Mechanical Systems and Signal Processing 184 (2023) 109677

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Framework for stochastic modelling of long-term non-homogeneous data with non-Gaussian characteristics for machine condition prognosis

Wojciech Żuławiński [b], Katarzyna Maraj-Zygmąt [b], Hamid Shiri [a], Agnieszka Wyłomańska [b], Radosław Zimroz [a,*]

[a] Faculty of Geoengineering, Mining and Geology, Wroclaw University of Science and Technology, Na Grobli 15, 50-421 Wroclaw, Poland
[b] Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wroclaw University of Science and Technology, Wyspianskiego 27, 50-370 Wroclaw, Poland

ARTICLE INFO

Communicated by J. Antoni

Keywords:
Long-term diagnostic data
Model identification
Trend extraction
Heavy-tailed distribution
Impulsive noise
Time-varying characteristics
Machine condition prognosis

ABSTRACT

To make prognosis one needs to build a model based on historical data. In the paper we propose a framework for modelling of long-term non-homogeneous data with non-Gaussian properties. These specific properties have been identified in real datasets describing the degradation process of the machine. The framework covers deterministic and random components separation, modelling of heavy-tailed, time-varying properties of random part as well as identification of possible autodependence hidden in the random sequence and identification of distribution for a random part. Due to non-linear trends, time-dependent scale (equivalent to the variance for Gaussian distributed data) and non-Gaussian characteristics present in the data, the final formula of the model is complex, its identification is challenging and requires specific, suitable to heavy-tailed processes, statistical methods. The paper provides two kind of novelties — first of all, it uses real data from condition monitoring systems and our findings may be novel and surprising to predictive maintenance community, secondly — processing such specific data opens new areas for general data modelling and highlight novel research directions.

1. Introduction

Progress with metrology, data transmission and storage, increasing demand for predictive maintenance are factors that boost development of condition monitoring systems. In the era of Big Data, collecting data during months or even years is an easy task. Unfortunately, end users (practitioners) are often confused. They have the data, but there is still a problem with decisions regarding actual condition and failure forecasting.

There are different methods of prognostics in machine health monitoring [1–4]. One of the classical approaches is the stochastic model-based methodology [5–7]. In the literature there are known different stochastic models used in the considered problem. The classical approach is based on the Gaussian processes, like Brownian motion (called also Wiener process [8]) or fractional Brownian motion (FBM) [9–11]. Especially the last process, considered as the classical Gaussian process with possible long-range dependence, seems to be useful for the description of non-linear characteristics of the data corresponding to the degradation process [12]. In the literature there are also considered non-Gaussian processes useful for diagnostic features (called health index, HI), like Gamma process [13], Generalised Cauchy (GC) process [14,15] or even processes with heavy-tailed distribution, such as fractional Lévy

* Corresponding author.
E-mail address: radoslaw.zimroz@pwr.edu.pl (R. Zimroz).

https://doi.org/10.1016/j.ymssp.2022.109677
Received 28 April 2022; Received in revised form 28 July 2022; Accepted 6 August 2022
Available online 19 August 2022

stable motion (FLSM), see e.g., [16,17]. The FBM uses the Hurst parameter to describe the degradation process, while the GC process uses the Hurst parameter and fractal dimension to describe the global and local characteristics of the degradation. However, the data corresponding to the degradation processes in practice are non-Gaussian, whereas FBM and GC processes obey a Gaussian distribution. Departure from Gaussian distribution is covered by the FLSM, considered as a natural extension of the FBM. FLSM also covers non-linear characteristics of the real data corresponding to HI. On the other hand, it has α-stable distribution, considered as the classical family of distributions with heavy-tailed (power law) behaviour. There are also methods based on the Kalman filter or its extended versions for HI data modelling [18–21]. Applications of the Markov jumps systems [22,23] or state-space models [24] are also used in that context. The alternative approaches are presented e.g., in [25]. For other reliability models describing the HI data variation see e.g., [26–29].

The mentioned above stochastic models are effective and suitable only under specific assumptions. Even if the model exhibits heavy-tailed behaviour (like FLSM), that is more adequate for real data, its scale (equivalent to the variance for the Gaussian models) is described by one specific function. For FLSM it is a power law function. Such an assumption may be too simplistic for complex HI data. Moreover, the α-stable distribution is only the exemplary one from the heavy-tailed distributions class. For different data, different non-Gaussian heavy-tailed distributions of the random part may occur. The model proposed in this paper can be considered as the general one and could be considered as the extension of approaches known from the literature, see e.g., [15,17], where authors propose specific non-Gaussian models with special time-dependent characteristics. The model proposed in this paper takes into account the characteristics discussed in the latest studies. It contains the time-dependent characteristics (like location and scale) as well as the possible non-Gaussian behaviour of the data. Moreover, it is much more general since we focus not only on one distribution of the noise (as was assumed in the Cauchy-based model [15]) and could be used for data with general time-dependent characteristics in contrast to fractional Lévy stable motion, where the power law behaviour of the data is assumed, see [17].

The preliminary analysis of real HI data has drawn our attention to a very specific behaviour of the corresponding time series. The analysis clearly indicates that the mentioned above stochastic models may be insufficient here. Thus, in this paper we propose a new framework for modelling of long-term data captured by machine monitoring systems. According to our experience, in many cases the data can be considered as a mixture of time series with different properties. We observe non-linear characteristics, non-homogeneous and non-Gaussian behaviour. Through non-homogeneous data we understand here the structure in which the data elements do not belong to the same data type in contrast to the homogeneous structure, where the data are of the same type (i.e. belong to the same distribution, the parameters of the distribution are the same etc.). From the engineering point of view, the analysed data has time-varying trend and scale, it belongs to non-Gaussian distribution so locally can contain some impulsive behaviour and may contain also internal dependency. The long-term observation (months, years) typically consists of deterministic part corresponding to the machine degradation called trend and random part corresponding to random nature of the data related to some fluctuation of degradation process, machine operating conditions, environmental impacts, measurement noise, etc. The random part in simplest cases may be just Gaussian or non-Gaussian heavy-tailed distributed noise or even as mentioned, a random process with some autodependence due to time-varying operational conditions. Moreover, even if random part is just Gaussian or non-Gaussian-distributed noise, the parameters of the distribution may be time-varying. The changing parameters may be just location or scale but also coefficients responsible for heavy-tailed (impulsive) behaviour. In many cases the transition between the regimes is very smooth, thus the proper identification of the data characteristics seems to be a very challenging task and application of one stochastic model seems to be insufficient. In case of the presence of non-Gaussian, non-homogeneous nature in the data, both modelling and prediction will be biased and may result in non-acceptable errors. Thus, it is necessary to precisely identify the structure of data and take into account all important components.

In this paper the framework for long-term complex data modelling is proposed. We describe in details, how to identify all components and how finally build a model for prognosis and Remaining Useful Life (RUL) estimation.

Some of RUL estimation procedures assume that the value of HI changes during the lifetime of the machine in a specific way, see Fig. 1. In this work, it is considered that the dataset contains three regimes (see panel (f) in Fig. 1) that may consist of trend and random components (see Fig. 2). Regime 1 assumes there is no degradation and some possible fluctuations might be present. Regime 2 describes linear growth of degradation with possible noise fluctuations. Regime 3 is exponential rapid growth of degradation. Such a model is frequently used in the literature [18,30]. Our preliminary analysis of available long-term data clearly indicates that the random part could be non-homogeneous (time-changing scale), strongly non-Gaussian and heavy-tailed. In such a case, even trend identification may be a challenging task. If the data contain also autodependence (periodic, seasonal behaviour), the presence of non-Gaussian noise makes identification of these components really complicated. For the non-Gaussian data, initially selected techniques (classical ones) may be inappropriate. Thus, it is critically important to develop a framework for how to process, analyse and model such data with time-varying properties.

The first goal of this paper is to introduce the framework for long-term data modelling. The second task is to demonstrate that the methodology can be considered in two versions, namely dedicated for Gaussian (classical approach) and non-Gaussian heavy-tailed (robust approach) distributed data. The general methodology is universal, the difference between these two approaches is related to the algorithms applied. The next goal is to prove that the classical methodology is efficient for Gaussian distributed HI data only. When the classical approach is applied for non-Gaussian heavy-tailed time series, the obtained results are not satisfactory. The last goal is to confirm that the analysed real HI datasets exhibit mentioned above characteristics and thus the known stochastic models may not cover all specific features of the time series.

The proposed approach has significant practical importance. First of all, the two considered datasets are from the real environment. Second, there are more examples confirming the presence of non-Gaussian noise in HI data, also time-varying scale

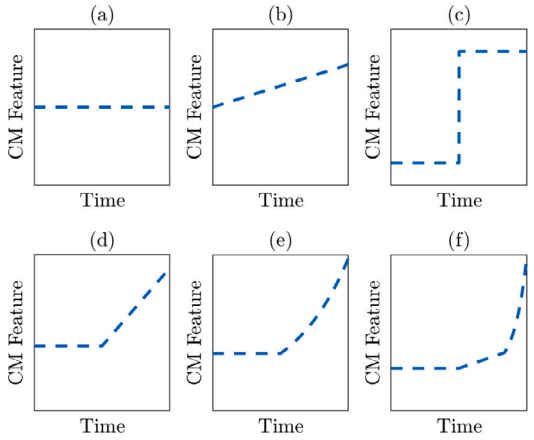

Fig. 1. Types of feature variations and degradation models: (a) good condition (constant trend), (b) good to gradual wear (linear trend) (c) switch from good to bad condition, (d) good to accelerated wear (linear trend), (e) good to accelerated wear (exponential trend), (f) three regimes model (good, linear progress and exponential progress of degradation).

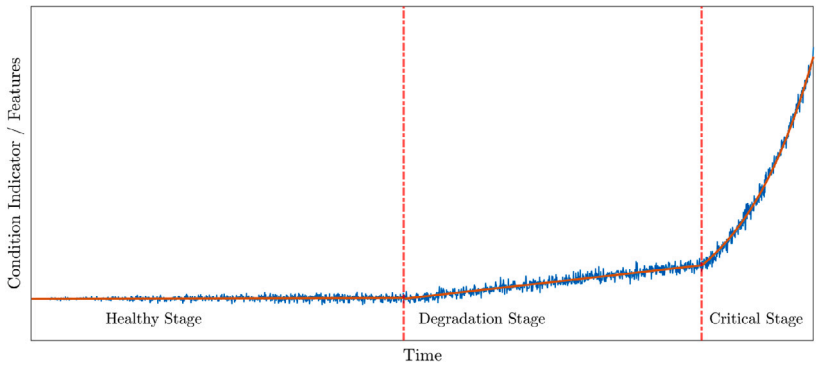

Fig. 2. The preliminary model used for our analysis. The red solid line represents the theoretical trend line, two vertical dashed lines point out the change of regime (Healthy Stage/Degradation/Critical stage). Note constant variance of random component.

is clearly seen. Finally, the difference between the basic approach (i.e., sum of trend and noise with constant variance) and the proposed novel methodology is significant. It may seriously influence the decision regarding the maintenance of machines.

The paper is organised as follows. First, we formulate the problem. Then, we propose a general framework for long-term data modelling for machine prognostics application. To illustrate the importance of the problem and the effectiveness of the proposed method, a deep simulation study is performed. Finally, two real datasets are used to prove our assumptions and show the efficiency of the framework. Two last sections cover discussion and conclusions.

2. Problem formulation

The ultimate goal of our research is to build the model based on the long-term historical data. The estimated model can be used to predict future values of given HI. The starting point is the dataset presented in Fig. 2. We extended this approach according to the adopted model (see Section 4.1 for a more detailed description). The model was built on the historical data. The preliminary analysis of the real HI clearly indicates that the data are non-homogeneous (their characteristics change in time), may contain some random components with significant autodependence and may exhibit non-Gaussian heavy-tailed behaviour. Moreover, we observe here three separate regimes that correspond to the three stages of the degradation process. The nature of the data may be different for different regimes. The nature of the deterministic component and scale of the data is different for the regimes but there is also a transition of the distribution for random component — from Gaussian (or almost Gaussian) distribution to strongly non-Gaussian distribution. In Fig. 3 we demonstrate the characteristics discussed above and their correspondence with the degradation process. Moreover, in Table 1 we show the main characteristics of the data corresponding to different regimes that result from the preliminary analysis of the real data.

All characteristics of the data mentioned above need to be investigated step-by-step and each component should be precisely identified in order to complete the model. Moreover, the procedure of the data modelling and identification needs to be automatic and dedicated for specific behaviour of the time series.

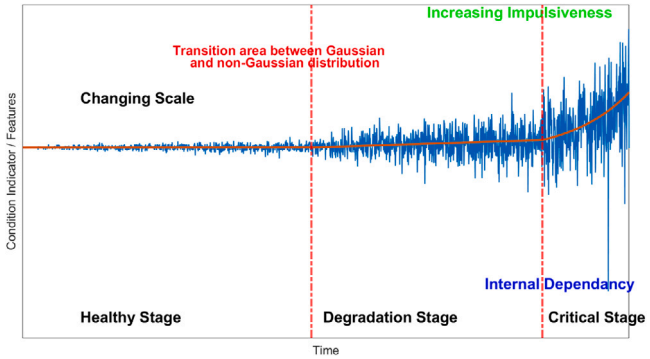

Fig. 3. Long-term data variation simulated from the adopted model.

Table 1
Main characteristics of the data for three regimes indicated in Fig. 3.

	Regime 1	Regime 2	Regime 3
Trend	Constant	Linear	Exponential
Scale	Nearly constant	Linearly growing	exp. growing
Autodependence of random component	Relatively small	Significant	Significant
Coefficients of the stochastic model	Negligible	Significant	Significant
Distribution of the random component	Nearly Gaussian	Non-Gaussian	Strongly non-Gaussian

The framework for data modelling is based on few important steps that are crucial in the classical data modelling. It is well known that mixture of random and deterministic components should be separated first and analysed using different approaches. There are well known techniques to perform such separation. However, in most cases, they assume the Gaussian nature of the random component. In Regime 1, for some cases the non-Gaussianity of the data might be considered as negligible, but for Regimes 2 or 3 — such assumption is not acceptable. Thus, in this case the dedicated algorithms need to be applied. After the separation of deterministic and random components, the next step is the analysis of the random component. In the simplest case, it is assumed that this component is a sequence of independent observations being realisations of identically distributed random variables. Moreover, in many cases it is assumed the data are Gaussian. However, the preliminary analysis of the real data indicates that the deterministic component cannot be considered as the sample corresponding to independent identically distributed (i.i.d.) random variables. As it was indicated, in the data we observe time-changing scale that needs to be eliminated before model fitting. To identify the scale for time-varying impulsive data, there is a need to apply robust algorithms. The final step after data normalisation (to make them homogeneous) is the analysis of its autodependence. Similar as for the previous steps, the possible non-Gaussian behaviour of the data requires to use the dedicated algorithms for the autodependence analysis and an optimal stochastic model fitting. The autodependence of the data is smaller for Regime 1 than for Regimes 2 and 3, thus the parameters of the fitted model (here the autoregressive model, AR model) in Regime 1 can be negligible, while for the two other regimes they are significant and cannot be ignored.

