a matter of minutes. A graphical representation for failure 11 can be seen in Figure 2. Prior to proceeding with analysis, data were scanned for missing values, a common phenomenon in industry, utilising SOM for imputation [72]. In order to mitigate the effect of external sources like speed or ambient temperature on the measurements, their ratios were employed instead. The ratios were calculated as follows:* $r_i = HE_{kjtemp}/CE_{kjtemp}$*, where* $i = 1, \dots, 8$ *the number of ratios,* j *indicates whether the temperatures are for suction (1) or discharge (2), and* $k = 1, \dots, 4$ *the number of cylinders.*

Table 1 Failure data sets time duration

Failure Case	1	2	3	4	5	6	7	8	9	10	11	12	13
Fault Duration (s)	333	119	280	245	125	242	114	233	494	131	246	73	254



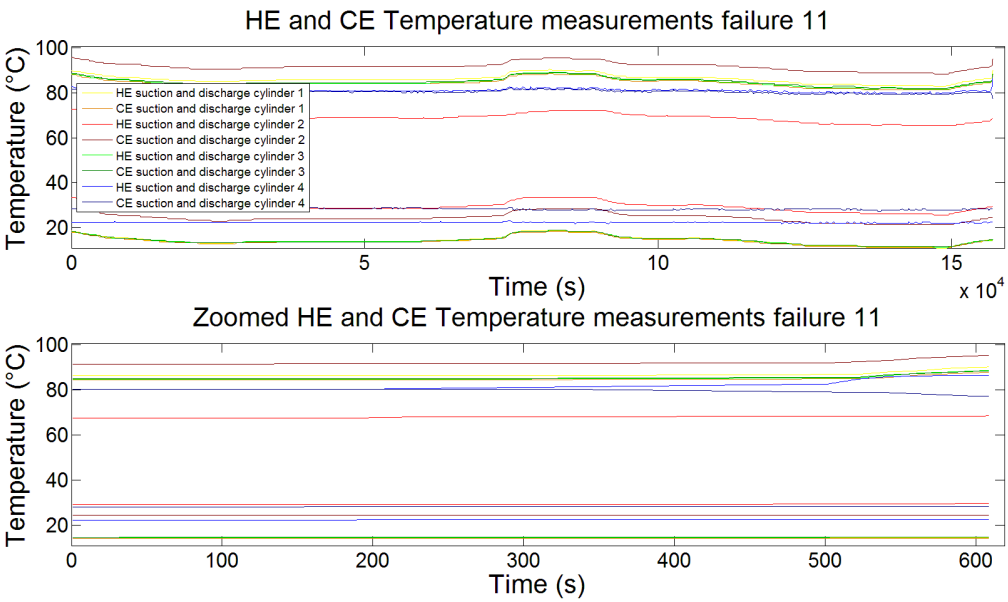


Figure 2 Failure 11 temperature measurements.

5. Results and discussion

After the process of collecting and cleansing the data was completed, they were employed in order to implement prognostics by creating health indicators and comparing the performance of several algorithms.

5.1 Health indicator

Healthy data from each case were centred and scaled to unit variance, and used to create a PCA model of 3 components ($CPV = 95\%$) while calculating T^2 and Q control limits. Failure data, after centring and scaling, were projected on the model calculating their T^2 and Q metrics creating HIs (Figure 3). Both metrics were divided with their respective statistical limits in order to be comparable.

