https://doi.org/10.33271/nvngu/2026-3/133

Yu. Holovko*,
orcid.org/0000-0001-6081-8072

Dnipro University of Technology, Dnipro, Ukraine
* Corresponding author e-mail: holovko.yu.m@nmu.one

NONPARAMETRIC HOMOGENEITY CRITERION FOR SELECTIVE FORMATION OF AN ENSEMBLE OF TIME SERIES SEGMENTS

Purpose. Development of a criterion for testing the homogeneity of timeseries segments that are sequentially combined into an ensemble of segments, which constitutes a prerequisite for obtaining statistically significant estimates of the series.

Methodology. The study is based on probabilistic analysis and statistical modelling. A substantial part of the research is devoted to demonstrating the adequacy of the proposed criterion, including its comparison with a wellknown criterion and advantages over it.

Findings. It is shown that when the proposed criterion is applied to assess the homogeneity of an ensemble with a fixed number of segments, it yields results close to those of the well-known Zhang criterion. The advantages of the proposed criterion become evident during the sequential assessment of the possibility of adding segments to the ensemble. In this case, the number of "lost" segments can be reduced several-fold. The scope of the criterion is limited by the fairly wide ranges of parameters that were used in statistical modelling.

Originality. A new criterion for forming an ensemble of homogeneous timeseries segments is developed. The criterion does not require statistical modelling to determine the distribution functions of the test statistic when varying the number of samples in the segments or the number of segments in the ensemble.

Practical value. The use of the proposed criterion can ensure a significant increase in the statistical reliability of estimates of the state of a process represented as a uniformly sampled time series.

Keywords: *nonparametric homogeneity test, time series, segment ensemble, random process*

Received: 27 December 2025
Accepted: 06 March 2026
Available online: 27 June 2026

Introduction. In nearly all areas of human activity and society, there exist phenomena that naturally and necessarily require investigation in their development over time. Financial analytics, sociology, monitoring and diagnostics of technical systems, climatology, ecology, medicine, robotics, and many other fields rely on sequences of values (samples) obtained at increasing, equally spaced points in time. Such sequences are referred to as *time series* or *discrete signals*; when emphasizing the random nature of the values, they are also called *discrete random processes*.

Methods for analyzing time series are largely determined by their sources of origin. For example, in many cases it is impossible to fully reproduce the conditions necessary for obtaining repeated measurements. That is, only a single series (one realization of a random process) is available. It is also important whether the entire series is accessible for analysis at once, or whether the data are received sequentially in the so-called "online" mode.

The study of time series typically aims at achieving at least one of the following objectives: assessing the current state of the process, detecting typical and/or anomalous events, and forecasting the process. The priority objective determines the choice of research methods.

The scope of research considered in this article is defined as follows. A single sufficiently long time series is available. The series has been detrended and its seasonal component removed. The task is to estimate the state or states of the process while disregarding short-term random influences. For example, such a series may originate from acoustic signals of industrial or natural origin, electromagnetic oscillations of specific devices, variations in informational or social indicators, and similar sources. The goal of the analysis is to obtain characteristics of the state of the corresponding systems, leaving the investigation of possible short-term excitations for subsequent research. Decisions made on the basis of such observations should be based on statistically reliable estimates. Obtaining such estimates is not a trivial procedure.

According to a fundamental principle of mathematical statistics, assessing the statistical significance of any characteristic requires independent repeated samples drawn from the same distribution. Real processes, however, often consist of fundamentally different stages. The presence of such stages is determined by changes in the state of the source or by external influences. This neces-

© Holovko Yu., 2026

sitates dividing the data into separate segments of consecutive samples and analyzing the feasibility of using these segments to compute statistically reliable estimates.

Segmentation may also be performed without the intention of computing statistical estimates. For instance, segmentation makes it possible to detect deviations from the overall trend, anomalous outliers, missing data, and other regime changes that cannot be identified through global analysis. In machine learning, dividing data into training and test segments is a prerequisite for verifying a model's learning capability, identifying regions with different forecast quality, and tuning algorithm parameters to local features. With large datasets, segmentation enables parallel processing and simplifies data storage and transmission in distributed systems.

Segmentation can be performed in various ways depending on its objectives. However, in all cases, statistically significant characteristics (estimates) are derived from multiple segments. This set of segments (the ensemble of segments) must possess identical probabilistic properties; otherwise, the statistical significance of the estimates may be negligible.

When regime changes occur, the same process with identical parameters may appear in different portions of the series that do not share common boundaries. Combining such corresponding segments can obviously also improve the statistical significance of the estimates. It is only necessary to make sure that the segments belong to the same process.

The problem lies in constructing a criterion for combining segments into an ensemble with similar probabilistic properties. The criterion must also verify whether an individual segment can be included in an already existing ensemble of segments.

Literature review. The formation of ensembles of segments with similar probabilistic properties is an inherent task in classification, clustering, outlier and missing-data detection, and the identification of switching points.

In recent years, the use of Machine Learning (ML) methods has expanded rapidly. The application of these methods to time series largely relies on ideas developed in image recognition [1, 2] and natural language processing [3]. The significant achievements in these fields have motivated the concentration of time-series research in combination with ML methods [4].

Three unsupervised ML tasks are most closely related to the topic of this article:

- identification of the samples that divide the series into non-overlapping segments with similar properties in a given sense – Change Point Detection (CPD);
- grouping segments based on the similarity of extracted characteristics – Segment Clustering (SC);
- determination of the system state – State Detection (SD).

ML includes a number of methods characteristic of this paradigm. For example, optimal clustering methods [5], density-based clustering algorithms [6], change-point detection algorithms employing decision trees, artificial neural networks, and Bayesian networks [7, 8], as well as generative learning models for segmentation and anomaly detection [9]. The cited reviews summarize results obtained in recent years through the application of these approaches.

One of the most widely used approaches to change-point detection remains the cumulative sum methodology (Cumulative Sum, CUSUM) introduced by Page (1954) [10, 11]. The cumulative sum accumulates deviations from a target value and identifies a change point once the sum exceeds a specified threshold. The popularity of CUSUM is due to its high computational efficiency, achieved through the recursive implementation of the cumulative-sum statistic. The statistical approach to change-point detection within the CUSUM framework is based on comparing probability distributions of samples: the logarithm of the likelihood ratio of two consecutive intervals separated by a potential change point is monitored [12]. Options for selecting an optimal sliding window are examined in [13].

The method continues to be developed, for example in [14, 15], but it has a fundamental limitation: it assumes that the probability distributions of the samples both before and after the change point are known. This assumption is precisely what enables the construction of efficient recursive computations.

A more general situation arises when the distribution after the change point is unknown. In this case, it is more appropriate to use the Generalized Likelihood Ratio (GLR) method (Siegmund & Venkatraman, 1995), which includes estimating the previously unknown parameters of the distribution after the change point. However, as in the case of CUSUM, the form of the distribution must still be specified in advance [16]. Moreover, the ability to compute the statistic recursively is completely lost, leading to a substantial decline in computational efficiency [17].

In studies [18, 19], regime changes are identified through variations in the logarithmic likelihood. The change detector is based on the Kullback–Leibler statistic (Kullback & Leibler, 1951). A detector output that exceeds a predefined threshold is interpreted as a regime change at the boundary between two segments. It is important to note that the accuracy of change-point methods deteriorates significantly in the presence of noise in the data [20].

The time-series segmentation method proposed in [21] (the ClaSP method – Classification Score Profile) is built upon the concept of self-supervision [22]. A change point is identified by training a binary *k*-nearest-neighbors classifier for each possible split point. The preferred split is the one that best distinguishes subsequences belonging to one of the two parts. The authors claim that a substantial advantage of the method lies in its domain independence, meaning that the method does not require the involvement of domain experts to specify hyperparameters. One such hyperparameter, which greatly influences performance and determines the size of the initial segments, is the window size. In the ClaSP method, the window size is set to the smallest window for which the aggregated window statistic best matches the statistic of the entire series.

Clearly, this approach is not flawless – it is easy to construct a synthetic series containing a relatively short interval with markedly different statistical properties. In such a case, segments constructed based on statistical similarity to the entire series may "overlook" the interval with significantly different characteristics. Demonstrating the method, even on a sufficiently large number

of examples, cannot be considered proof of its reliable applicability to time series of arbitrary origin.