One may consider the described above steps as a standard task from data analysis classes. However, for non-Gaussian heavy-tailed data almost each step requires advanced tools, including robust methods for trend and time-changing scale detection and robust approaches for random component modelling. Presence of large observations may significantly change the complexity of data analysis. In this paper, we define a step-by-step framework, how to analyse the data, in order to build a model for complex, long-term datasets. Although our inspiration comes from predictive maintenance, we believe that it could be interesting also for wider audience, especially for researchers dealing with non-Gaussian, non-homogeneous, time-varying data.

3. Methodology: a framework for long-term data modelling

The main steps of the procedure are presented in Fig. 4. First, we describe the general methodology for the proposed framework. The procedure consists of few sub-procedures that can be divided into two main groups, namely deterministic and random components separation, and random component analysis (normalisation, modelling and testing). We demonstrate the algorithms used in the scheme for the data with Gaussian distribution (Section 3.1) and non-Gaussian behaviour (Section 3.2).

Deterministic and random components separation

First, we identify the deterministic component (denoted as $T(t)$) in the long-term data (denoted as $\{S(t)\}$). Our preliminary research indicates that for the data analysed in this paper, this component cannot be described by one single deterministic function. More precisely, we observe here a much more complex situation, when the type of the deterministic component (called here trend) changes depending on the regime corresponding to the good state, warning and alarm. Thus, in this paper, to identify the deterministic component, without the segmentation of the data, we calculate the empirical location measure for overlapping

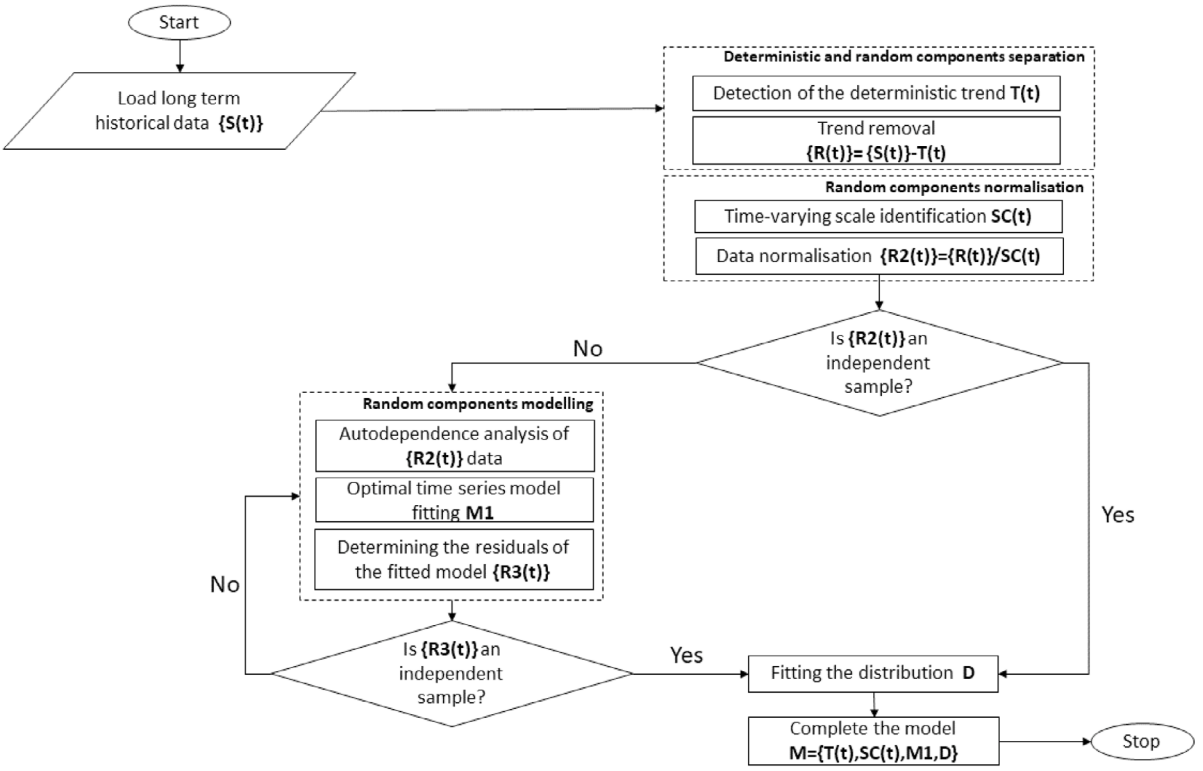

Fig. 4. The framework for modelling of long-term data. Each block of this diagram is described in details in Section 3.

segments from windows of given length w. As location measures we propose two statistics, see Section 3.1 for Gaussian distributed data and Section 3.2 for non-Gaussian case. This step is crucial for identifying the deterministic component that is further removed from the raw data. The series after removing the deterministic component is denoted as $\{R(t)\}$.

Random component normalisation

We assume that the time series $\{R(t)\}$ corresponds to the random component of the raw data. Our preliminary analysis clearly indicates that $\{R(t)\}$ has non-homogeneous structure that is mostly related to the time-varying scale. To make the data homogeneous, in the next step of our framework we identify the time-dependent empirical scale parameter and calculate it for segments of length w with overlapping o for the series corresponding to $\{R(t)\}$. As a result, we obtain the component $SC(t)$. Next, we normalise the data and obtain the series $\{R2(t)\}$ with homogeneous structure (time-constant characteristics). The problem of the time-varying scale identification is discussed separately in Sections 3.1 (for Gaussian case) and in Section 3.2 (for non-Gaussian case).

Random component modelling

As the next step, we analyse the autodependence of the time series $\{R2(t)\}$. This step is crucial to find the model describing such signal. If the sample autodependence measures indicate that the signal can be considered as a sequence of independent observations, then the next point in the framework is the proper distribution identification. Otherwise, there is a need first to fit the proper time series model and finally calculate its residuals, denoted as $\{R3(t)\}$. Autodependence measures can be also useful at this step to confirm that the residuals of the fitted model can be considered as independent observations. Finally, the proper distribution D is fitted to the series $\{R3(t)\}$ (or the series $\{R2(t)\}$). The sample autodependence measures used in the mentioned above steps are described in Section 3.1 (for Gaussian case) and in Section 3.2 (under the assumption of non-Gaussian distribution). We observe that the series $\{R2(t)\}$, obtained after the normalisation by the time-dependent scale parameter, exhibits behaviour adequate to the linear autoregressive model. In the framework, we call it $M1$. We remind that the time series $\{R2(t)\}$ is the stationary AR model with order $p > 0$ (AR(p)) if it satisfies the following equation

$$R2(t) - \phi_1 R2(t-1) - \cdots - \phi_p R2(t-p) = R3(t), \tag{1}$$

where the polynomial

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \tag{2}$$

for all $|z| = 1$ and $\{R3(t)\}$ is assumed to be i.i.d. time series [31]. In the classical approach, it is assumed that the residuals $\{R3(t)\}$ are Gaussian distributed, however there are also considered cases with non-Gaussian heavy-tailed distributed noises. The data analysed in this paper also exhibit such property.

To fit the proper distribution (and confirm the possible non-Gaussian behaviour), in this paper we propose to use the simple visual test based on the comparison of the empirical tail and the theoretical one corresponding to the tested distribution (with the estimated parameters from considered random sample). By this visual inspection, we can also conclude, which of the tested distributions is more proper for the examined data. We remind that the tail of the distribution of a random variable X is defined as $1-F_X(x)$, where $F_X(\cdot)$ is the cumulative distribution function (CDF) of X. The empirical tail of the random sample $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ is defined as $1 - \hat{F}_{\mathbf{x}}(x)$, where $\hat{F}_{\mathbf{x}}(\cdot)$ is the empirical CDF for x defined as [32]

$$\hat{F}_{\mathbf{x}}(x) = \frac{1}{n}\sum_{j=1}^{n} 1\{x_j \le x\}, \tag{3}$$

where $1\{A\}$ is the indicator of a set A. We consider Gaussian distribution, α-stable distribution and scaled Student's t distribution as the tested theoretical distributions. The α-stable and scaled Student's t distributions belong to the class of non-Gaussian heavy-tailed distributions. The definitions and the main properties of the considered distributions are presented in Appendix A. In the simulation study presented in Section 4.2, to confirm that the tested distribution is the proper one (or more adequate than the other distribution), we use the Kolmogorov–Smirnov (KS) statistic that is defined as

$$KS = \sup_x \left|\hat{F}_{\mathbf{x}}(x) - F_X(x)\right|, \tag{4}$$

where $F_X(\cdot)$ is the CDF of the tested theoretical distribution with the parameters estimated from the random sample $\mathbf{x}$. A small value of the KS statistic indicates that the empirical distribution is close to the tested one. For the estimation of the parameters, for all considered distributions, we use the maximum likelihood (MLE) method.

At the final step of the proposed framework, we synthesise the model: the deterministic trend $T(t)$, time-varying scale $SC(t)$, time series model $M1$ and the distribution D fitted to the independent observations corresponding to the series $\{R3(t)\}$ (or, alternatively, to $\{R2(t)\}$).

In the following subsections, we present in detail the algorithms used in the presented framework. As it was mentioned, for all of the steps of the procedure there is a need to replace the classical techniques (dedicated for Gaussian distributed models) by the robust versions that are more useful for non-Gaussian heavy-tailed data.

3.1. Algorithms for Gaussian distributed data

Deterministic and random components separation

In the general framework, the first crucial step is the identification of the deterministic trend $T(t)$. The classical statistic considered as the location measure is just a sample mean. In this case, we just calculate moving average (MA) for overlapping segments (with overlapping o) of length w and consider it as the deterministic component of the signal $\{S(t)\}$.

Random component normalisation

The classical estimator of the scale parameter is the sample standard deviation that for the signal $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ is defined as

$$SC_{\mathbf{x}} = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - \overline{\mathbf{x}}\right)^2}, \tag{5}$$

where $\overline{\mathbf{x}}$ is the sample mean for the signal $\mathbf{x}$. Because in the data we observe time-changing scale, in the Gaussian case we apply the empirical standard deviation for the segments of length w and overlapping o. In that case, the segments are considered as quasi-homogeneous sub-signals.

Random component modelling

In this part of the proposed framework, first we analyse the autodependence of the normalised signal corresponding to $\{R2(t)\}$. This part is crucial to confirm that the signal can be considered as an independent sequence or to find the proper time series model describing the data (i.e., $M1$). The most classical measure of autodependence used in the time series analysis is the autocorrelation function (ACF). Its sample version for the stationary signal $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ is defined as follows [31]

$$\hat{\rho}_{\mathbf{x}}(h) = \frac{\hat{\gamma}_{\mathbf{x}}(h)}{\hat{\gamma}_{\mathbf{x}}(0)}, \quad -n < h < n, \tag{6}$$

where $\hat{\gamma}_{\mathbf{x}}(\cdot)$ is the sample autocovariance function (ACVF) given by

$$\hat{\gamma}_{\mathbf{x}}(h) = \frac{1}{n}\sum_{j=1}^{n-|h|}(x_j - \overline{\mathbf{x}})(x_{j+|h|} - \overline{\mathbf{x}}) \tag{7}$$

and h is called a lag. In the case, when the analysed signal has a finite variance distribution (like Gaussian one), the $\hat{\rho}_{\mathbf{x}}(\cdot)$ is considered as the efficient estimator of the ACF. The visual inspection of its behaviour can give information if the considered signal exhibits (long- or short-range) dependence or if it can be considered as an independent sample. The significant level of the $\hat{\rho}_{\mathbf{x}}(\cdot)$ function for $h > 0$ indicates the autodependence of the considered signal. In the proposed framework, the sample ACF is applied for the Gaussian case for the signal corresponding to $\{R2(t)\}$ as well as to $\{R3(t)\}$, being a residual series of the fitted time series model.

In the random component's modelling part, after the confirmation that data are dependent, we can fit the optimal time-series model. Here we analyse the AR(p) model defined in Eq. (1). To find the optimal order p, we propose to use the criterion which is universal and can be used without the assumption of the data distribution. For given $p = 1, 2, \dots, p_{max}$, we estimate the AR(p) model's parameters using the approach presented below (i.e. in the Gaussian case, the Yule–Walker method). Then, we calculate the estimated model's residuals and analyse their sample ACF. More precisely, we calculate the statistic defined as

$$K_{\mathbf{x}}(p) = \max_{h=1,\dots,h_{max}} |\hat{\rho}_{\mathbf{x}}(h)|^2, \tag{8}$$

where $\hat{\rho}_{\mathbf{x}}(\cdot)$ is defined in Eq. (6) for residual series from fitted AR(p) model and h_{max} is the fixed maximum value of the lag. The optimal order p_{opt} is selected when the statistic $K_{\mathbf{x}}(p_{opt})$ takes the minimum value. This simple method is intuitive. The optimal model is considered when its residuals can be considered as an independent sample, i.e., when their sample ACF for $h > 0$ is equal (or very close) to zero.

After identification of the proper order p, we estimate the parameters of the AR(p) model. The universal approach for finite-variance models is the Yule–Walker method which utilises the ACVF of the time series model. In the algorithm, the theoretical ACVF is replaced by the empirical one defined in Eq. (7). More details of the Yule–Walker algorithm for the AR(p) model with finite-variance distribution one can find in [31].

3.2. Algorithms for non-Gaussian heavy-tailed distributed data (robust approach)

Deterministic and random components separation

When the data exhibits a non-Gaussian behaviour, then the MA may be insufficient to properly identify the deterministic component as it is sensitive to large observations. This is especially important in case when the window length w is relatively small and single outliers occur in the signal. The large values may significantly influence the sample mean and the identified trend may not correspond to the theoretical one. Thus, for the non-Gaussian data, it is recommended to use more robust statistic as the empirical location measure. In this paper, we apply the moving median (MM) used for segments from windows of length w with overlapping o. The sample median, similarly as the sample mean, is considered as an unbiased estimator of the theoretical mean for symmetric data. However, in contrast to the classical statistic, it is not so sensitive to large observations.

Random component normalisation

In case when the data exhibit non-Gaussian behaviour, the theoretical variance (and thus standard deviation) may be infinite (as for the α-stable distribution with $\alpha < 2$). Thus, in this case we propose to use the robust estimator of the scale parameter for the signal $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ which is defined through the kth order statistic [33–35]

$$SC^{Q}_{\mathbf{x}} = d\{|x_i - x_j|; i < j\}_{(k)}, \tag{9}$$

where d is a constant value and $k = \binom{h}{2} \approx \binom{n}{2}/4$, where $h = \lfloor n/2 \rfloor + 1$ and $\lfloor \cdot \rfloor$ denotes the integer part.

