An investigation of Outlier Detection Procedures for CMM Measurement Data

Petr Chelishchev[1], Aleksandr Popov[2] and Knut Sørby[1]

[1]*Department of Mechanical and Industrial Engineering, NTNU, NO-7491 Trondheim, Norway*
[2]*Department of Mathematics, Baltic State Technical University, St. Petersburg, Russia*

Abstract. The paper analyses methods for outlier detection in dimensional measurement. The cross sections of an internal cylinder were inspected by CMM (coordinate measuring machine), and received data sets were employed for further investigation. The efficiency of Rosner's and Grubbs' methods for excluding outliers from the measuring data had been estimated. The method of Rosner had been defined as the most effective for this case study. The simulation results were confirmed by experimental verification.

1 Introduction

The purpose of this work is to analyze the efficiency of outlier test procedures for particular type of data sets received from inspection with CMM (coordinate measuring machine). The following inspection conditions are considered:

- Varying sample size of measurements;
- Unknown number of outliers presented in the sample;
- A spectrum of different distributions of original data sets with unknown dispersion.

In the CMM inspection of the geometrical characteristics of components, the outliers are not necessarily incorrect measurements. The existence of an outlier could indicate that a further investigation of manufacturing processes, measurements procedure, or data analysis methods themselves is required.

The estimation of different statistical parameters (e.g. sample standard deviation, sample mean and so on) may be affected by outlier presence in the measuring data. As a result, it can lead to the invalid estimation of a confidence interval and inflate the random uncertainty estimates as well, thus a good component may be erroneously rejected. That especially yields the particular case, when contact fit methods such as MIC (maximum inscribed circle) or MZ (minimum-zone) are utilized, which are based on the most extreme points and hence very sensitive to outliers.

Outliers are extreme observations, which stay apart from the majority of other measurements. In a simple case, when only one outlier is presented, its inconsistence can be easily observed with respect to the rest of the data. However, when a group of outliers is present, it is difficult to detect them because of the masking effect, which will be described below. At the same time, an incorrect assumption about the original data distribution may lead to confusion of valid observations with outliers. According to ISO 16269-4 [1], the main causes for outliers are the following:

- a measurement or recording error (imprecise or/and incorrect);
- a distribution contamination (one or more contaminating distributions);
- an incorrect distributional assumption;
- rare observations (extreme observations from heavy-tailed original distribution).

In the particular case of measuring in manufacturing conditions, a contamination of a part surface is a frequent cause of outliers, even after attempt of surface cleaning.

In addition, a masking and a swamping effect can occur during the data analysis with parametric statistical test. The masking effect can happen when too few outliers are specified in the outlier detection procedure. Then the test performance can be influenced by the other outliers and as result, no outliers will be detected. On the other hand, if too many outliers are specified in the parameters of outlier test, then some valid observations can be incorrectly labeled as outliers, which is the so-called swamping effect. Therefore, to make a correct decision whether suspected observations are outliers or not can be a complicated task.

2 Methods

2.1 Graphical methods

The first step, before any analytical outlier detection algorithms are applied it is a visual analysis of measurement data. There are a number of graphical methods available such as histogram, scatter diagram, dot

plot and so on [2]. The pox plot become a very popular descriptive tool to reveal the most suspected measurements [3]. In fact, the box plot is a hybrid based on both a model and the graphical method. The graphical interpretation of data helps to choose the most appropriate analytical algorithm i.e. identify whether a single outlier or a group of outliers are present in order to prevent an influence of the masking or swamping effects, as described in previous section.

There are six data sets (with 475 observations in each) comparing with each other on Fig.1. The data sets denoted by A1, B1, C1 represent the first measurement results with outliers. After the contamination was physically removed from the workpiece surface, the measurements at the same sections and with the same point distribution were repeated with CMM. These data sets are denoted as A2, B2, C2 on Fig.1. The box plot gives a good demonstration of the influence on statistical parameters such as the sample mean, median, skewness, data spread, IQR (interquartile range). The relative displacement of these parameters can be easily observed.

