A Method for Signal Stationarity Estimation

Hristo ZHIVOMIROV
Department of Theory of Electrical Engineering and Measurements, Technical University of Varna, Bulgaria, hristo_car@abv.bg

Ivaylo NEDELCHEV
Department of Theory of Electrical Engineering and Measurements, Technical University of Varna, Bulgaria, iynedelchev@abv.bg

Abstract: - In the paper, a novel estimation method is proposed, concerning the wide-sense stationarity test of the signals. Background information is given about the concept of stationarity of the processes and signals. The problem of the signal stationarity estimation is addressed along with criticism of the available stationarity tests. Further, a new wide-sense stationarity estimation method is described, involving estimation of the mean-, variance- and autocovariance- stationarity of a signal. Finally, a few representative signals are tested and the results clearly indicate the consistence of the proposed test method. It is implemented in Matlab®-environment and can be download and use for free.

Keywords: - signal, stationarity, estimation, test, method

1. INTRODUCTION

The estimation whether a given process is stationary or not is a classic problem in the statistics, econometrics, signal processing *etc*. Multiple tests exist for this purpose (*e.g.*, Kwiatkowski–Phillips–Schmidt–Shin test [1], Dickey–Fuller test [1], Phillips–Perron test [2]), but they are somehow restricted to the hypothesis of either trend-stationarity or unit-root non-stationarity. However, it is possible a process to be deterministic and non-stationary and yet has no single unit-root (*e.g.*, linear frequency-modulated process), and so the different statistical tests could give contradictory results.

In the applied signal processing one is dealing primary with signals instead of the corresponding underlying processes, so the basic consideration is whether a signal is stationary or not. Moreover, the signal is considered "as it is" and hence the inferential statistical analysis is not of use.

In order to overcome the above, we propose a novel approach for signal stationarity estimation based on the assumption that if the signal is wide-sense stationary then the local statistical properties up to the second-order estimated for different parts of the signal are going to be approximately equal.

2. BACKGROUND

Let's consider a discrete or sampled and quantized continuous *a-priory* unknown process or phenomenon $\left\{X(s,t): s \in S \in \mathbb{R}, t \in T \in \mathbb{R}\right\}$ which is an object of observation (measurement) by means of an experiment. Without loss of generality one can assume that the process takes place in the time-domain *T*. One may define all possible results of the experiment as a sample space *S* [1], such that

$$S = \left\{s_1, s_2, \ldots, s_r, \ldots, s_R\right\}, \qquad (1)$$

where s_r is a possible outcome of the measurement experiment. All members of *S* (the aggregate observations) are named population [1].

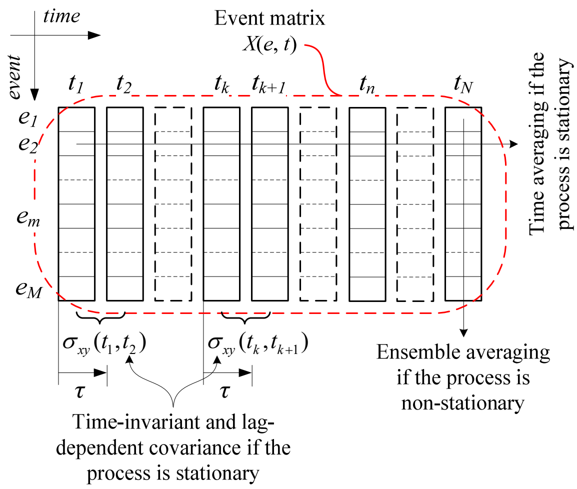

Figure 1. An event matrix $X(e,t)$ along with graphical representation of some related statistical concepts concerning its interpretation and processing.

For the time of observation (measurement) a set of outcome results (events) are collected termed event space $E \in S$. The last is actually consisted of observations (events) arranged in an event matrix $X(e,t) \in \mathbb{R}^{M \times N} : M, N \in \mathbb{N}$ as is shown in Fig. 1 (*e.g.*, *N* parallel measurements of body temperature in *M* points). For fixed event e_m, $X_e(t)$ is a time series $x(t)$ and for fixed time t_n, $X_t(e)$ is an ensemble of realizations $x(e)$.