Random component modelling

To identify autodependence, the classical ACF cannot be used for non-Gaussian heavy-tailed data. Thus, we propose to apply the robust version of the sample ACF defined in Eq. (6) for the classical (Gaussian) case. Based on the sorted signal $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ we construct two vectors, the vector $\mathbf{u}$ that contains the first $n - h$ observations and vector $\mathbf{v}$ which contains last $n - h$ observations. Then, the robust sample ACF is defined as follows [34]

$$\hat{\rho}^{Q}_{\mathbf{x}}(h) = \frac{(SC^{Q}_{\mathbf{u+v}})^2 - (SC^{Q}_{\mathbf{u-v}})^2}{(SC^{Q}_{\mathbf{u+v}})^2 + (SC^{Q}_{\mathbf{u-v}})^2}, \tag{10}$$

where $SC^{Q}_{\mathbf{x}}$ is defined in Eq. (9). The robust version of ACVF is given by

$$\hat{\gamma}^{Q}_{\mathbf{x}}(h) = \frac{1}{4}\left[(SC^{Q}_{\mathbf{u+v}})^2 - (SC^{Q}_{\mathbf{u-v}})^2\right]. \tag{11}$$

To find the optimal order p of the autoregressive model for the non-Gaussian case, we apply here the novel method that can be considered as the robust version of the similar methodology presented in the Gaussian case. More precisely, for given $p = 1, 2, \dots, p_{max}$, we calculate the statistic $K_{\mathbf{x}}(p)$ defined in Eq. (8), however the sample ACF defined in Eq. (6) is replaced here by its robust version given in Eq. (10). Moreover, to estimate the AR model's parameters, we use the dedicated methodology presented below (i.e. robust version of the Yule–Walker method). Finally, the optimal order p_{opt} is selected as the argument for which the statistic $K_{\mathbf{x}}(p)$ takes the minimum value.

Finally, we estimate the parameters of the optimal AR(p) model using the robust version of the Yule–Walker method mentioned in Section 3.2. However, here as the sample ACVF we use the robust estimator of this measure defined in Eq. (11).

3.3. Summary

In Table 2 we present the summary of the presented above framework. We indicate here the main differences between the proposed algorithms for the classical (dedicated for Gaussian distributed data) and robust (dedicated for non-Gaussian heavy-tailed distributed data) approaches.

Table 2
The summary of the framework for data modelling — classical and robust approaches.

	Classical approach	Robust approach
Deterministic and random component separation		
Deterministic component identification	Moving average	Moving median
Random component normalisation		
Time-varying scale identification	Sample standard deviation	Sample robust scale parameter
Autodependence identification	Sample ACF	Sample robust ACF
Random component modelling		
Identification of the model's order	Based on sample ACF	Based on sample robust ACF
Estimation of the model's parameters	Classical Yule–Walker approach	Robust Yule–Walker approach
Identification of the residuals autodependence	Sample ACF	Sample robust ACF
Testing distribution of the residual series	Gaussian distribution testing	Selected non-Gaussian distribution testing
Distribution fitting	MLE method for Gaussian distr.	MLE method for identified non-Gaussian distr.

4. Validation of the proposed approach by simulated data analysis

4.1. Model of the signal

Considering the data characteristics discussed in Section 2, we propose the following model of the signal $\{S(t)\}$ that will be used in simulation study

$$S(t) = R(t) + T(t), \tag{12}$$

where $\{R(t)\}$ and $T(t)$ are, respectively, random and deterministic (i.e. trend) components. Both these parts consist of three regimes, denoted as Regime 1, Regime 2 and Regime 3, which are related to three considered underlying states (healthy/warning/alarm) and determine the behaviour of the process with regard to both trend and noise's scale. Let us assume that we have a sample signal $S(1), \dots, S(N)$. The changepoint between Regimes 1 and 2 is denoted as τ_1, and the changepoint between Regimes 2 and 3 as τ_2, where $1 < \tau_1 < \tau_2 < N$. In other words, we can divide the signal regime-wise, so that the sequence $S(1), \dots, S(\tau_1)$ corresponds to Regime 1, the sequence $S(\tau_1 + 1), \dots, S(\tau_2)$ corresponds to Regime 2 and the sequence $S(\tau_2 + 1), \dots, S(N)$ corresponds to Regime 3. The same nomenclature we apply to the sequences $\{R(t)\}$ and the function $T(t)$.

The random component corresponding to $\{R(t)\}$ is constructed in the following way. First, we generate a trajectory of autoregressive time series of order p corresponding to the time series $\{R2(t)\}$ satisfying Eq. (1). The sequence $\{R3(t)\}$ consists of i.i.d. random variables from a "standard" version of considered distribution (i.e. the one with a unit scale). In the simulation study presented in the next subsection, we assume two distributions of the random component, namely Gaussian and α-stable. For the Gaussian distribution, we assume $R3(t) \sim \mathcal{N}(0, 1)$, and for the α-stable case $R3(t) \sim S(\alpha, 0, \frac{1}{\sqrt{2}}, 0)$. For the simplicity, we assume that the distribution in each regime is the same, however, as it was mentioned in Section 2, in practice it may be different for different regimes. Finally, the random component of the model is as follows

$$R(t) = SC(t)R2(t), \tag{13}$$

where the function $SC(t)$ represents the time-changing scale. As it was mentioned, its behaviour is different for each regime, namely, we assume that in Regime 1 the scale grows linearly, from σ_1 to σ_2 (where both values are relatively close to each other), then in Regime 2 it also increases linearly, from σ_2 to σ_3, and finally in Regime 3 it grows exponentially, from σ_3 to σ_4. The function $SC(t)$ is defined as follows

$$SC(t) = \begin{cases} a_1 t + b_1 & 0 < t \le \tau_1, \\ a_2 t + b_2 & \tau_1 < t \le \tau_2, \\ a_3 \exp(b_3 t) & \tau_2 < t \le N, \end{cases} \tag{14}$$

where constants a_1, b_1, a_2, b_2, a_3, b_3 are derived in such a way that $SC(1) = \sigma_1$, $SC(\tau_1) = \sigma_2$, $SC(\tau_2) = \sigma_3$ and $SC(N) = \sigma_4$.

As mentioned, the behaviour of the deterministic component $T(t)$ in Eq. (12) has different nature for different regimes. Let us recall that in the Regime 1 it is at the fixed constant level, denoted here as c_1. Then, in Regimes 2 and 3 we consider linear and exponential functions, respectively, with the same growth parameters as for the corresponding regimes of the scale function $SC(t)$. Moreover, we assume the function $T(t)$ has no discontinuities in regime changepoints τ_1 and τ_2. Under these assumptions, the deterministic term of the signal takes the form

$$T(t) = \begin{cases} c_1 & 0 < t \le \tau_1, \\ a_2 t + c_2 & \tau_1 < t \le \tau_2, \\ a_3 \exp(b_3 t) + c_3 & \tau_2 < t \le N, \end{cases} \tag{15}$$

where c_2 and c_3 are derived in such a way that $T(t)$ is continuous function. In Fig. 5 we present the deterministic component $T(t)$ and the scale function $SC(t)$ for the following values of the parameters: $\tau_1 = 600$, $\tau_2 = 900$, $N = 1000$, $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 10$, $\sigma_4 = 40$ and $c_1 = 10$. Moreover, in Fig. 6 we demonstrate the simulated signals corresponding to the sequence $\{S(t)\}$, where the random

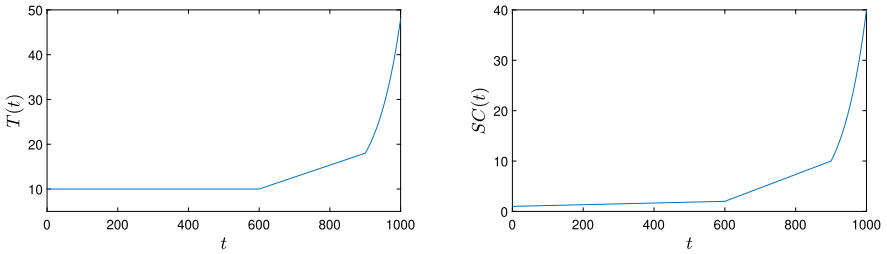

Fig. 5. The deterministic component $T(t)$ and the scale function $SC(t)$ for the following values of the parameters of the model: $\tau_1 = 600$, $\tau_2 = 900$, $N = 1000$, $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 10$, $\sigma_4 = 40$ and $c_1 = 10$.

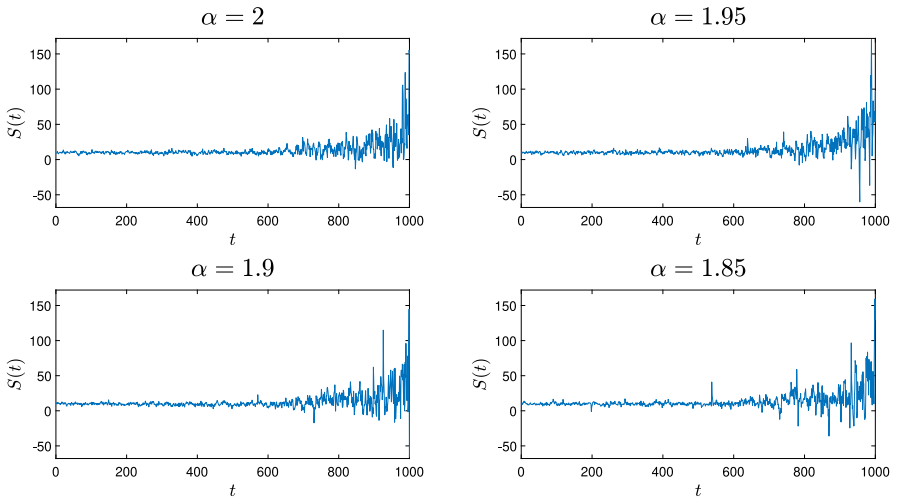

Fig. 6. The exemplary data from the model with deterministic component $T(t)$ and scale function $SC(t)$ presented in Fig. 5. The random components of the model correspond to the AR(1) model with $\phi = 0.5$. Four cases of distributions are considered: Gaussian (α-stable with $\alpha = 2$), and non-Gaussian α-stable with $\alpha = 1.95, 1.9$ and $\alpha = 1.85$.

component $\{R2(t)\}$ is the autoregressive model defined in Eq. (1) for $p = 1$ and $\phi = 0.5$ while the sequence $\{R3(t)\}$ is from Gaussian (i.e., α-stable with $\alpha = 2$) and non-Gaussian α-stable distribution. Here we consider three additional values of the α- parameter, namely $\alpha \in \{1.95, 1.9, 1.85\}$. As can be seen, for the Gaussian distributed signal we do not observe the large observations, while for the non-Gaussian heavy-tailed case, the large impulses occur in the data. The smaller the α, the higher impulses may occur in the signal.

4.2. Simulation study

In this section, we present the efficiency of the presented methodology for simulated signals from the model in Section 4.1. The parameters used in the model are as follows: $\tau_1 = 600$, $\tau_2 = 900$, $N = 1000$, $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 10$, $\sigma_4 = 40$ and $c_1 = 10$. Moreover, the random part corresponds to the AR(1) model with $\phi = 0.5$. In the simulation study and real data analysis, $t = 1, 2, \ldots, N$ denotes the time point (number of observation).

Here we analyse four cases of the distributions of the $\{R3(t)\}$ sequence, namely Gaussian (i.e. α-stable with $\alpha = 2$) and three cases of non-Gaussian α-stable distribution, $\alpha \in \{1.95, 1.9, 1.85\}$. In each of the considered cases, we simulate 100 datasets from the proposed model. The exemplary signals are presented in Fig. 6.

4.3. Application to simulated data

For each of the considered cases of $\{R3(t)\}$ sequence's distribution, and for each of the simulated datasets, we apply the methodology described in Section 3. We assume that the window length for trend identification is equal to 101 observations, while for the scale identification it is equal to 50. These values have been set experimentally (however, their selection is not critical). For each step of the methodology, we use the classical approach (dedicated for the Gaussian distributed signals) – see the left panels of Figs. 7–10 – and the robust approach (dedicated for the non-Gaussian heavy-tailed distributed signals) — see the middle panels of Figs. 7–10. In the left and middle panels, by grey lines we present the results for each of the simulated datasets while by the black line we show the corresponding theoretical functions.

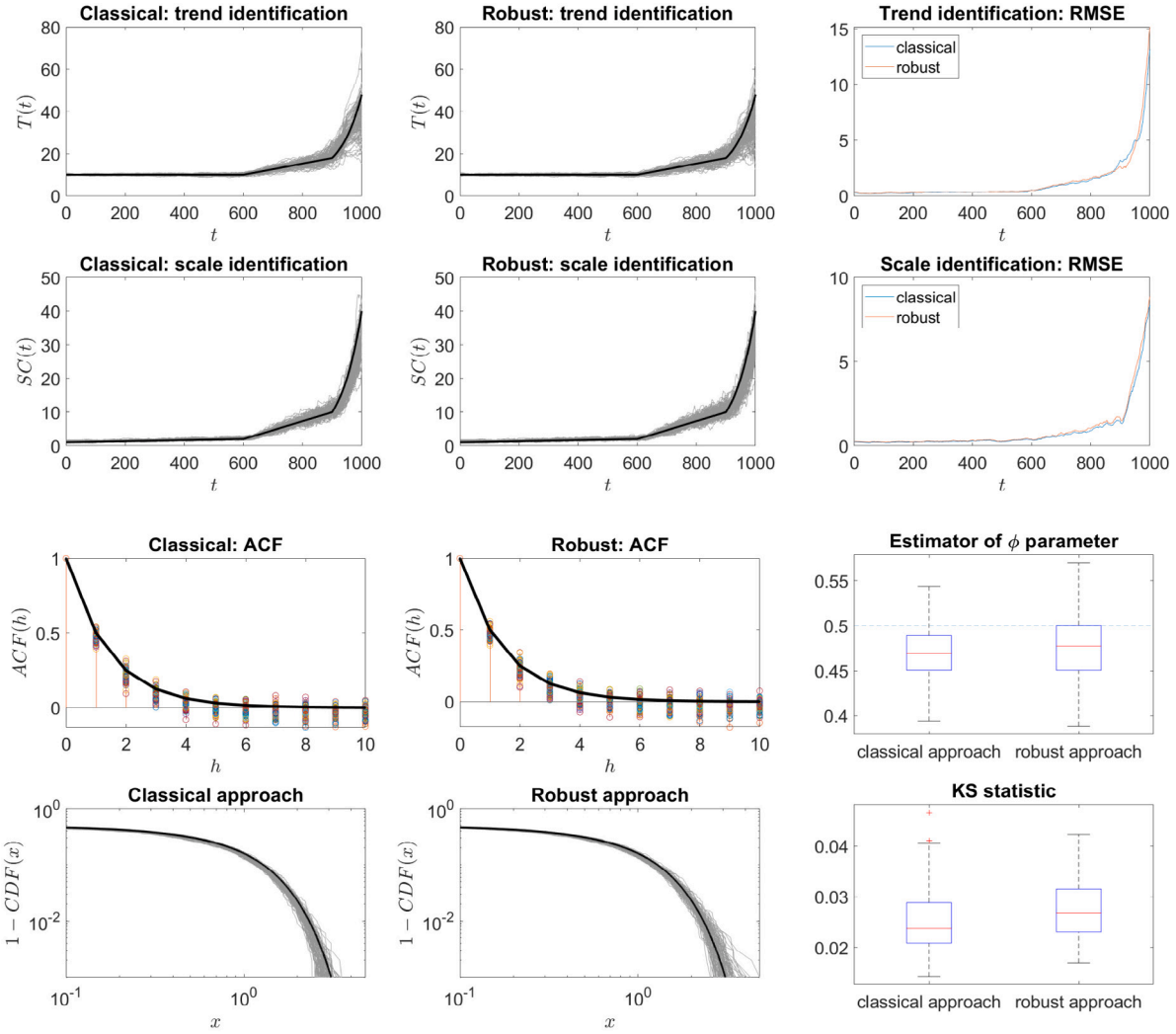

Fig. 7. Results of the proposed procedures for simulated data (trend, scale, AR model and distribution tail identification in rows 1–4, respectively). Left column: the classical approach (dedicated for Gaussian distributed signals), middle column: robust approach (dedicated for non-Gaussian heavy-tailed distributed signals), right column: evaluation of the results by RMSE (trend and scale) and boxplots (random components). Gaussian distributed (i.e. α-stable with $\alpha = 2$) random component $\{R3(t)\}$.

