International Journal of Automation and Computing

17(6), December 2020, 788-811
DOI: 10.1007/s11633-020-1243-2

Study on Statistical Outlier Detection and Labelling

Paweł D. Domański

Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw 00-665, Poland

Abstract: Outliers accompany control engineers in their real life activity. Industrial reality is much richer than elementary linear, quadratic, Gaussian assumptions. Outliers appear due to various and varying, often unknown, reasons. They meet research interest in statistical and regression analysis and in data mining. There are a lot of interesting algorithms and approaches to outlier detection, labelling, filtering and finally interpretation. Unfortunately, their impact on control systems has not been found sufficient attention in research. Their influence is frequently unnoticed, ignored or not mentioned. This work focuses on the subject of outlier detection and labelling in the context of control system performance analysis. Selected statistical data-driven approaches are analyzed, as they can be easily implemented with limited a priori knowledge. The study consists of a simulation study followed by the analysis of real control data. Different generation mechanisms are simulated, like overlapping Gaussian processes, symmetric and asymmetric, artificially shifted points and fat-tailed distributions. Simulation observations are confronted with industrial control loops datasets. The work concludes with a practical procedure, which should help practitioners in dealing with outliers in control engineering temporal data.

Keywords: Outlier detection, control loop quality, statistical analysis, robust estimation, heavy tails.

1 Introduction

An outlier is a strange phenomenon. Varying perspectives may give different interpretations. Simple definitions proposed by Dixon[1] define outliers as values, dubious in the eyes of the researcher or by Weiner[2] as contaminants. One of the most popular definitions has been formulated by Hawkins[3] naming, an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism to be an outlier. Johnson and Wichern[4] define an outlier, as an observation in a data set which appears to be inconsistent with the remainder of that set of data. Barnett and Lewis[5] say that, an outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs. As one can see there are various other names for the outliers, for instance anomalies, contaminants or fringeliers reflecting, unusual events which occur more often than seldom[2].

These strange phenomena may have disastrous effects on further data analysis, whatever it will be[6]. They may increase signal variance and reduce the power of statistical tests performed during analysis[7]. They destroy signal normality and introduce fat tails[8]. Finally, Rousseeuw and Leroy[9] point out that they significantly bias regression analysis.

Following presented definitions, we may try to investigate their origins[7, 10]. Generally, outliers may originate from a supposition about incorrect observations or from the inherent complex and non-linear variability of the data. Aberrant data can be caused by human errors or by intentional or motivated mis-reporting, by the erroneous operation of computer systems being in chain of the data measurement and collection process, by sampling errors or from standardization failures. Identification and correction of such incorrect values is not straightforward. Cautiousness, double checking, redundancy or recalculation may help. If incorrect observations cannot be reasonably corrected, they need to be eliminated, as they do not relate to valid data.

Complex, non-linear and often unknown process nature may lead to simplifications and mis-interpretations, like for instance incorrect assumptions about the data distribution leading to the presence of possible outliers[11]. Such processes can cause multi-modal, skewed, asymptotic, fat-tailed, flat or very strangely shaped distributions, which can depend on data sampling. As the process is complex, underlying data may have a different structure than originally assumed inherently characterized by the tails[12, 13] or there might just be more than one mechanism. Data may be affected by long term trends, cross-correlations with varying delays, self-similarity or multifractality[14]. On the other hand, an outlier might be rare, but a natural implication of the process itself. The above reasons cause problems and mislead engineers, who are accustomed to linear, quadratic and Gaussian simplifications.

We observe various responses to the outliers. From one side, we may continue the research pretending that everything is according to the assumed normal conditions, as nothing would happen. Actually, this is a dead end road. If we do not see the outlier, it does not mean that it

Research Article
Manuscript received April 3, 2020; accepted June 29, 2020; published online October 21, 2020
Recommended by Associate Editor Dong-Hua Zhou

does not exist. Once we are aware of outliers, we intentionally acknowledge their existence and their impact.

Finally, the outliers might be considered as a potential focus of inquiry. Data contamination might be considered as a source of an important information focusing on their analysis. Whatever the approach is, we need to detect them. Thereby, we may label them for further analysis. Such an analysis is a target during fraud detection, medical diagnosis, leakage detection, cyber security, etc[15, 16]. Current work focuses on specific applications to the analysis of control systems and to the detection of abnormal control loop performance.

Detected outliers might be removed or not. There are different policies in such a case. There are strong arguments for their removal. On the other hand, if they are suspected to be legitimate, they are representative for the population as a whole and should not be removed. Finally, they have to be isolated as they represent a wanted feature or an incident. Therefore, the analysis of outliers consists of several activities, such as detection, labelling, interpretation or removal.

The story of outlier detection starts with simple visual inspection methodology. The process dataset is visually inspected and any outlier is identified using the viewer′s expert knowledge and then manually removed. Time trends analysis has been soon supported by simple statistical review through the histogram plots. Such a simplified statistics allowed additional domains for outlier detection. This manual approach has been soon enhanced with structured research and scientific investigation. It has been quite natural that the initial research has been oriented towards statistical approaches[17–20]. These analyses have focused on the Gaussian approaches exploiting various properties of normal distributions. The literature on statistical outlier detection is very rich[3, 9, 21, 22]. Formally, statistical analysis should follow three steps[11]: 1) labelling (flagging for further investigation), 2) accommodation through robust statistical methods that are not unduly affected and 3) outlier identification, which tests if the observation is an outlier.

There have been proposed numerous statistical methods for outlier detection. Actually, the consideration that an observation is an outlier depends on the underlying distribution of the data. Most of the research is limited to univariate datasets assumed to follow an approximately normal probability density function. Control engineering and loop analysis perspective fits into the univariate assumption. Normality assumptions result in many algorithms, as for instance Z-scores and modified Z-scores[11], interquartile range (IQR)[23], Grubbs′ test[20], Tietjen-Moore test[24], minimum covariance determinant[25], extreme studentized deviate (ESD) test[26], and Thompson Tau test consisting of two steps[27]. Robust regression[9, 28] brought forward the fact that classical mean square method is sensitive even to a single outlier.

This observation led to further investigations utilizing new robust location and scale estimators, like Z-scores using median and MAD (median of the absolute deviations about the median[11]), Hampel filter[29], power law tail index estimates[30]. Finally, α-stable distribution can be considered as an underlying generation mechanism. This function exhibits a lot of attractive properties[31]. Research shows that it may be frequently validated as the real statistical process behind signal generation, also in control engineering[32]. It includes not only scale and location but also stability factor responsible for tails and skewness coefficient.

Recently, research started to exploit developments in data mining. There are dozens of approaches[16,33–37]. One may distinguish three different types of algorithms. Supervised methods utilize for training historical data about normal and abnormal objects. Semi-supervised approaches utilize for learning only normal or abnormal examples. No training data is utilized in the unsupervised methods. The next distinction takes into consideration method reference resolution, i.e., the difference between global versus local range. In fact, some approaches lie between. The next division uses method output. Labelling gives binary value, naming the objects either normal or abnormal. Scoring results in continuous output, like the probability of being an outlier.

Finally, the method can use direct information hidden in data (data-driven distance, density or angle-based approaches) for detection, or there exists an intermediate stage of modelling. Thus, labelling is performed according to the derived model (rational or sample model-based approaches).

The main goal of the presented research is to exploit an opportunity to use statistical outlier detection and labelling in an engineering task of control performance assessment (CPA). Furthermore, three novel modifications in already existing methods are proposed. The well-known MDist algorithm is improved with robust M-estimator using logistic ψ-function. The second MDist modification uses α-stable tail crossover estimation. The third proposal modifies the classical interquartile range algorithm with tail index α modelling. The proposed outlier detection task is validated on real industrial data, while properties of proposed algorithms are compared with classical ones.

The presented research starts with selection and description of the applied statistical methods (Section 2). Analysis included in section 3 consists of simulations and the discussion of obtained results. It is followed by industrial validation in Section 4. Section 5 summarizes the paper and addresses open issues.

2 Statistical outlier detection

The issue of outliers and their statistical detection can be traced even to the 19th century[17]. The research followed through the following decades. Thereby, the subject literature is quite extensive, but its main popularity has disappeared. Currently, data-mining approaches have

gained the largest publicity[38]. However, it should still be remembered that statistical approaches share formal simplicity and rigorousness. Additionally, recent findings in the area of non-Gaussian and robust statistics[39, 40] have brought a new impact and improved methods′ reliability. Comprehensive reviews of the statistical outlier analysis can be found in several studies[15, 41–43].

Statistical methods mostly depend on properties of the assumed probabilistic model of the underlying stochastic process. As there are so many methods, one has to select those which are the most appropriate in the considered situation. This paper focuses on control engineering applications, which may be further used in the CPA task or during the controller tuning procedure. Research shows that the majority of such time series data exhibit fat-tailed properties[32, 44], which can be efficiently modelled with stable distributions. Simultaneously, Gaussian models are still existent, although in the minority of observed situations. Thus, the selected methods should be appropriate in such cases. There is one more issue that should be taken into account: asymmetry. In several situations, the signals demonstrate asymmetric properties[45]. Consequently, the selection includes methods, which are able to address this issue. Concluding, we have to keep in mind an idea[46] that, the notion of outliers has to be considered not by itself but in connection with underlying scheme. The following six methods have been chosen for the analysis:

M.1 MDist-G: Elementary Z-score method assuming normality

M.2 MDist-rHub: A method utilizing robust location and scale M-estimators

M.3 MDist-α: α-stable tail crossover method

M.4 ESD: Generalized extreme studentized deviate

M.5 IQR: Interquartile range method

M.6 IQR-α: α-stable tail index modelling method.

The above methods are used for outlier detection. Three of them are well known from the literature (MDist-G, ESD and IQR), while three others (MDist-rHub, MDist-α and IQR-α) are new proposals. All addressed methods are compared with each other against for different types of outliers and finally validated using real industrial datasets.

These methods allow initial data preprocessing before control performance assessment, cleaning process data before the main CPA task. Furthermore, labelled outliers may be further analyzed, as they potentially bring into the picture additional knowledge about the process itself.

2.1 MDist-G

Z-scores approaches found in [11] (sometimes denoted as MDist[15]) seem to be the earliest proposed methods. They use the well-known normal distribution property that if X is distributed using $N(\bar{x}, \sigma^2)$, then $Z = (X - \bar{x})/\sigma$ is distributed with $N(0, 1)$. Thus, we can take the Z-scores of the observations $x_1, x_2, \cdots, X_n$,

$$z_i = \frac{x_i - \bar{x}}{\sigma} \tag{1}$$

as an algorithm to label outliers. The common rule uses a Z-score value of $D_G = 3$ and labels observations that exceed respective borders $\bar{x} \pm D_G\sigma$ as outliers. It indicates that 99.7% of data are considered inliers and only rare events at approximately 1 of 370 samples are identified as an outlier.

This approach is very simple and attractive. However, normal mean and standard deviation are very sensitive to outliers. The above statistics are evaluated for the complete dataset, also including outliers. Thus, we use the MDist approach to determine the outliers, while obtained results are simultaneously influenced by them. Modifications with robust estimators[9] should minimize that influence.

2.2 MDist-rHub

Robust statistics have proposed a lot of new estimators for time series affected by outliers[28] apart from classical median or MAD. M-estimators with Huber or logistic ψ-functions give such an opportunity[47]. Location M-estimator $x^M H_0$ is defined as a solution to the equation (2):

$$\sum_{i=1}^{N} \psi\left(\frac{x_i - \bar{x}}{\hat{\sigma_0}}\right) = 0 \tag{2}$$

where $\psi(.)$ is any non-decreasing odd function, μ is a location estimator and $\hat{\sigma_0}$ is a preliminary scale estimator, like MAD (3).

$$MAD = median_i \left\{|x_i - \bar{x}|\right\}. \tag{3}$$

M-estimators are affine equivariant and (2) can be solved using the sample median as a starting point. The logistic smooth ψ function is defined as (4).

$$\psi_{\log}(x) = \frac{\mathrm{e}^x - 1}{\mathrm{e}^x + 1} \tag{4}$$

which may be reformulated as $\psi_{\log}(x) = 2F(x) - 1$, with $F(x) = 1/(1 + \mathrm{e}^{-x})$ denoting a cumulative distribution function of the logistic distribution, known as the sigmoid function. The scale M-estimator can be defined as a solution to (5).

$$\frac{1}{N}\sum_{i=1}^{N} \rho\left(\frac{x_i - \bar{x}_0}{\sigma}\right) = \kappa \tag{5}$$

where $0 < \kappa < \rho(\infty)$, $\rho(.)$ is even, differentiable and non-decreasing on the positive numbers loss function, σ is a location estimator and $\bar{x}_0$ is a preliminary location

estimator, like the median. While the logistic ψ function (4) is taken as $\rho(.)$, we obtain the logistic ψ scale estimator. Functions implemented in Matlab LIBRA toolbox[48] have been utilized in the considered research. There is an ongoing discussion about estimator robustness and the appropriate multiplier, despite object $D_H = D_g = 3.0$ is used, i.e., the same values as in the MDist-G case.

2.3 MDist-α

In some situations, the underlying normal stochastic process may not be appropriate. CPA research and analysis of control error variables indicate that in many cases an α-stable distributions better represent the generation mechanism. Domański[31] shows that it exhibits many attractive features. An α-stable distribution does not exhibit closed probabilistic density function (PDF). It is expressed with a characteristic equation (6).

$$F^{stab}_{\alpha,\beta,\delta,\gamma}(x) = \exp\left\{\mathrm{i}\delta x - |\gamma x|^{\alpha}\left(1 - \mathrm{i}\beta l(x)\right)\right\} \tag{6}$$

$$l(x) = \begin{cases} \operatorname{sgn}(x)\tan\left(\dfrac{\pi\alpha}{2}\right), & \text{for } \alpha \neq 1 \\ -\operatorname{sgn}(x)\dfrac{2}{\pi}\ln|x|, & \text{for } \alpha = 1 \end{cases}$$

where $0 < \alpha \leq 2$ is called a characteristic exponent or a stability index, $|\beta| \leq 1$ is a skewness factor, $\delta \in \mathbf{R}$ is a location and $\gamma > 0$ is a scale or dispersion factor.

There are special cases with a closed form of the PDF (6):

1) $\alpha = 2$ reflects independent realizations, especially for $\alpha = 2$, $\beta = 0$, $\gamma = 1$ and $\delta = 0$, we get exact normal distribution equation.

2) $\alpha = 1$ and $\beta = 0$ denote the Cauchy case that is considered in details in the following paragraph.

3) $\alpha = 0.5$ and $\beta = \pm 1$ denote α=0.1, β=1 case, which is not considered in the analysis.

Estimation of α-stable PDF parameters can be done with different methods, like McCuloch's (percentile) approach[49], iterative Koutrouvelis method using characteristics function estimation[50], logarithmic moment method[51] or maximum likelihood algorithm[52]. The quantiles method has been utilized in the considered research.

The α-stable PDF exhibits crossovers only for symmetric shape ($\beta = 0$), while the scale is constant ($\gamma = \text{const}$) and stability index α varies (see Fig. 1). The outer ones (denoted as a_1 and a_4) might reflect positions, where the tail starts, so they might be used as the outlier detection thresholds. They are evaluated numerically.