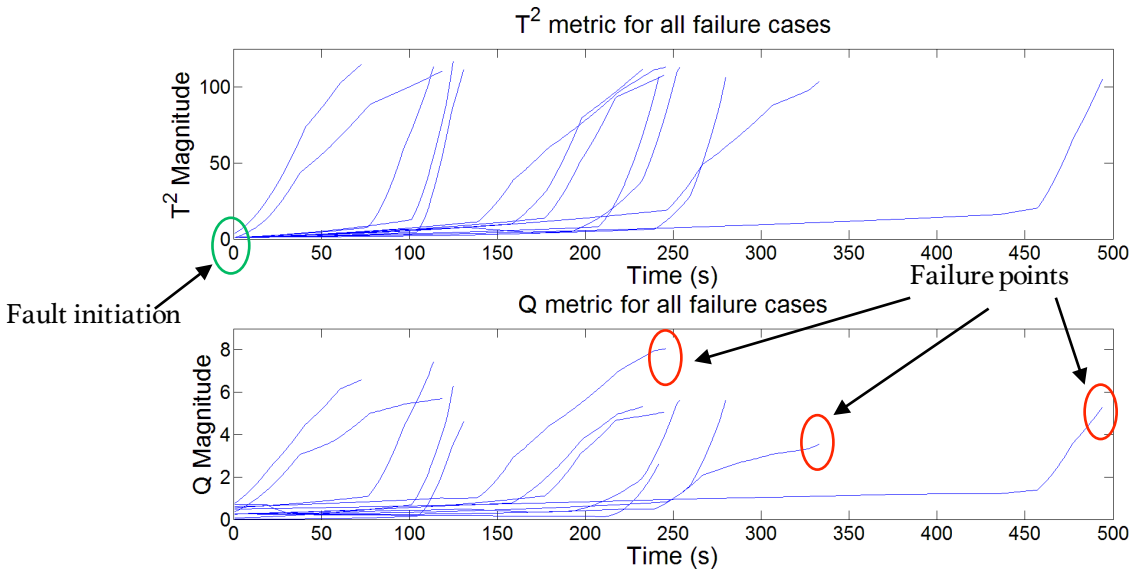


Figure 3 T^2 and Q health indicators

An appropriate HI, needs to be monotonic and encapsulate degradation evolution through time [21,23,30,34,39,40]. If this is satisfied, estimated RUL is expected to be accurate [73]. Furthermore, it is desired that HI is of low variability [30,34], present roughly same value during failure under same failure mode and operating conditions, and have resembling pattern [21,30]. Figure 3 confirms suitability of both metrics by fulfilling aforementioned perquisites, adequately reflecting fault propagation.

5.2 Prognostics methods configuration

RUL was estimated directly with T^2 and Q being independent variables and RUL dependent one. Prognostics methods attempted to model this relationship so that RUL could be calculated accurately. RUL was logarithmically transformed to improve model fit. During training, 12 cases were used for model building while the 13th was kept for testing. Results for cases 8 and 11 are presented. Training outcome of each method can be found below. All methods were implemented in Matlab [74–76].

Table 2 contains R^2 and $R^2_{adjusted}$ metrics for MLR and PR modes. For PR, third order was maximum order examined. Both methods have an adequate fit with PR being superior having greater values.

Table 2 MLR/PR R^2 and $R^2_{adjusted}$ metrics

Algorithm	MLR		PR	
Failure case	8	11	8	11
R^2	0,68	0,67	0,76	0.77
$R^2_{adjusted}$	0,67	0,67	0,76	0,77

Table 3 contains optimum K for KNNR 1, selected via 10-fold cross validation ranging from 1 to 200, while for KNNR 2 optimum K was decided a priori as K=1. Data were centred and scaled to unity, and Euclidean distance was used.

Table 3 Optimum K for KNNR 1

Failure case	8	11
k	15	14

5.3 Prognostics results

Figure 4, Figure 5, and Figure 6 contain prognostics results for both historical failures (8 and 11), giving a qualitative perspective of each method's performance. X-axis indicates time while y-axis RUL at each time stamp, with $t = 0$ the moment fault was detected (RUL=233 case 8, 246 case 11, and 494 case 9, Table 1) and $t = 233$, or $t = 246$, or $t = 494$ the moment of failure (RUL=0). Graphs consist of a number of lines. Black indicates actual RUL through time, as observed in-situ, and rest correspond to each algorithm's estimations. In failure cases 8 and 11 all methods perform comparably well with best performing being the ensemble technique (magenta line) as it tracks closely RUL evolution in both cases, followed by polynomial regression (continuous blue line), while worst performing seems to be MLR (continuous blue line) which struggles to follow actual RUL. KNNR 1 (dashed green line) and PR (continuous cyan line) perform adequately, while KKNR 2 (continuous green line) underestimate RUL. It can be noted that all methods converge to actual RUL as time passes. For failure 9, the methods employed estimate the RUL relatively less successfully whereas the ensemble techniques, PR, and KNNR 2 demonstrate better prediction. It should be noted that all methods eventually converge to the actual RUL.