ML methods make it possible to automatically detect hidden structures in complex, high-dimensional, and heterogeneous large-scale data. The capabilities of ML methods continue to grow. However, they have several systemic limitations and drawbacks that are often highlighted in the literature. Models may fit the training data too closely, losing their ability to generalize to new data (overfitting) [23, 24]. ML methods are sensitive to data quality (noise, missing values) [25, 26]. Complex models operate as a "black box," and even when producing highly accurate results, the inability to explain these results restricts their applicability [27, 28]. There are no universal rules for selecting the appropriate model or hyperparameter values, which results in domain dependence of the models.

Segmentation is an intermediate stage of the learning process and, in many cases, is constructed according to the principle of adapting to the data. This leads to segmentation drawbacks in ML that correspond to the general limitations of machine learning listed above.

Some ML problems can be addressed using classical statistical methods. Moreover, there are publications demonstrating that classical methods can provide higher accuracy [29, 30]. These methods are especially useful and effective when applied to time series with clearly expressed structure, ensuring statistical soundness and transparent interpretability of the estimates.

Classical methods include decomposition, stationarity testing, stationarization (when necessary), autocovariance analysis, spectral analysis, and the construction of autoregressive and moving-average models (AR, MA, ARMA) [31].

The use of the classical methods and the correct computation of time-series characteristics are possible only if the observations follow a single probabilistic law – that is, if the process is stationary, at least in the wide sense. Therefore, testing the stationarity of the series is a necessary prerequisite for applying classical analysis methods.

To assess stationarity, the two most commonly used tests are the Augmented Dickey–Fuller test (ADF-test) and the Kwiatkowski–Phillips–Schmidt–Shin test (KPSS-test). Both tests are parametric and are based on the concept of a unit root in the characteristic equation of an autoregressive model of the series. Their joint application is advantageous due to the opposite formulations of the null (H_0) and alternative (H_1) hypotheses: ADF – H_0: "the series has a unit root (the series is non-stationary)", H_1: "the series does not have a unit root"; KPSS – H_0: "the series is trend-stationary", H_1: "the series has a unit root (the series is non-stationary)".

Each of the tests has its own drawbacks. However, there are shortcomings inherent to both tests: errors in cases of "nearly non-stationary" series (root close to 1); dependence on window width and the number of lags; sensitivity to heteroskedasticity (non-constant variance); and errors in the presence of structural breaks [4, 32]. The latter two issues arise because both tests assume a constant process structure. If the series exhibits abrupt changes in level, trend, or variance, both tests typically indicate that the series is non-stationary, even though it may actually be piecewise stationary. Pre-segmenting the series and subsequently testing each segment individually does not solve the problem. First, each segment may require different actions for stationarization (trend removal, differencing), and second, with short segments, the sampling distributions of the test statistics deviate substantially from their asymptotic forms, making the critical values inaccurate [33].

Real-world series, which are almost always non-stationary, may be considered stationary over short or moderately long intervals only with a certain degree of confidence. Combining such intervals promises more reliable estimates. At the same time, the convenience of a stationary model often leads to its use even without sufficient justification. However, perfunctory uniting of data directly results in unacceptably wide confidence intervals in the computed estimates.

When regime changes occur, a stationary process with identical parameters may appear in different portions of the series that do not have common boundaries. Combining the corresponding segments can, evidently, improve the statistical significance of the estimates.

Purpose. To develop a criterion for assessing the homogeneity of time-series segments during the sequential construction of a segment ensemble. Such a criterion enables the formation of an ensemble containing a sufficient number of segments to ensure the required level of statistical reliability of the resulting estimates.

Methods. Since a discrete random process, or time series, is a set of ordered random variables, its most complete statistical description is given by the joint distribution function, which depends on a large (possibly infinite) number of arguments (twice the number of series elements). However, estimating such a function in the general case is not feasible.

The situation becomes significantly simpler if the assumption of stationarity (or ergodicity when only a single realization is available) is admissible. A wide-sense stationary time series can be effectively described by its mean and autocovariance function, or by its spectral density function. Statistical identity of these characteristics for any intervals of the series serves as evidence of wide-sense stationarity. These functions reflect different but equivalent properties of the series. Under stationarity, the value of the autocovariance function at zero lag, the integral of the spectral density from zero to the Nyquist frequency, and the variance must be identical for any intervals of a stationary series. In particular, this implies that if segments belong to the same stationary process, then their variances, as well as their means, must be equal.

In the further construction of the ensemble, we will require the homogeneity of the probability distributions of the sample values within the segments, which is a more general requirement than the equality of means and variances. Some clarifications are necessary here.

If two segments have completely identical distribution functions, then their means and variances (as well as all other existing moments) are equal. However, when we accept the statistical identity of the distribution functions of two segments, we only do not reject the hypothesis of their equality at a given significance level α. Therefore, in the general case, we can only state that, at the confidence level $1 - \alpha$, no grounds were found to distinguish the means and variances of the compared segments.

The assessment of correlation characteristics is not considered in this study. However, it is assumed that such analysis must be conducted at the next stage – after the ensemble has already been formed. The choice of methods depends on the objectives of the research. Classical methods include the construction of the autocovariance function, the power spectral density function (nonparametric methods), and regression models (parametric methods). The ensemble segments serve as the basis for calculating these functions and models.

The calculation of the autocovariance function or the power spectral density can be performed both for each segment individually and after the ensemble is fully arranged. In either case, the final form of the functions results from averaging. The variances and the confidence intervals of the functions' values may be regarded as indicators supporting the assumption of stationarity.

If the goal is to construct a process model, the model coefficients for the ensemble may be obtained by averaging the coefficients of models of identical orders fitted to each segment. Narrow confidence intervals for the coefficients may indicate the validity of the stationarity assumption.

Although no estimates of the time series are computed at this stage, it is taken into account that the ensemble of segments is being created specifically to obtain statistically reliable estimates.

Test statistic. We restrict our consideration to the case where each segment contains the same number of samples. Then the macro-parameters of the procedure for forming the ensemble estimates are:

n – the number of samples in a segment;

k – the number of segments in the ensemble.

As n increases, the sample estimates reflect a greater number of data features. As k increases, the sample estimates, according to the law of large numbers, approach their true values, and the confidence intervals become narrower.

Evidently, increasing both macro-parameters is limited by the volume and properties of the data as well as by processing time.

The value of k may be initially set based on expert assessment of the properties of the series and taking into account the fact that increasing the number of segments reduces confidence intervals proportionally to $\frac{1}{\sqrt{k}}$. Subsequently, the value of k may be adjusted depending on the required accuracy.

The number n is determined by the type of estimated characteristics. For example, suppose the task is to estimate the spectral characteristics of the series. Assume that the samples were obtained at a sampling frequency v_d, and that the desired spectral bin width must not exceed $\Delta\nu$. Then the minimum number of samples in a segment can be set immediately as $n = 1 + \left\lceil \frac{\nu_d}{\Delta\nu} \right\rceil$.

This work is devoted to constructing a criterion that determines whether a segment of n samples can be included in an ensemble of segments, each also containing n samples. In this setting, the hypothesis that all segments belong to the same parent population must remain acceptable at a specified significance level. Under such assumptions, the order of the samples within a segment is irrelevant, and each segment can be regarded as a selection of n independent realizations of a random variable.

Assuming that no prior information about the properties of the series is available, we will use nonparametric tests for comparing samples. The advantage of nonparametric tests lies primarily in the fact that their form remains unchanged for any continuous non-decreasing distribution of sample elements. This property arises because the test statistics are constructed not from the sample values $x_i, i = \overline{1,n}$, but from the values of the distribution function $F(x_i)$, which, due to

$$P(F(x_i) < x) = P(x_i < F^{-1}(x)) = F(F^{-1}(x)) = x, \quad (1)$$

are uniformly distributed for any continuous distribution of the sample elements. In the analysis of arbitrary data, especially when the underlying distribution deviates from normality, this advantage becomes decisive.

The verification of the homogeneity of the segments is directly determined by the ensembleconstruction procedure. The following strategy for building the ensemble is possible.

The signal samples are sequentially grouped into segments containing the same predetermined number of samples (n). After forming a specified number of segments (k), the homogeneity of the entire collection of segments (the ensemble) is tested. The problem may be formulated as follows.

