Predictive maintenance solution for industrial systems - an unsupervised approach based on log periodic power laws

Bogdan Łobodziński [a,*]

[a] *Burckhardt Compression AG, Franz-Burckhardt-Strasse 5, P.O. Box 3352, CH-8404 Winterthur, Switzerland*
E-mail: bogdan.lobodzinski@burckhardtcompression.com

Abstract. A new unsupervised predictive maintenance analysis method based on the renormalization group approach used to discover critical behavior in complex systems has been proposed. The algorithm analyzes univariate time series and detects critical points based on a newly proposed theorem that identifies critical points using a Log Periodic Power Law function fit. Application of a new algorithm for predictive maintenance analysis of industrial data collected from reciprocating compressor systems is presented. Based on the knowledge of the dynamics of the analyzed compressor system, the proposed algorithm predicts valve and piston rod seal failures well in advance.

Keywords: Failure Prediction, Predictive Maintenance, Unsupervised Analysis, Time Series, Log Periodic Power Law

1. Introduction

Detecting the symptoms of a failure and predicting when it will occur, in multivariate or univariate time series collected via the Internet of Things (IoT), is central to the concept of predictive maintenance (PM), which is now used in almost every area of industry. PM allows a company to better prepare for a potential failure by redesigning the production process in advance or creating a workaround when shutdown is not possible. Thus minimizing the costs and the effort of standard maintenance operations through predictive engineering.

Predicting failures, provided by PM, can be very profitable for a company, under the condition that PM minimizes the number of false warnings (false positives) and maximizes the number of correctly predicted events (true positives).

Let us define failure as the termination of the ability of a part of an industrial device to perform the required function with the required efficiency. Another aspect of a failure that needs to be specified is how it can manifest itself. A failure can develop gradually or appear as a sudden event. In this work we consider only those failures that develop over time.

Creating a properly working PM process faces two main problems:

(1) related to the determination of time when a failure occurred in the considered technological process or IoT data,

---

*Corresponding author. E-mail: bogdan.lobodzinski@burckhardtcompression.com.

(2) the development or the application of the best algorithm (based on physical description, machine learning, or statistical methods) to the data being analyzed.

Proper identification of the failure in IoT data (data labeling process) is a very difficult task and usually requires very specialized knowledge. For various reasons (economic, technological), not every failure requires corrective action. This is because the required efficiency that defines a failure in the monitored unit may depend on the degree of reliance of the process on the considered failure. Sometimes, a minor failure is not a sufficient reason to stop a monitored machinery or a production process. The determination of so-called good IoT data, i.e., the data when the monitored device behaves normally, is also a data labeling process that requires expertise. In this work, every action that defines the state of the data is considered a data labeling process.

Attempts to build a solution based on supervised Machine Learning (ML) models will encounter the following problems:

(1) the complexity of the data labeling process,
(2) the model degradation: due to the evolution of IoT data, it is necessary to update the supervised model because the current one starts to behave incorrectly. Such an update forces the creation of new training data necessary for a new supervised model (new data labeling process),
(3) the complex process of monitoring the predictions of the currently working model. This is a difficult task, especially when dealing with a small number of predicted failures [1].

In published works on failure prediction using supervised solutions [2], the authors omit the need for a continuous data labeling process, as well as the need to monitor the degradation of the running model. Thus, an important economic part of PM projects based on supervised models is passed over in silence. In review articles [2], methods based on Neural Network architecture are placed in the category of unsupervised methods. Hoever, these methods also require labeled training data. In this article, the term "unsupervised method" is understood only as a solution that does not require any knowledge of labeled data and any information about the state of the input data as good/healthy or anomalous.

The complex and expensive cycle of building and implementing a supervised method for a predictive maintenance process, sketched above, strongly suggests that unsupervised approaches are a better tool for building PM processes. An unsupervised solution removes the necessity of:

(1) a complex data labeling process running at all times as a part of a supervised model,
(2) monitoring of the deterioration of the model quality and thus re-creating models by training them on updated training data.

This paper describes an unsupervised failure prediction method [3] used to monitor reciprocating compressor systems based on a concept called Log-Periodic Power Law (LPPL) proposed in [4] and [5]. The LPPL mechanism is used to predict disasters in financial data [6], earthquakes [7], avalanches and similar phenomena [8].

However, due to the different nature of the data analyzed, the LPPL method cannot be applied in the same way as for bubble (or anti-bubble) detection in economic time series. Due to the different definitions of PM failure in industrial applications, modifications need to be made to the LPPL method. In a case of data describing the financial market, the variable that directly characterizes changes in the system under analysis is examined. Such a variable is, for example, the index of the analyzed financial market or the prices of shares.

In the case of a machine (e.g. compressor or other systems containing a number of subsystems), the analysis uses *indirect* data collected by sensors. It is not possible to measure the direct changes causing

the failure (e.g. material degradation, cracks etc.) but it is possible to measure changes resulting from the influence of deteriorating machine components on the measured values.

In the description of the application of the presented method, such indirect data are, for instance, changes in the opening angle of suction or discharge valves inside the compression chamber as a function of the volume in the cylinder chamber expressed by the angle of rotation of the crankshaft. In other words, degradative (unmonitored) changes affect measured variable changes in a less visible (more distorted) ways than in case of a direct measurement of variables.

The principle of operation of the proposed method is based on the detection of such behavior of the analyzed data, which is in accordance with the logarithmic periodic power law. Finding such a functional behavior in the data is equivalent to determining the corresponding point of change of the trend in the time series. The diagnosed trend is a signature of future failure. In turn, the identified point of change of the trend defines the beginning of the countdown to the predicted failure. Let us call the trend change point the *initial breakdown* (IB). The concept of detecting an IB point for use in predicting failure is similar to the concept of detecting a trend change point in a time series. Therefore, it is not necessary to predict the time of failure in the future. All that is needed is to determine whether a given point in the time series is or is not an IB starting point. If it is known that the current point in the analyzed time series is an IB point, then based on the knowledge of the dynamics of the monitored device, it is possible to determine the time window in the future (in units of the device's operating time) in which the failure will occur.

It should be emphasized that the determination of the IB point is not equivalent to the detection of anomalies. Anomalies in the behavior of the unit should be expected only in the time window of the expected failure. The concept of a time window in the future during which a failure is predicted is due to the fact that it is not possible to predict the exact time of failure of the monitored device using the presented method. This is mainly due to unpredictable events, such as: changes in the load of the device, periods when the device is off or independent maintenance activities.

The organization of the paper is as follows. Chapter 2 briefly describes the current status of unsupervised predictive maintenance in the industry. Chapter 3 answers the question of why the logarithmic-periodic oscillation detection method is suitable for detecting sudden failures in an industrial system. Chapter 4 describes a numerical method for determining failure time points from the data. The part 5 introduces the data used for the analysis presented and the following chapter 6 shows the results obtained using the new method and compares these results with those obtained using online statistical method. Finally, the part 7 discusses the results, advantages and disadvantages of the proposed method.

2. Related works

Given the introduced definition of "unsupervised method", as a methodology that does not require knowledge of the state of the input data, the list of existing methods for failure prediction is very limited. Following the work of [9] (which also includes references to other research works, as well as studies of specific solutions involving a broader understanding of the "unsupervised method"), the spectrum of available solutions can be generally divided into 2 categories.

(1) Solutions involving the use of prediction techniques to predict points in the future and then compare them with actual data to detect anomalies, as shown in [10]. Depending on predictive solutions used, this category of methods requires a large amount of data. In this case, a separate problem is the accuracy of the prediction part. Where the prediction part is a supervised method. Hence the

requirement to control it, which makes the whole solution complicated if one wants to use it in practical applications.

(2) Solutions based on measurements of distance or similarity used to evaluate the degree of data anomaly. In this case, methods using clustering algorithms [11] or determining similarity between time series or data are used. This approach sometimes leads to problems when encountering new data that did not exist in the past. It happens that it makes it impossible to use them for real-time data analysis.