Finally, the outlier threshold must be agreed with the Z-scores multiplier for normal case. It gives the relation $D_\alpha = 1.468$.

2.4 ESD

The generalized extreme studentized deviate test proposed in [26] can be applied to data drawn from an approximately normal distribution to find out one or more outliers. The test assumes the upper limit for the outliers number in contrast to the Grubb's[20] and the Tietjen-Moore[24] tests, which demand an exact definition of the outliers number. Knowing the upper limit of the outliers number n, the test performs n separate tests: for one outlier, for two outliers, and so on up to N outliers.

Therefore, we test the null hypothesis that there is no outlier versus the alternative one, saying that there are N outliers at most. For a data set X with N elements, we generate n test statistics $T_1, T_2, \cdots, T_n$, where each T_i forms a two-sided Grubbs' statistics, defined as $X_1 = X$, $\bar{x}_i$ is the mean of X_i and σ_i is a standard deviation of X_i.

$$T_i = \frac{\max\left\{|x - \bar{x}_i| : x \in X_i\right\}}{\sigma_i} \tag{7}$$

and $X_{i+1} = X_i - x_i$, where $x_i \in X_i$ such that $|x - \bar{x}_i|$ is maximized. ESD test is the sequentially applied Grubb's test, but adjusted for the critical values based on the

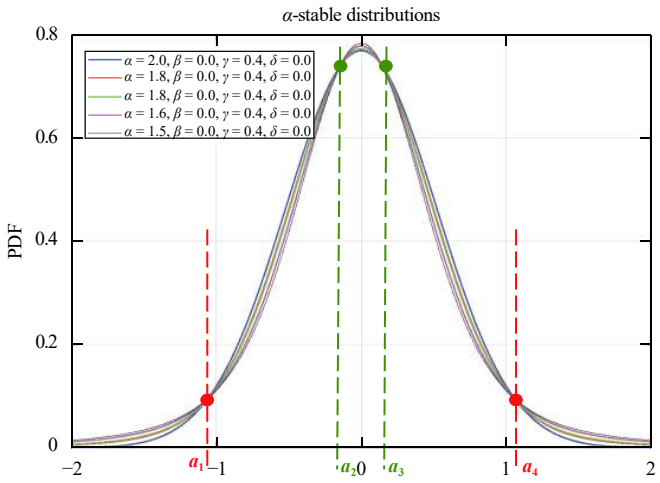


Fig. 1 α-stable distribution witch crossovers

number of tested outliers. The method is robust to the significant masking effect.

2.5 IQR

The majority of data is not normal enough to consider it as being drawn from a Gaussian distribution. A possible statistic in such case is the interquartile range (IQR) method[23]. It is calculated as the difference between upper 75th (denoted Q_3) and lower 25th (denoted Q_1) percentiles of the data. IQR may be used to find outliers. They are considered as observations that fall below $LL = Q_1 - 1.5$ IQR or above $HH = Q_3 + 1.5$ IQR. They are often presented in a box-plot: the highest and lowest occurring values are indicated by whiskers of the box and possible outliers are as individual points. The breakdown point for IQR is equal to 25%.

Once data are drawn from Gaussian distribution, the method gives more outliers than Gaussian MDist-G as the outliers lie outside the range $\mu \pm 2.698\sigma$.

2.6 IQR-α

The IQR-α method can be found under the name of Pareto tail modelling and is used for skewed data[53]. Actually, it is a general IQR approach, but it assumes α-stable PDF as an underlying outlier generating mechanism. Observations that are larger than a certain quantile of the fitted α-stable distribution are declared to be outliers. The IQR-α approach is acceptable for heavy-tailed distributions, but does not cope well with data that exhibit point-wise outliers lying far away from the centre[54].

Once the α-stable function is fitted to data, there remains only a question of what quantiles should be considered as thresholds. The literature is not clear at that point. Danielsson at al.[54] suggest to use 5% quantiles, while Alfons et al.[53] propose to use the 0.5% quantiles. Both variants are tested specified as IQR-$\alpha_{5\%}$ and IQR-$\alpha_{0.5\%}$, respectively.

3 Simulation study

The study presented in this paper focuses on control engineering, generally on the CPA preprocessing. There are two options to design simulation experiments. The first approach is to simulate complex and non-linear processes. It should be properly designed to reflect possible scenarios that may generate different kinds of outliers. The other way is to properly generate signals that reflect possible control errors that might be met in reality (also reflected by complex simulation). Both approaches would finally result in the same signals being analyzed, with the same generating mechanisms.

There is one more reason to support applied decision. During the realization of true industrial projects, an engineer is never sure what is behind the data and the control error variable. Thereby, it is important and practical to know what may be observed and why. Once we simulate different signals exhibiting different properties, we imitate certain situations met by an engineer. Thus, she/he may trace the reasons and perform root-cause analysis.

Both ways, i.e., complex process simulation and signal generation end up with the same signals, similar analysis and finally the observations. Because signal-based scenarios are closer to real CPA and allow faster root-cause analysis, this approach has been finally selected.

Consecutively simulations reflect phenomena observed in control practice. Univariate control loop sketched in Fig. 2, although very basic, reflects the majority of control structures in the process industry. Its performance may be observed and measured using various different measures and approaches[55]. These measures generally use control error signal $\varepsilon(t)$ during the calculation. The underlying hypothesis says that good controller design results in $\varepsilon(t)$ characterized by normal distribution $N(0, \sigma^2)$, i.e., with a zero mean and control error variance equal to σ^2. The best tuning exhibits the smallest variance.

However, nothing is ideal and perfect in reality. Observation of industrial variables show that the majority of signals do not agree with this hypothesis[32]. The signals are often heavy-tailed, skewed or even hold more strange shapes. In general, three main hypotheses can be observed and should be taken into account:

H1: Control error has underlying symmetric normal distribution.

H2: Control error has underlying skewed distribution.

H3: Underlying generation process comes from heavy-tailed distribution.

Hypothesis H1 assumes a symmetric main process. Tested contamination may be of different origins. It might be another Gaussian noise with varying share or with differing variances. It can be symmetrical or skewed. Such scenarios reflect impact of external processes (disturbances) on the control loop. Other contamination

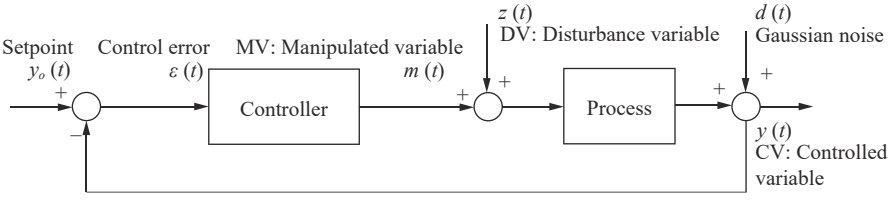


Fig. 2 Univariate control loop

schemes will represent erroneous observation (measurement errors, data transfer artifacts, human interventions). In such cases, some points are artificially shifted away from the centre. In H2 hypothesis, the main data is skewed, while contamination may be different.

Hypothesis H3 means that control error is drawn from the process similar to the fat-tailed distribution. The heavy tails hypothesis requires more attention. Actually, the heavy tailed distribution has a tail that is heavier than an exponential distribution[56]. Two stochastic processes are tested: Laplace double exponential distribution and α-stable distribution with varying stability index α reflecting tail heaviness and varying skewness β representing asymmetric behavior. Tails simulate unknown, varying and uncoupled correlations and persistent disturbing processes, while skewness reflects process non-linearities, constraints or misfit in the operating point.

It should be noted that there are some differences in views in the literature about a connection between outliers and tail heaviness. The rather common understanding that tails represent outliers is contested by some researchers. Klebanov and Volchenkova[57] say that, the idea on connection of the presence of outliers with the heaviness of distributional tails is wrong. Thereby, this aspect remains open in the conducted research. The focus remains on the characteristics of time series with cautious results interpretation.

3.1 Gaussian control error (H1)

This section will be considered as the reference hypothesis. It is assumed that Gaussian control error represents a well tuned loop. The analysis in all cases is conducted for a simulated time series $X = \{x_1, x_2, \cdots, x_N\}$. We assume that the underlying normal data are distributed with $N(0, \sigma_M^2)$ and $\sigma_M = 0.6$. Dataset length $N = 50\,000$ is kept constant in each simulation run. The time trend for the first 2000 points of original signal is shown in Fig. 3. Histograms and fitted PDFs are shown in Fig. 4(a).

Results of outlier detection are shown in Table 1. Graphical representation is sketched in Fig. 4(b). We notice that MDist methods (M.1, M.2 and M.3) give similar indications minimizing detected outliers' number, what agrees with expectations. ESD (M.4) detects zero outliers. Other approaches are less conservative. IQR-α with larger confidence lever 5% is highly relaxed and labels a lot of outliers.

Four types of scenarios are selected for further analysis, as they reflect situations with symmetric or skewed contamination:

Table 1 Detected outliers for normal process (H1)

	minTh	maxTh	leftPts	rightPts	outPerc
M.1	−1.80	1.80	74	62	0.27
M.2	−1.81	1.81	74	62	0.27
M.3	−1.77	1.78	82	66	0.30
M.4	n/d	n/d	0	0	0.00
M.5	−1.62	1.62	177	166	0.69
$M.6_{0.5\%}$	−1.55	1.54	272	234	1.01
$M.6_{5\%}$	−0.98	0.98	2656	2554	10.42

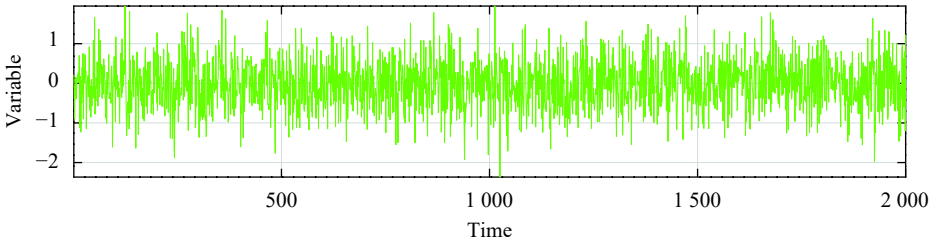
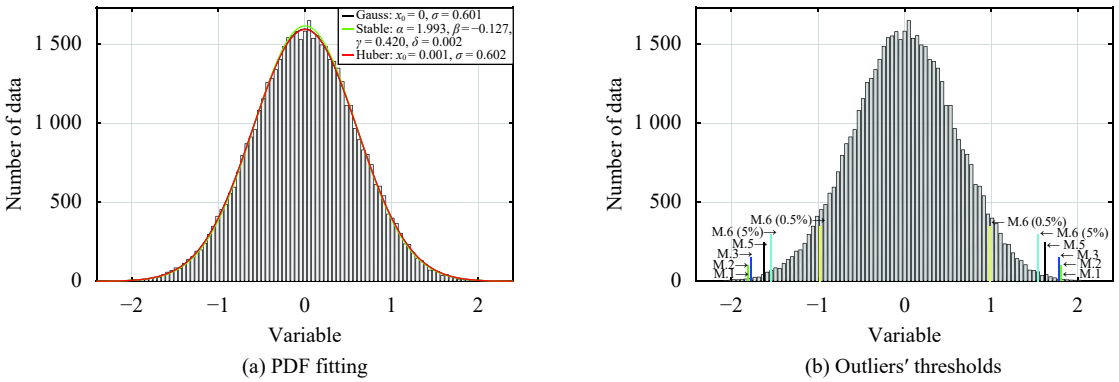


Fig. 3 Normally distributed control error (H1)



(a) PDF fitting (b) Outliers' thresholds

Fig. 4 Control error histograms (H1)

H1.1: Gaussian contamination – varying share

H1.2: Gaussian contamination – varying standard deviation

H1.3: Skewed contamination derived from gamma distribution

H1.4: Injected erroneous observations: one-sided and two-sided.

3.1.1 Contamination with Gaussian (H1.1)

In this section, the effect of contamination with another normal distribution is checked. The contaminating process $N(0, \sigma_c^2)$ has standard deviation three times larger than the main process, i.e., $\sigma_c = 3\,\sigma_M = 1.8$. The share of the induced outliers, points of different stochastic process, is tested. Seven contamination shares are used: $c_{sh} = [0.5\%, 1\%, 2\%, 5\%, 10\%, 20\%, 25\%]$. Aggregated results are presented to save the space and show major effects.

Contamination changes statistical properties of the obtained time series. It affects estimation of the standard deviation and tail properties. Therefore, it affects outlier detection results. Two types of the relation are presented to capture these effects. At first, relationship of fitted normal standard deviation σ, robust standard deviation σ_{rob}, and characteristic exponent α and scale γ of α-stable distributions are presented in Fig. 5(a). Observation of the statistical properties relations agrees with expectations. Increasing share of the contamination increases normal standard deviation. Scale robust estimators are less sensitive to the above effect, similarly to the stable distribution scale (see Table 2). Normal standard deviation σ changes by approximately 66%, while σ_{rob} by approximately 22% and γ only by approximately 16%. We see that larger contamination increases tails, which is observed by decreasing value of stability factor α (approximately 23%).

Further relations present how many new outliers are detected versus a contamination share. The number of new outliers is a difference between the total number of outliers found for some selected contamination and a number for zero contamination. Fig. 5(b) presents this relation for an original dataset.

Table 2 Scope of changes for statistical factors (H1), robust results in bold

Hypothesis	σ	MAD	σ_{rob}	α	γ
H1.1	66.3%	44.0%	22.7%	−23.5%	**17.0%**
H1.2	94.8%	61.3%	30.5%	−28.1%	**23.3%**
H1.3	92.5%	59.0%	22.1%	−34.2%	**6.6%**
H1.4 (one)	58.2%	37.8%	11.0%	−42.5%	**−8.5%**
H1.4 (two)	61.6%	35.4%	10.5%	−18.6%	**6.3%**

Observation of the simulation results allows us to formulate initial observations. First of all, IQR method (M.5) is the most sensitive to the range of the contamination share. Moreover, two other robust MDist threshold methods: MDist-α and MDist-rHub are also significantly sensitive. These methods tend to detect more outliers as the original time series is more contaminated, in contrary to the ESD algorithm and IQR-$\alpha_{5\%}$. These two approaches are fully robust, and their detection remains unchanged. IQR-$\alpha_{0.5\%}$ goes even further, as it starts to detect even fewer outliers. We notice that these approaches consider observations constituting the tail as normal occurrences (inliers). Classical 3σ threshold MDist-G method increases detection up to 10% share, while for larger contamination, the detection ability saturates.

3.1.2 Contamination with Gaussian (H1.2)

In this section, the contamination share is kept constant at $c_{sh} = 5\%$, while the variance of the contaminating signal varies. As in the previous case, standard deviation of the original signal equals to $\sigma_M = 0.6$ and this signal remains exactly the same. The following standard deviations of the contaminating signal are simulated: $\sigma_c = [0.9, 1.2, 1.5, 1.8, 2.1, 2.4, 2.7]$.