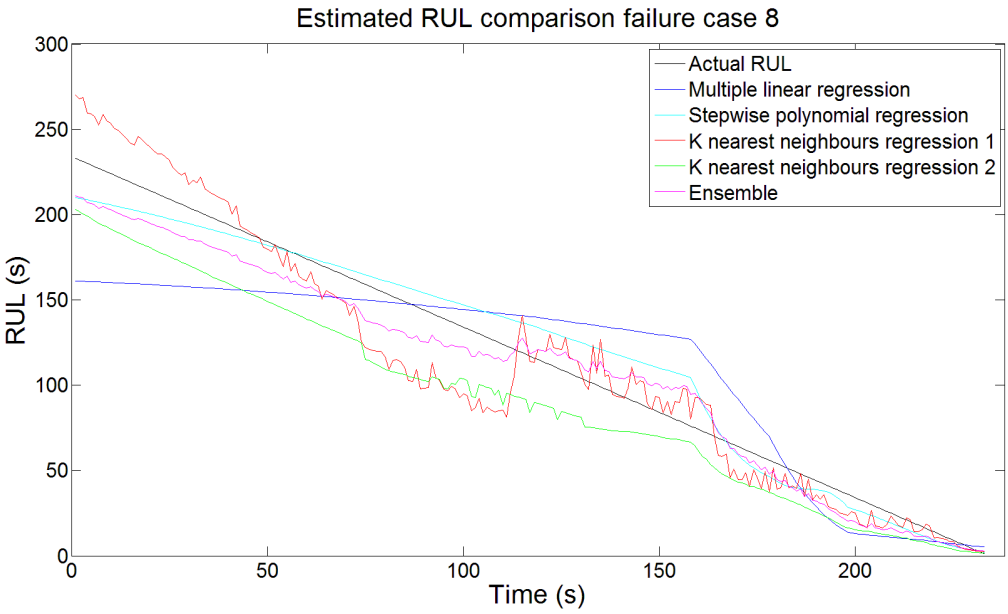


Figure 4 RUL estimation for failure case 8

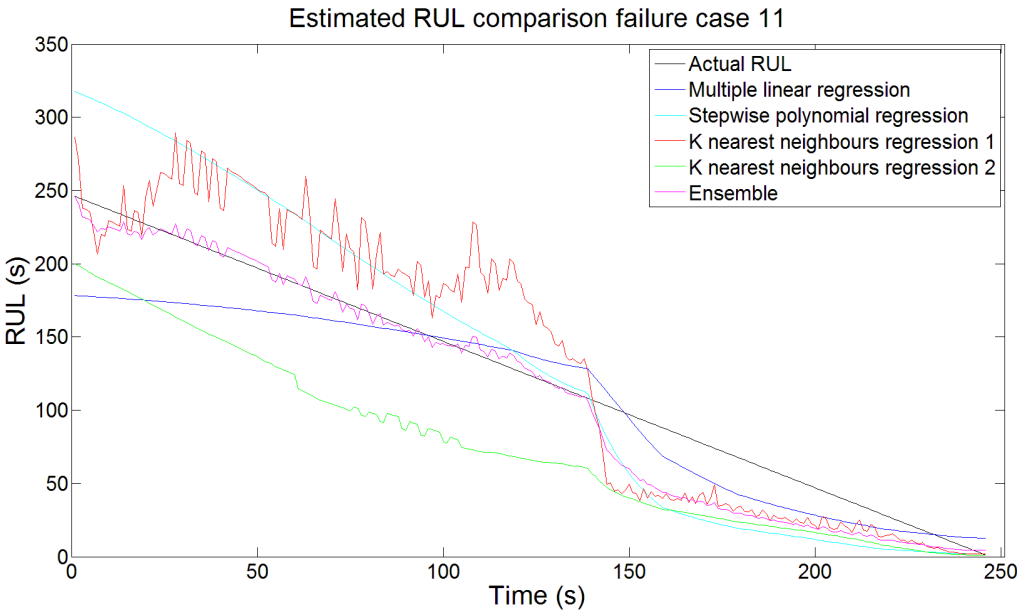


Figure 5 RUL estimation for failure case 11

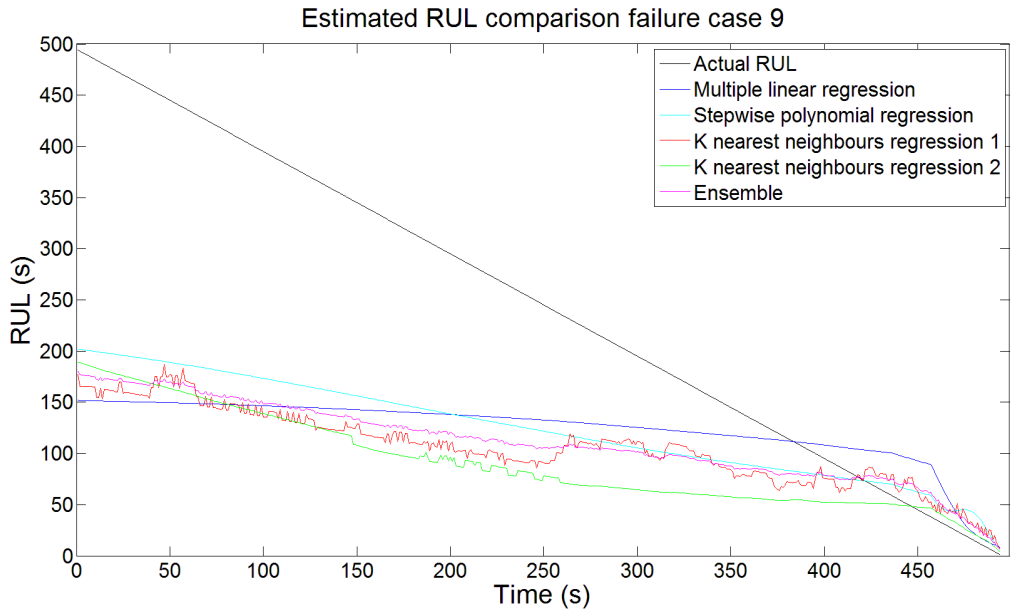


Figure 6 RUL estimation for failure case 9

Quantitative inspection of methods' performance can be done via metrics found in Table 4. The prognostics horizon for all metrics is from moment of fault detection until failure, meaning all available samples were considered in calculation. From results, it is evident that ensemble method consistently outperforms the rest being superior in most metrics for both failures, while in cases where another technique prevails, ensemble follows closely. Although PR performs well in case 8, in 11 it is outperformed by others. KNNR 2 displays, lowest variability, an attribute highly desired, followed closely by ensemble one. Overall, quantitative results are in accordance with qualitative ones. Furthermore, results confirmed the claim of lack of universal metrics since the same method might be suitable or not depending on metric used. This calls for more effort to be put towards this direction.

Table 4 Evaluation metrics

Performance metrics	NMSE			MAPER			CRA			MAD		
Failure case	8	11	9	8	11	9	8	11	9	8	11	9
Method												
MLR	0.77	0.86	0.09	30.50	29.39	53.51	0.70	0.71	0.46	26.10	16.90	103.58
PR	**0.96**	0.66	**0.12**	**14.17**	40.14	**52.72**	**0.86**	0.60	0.47	11.3	35.61	**83**
KNNR 1	0.90	0.71	0.05	16.94	35.53	58.53	0.83	0.65	0.41	16.43	33.53	97.79
KNNR 2	0.83	0.46	0.02	27.96	49.13	59.05	0.72	0.51	0.41	**9.33**	**14.36**	87.06
ENSEMBLE	**0.96**	**0.92**	0.11	16.14	**25.17**	52.86	0.84	**0.75**	**0.48**	9.35	15.21	85.29

Based on prognostics results presented in this section, there are some comments that can be made:

- *PSR and MLR performed similarly well with PSR being superior based on both qualitative and quantitative results, as it could better reflect the complex relationship between RUL and HIs by including interaction and higher order terms of HIs, overcoming MLR's rigidness.*
- *KNNR 1 and KNNR 2 displayed similar performance, both belonging to similarity based prognostics family and using the same distance metric (Euclidean) for similarity analysis.*
- *KNNR 1 tended to outperform KNNR 2 indicating that even when considering more than one similar case from the same failure during RUL estimation can increase accuracy. On the other*

hand, KNNR 2 displayed lower variation hinting that considering each case separately can reduce volatility.