The solution proposed in this paper is an attempt to solve the issues identified in both of the above categories: the problem of demand for large amounts of data and the problem of data changes that do not occured in the past.

The essence of the proposed method is to determine the initial breakdown (IB) points in the time series of input data. Thus, it is natural to compare the proposed model with online methods that determine trend change points in time series in fully unsupervised mode. Latest advances in statistical online trend change point detection can be found in studies [12, 13]. The difference between the statistical methods of determining the trend change points and the IB points determined by the proposed method is that in the proposed solution we are based on the search for a specific functional behavior (pattern) in the data. As will be shown later, this pattern is characteristic for physical processes leading to a degradation of device parts. A comparison of the results of the two methods, based on the same input data (statistical [13, 14] and the proposed model), will be presented and discussed in the section 6.

3. The occurrence of log-periodic oscillations as a prelude to failure.

The purpose of the PM method is to provide predictions for future failures of the system described as a set of various components cooperating together. This section illustrates why the LPPL-based algorithm is applicable to failure prediction and describes the basics of the LPPL-based method.

The generalized relation describing the hazard rate $h(t)$ (or hazard function) of a certain physical quantity at time t prior to material destruction by degeneration [15] is of the form

$$\dot{h}(t) = Gh(t)^{\delta} \tag{1}$$

where the $\dot{h}(t)$ denotes the derivative of the function $h(t)$ in time t. The δ and G are the parameters of model. The G is a positive constant parameter. In the following, it is assumed that $h(t)$ corresponds to changes in the variable that is tracked and from which the failure is attempted to be predicted. The hazard function, based on conditional probability theory, measures the probability that the relevant variable will show signs of a critical behaviour, given that the critical behaviour has not occurred prior to time t.

The equation (1) has 3 classes of solutions depending on the value of δ. For $\delta = 1$ the exponential function is obtained

$$h(t) = h(t_0) e^{G(t-t_0)} \tag{2}$$

where t_0 is a value of the initial time.
For $\delta < 1$:

$$h(t) = [(1-\delta)(t-t_0)G]^{\frac{1}{1-\delta}} \tag{3}$$

and for $\delta > 1$:

$$h\left(t\right) = \left[\left(\delta - 1\right)\left(t_c - t\right)G\right]^{\frac{1}{1-\delta}} \tag{4}$$

with t_c as a constant corresponding to time in a future. As can be seen, the solutions found for $\delta = 1$ (2) and for $\delta < 1$ (3) do not converge for times $t > 0$. Therefore, the time of critical behaviour cannot be determined. The most interesting case is $\delta > 1$ (4), where the solution has a converging point at a finite time t_c in future. In our analysis, the time t_c will denote a critical time point and can be determined by a choice of $G = \left(\frac{1}{t_c}\right)^{\frac{1}{1-\delta}}$. Then $t_c = \frac{1}{(\delta-1)h(t_0)^{\delta-1}}$.

Let us assume that the degradation of a working part, can be treated as a discontinuous stochastic process associated with a given monitored variable. To simplify the analysis, the second assumption is to treat the degradation changes of physical quantity $p\left(t\right)$ over time t as a non-homogeneous Poisson process in which the changes occur according to the hazard rate function $h\left(t\right)$. The dynamics of such a process can be described by the equation

$$dp\left(t\right) = -p\left(t\right)h\left(t\right)dt \tag{5}$$

the solution of which can be written as

$$\log\left[\frac{p\left(t\right)}{p\left(t_0\right)}\right] = -\int_{t_0}^{t} h\left(u\right)du = P\left(t\right). \tag{6}$$

In our case, $P\left(t\right)$ has an approximate form

$$P\left(t\right) \approx \frac{h_0}{\eta + 1}\left(t_c - t\right)^{\eta+1} \tag{7}$$

where $\eta = \frac{1}{1-\delta}$ and the $P\left(t\right)$ is shifted by the integration constant $\frac{h_0}{\eta+1}\left(t_c - t_0\right)^{\eta+1}$. The result obtained (7) coincides with work [16] (compare with equation (3) in the reference [16])

Solution (7) is invariant under continuous scale invariance (CSI), which manifests itself through the scaling property of the solution $P\left(t\right)$ if the argument of the function $P\left(t\right)$ is scaled to $\left(t_c - t\right)$. Rescaling by some factor v the argument $t_c - t \to \left(t_c - t\right) \times v$ changes the solution $P\left(t\right)$ to the form $P\left(t\right) \times \mu$ where $\mu = v^{-1-\eta}$ [17]. The CSI feature, around the critical points $t = t_c$, is common to systems demonstrating a continuous phase transition (second order phase transition).

The basic assumption of the LPPL method is that, the described process is near the critical point of the second-order phase transition. With this assumption, the final equation is obtained, the use of which for fitting is known in literature as the LPPL method [18], [19].

Let $W\left(t\right) = \log\left(p\left(t\right)\right)$ and $p\left(t\right)$ corresponds to the variable by which our industrial system is analyzed as a function of time t. Let t_c be the time of event that defines our phase transition in physical framework (critical time point). Then the argument x and the real function $F\left(x\right)$ are defined as

$$x = t_c - t \text{ with } t < t_c \text{ and } F\left(x\right) = W\left(t_c\right) - W\left(t\right) \tag{8}$$

As a starting point for our derivation, the CSI is assumed to exist around critical points.

This allows us to use the renormalization group approach, which permits us to write a certain real function $F(x)$ around a critical point $x \approx 0$ through the rescaled argument x expressed by a scaling function $\phi(x)$ in the form

$$F(x) = \frac{1}{\mu} F[\phi(x)] \tag{9}$$

where μ is a constant and its argument x is invariant under arbitrary linear transformation of the x:

$$\phi(x) = \nu x \text{ and } x > 0 \tag{10}$$

with ν as a constant.
The solution of equation (9) is the function

$$F(x) = C x^\alpha \tag{11}$$

where C and α are constants to be determined.

The condition of scaling invariance (9) with the general form of the solution postulated by eq. (11) can be rewritten as

$$C x^\alpha = C \frac{\nu^\alpha}{\mu} x^\alpha \tag{12}$$

which, after taking into account the identity,

$$1 = e^{i 2\pi n} \text{ with } n \in \mathbb{N}, \tag{13}$$

leads to the equation

$$e^{i 2\pi n} = \frac{\nu^\alpha}{\mu} \tag{14}$$

what allows to calculate the α exponent in a most general form as

$$\alpha = \frac{\log(\mu)}{\log(\nu)} + i \frac{2\pi}{\log(\nu)} n. \tag{15}$$

Therefore, the solution of equation (9) can be expressed, according to [20], as

$$F(x) = \frac{C}{\mu} (\nu x)^{\frac{\log(\mu)}{\log(\nu)} + i \frac{2\pi}{\log(\nu)} n} = \frac{C}{\mu} (\nu x)^{\frac{\log(\mu)}{\log(\nu)}} \Pi\left(\frac{\log(x)}{\log(\nu)}\right) \tag{16}$$

where $\Pi(\cdot)$ is a periodic function with period 1, i.e. $\Pi(y) = \Pi(y+1)$. The index n should be treated as one of the parameters characterizing the described physical system. Since n and other parameters appearing in the function (16) are unknown, it is necessary to reformulate the function (16) in such a way that it can be used to fit existing data and thus determine whether a given point t_c is a critical point.

An additional necessary condition that must be satisfied by function (16), or more precisely by its real part, is the trend that is determined by the power law ($n = 0$), which is the leading order term and the oscillations associated with $n \neq 0$ will contribute as next-to-leading order corrections.