The lower and upper fences (lower and upper outlier cut-off) is expressed by following:

$$LF = q_1 - w(q_3 - q_1) \qquad (1.1)$$

$$UF = q_3 + w(q_3 - q_1), \qquad (1.2)$$

where q_1, q_3 are the first (lower) and the second (upper) quartiles of data sample, and w is the significant factor [4]. The extreme points, which are outside of these fences, are indicated by red dots. For example, with significant factor $w = 1.5$ red dots can be classified as suspected outliers (Fig.1, left), with $w = 3$ as extreme outliers (Fig.1, right) [5]. The vertical box represents IQR of the data, the different between the lower and upper quartiles. Thus, we do not have extreme outliers in the studied case (Fig. 1, right), but there are some suspected observations in all sections. The section A represents the case with multiple potential outliers, section B with two, and section C with a single potential outlier (Fig.1, left). From now on, we can precede with selection of the most suitable outlier detection analytical algorithm for our particular problem.

2.2 Analytical algorithms

There are many outlier methods proposed in the last decade [6]. The difference between them can shortly formulated by following.

- What a sample size can method be applied for (only for small, only for large, or both)?
- How strict is a requirement to the distribution of data set?
- Can method be exploited whether for a single or multiple outliers?
- In a case of multiple outliers method, it is either necessary to provide exact number of outliers or only an upper amount.

Two suitable outlier methods according to these conditions are considered in this research. These methods are Grubbs and Rosner/GESD (Generalized Extreme Studentized Deviate) tests, which are recommended by

ISO [1, 7]. Both methods are based on an estimation of a distance deviation from the sample mean with assumption about an approximately normal distribution. The strictness of this normality assumption is examined in this paper.
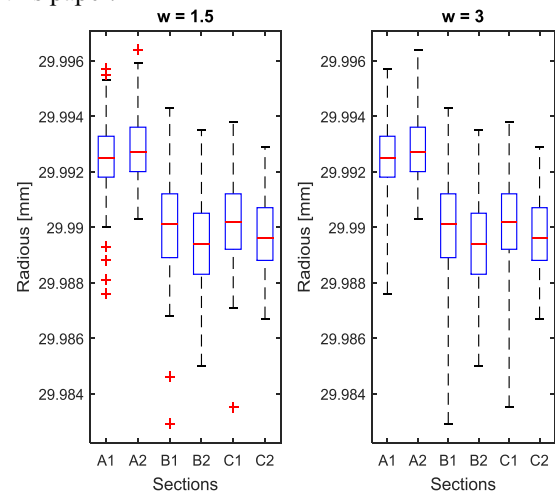

Figure 1. Boxplot of CMM measurements with and without suspected observations.

2.2.1 Grubbs method

The Grubbs method is used to determine a single outlier in a normally distributed data set and can be utilized as sequentially outlier detection procedure for multiple outliers [7, 8]. It tests two types of hypothesis: null hypothesis H_0 – no outliers in the sample, alternative H_1 – the sample has a one outlier. The test statistic for two-sided case computed by (2):

$$G = \frac{\max |\rho_i - \bar{\rho}|}{s} \qquad (2)$$

where $\bar{\rho}$ is the sample mean and s is the sample standard deviation. The G statistic shows how many standard deviations are in an absolute distance of an individual observation from the sample mean. The null hypothesis must be rejected with a significance level α, if the following condition is met:

$$G > \frac{n-1}{\sqrt{n}} \sqrt{\frac{\left(t_{\left(\frac{\alpha}{2n}, n-2\right)}\right)^2}{n-2 + \left(t_{\left(\frac{\alpha}{2n}, n-2\right)}\right)^2}} \qquad (3)$$