- *Ensemble method's performance is highly dependable on individual performance of compromising methods. Its components performed well thus it displayed the best overall performance based on both qualitative and quantitative results. Its output could be seen as refinement of prognostics estimations of its elements.*
- *In relation to the relative underperformance of prognostic methods in failure 9, this can be attributed to the fact that information used for this investigation was obtained from an operational industrial compressor that had been running under various working conditions and most importantly, for different applications (compressing gases with different composition) during the time span the data was made available (one and a half years). The deterioration of valve health is highly dependent on operating environment. In cases where the machine compressed gas that was impure or had high corrosive properties, the valves would degrade at a faster rate. As a result, different algorithms may capture and better represent the relationship of RUL and HIs depending on the case.*

Importance of HI quality should also be noted, as performance of algorithms is heavily dependent on quality of HIs used since they reflect degradation process. The HIs that were used (T^2 and Q) encapsulated adequately failure evolution confirmed by good results, tracking closely fault propagation through time and enabling successful prognostics on an instantaneous failure mode.

6. Conclusions

In this work, three prognostics techniques (MLR, PR, and KNNR 1), along with a RUL estimation variations (KNNR 2), and an ensemble method combining aforementioned algorithms' output, were benchmarked using valve failure data from an operational industrial reciprocating compressor. To the authors' knowledge this was the first attempt of RUL estimation on reciprocating compressor valves. Furthermore, use of actual data addressed lack of works regarding implementation of prognostics in industrial applications demonstrating PHM's potency. Moreover, it was the first known implementation of that T^2 and Q metrics as HIs and utilised in direct RUL estimation process.

Results showed that all methods performed comparably well both in qualitative (graphs) and quantitative (metrics) analysis, with ensemble outperforming the rest by better tracking RUL evolution and having high metric values. PR was shown to be superior to PR while KNNR 1, and KNNR 2 performed closely being both similarity based methods using the same distance metric, with KNNR 1 performing the best. Also, quality of HIs used was deemed satisfactory given good results of techniques, confirming suitability of T^2 and Q metrics to be used as such. Moreover, results demonstrated that all methods were able to cope with instantaneous nature of failure mode under study.