The periodic function $\Pi(\cdot)$ can be expressed by means of the Fourier series with respect to the variable y with period T

$$\Pi(y) = \exp\left[i2\pi n\left(\frac{y}{T}\right)\right] = \sum_{k=-\infty}^{+\infty} c_k e^{i2\pi k\left(\frac{y}{T}\right)} \tag{17}$$

with

$$c_k = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}\left\{\exp\left[i\frac{2\pi}{T}yn\right]\right\} e^{-i\frac{2\pi}{P}ky}dy = \frac{1}{\pi}\frac{\sin\left[\pi(k-n)\right]}{k-n} \text{ for } n,k \in \mathbb{N}. \tag{18}$$

However, non-zero coefficients c_k of the Fourier expansion (17) are obtained only for non-integer differences $(k-n)$, which contradicts our claim for the existence of non-zero expressions associated with $n \in \mathbb{N}$. The problem of zero coefficients of the fourier expansion of (17) can be solved by rewriting the identity (13) to the form

$$1 = e^{i2\pi n} \rightarrow 1 = e^{i\frac{2\pi}{q}(nq)} \tag{19}$$

where, the variable q denotes a parameter associated with the physical degradation mechanism of the described system.

In that formulation, α (15) can be rewritten as

$$\alpha = \frac{\log(\mu)}{\log(\nu)} + i\frac{2\pi}{\frac{\log(\nu)}{q}}\frac{n}{q} \tag{20}$$

what allows us to rewrite the equation (16) to the form

$$F(x) = \frac{C}{\mu}(\nu x)^{\frac{\log(\mu)}{\log(\nu)}+i\frac{2\pi}{\frac{\log(\nu)}{q}}\frac{n}{q}} \tag{21}$$

In this case, the expansion of $F(x)$ into Fourier series gives us the following result.

$$F(x) = \frac{C}{\mu}(\nu x)^{\frac{\log(\mu)}{\log(\nu)}} e^{i\frac{2\pi}{\frac{\log(\nu)}{q}}\frac{n}{q}\log(x)} = \frac{C}{\mu}(\nu x)^{\frac{\log(\mu)}{\log(\nu)}}\sum_{k\in\mathbb{N}} c_k e^{i\frac{2\pi}{\frac{\log(\nu)}{q}}\log(x)k} \tag{22}$$

where

$$c_k = \frac{1}{T}\int_{-T/2}^{T/2} e^{i\frac{2\pi}{T}\left(\frac{n}{q}-k\right)y}dy = \frac{1}{\pi}\frac{\sin\left[\pi\left(k-\frac{n}{q}\right)\right]}{k-\frac{n}{q}} \tag{23}$$

with a redefined variable $y = \frac{\log(x)}{\frac{\log(v)}{q}}$.

Given the denominator $\left(k - \frac{n}{q}\right)$ in the coefficients of c_k (23), the main dominant terms of the series (22) are defined by the index k from the set $\{-1 + [n/q], [n/q], 1 + [n/q]\}$ where $[n/q]$ is the integer part of division n/q. To simplify the notation, let us redefine the index values from $\{-1 + [n/q], [n/q], 1 + [n/q]\}$ to $\{-1, 0, +1\}$.
This allows us to approximate the final form of the function $F(x)$ (22) by the first 3 largest components of the Fourier series (c_0 and c_{-1} or c_{+1})

$$F(x) \approx \frac{C}{\mu}(vx)^{\frac{\log(\mu)}{\log(v)}}\left[c_0 + c_{\pm 1}\cos\left(\frac{2\pi}{\frac{\log(v)}{q}}\log(x)\right) \pm ic_{\pm 1}\sin\left(\frac{2\pi}{\frac{\log(v)}{q}}\log(x)\right)\right] \tag{24}$$

where the notation $c_{\pm 1}$ was used to denote ambiguity as to which coefficient is the second dominant one for $k = -1$ or $k = +1$. Since only the real part of the expression is of interest (our measurements are real values), using the previous definition of the variable $F(x)$ (8) and generalizing our unknown parameters (μ, v, C, q, n, k and $W(t_c)$) by adding constant A and phase Φ to the formula (24) to their new representations (A, B, m, C, ω, Φ) one obtains the final formula which is referred to as a first-order model and used in LPPL literature [18, 19]

$$W(t) \approx A + |t_c - t|^m [B + C\cos(\omega\log|t_c - t| + \Phi)]. \tag{25}$$

For the purpose of numerical fitting the LPPL function to the data, a transformed version of the formula (25) is used, in the form of

$$W(t) \approx A + |t_c - t|^m [B + C_1\cos(\omega\log|t_c - t|) + C_2\sin(\omega\log|t_c - t|)] \tag{26}$$

where $C_1 = C\cos(\Phi)$ and $C_2 = -C\sin(\Phi)$. Both equations (25, 26) can be used to find critical time points in the time series of input data.

4. Fitting method of the LPPL model to the data

Due to the number of parameters (A, B, m, C_1, C_2, ω) necessary to be determined during the fitting procedure of the LPPL function (26) and the presence of many local extremes, the procedure of obtaining the best fit is difficult and computationally expensive.

Instead of trying to determine the critical time t_c in the future, as is determined in the case of predictions of crashes in financial time series [16, 19, 21], it is assumed that the critical time point is "now". With this assumption, the calculations involved in fitting the function (26) to the data are performed for a range of time windows of different lengths from the past to "now". This corresponds to the hypothesis that the time point t_c (time point of the phase transition) is "now" and corresponds to the last point in the input time series t_{inp}. Therefore, it is necessary to add to the set of parameters, an additional parameter specifying the length of the subset of time series points t_{inp} preceding the time point t_c for which the best fit of the function $W(t)$ (26) was found. The length of this subset is denoted as l_{max}.
In particular, our redefined set of arguments $x = t_c - t$ in the matching procedure is the set of points

$\langle 1, x_{l_{max}} \rangle$ in time units characteristic to the t_{inp} series. The definition of the argument set $x = t_c - t$ in the matching procedure is replaced by the set of points $\langle 1, x_{l_{max}} \rangle$, where 1 corresponds to the present time and l_{max} corresponds to the time of l_{max} from the past. The index l_{max} is the one of the parameters of the matching function (26).

As parameter fitting, the method described in the work of [21] is used, appropriately modified for our purposes, i.e., by excluding the parameter corresponding to the t_c and adding the parameter l_{max} to the fitting procedure.

The constraints imposed on the fitting parameters (l_{max},A,B,m,C_1,C_2,ω) are as follows:

(1) l_{max}: the number of past data used for the best fit. For a small number of data there may be too many good fits (with very small fitting error), which may correspond to random correlations of the data with the form of the fitted function (26).
(2) $A > 0$: since in our case there are always positive values. It is determined by the character of the input data.
(3) $0 < m < 1$: to ensure that the fitting value for the critical time t_c was greater than zero ($m > 0$) and changed faster than exponentially for times close to the critical time t_c ($m < 1$).
(4) $2 < \omega < 8$: this condition avoids too fast log-period oscillations (otherwise they would fit the random component of the input data) and too slow log-period oscillations (otherwise they would contribute to the power law behaviour $\approx A + B |t_c - t|^m$)
(5) B, C_1,C_2: these parameters are fitted without additional constraints.

Depending on the type of input data to be analyzed, the limits of variation of the parameters to be fitted require careful adjustment.

Having determined all parameters of the fitting LPPL function, in the next step, it is necessary to determine trends based on the identified local maxima and minima of the shape of the fitted function: T_{max} for local maxima and T_{min} for local minima. The procedure of finding trends is carried out separately for maxima and minima in the following way:

(1) all local extrema (N) are found,
(2) from this set, the N-1 extreme values closest to the current time point (i.e., the point for which an attempt is made to determine whether phase transition has occurred or not) are selected,
(3) To this set of points a straight line is fitted by linear regression. The slope of the line determines the trend for a given category of extremes (maxima or minima).

Then, using the calculated trends of the extreme values of the best LPPL fit, it is determined whether a given point (defined as a pair: {datetime, value}) corresponds to a phase transition or not. For this purpose, a theorem describing the critical time point was formulated. Its proof will be the aim of the next publication.