In the fourth row, on the left and middle panels, we demonstrate the empirical tails (i.e. $1 - \hat{F}_{\mathbf{x}}(x)$) for the residual series corresponding to the sequence $\{R3(t)\}$ obtained by the Yule–Walker method with the classical (left panel) and robust version (middle panel), while by the black lines we present the tails of the theoretical (tested) distribution (i.e. $1 - F_X(x)$).

In the right panels (last columns) of Figs. 7–10, we demonstrate the statistics that give us a possibility to parametrise and compare the results for classical and robust algorithms. In the first and second rows of Figs. 7–10, we use the root mean square errors (RMSE) of the identified trends (first rows) and scales (second rows) calculated for each $t = 1, 2, \ldots, N$ separately. We remind that the RMSE is the root of square difference between the theoretical value and its estimator. RMSE is a frequently used measure of the differences between predicted values (here appropriate functions estimated) and the theoretical values (represented by black lines on the plots). The RMSE serves to aggregate the magnitudes of the errors in predictions for various data points into a single measure of predictive power. RMSE is always non-negative, and a value of 0 (almost never achieved in practice) would indicate a perfect fit. In general, a lower RMSE is better than a higher one. Here the RMSE is calculated along the datasets separately for each time point $t = 1, 2, \ldots, N$. In the third rows of Figs. 7–10, we present the boxplots of the estimators of ϕ parameter that, according to the Yule–Walker methodology, is calculated based on the sample (or robust sample) ACVF. In our case, by the assumption $\phi = 0.5$, the box closer to the theoretical value corresponds to the more efficient technique. In descriptive statistics, a boxplot is a method for graphically demonstrating the locality, spread and skewness groups of numerical data through their quartiles. Boxplots are non-parametric: they display variation in samples of a statistical population without making any assumptions about the underlying statistical distribution. The spacings in each subsection of the boxplot indicate the degree of dispersion (spread) and skewness of the

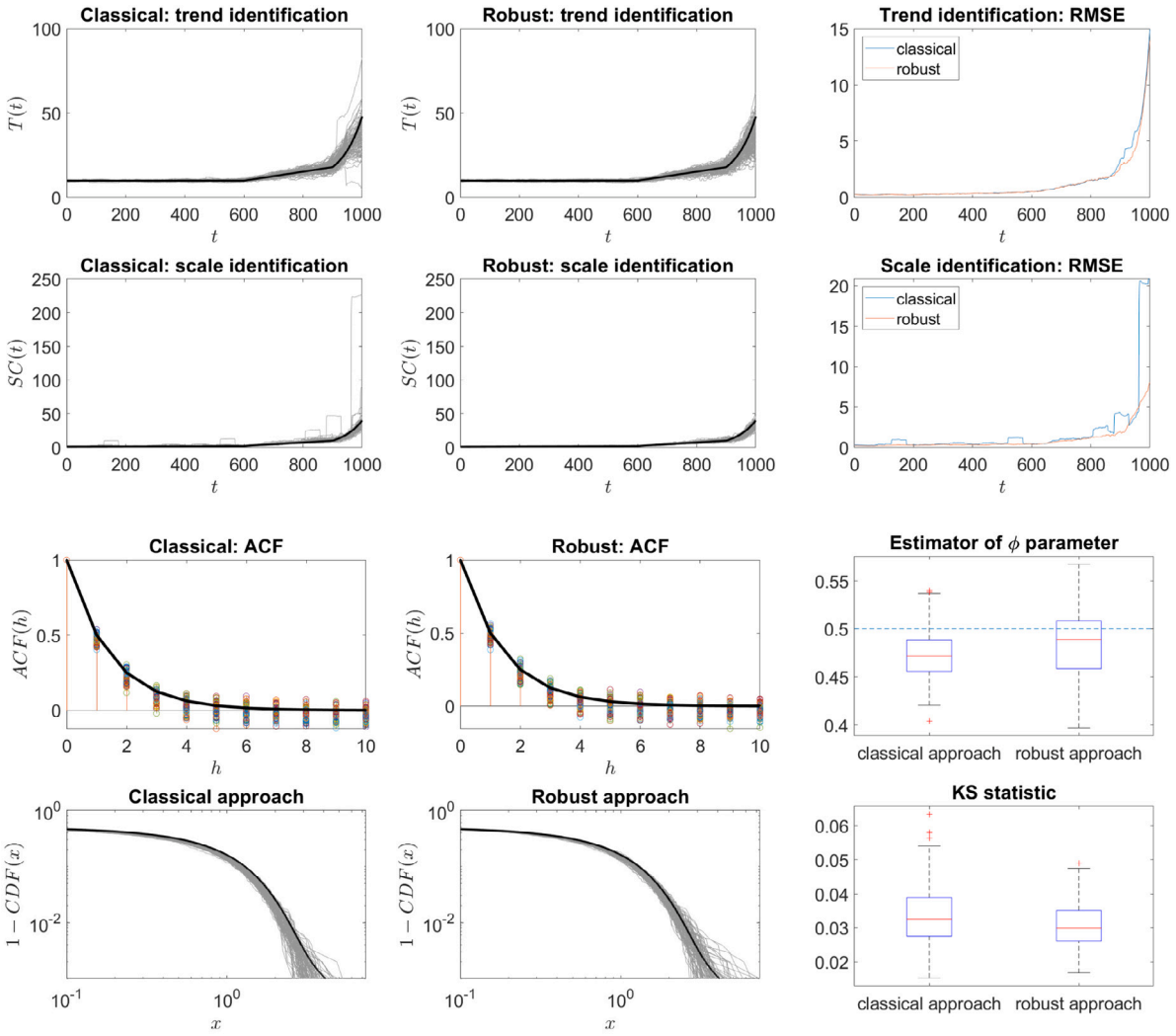

Fig. 8. Results of the proposed procedures for simulated data (trend, scale, AR model and distribution tail identification in rows 1–4, respectively). Left column: the classical approach (dedicated for Gaussian distributed signals), middle column: robust approach (dedicated for non-Gaussian heavy-tailed distributed signals), right column: evaluation of the results by RMSE (trend and scale) and boxplots (random components). The α-stable distributed random component $\{R3(t)\}$ with $\alpha = 1.95$.

data, which are usually described using the five-number summary. The boxplot is often used to compare various methods visually. The boxplot with median closer to the theoretical value and with smaller box length indicates the more efficient method.

Finally, in the fourth rows and right panels of Figs. 7–10, we demonstrate the KS statistic defined in Eq. (4) for the tested distribution defined through the CDF $F_X(x)$. Smaller values of the KS statistic indicate that the empirical distribution of given sample is closer to the tested theoretical distribution. The KS statistic is considered as one of the most common methods to quantify a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. It is also used as a test statistic for the KS goodness-of-fit test.

In Fig. 7, we demonstrate the results for the Gaussian distributed model. In this case, the classical and robust methods give comparable results. The RMSE values for trend and scale identification are at comparable levels. Moreover, the estimators of ϕ parameter by using Yule–Walker methodologies are close to each other. Finally, the KS statistics for Gaussian distribution testing applied to the series corresponding to $\{R3(t)\}$ are relatively small and take similar values. This indicates that for the Gaussian distributed signals the methodology utilising the classical algorithms is efficient.

The situation changes in case when the analysed signals exhibit non-Gaussian heavy-tailed behaviour, i.e. when the random component $\{R3(t)\}$ is α-stable distributed with $\alpha < 2$. The corresponding cases are demonstrated in Figs. 8–10. One can observe here that the robust algorithms are more effective than the classical techniques. The RMSE values for the trend and the scale identification are the smallest for the algorithms dedicated for the non-Gaussian heavy-tailed distributed signals. Moreover, the estimators of the

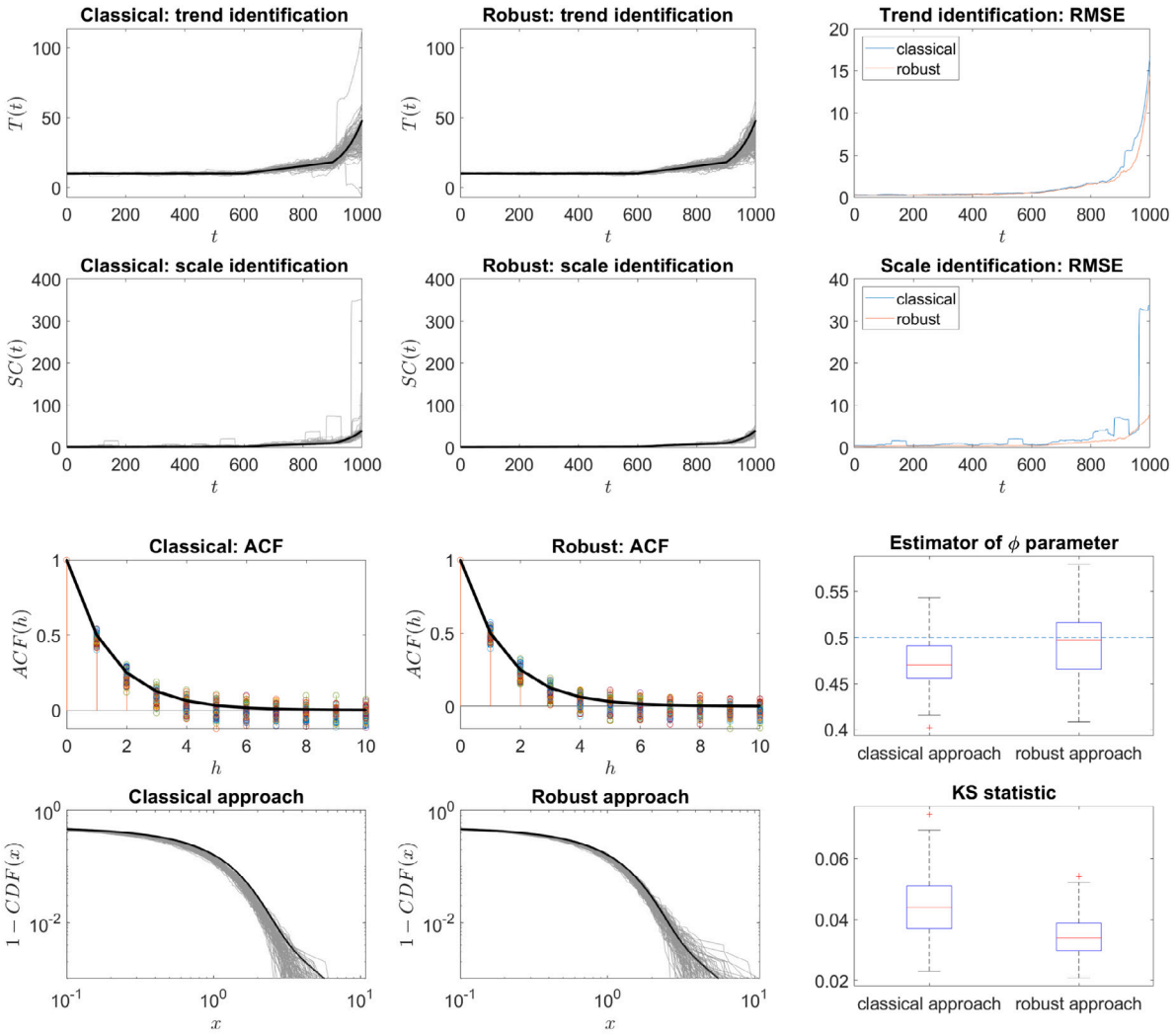

Fig. 9. Results of the proposed procedures for simulated data (trend, scale, AR model and distribution tail identification in rows 1–4, respectively). Left column: the classical approach (dedicated for Gaussian distributed signals), middle column: robust approach (dedicated for non-Gaussian heavy-tailed distributed signals), right column: evaluation of the results by RMSE (trend and scale) and boxplots (random components). The α-stable distributed random component $\{R3(t)\}$ with $\alpha = 1.9$.

ϕ parameter are closer to the theoretical value $\phi = 0.5$ which indicates the robust algorithm outperforms the classical Yule–Walker methodology. Finally, the comparison with the tested distributions (i.e. α-stable with $\alpha = 1.95, 1.9, 1.85$, respectively) presented in the fourth rows of Figs. 8–10 clearly indicates that the empirical distribution of the signal corresponding to $\{R3(t)\}$ sequence coincides with the tested one in case when the robust version of the procedure is applied. The value of the KS statistic is closer to zero for the robust approach in comparison to the classical methodology. The smaller the α parameter, the efficiency of the robust algorithms is more evident. The presented simulation study clearly confirms the justification for applying the dedicated algorithms for the heavy-tailed distributed signals. Moreover, the robust algorithms can be considered as universal as they are also effective for Gaussian data. The procedure was also applied for the data with t location-scale distributed random component $\{R3(t)\}$ for two values of ν parameters, namely $\nu = 9$ and $\nu = 6$ under the assumption $\mu = 0$ and $\sigma = 1$. The results are presented in Appendix B in Figs. B.20 and B.21. The selected values of ν parameters correspond to the estimated values of the number of degrees of freedom for the real data analysed in the next section. Similarly, as for the α-stable case, we observe that for the smaller value of the ν parameter (i.e. when the t location-scale distribution is far from the Gaussian one) the results obtained using the robust approach are closer to the reality. This is especially visible on the level of the ϕ parameters estimation and the distribution fitting. On the other hand, when the distribution of random component $\{R3(t)\}$ is closer to Gaussian, the results for classical and robust approaches become comparable. This confirms the universality of the proposed robust approach.