Test the validity of the null hypothesis

$$H_0 : \ F_i(x) = F_0(x), \ i = \overline{1,k},$$

against the alternative hypothesis

$$H_1 : \exists\, j \in \overline{1,k} \Rightarrow F_j(x) \neq F_0(x), \quad (2)$$

where $F_i(x)$ is the distribution function of the i^{th} sample; $F_0(x)$ is an arbitrary unknown distribution function.

A general approach to testing the homogeneity of the distribution functions of several samples is based on the use of twosample tests [34]. The test statistic measures the deviation of the empirical distribution functions of the samples from the empirical distribution function constructed from the entire collection of samples. That is, in formulation (2), the function F_0(x) is taken to be the distribution function of the combined samples, and the statistic is computed using the corresponding empirical functions $F_i^e(x)$ and $F_0^e(x)$. As a result of comparing the statistic with its critical value, the hypothesis of homogeneity of the ensemble segments H_0 either cannot be rejected and is therefore accepted, or must be rejected.

k – sample versions of the Kolmogorov-Smirnov and Anderson-Darling (W. J. Conover, 1999) tests, as well as Zhang's tests (J. Zhang, 2006), are built upon a similar methodology. The critical values of their statistics rely on asymptotic approximations with respect to sample size and, in most cases, are known only for certain values of the number of samples in the ensemble. For arbitrary values of n and k, determining the distribution of the test statistics requires statistical modelling.

If H_0 is rejected, the homogeneity of the ensemble can be violated by only a single segment. For example, this may occur when a point anomaly (an outlier) is

superimposed on an otherwise stationary ergodic series. Even if the procedure used to compute the statistic makes it possible to identify which particular segment violates homogeneity, after removing this "offending" segment, one must again test the homogeneity of the remaining ($k - 1$) segments. For the updated ensemble, the homogeneity hypothesis may again be rejected, and once more, after removing the segment that breaks homogeneity, one must test the remaining ($k - 2$) segments. And so on. When the homogeneity hypothesis is finally accepted, the required statistical estimates are computed using the segments that remain in the ensemble.

Even if a reliable k – sample homogeneity test is available, the described procedure has the following drawbacks:

- after removing a segment (or several segments) from the ensemble, the homogeneity of the same remaining segments must be tested again;
- different ensembles derived from the same series, due to the removal of certain segments, will contain different numbers of segments. Consequently, the estimates obtained from different ensembles will have different accuracy levels;
- the number of segments remaining in the ensemble may be too small to obtain estimates with the desired confidence level.

It is natural to attempt to construct k-sample homogeneity tests based on known twosample tests. The hypothesis to be tested can be written as

$$H_0 : F_i\left(x\right) = F_j\left(x\right),\ \forall i, j \in \overline{1,k},$$

against the alternative hypothesis

$$H_1 : \exists i_*, j_* \in \overline{1,k} : F_{i_*}\left(x\right) \neq F_{j_*}\left(x\right),$$

and, accordingly, the statistic may be constructed by comparing the values of the twosample statistics for all possible pairs of samples.

Based on twosample tests, the statistic for a k-sample homogeneity test can be defined as

$$M_k = \max_{1\le i,j\le k}\left\{S_{i,j}\right\}, \tag{3}$$

where $S_{i,j}$ is the statistic of a twosample test computed for the i^{th} and j^{th} samples. The statistics $S_{i,j}$ are computed for all possible pairs of samples in the ensemble, and their total number is $C_k^2 = \dfrac{k(k-1)}{2}$.

Any known statistics of two sample tests may be used as $S_{i,j}$ (A. I. Kobzar, 2006). In this study, the calculations were performed using the Smirnov test (Bolshev & Smirnov, 1983). We present the main relations employed in this case.

For an arbitrary sample $\theta_1, \theta_2, ..., \theta_n$, whose elements are ordered in increasing order, the empirical distribution function $F^e(x)$ is computed as

$$F^e\left(x\right) = \frac{1}{n}\sum_{i=1}^{n} H\left(x - \theta_i\right), \tag{4}$$

where $H\left(x\right) = \begin{cases} 0, & x < 0 \\ 1, & x \ge 0 \end{cases}$ is the Heaviside function and x is a continuous variable.

The test statistic (the twosided twosample Smirnov statistic) is

$$D = \max_x \left| F_1^e(x) - F_2^e(x) \right|, \tag{5}$$

where $F_1^e(x)$, $F_2^e(x)$ are the empirical distribution functions of the two samples with the same number of elements n. The statistic is evaluated at the jump points of the empirical distribution function of the pooled sample. Asymptotically, the distribution of this statistic is described by the Kolmogorov function $K(t)$

$$P\left(\eta D \le t\right) = K(t) = 1 - 2\sum_{j=1}^{\infty}(-1)^{j+1} e^{-2(jt)^2}, \tag{6}$$

which makes it possible to compute the achieved significance level

$$p_{value} = \min\left(\max\left(2\sum_{j=1}^{\infty}(-1)^{j+1} e^{-2(jt)^2}, 0 \right), 1 \right).$$

The hypothesis that all samples originate from the same source is rejected if $p_{value} < \alpha$, where α is the specified significance level.

The quantity η in (6) was computed using Stephens' formula (Stephens, M. A., 1970)

$$\eta = \sqrt{0.5n} + 0.12 + \frac{0.11}{\sqrt{0.5n}}. \tag{7}$$

Let us denote the corrected statistic by $S = \eta D$. Accordingly (6) in the asymptotics

$$P(S \le t) = K(t).$$

The statistic $F^e(x)$ can take only discrete values, increasing in steps of $\dfrac{1}{n}$. Consequently, D, ηD and the corrected statistic S are also discrete, and the values of S change in increments

$$\Delta S = \frac{\eta}{n} = \frac{1}{\sqrt{2n}}\left(1 + \frac{0.12}{n} + \frac{0.11\cdot\sqrt{2}}{n\sqrt{n}}\right) \approx \frac{1}{\sqrt{2n}}.$$

This results in a steplike form of the distribution function graph of the statistic S. The lengths of the horizontal segments are equal to ΔS (or multiples of ΔS for small n), while the lengths of the vertical segments multiples of $\dfrac{1}{n}$.

It is important that approximations (6) and (7) can be used satisfactorily when $n > 50$, and the approximation error decreases with increasing n as $\dfrac{1}{n}$ (Bolshev & Smirnov, 1983). In the following, we assume that the number of samples in the segments satisfies this condition. However, the steplike shape of the distribution function remains significant for all practically feasible values of n.

The difficulty in using the criterion based on the statistic M_k defined in (3) lies in the necessity of having the distribution function of this statistic.

Statistical modelling of the distribution function of the test statistic. As is well known, the methodology of statistical modelling is based on the Monte Carlo method.

The search for the distribution function of the statistic M_k was carried out as follows.

Multiple simulations of independent realizations of sequences of random variables (samples) with a specified distribution function were performed. Each realization consisted of *nk* samples. The samples were sequentially grouped into *k* samples (an ensemble). For each simulated ensemble, the test statistic was calculated. Based on the values obtained from a large number of simulations (the number of simulations is denoted by *N*), the distribution function of the statistic was constructed.

It is important to assess the accuracy of the distribution function obtained in this way and its dependence on the number of simulations.

Let us use the notion of an indicator function in the form

$$1_{s<x} = \begin{cases} 1, & \text{if} \quad s \le x \\ 0, & \text{if} \quad s > x \end{cases}, \tag{8}$$

where s is a random variable; x is a parameter.