Statistical properties behave according to expectations (see Fig. 6(a)) similarly to the previous case. Growing polluting variance increases normal standard deviation. Scale robust estimators are less sensitive to the above effect, similarly to the stable distribution scale. Larger contamination increases tails, which is observed by the diminishing value of the exponential factor. The range of this effect is indicated in a summarizing Table 2.

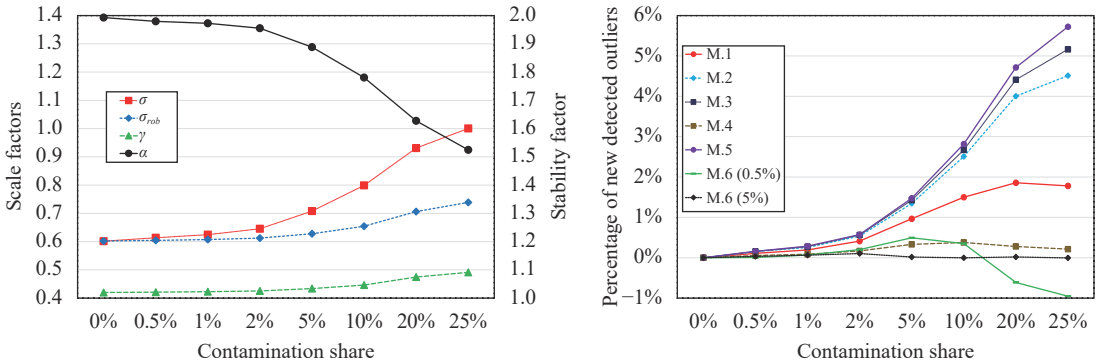
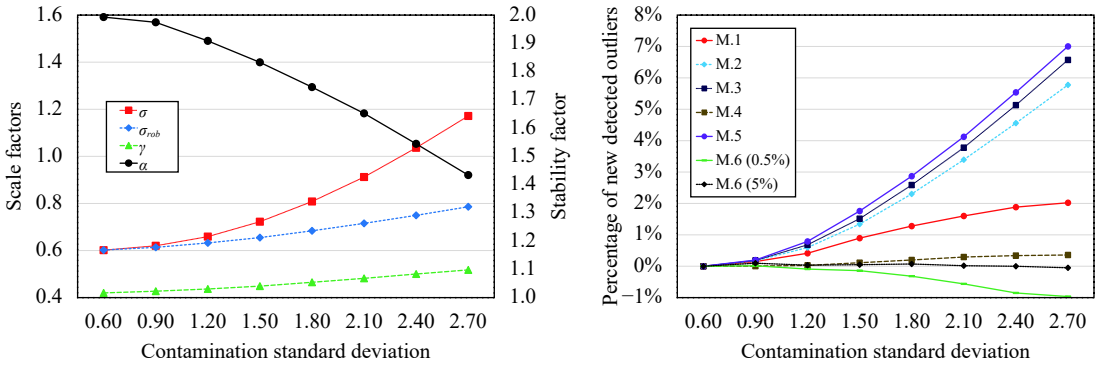


Fig. 5 Contamination share relations



(a) Statistical factors (b) New detected outliers

Fig. 6 Relations depending on varying contamination σ

The diagram in Fig. 6(b) presents how many new outliers are detected versus a contamination varying standard deviation for a resulting dataset.

Observations are exactly the same as previously. The ESD algorithm and IQR-$\alpha_{5\%}$ do not label new contaminating observations as outliers and tend to consider them as inliers. IQR, MDist-α and MDist-rHub are significantly more sensitive, especially IQR. We notice that this dependence is linear with the contaminating standard deviation. MDist-G method initially increases detection, but larger σ results in constant robustness. IQR-$\alpha_{0.5\%}$ detects even fewer outliers.

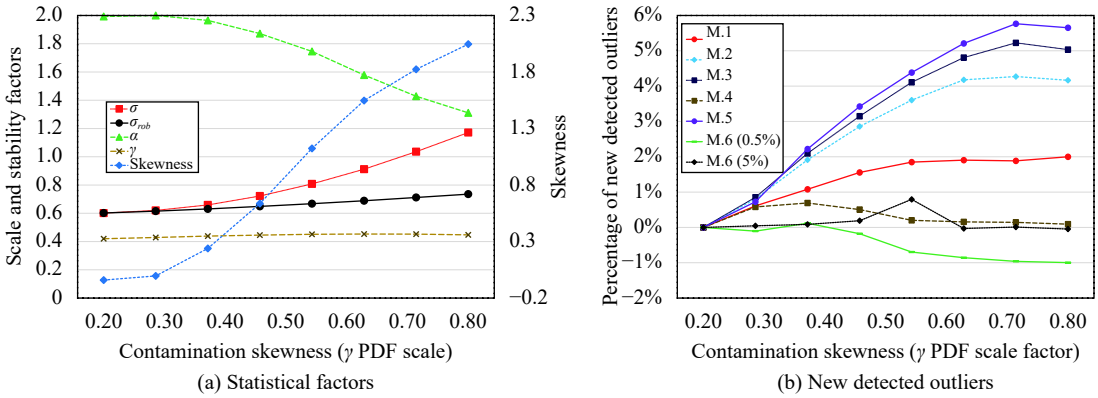
3.1.3 Asymmetric contamination (H1.3)

Asymmetry in the contaminating process is analyzed in this section. The underlying time series remains the same as $N(0, \sigma_M^2)$. However, time series pollution is generated using asymmetric mechanism drawn from the gamma distribution[58] characterized by shape parameter k and scale θ. Shape factor is kept constant as $k = 5.0$. Different scales, which increase contaminated data skewness are analyzed, i.e., $\theta = [0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8]$.

Statistical properties of the asymmetrically contaminated time series for an original dataset are summarized in Fig. 7(a). We see that contamination with skewed stochastic processes originating from gamma distributions have direct effects on the total skewness. The most proper reaction is visible with a γ scale factor of α-stable distribution, which remains fully robust to asymmetry and one-sided tails. Gaussian standard deviation significantly increases, what may cause misinterpretation and in consequence improper outlier detection. Robust σ_{rob} estimate behaves also properly, though it is not as constant as γ. Skewed tail causes decrease in stability factor α.

Differences between distributions are reflected in the outlier detection results. Observing the definition of these algorithms, we notice that some of them are symmetrical in a definition, like MDist approaches (M.1, M.2 and M.3) and ESD. The algorithms using quantiles, i.e., IQR variants (M.5 and M.6), take into account skewness of data. It is better visible on time series histograms with added detection threshold shown in Fig. 8.

The Gaussian MDist-G method is mostly biased by the tails in both domains and it labels as outliers only the most extreme observations. The other two robust versions of the MDist approach behave slightly better. The ESD method (M.4), as in all previous cases, is the most conservative always detecting the most extreme realizations. Quantile-based algorithms, i.e., IQR (M.5) and IQR-α (M.6) work quite properly in both domains, however quantiles selection impedes results. The 0.5% margin is extremely conservative detecting only the most extreme observations. Moreover, it increases its conservativeness with increasing contaminating skewness. In



(a) Statistical factors (b) New detected outliers

Fig. 7 Asymmetry impact

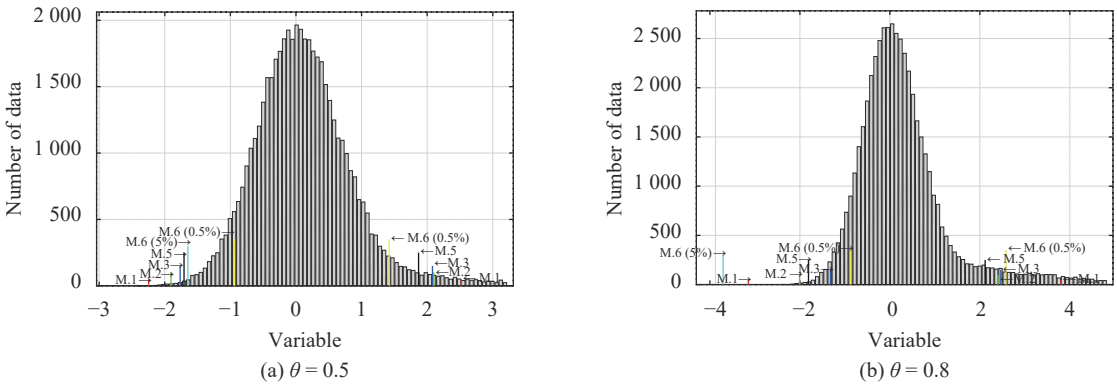


Fig. 8 Histograms with outliers' thresholds

contrary, the 5% confidence margin is not affected by the induced outliers.

Concluding, we may notice that skewed contamination creates bigger challenges for outlier detection than a symmetrical one. In such case, it is suggested to compare the results between IQR approach applied to original skewed data with robust MDist methods (MDist-rHub and MDist-α). A serious question arises. How to detect opposite situation, i.e., skewed original data and symmetric contamination?

3.1.4 Contamination with erroneous observations (H1.4)

The issue of the erroneous observations is addressed in these simulations. Once there are some unknown, artificial problems with the measurement unit, data collection system or communication loss, target data may include some clearly invalid observations, which often appear on system limits and always have the same value. We may distinguish two such cases: artificial outlier (errors) appear on one side (upper or lower limit) or on both sides simultaneously. These effects are investigated in the following paragraphs, analyzing the effect of varying number (share) of such a contamination c_{sh}.

One-sided erroneous observations.

To address this issue, the data are contaminated with randomly injected constant values $x_i = 3$. Seven different contamination shares are investigated: c_{sh} = [1%, 2%, 3%, 4%, 5%, 6%, 7%]. Aggregated simulation results are presented below. Fig. 9(a) shows the relationship between the number of injected error observations and the main statistical factors for both datasets. Robustness of scale indexes is confirmed. We notice that stable distribution scaling factor γ even decreases (Table 2).

The signal histogram shows these erroneous observations clearly. However, the use of MDist scores might not work properly. It is shown clearly in Fig. 10. The Gaussian method works properly for shares $c_{sh} \leq 6\%$, while for $c_{sh} = 7\%$ the injected wrong observations are not detected. It is also visible in Fig. 9(b) showing the aggregated summary. It shows that the use of low confidence IQR-α may fall under the same risk.

The review of results shows that the robust MDist-rHub approach is the safest and unbiased, minimizing the risk around the peak of outliers. The standard IQR method might be useful as well, but with a tendency to label more observations as outliers. This simulation shows that even detection of the simplest looking outliers is not so straightforward using automatic measures. Additional risk lies in the fact that one sided outliers affect mean value and robust estimates are required. Manual inspections are inevitable.

Two-sided erroneous observations.

Next, data are contaminated with randomly injected constant values $x_i = \pm 3$. Seven contamination shares are

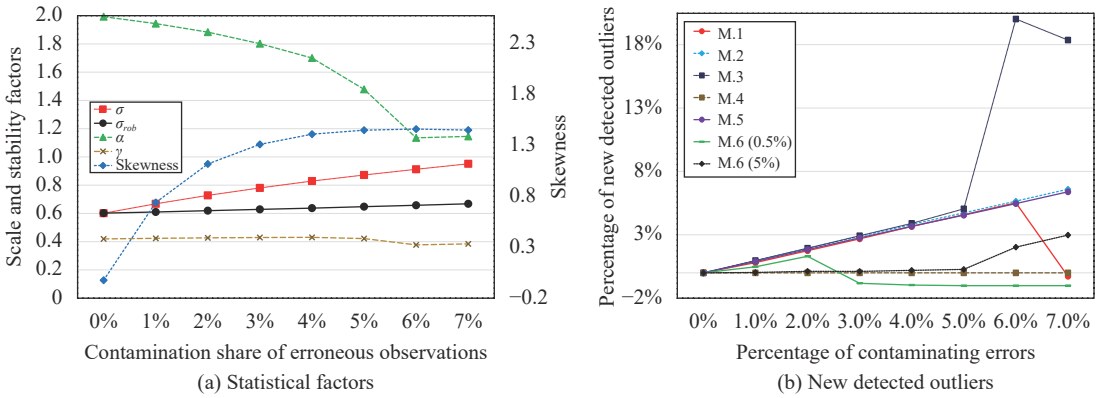
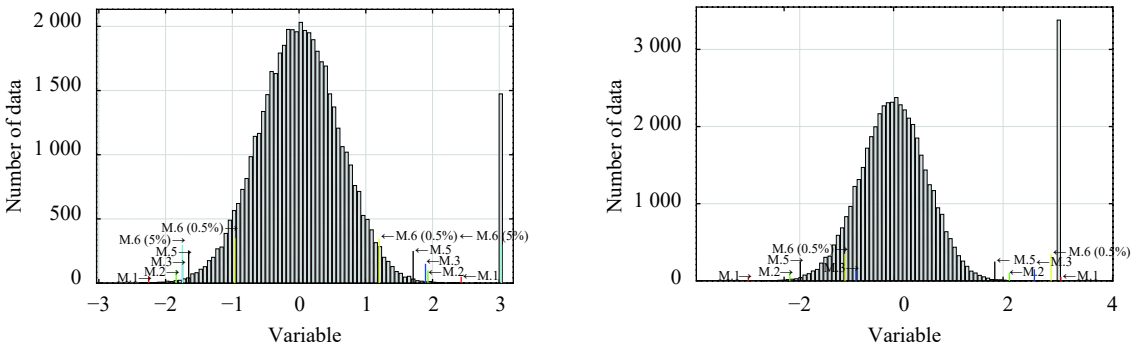


Fig. 9 One-sided errors impact



(a) $c_{sh} = 3\%$ (b) $c_{sh} = 7\%$

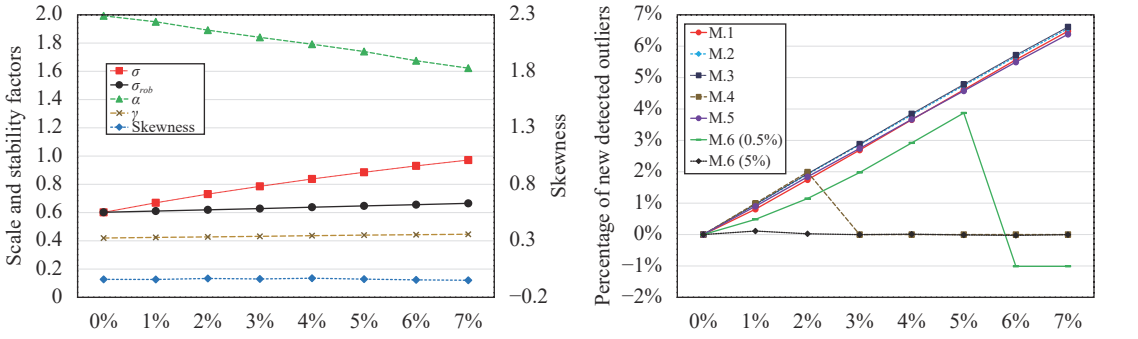
Fig. 10 Histograms with outliers' thresholds (one-sided errors)

investigated: $c_{sh} = [1\%, 2\%, 3\%, 4\%, 5\%, 6\%, 7\%]$. Aggregated results and observations are presented below. The statistical relationship is investigated at first. Fig. 11(a) summarizes the main statistical factors for a varying share of injected errors. Two-sided extremes increase standard deviation, while other estimators (especially γ) give robust estimations, despite the number of injected outliers. Furthermore, they cause tails, which are detected by diminishing stability factor α.