Theorem 1. *Assume that the function $f(x)$ corresponds to the best found fit of the LPPL function (26) to the analyzed input time series $ts = \{(t_n - t_{l_{max}}, y_{n-l_{max}}), ..., (t_n - t_{n-1}, y_{n-1})\}$ where $l_{max} > 0$ is one of the fit parameters of the function $f(x)$ (26) specifying the length of the sequence preceding the actual values $(t_n - t_{n-1}, y_{n-1})$ defined by the index n. The function $f(x)$, have $N_{max} > 2$ local maxima and $N_{min} > 2$ local minima. Let T_{max} denote the slope of the linear fit determined by the last N_{max} values of the local maxima and T_{min} denote the slope of the linear fit determined by the last N_{min} values of local minima.*

If both trends T_{max} and T_{min} determined for the function $f(x)$ have the same behavior (both increasing or decreasing) for the last point $(t_n - t_{n-1}, y_{n-1})$ of the series ts, then the point (t_n, y_n) is the critical point

for the series ts. The trend of the series ts will change to the opposite for next points (t_{n+k}, y_{n+k}) (where $k > 0$) with respect to the trend for points preceding t_n.

In the presented method, any point that satisfies the Theorem (1) is treated as a point initiating a future failure - the initial breakdown (IB) point.

The formalism described above can be summarized as follows: the appearance of a failure due to deterioration of components in a complex system is preceded by the appearance of a phase transition of the 2nd kind, defined by point IB (critical time point). From this time point, the process of failure development begins. This period (after the time defined by the IB point) can be monitored by detecting anomalies. Anomalies (as a result of reduced process efficiency) occur when the effects of the failure become large enough and begin to affect the behavior of the monitored device.

Determining the time of failure requires knowledge of the dynamics of the monitored system and will be discussed in section 6.

5. Data description

Monitoring data from a reciprocating compressor, describing the PV diagram of one of the compression chambers, was used to demonstrate the operation of the method. Based on these, the values of the angle of opening of the suction valve (OSV) expressed by the angle of rotation of the crankshaft were determined. The data is available at [22].

This value, for a given cycle described by the PV diagram, is very sensitive to changes in the amount of gas in the chamber, for example, due to its leakage through a broken valve or piston rod seal system. OSV changes are directly dictated by the thermodynamics of the compression process in the compressor chamber. Monitoring the changes in OSV provides a base for compressor diagnostics and allows to determine compressor efficiency, valve operation, or the condition of the piston seals or piston rod sealing elements [23].

The data has been averaged to daily values and covers a period of time between 2019-08-23 and 2022-01-19.

In order to compare the results of failure prediction, data identifying the dates of repair interventions with their respective reasons for failure and the dates of observation of anomalous compressor behavior without interruption of operation were used.

6. Prediction of failures: methodology and results

To test the effectiveness of our algorithm, a backtest of the detection method was conducted on the OSV data, calculating initial breakdown (IB) points. The range of the variable length of the time series l_{max} was assumed to be $30 < l_{max} < 101$ in time units of days. Given the minimum number of observations (101 days) needed to perform calculations of the IB points, the backtest is started for time $t = t_{start} + 101$ (in daily units). Then, moving forward in time to the future, the best-fit LPPL function (26) is calculated for each subsequent time t by determining the goodness of fit of the LPPL function using the mean squared error (*mse*).

Intuitively, one can expect that the accuracy of determining the critical points using Theorem (1) will strongly depend on the error of fitting the LPPL curve (26) to the data. The smaller the fitting error, the greater the confidence that a given point of the input time series is indeed an IB point according to

Theorem (1). In addition, it is expected that in the vicinity of the true IB breakpoint (before and after it), the method should find some good fits of the LPPL function with a small error, but this is not a necessary criterion for the existence of an IB point for days as time units.

Figure (1) shows 2 examples of the fit function (26) to data along with calculated criteria for trends determined from maxima and minima of the fitting function satisfying the criteria of Theorem (1).
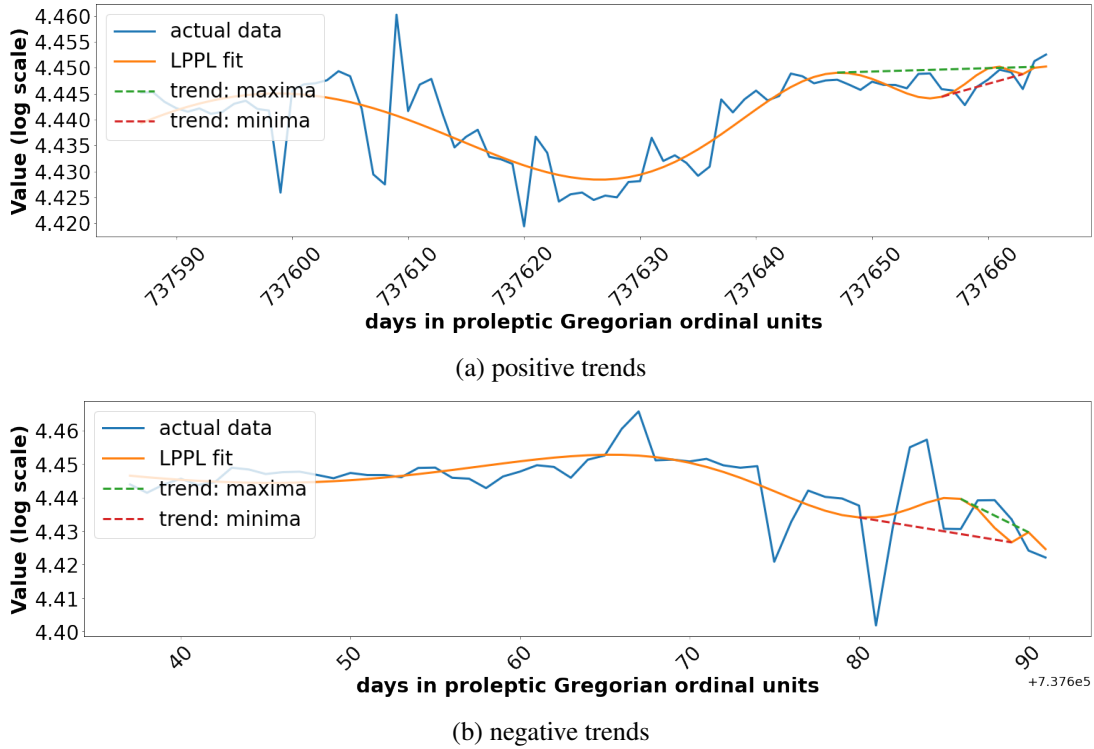

(a) positive trends

(b) negative trends

Fig. 1. Examples of LPPL function (26) fits for cases with positive trends (a) and with negative trends (b) for selected current times (dates in proleptic Gregorian ordinal units). Fit parameters calculated for both trend categories: (a) positive trend: current date: $t_c = 2020 - 08 - 28$, length of the time series: $l_{max} = 79$, mean squared error: $mse = 2.6 \cdot 10^{-5}$; (b) negative trend: current date: $t_c = 2020 - 09 - 23$, length of the time series: $l_{max} = 55$, mean squared error: $mse = 5.7 \cdot 10^{-5}$.

Figure (2) shows the application of this procedure to the diagnosis of the critical points with additional information about the dates of failure repairs (see the description of Figure (2)).

Figure (2) confirms our initial hypothesis very well. It shows:

(1) groups of points with similar fitting errors at least 14 days before the time of failure identification (repair is usually performed with additional delay due to the compressor operating conditions),

(2) dependence of the matching error on the criticality of the failure: the signaling of the grouping of critical points for the beginning of 2020-12-14 is characterized by a large error ($mse = 24.9 \cdot 10^{-5}$), much larger than for the group of points for which the failure has been confirmed ($mse <= 5.7 \cdot 10^{-5}$).