If $S_1, S_2, ..., S_N$ are the N values of the statistic, each computed from the corresponding simulated ensembles, then an estimate of the distribution function of the statistic at the point x can be taken as

$$\tilde{S}_N(x) = \frac{1}{N}\sum_{j=1}^{N} 1_{S_j<x}.$$

Since the expectation and variance of the indicator function (8) are

$$E(1_{s\le x}) = F(x); \quad D(1_{s\le x}) = F(x)(1 - F(x)),$$

the expectation and variance of $\tilde{S}_N(x)$ are

$$E\left(\tilde{S}_N(x)\right) = F(x); \quad D\left(\tilde{S}_N(x)\right) = \frac{F(x)\left(1-F(x)\right)}{N}.$$

For fixed x, according to the Central Limit Theorem, the distribution of $\tilde{S}_N(x)$ is asymptotically normal, that is,

$$\frac{\tilde{S}_N(x) - F(x)}{\sqrt{\frac{F(x)\left(1-F(x)\right)}{N}}} = \sqrt{N}\,\frac{\tilde{S}_N(x) - F(x)}{\sqrt{F(x)\left(1-F(x)\right)}} \sim \mathrm{N}(0,1).$$

Then

$$\mathrm{P}\left\{\left|\sqrt{N}\,\frac{\tilde{S}_N(x) - F(x)}{\sqrt{F(x)\left(1-F(x)\right)}}\right| < t_\gamma\right\} = 1 - 2\Phi(t_\gamma) = \gamma,$$

or

$$\mathrm{P}\left\{\left|\tilde{S}_N(x) - F(x)\right| < t_\gamma\sqrt{\frac{F(x)\left(1-F(x)\right)}{N}}\right\} = \gamma,$$

where $\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_0^t \exp\left(-\frac{z^2}{2}\right)dz$ is the Laplace (standard normal CDF) function; γ is the confidence level; t_γ is the quantile of the standard normal distribution corresponding to confidence level γ; value $\tilde{S}_N(x)$ may be used as the value of $F(x)$ at the confidence level γ, with absolute

$$\Delta = \Delta(x, N, \gamma) = t_\gamma\sqrt{\frac{F(x)\left(1-F(x)\right)}{N}},$$

and relative

$$\delta = \delta(x, N, \gamma) = \frac{\Delta(x, N, \gamma)}{F(x)} = t_\gamma\sqrt{\frac{\left(1-F(x)\right)}{NF(x)}} \tag{9}$$

errors.

Since the maximum value of $F(x)(1 - F(x))$ is $\frac{1}{4}$, one can sometimes use the rather crude upperbound estimate for the absolute error $\Delta \le \frac{t_\gamma}{2\sqrt{N}}$.

It is most important to evaluate the values of the empirical distribution function and the corresponding statistic values (percentiles) in its tails. In the case of a lefttail critical region, when the region of interest is where $F(x)$ is small, the use of formula (9) is fully justified.

For a righttail critical region, where $F(x)$ is close to 1, the accuracy of the distribution function itself is less important than the accuracy of its complement to one. Therefore, in this case, the relative error must be determined by relating the absolute error not to $F(x)$, but to $1 - F(x)$. Thus,

$$\delta_+(x, N, \gamma) = \frac{\Delta(x, N, \gamma)}{1 - F(x)} = t_\gamma\sqrt{\frac{F(x)}{N[1-F(x)]}}.$$

For a given confidence level γ, the number of simulations N_γ required to ensure a relative error δ_+ (righttail critical region) is

$$N_\gamma = \frac{t_\gamma^2 F(x)}{\delta_+^2[1-F(x)]}.$$

Table 1 illustrates the relationship between the desired accuracy and the required number of simulations.

The calculations presented below were obtained with no fewer than 4,000 simulations. This ensures that, at confidence level $\gamma = 0.95$, the relative error of the estimated distribution function of the statistic at the level 0.9 does not exceed $\delta_+ = 0.078$, and the absolute error does not exceed $\Delta = 0.0078$. For the distributionfunction value 0.95 at the same confidence level, the relative error does not exceed $\delta_+ = 0.113$, and the absolute error does not exceed $\Delta = 0.006$.

Fig. 1 shows, as an example, the plots of the distribution function of the statistic M_k (solid lines), obtained for different numbers of samples in the segments (n = 64, 32) and different numbers of segments in the ensemble (k = 4, 64). Small values of $n$were selected intentionally to demonstrate more clearly the discrete nature of the dependencies. The points corresponding to the empirical distributionfunction values lie only on the vertical segments of the graph. The horizontal segments contain no such points and are drawn solely to provide a continuous curve.

Approximation of the test statistic. In addition to the empirical distribution functions, Fig. 1 also shows the plot of the Kolmogorov distribution (black solid line). The shape of the imaginary envelope curves of the empirical distribution functions (one can draw them on the

Table 1

Required number of simulations to ensure the relative error of the estimated distribution function (δ_+) at a given confidence level (γ)

γ		0.99				0.95				0.9			
δ_+		0.5	0.2	0.1	0.05	0.5	0.2	0.1	0.05	0.5	0.2	0.1	0.05
$F(x)$	0.85	123	767	3,066	12,264	62	384	1,534	6,134	38	233	932	3,726
	0.90	195	1,218	4,870	19,477	98	609	2,436	9,742	60	370	1,480	5,917
	0.95	412	2,570	10,280	41,119	206	1,286	5,142	20,566	125	781	3,123	12,491
	0.99	2,143	13,391	53,562	214,247	1,072	6,698	26,790	107,159	651	4,068	16,271	65,084

left and on the right) is close to the graph of the Kolmogorov distribution.

Furthermore, the following observation can be made. Suppose that the statistic is constructed as

$$M_k^0 = \max_{i=1,3,..k-1}\{S_{i,i+1}\} = \max\{S_{1,2}, S_{3,4}, ..., S_{k-1,k}\}, k \text{ even},$$

that is, as the maximum of twosample statistics computed for different independent pairs of segments (assuming H_0 holds). Since each $S_{i,i+1}$ is distributed according to the Kolmogorov distribution, the distribution function of M_k^0 would be

$$\begin{aligned} F_{M_k^0}(z) &= P\{M_k^0 \le z\} = \\ &= P\{S_{1,2} \le z, S_{3,4} \le z, \ldots S_{k-1,k} \le z\} = \\ &= F_{S_{1,2}}(z) F_{S_{3,4}}(z) \ldots F_{S_{k-1,k}}(z) = K^{\frac{k}{2}}(z). \end{aligned} \quad (10)$$

When computing the statistic M_k from (3), each k of the samples of the ensemble is involved in the construction of several twosample statistics (specifically, in constructing $(k-1)$ statistics). Therefore, the statistics $S_{i,j}$ in (3) are not independent, and formula (10) cannot be applied directly in this case.

As an alternative, it may be reasonable to construct a k-sample statistic based on the maxima of twosample statistics for all possible pairwise combinations of samples in the ensemble. All of these statistics are also distributed as $K^{\frac{k}{2}}(z)$. However, this approach encounters a very rapid growth in the number of possible pairings as the number of elements increases [35]

$$n_k^{(2)} = \frac{k!}{2^{\frac{k}{2}}\left(\frac{k}{2}\right)^2}, k \text{ even}.$$

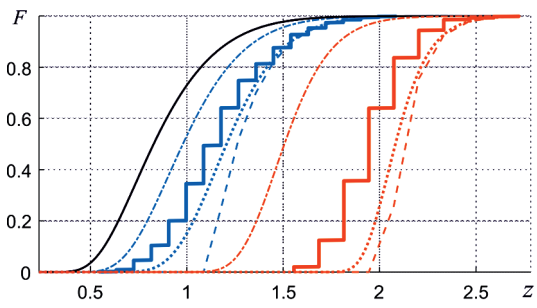

Fig. 1. Distribution functions obtained by statistical modelling and their approximations:

colors: blue — n = 64, k = 4; red — n = 32, k = 64; black — Kolmogorov distribution

Thus, for example, $n_6^{(2)} = 15$, $n_{10}^{(2)} = 945$, $n_{50}^{(2)} = 5.84\ldots\cdot 10^{31}$. Such an approach may be reasonable only for small values of k.

Returning to the distribution function M_k

$$F_{M_k}(z) = \mathrm{P}\{S_{1,2} \le z, S_{1,3} \le z, \ldots S_{k-2,k} \le z, S_{k-1,k} \le z\}.$$

It is impossible to determine this function without additional information about the mutual dependencies among all twosample statistics $S_{i,j}$. However, it is possible to specify bounds for the distribution function using the generalized Fréchet inequality

$$F_{M_k}^{\inf}(z) = \max\left(0, kK(z) - (k-1)\right) \le F_{M_k}(z) \le K(z).$$

In Fig. 1, in addition to the empirical distribution function of the statistic (3), the bounds of the possible values of the distribution function are shown — $K(z)$ (black solid line) and $F_{M_k}^{\inf}(z)$ (dashed line), as well as the curves $F = K^{\frac{k}{2}}(z)$ (dashdot) and $F = K^{\frac{k(k-1)}{2}}(z)$ (dotted). The latter two curves are constructed under the assumption that all twosample statistics involved in the construction of the statistic are independent. Their shapes are close to the envelope of the empirical distribution functions.