References

[1] H.P. Bloch, A Practical Guide to Compressor Technology, Second, John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

[2] H.P. Bloch, J.J. Hoefner, Reciprocating Compressors: Operation & Maintenance, Butterworth-Heinemann, 1996. http://books.google.co.uk/books/about/Reciprocating_Compressors.html?id=VreR8GlfCCIC&pgis=1.

[3] V.T. Tran, F. AlThobiani, A. Ball, An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks, Expert Syst. Appl. 41 (2014) 4113–4122. doi:10.1016/j.eswa.2013.12.026.

[4] W.A. Griffith, E.B. Flanagan, Online continuous monitoring of mechanical condition and performance for critical reciprocating compressors, in: 30th Turbomach. Symp. Texas Houst., 2001.

[5] Keerqinhu, G. Qi, W.-T. Tsai, Y. Hong, W. Wang, G. Hou, Z. Zhu, Fault-Diagnosis for Reciprocating Compressors Using Big Data, 2016 IEEE Second Int. Conf. Big Data Comput. Serv. Appl. (2016) 72–81. doi:10.1109/BigDataService.2016.27.

[6] G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, B. Wu, Intelligent Fault Diagnosis and Prognosis for Engineering Systems, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2006. doi:10.1002/9780470117842.

[7] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mech. Syst. Signal Process. 20 (2006) 1483–1510. doi:10.1016/j.ymssp.2005.09.012.

[8] A.J. Guillén, A. Crespo, M. Macchi, J. Gómez, On the role of Prognostics and Health Management in advanced maintenance systems, Prod. Plan. Control. 7287 (2016) 1–14. doi:10.1080/09537287.2016.1171920.

[9] J. Yan, Machinery Prognostics and Prognosis Oriented Maintenance Management, John Wiley & Sons Singapore Pte. Ltd, Singapore, 2014. doi:10.1002/9781118638743.

[10] S. Kadry, Diagnostics and Prognostics of Engineering Systems, IGI Global, 2013. doi:10.4018/978-1-4666-2095-7.

[11] B.-S. Yang, W.-W. Hwang, D.-J. Kim, A. Chit Tan, Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines, Mech. Syst. Signal Process. 19 (2005) 371–390. doi:10.1016/j.ymssp.2004.06.002.

[12] C. Annicchiarico, A. Babbini, R. Capitani, P. Tozzi, Numerical and Experimental Testing of Composite Rings for Reciprocating Compressor Valves, in: ASME 2013 Press. Vessel. Pip. Conf., ASME, 2013. doi:10.1115/PVP2013-97984.

[13] F. Gu, Y. Shao, N. Hu, A. Naid, A.D. Ball, Electrical motor current signal analysis using a modified bispectrum for fault diagnosis of downstream mechanical equipment, Mech. Syst. Signal Process. 25 (2011) 360–372. doi:10.1016/j.ymssp.2010.07.004.

[14] H. Cui, L. Zhang, R. Kang, X. Lan, Research on fault diagnosis for reciprocating compressor valve using information entropy and SVM method, J. Loss Prev. Process Ind. 22 (2009) 864–867. doi:10.1016/j.jlp.2009.08.012.

[15] K. Feng, Z. Jiang, W. He, B. Ma, A recognition and novelty detection approach based on Curvelet transform, nonlinear PCA and SVM with application to indicator diagram diagnosis, Expert Syst. Appl. 38 (2011) 12721–12729. doi:10.1016/j.eswa.2011.04.060.

[16] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, D. Siegel, Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications, Mech. Syst. Signal Process. 42 (2014) 314–334. doi:10.1016/j.ymssp.2013.06.004.

[17] J.Z. Sikorska, M. Hodkiewicz, L. Ma, Prognostic modelling options for remaining useful life estimation by industry, Mech. Syst. Signal Process. 25 (2011) 1803–1836. doi:10.1016/j.ymssp.2010.11.018.

[18] B. Sun, S. Zeng, R. Kang, M.G. Pecht, Benefits and challenges of system prognostics, IEEE Trans. Reliab. 61 (2012) 323–335. doi:10.1109/TR.2012.2194173.

[19] A. Heng, S. Zhang, A.C.C. Tan, J. Mathew, Rotating machinery prognostics: State of the art, challenges and opportunities, Mech. Syst. Signal Process. 23 (2009) 724–739. doi:10.1016/j.ymssp.2008.06.009.

[20] M.S. Kan, A.C.C. Tan, J. Mathew, A review on prognostic techniques for non-stationary and non-linear rotating systems, Mech. Syst. Signal Process. 62–63 (2015) 1–20. doi:10.1016/j.ymssp.2015.02.016.

[21] E. Zio, Prognostics and Health Management of Industrial Equipment, in: S. Kadry (Ed.), Diagnostics Progn. Eng. Syst., IGI Global, 2013: pp. 333–356. doi:10.4018/978-1-4666-2095-7.ch017.