where $t_{\left(\frac{\alpha}{2n}, n-2\right)}$ - the Student's quantile given at probability $\frac{\alpha}{2n}$ and $n-2$ degrees of freedom in the data set with number of observations n. One of the weaknesses of the method is the influence by the masking effect.
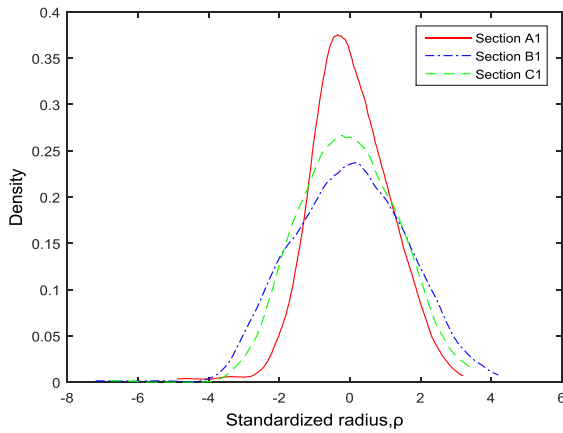
Figure 2. Kernel estimates $\hat{f}_{A1}(\rho), \hat{f}_{B1}(\rho), \hat{f}_{C1}(\rho)$ based on 475 measuring points with outliers.
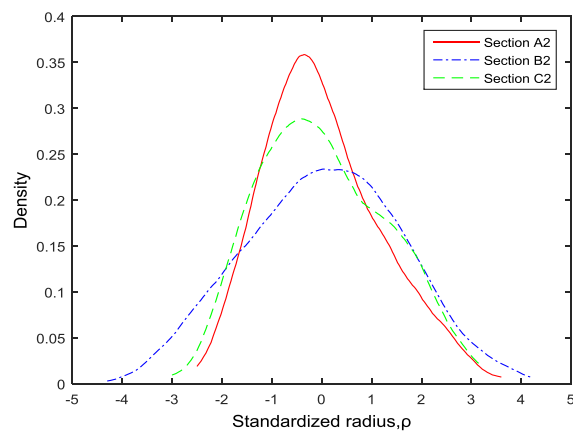

Figure 3. Kernel estimates of $\hat{f}_{A2}(\rho), \hat{f}_{B2}(\rho), \hat{f}_{C2}(\rho)$ based on the repeated 475 measuring points after surface cleaning.

2.2.2 Rosner method

The Rosner (GESD) method is exploiting for detection of single and multiple outliers in nearly normal distributed data, when exact number of outliers are unknown. The only upper limit m of expecting outliers is required to indicate.

In order to avoid the masking effect m should not be chosen too small. There are two hypothesis types: null hypothesis H_0 – no outliers in the sample, alternative hypothesis H_1 – the sample has up to m outliers. For two sided case, the ESD test statistic is computed as a following [9]:

$$R_i = \frac{\max_i |\rho_i - \overline{\rho}|}{s}, \qquad (4)$$

where $\overline{\rho}$ and s are the sample mean and the standard deviation, respectively. Excluding one observation, which maximized $|\rho_i - \overline{\rho}|$, the test statistic in (4) is recalculated again for $n-1$ data sample. This procedure repeats m times until all extreme measurements are removed from the data set. The output of this computation will be the array of $R_1, R_2, ... R_m$. Then the critical value k_i for each single element of the vector R_i is calculated:

$$k_i = \frac{t_{(p, n-i-)}(n-i)}{\sqrt{\left(n-i-1+\left(t_{(p,n-i-1)}\right)^2\right)(n-i+1)}}, i = 1, 2, ... m \quad (5)$$

where $t_{(p, v)}$ is quantile of Student's distribution with degrees v of freedom and probability $p = 1 - \frac{\alpha/2}{n-i+1}$. Thus the total number of outliers is the largest i such that $R_i > k_i$. Opposite to Grubbs test, GESD can be influenced by the swamping effect (described above), but the influence of the masking effect is relatively neglected.

Table 1. Efficiency rate of outlier detection (randomly located outliers, medium values, 100 observations)

Number of outliers	Method	Section A $\overline{\rho}_m^A = 4.49$	Section B $\overline{\rho}_m^B = 6.00$	Section C $\overline{\rho}_m^C = 4.94$
1	Grubbs	0.66	0.65	0.67
	Rosner	0.66	0.65	0.67
2	Grubbs	0.36	0.33	0.34
	Rosner	0.60	0.58	0.61
3	Grubbs	0.12	0.09	0.09
	Rosner	0.58	0.57	0.58
4	Grubbs	0.02	0.01	0.01
	Rosner	0.57	0.55	0.57

Table 2. Efficiency rate of outlier detection (randomly located outliers, large values, 100 observations)

Number of outliers	Method	Section A $\overline{\rho}_l^A = 5.2$	Section B $\overline{\rho}_l^B = 6.9$	Section C $\overline{\rho}_l^C = 5.7$
1	Grubbs	0.99	0.99	1.00
	Rosner	0.99	0.99	1.00
2	Grubbs	0.95	0.94	0.96
	Rosner	0.99	0.99	1.00
3	Grubbs	0.77	0.72	0.76
	Rosner	0.99	0.99	0.99
4	Grubbs	0.37	0.32	0.34
	Rosner	0.99	0.99	0.99