Multiple statistical properties could be assigned to the event matrix $X(e,t)$ but of particular interest in this paper are the first and second order statistical moments – the mean and the covariance of the process (the last one includes also the variance).

In general, most of the processes are not fully cognizable since one cannot obtain all possible realizations of a given process *i.e.*, the sample space *S* and the corresponding population are not fully available. In such case only the subset of the observed (measured) events *E* is of use and hence one could estimate only the parameters of the process itself via sample (empirical) mean $\hat{\mu}_x$ and sample cross-covariance $\hat{\sigma}_{xy}$ of the acquired data. In the common case, $\hat{\mu}_x(t)$ and $\hat{\sigma}_{xy}(t)$ are functions of the time and they are calculated over ensembles of realizations (*i.e.*, column-wise on Fig. 1) [3, 4]:

$$\hat{\mu}_x(t_n) \triangleq \frac{1}{M}\sum_{m=1}^{M} x(e_m, t_n), \quad (2)$$

$$\hat{\sigma}_{xy}(t_k, t_n) \triangleq \frac{1}{M}\sum_{m=1}^{M} x(e_m, t_k)\, x(e_m, t_n). \quad (3)$$

A special case of process is one for which [3, 4]:

$$\hat{\mu}_x(t_n) \equiv \hat{\mu}_x = const., \forall n, \quad (4)$$

$$\hat{\sigma}_{xy}(t_k, t_n) \equiv \hat{\sigma}_{xy}(\tau) : \tau = t_n - t_k, \forall k, n, \quad (5)$$

so that $\hat{\mu}_x$ is not function of time *t* (*i.e.*, it is time-invariant), and $\hat{\sigma}_{xy}(\tau)$ is a function only of the time-shift τ between the considered ensembles of realizations. This type of process is termed a wide-sense (also weakly, covariance or second-order) stationary (WSS) process, which means that its statistical properties up to the second-order do not vary over time [3, 4].

If the process is WSS it is possible to evaluate the sample mean and cross-covariance using only two ensembles of realizations (*e.g.*, $X_{t_1}(e)$ and $X_{t_2}(e)$, *cf.* Fig. 1).

Further, if the WSS process meets the following conditions [4, 5]:

$$\hat{\mu}_x(t_n) \equiv \hat{\mu}_x(e_m) \equiv \hat{\mu}_x, \quad (6)$$

$$\hat{\sigma}_{xy}(\tau) \equiv \hat{\sigma}_{xx}(\tau), \quad (7)$$

and more precisely

$$\frac{1}{M}\sum_{m=1}^{M} x(e_m, t_n) = \frac{1}{N}\sum_{n=1}^{N} x(e_m, t_n) = const., \forall n, m, \quad (8)$$

$$\begin{aligned} &\frac{1}{M}\sum_{m=1}^{M} x(e_m, t_n)\, x(e_m, t_n + \tau) = \\ &= \frac{1}{N}\sum_{n=1}^{N-h} x(e_m, t_n)\, x(e_m, t_n + \tau), \forall n, m \end{aligned}, \quad (9)$$

then it is called to be an ergodic process. When the process is ergodic one can estimate the mean $\hat{\mu}_x$ and the autocovariance $\hat{\sigma}_{xx}(\tau)$ (instead the cross-covariance $\hat{\sigma}_{xy}(t_k, t_n)$) using only one time sequence of the event matrix $X(e,t)$ (*e.g.*, $X_{e_1}(t)$, *cf.* Fig. 1) and time averaging instead of ensemble averaging. However, in order to proof the stationarity and eventually the ergodicity of the process one must examine the whole event matrix $X(e,t)$.

Here must be noted that the autocovariance function (ACvF) $\hat{\sigma}_{xx}(\tau)$ is actually an unscaled variant of the autocorrelation function (ACrF) $\hat{\rho}_{xx}(\tau)$ of the process [6].