[22] T. Wang, Jianbo Yu, D. Siegel, J. Lee, A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems, in: 2008 Int. Conf. Progn. Heal. Manag., IEEE, 2008: pp. 1–6. doi:10.1109/PHM.2008.4711421.

[23] A. Mosallam, K. Medjaher, N. Zerhouni, Bayesian approach for remaining useful life prediction, Chem. Eng. Trans. 33 (2013) 139–144. doi:10.3303/CET1333024.

[24] E. Zio, F. Di Maio, A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system, Reliab. Eng. Syst. Saf. 95 (2010) 49–57. doi:10.1016/j.ress.2009.08.001.

[25] F. DI MAIO, E. ZIO, Failure prognostics by a data-driven similarity-based approach, Int. J. Reliab. Qual. Saf. Eng. 20 (2013). doi:10.1142/S0218539313500010.
[26] M.-Y. You, G. Meng, A generalized similarity measure for similarity-based residual life prediction, Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng. 225 (2011) 151–160. doi:10.1177/0954408911399832.
[27] L.L. Li, D.J. Ma, Z.G. Li, Residual Useful Life Estimation by a Data-Driven Similarity-Based Approach, Qual. Reliab. Eng. Int. (2016). doi:10.1002/qre.2001.
[28] M. Ahmed, F. Gu, A.D. Ball, Fault Detection of Reciprocating Compressors using a Model from Principles Component Analysis of Vibrations, J. Phys. Conf. Ser. 364 (2012). doi:10.1088/1742-6596/364/1/012133.
[29] M. Ahmed, M. Baqqar, F. Gu, A.D. Ball, Fault detection and diagnosis using Principal Component Analysis of vibration data from a reciprocating compressor, in: 2012 UKACC Int. Conf. Control, IEEE, 2012: pp. 461–466. doi:10.1109/CONTROL.2012.6334674.
[30] M. Zhao, B. Tang, Q. Tan, Bearing remaining useful life estimation based on time–frequency representation and supervised dimensionality reduction, Measurement. 86 (2016) 41–55. doi:10.1016/j.measurement.2015.11.047.
[31] Y.G. Li, P. Nilkitsaranont, Gas turbine performance prognostic for condition-based maintenance, Appl. Energy. 86 (2009) 2152–2161. doi:10.1016/j.apenergy.2009.02.011.
[32] M. Alamaniotis, A. Grelle, L.H. Tsoukalas, Regression to fuzziness method for estimation of remaining useful life in power plant components, Mech. Syst. Signal Process. 48 (2014) 188–198. doi:10.1016/j.ymssp.2014.02.014.
[33] Y. Xing, E.W.M. Ma, K.L. Tsui, M. Pecht, An ensemble model for predicting the remaining useful performance of lithium-ion batteries, Microelectron. Reliab. 53 (2013) 811–820. doi:10.1016/j.microrel.2012.12.003.
[34] T.H. Loutas, D. Roulias, G. Georgoulas, Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic E-support vectors regression, IEEE Trans. Reliab. 62 (2013) 821–832. doi:10.1109/TR.2013.2285318.
[35] F. Lasheras, P. Nieto, F. de Cos Juez, R. Bayón, V. Suárez, A Hybrid PCA-CART-MARS-Based Prognostic Approach of the Remaining Useful Life for Aircraft Engines, Sensors. 15 (2015) 7062–7083. doi:10.3390/s150307062.
[36] A. Saxena, J. Celaya, B. Saha, S. Saha, K. Goebel, Evaluating algorithm performance metrics tailored for prognostics, IEEE Aerosp. Conf. Proc. (2009). doi:10.1109/AERO.2009.4839666.
[37] M.A.A. Wahab, M.M. Hamada, A. Mohamed, Artificial neural network and non-linear models for prediction of transformer oil residual operating time, Electr. Power Syst. Res. 81 (2011) 219–227. doi:10.1016/j.epsr.2010.08.010.
[38] E. Zio, F. Di Maio, A fuzzy similarity-based method for failure detection and recovery time estimation, Int. J. Performability Eng. 6 (2010) 407–424.
[39] A. Mosallam, K. Medjaher, N. Zerhouni, Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction, J. Intell. Manuf. (2014). doi:10.1007/s10845-014-0933-4.
[40] A. Mosallam, K. Medjaher, N. Zerhouni, Component based data-driven prognostics for complex systems: Methodology and applications, in: 2015 First Int. Conf. Reliab. Syst. Eng., IEEE, 2015: pp. 1–7. doi:10.1109/ICRSE.2015.7366504.
[41] Q. Zhang, P.W.-T. Tse, X. Wan, G. Xu, Remaining useful life estimation for mechanical systems based on similarity of phase space trajectory, Expert Syst. Appl. 42 (2015) 2353–2360. doi:10.1016/j.eswa.2014.10.041.
[42] P. Wang, B.D. Youn, C. Hu, A generic probabilistic framework for structural health prognostics and uncertainty management, Mech. Syst. Signal Process. 28 (2012) 622–637. doi:10.1016/j.ymssp.2011.10.019.
[43] R. Khelif, S. Malinowski, B. Chebel-Morello, N. Zerhouni, RUL prediction based on a new similarity-instance based approach, in: 2014 IEEE 23rd Int. Symp. Ind. Electron., IEEE, 2014: pp. 2463–2468. doi:10.1109/ISIE.2014.6865006.
[44] F. Xue, P. Bonissone, A. Varma, W. Yan, N. Eklund, K. Goebel, An Instance-Based Method for Remaining Useful Life Estimation for Aircraft Engines, J. Fail. Anal. Prev. 8 (2008) 199–206. doi:10.1007/s11668-008-9118-9.
[45] J. Lam, S. Sankararaman, B. Stewart, Enhanced Trajectory Based Similarity Prediction with