Based on these observations and on the analysis of the results of statistical modelling, it is proposed to approximate the distribution function of the statistic M_k in the following form

$$F_{M_k}(z) = K^{\left(1+\frac{2}{\sqrt{n}}\right)\left(\frac{k}{2}\right)^2}(z). \quad (11)$$

Relation (11) is proposed for use when $n \in [64, 2048]$, $k \in [2, 128]$, and $F_{M_k}(z) \in [0.8; 0.995]$. These ranges of n and k are determined by the volume of the statistical modelling performed and may likely be extended toward larger values.

The criterion based on the statistic M_k in (3) and the approximation of its distribution function in the form (11) will be referred to as the M_k-criterion. This is a righttailed test: the null hypothesis is rejected for large values of the statistic M_k. The achieved significance level is computed from

$$p_{value} = \min(\max(1 - F_{M_k}(M_k^*), 0), 1), \quad (12)$$

where M_k^* is the computed value of the statistic.

Assessment of the adequacy of the M_k-criterion. The adequacy of the proposed criterion will be evaluated by comparing the significance levels (probabilities of Type I errors) α and α_M. The true or nominal significance levels

α are determined by the parameters of the generated population of random variables. The empirical significance levels α_M are computed as the ratio of the number of ensembles for which, according to the criterion $p_{value} < \alpha$, to the total number of generated ensembles.

Ideally, $\alpha_M = \alpha$ for all possible values of α. Furthermore, in the case of an ideal criterion, the values of p_{value} are uniformly distributed on the interval [0, 1] (1). In reality, due to the stochastic nature of the data, these ideal properties are not observed, but the closeness of α and α_M, as well as a relatively uniform histogram of the empirical p_{value}, can be regarded as indicators of the quality of the criterion.

The discreteness of the proposed statistic complicates the use of histograms for assessing adequacy, therefore we will rely only on the deviations of α_M from α.

If $\alpha_M \gg \alpha$, the use of the criterion leads to an inflated Type I error rate; that is, the null hypothesis may be rejected when it is actually true.

If $\alpha_M \ll \alpha$, the acceptance region of the null hypothesis becomes unjustifiably wide, which almost certainly leads to accepting H_0 when the alternative is actually true – thus the probability of Type II errors increases.

Establishing an acceptable relationship between α and α_M is generally subjective and must take into account the properties of specific data. Nevertheless, following Bradley (Bradley, J. V., 1978), we will consider a criterion "stringent" if $0.9\alpha \le \alpha_M \le 1.1\alpha$, and "liberal" if $0.5 \le \alpha_M \le 1.5\alpha$. Additionally, we will consider the criterion "mediocre" if $0.2 \le \alpha_M \le 2\alpha$. Only if none of these conditions are met will the criterion be regarded as "unacceptable".

It should be noted that these characterizations of the criterion are tied to the specific significance level α. The same criterion may be unacceptable at one level and stringent at another.

For an additional evaluation of adequacy, all generated ensembles were also tested by Zhang's Atest, which, according to many researchers, is the best known k-sample homogeneity test [36, 37].

Below we present the main features of applying Zhang's Atest, taking into account a simplification enabled by the equal number of observations in all samples.

The combined ordered sample $X_1, X_2, \ldots, X_{\hat{n}}$, where $\hat{n} = nk$, is augmented with the elements $X_0 = -\infty$, $X_{\hat{n}+1} = \infty$.

Let the ranks of the elements of the i^{th} sample in the combined sample be denoted by $R_{i,j} (i = \overline{1,k};\ j = \overline{1,n})$, with the additional ranks of X_0 and $X_{\hat{n}+1}$ defined as $R_{i,0} = 1$ and $R_{i,n+1} = \hat{n} + 1$, respectively.

The values of the empirical distribution function of the i^{th} sample at the points $X_1, X_2, \ldots, X_{\hat{n}}$ are denoted by $F_{i,m}\left(i = \overline{1,k};\ m = \overline{1,\hat{n}}\right)$, and, taking $j_0 = 0$, we compute sequentially for $m = 1, 2, \ldots, \hat{n}$

$$\left(j_m, F_{i,m}\right) = \begin{cases} \left(j_{m-1} + 1,\ \dfrac{j_{m-1} + 0.5}{n}\right) & \text{if } R_{i,j_{m-1}+1} = m \\ \left(j_{m-1},\ \dfrac{j_{m-1}}{n}\right) & \text{if } R_{i,j_{m-1}+1} \ne m \end{cases}.$$

The test statistic Z_A is computed as

$$Z_A = -n \sum_{m=1}^{\hat{n}} \sum_{i=1}^{N_i} \frac{F_{i,m} \ln F_{i,m} + (1 - F_{i,m}) \ln(1 - F_{i,m})}{(m - 0.5)(\hat{n} - m + 0.5)}.$$

This is a lefttailed test: the null hypothesis is rejected for small values of the statistic Z_A.The achieved significance level is computed as

$$p_{value} = \max(F_{Z_A}(Z_A^*), 0),$$

where $F_{Z_A}(z)$ is the distribution function of the statistic and Z_A^* is the observed value.

Use of the criterion requires prior statistical modelling to determine the distribution of the statistic $F_{Z_A}(z)$.

A comparison of the Type I errors obtained using the proposed M_k-criterion (denoted α_M) and Zhang's Acriterion (denoted α_Z) with the nominal Type I error level α is presented in Table 2. The data are given for various values of the number of samples in a segment (n) and the number of segments in the ensemble (k), at nominal significance levels α = 0.005, 0.01, 0.05, 0.1, 0.15, 0.2, 0.25. In total, the table contains 203 comparison points.

According to the classification described earlier, the M_k-criterion can be considered *liberal* in 189 cases (the corresponding values in the table are shown in italics), of which 115 cases qualify as ***stringent*** (bold italics). No cases were found in which the criterion should be classified as unacceptable. Only in 14 cases can the criterion be regarded as mediocre, and these cases occur at rather small significance levels: $\alpha = 0.005$ and $\alpha = 0.01$.

Using Zhang's Acriterion yields somewhat better results. Specifically, it can be considered *liberal* in 199 cases (compared to 189 for the M_k-criterion) and ***stringent*** in 145 cases (compared to 115 for the M_k-criterion).

Another important advantage of Zhang's criterion should be noted – the uniformity of the distribution of its p_{value}.

At the same time, the M_k-criterion, within the interval of the most practically relevant significance levels, demonstrates nearly comparable results, while having the following essential advantages:

1. No statistical modelling is required to determine the distribution function of the statistic for each new number of samples in the segments or each new number of segments in the ensemble.

2. During sequential ensemble growth, it is not necessary to reevaluate the homogeneity of the entire updated ensemble every time. It is sufficient to compare only the new segment with those already included in the ensemble.

These advantages of the M_k-criterion enable selective formation of a segment ensemble from timeseries samples according to the following algorithm.

Algorithm for using the criterion to form an ensemble of homogeneous segments.