3. PROBLEM STATEMENT

Let's consider a discrete or sampled and quantized (according the Nyquist–Shannon sampling theorem) continuous physically realizable and observable signal (*e.g.*, time series)

$$x[nT_s], \quad (10)$$

where $n = \{1, ..., N\} \in \mathbb{N}$ is the sample number; N – number of all samples; $T_s = 1/f_s$ – sampling time interval and f_s is the sampling frequency. One must be aware that in this case there is no event matrix $X(e,t)$ but only one time series and so Eqs. (2) ÷ (5) cannot be used. In this light, no meaningful decision could be drawn about the stationarity and ergodicity of the underling process $X(s,t)$.

From the applied signal processing point of view, the main question is "Does the signal $x[nT_s]$ itself is WSS?". A signal is said to be WSS if a time shift does not affects its local statistical properties up to the second-order [4], or which is the same – the local spectral content of the signal not vary by time [7], as far as the ACrF (and hence the ACvF) and the power spectrum density (PSD) of a given signal are bounded by the Wiener–Khinchin theorem.

In the practice, the stationarity of a signal concerns actually its DC-value, RMS-value and its spectral content (by the ACvF/ACrF). Here must be noted that for the signals the ergodicity property is not defined at all.

This problem is not trivial since it directly affects the proper choice of the signal processing techniques used in the post-processing of the signal. For instance, the PSD could be calculated using the ACrF or by Bartlett's or Welch's methods only if the signal is WSS. Otherwise, a proper time-frequency analysis technique should be used (*e.g.*, Short-time Fourier transform) [7].

Moreover, the standard statistical tests for time-series stationarity tend to fail when they are applied on some types of signals. As an example, in case of operation with a linear chirp (*i.e.*, linear FM) signal, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test fails to reject the null-hypothesis that the signal is trend-stationary, although the chirp is a classic non-stationary signal. This failure persists when Phillips–Perron (PP) or augmented Dickey–Fuller (ADF) tests for a unit-root is used *i.e.*, all three test undoubtedly but erroneously indicate that the chirp signal is stationary.

Another way of estimation is to plot the signal oscillogram or spectrogram and perform a visual inspection (Eyeball test) on them. Most of the non-stationarities could be detected by this procedure (however, the actually non-stationary red noise signal appears to be stationary in the time-frequency domain) but when a massive number of signals are under test, this approach is not relevant.

Obviously, there is need of another approach when a determination of a signal stationarity is a must.

4. PROPOSED SOLLUTION

We proposed a novel solution of the problem, based on the definition for signal stationarity. If the signal is really covariance stationary, then the local statistical properties up to the second-order estimated for different parts of the signal are going to be approximately equal. Here must be noted that one does not try to make any decisions about the underlying process itself.

The estimation algorithm is shown in Fig. 2 and contains:

1) Signal detrending in order to ensure that any non-stationary behavior due to a linear trend is avoided. So actually, the algorithm checks if the signal is trend-stationary *i.e.*, if an underlying trend can be removed, leaving a stationary signal [1]. This does not compromise the practical implementation of the stationarity estimation, since the signal detrending is advisable in the common case;

2) Splitting of the signal $x[n]$ on two equal-length parts $x_1[n]$ and $x_2[n]$ (*cf.* Fig. 3), for the sake of comparison of the statistical properties of interest regarding these signals;

3) Estimation of the signal stationarity about its (i) mean, (ii) variance and (iii) autocovariance. If these properties are approximately similar for the two partial signals $x_1[n]$ and $x_2[n]$, then the whole signal $x[n]$ is estimated as stationary.