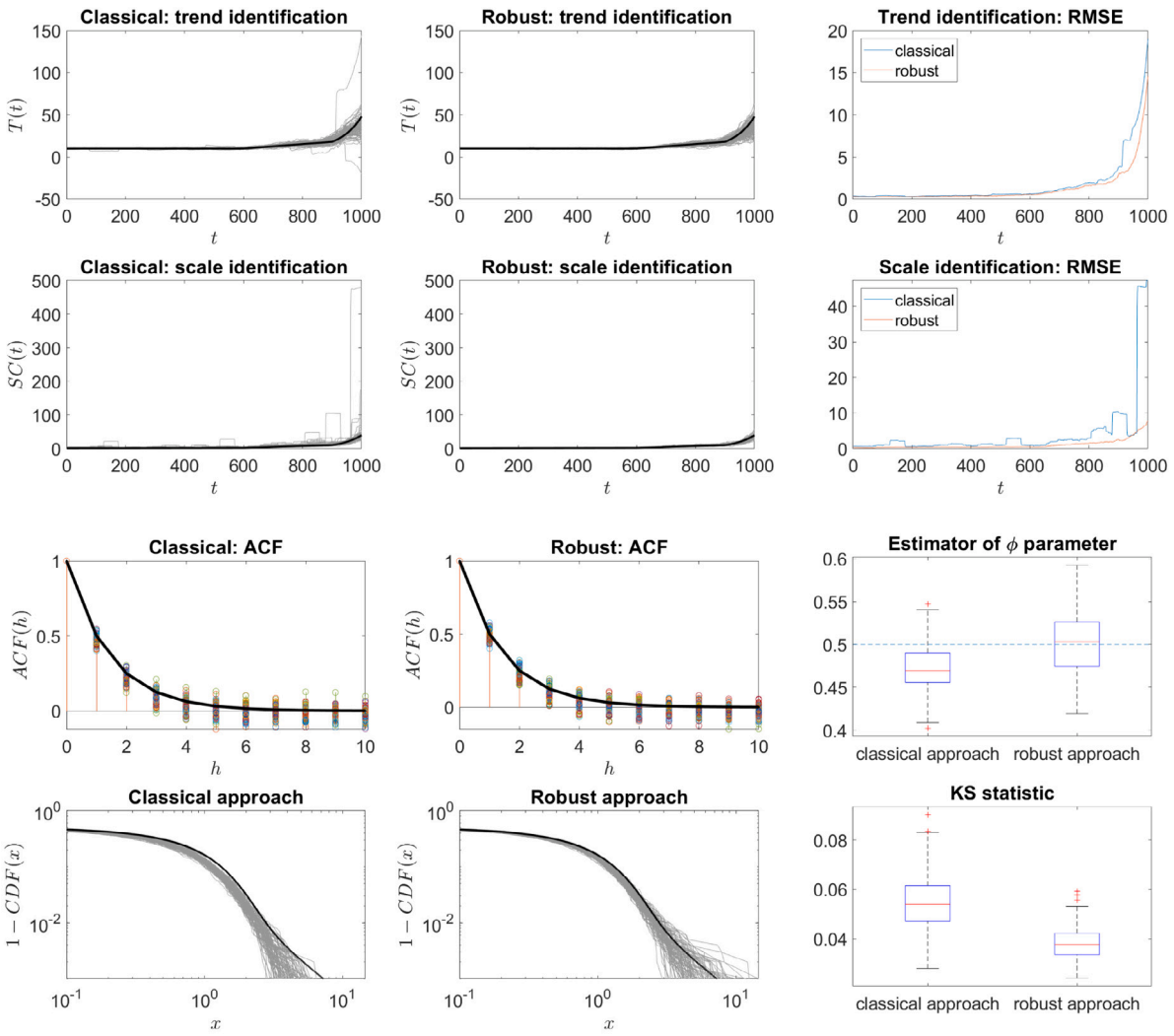

Fig. 10. Results of the proposed procedures for simulated data (trend, scale, AR model and distribution tail identification in rows 1–4, respectively). Left column: the classical approach (dedicated for Gaussian distributed signals), middle column: robust approach (dedicated for non-Gaussian heavy-tailed distributed signals), right column: evaluation of the results by RMSE (trend and scale) and boxplots (random components). The α-stable distributed random component $\{R3(t)\}$ with $\alpha = 1.85$.

5. Real data analysis

In this section, the novel procedure will be applied and evaluated for available real datasets. These data are commonly used as benchmark datasets for various papers and competitions and have specific behaviour related to the properties of the noise. Below basic information about objects, experiments, and data will be recalled and suitable references will be provided.

5.1. Description of the analysed datasets

5.1.1. FEMTO dataset

The FEMTO dataset was acquired by Franche-Comté Electronics Mechanics Thermal Science and Optics–Sciences and Technologies institute from a PRONOSTIA platform, see Fig. 11. Also, it is published as a prognosis challenge at the IEEE International Conference of PHM 2012. This collection of data includes 17 lifetime curves that cover historical degradation of bearing. Two accelerometers and a temperature sensor are used to acquire acceleration and temperature. The speed of the shaft was kept stable during the test. In the considered dataset, it is assumed that the failure of the bearing occurs when the amplitude of the vibration signal has arrived above 20 g [29].

The FEMTO time series is a realistic dataset acquired in the lab. The bearings were naturally degraded without any initial artificial fault. In this dataset, the fault pattern, as well as the degradation trend of bearings, are different under the same operating

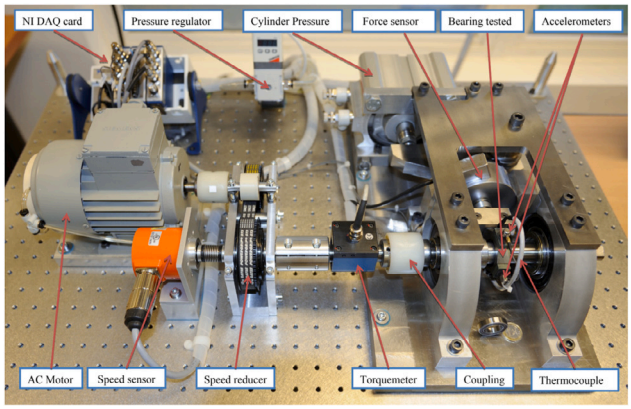

Fig. 11. Overview of PRONOSTIA FEMTO testrig [36].

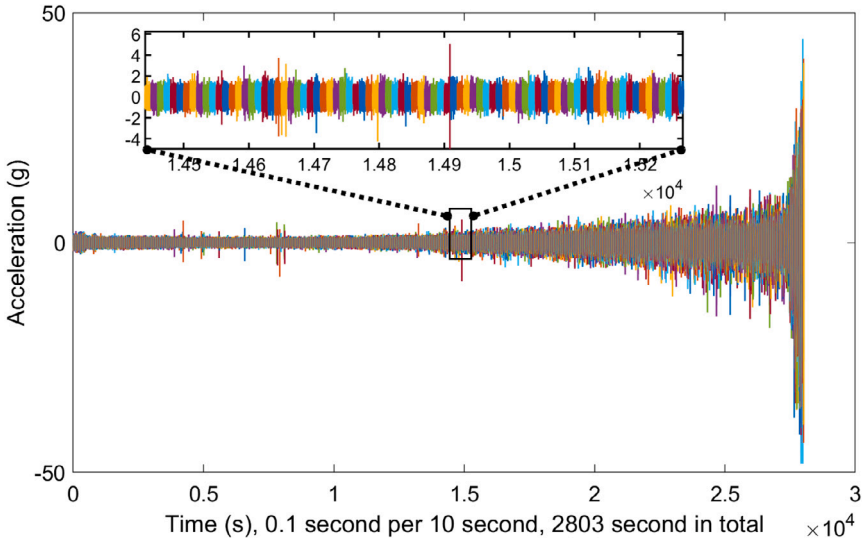

Fig. 12. Raw bearing run-to-failure vibration signals for FEMTO dataset. Main plot — whole vibration signal from run-to-failure test, rectangular box inside: its exemplary part (zoom). Note that each segment in colour means different measurement of 10 s signal — see description in the section.

condition. It should be noted that the FEMTO has been frequently used in various application as feature extraction [20,37–43], segmentation [44–48], RUL prediction [10,49–53], etc.

In Fig. 12, we present the historical data variation i.e. vibration signal under test conditions for one of the case studies in the FEMTO dataset. The shaft speed is approximately 1800 rpm and the load is equal to 4000 N. In this study, for following the degradation of bearing, one of the classical extracted features, root mean square (RMS), is used as HI. The RMS values for the mentioned test conditions are demonstrated in Fig. 13. Note that both subplots present the same data, the top subplot as a cloud of points, the bottom subplot as a continuous function. The reason to present the data in two different ways is to highlight that the impulse in the bottom plot is just a single outlier in the time series. In panel (b), we present the zoom of the first 900 data to highlight the outliers' occurrence. They do not have an influence on the fitted model but may disturb the overall pattern of the data. Thus, for the further analysis we take the cut data (from the 812th observation).

5.1.2. Wind turbine dataset

The second considered dataset consists of vibration-based HI. The sensor has been installed on a high-speed bearing shaft of the 2.2 MW power wind turbine. The HI examined covers more than 50 days of data acquisition. In the end, the inner race bearing fault has appeared, which was confirmed by inspection. The bearing type used during the test is $32222 - J2$- SKF. It should be mentioned that this dataset has been used for prognosis by several authors, see e.g., [54–57].

In this paper, the considered HI is a feature proposed in the paper [57]. The values of inner race energy for a group of measurements for one particular speed (high-speed) are demonstrated in Fig. 14. Note that both subplots present the same data, the top subplot as a cloud of points, and the bottom subplot as a continuous function.

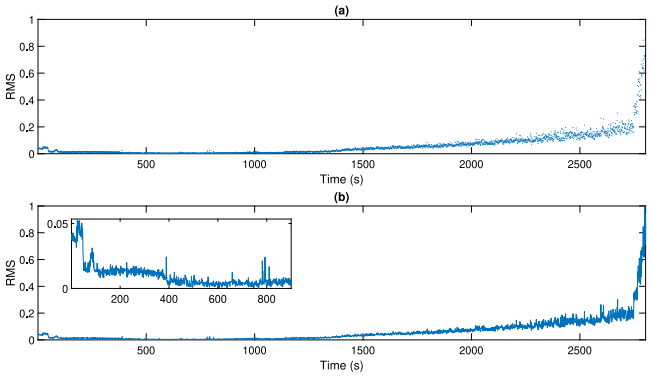

Fig. 13. Health Index (RMS) for FEMTO dataset. Top panel — cloud of points, bottom panel — data represented as continuous function. Box inside the bottom figure presents the zoom of the data for the first 900 observations to highlights the outliers occurrence.

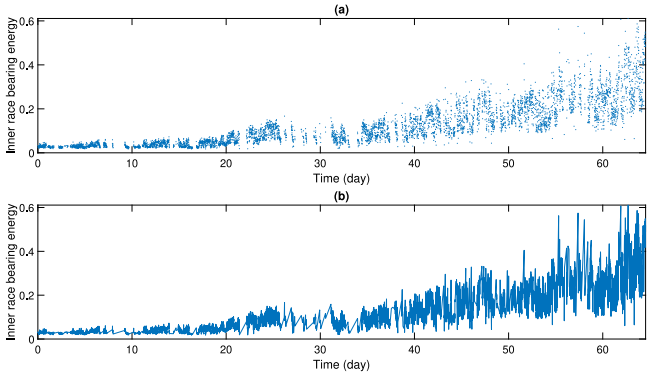

Fig. 14. Health Index presentation for Wind Turbine dataset. Top panel — cloud of points, bottom panel — data represented as continuous function.

5.2. Results for the real datasets

In this section, we present the results of the application of the proposed methodology for two real datasets (FEMTO data and Wind Turbine data). For both datasets, we use the robust approach as the universal methodology appropriate for Gaussian and non-Gaussian-distributed time series.

5.2.1. FEMTO dataset

The results for the FEMTO data are presented in Fig. 15. To identify the trend component (panel (a)), we use here the window of length 51 observations, while to identify the scale (panel (c)), we take the window of length 100. As can be seen, the estimated trend changes along time; i.e., for the first regime, it is almost constant; then it changes its nature and behaves as a linear function, while in the last regime it behaves as an exponential function. This corresponds to the trend component used in the simulated model described in Section 4.1. After removing the deterministic trend, we receive the random component (panel (b)). As can be seen, it corresponds to a non-homogeneous sequence with a non-constant scale. These partial results confirm our assumption and it is one of our findings that a random part scale increases over time. Thus, before further analysis, the data needs to be rescaled. The identified scale function is presented in panel (c). One can notice that the scale of a random part increases in a non-linear way. The last part of the curve describing scale variation confirmed significantly higher amplitudes of the noise. This could impact prognosis quality at the end of the lifetime curve. The random component after normalisation is presented in panel (d). This component corresponds to the sequence $\{R2(t)\}$. The empirical robust ACF presented in panel (e) clearly indicates the short-range dependence of the analysed data; however, we may assume that the coefficients of the model do not take significant values (the values of the empirical robust ACF are relatively small). As this dataset comes from a test rig and the experiment has been done during constant load/speed condition, through the ACF analysis one can conclude that the autodependence of the data is almost negligible. Next, we fit the optimal AR model. First, the optimal order p was identified (here $p = 10$) however, the coefficients of the model are relatively small ($\phi_1 = 0.1683, \phi_2 = 0.0385, \phi_3 = 0.0114, \phi_4 = 0.004, \phi_5 = 0.0093, \phi_6 = -0.0412, \phi_7 = -0.0378, \phi_8 = 0.0034, \phi_9 = -0.0551, \phi_{10} = 0.0408$). The residual series corresponding to the sequence $\{R3(t)\}$ is presented in panel (f), while its empirical robust ACF is demonstrated in panel (g). The plot of the empirical robust autocorrelation function indicates that the data can be considered as independent observations. Finally, by comparing the empirical and theoretical tails (panel (h)), we conclude that the residual series corresponds

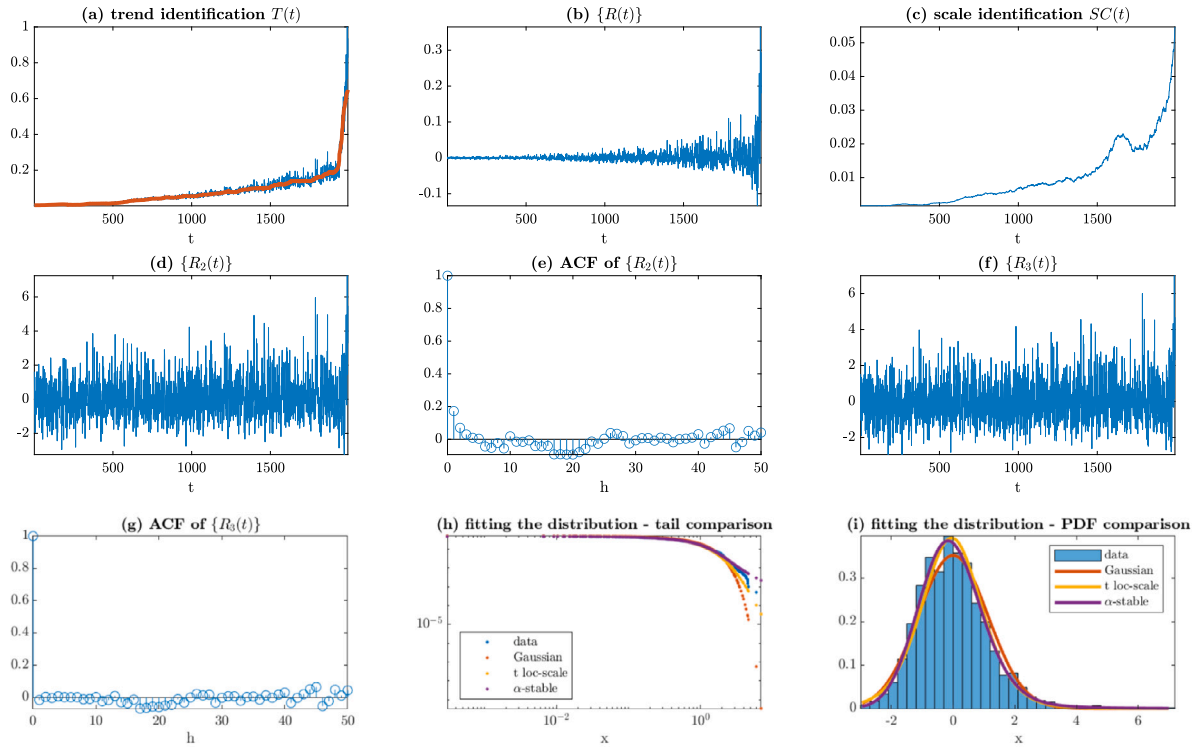

Fig. 15. FEMTO data modelling — the robust approach. Results of each step of the proposed framework: (a) identified trend $T(t)$, (b) random component $\{R(t)\}$, (c) identified scale $SC(t)$, (d) normalised random component $\{R_2(t)\}$, (e) autocorrelation of normalised random component, (f) residuals of AR model, (g) autocorrelation of the residual series, (h) comparison of fitted tails, (i) comparison of fitted distributions PDF.