3 Data simulation and case study

The data sets used in this study were derived from CMM (Leitz PMM-C-600) measurements (x_i, y_i) taken from three cross-sections (A, B, C) of an internal cylindrical surface. The cylinder axis was aligned with z axis. The circle center coordinates (x_c, y_c) for each section were estimated with LSC (least squired circle) method by PC-DMIS software based on 475 measured points. Then the radius variable for each measured point r_i was calculated by:

$$r_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \quad . \qquad (6)$$

For practical convenience, the result data arrays of r_i were standardized by $\rho_i = (\overline{r} - r_i) \times 1000$, where $\overline{r}$ is

the average radius in the cross section. The standardized radius variable ρ_i were further used in simulation tests.

The data sets of repeated measurements A2, B2, and C2 were tested for normality distribution with Anderson-Darling method [10]. This method is more sensitive to outliers and especially effective for detecting any departure in the tails of data distribution. Only one of three data sets (p-values: 0.001, 0.138, 0.001 for A2, B2, C2 correspondently) had p-value over specified significance level 0.05, thus the data are quite unlikely from a population with normal distribution.

In order to obtain a form distribution of $f(\rho)$ for standardized variable ρ_i, the kernel density estimator $\hat{f}(\rho)$ was applied [11]:

$$\hat{f}(\rho) = \frac{1}{bn}\sum_{i=1}^{n} K\left(\frac{\rho - \rho_i}{b}\right), \quad -\infty < \rho < \infty \quad (7)$$

where K is the kernel smoothing function, b is a bandwidth and n is the sample size. Epanechnikov kernel was used as the smoothing function K with default MATLAB bandwidth b and the sample size n with 475 observations. The estimates of pdf (probability density function) for sections A1, B1, and C1 with outliers are illustrated on Fig. 3, and Fig. 4 for repeated measurements of the same sections A2, B2, and C2, after the outlier issue was physically eliminated (no analytical algorithm were used so far).

The estimated pdf objects $\hat{f}_{A2}(\rho), \hat{f}_{B2}(\rho), \hat{f}_{C2}(\rho)$ were further used to generate random data samples to simulate workpiece measurements without outliers (Fig. 4). In addition, some of the data points were replaced by simulated outliers. The simulation of outliers was based on a uniform distribution around a specified deviation from the mean value of the random data sample. The effectiveness of outlier detection of the Grubbs and Rosner methods with different combination of correlated factors were estimated from 10^5 iterations with summing two possible results: 0 – failure; 1 – success. In order to meet success requirements the same number of outliers with identical indexes must be detected (e.g. if only three outliers from four detected correctly then result is considered as a failure). The efficiencies e_G, e_R (Grubbs and Rosner method, respectively) were estimated simultaneously as a rate of number of success iteration Sum_G, Sum_R to total iteration number M, then $e_G = Sum_G / M$ and $e_R = Sum_R / M$.

The following factors were considered in the test of the efficiency of outlier detection procedures:

- non-normal distribution of random data samples;
- size variation of the random data samples;
- outliers randomly distributed around a mean value of the random data sample with specified deviation values;
- a defined number of outliers in each data set (from 1 to 4);
- outliers as randomly distributed data points or as a block of data points.

The more detail description of these factors is given in the next section.

Table 3. Efficiency rate of outlier detection (located as a block, large values, 100 observations)

Number of outliers	Method	Section A $\overline{\rho}_l^A = 5.2$	Section B $\overline{\rho}_l^B = 6.9$	Section C $\overline{\rho}_l^C = 5.7$
2	Grubbs	0.95	0.92	0.96
	Rosner	1.00	1	1
3	Grubbs	0.6	0.52	0.55
	Rosner	1.00	0.99	1
4	Grubbs	0.07	0.07	0.06
	Rosner	1.00	0.99	1

Table 4. Efficiency Rate of Outlier Detection for Various Sample Sizes (2 Outliers with Random Locations, Large Values)

| Sample size	Method	Section A $\overline{\rho}_l^A = 5.2$	Section B $\overline{\rho}_l^B = 6.9$	Section C $\overline{\rho}_l^C = 5.7$
15	Grubbs	0.07	0.06	0.06
	Rosner	0.75	0.74	0.76
30	Grubbs	0.44	0.42	0.42
	Rosner	0.92	0.92	0.94
60	Grubbs	0.84	0.83	0.85
	Rosner	0.98	0.98	0.99
100	Grubbs	0.95	0.94	0.96
	Rosner	0.99	0.99	1.00