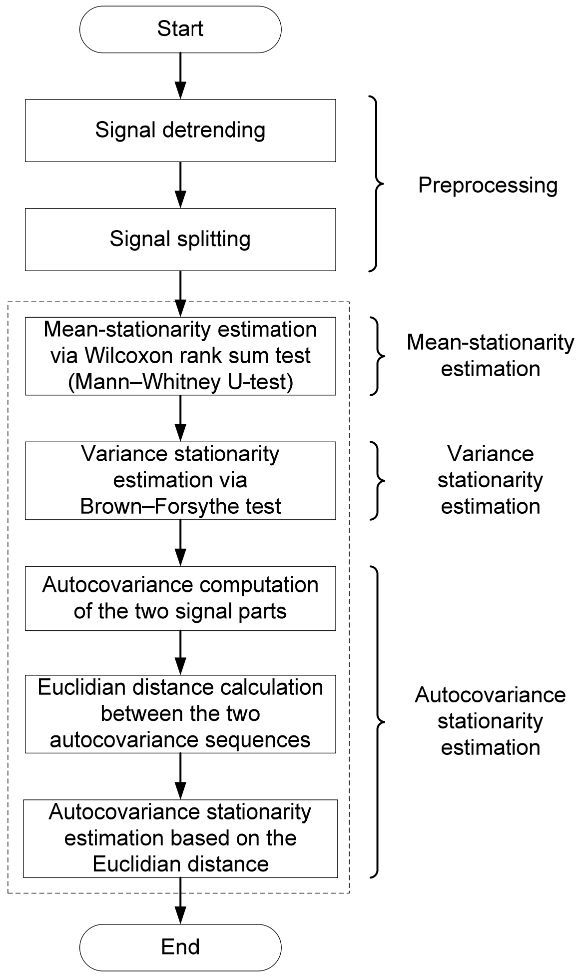

Figure 2. The algorithm of the proposed WSS estimation method. Three different parameters are under stationarity test – the signal mean, variance and autocovariance.

Further in the paper, without loss of generality one assumes:

$$x[nT_s] \to x[n];\ t_n = nT_s \to n;\ \tau = hT_s \to h\ , \quad (11)$$

in order not to be restricted to the time-domain only (*e.g.*, the signal could be spatial-related).

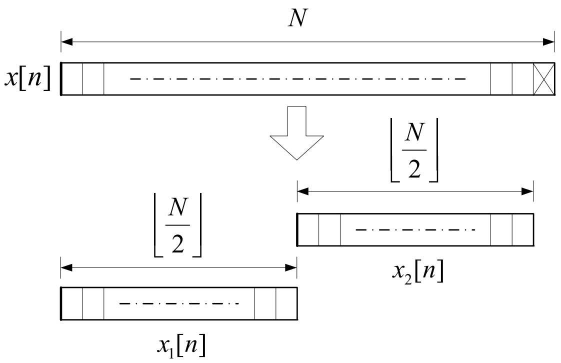

Figure 3. Signal splitting for comparison of the mean, variance and autocovariance of the two parts. The last sample is discharged if the overall length N is odd.

For the sake of the WSS-estimation we proposed the following modified formulae, based on the classic ones in [5], for calculation of the local mean, variance and ACvF of the signal $x[n]$:

$$\hat{\mu}_x(Q,P) \triangleq \frac{1}{Q-P+1}\sum_{n=P}^{Q} x[n], \tag{12}$$

$$\hat{\sigma}_x^2(Q,P) \triangleq \frac{1}{Q-P+1}\sum_{n=P}^{Q} x[n]^2, \tag{13}$$

$$\hat{\sigma}_{xx}(Q,P,h) \triangleq \frac{1}{(Q-h)-P+1}\sum_{n=P}^{Q-h} x[n]x[n+h], \tag{14}$$

where P and Q are the indexes of the first and last sample of the considered signal segment, respectively and h is the current hop size (shift) when the ACvF is calculated. Equation (14) is actually the discrete form of the Eq. (1) given at [8], and Eqs. (12) and (13) are derived similarly.