Uncertainty Quantification, in: PHM 2014 - Proc. Annu. Conf. Progn. Heal. Manag. Soc. 2014, 2013: pp. 623–634.

[46] *W. Rezgui, N.K. Mouss, L.-H. Mouss, M.D. Mouss, M. Benbouzid, A regression algorithm for the smart prognosis of a reversed polarity fault in a photovoltaic generator, in: 2014 First Int. Conf. Green Energy ICGE 2014, IEEE, 2014: pp. 134–138. doi:10.1109/ICGE.2014.6835411.*

[47] *C. Hu, G. Jain, P. Zhang, C. Schmidt, P. Gomadam, T. Gorka, Data-driven method based on particle swarm optimization and k-nearest neighbor regression for estimating capacity of lithium-ion battery, Appl. Energy. 129 (2014) 49–55. doi:10.1016/j.apenergy.2014.04.077.*

[48] *Jianguang Zhao, Hongbo Li, Fanjing Zeng, Tiefeng Li, Prognostics of high frequency receiver based on evidential regression, in: Proc. IEEE 2012 Progn. Syst. Heal. Manag. Conf. (PHM-2012 Beijing), IEEE, 2012: pp. 1–5. doi:10.1109/PHM.2012.6228858.*

[49] *G. Chirici, A. Barbati, P. Corona, M. Marchetti, D. Travaglini, F. Maselli, R. Bertini, Non-parametric and parametric methods using satellite images for estimating growing stock volume in alpine and Mediterranean forest ecosystems, Remote Sens. Environ. 112 (2008) 2686–2700. doi:10.1016/j.rse.2008.01.002.*

[50] *A. Haara, A. Kangas, Comparing k nearest neighbours methods and linear regression – is there reason to select one over the other?, Math. Comput. For. Nat. Sci. 4 (2012) 50–65. http://mcfns.com/index.php/Journal/article/view/MCFNS.4%3A50.*

[51] *R.E. McRoberts, E.O. Tomppo, A.O. Finley, J. Heikkinen, Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery, Remote Sens. Environ. 111 (2007) 466–480. doi:10.1016/j.rse.2007.04.002.*

[52] *S. Li, Z. Shen, G. Xiong, A k-nearest neighbor locally weighted regression method for short-term traffic flow forecasting, in: 2012 15th Int. IEEE Conf. Intell. Transp. Syst., IEEE, 2012: pp. 1596–1601. doi:10.1109/ITSC.2012.6338648.*

[53] *Tao Zhang, Lifang Hu, Zhixin Liu, Yuejie Zhang, Nonparametric regression for the short-term traffic flow forecasting, in: 2010 Int. Conf. Mech. Autom. Control Eng., IEEE, 2010: pp. 2850–2853. doi:10.1109/MACE.2010.5536292.*

[54] *Z.-W. Yuan, Y.-H. Wang, Research on K Nearest Neighbor Non-parametric Regression Algorithm Based on KD-Tree and Clustering Analysis, in: 2012 Fourth Int. Conf. Comput. Inf. Sci., IEEE, 2012: pp. 298–301. doi:10.1109/ICCIS.2012.246.*

[55] *I.T. Jolliffe, Principal Component Analysis, Second, Springer-Verlag, New York, 2002. doi:10.1007/b98835.*

[56] *S. Bersimis, S. Psarakis, J. Panaretos, Multivariate statistical process control charts: an overview, Qual. Reliab. Eng. Int. 23 (2007) 517–543. doi:10.1002/qre.829.*

[57] *T. Kourti, Application of latent variable methods to process control and multivariate statistical process control in industry, Int. J. Adapt. Control Signal Process. 19 (2005) 213–246. doi:10.1002/acs.859.*