Table 3
Identified characteristics for the FEMTO data set.

Trend	Time-varying (transition from linear to exponential function)
Scale	Time-varying
Autodependence of random component	Negligible
Coefficients of AR model	Negligible
Distribution of the random component	t location-scale with $\nu = 8.57$.

to t location-scale distribution with $\nu = 8.57$. The empirical tail is closer to the tail of the t location-scale distribution than to the α-stable one with the estimated α parameter. Obviously, the Gaussian distribution is rejected here, as the empirical tail and the tail of the Gaussian distribution do not coincide. The t location-scale distribution is also confirmed by comparing the normalised histogram of the residual signal and the probability density functions (PDF) of the theoretical distributions presented in panel (i). This result also is in line with our assumption and motivation of this work. The random part is not a Gaussian noise. However, the level of non-Gaussianity is relatively small.

In Table 3 we summarise the characteristics identified for the FEMTO dataset.

5.2.2. Wind turbine dataset

The results for the modelling of Wind Turbine data are presented in Fig. 16. In panel (a), we demonstrate the estimated trend component. It is clearly seen that it changes its nature over time, which corresponds to our assumptions. After removing the deterministic trend, we analyse the random part (see panel (b)). It exhibits non-homogeneous behaviour, which is related to the fact that its scale changes over the time. The identified component corresponding to the sequence $SC(t)$ is presented in panel (c). Definitely, we can see some seasonality in the random part; moreover, the model proposed by us is validated here. The scale is increasing in a different way for each regime (0 – 2000 and 2000 – 4000), after which the situation is even more complicated.

Data after normalisation are demonstrated in panel (d). As mentioned, we observe here a significant autodependence of the data. The corresponding empirical robust ACF is shown in panel (e). Here, in contrast to the FEMTO data, the empirical robust ACF takes higher values, and thus the fitted AR model has parameters significantly different from zero. The fitted model is AR(4) and the estimated parameters are: $\phi_1 = 0.54, \phi_2 = 0.1, \phi_3 = 0.036, \phi_4 = 0.016$. The residual series corresponding to the sequence $\{R3(t)\}$ is presented in panel (f) and the corresponding empirical ACF is demonstrated in panel (g). One can see that the series can be considered as sequences of independent observations. The inverse AR filter has significantly reduced the dependence of the data. Finally, by

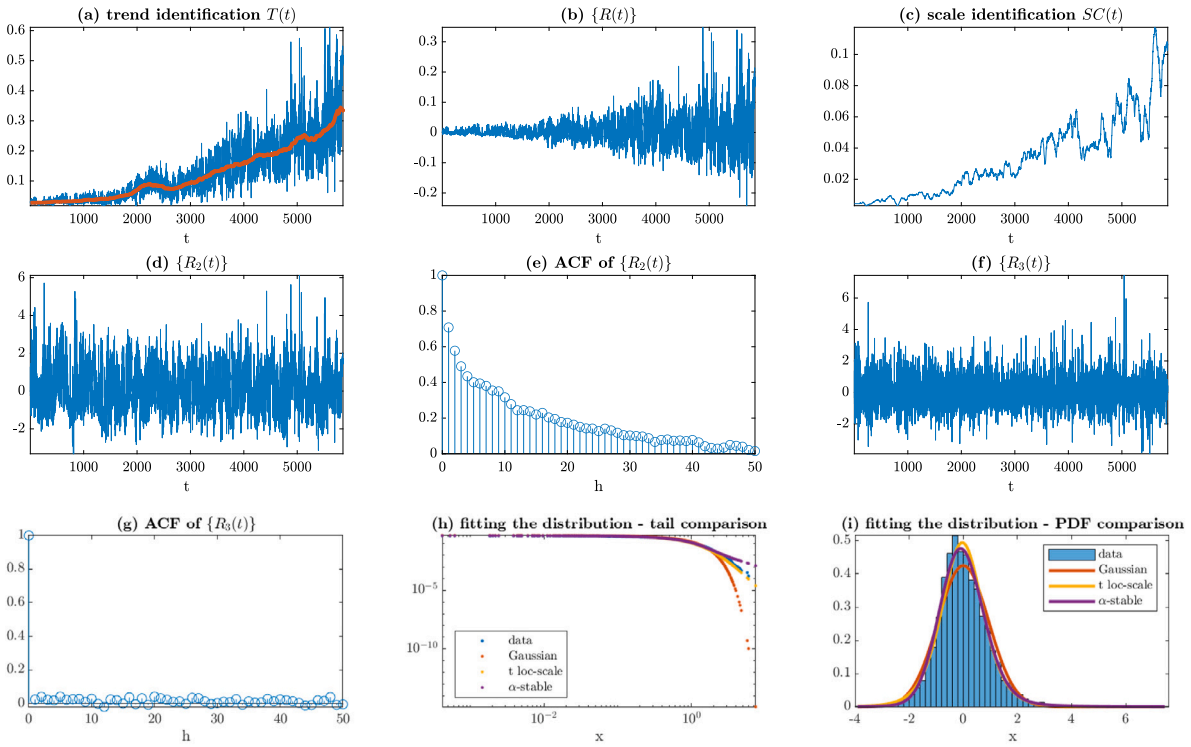

Fig. 16. Wind Turbine data modelling — the robust approach. Results of each step of the proposed framework: (a) identified trend $T(t)$, (b) random component $\{R(t)\}$, (c) identified scale $SC(t)$, (d) normalised random component $\{R_2(t)\}$, (e) autocorrelation of normalised random component, (f) residuals of AR model, (g) autocorrelation of the residual series, (h) comparison of fitted tails, (i) comparison of fitted distributions PDF.

Table 4
Identified characteristics for the Wind Turbine dataset.

Trend	Time-varying
Scale	Time-varying
Autodependence of random component	Significant
Coefficients of AR model	Significantly non-zero
Distribution of the random component	t location-scale with $\nu = 6.33$

comparing the empirical and the theoretical tails (corresponding to the three tested distributions with parameters estimated from the data) we can conclude the distribution of the residual series is closer to the t location-scale distribution (with $\nu = 6.33$) than to the α-stable one. The Gaussian distribution is unequivocally rejected. The selection of the t location-scale distribution as appropriate for the residual series is also confirmed by comparing the empirical and theoretical PDFs presented in panel (i). The identified characteristics of the appropriate components are summarised in Table 4.

6. Discussion

The presented results for both real datasets confirm the preliminary assumption about the corresponding models. For both datasets, we identified time-varying deterministic trends and scale functions. Moreover, the autodependence for both datasets is detected. However, as was indicated, the autodependence for the FEMTO data could be negligible, whereas for the Wind Turbine data we see the strong autodependence. The detected distribution of the random component for both cases has a non-Gaussian distribution. However, the distribution for the first dataset (i.e., FEMTO data) is closer to the Gaussian distribution (as the ν parameter is higher) than the Wind Turbine dataset, where the fitted t location-scale distribution has the smaller value of the parameter responsible for the impulsive (non-Gaussian) behaviour. In this paper, we did not consider separately models corresponding to three regimes discussed in Section 2. We take here the simplifying assumption that the random part for the analysed datasets corresponds to one stochastic model (here AR time series) and its distribution does not change over time.

To verify the results we analyse quantile lines constructed based on the fitted models for two real datasets. We remind that the presentation of quantile lines constructed based on the fitted model is one of the most common approaches for confirmation that given model is properly fitted to the data. This approach is widely used in both the literature and practical applications. The construction of the quantile lines is as follows: first we fit the model to real data, then we simulate number of trajectories from the

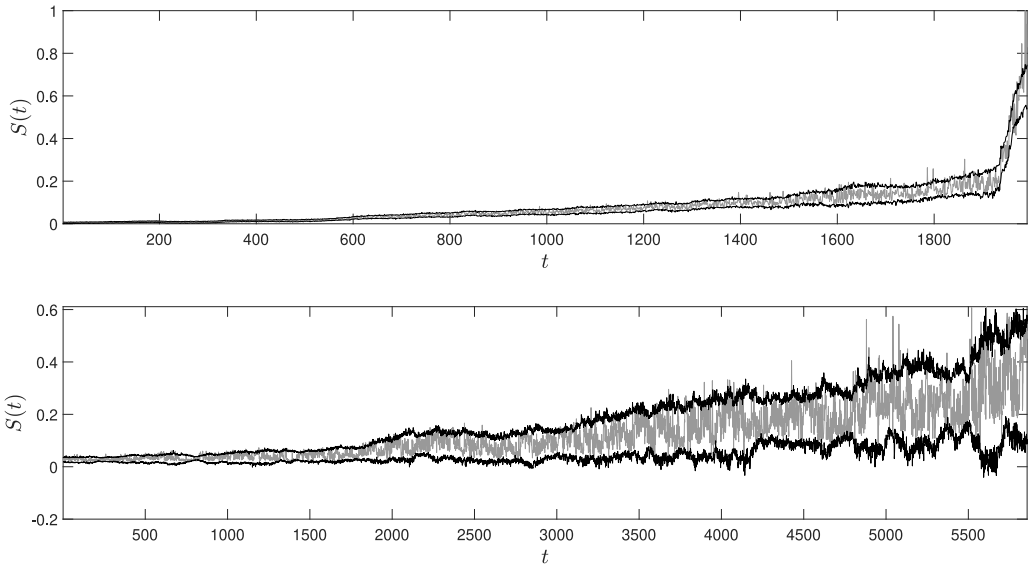

Fig. 17. Validation of the approach for real datasets (grey lines): top panel — FEMTO data, bottom panel — Wind Turbine data and the constructed quantile lines (black lines) on the level of 5% and 95% constructed on the basis of 100 simulated trajectories corresponding to the fitted non-Gaussian models.

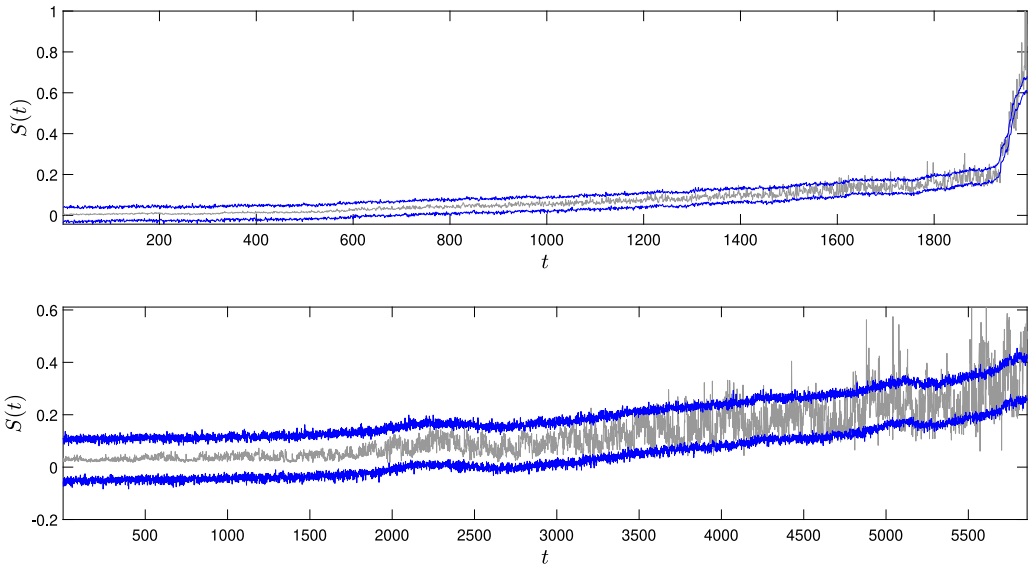

Fig. 18. Validation of the approach for real datasets (grey lines): top panel — FEMTO data, bottom panel — Wind Turbine data and the constructed quantile lines (blue lines) on the level of 5% and 95% constructed on the basis of 100 simulated trajectories corresponding to the fitted Gaussian models.

fitted model, and for each time point we calculate the quantiles (at appropriate levels). They are called quantile lines. If the real data fall into the constructed intervals (with appropriate probability), then we can confirm that the fitted model is a proper one. The quantile lines constructed based on the various fitted models can also be used to select the optimal model or to indicate which model from the considered ones is the best for the real data.

The models have been synthesised and simulation results are as demonstrated in Fig. 17. In this figure, by grey lines the real datasets are presented (top panel — FEMTO data, bottom panel — Wind Turbine data) and by black lines — the constructed quantile lines on the levels of 5% and 95% obtained on the basis of 100 simulated trajectories corresponding to the fitted models.