Based on the above equations, the WSS conditions could be written as:

$$\hat{\mu}_x \triangleq \frac{1}{Q-P+1}\sum_{n=P}^{Q} x[n] = const., \forall Q, P, \tag{15}$$

$$\hat{\sigma}_{xx}(h) \triangleq \frac{1}{(Q-h)-P+1}\sum_{n=P}^{Q-h} x[n]x[n+h] = = const., \forall Q, P, \tag{16}$$

$$\hat{\sigma}_x^2 \triangleq \frac{1}{Q-P+1}\sum_{n=P}^{Q} x[n]^2 = const., \forall Q, P, \tag{17}$$

where the following inequality must be met:

$$0 \le P < Q-h \le Q \le N. \tag{18}$$

Equations (15) ÷ (17) are newly proposed, but a similar idea is also given in [4].

For the split signal $x[n]$, we set $P_1 = 1$, $Q_1 = \lfloor N/2 \rfloor$ and $P_2 = \lfloor N/2 \rfloor + 1$, $Q_2 = 2\lfloor N/2 \rfloor$, $h = 0, \ldots, \lfloor N/2 \rfloor - 1$ (*cf.* Fig. 4). In this manner we will check whether the ACvF is only time-shift-dependent and independent of the particular location defined by P and Q, and also if the two partial signals $x_1[n]$ and $x_2[n]$ obtain equal mean and variance.

The test of the separate WSS conditions (Eqs. (15) and (16)) allows not only quantity but also quality WSS-estimation of a given signal.

4.1. Mean-stationarity check

For comparison of the statistical locations (mean, median, mode) of the two partial signals $x_1[n]$ and $x_2[n]$ a special test is used – Wilcoxon rank-sum test (*a.k.a.* Mann–Whitney U test) [9, 10] which is a non-parametric distribution-free test, instead of the classic Student's t-test, since the last is appropriate choice only for samples with Normal probability density function (PDF) [1]. The test is used for assessing whether the two signals $x_1[n]$ and $x_2[n]$ come from PDFs with one and the same location parameters, in this particular case – mean. Hence, it is a location test for equality of mean, with null hypothesis that the two samples come from distributions with equal locations.

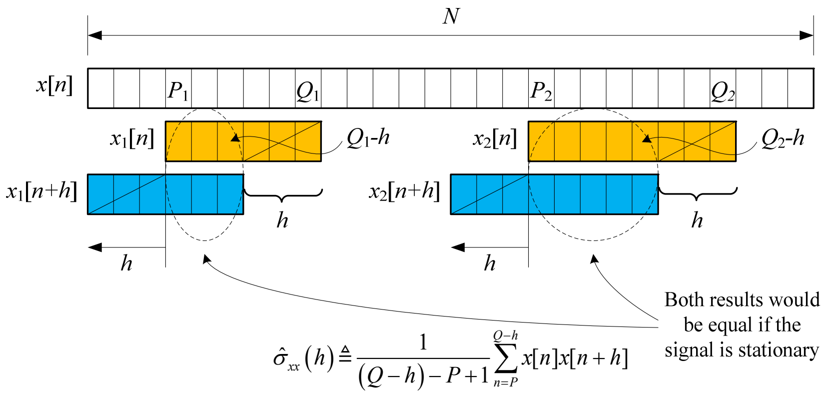

Figure 4. Graphical representation of the idea for local ACvF calculation and comparison.

As is described earlier, this mean-stationary check is actually a trend-stationary test, because of the initial detrending of the signal under test.

The mean-stationarity check test is implemented in the Matlab®-environment via the built-in function *ranksum* [11].

4.2. Variance stationarity check

As a dispersion test for equality of variances the Brown–Forsythe test is used. The test assesses whether the two signals $x_1[n]$ and $x_2[n]$ come from PDFs with equal variances, based on applying of a specific F-test [12]. The test has been preferred instead of the Bartlett's test, since the last is more sensitive to non-normality of the PDFs [1]. The null hypothesis is that the populations obtain same variances.

The test is implemented in the Matlab®-environment via the built-in function *vartestn* [11].

4.3. Autocovariance stationarity check

In order to check the hypothesis that the two signals $x_1[n]$ and $x_2[n]$ obtain nearly identical ACvFs we compare them using the similarity check based on their Euclidian distance (ED).