Taking these conclusions into account and comparing the recorded failures and behaviors suggesting problems in compressor operation, classified by experts as insignificant, with the predictions made by
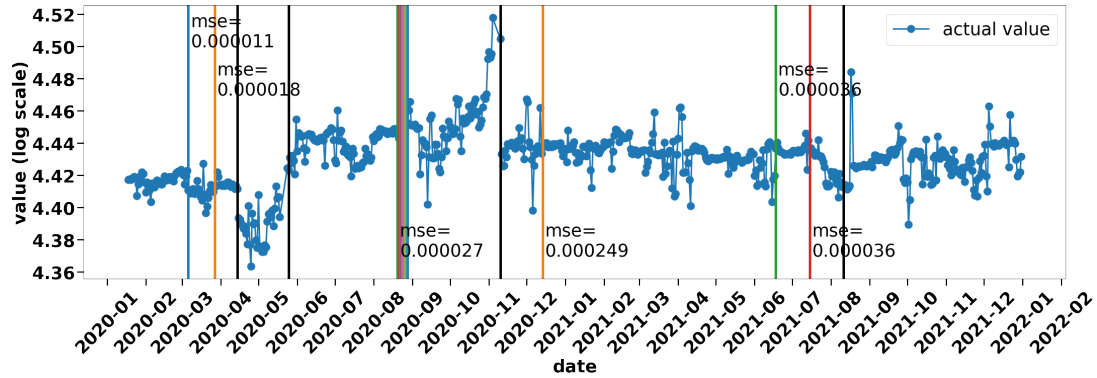
Fig. 2. Calculated IB points for the analyzed input time series (marked in figure by a solid line with measurement points). The vertical lines of different colors indicate the diagnosed by the algorithm IB points. For each group of the IB points, the mean value of the matching error (*mse*) of the LPPL function (26) is annotated. The black vertical lines indicate the dates of compressor repairs, which typically took place a few to several weeks after the algorithm detected the fault (IB points). The correlation with the diagnosed breakpoints is clearly visible, except for the prediction determined with the largest error (mse=0.000249) for 2020-12-14.

the model, it is possible to determine the threshold values of the fitting error (*mse*) and the corresponding categories of predictions:

Definition 1. *Classification of calculated initial breakdown points:*

(1) **Critical event***: (mse $< 6 \cdot 10^{-5}$) severe failure expected, checking of the compressor required and preparation for repair,*

(2) **Monitoring event***: ($6 \cdot 10^{-5}$ $<= mse < 10 \cdot 10^{-5}$) distinct possibility of issues, monitoring of compressor behavior required.*

(3) **Irrelevant event***: ($10 \cdot 10^{-5}$ $<= mse$) no significant issues predicted, compressor behaviour can be monitored though.*

As shown, sometimes the algorithm detects a larger number of IB points located very close to each other. Such cases were simplified by choosing a single representation (the first IB point of the group) for each group of signals. This was done by assuming that the signal belongs to a group if its distance from the preceding signal is smaller than or equal to 3 days.

6.1. Determining the time window of predicted failures

By comparing the failure times predicted by the algorithm with their actual occurrence, a criterion for predicting the time window in which the failure will occur can be also determined. One of the parameters for fitting the LPPL function (26) to the data is the length of the chosen sequence of data preceding the analyzed time point l_{max}. For the data analyzed, the time window for the occurrence of a predicted failure was defined as:

Definition 2. *The predicted time period of a failure occurrence is defined as an interval $\left\langle n + \frac{l_{max}}{2}, n + 90 \right\rangle$, where n and l_{max} are the indices of the actual time point x_n and the parameter defining the length of the input time series t_{inp} used to find the best fit of the LPPL function (26) to a given value of t_{inp} (x_n, y_n), respectively.*

Definition (2) is based on a knowledge of the dynamics of the device for which the algorithm parameters have been defined. For other devices, all parameters should be selected based on the dynamics of their behavior. Duration of the time window with predicted failure is assumed to be valid for a certain period of time and is up to 90 days.

In a case of compressors, due to the different criticality of failures, some of them may be accepted for a longer period of time (even several months in the case of valve failures) while waiting for a convenient moment for repair.

Considering:

- the classification of alerts specified in Definition (1),
- the selection of the representative of the groups of warnings (by selection of the initial signal for the common group of calculated IB points),
- Definition (2) specifying the expected time window in which the failure will occur,

the raw results shown in Figure (2) can be redrawn to a new form, as shown in Figure (3). The correlation between the predicted failure times and the actual repair times, or the time periods when experts detected abnormal compressor behavior is very good for the **Critical event** and **Monitoring event** categories. Predictions in the **Irrelevant event** category were not confirmed by any repair and diagnosis records.

6.2. *Root cause of predicted failures*

In the analysis presented here, the input data monitors the change in the opening angle of the suction valves in the compression chamber expressed by the angle of rotation of the crankshaft. The trend identified from the determination of the IB points can be used to guess the type of future failure in the compression chamber [23]. By predicting the trend for times after the IB point, it is possible to try to determine approximately which part will fail - the valve or the piston rod sealing rings. Thus, when the predicted trend of the suction valve opening angle is decreasing, it is likely that the suction valve or piston rod seal rings are failing. If the trend suggests an increase in angle, this behavior indicates a leak in the discharge valve.

Thus, based on Theorem (1) and a physical interpretation of the behavior of the time series, the algorithm is able to predict not only the time window of failure, but also the group of parts that may fail. This provides an opportunity to verify the prediction not only on the basis of event times, but also on the basis of identifying the parts that can fail.

To better illustrate the additional information regarding the location of the future failure, Figure (4) shows the same data as Figure (3) with additional information about the parts that actually failed, the parts in which experts have observed problems and prognosis of failures predicted by the algorithm. Details are given in the description of Figure (4).

For the entire time period analyzed, 2 cases deviating from this diagnosis are visible

(1) in the category **Monitoring event**, for date $2021 - 07 - 15$, there is a disagreement between the predicted failure type *suction valve or sealing - leakage* and the diagnosed one *indiation of discharge valve leakage*.

(2) the perturbation identified by experts, started on $2021 - 11 - 30$ and identified as *indiation of discharge valve leakage* was not predicted by the algorithm at all.
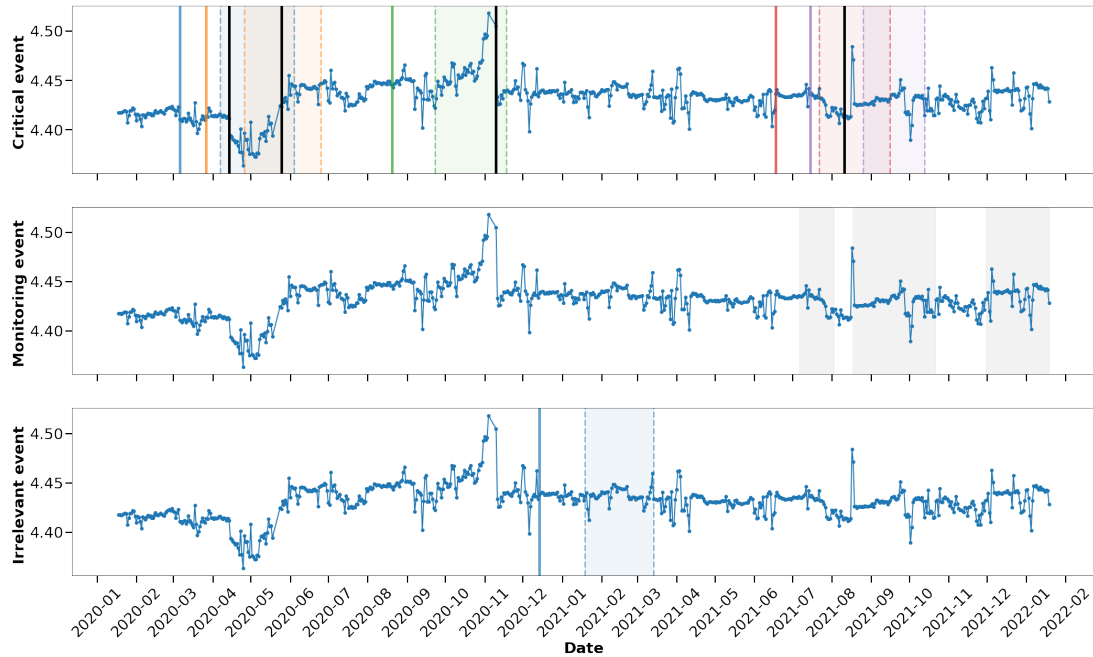
Fig. 3. Redrawn prediction of failures by the algorithm in comparison with reparation times and problems detected by experts. Figure shows representations of groups of discovered IB points (colored and continuous single vertical lines) and predicted failure time periods corresponding to them according to Definition (2) (corresponding colored areas bounded by vertical dashed lines). The categorization of the criticality of the forecasted problems (Definition (1)) is represented by the splitting into 3 separate figures, each for a separate criticality category. The **Monitoring event** class remains without any marked vertical lines. The algorithm found no prediction for this category. The reparation (maintenance) dates are indicated in figures by black vertical lines. Time periods of abnormal compressor behavior requiring monitoring and not qualified for repair by experts are indicated by the gray colour without additional delimiting lines. The intersection of the areas determined by the algorithm as a period of failure occurrence with periods of abnormal behavior of the compressor can be seen for periods starting from 2021-07-15 in the **Critical event** category. The raw data of the input time series t_{inp} are indicated by the solid blue line with the measurement points (*o*).