Two-sided errors seem to be simpler in detection than previous asymmetric case. Although two-sided outliers do not affect mean estimation, it is still safer to use approaches utilizing robust estimators (Fig. 11(b)). Observing exemplary histograms sketched in Fig. 12, MDist-rHub and MDist-α are suggested to be used in such situations. It is further confirmed by a summary of the results. Thereby, IQR-α is very sensitive to the confidence ratio selection. The literature suggestion 0.5% seems to be too conservative, while 5% is over-relaxed.

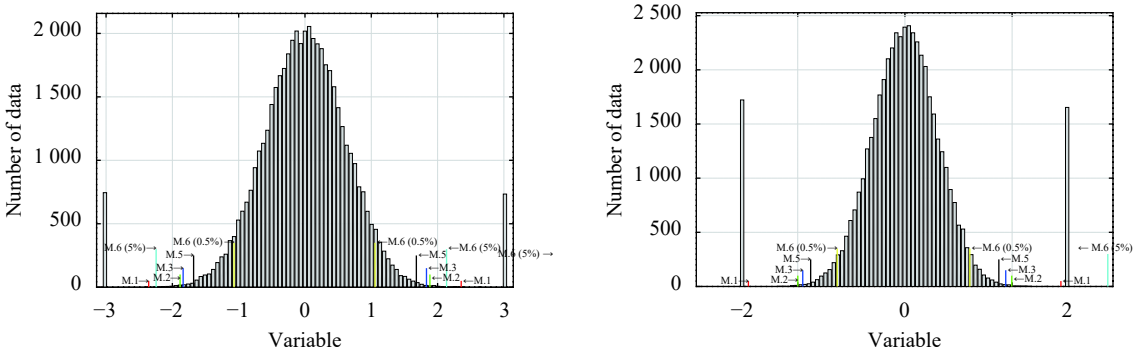
3.1.5 Summary of H1 hypothesis

Results of the H1 hypothesis analysis are summarized in Table 2. We see that scaling factor γ drawn from the α-stable distribution is the most robust in all analyzed scenarios. It suggests that γ is the most promising coefficient for further consideration.



(a) Statistical factors (b) New detected outliers

Fig. 11 Two-sided errors impact



(a) $c_{sh} = 3\%$ (b) $c_{sh} = 7\%$

Fig. 12 Two-sided errors histograms with thresholds

3.2 Asymmetric underlying process (H2)

Previously considered scenarios assumed that the underlying process is symmetric, while the contaminating generator might be any. Now, the roles are inverted. The basic process is skewed. There exist many scenarios, with different combinations between the main data and contamination. The following contamination scenarios are selected:

H2.1: Gaussian with different shares

H2.2: Gaussian with varying standard deviation

H2.3: Skewed process derived from gamma distribution

H2.4: One-sided injected erroneous observations.

The underlying stochastic process is generated with Pearson distribution[59] having parameters: mean $\bar{x} = 0$, standard deviation $\sigma_M = 0.6$, skewness $\beta = 0.7$ and kurtosis equal to 5.4. Dataset length $N = 50\,000$ is kept constant during each simulation. Time trend for first 2 000 points is shown in Fig. 13. Histogram together with fitted probabilistic density functions is sketched in Fig. 14 (a).

Results of outlier detection are shown in Table 3. Graphical representation of thresholds for each method is indicated in Fig. 14 (b).

We notice that MDist methods (M.1, M.2 and M.3) give similar indications minimizing detected outliers' number, which agrees with expectations. ESD (M.4) detects only extreme outliers. Other approaches, especially classical IQR (M.5) and high confidence IQR-$\alpha_{5\%}$, are less conservative.

3.2.1 Gaussian with different shares (H2.1)

In this section, the simplest contamination scheme is analyzed. Original asymmetric data are affected by normal Gaussian process $N(0, \sigma_c^2)$ with its standard deviation three times larger than the one of main process, i.e., $\sigma_c = 3\sigma_M = 1.8$. Seven contamination shares are used: $c_{sh} = [2\%, 4\%, 6\%, 8\%, 10\%, 12\%, 14\%]$. The impact of analyzed contamination on final data statistics is shown in Fig. 15(a).

Table 3 Detected outliers for skewed process (H2)

	minTh	maxTh	leftPts	rightPts	outPerc
M.1	−1.81	1.81	62	435	0.99
M.2	−1.65	1.61	110	690	1.60
M.3	−1.52	1.59	164	728	1.78
M.4	3.33	3.38	1	45	0.09
M.5	−1.48	1.43	196	1 056	2.50
$M.6_{0.5\%}$	−1.30	3.06	435	36	0.94
$M.6_{5\%}$	−0.85	1.05	2 823	2 511	10.67

First of all, we see that the increasing number of contaminating data increases final scaling estimates. This relation seems to be linear. Furthermore, the robustness of stable distribution scaling γ is confirmed. We also see that an increasing share of points drawn from a symmetrical contamination process decreases final skewness. A summary of these results is sketched in Table 4.

Changes in the properties of the statistical factors affect outlier detection methods, as they use them. Obtained results are quite in line with previously observed relations. Both robust MDist methods (M.2 and M.3) together with IQR increase number of detected outliers as the contaminating share c_{sh} increases. In contrary, the

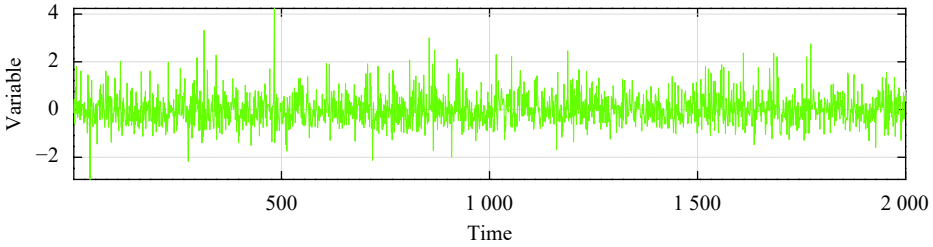
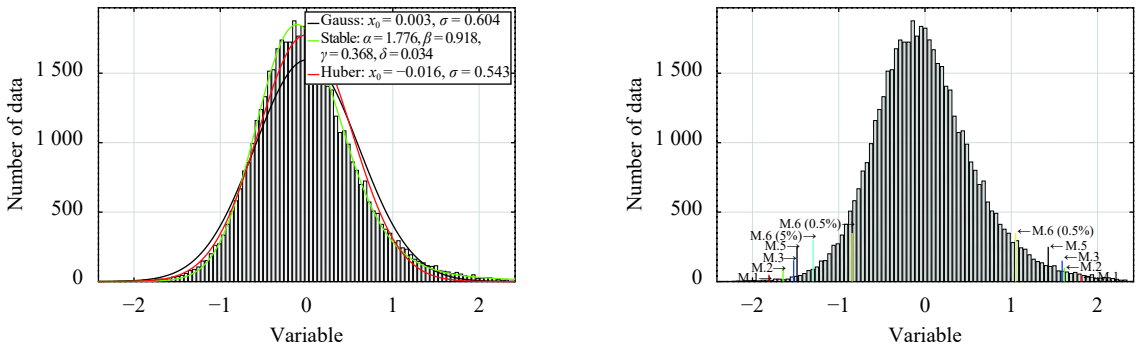
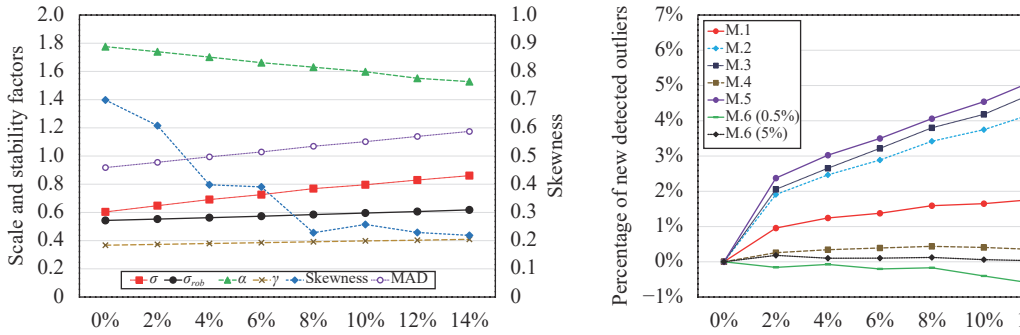


Fig. 13 Normally distributed control error (H2)



(a) PDF fitting (b) Outliers' thresholds

Fig. 14 Histogram for skewed control error (H2)



(a) Statistical factors (b) New detected outliers

Fig. 15 Contamination share impact (skewed data)

Table 4 Scope of changes for statistical factors (H2), robust results in bold

Hypothesis	σ	MAD	σ_{rob}	α	γ
H2.1	42.6%	27.9%	13.8%	−14.0%	**11.4%**
H2.2	80.5%	53.6%	26.8%	−22.4%	**22.2%**
H2.3-right	93.4%	67.8%	25.7%	−28.5%	**10.5%**
H2.3-left	93.2%	60.1%	21.0%	−33.0%	**5.6%**
H2.4-right	245.9%	147.8%	**12.4%**	−59.6%	−49.7%
H2.4-left	92.4%	55.5%	**10.7%**	−47.1%	−16.6%

ESD method considers all new observations as inliers. Performance of the IQR-α strongly depends on the selected confidence value. Small confidence 0.5% does not detect new outliers, while 5% behaves similarly to robust MDist approaches. Exemplary histograms with detection thresholds are presented in Fig. 16.

3.2.2 Gaussian with varying variance (H2.2)

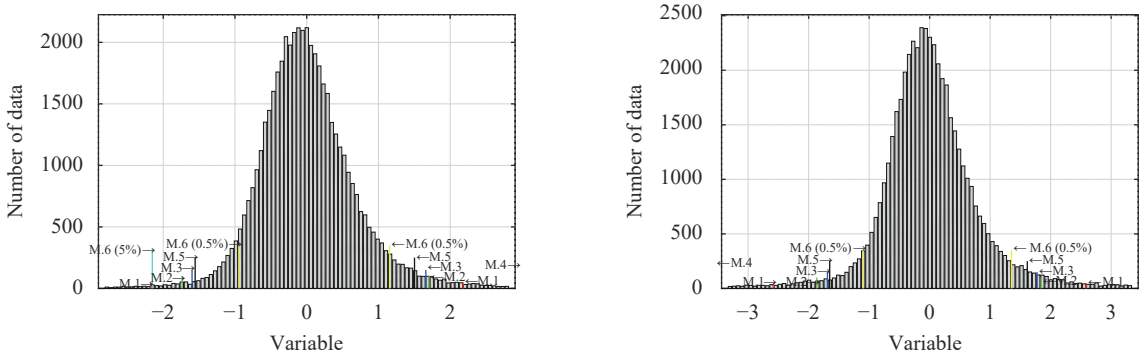
Now contamination share is kept constant at $c_{sh} = 4\%$, while the variance of the contaminating signal varies. The standard deviation of the original signal equals to $\sigma_M = 0.6$ and this signal remains exactly the same. The following standard deviations of the contaminating signal are simulated: $\sigma_c = [0.9, 1.2, 1.5, 1.8, 2.1, 2.4, 2.7]$.

Statistical properties behave according to expectations and are shown in Fig. 17(a). Relations are similar to H2.1 scenario. Growing polluting variance increases normal standard deviation. Scale robust estimators are less sensitive to the above effect, especially stable distribution scale γ. Their increase is no longer linear and starts to be exponential. Larger contamination increases tails, which is observed by the diminishing value of the exponential factor. Simultaneously, an overall skewness decreases toward symmetric behavior. The range of this effect is indicated in Table 4. The diagram in Fig. 17 (b) presents the number of new detected outliers versus standard deviation for resulting data.

The results are quite in line with previously observed relations. The only difference is at the start. The contaminating signal is hidden inside of the main process, i.e., σ_c is close to the main signal σ_M. Fluctuations are faster but the relationship becomes linear for larger variances. Detection methods' performance exhibits similarly to the previous scenario. Exemplary histograms with thresholds are sketched in Fig. 18.

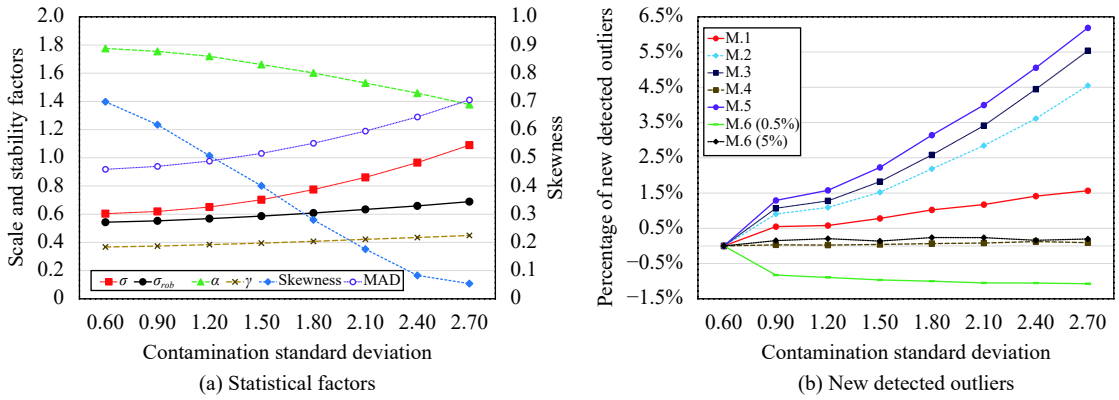
3.2.3 Skewed contamination (H2.3)

Asymmetry in the contaminating process is analyzed now. The underlying time series remains normal. However, time series pollution is generated using an asymmetric mechanism drawn from the gamma distribution characterized by shape parameter k and scale θ. Shape factor is kept constant as $k = 5.0$. Different scales,



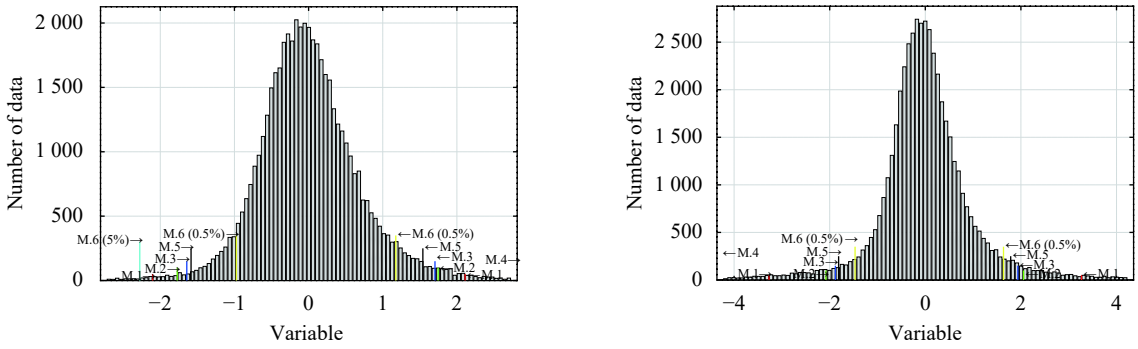
(a) $c_{sh} = 6\%$ (b) $c_{sh} = 14\%$

Fig. 16 Skewed data varying share histograms with thresholds



(a) Statistical factors (b) New detected outliers

Fig. 17 Contamination variance impact (skewed data)



(a) $\sigma_c = 1.2$ (b) $\sigma_c = 2.7$

Fig. 18 Skewed data varying variance histograms with thresholds

which increase contaminated data skewness are analyzed, i.e., $\theta = [0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8]$. While the underlying process is asymmetric and skewed to the right, there is expected difference depending on the side of contamination. Thereby, two versions of the simulations have been performed, with right-skewed and left-skewed contaminating signal.