In Cartesian coordinates, if $a=(a_1,a_2,...,a_N)$ and $b=(b_1,b_2,...,b_N)$ are two points in Euclidean N-space, the distance from a to b is given by the L^2 -norm *i.e.*, the ED is calculated as [1]

$$Ed(a,b)=\|a-b\|_2=\sqrt{\sum_n (a_n-b_n)^2}\ . \quad (19)$$

In the particular case:

$$Ed(\hat{\sigma}_{x_1x_1},\hat{\sigma}_{x_2x_2})=\sqrt{\sum_{l=1}^{2\lfloor N/2\rfloor-1}\left(\hat{\sigma}_{x_1x_1}[l]-\hat{\sigma}_{x_2x_2}[l]\right)^2}\ . \quad (20)$$

If the ED is small enough, then the two ACvF sequences are nearly identical, and hence the overall signal's ACvF could be considered time-invariant.

Table 1. Mathematical description of some representative types of signals for test purposes.

Test signal	Signal type	Signal description
TS1	Sine-wave (monocomponent trend-stationary) $x(t)=a+b\cdot t+U_{m0}\sin(2\pi f_0 t)$	$x(t)=1.0+0.1\cdot t+1.0\sin(2\pi 440t)$ $t=\{0,...,5\}$
TS2	White noise $(PSD\sim const.)$ [14] (multicomponent stationary) $x(t)=\mathcal{WN}(\mu,\sigma^2)$	$x(t)=\mathcal{WN}(0,1)$ $t=\{0,...,5\}$
TS3	Violet noise $(PSD\sim f^2)$ [14] (multicomponent stationary) $x(t)=\mathcal{VN}(\mu,\sigma^2)$	$x(t)=\mathcal{VN}(0,1)$ $t=\{0,...,5\}$
TS4	Linear chirp (monocomponent non-stationary) $x(t)=U_m\sin\left(2\pi f_1 t+\pi\frac{f_2}{T}t^2\right)$	$x(t)=1.0\sin\left(2\pi 1000t+\pi\frac{10000}{5}t^2\right)$ $t=\{0,...,5\}$
TS5a TS5b	Sequence of sine-waves (multicomponent non-stationary) $x(t)=\begin{cases}U_{m1}\sin(2\pi f_1 t),\ 0\le t<t_1\\U_{m2}\sin(2\pi f_2 t),\ t_1\le t<t_2\\...\end{cases}$	$x(t)=\begin{cases}1.0\sin(2\pi 440t),\ 0\le t<1\\2.0\sin(2\pi 440t),\ 1\le t\le 5\end{cases}$ $x(t)=\begin{cases}1.0\sin(2\pi 440t),\ 0\le t<1\\1.0\sin(2\pi 1000t),\ 1\le t\le 5\end{cases}$
TS6	Red noise $(PSD\sim f^{-2})$ [14] (multicomponent non-stationary) $x(t)=\mathcal{RN}(\mu,\sigma^2)$	$x(t)=\mathcal{RN}(0,1)$ $t=\{0,...,5\}$
TS7	Human speech (multicomponent non-stationary)	Record "DR2_FRAM1_SI522" from the TIMIT database [15]
TS8	Music (multicomponent non-stationary)	Built-in Matlab audio sample of the "Hallelujah chorus" from the Handel's "Messiah"