To confirm the validity of using the non-Gaussian model with non-homogeneous characteristics, in Fig. 18 we present the quantile lines constructed based on the simple Gaussian model that contains the time-dependent trend and the Gaussian noise. This model could be considered as the classical approach for such data modelling. Similarly as previously, the quantile lines are constructed using 100 trajectories of the fitted model and the quantile levels are 5% and 95%. The deterministic trend was fitted here by using the moving median method, and the parameters of the Gaussian distribution were estimated for the detrended data by using the

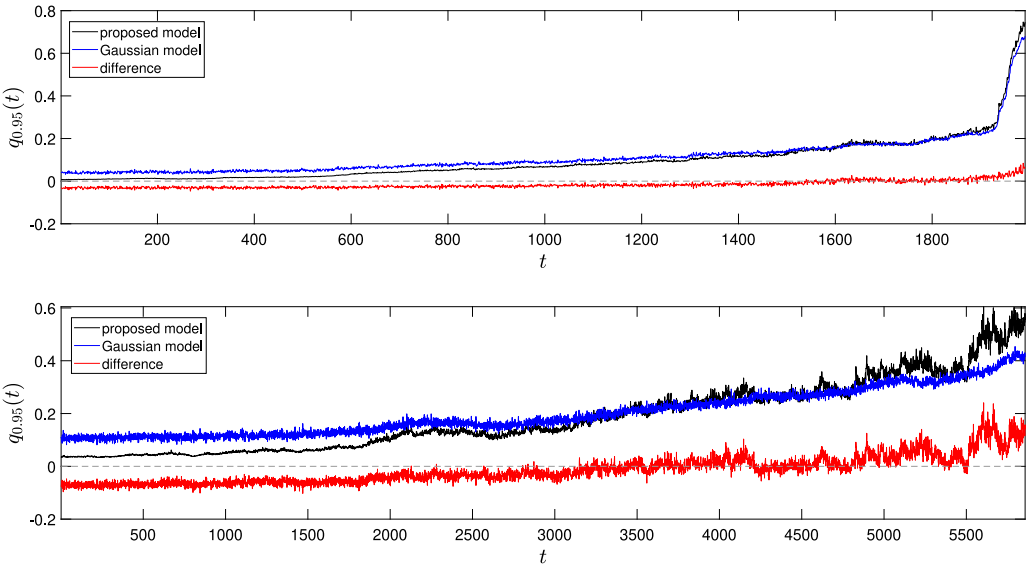

Fig. 19. The comparison of quantile lines on the level of 95% constructed for FEMTO data (top panel) and Wind Turbine data (bottom panel) for the proposed model (non-Gaussian non-homogeneous model) and Gaussian-based model (model containing Gaussian noise and the deterministic trend). By red lines we indicate the differences between the constructed quantile lines. In all cases the quantile lines were constructed by using 100 trajectories of the simulated models.

maximum likelihood method. Comparing Figs. 17 and 18 one can clearly see that the classical model is inappropriate for the real datasets. For the data from the first regime (healthy stage) the fitted Gaussian models seem to be overestimated while for the data from the last part (critical stage) the models are underestimated.

In consequence, the fitted models do not keep the specific nature of the datasets and cannot be considered as the optimal ones for instance for the prognosis. Considering the quantile lines for the proposed non-Gaussian non-homogeneous model, one can clearly see that the simulated and real data coincide for both datasets.

To highlight that the proposed model outperforms the Gaussian-based approach, in Fig. 19 we present the quantile lines on the level 95% for both datasets constructed using both approaches (marked by black and blue lines, respectively) as well as their difference. One can conclude that the quantiles for the first regime (healthy stage) take higher values for the Gaussian-based model in contrast to the non-Gaussian-based approach, while for the critical stage we have the opposite situation. The differences between quantiles for non-Gaussian and Gaussian-based models are significantly higher than zero which confirms justification for the application of the proposed model. Finally, we have calculated the RMSE between the quantiles for non-Gaussian and Gaussian-based models. For FEMTO data we have 0.0239 while for Wind Turbine data - 0.0540. The difference in RMSE values is related to the variation of scale function and non-Gaussian character of the residuals. In fact, the FEMTO data were identified as t location-scale distributed with higher number of degrees of freedom in comparison to Wind Turbine data. This indicates that FEMTO data is closer to Gaussian distribution than Wind Turbine one.

It should be highlighted that selection of window size (using trial and error method) applied in trend estimation and scale identification is important but not critical. Window size should be large enough to assure minimum bias for calculated statistics but from the other hand it should be small enough to provide precision in time-domain. Window size for trend and scale is different as variation for trend is slower than for scale. Window size does not influence directly on ACF and applied AR model estimation.

7. Conclusions

In this paper, we discuss the problem of modelling long-term data from a condition monitoring system. The main goal was to introduce new components of the classical model corresponding to such data. Based on real examples, we showed the presence of specific features in two real datasets that may be important during modelling and should not be neglected. We proposed to generalise the classical model (that assumes non-linear trend and Gaussian distribution), assuming non-Gaussian distribution of the noise. Moreover, we assumed that the characteristics of the random part may be time-varying, like increasing variance (in the case of the Gaussian distribution) or scale (in the case of the non-Gaussian distribution). The last novelty was the modelling of random component by autoregressive time series (AR model describing HI fluctuation due to operational or environmental factors).

Next, we proposed dedicated algorithms for the identification of the above-mentioned components. Finally, we introduced a framework to identify each component and synthesise all of them to simulate the exemplary trajectory of the proposed model.

The presented research brings new knowledge about the long-term data models used in condition monitoring. The new model and the proposed novel methodology for such specific data analysis may help to construct more accurate prognosis. Taking into account the non-Gaussian character of the noise focused our attention on robust techniques for trend estimation, AR model identification,

and finally estimation of the distribution of the random part. The proposed robust methods are not sensitive to outliers present in long-term data, and thus they allow to obtain more accurate estimation of each component. We would like to highlight that time-varying variance (or scale) in a random part is also very important in the context of data modelling. The variance/scale takes the highest values in the last segment (exponential growth of diagnostic feature). It is obvious that higher variation provides more complicated identification of the trend and in consequence less precision in RUL estimation.

The proposed procedure has been applied to two real data (FEMTO and Wind Turbine datasets). These raw data and partial results confirm the preliminary assumption about non-Gaussian, non-homogeneous behaviour and autodependence in the time series.

We believe that there is room for enhancement of the proposed approach. We plan to continue this research for more complex data related to various types of HI from monitoring systems. To simplify the problem in this paper, we assumed that the random part of the data will behave in the same way in the whole data (i.e. model will be the same, some parameters may change). There is a hypothesis, that again comes from observation of various real datasets that in the first part of the input data (Regime 1, i.e. good condition), the AR component may be neglected, but in Regime 3 it may be significantly stronger than for Regime 2. The "stronger" AR component may mean higher amplitudes, but also higher coefficients of the AR model and higher order of the model (higher complexity). This issue needs to be studied deeper and validated wider for various data sets.

Building a precise model for the prognosis of the data is a critical issue. The assumption of three regimes corresponding to the degradation process means that we need to develop three models, separately for each phase of machine lifetime. Thus, at the pre-processing step, we need to segment the data. In case of non-Gaussian data with time-varying characteristics, the segmentation seems to be another class of signal processing problems, and for sure it requires dedicated segmentation methods (the preliminary analysis indicates that the classical algorithms for the structure break point detection are ineffective for automatic segmentation of our data). Thus, the future study will be related also to this part.

In this paper, we highlight that knowledge about the data and a precise model is crucial to the prognosis. The proposed method brings new knowledge to the community, and the developed framework allows the building of an advanced model that can generate (simulate) data with very high similarity to real trajectory. Thanks to simulation, we can use new (replicated) data to train machine learning-based prognostic systems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

H. Shiri gratefully acknowledges the European Commission for its support of the Marie Sklodowska Curie program through the ETN MOIRA project (GA 955681).

The work of W. Żuławiński, K. Maraj-Zygmąt, R. Zimroz and A. Wyłomańska is supported by project No. POIR.01.01.01-00-0350/21 entitled "A universal diagnostic and prognostic module for condition monitoring systems of complex mechanical structures operating in the presence of non-Gaussian disturbances and variable operating conditions" co-financed by the European Union from the European Regional Development Fund under the Intelligent Development Program. The project is carried out as part of the competition of the National Center for Research and Development no: 1/1.1.1/2021 (Szybka Ścieżka).

Appendix A. Considered distributions

- The Gaussian distribution $\mathcal{N}(\mu,\sigma)$ is defined through the PDF [32,58]

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},\quad x\in\mathbb{R}, \tag{A.1}$$

  where $\mu\in\mathbb{R}$ is the mean and $\sigma>0$ is the standard deviation.
- The α-stable distribution $S(\alpha,\beta,\gamma,\delta)$ is defined through the characteristic function [59,60]

$$\boldsymbol{\Phi}(t)=\begin{cases}\exp\left(-\gamma^{\alpha}|t|^{\alpha}\left[1+i\beta\,\mathrm{sign}(t)\tan\frac{\pi\alpha}{2}((\gamma|t|)^{1-\alpha}-1)\right]+i\delta t\right) & \text{for } \alpha\neq 1,\\ \exp\left(-\gamma|t|\left[1+i\beta\,\mathrm{sign}(t)\frac{2}{\pi}\ln(\gamma|t|)\right]+i\delta t\right) & \text{for } \alpha=1,\end{cases} \tag{A.2}$$

  where $0<\alpha\leq 2$ — stability parameter, $-1\leq\beta\leq 1$ — skewness parameter, $\gamma>0$ — scale parameter and $\delta\in\mathbb{R}$ — location parameter. Notice that Gaussian distribution is a special case of the α-stable distribution. The α-stable distribution with $\alpha=2$ corresponds to the normal distribution. It is the only case when the variance is finite.

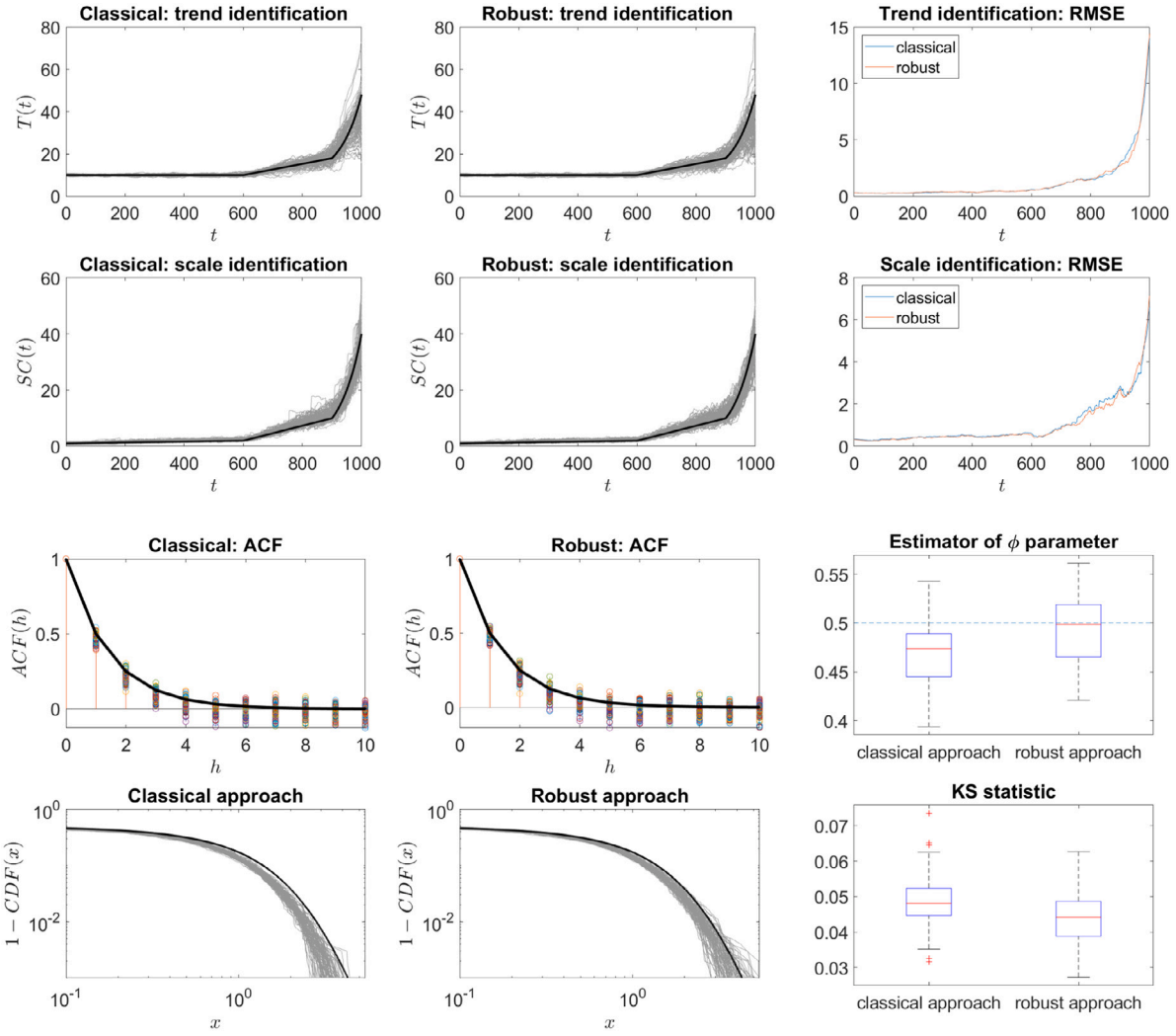

Fig. B.20. Results of the proposed procedures for simulated data (trend, scale, AR model and distribution tail identification in rows 1–4, respectively). Left column: the classical approach (dedicated for Gaussian distributed signals), middle column: robust approach (dedicated for non-Gaussian heavy-tailed distributed signals), right column: evaluation of the results by RMSE (trend and scale) and boxplots (random components). The t location-scale distributed random component $\{R3(t)\}$ with $\nu = 9$.

- The t location-scale distribution $\mathbf{t}(\mu, \sigma, \nu)$ is defined through the PDF [61,62]

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{\nu + \left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right]^{\left(\frac{\nu+1}{2}\right)}, \tag{A.3}$$

where $\Gamma(\cdot)$ is gamma function, $\mu \in \mathbb{R}$ — location parameter, $\sigma > 0$ — scale parameter and $\nu > 0$ — shape parameter. The variance is only defined for values of $\nu > 2$. If a given random variable $X \sim \mathbf{t}(\mu, \sigma, \nu)$, then $(X - \mu)/\sigma$ has a Student's t distribution with ν degrees of freedom. When ν tends to infinity, then Student's t distribution tends to a Gaussian distribution.

Appendix B. Simulation study for data with t location-scale distribution of the random component

See Figs. B.20 and B.21.