6.3. Comparison with statistical method

The described method generates IB points before the appearance of anomalies associated with a future failure (section 4). This is why the LPPL-based solution has an advantage over solutions based on anomaly detection as a signature of a forthcoming failure.

It cannot be excluded that the IB points detected by the described method, may be the same as the statistically defined change points in the input trend of the time series. For comparison of the results determined by the discussed solution with the trend change points determined statistically, the online changepoint detection method from the work of [13] was used. This solution is available in the form of the Python library [14] (changepoint_online).

I ran a model based on detection of two states of the analyzed data: one focusing on the detection of a decreasing trend, and the other focusing on the changes in the increasing trend. This way, it is possible to track when the analyzed variable increases and when it decreases. This corresponds to the ability to detect the type of future failure, as described in subsection 6.2.

In each iteration of the historical test, two changepoint detectors are run, one focusing on decreasing changes (left detector) and the other on increasing changes (right detector). In the source of the
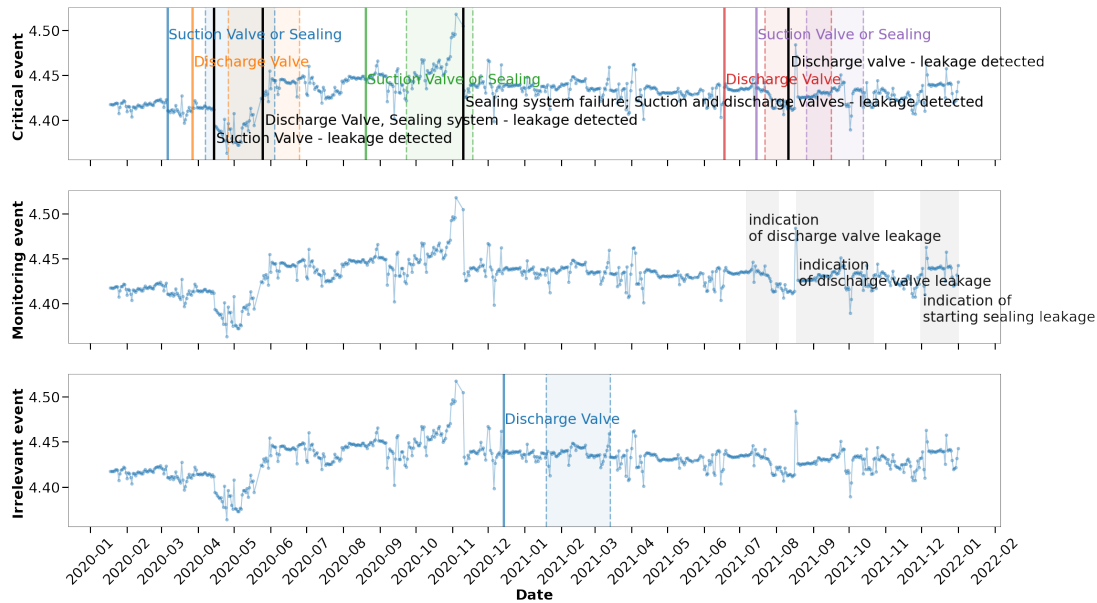
Fig. 4. The same graphs as in Figure (3) with added annotations describing the failures predicted by the algorithm - colored texts (different from gray and black), reasons for compressor reparations - black text and diagnosed abnormal compressor behavior requiring monitoring - gray text. For better visualization, areas of the diagnosed abnormal behaviour of the compressor that require monitoring are displayed in the "Monitored Event" category.

changepoint_online [14], this corresponds to the terminology: left and right detectors, respectively. The two detectors work independently of each other. When either model exceeds a threshold value, it is reinitialized.

A comparison of the results of the two methods at the level of determining IB points and trend change points is shown in Figure 5.

As shown in Figure 5, the set of IB points determined by the LPPL method (14 alerts) is much smaller than the set of trend change points found by he online statistical method (changepoint_online) (44 left and 33 right detections) . This is consistent with intuition because not every trend change point is a signature of change due to device degradation. The selection of all trend change points will create a lot of false signals. Therefore, it is necessary to correctly identify the threshold value to be used for selection of the best trend change points, which may suggest future failure. The choice of the threshold value should demonstrate the relation between changes in the behavior of the analyzed variable with degradation processes in the monitored device. Such a correlation requires a separate analysis and knowledge of the labeled data. The applied statistical method is unable to predict future failures based solely on determining trend change points in a fully unsupervised manner.

7. Discussion

The comparison of information coming from the failure predictions and comparing it with the knowledge from maintenance logs and expert detection of periods of anomalous compressor operation, including the prediction classification, is provided in Table 1.

B. Łobodziński /

	Predictions				Maintenance logs		Expert diagnoses		
Label	Alert date	mse	Predicted time window of failure	Predicted failure	Maintenance date	Diagnosis	Start date	End date	Diagnosis
TP	2020-03-06	$11 \cdot 10^{-5}$	2020-04-07 2020-06-04	SV or Sealing	2020-04-14	SV: leakage detected	-	-	-
TP	2020-03-27	$18 \cdot 10^{-5}$	2020-04-26 2020-06-25	DV	2020-05-25	DV, Sealing system: leakage detected	-	-	-
TP	2020-08-20	$26 \cdot 10^{-5}$	2020-09-23 2020-11-18	SV or Sealing	2020-11-10	Sealing system failure; SV and DV: leakage detected	-	-	-
FP	2020-12-14	$249 \cdot 10^{-4}$	2021-01-19 2021-03-14	DV	-	-	-	-	-
TP	2021-06-18	$36 \cdot 10^{-5}$	2021-07-22 2021-09-16	DV	2021-08-11	DV - leakage detected	2021-07-06	2021-08-02	DV leakage
FP	2021-07-15	$36 \cdot 10^{-5}$	2021-08-26 2021-10-13	SV or Sealing	-	-	2021-08-17	2021-10-21	DV leakage
FN	-	-	-	-	-	-	2021-11-30	2022-03-02	starting sealing leakage