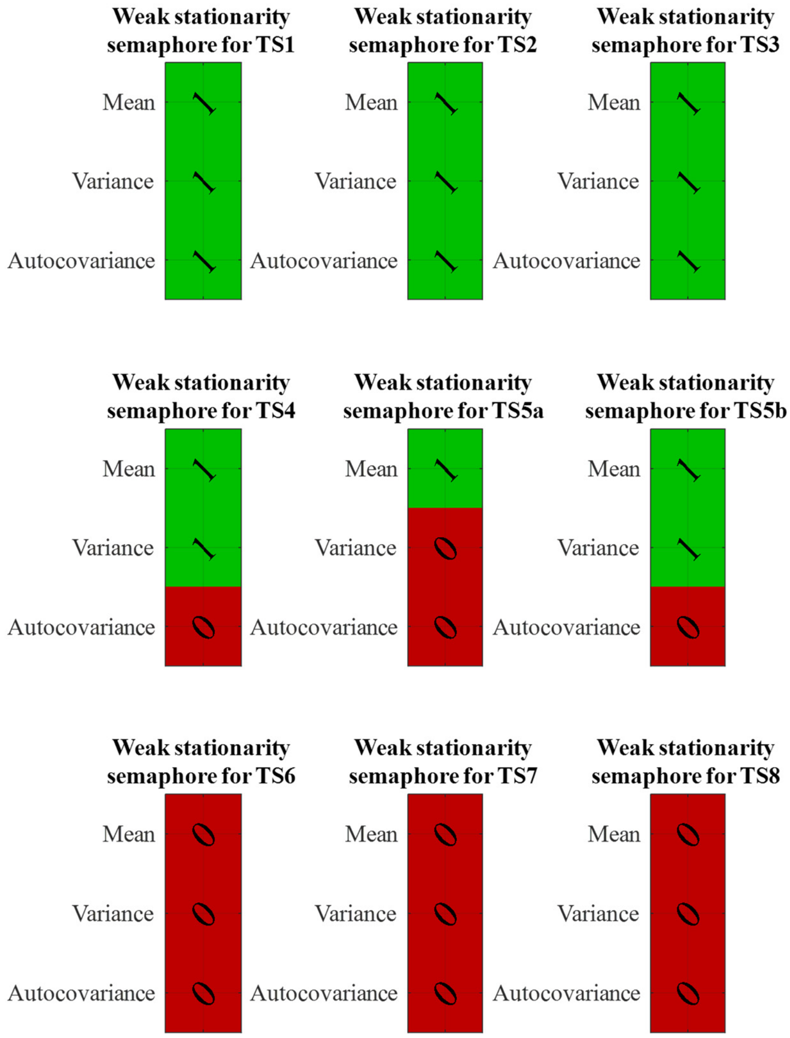

Figure 5. Simulation results for different representative type of signals under test for stationarity. The mathematical descriptions of test signals and their properties are given in Tab. 1. TS1 – sine wave with linear trend; TS2 – white noise; TS3 – violet noise; TS4 – Linear chirp; TS5 – sequence of two sine waves (a) with different RMS-values and (b) with different frequencies; TS6 – red noise; TS7 – sample of human speech; TS8 – music sample.

5. SIMULATION RESULTS

The proposed signal stationarity estimation method is implemented in the Matlab® software environment as a custom-made function and a few computer simulations were performed in order to examine the performance of the method.

The accessibility of the Matlab® simulation files [13] allows the presented experimental results to be reproduced by other researchers for sake of comparison, validation and further development.

Further, for the Wilcoxon rank-sum test and the Brown–Forsythe test a significance level of $\alpha = 0.05$ is chosen. Finally, the proposed stationarity test is applied on all test signals and the results are shown on Fig. 5. As one can see, the results are in perfect agreement with the actual state. It is interesting to note that for all non-stationary signal types the KPSS-, PP- and ADF tests show erroneous positive stationary decision, which confirms the need for new type of signal stationarity estimation method.

Table 2. Euclidian distances between the autocovariance functions of the two halves of the split test signal.

Test signal	Stationary	Autocovariance ED
TS1	yes	$\cong 10.10^{-14}$
TS2		$\sqrt{2}$
TS3		$\cong 2$
TS4	no	$\cong 3.15$
TS5a		$\cong 19$
TS5b		$\cong 123$
TS6		$\cong 91.5$
TS7		$\cong 5.5$
TS8		$\cong 6$

For better visualization and interpretation of the results, an original graphic representation named

"Stationary semaphore" is proposed, as one can see on Fig. 5. The idea is to mimic the street traffic light, with three light sections – for the mean stationary flag, for the variance stationary flag and one for the autocovariance. The signal is determined as WSS if all flags are raised *i.e.*, if there is full "green light" on the semaphore (*cf.* TS1, TS2 and TS3 in Fig. 5). When the signal under test is highly non-stationary, all three sections are in red (*cf.* TS6, TS7 and TS8 in Fig. 5).