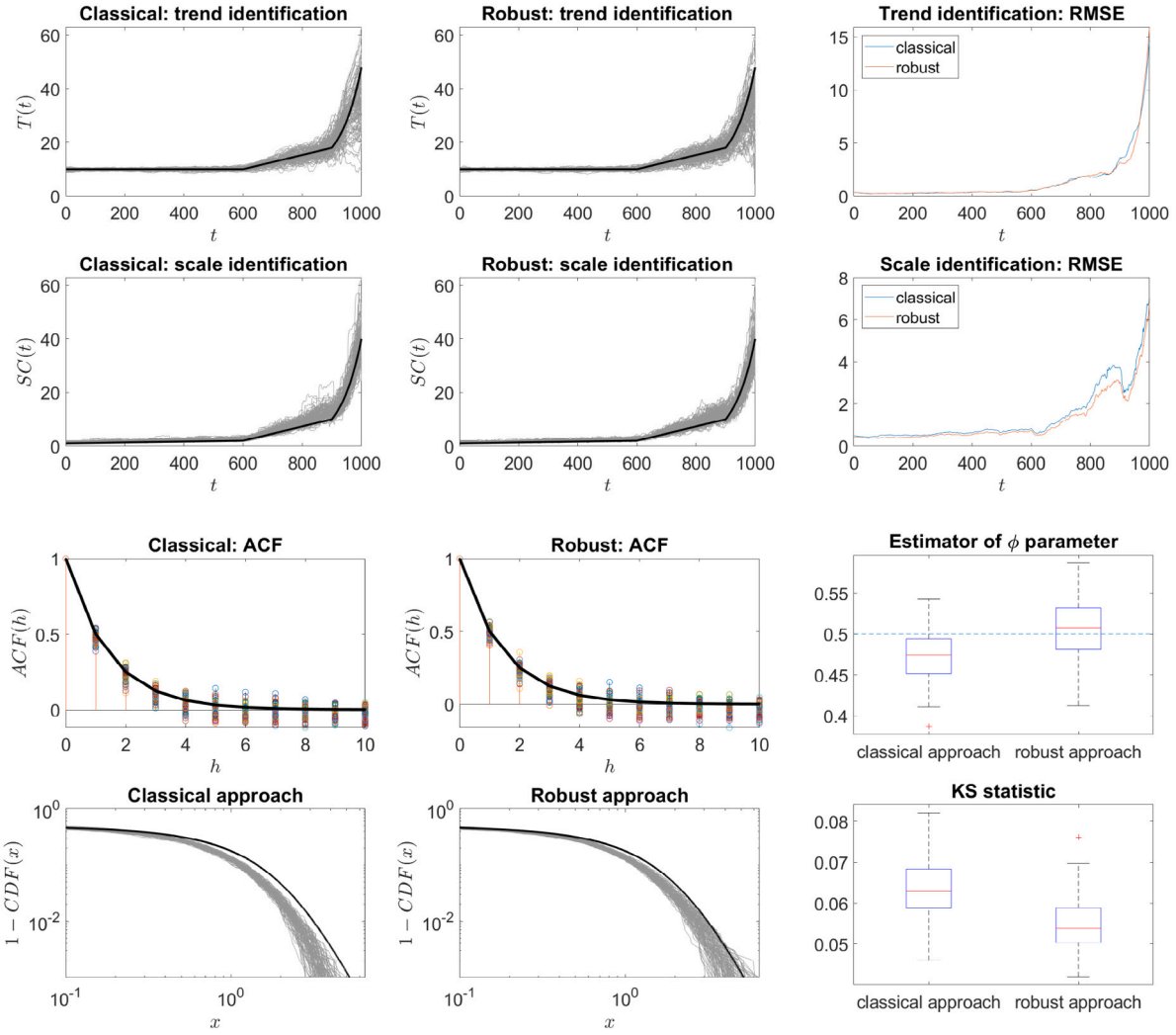

Fig. B.21. Results of the proposed procedures for simulated data (trend, scale, AR model and distribution tail identification in rows 1–4, respectively). Left column: the classical approach (dedicated for Gaussian distributed signals), middle column: robust approach (dedicated for non-Gaussian heavy-tailed distributed signals), right column: evaluation of the results by RMSE (trend and scale) and boxplots (random components). The t location-scale distributed random component $\{R3(t)\}$ with $\nu = 6$.

References

[1] M. Kan, A. Tan, J. Mathew, A review on prognostic techniques for non-stationary and non-linear rotating systems, Mech. Syst. Signal Process. 62 (2015) 1–20.

[2] J. Lee, F. Wu, Z.W. Hao, M. Ghaffari, L. Liao, D. Siegel, Prognostics and health management design for rotary machinery systems - reviews, methodology and applications, Mech. Syst. Signal Process. 42 (1–2) (2014) 314–334.

[3] A. Heng, S. Zhang, A. Tan, J. Mathew, Rotating machinery prognostics: State of the art, challenges and opportunities, Mech. Syst. Signal Process. 23 (3) (2009) 724–739.

[4] M. Zhao, B. Tang, Q. Tan, Bearing remaining useful life estimation based on time-frequency representation and supervised dimensionality reduction, Measurement: J. Int. Meas. Confederation 86 (2016) 41–55.

[5] J. Sikorska, M. Hodkiewicz, L. Ma, Prognostic modelling options for remaining useful life estimation by industry, Mech. Syst. Signal Process. 25 (5) (2011) 1803–1836.

[6] E. Ramasso, A. Saxena, Performance benchmarking and analysis of prognostic methods for cmapss datasets, Int. J. Prognostics Health Manage. 5 (2) (2014).

[7] S. Hong, Z. Zhou, E. Zio, K. Hong, Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method, Digital Signal Process. Rev. J. 27 (1) (2014) 159–166.

[8] B. Yan, X. Ma, G. Huang, Y. Zhao, Two-stage physics-based Wiener process models for online RUL prediction in field vibration data, Mech. Syst. Signal Process. 152 (2021).

[9] Y. Wen, J. Wu, D. Das, T.-L. Tseng, Degradation modeling and RUL prediction using Wiener process subject to multiple change points and unit heterogeneity, Reliab. Eng. Syst. Saf. 176 (2018) 113–124.

[10] Z. Huang, Z. Xu, X. Ke, W. Wang, Y. Sun, Remaining useful life prediction for an adaptive skew-Wiener process model, Mech. Syst. Signal Process. 87 (2017) 294–306.
[11] X. Xi, M. Chen, H. Zhang, D. Zhou, An improved non-Markovian degradation model with long-term dependency and item-to-item uncertainty, Mech. Syst. Signal Process. 105 (2018) 467–480.
[12] H. Zhang, D. Zhou, M. Chen, J. Shang, FBM-based remaining useful life prediction for degradation processes with long-range dependence and multiple modes, IEEE Trans. Reliab. 68 (3) (2019) 1021–1033.
[13] M. Ling, H. Ng, K. Tsui, Bayesian and likelihood inferences on remaining useful life in two-phase degradation models under gamma process, Reliab. Eng. Syst. Saf. 184 (2019) 77–85.
[14] H. Liu, W. Song, Y. Niu, E. Zio, A generalized Cauchy method for remaining useful life prediction of wind turbine gearboxes, Mech. Syst. Signal Process. 153 (2021) 107471.
[15] H. Liu, W. Song, Y. Zhang, A. Kudreyko, Generalized Cauchy degradation model with long-range dependence and maximum Lyapunov exponent for remaining useful life, IEEE Trans. Instrum. Meas. 70 (2021) 1–12.
[16] W. Song, H. Liu, E. Zio, Long-range dependence and heavy tail characteristics for remaining useful life prediction in rolling bearing degradation, Appl. Math. Model. 102 (2022) 268–284.
[17] S. Duan, W. Song, E. Zio, C. Cattani, M. Li, Product technical life prediction based on multi-modes and fractional Lévy stable motion, Mech. Syst. Signal Process. 161 (2021) 107974.
[18] C. Lim, D. Mba, Switching Kalman filter for failure prognostic, Mech. Syst. Signal Process. 52–53 (1) (2015) 426–435.
[19] M. Kordestani, M.F. Samadi, M. Saif, K. Khorasani, A new fault prognosis of MFS system using integrated extended Kalman filter and Bayesian method, IEEE Trans. Ind. Inf. (2018).
[20] R. Singleton, E. Strangas, S. Aviyente, Extended Kalman filtering for remaining-useful-life estimation of bearings, IEEE Trans. Ind. Electron. 62 (3) (2015) 1781–1790.
[21] L. Reuben, D. Mba, Diagnostics and prognostics using switching Kalman filters, Struct. Health Monit. 13 (3) (2014) 296–306.
[22] P. Cheng, H. Wang, V. Stojanovic, S. He, K. Shi, X. Luan, F. Liu, C. Sun, Asynchronous fault detection observer for 2-D Markov jump systems, IEEE Trans. Cybern. (2021) 1–12.
[23] P. Cheng, M. Chen, V. Stojanovic, S. He, Asynchronous fault detection filtering for piecewise homogenous Markov jump linear systems via a dual hidden Markov model, Mech. Syst. Signal Process. 151 (2021) 107353.
[24] J. Sun, H. Zuo, W. Wang, M. Pecht, Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance, Mech. Syst. Signal Process. 28 (2012) 585–596.
[25] N. Nedić, D. Pršić, C. Fragassa, V. Stojanović, A. Pavlovic, Simulation of hydraulic check valve for forestry equipment, Int. J. Heavy Vehicle Syst. 24 (3) (2017) 260–276.
[26] X.-S. Si, W. Wang, C.-H. Hu, D.-H. Zhou, Remaining useful life estimation–a review on the statistical data driven approaches, European J. Oper. Res. 213 (1) (2011) 1–14.
[27] C. Okoh, R. Roy, J. Mehnen, L. Redding, Overview of remaining useful life prediction techniques in through-life engineering services, Procedia Cirp 16 (2014) 158–163.
[28] Z. Zhang, X. Si, C. Hu, Y. Lei, Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods, European J. Oper. Res. 271 (3) (2018) 775–796.
[29] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: A systematic review from data acquisition to RUL prediction, Mech. Syst. Signal Process. 104 (2018) 799–834.
[30] L. Cui, X. Wang, H. Wang, J. Ma, Research on remaining useful life prediction of rolling element bearings based on time-varying Kalman filter, IEEE Trans. Instrum. Meas. 69 (6) (2019) 2858–2867.
[31] P.J. Brockwell, R.A. Davis, Introduction to Time Series and Forecasting, Springer, 2016.
[32] J.F. Lawless, Statistical models and methods for lifetime data, Canad. J. Statist. 10 (4) (1982) 316–317.
[33] P.J. Rousseeuw, C. Croux, Alternatives to the median absolute deviation, J. Amer. Statist. Assoc. 88 (424) (1993) 1273–1283.
[34] Y. Ma, M.G. Genton, Highly robust estimation of the autocovariance function, J. Time Series Anal. 21 (6) (2000) 663–684.
[35] C. Croux, P.J. Rousseeuw, Time-efficient algorithms for two highly robust estimators of scale, in: Y. Dodge, J. Whittaker (Eds.), Computational Statistics, Physica-Verlag HD, Heidelberg, 1992, pp. 411–428.
[36] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, C. Varnier, PRONOSTIA: An experimental platform for bearings accelerated degradation tests., in: IEEE International Conference on Prognostics and Health Management, PHM'12, IEEE Catalog Number: CPF12PHM-CDR, 2012, pp. 1–8.
[37] A. Mosallam, K. Medjaher, N. Zerhouni, Time series trending for condition assessment and prognostics, J. Manufact. Technol. Manage. (2014).
[38] 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 (4) (2013) 821–832.
[39] K. Javed, R. Gouriveau, N. Zerhouni, P. Nectoux, Enabling health monitoring approach based on vibration data for accurate prognostics, IEEE Trans. Ind. Electron. 62 (1) (2014) 647–656.
[40] B. Zhang, L. Zhang, J. Xu, Degradation feature selection for remaining useful life prediction of rolling element bearings, Qual. Reliab. Eng. Int. 32 (2) (2016) 547–554.
[41] S. Hong, Z. Zhou, E. Zio, W. Wang, An adaptive method for health trend prediction of rotating bearings, Digit. Signal Process. 35 (2014) 117–123.
[42] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, J. Dybala, A model-based method for remaining useful life prediction of machinery, IEEE Trans. Reliab. 65 (3) (2016) 1314–1326.
[43] Y. Nie, J. Wan, Estimation of remaining useful life of bearings using sparse representation method, in: 2015 Prognostics and System Health Management Conference (PHM), IEEE, 2015, pp. 1–6.
[44] Z. Liu, M.J. Zuo, Y. Qin, Remaining useful life prediction of rolling element bearings based on health state assessment, Proc. Inst. Mech. Eng. Part C 230 (2) (2016) 314–330.
[45] J.K. Kimotho, C. Sondermann-Wölke, T. Meyer, W. Sextro, Machinery prognostic method based on multi-class support vector machines and hybrid differential evolution–particle swarm optimization, Chem. Eng. Trans. 33 (2013).
[46] D. Zurita, J.A. Carino, M. Delgado, J.A. Ortega, Distributed neuro-fuzzy feature forecasting approach for condition monitoring, in: Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA), IEEE, 2014, pp. 1–8.
[47] L. Guo, H. Gao, H. Huang, X. He, S. Li, Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring, Shock Vib. 2016 (2016).
[48] X. Jin, Y. Sun, Z. Que, Y. Wang, T.W. Chow, Anomaly detection and fault prognosis for bearings, IEEE Trans. Instrum. Meas. 65 (9) (2016) 2046–2054.
[49] H. Li, Y. Wang, Rolling bearing reliability estimation based on logistic regression model, in: 2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE), IEEE, 2013, pp. 1730–1733.
[50] Y. Wang, Y. Peng, Y. Zi, X. Jin, K.-L. Tsui, A two-stage data-driven-based prognostic approach for bearing degradation problem, IEEE Trans. Ind. Inf. 12 (3) (2016) 924–932.

[51] Y. Pan, M.J. Er, X. Li, H. Yu, R. Gouriveau, Machine health condition prediction via online dynamic fuzzy neural networks, Eng. Appl. Artif. Intell. 35 (2014) 105–113.
[52] L. Wang, L. Zhang, X.-z. Wang, Reliability estimation and remaining useful lifetime prediction for bearing based on proportional hazard model, J. Central South Univ. 22 (12) (2015) 4625–4633.
[53] L. Xiao, X. Chen, X. Zhang, M. Liu, A novel approach for bearing remaining useful life estimation under neither failure nor suspension histories condition, J. Intell. Manuf. 28 (8) (2017) 1893–1914.
[54] L. Saidi, J.B. Ali, E. Bechhoefer, M. Benbouzid, Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR, Appl. Acoust. 120 (2017) 1–8.
[55] L. Saidi, E. Bechhoefer, J.B. Ali, M. Benbouzid, Wind turbine high-speed shaft bearing degradation analysis for run-to-failure testing using spectral kurtosis, in: 2015 16th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), IEEE, 2015, pp. 267–272.
[56] J.B. Ali, L. Saidi, S. Harrath, E. Bechhoefer, M. Benbouzid, Online automatic diagnosis of wind turbine bearings progressive degradations under real experimental conditions based on unsupervised machine learning, Appl. Acoust. 132 (2018) 167–181.
[57] E. Bechhoefer, R. Schlanbusch, Generalized prognostics algorithm using Kalman smoother, IFAC-PapersOnLine 48 (21) (2015) 97–104.
[58] G. Trenkler, Statistical distributions, Comput. Statist. Data Anal. 19 (4) (1995) 483–484, M. Evans and N. Hastings and B. Peacock, (1993): (2nd Edition).
[59] J.P. Nolan, Numerical calculation of stable densities and distribution functions, Commun. Stat. Stoch. Models 13 (4) (1997) 759–774.
[60] A. Weron, R. Weron, Computer Simulation of Lévy α-Stable Variables and Processes, Springer Berlin Heidelberg, 1995, pp. 379–392.
[61] B.L. Welch, 'Student' and small sample theory, J. Amer. Statist. Assoc. 53 (284) (1958) 777–788.
[62] Student, The probable error of a mean, Biometrika 6 (1) (1908) 1–25.