Input data:

- $x_1, x_2, \ldots, x_N$ – the sequence of observations;
- n – the number of samples in each segment;
- k – the number of segments in the ensemble;
- α – the significance level;
- N – the total number of observations ($N \gg nk$);

Table 2

Nominal and experimental significance levels

	Nominal significance level, α						
	0.005	0.01	0.05	0.1	0.15	0.2	0.25
$n = 64; k = 4$							
α_M	***0.0052***	***0.0104***	***0.061***	***0.1036***	*0.1804*	*0.1604*	***0.2366***
α_Z	***0.0046***	*0.0072*	***0.0504***	***0.1006***	***0.1518***	***0.2002***	***0.2486***
$n = 64; k = 8$							
α_M	***0.005***	*0.012*	*0.04*	*0.0776*	*0.1276*	***0.2046***	*0.2046*
α_Z	*0.006*	*0.0118*	*0.0558*	***0.0988***	***0.1472***	***0.199***	***0.2442***
$n = 64; k = 16$							
α_M	*0.004*	***0.0108***	*0.0424*	*0.0794*	***0.1376***	*0.1376*	***0.2256***
α_Z	*0.044*	***0.0108***	*0.0584*	***0.1028***	***0.1544***	***0.1988***	***0.2464***
$n = 64; k = 32$							
α_M	0.0088	0.0168	*0.0406*	*0.0706*	*0.1272*	*0.1272*	*0.2186*
α_Z	***0,0054***	*0.0114*	***0.045***	***0.096***	***0.1516***	***0.198***	***0.2508***
$n = 64; k = 64$							
α_M	***0.0045***	***0.0098***	***0.0516***	***0.108***	*0.108*	***0.2006***	*0.2006*
α_Z	***0.0054***	***0.0092***	***0.0508***	***0.1032***	***0.152***	***0.2002***	***0.2502***
$n = 128; k = 4$							
α_M	***0.0054***	***0.0096***	***0.0498***	***0.1008***	***0.1376***	***0.19***	***0.265***
α_Z	*0.0062*	*0.0122*	*0.566*	***0.1086***	***0.1584***	***0.2068***	***0.2554***
$n = 128; k = 8$							
α_M	*0.068*	0.0166	*0.0554*	*0.1226*	*0.175*	***0.175***	*0.2408*
α_Z	*0.006*	*0.0134*	***0.0514***	***0.0944***	***0.1484***	***0.193***	*0.236*
$n = 128; k = 16$							
α_M	*0.0068*	*0.0134*	***0.0544***	*0.128*	***0.128***	***0.1806***	***0.2512***
α_Z	*0.0052*	*0.0112*	***0.0502***	***0.1028***	***0.1568***	***0.2106***	***0.2588***
$n = 128; k = 32$							
α_M	***0.0052***	0.0166	***0.044***	*0.112*	*0.1792*	***0.1792***	***0.2584***
α_Z	***0.004***	***0.0078***	***0.0472***	***0.095***	***0.1462***	***0.1922***	***0.2457***
$n = 128; k = 64$							
α_M	*0.006*	***0.0104***	*0.0602*	***0.0938***	***0.1448***	***0.212***	*0.212*
α_Z	*0.0076*	*0.0146*	***0.0498***	***0.1076***	***0.1554***	***0.206***	***0.2594***
$n = 256; k = 4$							
α_M	*0.0062*	*0.0116*	*0.065*	***0.1084***	***0.1652***	***0.2042***	***0.2426***
α_Z	0.0076	*0.0136*	***0.0478***	***0.1008***	***0.1498***	***0.1952***	*0.2396*
$n = 256; k = 8$							
α_M	*0.006*	*0.012*	*0.0604*	***0.0978***	***0.1652***	***0.2058***	***0.2532***
α_Z	***0.0044***	***0.0102***	***0.0486***	***0.101***	***0.1492***	***0.2074***	***0.2608***
$n = 256; k = 16$							
α_M	*0.007*	*0.0152*	*0.058*	***0.1086***	***0.143***	***0.187***	***0.242***
α_Z	***0.0044***	***0.0102***	***0.044***	***0.0972***	***0.1522***	***0.2024***	***0.2486***
$n = 256; k = 32$							
α_M	0.0088	*0.0138*	*0.058*	***0.1046***	***0.1422***	***0.1868***	***0.2426***
α_Z	*0.0068*	*0.0126*	*0.0618*	*0.1148*	*0.1658*	***0.2174***	***0.2634***
$n = 256; k = 64$							
α_M	*0.006*	*0.0124*	*0.0638*	***0.0872***	***0.1628***	*0.1628*	*0.219*
α_Z	*0.0056*	*0.0136*	***0.0504***	***0.1024***	***0.146***	***0.197***	***0.2516***
$n = 256; k = 128$							
α_M	***0.0046***	***0.0122***	***0.0474***	***0.0988***	***0.1406***	***0.1908***	***0.2514***
α_Z	*0.0068*	*0.0126*	***0.0508***	***0.0984***	***0.1488***	***0.203***	***0.2556***
$n = 512; k = 4$							
α_M	***0.005***	***0.0106***	***0.0536***	***0.0946***	***0.1554***	***0.1852***	***0.2408***
α_Z	*0.0042*	*0.0112*	***0.0506***	***0.1006***	***0.1478***	***0.1964***	***0.2482***
$n = 512; k = 8$							
α_M	*0.0064*	***0.0106***	*0.06*	***0.1046***	***0.1484***	***0.205***	***0.2382***
α_Z	***0.0048***	***0.0102***	***0.048***	***0.0998***	***0.145***	***0.196***	***0.244***
$n = 512; k = 16$							

Table 2 (continued)

	Nominal significance level, α						
	0.005	0.01	0.05	0.1	0.15	0.2	0.25
α_M	*0.0066*	*0.0116*	***0.0552***	***0.1086***	***0.1572***	***0.2194***	***0.2612***
α_Z	*0.007*	*0.0122*	*0.0584*	*0.112*	***0.159***	***0.2082***	***0.2546***
$n = 512$; $k = 32$							
α_M	0.0084	*0.013*	*0.0568*	*0.113*	***0.1394***	***0.2104***	***0.252***
α_Z	*0.0058*	*0.0118*	***0.0528***	*0.093*	***0.1466***	***0.1982***	***0.2522***
$n = 512$; $k = 64$							
α_M	0.0096	0.0174	*0.0622*	*0.1148*	***0.1442***	***0.1808***	*0.2188*
$n = 1{,}024$; $k = 4$							
α_M	*0.007*	*0.0128*	*0.0612*	***0.1122***	***0.158***	***0.2208***	***0.2676***
α_Z	0.009	*0.0154*	***0.0522***	***0.0988***	***0.1454***	***0.1998***	***0.249***
$n = 1{,}024$; $k = 8$							
α_M	0.009	0.0162	*0.0564*	***0.1068***	***0.1586***	***0.2044***	***0.2544***
α_Z	0.0096	0.158	*0.061*	***0.106***	***0.153***	***0.2078***	***0.259***
$N = 1{,}024$; $k = 16$							
α_M	0.0092	*0.0146*	*0.0586*	***0.1044***	***0.163***	***0.2106***	***0.2702***
α_Z	*0.006*	***0.0108***	*0.0568*	***0.105***	***0.1492***	***0.1964***	***0.2452***
$n = 1{,}024$; $k = 32$							
α_M	0.0094	0.0174	*0.068*	***0.1102***	***0.165***	***0.212***	***0.2438***
α_Z	*0.006*	***0.01***	***0.0506***	***0.982***	***0.1508***	***0.1996***	***0.2516***
$n = 1{,}024$; $k = 64$							
α_M	*0.0074*	*0.0148*	*0.067*	***0.114***	***0.1498***	***0.1956***	***0.228***
$n = 2{,}048$; $k = 4$							
α_M	*0.0058*	*0.0114*	*0.059*	*0.1108*	*0.1674*	***0.206***	***0.2554***
α_Z	***0.005***	***0.0108***	***0.0512***	***0.1034***	***0.1512***	***0.1958***	***0.248***
$n = 2{,}048$; $k = 8$							
α_M	0.008	*0.013*	*0.063*	*0.1132*	***0.1562***	***0.2176***	***0.254***
α_Z	***0.0052***	***0.0106***	***0.0518***	***0.1006***	***0.1514***	***0.2012***	***0.2502***
$n = 2{,}048$; $k = 16$							
α_M	0.008	0.0164	*0.0602*	*0.1142*	*0.174*	***0.2106***	***0.2662***
α_Z	*0.0056*	***0.0086***	***0.0514***	***0.1044***	***0.1638***	*0.2246*	*0.2822*
$n = 2{,}048$; $k = 32$							
α_M	*0.0058*	0.013	*0.0564*	***0.1086***	***0.1482***	***0.2024***	***0.2424***
α_Z	0.0076	*0.144*	***0.052***	***0.1036***	***0.1602***	*0.2662*	***0.2662***
$n = 2048$; $k = 64$							
α_M	0.0098	0.0164	***0.0544***	*0.1148*	***0.1508***	***0.1874***	***0.2324***

- $f_S2(s_1, s_2)$ – a function that computes the corrected statistic of the twosample Kolmogorov–Smirnov test for segments s_1 and s_2 (4, 5, 7);

- $f_Pvalue(M, k, n)$ – a function that computes the achieved significance level for the statistic M, obtained from k segments with n samples each (11, 12).

Output data:

- X – a matrix of size $n \times k$, where each column contains the samples of one segment.