Right-sided skewed contamination. Statistical properties of the right-sided contaminated data are summarized in Fig. 19. We see that contamination with skewed stochastic process originating from gamma distribution have direct effects on the total skewness. The most proper reaction is visible with γ scale factor of α-stable distribution, which remains fully robust to asymmetry and one-sided tails. Gaussian standard deviation increases significantly. It may cause misinterpretation and improper outlier detection. Robust σ_{rob} estimate behaves also properly, though it is not as constant as γ. Skewed tail causes decrease in stability factor α. The range of changes is summarized in Table 4. The resulting skewness coefficient reflects impact of the asymmetric contamination process.

Outlier detection methods perform similarly to the previous cases with MDist-rHub, MDist-α and IQR working consistently.

Left-sided skewed contamination. Opposite-side contamination does not make changes in the outlier detection properties. Statistical properties behave as previously, with an exception for the skewness factor which changes in the opposite direction (Fig. 20(a)). Such a performance causes similar detection represented in Fig. 20(b).

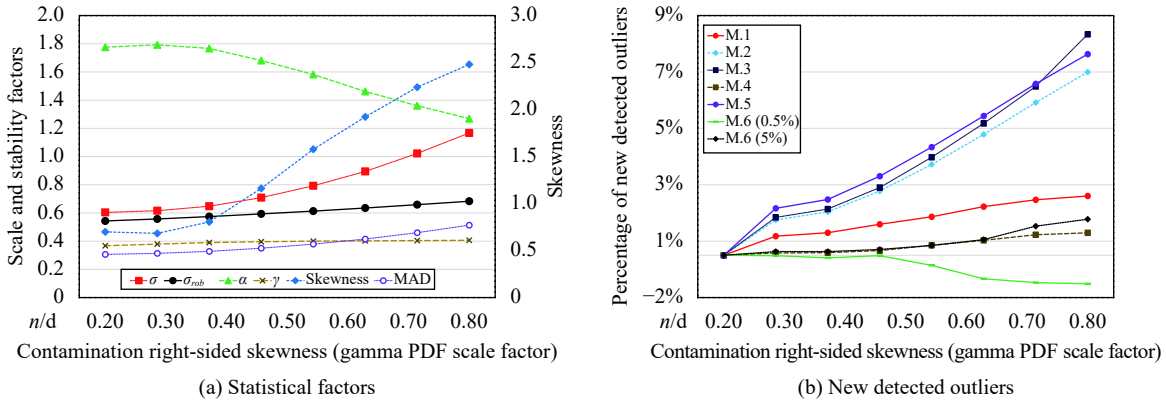
Comparison between the impact of opposite sides is shown in Fig. 21. Observation of the methods' performance suggests that robust MDist methods (M.2 and M.3) seem to be the most reliable with M.2 (MDist-rHub) being more conservative.

3.2.4 One-sided erroneous observations (H2.4)

Erroneous observations are addressed in this section. As the main process is not symmetric, it is interested in how one-sided errors would affect detection. Thus, we consider two cases: artificial observations (errors) appear on the left or on the right side.

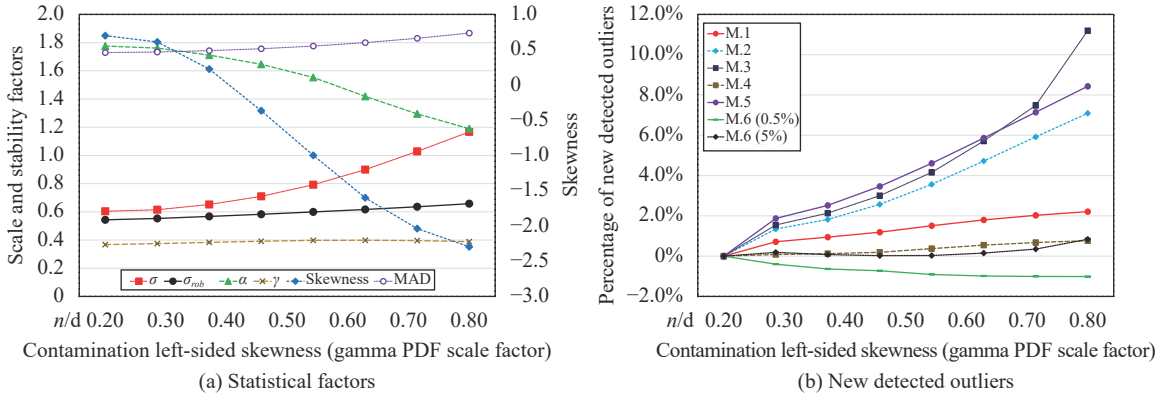
Right-sided errors. To address this issue, the data are contaminated with randomly injected constant values $x_i = +8$. Seven different contamination shares are investigated: $c_{sh} = [1\%, 2\%, 3\%, 4\%, 5\%, 6\%, 7\%]$. Aggregated simulation results are presented in Fig. 22, which shows the relationship between the number of injected error observations and the main statistical factors. Robustness of scale indexes is confirmed (stable distribution scaling factor γ even decreases) and summarized in Table 4.

Outlier detection methods perform similarly to the symmetric considerations (hypothesis H1.4). The risk of missing the peak by the threshold is serious for the most relaxed methods, i.e., MDist-α and IQR-α with large con-



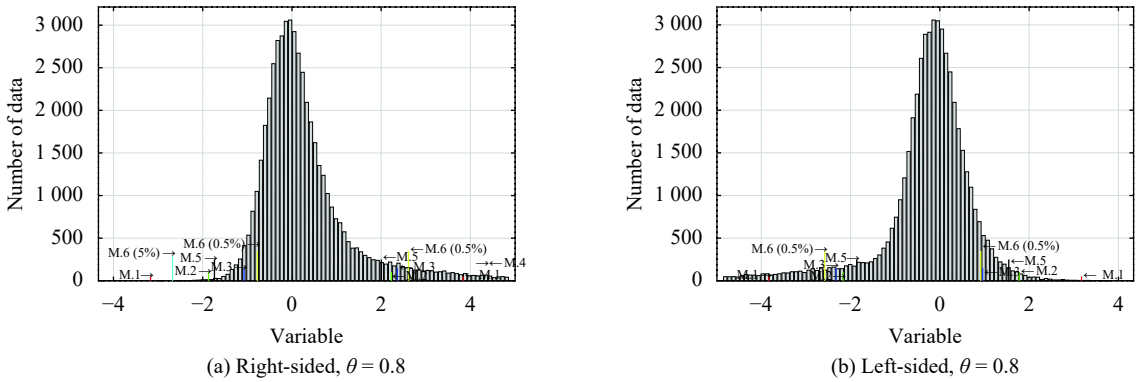
(a) Statistical factors (b) New detected outliers

Fig. 19 Skewed data with right-sided skewness impact (H 2.3)



(a) Statistical factors (b) New detected outliers

Fig. 20 Skewed data with left-sided skewness impact (H 2.3)



(a) Right-sided, $\theta = 0.8$ (b) Left-sided, $\theta = 0.8$

Fig. 21 Skewed data and asymmetric contamination thresholds

fidence interval 5%. It is caused by biased location estimate used by the method.

Left-sided errors. Opposite-side contamination does not make any fundamental change in outlier labelling. Statistical properties behave as previously, with an exception of the skewness factor which changes in the opposite direction (Fig. 23(a)). Such a performance causes similar detection as in Fig. 23(b). The risk of wrong labelling appears, but to a smaller extent. It is due to the lesser impact of the biased location estimate.

Comparison between opposite sides impact is shown in Fig. 24. Observation of the methods' performance suggests that MDist-rHub method is the most reliable.

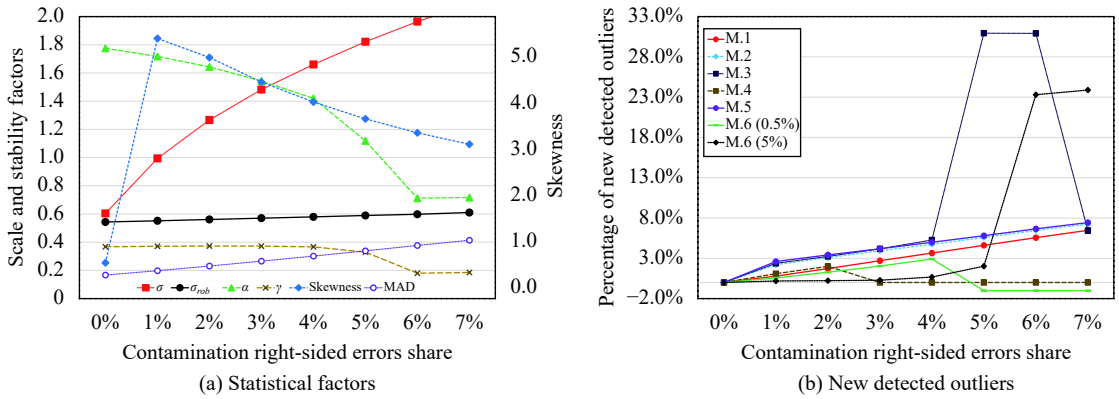
Similarly to the symmetric underlying process results analyzed with the hypothesis H1.4, we may conclude that automatic statistical detection of the constant erroneous observations may pose the biggest challenge.

3.2.5 Summary of H2 hypothesis

Observations done during the analysis of H2 hypothesis are summarized in Table 4. We see that scaling factor γ drawn from the α-stable distribution is the most robust in all analyzed scenarios. It suggests that γ is the most promising coefficient to be further considered in evaluations.

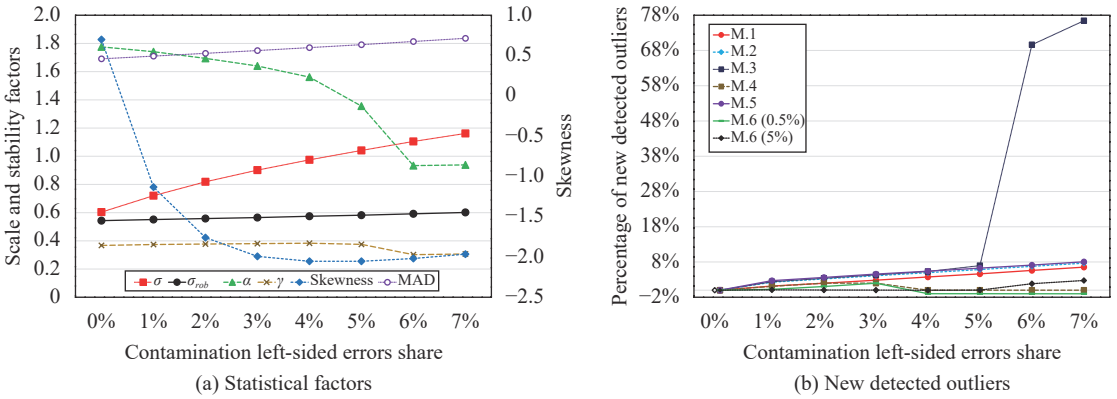
3.3 Fat-tailed underlying process (H3)

All above considered contamination scenarios as-



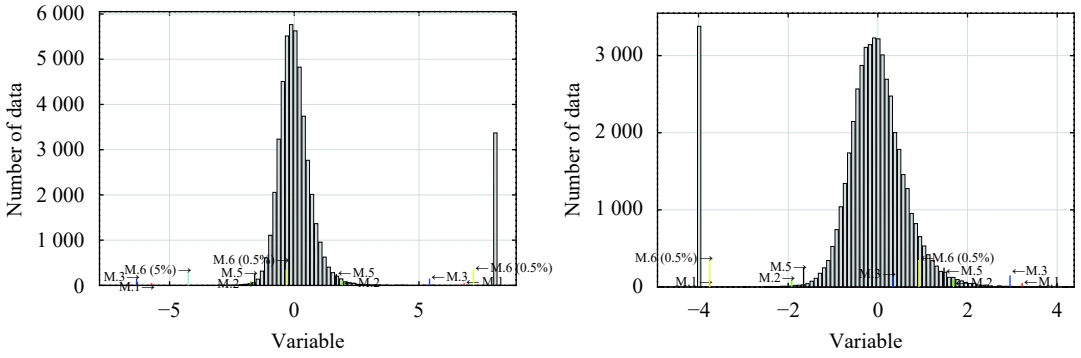
(a) Statistical factors (b) New detected outliers

Fig. 22 Skewed data with right-sided errors impact (H 2.4)



(a) Statistical factors (b) New detected outliers

Fig. 23 Skewed data with left-sided skewness impact (H 2.4)



(a) Right-side, $\theta = 0.8$ (b) Left-side, $\theta = 0.8$

Fig. 24 Skewed data and asymmetric contamination thresholds (H 2.4)

sumed that the underlying basic stochastic process is thin-tailed. In this case, simple approaches using fat-tailed underlying generating mechanisms are analyzed:

H3.1: Laplace double exponential process

H3.2: α-stable stochastic process

H3.3: α-stable stochastic process with varying characteristic exponent α

H3.4: α-stable stochastic process with varying skewness β

In these cases, there is no contamination artificially induced, as it is assumed that the outliers are hidden within the tails.

3.3.1 Laplace underlying generation process (H3.1)

Laplace distribution is also called double exponential. It is symmetrical and forms a function of a difference between two independent variables with identical exponential distributions. Its probability density function is given by (8).

$$PDF_{\mu,b} = \frac{1}{2b} \mathrm{e}^{-\frac{|x-\mu|}{b}} \tag{8}$$

where $\mu \in \mathbf{R}$ is location factor and $b > 0$ is scale parameter. Its shape decays exponentially and depends on parameter b.

Dataset length $N = 50\,000$ is kept unchanged. Time trend for first 2 000 points is shown in Fig. 25. Histogram together with fitted normal, robust normal and α-stable probabilistic density functions is sketched in Fig. 26(a).

Estimated statistical factors are presented in Table 5. We see that double exponential distribution has heavy tails. It is expected that outlier labelling using a Gaussian approach might not be efficient.

Results of outlier detection are shown in Table 6 and a graphical diagram representing detection thresholds is sketched in Fig. 26(b). We notice that Gaussian MDist method minimizes detected outliers' number, while robust estimators label more observations. The ESD (M.4) detects only a few outliers. ESD approach is less conservative. IQR-α with larger confidence level 5% is extremely relaxed and labels a very high number of observations as outliers. In contrast, the 0.5% confidence level is very conservative. It labels only the most extreme observations as outliers.