Table 1: Detailed comparison of model predictions with maintenance logs entries and expert diagnoses. The "Predictions" section contains a description of the alert: "Alert date" - date of alert occurrence, "mse" - goodness of the fit, "Predicted failure time window" - the predicted time window in which the failure may occur, and "Predicted failure" - the predicted failure diagnosis. The "Maintenance logs" section contains columns: "Maintenance date" - repair date and "Diagnosis" - the reason for the failure. The "Expert diagnoses" columns contain information about: the beginning date of the anomaly notice - "Start date", the end date of the anomaly - "End date" and the symptom of the anomaly - "Diagnosis". The "Label" column contains the identification of the predicted failure. The following convention is used: True Positive (TP) event, with correctly predicted compressor failures or misbehavior, False Positives (FP), an event predicted by the algorithm that turned out to be false, and False Negatives (FN), i.e. compressor failures or instabilities that were not predicted.
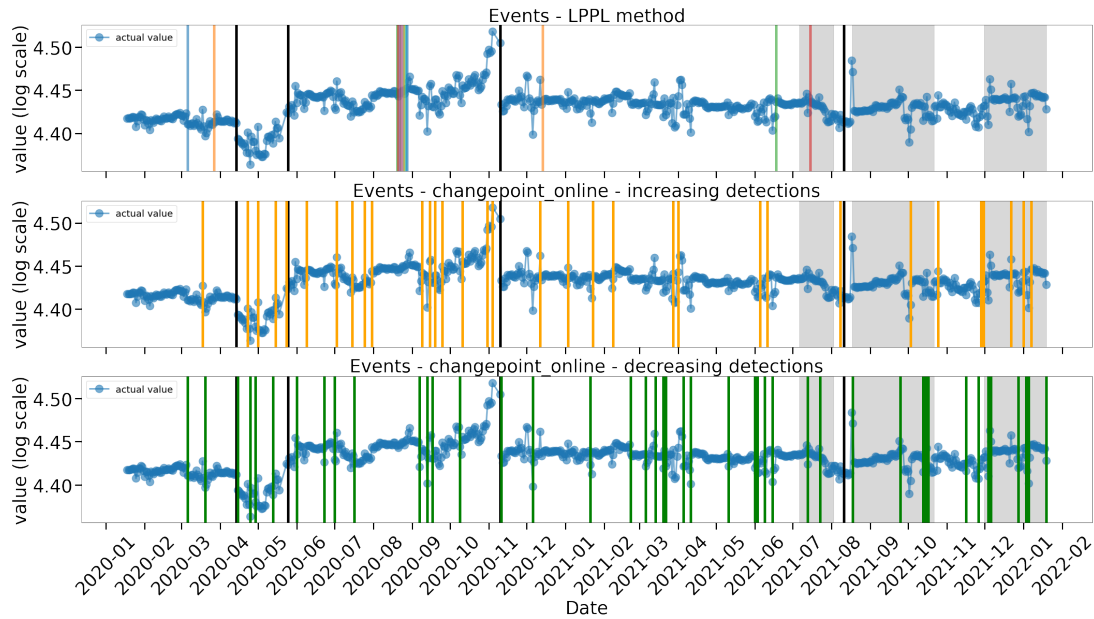
Fig. 5. Comparison of the results of the presented method (LPPL method - top figure) with the statistical method (change-point_online - middle and bottom figures) at the level of determining IB points and trend change points. The analyzed time series t_{inp} is shown by the solid blue line with the measurement points (o). The black vertical lines indicate the compressor repair dates. Time periods when abnormal compressor behavior requiring monitoring was not qualified for repair by experts are indicated by the gray colour without additional delimiting lines.

Top figure: for the LPPL method, the vertical lines in different colors indicate the IB points diagnosed by the algorithm for the LPPL method (as in Figure 2).

Middle and bottom figure: the found points of upward (middle figure) and downward trend change (bottom figure). Calculations of trend change points are shown for the threshold value of the 75th percentile.

Major challenges in the field of industrial application of failure prediction, especially in the unsupervised version, is the number of issues corresponding to proper identification of failures and behaviors of monitored devices. These difficulties mainly stem from:

- the large variety in the types of failures,
- the small number of failures compared to the amount of data,
- the large variety of behaviors leading to the same type of failure,
- the difference in operating conditions, which can cause ambiguity in labeling accurate data.

The problems are very difficult to solve if the methods used to predict failures are based on the analysis of numerical values of input data (by calculating similarities, correlations, logic trees, building naural networks, etc.). Normalization and/or standardization procedures only introduce a common scale to the analyzed data.

The proposed method introduces a new type of procedure, which is based on the search for common functional behavior (26). From the point of view of the numerical values of the analyzed data, the course of the fitted function can be very different for different events - different patterns of the function for different values of fitted parameters. Even when new data appears with values that did not exist in the past, it is possible to determine the initial breakdown (IB) for a potential event, as the model fits functions to the data.

The IB points determined by this method, upon which the time window of failure occurrence is predicted in the next step, are the trend change points in the data.

To calculate the key performance indicators (KPIs) of the presented algorithm, the generally known indicators can be used: $Precision = TP/(TP + FP)$ and $Recall = TP/(TP + FN)$, where TP - defines *True Positive* events, i.e. correctly predicted failures or abnormal behavior in the operation of the compressor, FP - *False Positives*, i.e. events predicted by the algorithm that turned out to be false and FN - *False Negatives*, i.e. failures and instabilities of the compressor that were not predicted.

In the analysis of results (section 6), the limits of acceptance of errors of fitting the LPPL function (26) determining the criticality of the predicted events (Definition 1) were defined on the basis of the results. Therefore, in order to calculate the *Precision* and *Racall* indicators, all results are taken into account without distinguishing them due to the defined criticality.

Comparing the predicted failure periods, taking into account the dates and predicted types of failures with the dates and descriptions of Maintenance logs or recorded faults (Figures 3 and 4 and the table 1), the calculated values of TP, FP and FN are as follows: $TP = 4$, $FP = 2$, $FN = 1$. Hence *Precision* $= 0.67$, *Recall* $= 0.8$. Given that, this result takes into account the dating of the predicted failure period along with the prediction of the cause of failure, such an outcome is considered very good.

In summary, the list of advantages and disadvantages of the presented method is a consequence of a paradigm change in data behavior classification, from the one based on numerical values to the one based on functional similarity.

Advantages of the method:

- The proposed model can be applied to very short time series (in our case, the minimum length of the series is only 101 points).
- There are no problems with data that appears for the first time. In the proposed solution, the part that qualifies certain data as IB is based solely on functional behavior. This universality is due to the renormalization group approach.
- The simplicity of the final production solution. The most difficult part of the algorithm is fitting the function to the data. Since the method does not contain components based on supervised methods, there is no need to monitor their quality.

Disadvantages of the method:

- The method is applicable to data that describe a physical process that degrades/changes due to perturbations introduced by interacting elements. This is because the method searches for behavior characteristic of phenomena in which phase transitions can be observed. Hence, not all data are appropriate for the described method.
- Matching the function to the data is based on the proper determination of boundaries of the parameters to be matched (26). This requires individual adjustment of the ranges of change of these parameters and the unit of time in the data to the device being monitored.

8. Conclusions

This paper presents the application of a methodology for describing critical behavior in complex systems based on the renormalization group approach in unsupervised predictive maintenance. The proposed algorithm analyzes the behavior of a complex system based on a time series representing the physical behavior of the system. To demonstrate the effectiveness of the algorithm for industrial applications,

predictive results are presented for time series describing the thermodynamics of the gas compression process in a monitored reciprocating compressor in one of the compression chambers.

It was shown that failures in the analyzed industrial system can be treated as critical behavior in complex systems. Then the symptoms of future failure appear in the form of Log Periodic Power Law structures for phase transitions in the analyzed time series. Based on the most generalized scheme for describing the behavior of an analyzed system in the vicinity of phase transitions of the 2nd kind based on the Log Periodic Power Law, a new way of predicting failures in compressor systems is proposed.

The presented algorithm is based on 3 steps. In the first step, the algorithm determines the IB points in the analyzed time series by means of fitting the LPPL function (equation (26)) using the proposed Theorem (1). The second step of the proposed method is based on the knowledge of the dynamics of the monitored system, and specifies the time window in which the predicted failure may occur. In the last step, a criticality classification of the predicted failure is carried out, based on the goodness of fit of the LPPL curve to the data (*critical event*, *monitoring event*, *insignificant event*).