Algorithm:

1. Form the first segment from the observations: $s_1 = \{x_1, ..., x_n\}$.
2. Set $M = -\infty$; $I = 1$; $i = 1$; $X(1 : n, 1) = s_1$; $X(1 : n, 2 : k) = 0$.
3. Set $I = I + 1$.
4. If $nI > N$, ***terminate***.
5. Form the segment from n consecutive observations

$$s = \{x_{n(I-1)+1}, ..., x_{nI}\}.$$

6. Compute $S_j = f_S2(s, X(1 : n, j)),\ j = \overline{1,i}$.
7. Compute $S_{\max} = \max\left(\max_{j=1,i} S_j, M\right)$.
8. Compute $p = f_Pvalue(S_{\max}, i + 1, n)$.
9. If $p < \alpha$, then go to step 3.
10. Set $i = i + 1$, $M = S_{\max}$.
11. Insert the segment into the output matrix

$$X(1 : n, i) = s.$$

12. If $i < k$, then go to step 3.
13. Terminate.

Example of using the criterion. The described algorithm was applied to a synthetic time series. The synthetic series was constructed as a sequence of generated normally distributed samples, to which identical additional excitation impulses were added at random uniformly distributed time moments (with intensity λ). The samples were divided into nonoverlapping segments of n samples each. Segments were then grouped into ensembles according to the algorithm presented above. As an evaluation characteristic, a loss coefficient k_{out} was

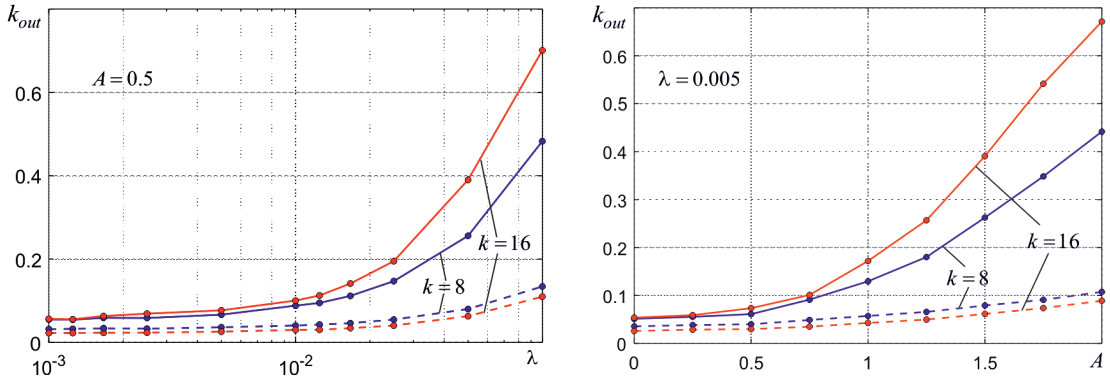

Fig. 2. Loss coefficients as functions of excitation amplitude (a) and excitation intensity (b): dashed lines – M_k-criterion; solid lines – Zhang's Acriterion

used, defined as the ratio of the number of segments that did not enter the constructed ensembles to the total number of segments. At the same time, ensembles were also formed from the same series using Zhang's criterion, and these ensembles were evaluated using the same loss coefficient k_{out}.

During statistical modelling, the number of simulations depended on the rate of stabilization of the value k_{out} and did not exceed 500 simulations. The number of samples in each simulation was 50 kn, which corresponds to a maximum possible number of 50 ensembles.

A single excitation impulse was defined as

$$Ae^{-\beta i}\sin(\varpi i), \quad i = 0, \left\lceil \frac{4\pi}{\sqrt{\varpi^2 - \beta^2}} \right\rceil.$$

where A is the amplitude; ϖ is the frequency; β is the damping coefficient.

Fig. 2 presents the results obtained for the following parameter values: $\varpi = 0.1$, $\beta = 0.2$, $n = 128$, $\alpha = 0.05$. Other parameters are shown in the figure.

Conclusions. The statistic of the criterion was defined as the maximum among all possible twosample statistics that can be constructed from the segments of the ensemble. As the twosample statistic, the statistic of the Kolmogorov–Smirnov homogeneity test was used.

Statistical modelling was used to study the distributions of the proposed statistic for various segment and sample sizes. Based on the considerations presented in this work and on the analysis of the modelling results, a function was proposed that approximates the distribution function of the statistic. The approximation represents the Kolmogorov function raised to a power that depends on the number of samples in the segments and the number of segments in the ensemble. The adequacy of the proposed approximation was assessed by comparing the nominal significance level with the significance levels obtained using the proposed criterion and Zhang's criterion. The comparison results showed that the criterion based on the proposed approximation (referred to as the M_k-criterion) can be used as a k-sample homogeneity test, at least for $n \in [64; 2048]$, $k \in [2; 128]$, and $\alpha \in [0.005; 0.2]$.

At the same time, the advantage of the M_k-criterion lies in its ability to dynamically test the possibility of including each new segment in the ensemble. This eliminates the need to precompute (via statistical modelling) distribution functions of the statistic for all intermediate values of k, as would be required when using known k-sample tests such as Zhang's criterion or the Anderson–Darling criterion.

An example was presented showing the formation of an ensemble of segments for a synthetic signal. The signal was generated as a sequence of normally distributed samples, to which identical impulses were added at random time moments. When ensembles were formed using the M_k-criterion, the number of "lost" segments was several times smaller than when using Zhang's k-sample criterion, which consequently increases the statistical reliability of the estimates obtained from the constructed ensemble.

The proposed criterion is most appropriate for monitoring the state of systems. In many cases, systems may operate in several states that are close to quasistationary. State identification requires statistical processing of the signals generated by the system. This, in turn, demands the availability of a set of statistically similar samples (ensembles of segments) that correspond to a limited time interval (the identification interval) [38, 39].

Random external influences may substantially alter the characteristics of the recorded signals. In such cases, the formal merging of consecutive segments into an ensemble leads to unacceptably large confidence intervals, while rejecting ensembles based on strict statistical requirements results in an undesirably wide identification interval. The use of the Mkcriterion during ensemble formation significantly alleviates this issue by increasing the proportion of segments that can be included in the ensemble (i.e., by reducing the loss coefficient). At the same time, the rejected segments may be analysed separately, including the possibility of treating them as data for constructing a new ensemble.

The conceptual design of the criterion naturally supports both offline and online use.

References.

1. Wu, Y., Zhang, L., Gu, Z., Lu, H., & Wan, S. (2023). Edge-AI-driven framework with efficient mobile network design for facial expression recognition. *ACM Transactions on Embedded Computing Systems, 22*(3), 1-17. https://doi.org/10.1145/3587038

2. Yuan, B., & Zhao, D. (2024). A survey on continual semantic segmentation: Theory, challenge, method and application. *IEEE Trans-*