3.3.2 α-stable underlying generation process (H3.2)

α-stable distribution belongs to the family of stable distributions. It has more degrees of freedom as it is parametrized by four parameters: $0 < \alpha \leq 2$ called stability index, $|\beta| \leq 1$ called factor, $\delta \in \mathbf{R}$ as distribution location and $\gamma > 0$ as its scale.

Dataset length $N = 50\,000$ is kept the same. The following parameters are set: $\alpha = 2.0$, $\beta = 0.0$, $\gamma = 0.6/\sqrt{2}$ and $\delta = 2.0$. Time trend for first 2 000 points is shown in Fig. 27. Histogram together with fitted normal, robust normal and α-stable probabilistic density functions is sketched in Fig. 28(a). The extreme values and resulting heavy tails are well visible.

Table 5 Estimated statistical properties of Laplace process

Min	Max	Mean $\bar{x}$	Median	Robust $\bar{x}_{rob}$
−10.48	10.76	0.005	0.004	0.006
σ	σ_{rob}	MAD	Kurtosis	Skewness
1.415	1.062	0.998	6.13	−0.01
α	β	γ	δ	
1.458	−0.002	0.701	0.003	

Table 6 Detected outliers for Laplace process (H3.1)

	minTh	maxTh	leftPts	rightPts	outPerc
M.1	−4.25	4.25	364	378	1.48
M.2	−3.19	3.20	1031	1012	4.09
M.3	−2.97	2.96	1299	1287	5.17
M.4	−6.98	6.92	20	30	0.10
M.5	−2.78	2.78	1581	1534	6.23
$M.6_{0.5\%}$	−9.75	9.56	3	4	0.01
$M.6_{5\%}$	−2.26	2.23	2668	2645	10.63

Estimated statistical factors are presented in Table 7. We see significant impact of extreme values influencing the tails. It is expected that outlier labelling using Gaussian approach might not work as wished.

Outlier detection numerical results are sketched in Table 8 and graphical diagram showing detection thresholds is presented in Fig. 28(b). Still Gaussian based MDist-G approach significantly minimizes detected outliers' number, which agrees with expectations. ESD (M.4)

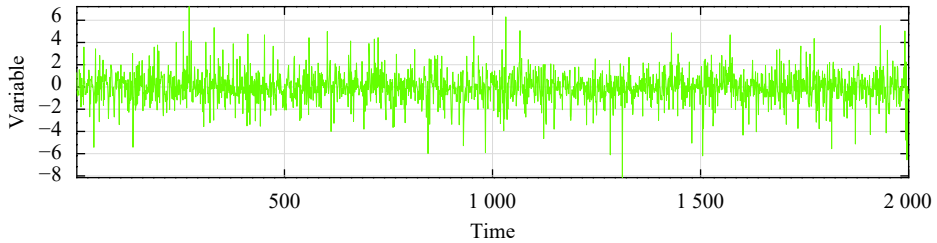
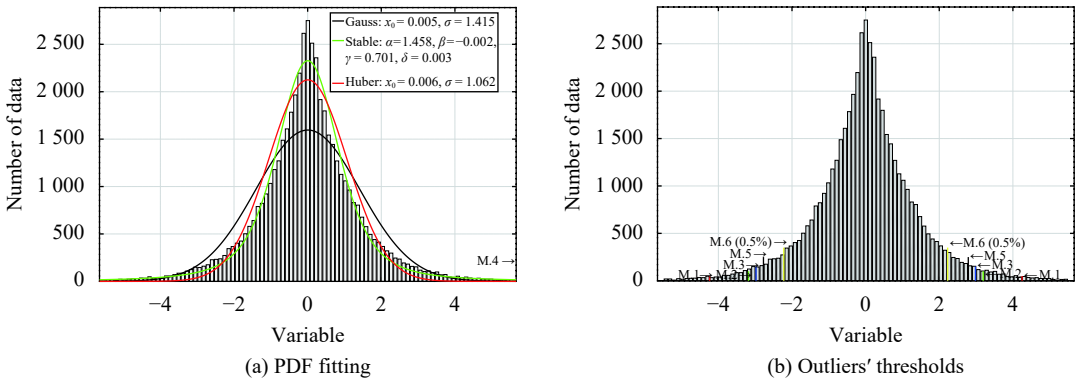


Fig. 25 Normally distributed control error (H3.1)



(a) PDF fitting (b) Outliers' thresholds

Fig. 26 Histogram for Laplace process (H3.1)

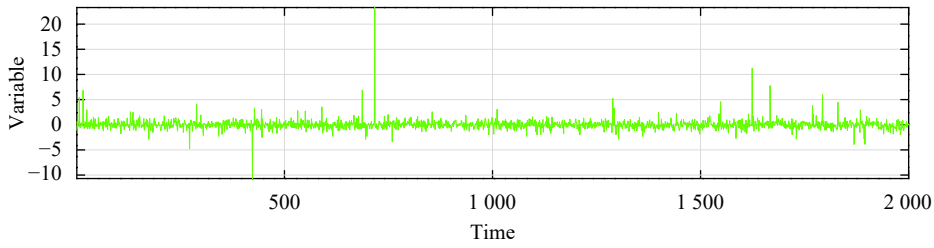
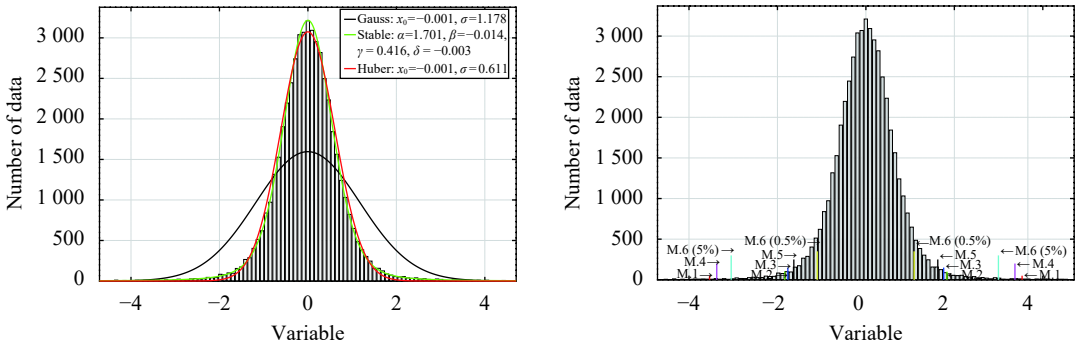


Fig. 27 Normally distributed control error (H3.2)



(a) PDF fitting (b) Outliers′ thresholds

Fig. 28 Histogram for α-stable process (H3.2)

Table 7 Estimated statistical properties of α-stable process

Min	Max	Mean $\bar{x}$	Median	Robust $\bar{x}_{rob}$
−143.83	29.38	−0.001	−0.001	−0.001
σ	σ_{rob}	MAD	Kurtosis	Skewness
1.178	0.611	0.568	4525.56	−36.11
α	β	γ	δ	
1.701	−0.014	0.416	−0.003	

starts to be quite comparable with low confidence IQR-α. IQR-α with large 5% level is extremely relaxed and labels a very high number of observations as outliers. Robust MDist approaches together with IQR give similar results.

At this point, it is important to remember, whether heavy tails are outliers or not, whether they are drawn from the underlying scheme or not[46]. This decision has large consequences. Once we assume that they are in connection with the underlying scheme we should use conservative approaches, like ESD or IQR$\alpha_{5\%}$. If we have any external indications or knowledge that the tails do not originate from the main generating mechanism, we should use other methods, as MDist-rHub, MDist-α or IQR.

Table 8 Detected outliers for α-stable process (H3.2)

	minTh	maxTh	leftPts	rightPts	outPerc
M.1	−5.18	5.20	97	99	0.39
M.2	−1.83	1.83	722	733	2.91
M.3	−1.75	1.75	806	807	3.23
M.4	−3.38	3.38	208	219	0.85
M.5	−1.61	1.61	983	981	3.93
M.6$_{0.5\%}$	−3.13	3.13	238	244	0.96
M.6$_{5\%}$	−1.10	1.10	2598	2615	10.43

3.3.3 α-stable process with varying stability factor (H3.3)

In this section, the analysis using the underlying α-stable stochastic generating process is continued. The effect of tail heaviness is then analyzed. Thereby, all parameters are kept constant (as in scenario H3.2) and characteristics exponent varies from the uncorrelated stochastic process reflected by value $\alpha = 2.0$ up to the one reflecting Cauchy distribution $\alpha = 1.0$ with an increment of 0.1. Thus, eleven simulations run have been performed and analyzed.

Results presented in Fig. 29 are very interesting. We may observe that there are two kinds of detection. The robust detection approaches, i.e., MDist-rHub (M.2), MDist-α (M.3) and IQR (M.5) have a tendency to consider any new observations in the tails (decreasing α increases tails). They take into account the idea that the fat-tailed distribution is not an underlying stochastic generation mechanism.

In contrast, Gaussian MDist and IQR-α approaches consider observations in tails as samples generated by a valid underlying process. Thereby, the number of labelled outliers is kept quite constant. The generalized extreme studentized deviate test exhibits somehow balanced results. Relatively thin tails ($\alpha > 1.5$) increases detection only a little bit, while the number of newly detected outliers saturates for small stability factors ($\alpha < 1.5$).

3.3.4 α-stable process with varying skewness (H3.4)

In this section, the analysis using an underlying α-

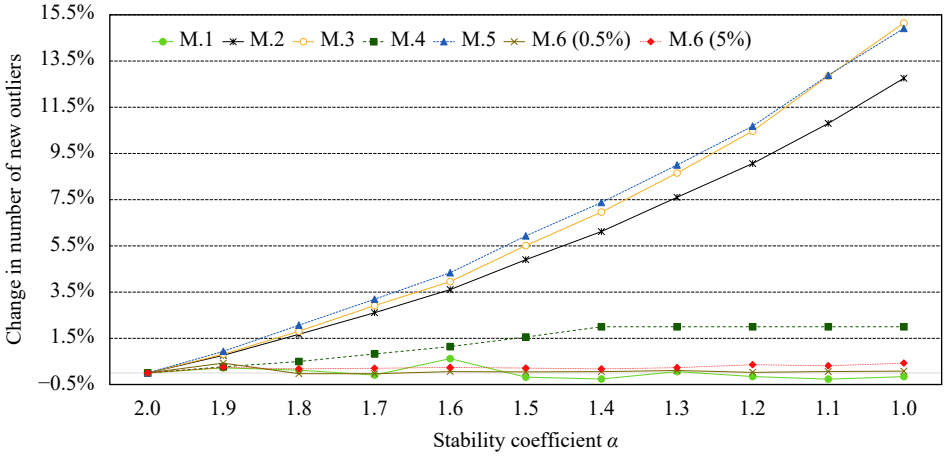


Fig. 29 Impact of varying stability α on new detected outliers

stable stochastic generating process is continued. The effect of asymmetric behavior is under consideration. Thereby, all parameters are kept constant (as in scenario H3.2), while the skewness coefficient changes from the left-skewed $\beta = -1.0$ up to the right-skewed $\beta = +1.0$ with an increment of 0.2. Simulation results are summarized in the relations depicted in Fig. 30(a). These simulations show that methods distinguish between different properties and an interpretation of the underlying generation mechanism. ESD is robust to skewness. We may interpret it in such a way that skewed observations come from valid underlying processes. Similar behavior is exhibited for the IQR-$\alpha_{0.5\%}$ approach. The IQR-$\alpha_{5\%}$ method exhibits opposite properties. It labels skewed observations as outliers.

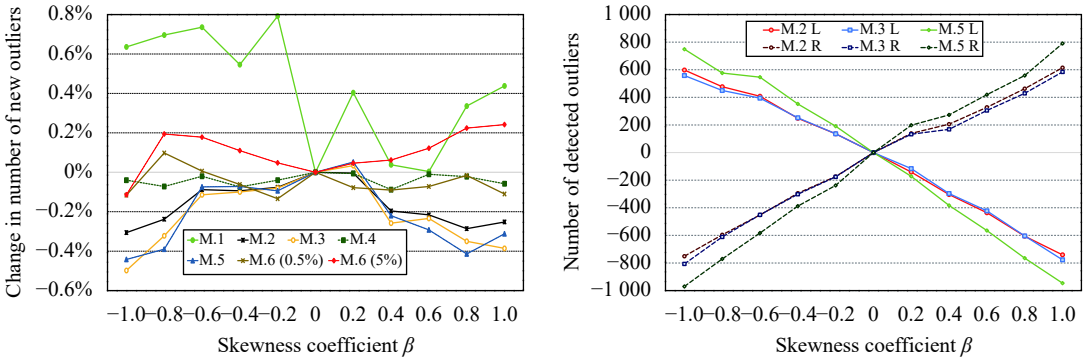
Robust methods give rather unintuitive conclusions. Left-sided or right-sided skewness decrease the number of new labelled outliers. It is caused by the asymmetry effect being overlapped with the tails. Fig. 30(b) presents separate relationships between left-sided and right-sided numbers of the detected outliers versus the skewness factor β.

We see that each robust method performs asymmetrically. When the time series is left skewed, the increase in number of detected outliers in the left side is smaller than the number of lost (not detected any more) outliers in a right-sided tail, and opposite. Skewness makes the tail thinner from one side than the opposite one. This observation shows that outliers analysis should consider both sides. The observations in tails are considered as outliers (labelled and truncated), while skewed observations fall under the label of inliers.

3.4 Simulation study conclusions

The first conclusion is relatively simple, as there is no single universal outlier detection algorithm. We may group the methods according to their robustness. MDistrHub, MDist-α and IQR are the robust ones, as MDist is very sensitive to the outliers. The generalized extreme studentized deviate test ESD performs very conservatively and detects only extreme values. Interquartile method based on the α-stable distribution (IQR-α) strongly depends on the chosen value of the confidence level. Actually, this effect requires further analysis. Especially, the impact of difference confidence values need to be investigated. Values suggested in the literature seem to be case dependent, and their generalization to any case is questionable. We may identify four main issues that have appeared during the simulation study:

1) Interpretation of tails, i.e., generated by the main mechanism or a contamination



(a) New detected outliers (b) Detected outliers in both sides

Fig. 30 Impact of skewness β on detected outliers (H3.4)

2) Asymmetry in data
3) The most extreme samples
4) Erroneous observations.

The tail interpretation requires further process knowledge, as simple applications of some methods may lead towards incorrect interpretations, i.e., inliers might be labelled as outliers or they could not be detected. Asymmetry poses a serious challenge, because the user should know, whether the observed skewness represents contamination or the underlying generating mechanism. In the case of skewed data, it is suggested to compare observations between IQR and robust MDist-rHub or MDist-α.

Constant errors that appear in data, although they are the simplest in manual detection, can pose big problem for the analysis. First of all, one has to remember that such extreme values may bias estimation of location and scale. In particular, the location estimator has to be reviewed and it is suggested to apply robust ones. In such cases, the α-stable scale factor should be considered as it seems to be the most robust. Once a researcher focuses on the extreme values, the ESD test is suggested.