Taking into account the specificity of problem detection in industrial systems (the demand to reduce the number of false alarms and to minimize the number of unpredicted events), it has been demonstrated that it is possible to experimentally determine such an error threshold of fitting the LPPL function to the data that all serious failures can be predicted if the fitting error is smaller than the threshold. In addition, it is also possible to define such thresholds for the LPPL curve-fit error to the data, for which the area of occurrence of less critical failures, that do not require rapid intervention, can be defined.

The method can also be applied to predictive IoT analysis of other industrial systems.

References

[1] J. Białek, W. Kuberski, N. Perrakis, and A. Bifet, "Estimating Model Performance Under Covariate Shift Without Labels", https://arxiv.org/abs/2401.08348;
J. Kivimäki, J. Białek, J. K. Nurminen, and W. Kuberski, "Confidence-based Estimators for Predictive Performance in Model Monitoring", in *Journal of Artificial Intelligence Research*, 82, (2025), 209-240;
R. Müller, M. Abdelaal, and D. Stjelja, "Open-Source Drift Detection Tools in Action: Insights from Two Use Cases", https://arxiv.org/abs/2404.18673

[2] Soori, M., B. Arezoo, and R. Dastres, "Internet of things for smart factories in industry 4.0, a review" in *Internet of Things and Cyber-Physical Systems*, 2023;
Z. M. Çınar, A. Abdussalam Nuhu, Q. Zeeshan, O. Korhan, M. Asmael, B. Safaei, "Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0" in *Sustainability* 12.19 (Oct. 2020), p. 8211. issn: 2071-1050. doi: 10.3390/su12198211. url: http://dx.doi.org/10.3390/su12198211;
T. Carvalho, F. Soares, R. Vita, R. Francisco, J. Basto, S. G. Soares Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance" in *Computers & Industrial Engineering* 137 (Sept. 2019), p. 106024. doi: 10.1016/j.cie.2019.106024;
Y. Ran, X. Zhou, P. Lin, Y. Wen, R. Deng. "A Survey of Predictive Maintenance: Systems, Purposes and Approaches", https://arxiv.org/abs/1912.07383

[3] B. Lobodzinski, A. Cuquel, "Method for predicting failures in industrial systems", May 15 2024, European Patent Application EP 4 369 127 A1, https://data.epo.org/publication-server/document/pdf/4369127/A1/2024-05-15

[4] A. Johansen, D. Sornette, "Evaluation of the quantitative prediction of a trend reversal on the Japanese stock market in 1999" in *International Journal of Modern Physics C*, vol. 11, no. 2, 2000, pp. 359–364.

[5] D. Sornette, "Why Stock Markets Crash (Critical Events in Complex Financial Systems)" in *Princeton Science Library*, 2003, p. 49.; D. Sornette, "Critical market crashes", in *Physics Reports*, vol. 378, no. 1, 2003, pp. 1–98.

[6] L., Lin, RE., Ren RE., D., Sornette, "The volatility-confined LPPL model: A consistent model of explosive financial bubbles with mean-reverting residuals" in *Int Rev Financ Anal.*, 2014, 33, pp. 210–225 and references therein

[7] D., Sornette, G., Sammis, "Complex Critical Exponents from Renormalization Group Theory of Earthquakes: IMplications for Earthquake Predictions", in *J. Phys. I France*, 5, 1995, p. 607-619

[8] Q. Lei, D. Sornette, "Generalized failure law for landslides, rockbursts, glacier breakoffs, and volcanic eruptions", https://arxiv.org/abs/2409.11455

[9]   Giannoulidis, A., Gounaris, A., Naskos, A. et al., "Engineering and evaluating an unsupervised predictive maintenance solution: a cold-forming press case-study" in *Information and Software Technology*, 122, (2024), 106287 https://doi.org/10.1007/s10845-024-02352-z

[10] Hundman, K., Constantinou, V., Laporte, C., Colwell, I., Soderstrom, T., "Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '18" in *Association for computing machinery*, New York, NY, USA, (2018), pp. 387–395. https://doi.org/10.1145/3219819.3219845.

[11] Diez, A., Khoa, N. L. D., Makki Alamdari, M., Wang, Y., Chen, F., & Runcie, P., "A clustering approach for structural health monitoring on bridges" in *Journal of Civil Structural Health Monitoring]*, 6(3), (2016), 429–445. https://doi.org/10.1007/s13349-016-0160-0

[12] Edward, A., Romano, G., Eckley, I., A., and Fearnhead, P., 2023. "Online Non-Parametric Changepoint Detection with Application to Monitoring Operational Performance of Network Devices", in *Computational Statistics & Data Analysis*, 177, 107551;
Pishchagina, Liudmila, Gaetano Romano, Paul Fearnhead, Vincent Runge, and Guillem Rigaill. 2023. "Online Multivariate Changepoint Detection: Leveraging Links with Computational Geometry", in *arXiv* Preprint arXiv:2311.01174;
Gaetano, R., Eckley, I., A., and Fearnhead, P., "A Log-Linear Nonparametric Online Changepoint Detection Algorithm Based on Functional Pruning", in *IEEE Transactions on Signal Processing*, 72, 2024, 594–606, https://doi.org/10.1109/tsp.2023.3343550;
Kes, W., Romano, G., Eckley, I., and Fearnhead, P., "A Constant-Per-Iteration Likelihood Ratio Test for Online Changepoint Detection for Exponential Family Models", in *Statistics and Computing*, 34, (3), 2024, 1–11

[13] Gaetano, R., Eckley, I., A., Fearnhead, P. and Rigaill, G., "Fast Online Changepoint Detection via Functional Pruning CUSUM Statistics", in *Journal of Machine Learning Research*, 24, (81), 2023, 1–36, https://www.jmlr.org/papers/v24/21-1230.html

[14] Code: Romano, G., Grose, D., Ward, K., Edward, A., Pishchagina, L., Rigaill, G., Runge, V., Fearnhead, P., Eckley, I., A., "changepoint_online: A Collection of Methods for Online Changepoint Detection", April 2024, version v1.2.1, https://https://github.com/grosed/changepoint_online

[15] B. Voight, "A method for prediction of volcanic eruptions" in *Nature*, 332, 1988, pp. 125–130.

[16] A. Johansen, O. Ledoit, and D. Sornette, "Crashes As Critical Points" in *International Journal of Theoretical and Applied Finance (IJTAF)*, World Scientific Publishing Co. Pte. Ltd., vol. 3(02), 2000, pp. 219–255.

[17] K. Ide, D. Sornette, "Oscillatory Finite-Time Singularities in Finance, Population and Rupture" in *Physica A*, vol. 307(1-2), 2002, pp. 63–106.

[18] J. A. Feigenbaum, P. G. O. Freund, "Discrete scale invariance in stock markets before crashes" in *International Journal of Modern Physics B* vol. 10, 1996, pp. 3737-–3745.

[19] D. Sornette, A. Johansen, and J.-P. Bouchaud, "Stock market crashes, precursors and replicas" in *Journal de Physique I*, vol. 6 (1), 1996 ,pp. 167–175.

[20] J. Nauenberg, "Scaling representation for critical phenomena" in *J. Phys. A*, vol. 8, 1975, p. 925.

[21] M. Shu, W. Zhu, "Real-time Prediction of Bitcoin Bubble Crashes" in *Physica A: Statistical Mechanics and its Applications*, vol. 548, 2019, p. 124477; (with code github: https://github.com/Boulder-Investment-Technologies/lppls).

[22] Open Science Framework https://osf.io/besak/?view_only=e5cab89e58724998848c8405bceb742d

[23] H. P. Bloch, J. J. Hoefner, "Reciprocating compressors: operation and maintenance", Gulf Professional Publishing, 1996.