4 Simulation and experiment results

The distribution of outliers in the simulations was based on a *medium* and a *large* deviation from the mean value of the simulated measurements. The medium value for the outliers are generated in the interval $\rho_m \in 3.90s \pm 0.01$ (Table 1) and the large value of the outliers from an interval $\rho_l \in 4.5s \pm 0.1$ (Table 2, 3), where s is the estimated standard deviation of the simulated measurements ($s_A = 1.15$, $s_B = 1.54$, $s_C = 1.27$ for measurement sets A2, B2, and C2, correspondingly). There are different numbers of outliers tested both with randomly distributed locations (Table 1, 2) and with location as a block (Table 3). For the random location, two discrete values, $\pm \rho_m$ and $\pm \rho_l$, were used, while for the block location only negative $-\rho_l$ values were integrated to meet the most typical conditions (associated with contamination). Due to low skewness, the simulation results for $+\rho_l$ were very similar thus, they are not shown here. The simulation results of the influence of the sample size on an outlier detection performance were specified in Table 4. The significance level $\alpha = 0.05$ was applied in all tests. All simulation tests were carried out in MATLAB and results are tabulated in Table 1, 2, 3, and 4. The simulation results were rounded up to the second digit from decimal point.

Both methods were also applied with the experimental measurements. The detected outliers were tabulated in Table 5. The comparing boxplot of data set after removing of outliers (A1*, B1*, C1*) with data samples of repeated measurements (A2, B2, C2) are illustrated on Fig. 4. There are two large outliers were removed in

Table 5. Appliance of the outlier methods with real measurement data (475 observations)

Outliers no	Outlier parameters	Section A1*		Section B1*		Section C1*	
		Rosner	**Grubbs**	**Rosner**	**Grubbs**	**Rosner**	**Grubbs**
1	**Index**	60	60	459	459	2	2
	Values [mm]	29.9881	29.9881	29.9829	29.9829	29.9835	29.9835
2	**Index**	61	61	-	-	-	-
	Values [mm]	29.9876	29.9876	-	-	-	-
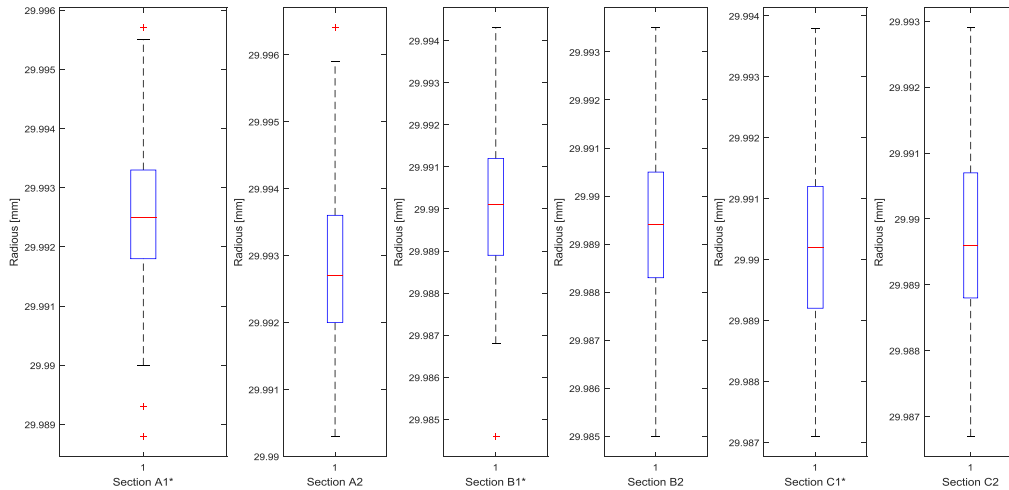
Figure 4. Boxplot of the measuring data: A1*, B1*, C1* - after removing of outliers; the repeated measurements A2, B2, C2 – after cleaning of workpiece.

section A1 with both Rosner and Grubbs tests and by one outlier in section B1 and C1. There were number of medium outliers which not detected by any of the methods (Sections: A1* and B1*) though some of these suspected points disappeared after measurement were repeated (Section A2). This fact confirm the simulation results, which were obtained in Table 1, for medium values of outliers.