4 Industrial data validation

Industrial validation is based on real process control error time series data. Data length is $N = 65\,521$ recorded with sampling interval of 10 s. Examples of data are presented in Fig. 31 on time trends showing the first 1 000 points. Data are selected in such a way that they reflect different properties. Var2 data reflects control error of a relatively good loop, with a shape similar to the normal bell shape ($\alpha = 2.0$), with only a few potential outliers. Var1 has fatter tails ($\alpha = 1.76$), while Var3 reflects even heavier ones ($\alpha = 1.69$). The last signal Var4 has the most heavy tails with stability index $\alpha = 1.46$.

The first step of analysis requires simple statistical validation of data, i.e., quantitative evaluation of the basic statistics and qualitative review of the histograms. Statistical factors can be found in Table 9 and validation of the histograms are shown in Fig. 32. Table 9 includes simple statistics like min and max, Gaussian mean $\bar{x}$, standard deviation σ, kurtosis and skewness, two robust location estimates as median and M-estimator using log ψ function denoted as median and $\bar{x}_{rob}$, and two robust scale estimators, i.e., MAD and M-estimator using log ψ function denoted as σ_{rob}. Additionally, there are attached factors of the fitted α-stable distribution estimated with percentile method, denoted as α, β, γ and δ.

We see that data are not skewed significantly. The skewness factor from the α-stable distribution (β^{stab}) shows values rather close to zero with Var1 being the most skewed (see Fig. 32). Three of the datasets exhibit tails, i.e., Var1, Var3 and Var4 characterized by the fattest tails. Var2 time series is the closest to the Gaussian density function, although detailed normality hypothesis tested with three tests (Kolmogorov-Smirnov, Lilliefors and Shapiro-Wilk) rejects time series normality. Var3 and Var4 histograms are symmetrical and very well fitted with stable distribution. It also shows good effect of using robust scale estimators which effectively neglect tails.

Table 9 Statistic factors for industrial control error data

	Var1	Var2	Var3	Var4
Min	−6.71	−3.95	−6.75	−3.63
Max	7.15	7.06	6.35	4.47
$\bar{x}$	0.005	−0.006	0.001	0.005
Median	−0.014	0.003	0.004	−0.003
σ	0.705	1.073	0.411	0.324
Kurtosis	18.63	2.54	66.17	48.30
Skewness	0.086	−0.005	−1.044	1.828
$\bar{x}_{rob}$	0.000	−0.005	0.003	−0.001
σ_{rob}	0.540	1.166	0.193	0.137
MAD	0.478	0.886	0.203	0.157
α	1.762	2.000	1.686	1.462
β	0.39	−0.11	−0.06	0.10
γ	0.367	0.835	0.131	0.090
δ	0.022	0.003	0.001	0.005

Initial observations of the industrial control errors and their statistical properties show that the signals exhibit various properties. Their investigation should give fruitful comments about applied outliers detection methods. Detection results are grouped and presented in Tables 10 and 11. Table 10 shows thresholds determined by each method together with the number of labelled outliers. Table 11 presents the percentage of points labelled as outliers. Visual representation of the methods performance can be reviewed with the thresholds put on the histogram plots shown in Fig. 33.

We see that all the methods label the least number of outliers for data exhibiting almost normal properties, i.e., for Var2 dataset in the considered case. Actually following general Gaussian assumptions, it is expected that almost zero outliers should be detected in this case as the underlying generation mechanism is normal. ESD is still the most conservative, while IQR-$\alpha_{5\%}$ is the most relaxed. It is also interesting to notice that Var1 being not well fitted by any of the PDFs delivers quite a lot of outliers, derived with each method. More right-sided outliers is the most tempting hypothesis. But the question is whether the skewness originates from the baseline underlying data generation mechanism or from the anomalies, i.e., they are generated by the same mechanism as the outliers. Thereby in the first case, the robust and non-robust methods should indicate opposite directions, while the common one would be suggested in the other one. In our case, we get the second observation: asymmetry originates from the baseline statistical process. Effect of tails is compliant with simulations. We notice a difference

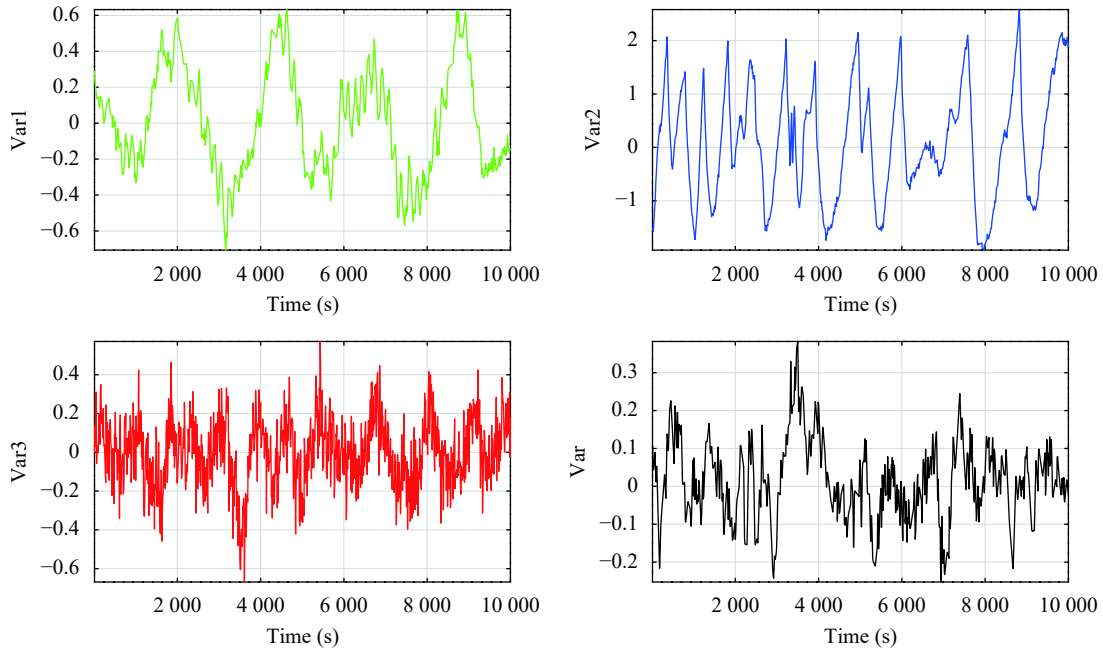
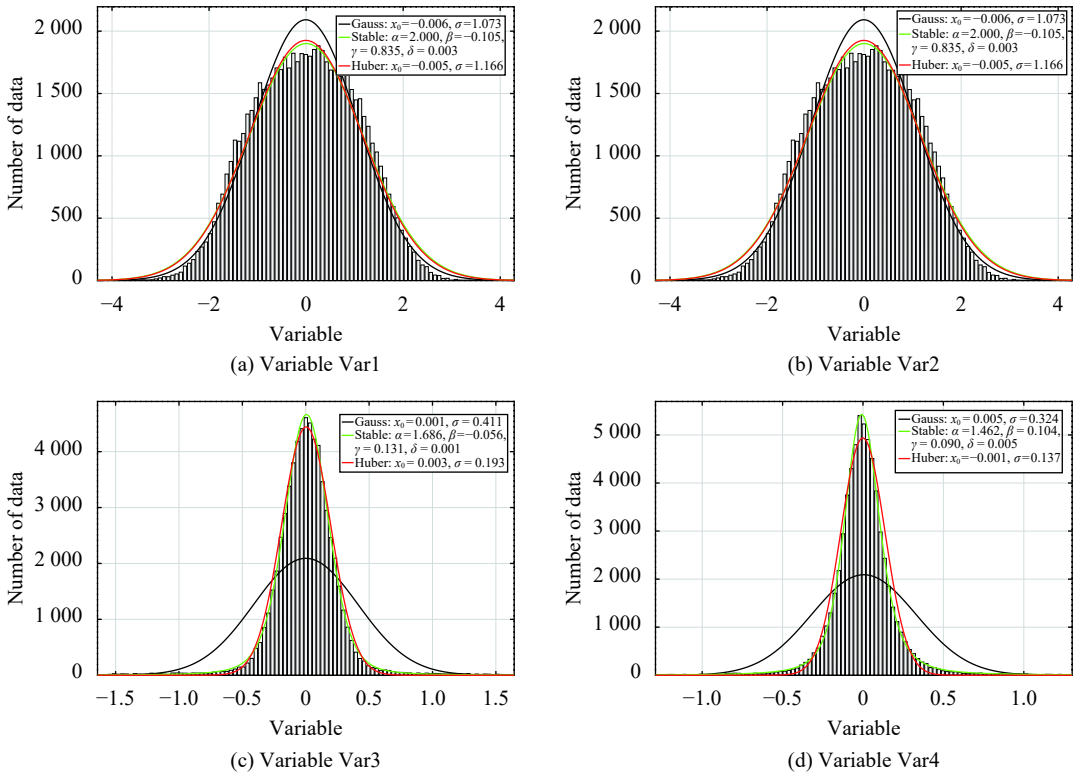


Fig. 31 Industrial data histograms



(a) Variable Var1 (b) Variable Var2

(c) Variable Var3 (d) Variable Var4

Fig. 32 Histograms for datasets originating from industry

between MDist-rHub and MDist-α. The first one considers tails as outliers, while MDist-α considers them as inliers.

5 Conclusions and further research

The presented work consists of extended simulations of different outlier labelling approaches and their application to the CPA task. Furthermore, three novel modifications into already existing methods are proposed and validated on real industrial data. Properties of proposed algorithms are compared with classical ones.

Methods are applied to the engineering task of outlier detection in control engineering and the assessment of

Table 10 Outlier labelling: minimum (minT) and maximum (maxT) thresholds and number of outliers minN and maxN

Methods	Error1				Error2				Error3				Error4			
	minT	minN	maxT	maxN	minT	minN	maxT	maxN	minT	minN	maxT	maxN	minT	minN	maxT	maxN
M.1	−2.110	457	2.119	461	−3.226	25	3.214	30	−1.232	628	1.234	585	−0.967	630	0.977	711
M.2	−1.620	743	1.619	758	−3.502	9	3.492	22	−0.577	1490	0.584	1433	−0.412	1649	0.411	1972
M.3	−1.534	817	1.578	793	−3.535	8	3.541	22	−5.871	24	5.873	12	−5.867	0	5.877	0
M.4	−2.944	239	2.965	213	6.165	3	7.065	1	−1.397	520	1.408	480	−1.237	479	1.237	521
M.5	−1.444	921	1.431	1038	−3.231	24	3.222	30	−0.511	1742	0.517	1688	−0.360	2043	0.357	2492
$M.6_{0.5\%}$	−1.836	592	2.697	234	−3.039	47	3.045	42	−1.018	814	0.963	803	−1.108	550	1.275	502
$M.6_{5\%}$	−0.885	3498	0.995	3397	−1.940	1888	1.945	1788	−0.350	3382	0.348	3439	−0.273	3436	0.298	3414

Table 11 Percentage of data labelled as outliers

Methods	Error1	Error2	Error3	Error4
M.1	1.40	0.08	1.85	2.05
M.2	2.29	0.05	4.46	5.53
M.3	2.46	0.05	0.05	0.00
M.4	0.69	0.01	1.53	1.53
M.5	2.99	0.08	5.23	6.92
$M.6_{0.5\%}$	1.26	0.14	2.47	1.61
$M.6_{5\%}$	10.52	5.61	10.41	10.45

control loops quality. The analysis focuses on four main aspects that are frequent on those applications: interpretation of tails, asymmetry, extreme samples and artificial erroneous observations.

These effects reflect common effects in control systems. Tails are often generated by non-linearities and signal limitations, interactions with internal and external disturbances and an influence of other stochastic processes which may be internally fat-tailed. Asymmetry often originates from non-linear process properties, active constraints or improper operating point selection. Extreme samples may have different origins: system errors, process breakdowns, human interventions or an impact of persistent disturbances, like weather conditions. Data errors are often caused by a system (communication and storage) or a human (somebody may cut a cable).

There is one main observation. There is no single ideal method and it is suggested to use a combination of different methods. This selection can consists of: classical Gaussian MDist Z-scores method, its robust MDist-rHub variant based on location and scale M-estimators, interquartile range (IQR) method and MDist-α. Balancing these methods should enable reasonable detection in connection with external knowledge about the process. Such information is crucial, as any hint about existing mechanisms, disturbances and interconnections is inevitable.

The generalized extreme studentized deviate test exhibits otherwise and should be used with a different goal, as it focuses on labelling of the most extreme observations.

Finally, some discussions are required about IQR-α. This method is potentially very powerful, but an impact and interpretation of the confidence interval is required and should be further investigated. Current literature is not comprehensive and some suggestions of the threshold selection would be very helpful.

The practical application procedure should start with the visual inspection of time trends that should be followed by simple statistical analysis. These tests ought to include evaluation of normal distribution parameters (min, max, mean, standard deviation) together with robust estimators (median, MAD, M-estimators). Histograms should be further reviewed accompanied with fitting of potential probabilistic density functions (Gaussian, robust, stable, Laplace). Observations brought up from this analysis should allow to formulate subsequent actions:

1) The ESD and IQR-α with a very low confidence coefficient methods detect and remove only the most extreme values.

2) Erroneous observations can be manually removed quite easily with the support of the histogram plot.

3) Once the time series exhibits close to the normal process, MDist-G is enough.

4) IQR method combined robust estimators helps for skewed data.

5) Heavy tailed data require the coordinated use of IQR, MDist-rHub and MDist-α.

Unluckily, obtained results cannot be considered as a closed subject. There is still a need to understand the tails and the mechanisms behind them. This subject is rather rare in control engineering research. Therefore, there is a need to investigate approaches and methods proposed in other scientific contexts, like economy or statistics. The authors intentionally excluded from the analysis data-mining approaches as they often require models. In non-linear, complex and cross-correlated environments, biased by human interventions, this approach seems to be practically less promising.