[58] *P. Nomikos, J.F. MacGregor, Multivariate SPC Charts for Monitoring Batch Processes, Technometrics. 37 (1995) 41–59. doi:10.1080/00401706.1995.10485888.*

[59] *B.D.E. Ketelaere, M.I.A. Hubert, E. Schmitt, Overview of PCA-Based Statistical Process-Monitoring Methods for Time-Dependent , High-Dimensional Data, J. Qual. Technol. 47 (2015) 318–335.*

[60] *U. Kruger, L. Xie, Statistical Monitoring of Complex Multivariate Processes, John Wiley & Sons, Ltd, Chichester, UK, 2012. doi:10.1002/9780470517253.*

[61] *T. Kourti, J.F.J.F. MacGregor, Process analysis, monitoring and diagnosis, using multivariate projection methods, Chemom. Intell. Lab. Syst. 28 (1995) 3–21. doi:10.1016/0169-7439(95)80036-9.*

[62] *D.C. Montgomery, G.C. Runger, Applied Statistics and Probability for Engineers, Third, John Wiley & Sons, Inc., 2003.*

[63] *L. Györfi, M. Kohler, A. Krzyżak, H. Walk, A Distribution-Free Theory of Nonparametric Regression, Springer New York, New York, NY, 2002. doi:10.1007/b97848.*

[64] *N.S. Altman, An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression, Am. Stat. 46 (1992) 175–185. doi:10.1080/00031305.1992.10475879.*

[65] *J.M. Ver Hoef, H. Temesgen, A Comparison of the Spatial Linear Model to Nearest Neighbor (k-NN) Methods for Forestry Applications, PLoS One. 8 (2013) e59129. doi:10.1371/journal.pone.0059129.*

[66] *X. Tian, Z. Su, E. Chen, Z. Li, C. van der Tol, J. Guo, Q. He, Estimation of forest above-ground*

biomass using multi-parameter remote sensing data over a cold and arid area, Int. J. Appl. Earth Obs. Geoinf. 14 (2012) 160–168. doi:10.1016/j.jag.2011.09.010.

[67] H. Gu, L. Dai, G. Wu, D. Xu, S. Wang, H. Wang, Estimation of forest volumes by integrating Landsat TM imagery and forest inventory data, Sci. China Ser. E Technol. Sci. 49 (2006) 54–62. doi:10.1007/s11431-006-8107-z.

[68] H. Sun, H. Liu, H. Xiao, R. He, B. Ran, Short term traffic forecasting using the local linear regression model, Transp. Res. Rec. (2003) 143–150. http://www.its.uci.edu/its/publications/papers/CTSS/UCI-ITS-TS-WP-02-5.pdf.

[69] S. Arlot, A. Celisse, A survey of cross-validation procedures for model selection, Stat. Surv. 4 (2010) 40–79. doi:10.1214/09-SS054.

[70] L. Ljung, System Identification Toolbox ™ User â€™ s Guide, The MathWorks, Inc., 2015.

[71] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, M. Schwabacher, Metrics for evaluating performance of prognostic techniques, in: 2008 Int. Conf. Progn. Heal. Manag., IEEE, 2008: pp. 1–17. doi:10.1109/PHM.2008.4711436.

[72] L. Folguera, J. Zupan, D. Cicerone, J.F. Magallanes, Self-organizing maps for imputation of missing data in incomplete data matrices, Chemom. Intell. Lab. Syst. 143 (2015) 146–151. doi:10.1016/j.chemolab.2015.03.002.

[73] Fang Qian, Gang Niu, Remaining useful life prediction using ranking mutual information based monotonic health indicator, in: 2015 Progn. Syst. Heal. Manag. Conf., IEEE, 2015: pp. 1–5. doi:10.1109/PHM.2015.7380042.

[74] Statistics and Machine Learning Toolbox User's Guide, The MathWorks, Inc., 2016.

[75] SOM Toolbox 2.0, (n.d.). http://www.cis.hut.fi/projects/somtoolbox/.

[76] Bioinformatics Toolbox ™ User â€™ s Guide, The MathWorks, Inc., 2015.

Cranfield University

CERES https://dspace.lib.cranfield.ac.uk

School of Aerospace, Transport and Manufacturing (SATM) Staff publications (SATM)

2017-12-14

Reciprocating compressor prognostics of an instantaneous failure mode utilising temperature only measurements

Loukopoulos, Panagiotis

Elsevier

Panagiotis Loukopoulos, George Zolkiewski, Ian Bennett, et al., Reciprocating compressor prognostics of an instantaneous failure mode utilising temperature only measurements, Applied Acoustics, Volume 147, April 2019, Pages 77-86

https://doi.org/10.1016/j.apacoust.2017.12.003

Downloaded from Cranfield Library Services E-Repository