5 Discussion

In the case of a single outlier, the Grubbs and Rosner tests have similar performance. For more than two outliers cases there was significant difference in the outlier detection efficiency. Both procedure had a lower efficiency rate for the medium outlier values (Table 1). Therefore, the parametric outlier tests must be used very carefully when small value of outliers are presented. However, Rosner method had at least 0.98 efficiency in whole range 1 – 4 of outliers in case of large outlier value, while the Grubbs method has 0.95 efficiency for two outliers, but even lower for larger number of outliers (Table 2). This is a good demonstration of the influence of the masking effect on the Grubbs procedure and the very low influence on Rosner method.

According to the Table 3, the additional simulation test showed that outliers distribution either as the block or random had no notable influence on Rosner method performance what is opposite to the Grubbs method, which efficiency was fairly lower for the block location of outliers than for random distributed among the sample set. Meanwhile, there was a great influence of the sample size on efficiency rate observed for both of the methods as shown in Table 4. The consecutive outlier detection procedure (Grubbs) had efficiency below 0.5 for sample size with 30 observations or lower, while Rosner's test could provide at least 0.75 efficiency rate even for 15 observations sample. In spite of some differences in distribution, form and variation range between all three data sets the test performance did not distinguished so much within each individual method. That leads us to a conclusion that both methods have no any strict requirements to the normal distribution.

There were no additional tests of masking or swamping effects presented for Rosner method in this paper. It is, however, a well-known fact that too small number of outliers initially applied in the Rosner test (relatively to the actual number of outliers in the sample) can lead to the masking effect. However, when the extra two outliers had been initialized with Rosner test (additionally to the actual number of outliers) during of simulation the swamping effect was not observed.

6 Conclusion

There are many different outlier procedures available for data analysis, but it is a difficult task for an unexperienced operator to choose the most suitable test for a particular problem. The following specific conditions were met for this study with considered methods:

- the ability of the methods work with various sample sizes;

- the ability to detect multi outliers when maximum outlier number is unknown;
- the stable efficiency (over 0.9) to detect the large outliers, which bring the most significant influence;
- the applicability for data from unknown non-normal distributions;
- the stability to the masking and swamping effects.

The outlier detection procedures such as Grubbs and Rosner can be successfully applied even with real workpiece measurements, which are difference from the normal distribution. However, the Rosner method is more reliable and hence preferable. Meanwhile the medium outliers should be double-checked before removing/accepting for further analysis. It is not recommended to use samples below 30 measuring points to avoid the low efficiency outlier detection procedure. The measurement tests conducted with CMM confirm the simulation results and all conclusions above. The research of experimental measurements also revealed that multiple outliers groups can be expected with CMM measurements. Therefore, the automated outlier detection procedure based on the Rosner / GESD method can be effectively applied with a geometry inspection routine.

References

1. ISO16269-4, Detection and treatment of outliers, in Part 4: *Statistical interpretation of data*, (2010)
2. Chambers John, William Cleveland, Beat Kleiner, and Paul Tukey, *Graphical Methods for Data Analysis*, Wadsworth, (1983)
3. Sim C.H., Gan F.F. and Chang T.C., *Outlier Labeling with Boxplot Procedures*, Journal of the American Statistical Association, **100**: p. 642-652, (2005)
4. Langford E., *Quartiles in Elementary Statistics*, Journal of Statistics Education, **14**, (2006)
5. Tukey J.W., *Exploratory data analysis*, Massachusetts: Addison-Wesley, (1977)
6. Barnett V., Lewis T., *Outliers in Statistical data*, ed. 3rd. 1994, New York: Wiley.
7. ISO5725-2, Accuracy of measurement methods and results, in Part 2: *Basic method for the determination of repeatability and reproducibility of a standard measurement method*, (1994)
8. Grubbs F., *Procedures for Detecting Outlying Observations in Samples*. Technometrics, **11**: p. 1-21, (1969)
9. Rosner B., *Percentage Points for a Generalized ESD Many-Outlier Procedure*. Technometrics, **25**: p. 165-172, (1983)
10. Anderson T.W., Darling D.A., *A Test of Goodness-of-Fit*. Journal of the American Statistical Association, **49**: p. 765–769, (1954)
11. Alexandre B.T., *Introduction to Nonparametric Estimation*, New York: Spring-Verlag, (2009)