Successive research is required for multivariate analysis. For instance, robust multi-variable relations analysis will help to model adaptive equipment static curves,

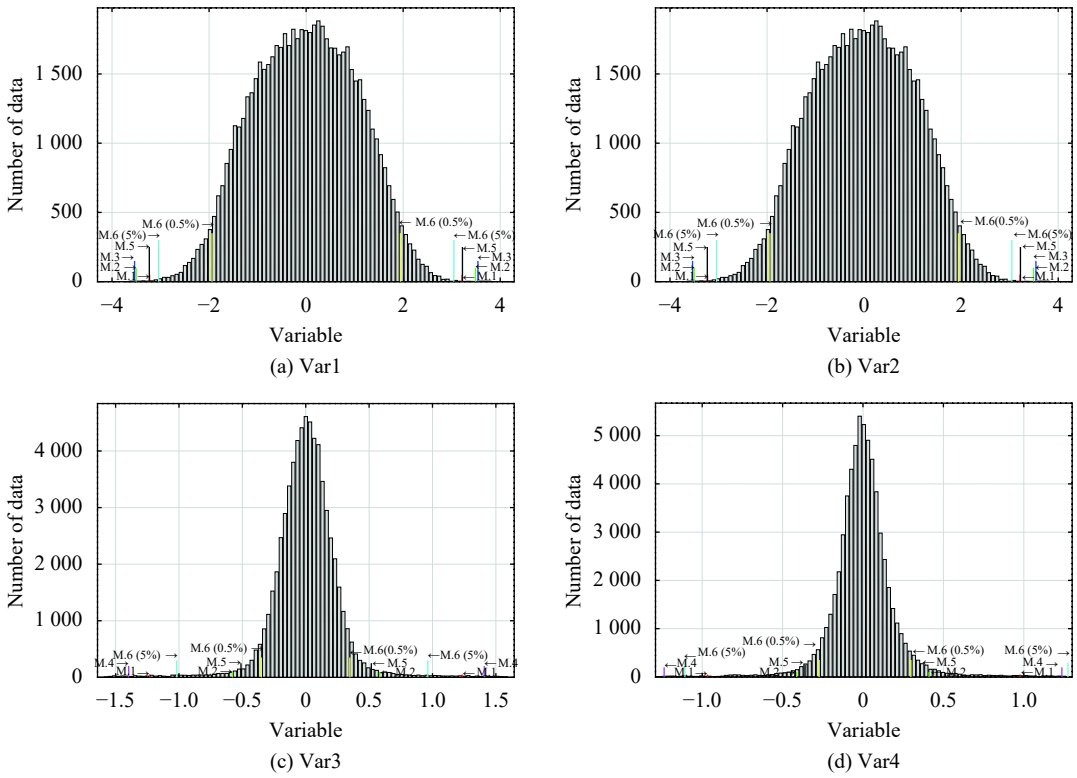


Fig. 33 Industrial data histograms

which further results in better maintenance policies. It would be useful for the life-cycle analysis, predictive maintenance and fault detection. Finally, an aspect of cyber security of control systems must be seriously considered through anomaly detection.

References

[1] W. J. Dixon. Analysis of extreme values. *The Annals of Mathematical Statistics*, vol. 21, no. 4, pp. 488–506, 1950. DOI: 10.1214/aoms/1177729747.

[2] H. Wainer. Robust statistics: A survey and some prescriptions. *Journal of Educational Statistics*, vol. 1, no. 4, pp. 285–312, 1976. DOI: 10.3102/10769986001004285.

[3] D. M. Hawkins. *Identification of Outliers*, Dordrecht, The Netherlands: Springer, 1980. DOI: 10.1007/978-94-015-3994-4.

[4] R. A. Johnson, D. W. Wirchern. *Applied Multivariate Statistical Analysis*, 3rd ed., Englewood Cliffs, USA: Prentice-Hall, 1992.

[5] V. Barnett, T. Lewis. *Outliers in Statistical Data*, 3rd ed., Chichester, UK: Wiley, 1994.

[6] J. R. Xue, J. W. Fang, P. Zhang. A survey of scene understanding by event reasoning in autonomous driving. *International Journal of Automation and Computing*, vol. 15, no. 3, pp. 249–266, 2018. DOI: 10.1007/s11633-018-1126-y.

[7] J. W. Osborne, A. Overbay. The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, vol. 9, no. 9, Article number 6, 2004. DOI: 10.7275/qf 69-7k43.

[8] N. N. Taleb. Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications. USA: STEM Academic Press, 2020.

[9] P. J. Rousseeuw, A. M. Leroy. *Robust Regression and Outlier Detection*, New York, USA: John Wiley & Sons, 1987.

[10] I. Ben-Gal. Outlier detection. *Data Mining and Knowledge Discovery Handbook*, O. Maimon, L. Rokach, Eds., Boston, USA: Springer, pp. 131–146, 2005. DOI: 10.1007/0-387-25465-X_7.

[11] B. Iglewicz, D. C. Hoaglin. *How to Detect and Handle Outliers*, Milwaukee, USA: ASQ Quality Press, 1993.

[12] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, H. E. Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, vol. 316, no. 1–4, pp. 87–114, 2002. DOI: 10.1016/S0378-4371(02)01383-3.

[13] J. Barunik, T. Aste, T. Di Matteo, R. P. Liu. Understanding the source of multifractality in financial markets. *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 17, pp. 4234–4251, 2012. DOI: 10.1016/j.physa.2012.03.037.

[14] B. Mandelbrot, R. L. Hudson. *The Misbehavior of Markets: A Fractal View of Financial Turbulence*, New York, USA: Basic Books, 2005.

[15] H. P. Kriegel, P. Kröger, A. Zimek. Outlier detection techniques. In *Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Washington, USA, 2010.

[16] K. G. Mehrotra, C. K. Mohan, H. M. Huang. *Anomaly Detection Principles and Algorithms*, Cham, Germany: Springer, 2017. DOI: 10.1007/978-3-319-67526-8.

[17] B. Peirce. Criterion for the rejection of doubtful observations. *Astronomical Journal*, vol. 2, no. 45, pp. 161–163,

1852. DOI: 10.1086/100259.

[18] J. Irwin. On a criterion for the rejection of outlying observations. *Biometrika*, vol. 17, no. 3–4, pp. 238–250, 1925. DOI: 10.1093/biomet/17.3-4.238.

[19] E. S. Pearson, C. C. Sekar. The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika*, vol. 28, no. 3–4, pp. 308–320, 1936. DOI: 10.1093/biomet/28.3-4.308.

[20] F. E. Grubbs. Sample criteria for testing outlying observations. *The Annals of Mathematical Statistics*, vol. 21, no. 1, pp. 27–58, 1950. DOI: 10.1214/aoms/1177729885.

[21] N. A. Heckert, J. J. Filliben, C. M. Croarkin, B. Hembree, W. F. Guthrie, P. Tobias, J. Prinz. NIST/SEMATECH e-Handbook of Statistical Methods, 2012, [Online], Available: http://www.itl.nist.gov/div898/handbook/, February 08, 2020.

[22] F. Rosado. Outliers: The strength of minors. *New Advances in Statistical Modeling and Applications*, A. Pacheco, R. Santos, M. D. R. Oliveira, C. D. Paulino, Eds., Cham, Germany: Springer, 2014.

[23] D. L. Whaley III. The Interquartile Range: Theory and Estimation, Master dissertation, Faculty of the Department of Mathematics, East Tennessee State University, USA, 2005.

[24] G. L. Tietjen, R. H. Moore. Some grubbs-type statistics for the detection of several outliers. *Technometrics*, vol. 14, no. 3, pp. 583–597, 1972. DOI: 10.1080/00401706.1972.10488948.

[25] M. Hubert, M. Debruyne. Minimum covariance determinant. *WIREs Computational Statistics*, vol. 2, no. 1, pp. 36–43, 2010. DOI: 10.1002/wics.61.

[26] B. Rosner. Percentage points for a generalized ESD many-outlier procedure. *Technometrics*, vol. 25, no. 2, pp. 165–172, 1983. DOI: 10.1080/00401706.1983.10487848.

[27] R. Thompson. A note on restricted maximum likelihood estimation with an alternative outlier model. *Journal of the Royal Statistical Society*: *Series B* (*Methodological*), vol. 47, no. 1, pp. 53–55, 1985. DOI: 10.1111/j.2517-6161.1985.tb01329.x.

[28] P. J. Huber Hoboker, E. M. Ronchetti. *Robust Statistics*, 2nd ed., Hoboken, USA: Wiley, 2009. DOI: 10.1002/9780470434697.

[29] R. K. Pearson. *Mining Imperfect Data: Dealing with Contamination and Incomplete Records*, Philadelphia, USA: SIAM, 2005.

[30] N. N. Taleb. *Real-world Statistical Consequences of Fat Tails: Papers and Commentary*, UK:STEM Academic Press, 2018.

[31] P. D. Domański. Statistical measures. *Control Performance Assessment: Theoretical Analyses and Industrial Practice*, P. D. Domański, Ed., Cham, Germany: Springer, pp. 53–74, 2020. DOI: 10.1007/978-3-030-23593-2_4.

[32] P. D. Domański. Non-Gaussian properties of the real industrial control error in SISO loops. In *Proceedings of the 19th International Conference on System Theory, Control and Computing*, IEEE, Cheile Gradistei, Romania, pp. 877–882, 2015. DOI: 10.1109/ICSTCC.2015.7321405.

[33] K. Malik, H. Sadawarti, G. S. Kalra. Comparative analysis of outlier detection techniques. *International Journal of Computer Applications*, vol. 97, no. 8, pp. 12–21, 2014. DOI: 10.5120/17026-7318.

[34] S. A. Shaikh, H. Kitagawa. Top-k outlier detection from uncertain data. *International Journal of Automation and Computing*, vol. 11, no. 2, pp. 128–142, 2014. DOI: 10.1007/s11633-014-0775-8.

[35] Z. G. Ding, D. J. Du, M. R. Fei. An isolation principle based distributed anomaly detection method in wireless sensor networks. *International Journal of Automation and Computing*, vol. 12, no. 4, pp. 402–412, 2015. DOI: 10.1007/s11633-014-0847-9.

[36] S. Banerjee, T. Chattopadhyay, U. Garain. A wide learning approach for interpretable feature recommendation for 1-d sensor data in iot analytics. *International Journal of Automation and Computing*, vol. 16, no. 6, pp. 800–811, 2019. DOI: 10.1007/s11633-019-1185-8.

[37] N. N. R. Ranga Suri, N. Murty M, G. Athithan. *Outlier Detection: Techniques and Applications: A Data Mining Perspective*, Cham, Germany: Springer, 2019. DOI: 10.1007/978-3-030-05127-3.

[38] A. Zimek, P. Filzmoser. There and back again: Outlier detection between statistical reasoning and data mining algorithms. *WIRES Data Mining and Knowledge Discovery*, vol. 8, no. 6, Article number e1280, 2018. DOI: 10.1002/widm.1280.

[39] P. J. Rousseeuw, M. Hubert. Anomaly detection by robust statistics. *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 2, Article number e1236, 2018. DOI: 10.1002/widm.1236.

[40] M. Templ, J. Gussenbauer, P. Filzmoser. Evaluation of robust outlier detection methods for zero-inflated complex data. *Journal of Applied Statistics*, vol. 47, no. 7, pp. 1144–1167, 2020. DOI: 10.1080/02664763.2019.1671961.

[41] M. P. J. Van Der Loo. Distribution based Outlier Detection in Univariate Data. Technical Report Discussion Paper 10003, Statistics Netherlands, The Hague/Heerlen, Netherlands, 2010.

[42] G. Barbato, E. M. Barini, G. Genta, R. Levi. Features and performance of some outlier detection methods. *Journal of Applied Statistics*, vol. 38, no. 10, pp. 2133–2149, 2011. DOI: 10.1080/02664763.2010.545119.

[43] M. Gupta, J. Gao, C. Aggarwal, J. W. Han. *Outlier Detection for Temporal Data*, San Rafael, USA: Morgan & Claypool Publishers, 2014. DOI: 10.2200/S00573ED1V01Y201403DMK008.

[44] P. D. Domański. Statistical measures for proportional–integral–derivative control quality: Simulations and industrial data. *Proceedings of the Institution of Mechanical Engineers*, *Part I*: *Journal of Systems and Control Engineering*, vol. 232, no. 4, pp. 428–441, 2018. DOI: 10.1177/0959651817754034.

[45] P. D. Domański, S. Golonka, P. M. Marusak, B. Moszowski. Robust and asymmetric assessment of the benefits from improved control – industrial validation. *IFAC-PapersOnLine*, vol. 51, no. 18, pp. 815–820, 2018. DOI: 10.1016/j.ifacol.2018.09.260.

[46] L. B. Klebanov. Big outliers versus heavy tails: What to use? https://arxiv.org/abs/1611.05410.

[47] C. Croux, C. Dehon. Robust estimation of location and scale. *Encyclopedia of Environmetrics*, A. H. El-Shaarawi, W. W. Piegorsch, Eds., Hoboken, USA: Wiley, 2013. DOI: 10.1002/9780470057339.vnn093.

[48] S. Verboven, M. Hubert. LIBRA: A MATLAB library for

robust analysis. *Chemometrics and Intelligent Laboratory Systems*, vol. 75, no. 2, pp. 127–136, 2005. DOI: 10.1016/j.chemolab.2004.06.003.

[49] J. H. McCulloch. Simple consistent estimators of stable distribution parameters. *Communications in Statistics – Simulation and Computation*, vol. 15, no. 4, pp. 1109–1136, 1986. DOI: 10.1080/03610918608812563.

[50] I. A. Koutrouvelis. Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, vol. 75, no. 372, pp. 918–928, 1980. DOI: 10.1080/01621459.1980.10477573.

[51] E. E. Kuruoglu. Density parameter estimation of skewed α-stable distributions. *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2192–2201, 2001. DOI: 10.1109/78.950775.

[52] S. Borak, A. Misiorek, R. Weron. Models for heavy-tailed asset returns. *Statistical Tools for Finance and Insurance*, 2nd ed., P. Cizek, W. K. Härdle, R. Weron, Eds., Berlin, Heidelberg, Germany: Springer, pp. 21–55, 2011. DOI: 10.1007/978-3-642-18062-0_1.

[53] A. Alfons, M. Templ, P. Filzmoser. Robust estimation of economic indicators from survey samples based on pareto tail modelling. *Journal of the Royal Statistical Society: Series C*, vol. 62, no. 2, pp. 271–286, 2013. DOI: 10.1111/j.1467-9876.2012.01063.x.

[54] J. Danielsson, L. M. Ergun, L. De Haan, C. G. De Vries. Tail Index Estimation: Quantile Driven Threshold Selection, Bank of Canada Staff Working Paper 2019–28, Bank of Canada.

[55] P. D. Domański. *Control Performance Assessment: Theoretical Analyses and Industrial Practice*, Cham, Germany: Springer, 2020. DOI: 10.1007/978-3-030-23593-2.

[56] M. C. Bryson. Heavy-tailed distributions: Properties and tests. *Technometrics*, vol. 16, no. 1, pp. 61–68, 1974. DOI: 10.1080/00401706.1974.10489150.

[57] L. B. Klebanov, I. Volchenkova. Outliers and the ostensibly heavy tails. https://arxiv.org.abs/1807.08715vl.

[58] G. Marsaglia, W. W. Tsang. A simple method for generating gamma variables. *ACM Transactions on Mathematical Software*, vol. 26, no. 3, pp. 363–372, 2000. DOI: 10.1145/358407.358414.

[59] N. L. Johnson, S. Kotz, N. Balakrishnan. *Continuous Univariate Distributions*, 2nd ed., New York, USA: Wiley, 1995.

Paweł D. Domański received the M.Sc. degree, Ph.D. degree and D.Sc. degree in control engineering from Faculty of Electronics and Information Technology, Warsaw University of Technology, Poland in 1967, 1991 and 1996, respectively. He works in the Institute of Control and Computational Engineering, Warsaw University of Technology, Poland from 1991. He is the author of one book and more than 100 publications. Apart from scientific research, he participated in dozens of industrial implementations of advanced process control and optimization in power and chemical industries all over the world.

His research interests include industrial advanced process control applications, control performance quality assessment and optimization.

E-mail: p.domanski@ia.pw.edu.pl
ORCID iD: 0000-0003-4053-3330