actions on Pattern Analysis and Machine Intelligence. Retrieved from https://arxiv.org/pdf/2310.14277
3. Lauriola, I., Lavelli, A., & Aiolli, F. (2022). An introduction to deep learning in natural language processing: Models, techniques, and tools. *Neurocomputing, 470*, 443-456. https://doi.org/10.1016/j.neucom.2021.05.103
4. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice. OTexts*. Retrieved from https://otexts.com/fpp2/
5. Plaga, L. S., & Bertsch, V. (2025). Optimization based priority clustering for time series in power system models-improving system robustness and storage deployment. *Sustainable Energy, Grids and Networks,* 101802. https://doi.org/10.2139/ssrn.5135673
6. Abdulsahib, A. K., Balafar, M. A., & Baradarani, A. (2024). DG-BPSO-DBSCAN: An optimized clustering technique based on supervised/unsupervised text representation. *IEEE Access, 12*, 110798-110812.
7. Bisharat, A. (2023). Comparing Algorithm Performance in Machine Learning for Landslide Susceptibility Studies: An Overview. *Journal of Artificial Intelligence Research & Advances*. https://doi.org/10.37591/JOAIRA
8. Babets, D., Opilski, Z., Hnatushenko, V., Kashtan, V., Michalczuk, A., Sdvyzhkova, O., Maciak, E., & Cyran, K.A. (2025). Isolation Forest as a Tool for Entangled Photon Detection. *2025 11th International Conference on Control, Decision and Information Technologies (CoDIT), 1*, 930-935, https://doi.org/10.1109/CoDIT66093.2025.11321541
9. Cao, J., Miao, J., Tao, H., Wang, Y., Wu, J., Wang, Z., & Wu, X. (2025). Generative Models for Time Series Anomaly Detection: A Survey. *IEEE Transactions on Artificial Intelligence*. Retrieved from https://ieeexplore.ieee.org/document/11184472
10. Liang, X., Zhang, L., & Zhao, J. (2025). AT-LSTM-CUSUM Digital Intelligent Model for Seepage Safety Prediction of Concrete Dam. *Structural Control and Health Monitoring, 2025*(1), 8518538.
11. Peerajit, W. (2025). A Novel Numerical Integration Equation Approach using the Gauss-Legendre Quadrature to Approximate the Average Run Length of Time-Series Model Running on a CUSUM Control Chart. *WSEAS Transactions on Systems, 24*, 345-358.
12. Liu, S., Yamada, M., Collier, N., & Sugiyama, M. (2013). Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks, 43*, 72-83. https://doi.org/10.48550/arXiv.1203.0453
13. Xie, L., Moustakides, G. V., & Xie, Y. (2023). Window-limited CUSUM for sequential change detection. *IEEE Transactions on Information Theory, 69*(9), 5990-6005. https://doi.org/10.48550/arXiv.2206.06777
14. Heinrichs, F. (2025). Self-Normalization for CUSUM-based Change Detection in Locally Stationary Time Series. *arXiv preprint arXiv* :2509.07112. https://doi.org/10.48550/arXiv.2509.07112
15. Lopes, J. F., Barbon Junior, S., & de Melo, L. F. (2025). Online Meta-Recommendation of CUSUM Hyperparameters for Enhanced Drift Detection. *Sensors*, *25*(9), 2787. https://doi.org/10.3390/s25092787
16. Liang, Y., Tartakovsky, A. G., & Veeravalli, V. V. (2022). Quickest change detection with non-stationary post-change observations. *IEEE Transactions on Information Theory, 69*(5), 3400-3414.
17. Wang, H., & Xie, Y. (2023). Sequential change-point detection: Computation versus statistical performance. *Wiley Interdisciplinary Reviews: Computational Statistics, 16*(1), e1628. https://doi.org/10.48550/arXiv.2210.05181
18. Kuncheva, L. I., & Faithfull, W. J. (2013). PCA feature extraction for change detection in multidimensional unlabeled data. *IEEE transactions on neural networks and learning systems, 25*(1), 69-80.
19. Faithfull, W. J. (2018). *Unsupervised change detection in multivariate streaming data. Bangor University (United Kingdom)*. Retrieved from https://www.researchgate.net/publication/328580385_Unsupervised_Change_Detection_in_Multivariate_Streaming_Data
20. Yamada, M., Kimura, A., Naya, F., & Sawada, H. (2013, August). Change-Point Detection with Feature Selection in High-Dimensional Time-Series Data. *IJCAI, 13*, 1827-1833.
21. Ermshaus, A., Schäfer, P., & Leser, U. (2023). ClaSP: parameter-free time series segmentation. *Data Mining and Knowledge Discovery, 37*, 1262-1300 https://doi.org/10.1007/s10618-023-00923-x
22. Tsai, YHH., Wu Y., Salakhutdinov, R., & Morency, L.-P. (2020). Demystifying self-supervised learning: an information theoretical framework. *arXiv preprint*. https://doi.org/10.48550/arXiv.2006.05576
23. Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM, 55*(10), 78-87. https://doi.org/10.1145/2347736.2347755
24. Aliferis, C., & Simon, G. (2024). Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and AI. *Artificial intelligence and machine learning in health care and medical sciences*, 477-524. https://doi.org/10.1007/978-3-031-39355-6_10
25. Northcutt, C., Athalye, A., & Mueller, J. (2021). Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks. *arXiv.*2103.14749. https://doi.org/10.48550/arXiv.2103.14749
26. Gong, Y., Liu, G., Xue, Yu., Li, R., & Meng, L. (2023). A survey on dataset quality in machine learning. *Information and Software Technology, 162*, 107268, https://doi.org/10.1016/j.infsof.2023.107268.
27. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, *1*(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x
28. Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, *627*(8002), 49-58. https://doi.org/10.1038/s41586-024-07146-0
29. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one, 13*(3), e0194889. https://doi.org/10.1371/journal.pone.0194889
30. Cerqueira, V., Torgo, L., & Soares, C. (2019). Machine learning vs statistical methods for time series forecasting: Size matters. *arXiv preprint arXiv*:1909.13316. https://doi.org/10.48550/arXiv.1909.13316
31. Shumway, R. H., & Stoffer, D. S. (2016). *Time series analysis and applications*. Retrieved from https://swmprats.net/stoffer/tsa4/tsaEZ.pdf
32. Afriyie, J. K., Twumasi-Ankrah, S., Gyamfi, K. B., Arthur, D., & Pels, W. A. (2020). Evaluating the performance of unit root tests in single time series processes. *Mathematics and Statistics, 8*(6), 656-664. https://doi.org/10.13189/MS.2020.080605
33. Herranz, E. (2017). Unit root tests. *Wiley Interdisciplinary Reviews: Computational Statistics, 9*(3), e1396. https://doi.org/10.1002/wics.1396
34. Smeeton, N., Spencer, N. H., & Sprent, P. (2025). *Applied nonparametric statistical methods*. CRC press. https://doi.org/10.1201/9780429326172
35. Stanley, R. P. (2012). *Enumerative Combinatorics: Volume 1* (*2nd* ed.). Cambridge University Press, ISBN:978-1-107-60262-5.
36. Rolke, W. (2024). Simulation Studies For Goodness-of-Fit and Two-Sample Methods For Univariate Data. *arXiv preprint arXiv*: 2411.05839. https://doi.org/10.48550/arXiv.2411.05839
37. Cui, X., Geng, H., Wang, G., Wang, Z., & Zou, C. (2025). ART: Distribution-Free and Model-Agnostic Changepoint Detection with Finite-Sample Guarantees. *arXiv preprint arXiv*: 2501.04475. https://doi.org/10.48550/arXiv.2501.04475
38. Holovko Yu. (2023). Spectral estimation of a broadband time-limited noise signal. *Scientific Journal "Mathematical Modeling", 2*(49), 86-97. https://doi.org/10.31319/2519-8106.2(49)2023.292638
39. Golovko, Yu., & Sdvyzhkova, O. (2024). *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, (4), 114-120. https://doi.org/10.33271/nvngu/2024-4/114

Непараметричний критерій однорідності для селективного формування ансамблю сегментів часового ряду

*Ю. М. Головко**

Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна
* Автор-кореспондент e-mail: holovko.yu.m@nmu.one

Мета. Розробка критерію перевірки однорідності сегментів часового ряду, які послідовно об'єднуються в ансамбль сегментів, що є передумовою отримання статистично значущих оцінок ряду.

Методика. Основою дослідження є ймовірнісний аналіз і статистичне моделювання. Значну частину дослідження складає доведення адекватності запропонованого критерію, що включає співставлення й демонстрацію переваг відносно відомого критерію. Статистика запропонованого критерію базується на статистиці двовибіркового критерію однорідності Колмогорова-Смірнова.

Результати. Показано, що при застосуванні запропонованого критерію для перевірки однорідності ансамблю із фіксованою кількістю сегментів, цей критерій дає результати близькі до результатів відомого критерію Жанга. Переваги запропонованого критерію виявляються при послідовному оцінюванні можливості включення сегментів у ансамбль. У такому випадку кількість «втрачених» сегментів може буте зменшена в рази. Область використання критерію обмежена наведеними досить широкими інтервалами параметрів, що приймалися при статистичному моделюванні.

Наукова новизна. Розроблено новий критерій формування ансамблю однорідних сегментів часового ряду. Критерій не вимагає проведення статистичного моделювання для визначення функцій розподілу статистики критерію при зміні кількості відліків у сегментах і кількості сегментів у ансамблі.

Практична значимість. Використання запропонованого критерію може забезпечити суттєве підвищення статистичної значущості оцінок стану процесу, що представлений у вигляді рівномірного часового ряду.

Ключові слова: *непараметричний критерій однорідності, часовий ряд, ансамбль сегментів, випадковий процес*