The results of simulations clearly indicate the consistence of the proposed WSS-estimation test.

6. CONCLUSIONS

In the paper we presented a novel method for signal wide-sense stationarity estimation. It is based on the fact that if the signal under consideration is covariance stationary, then the local statistical properties up to the second-order estimated for different locations of the signal are going to be approximately equal. One must note that the method does not address the underlying process itself, but concerning only the corresponding signal.

The estimation is based on splitting of the signal on two equal-length partial signals for comparison of their statistical properties of interest: mean, variance and autocovariance. If these properties are approximately equal or similar, then the initial signal is estimated as stationary. The comparisons are conducted via Wilcoxon rank-sum test for equality of means, Brown–Forsythe test for equality of variances and ACvFs comparison based on Euclidian distance, respectively. For experimental purposes, several representative stationary and non-stationary signals are generated and tested for stationarity and the results clearly indicate the consistence of the test method. Also, for better visualization and interpretation of the test results, an original graphic image named "Stationary semaphore" is proposed, which is quite intuitive.

The proposed method for signal stationary estimation (including trend-stationarity) is a new contribution to the signal estimation theory. The possible applications are in the fields of statistical signal processing, measurement and instrumentation, econometrics, biomedicine *etc*.

REFERENCES

[1] Everitt, B., Skrondal, A., *The Cambridge Dictionary of Statistics,* Cambridge University Press, 2010.

[2] Phillips, P., Perron, P., Testing for a Unit Root in Time Series Regression, *Biometrika*, Vol. 75, No. 2, 1988, pp. 335-346. doi:10.1093/biomet/75.2.335.

[3] Woodward, W., Gray, H., Elliott, A., *Applied time series analysis, with R*, Taylor & Francis, 2017.

[4] Jan, J., *Digital Signal Filtering, Analysis and Restoration*, The Institution of Engineering and Technology, 2000.

[5] Manolakis, D., Ingle, V., *Applied Digital Signal Processing*, Cambridge University Press, 2011.

[6] van Etten, W., *Introduction to Random Signals and Noise*, John Wiley & Sons Ltd., 2005.

[7] Boashash, B., *Time Frequency Signal Analysis and Processing: A Comprehensive Reference*, Elsevier, 2015.

[8] Steinberg, I., On the time reversal of noise signals, *Biophysical Journal*, Vol. 50, No. 1, 1986, pp. 171-179. doi:10.1016/S0006-3495(86)83449-X.

[9] Wilcoxon, F., Individual comparisons by ranking methods, *Biometrics Bulletin*, Vol. 1, No. 6, 1945, pp. 80-83. doi:10.2307/3001968.

[10] Mann, H., Whitney, D., On a Test of whether one of Two Random Variables is Stochastically Larger than the Other, *Annals of Mathematical Statistics*, Vol. 18, No. 1, 1947, pp. 50-60. doi:10.1214/aoms/1177730491.

[11] *Matlab R2019b Statistics and Machine Learning Toolbox User's Guide*, The MathWorks Inc., 2019.

[12] Brown, M., Forsythe, B., Robust tests for the equality of variances, *Journal of the American Statistical Association*, Vol. 69, 1974, pp. 364-367. doi:10.1080/01621459.1974.10482955.

[13] Zhivomirov, H., Signal Stationarity Estimation with Matlab, Online at: https://www.mathworks.com/matlabcentral/fileexchange/75118-stationarity-estimation-of-a-signal-with-matlab, Last accessed on September 30th, 2020.

[14] Zhivomirov, H., A Method for Colored Noise Generation, *Romanian Journal of Acoustics and Vibration*, Vol. XV, No. 1, 2018, pp. 14-19.

[15] Garofolo, J. et al., *TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1*, Linguistic Data Consortium, 1993.