ROBUST TESTING FOR STATIONARITY IN THE PRESENCE OF OUTLIERS

Jack Dagdagan, Michael Muma, and Abdelhak M. Zoubir

Signal Processing Group
Technische Universität Darmstadt
Merckstraße 25, 64283 Darmstadt, Germany
Email: {dagdagan, muma, zoubir}@spg.tu-darmstadt.de

ABSTRACT

Testing the stationarity of stochastic processes is required in a variety of signal processing applications. When dealing with real-world problems, the presence of outliers and impulsive (heavy-tailed) noise causes classical stationarity tests to break down. In this work, a set of robust stationarity tests that are based on a sphericity statistic test (SST) in the frequency domain is proposed. Different possible approaches are investigated and compared to existing robust and non-robust stationarity tests in terms of the receiver operating characteristic (ROC). In addition to extensive simulations, a real-world data example of a malfunctioning window regulator motor, for which the dominant frequencies show a modulating character that results in a non-stationary signal, is investigated. Both for simulated and real-world data, the proposed methods significantly outperform existing approaches.

***Index Terms*—** stationarity, KPSS, SST, robustness, M-estimator, hypothesis testing.

1. INTRODUCTION

Stationarity is an important property of stochastic processes and testing for stationarity is required in many signal processing applications. For example, in automotive industry, an indication of malfunction of window regulator motors is the modulating character of the frequencies that contain most of the sound power. This modulating character results in a non-stationary signal. In the literature, to test for the stationarity of a stochastic process, there exist two important definitions of stationarity: *strict stationarity*, which constrains all statistical properties to be time invariant. This, in practice, is too difficult to test for. *second-order stationarity* or *wide-sense stationarity* (WSS), which implies the first and second order moments and the covariance function to exist and to be time-invariant.

Different methods have been developed which test for WSS [1–3]. An early contribution that is based on testing for a unit root was given in [1]. However, the power and size of this test depend on the assumed model of the process of interest. Testing for the null hypothesis of stationarity against the alternative of a unit root [2], which overcomes this problem, has become a standard approach. A new stationarity test was presented in [3]. It is based on a sphericity statistic in the frequency domain (SST) that compares favorably to [2] for non-stationary processes other than the random walk.

In real-world scenarios, measurements are often made in a harsh environment, e.g., in manufacturing halls, which produces heavy-tailed noise in the measurements. The presence of outliers and impulsive (heavy-tailed) noise has also been reported in many other important signal processing applications [4–9] and can cause classical stationarity tests to break down. Recently, robust stationarity tests have been proposed [10, 11] that are based on [2]. They replace the centered and detrended residuals by the indicator function of the difference of these residuals and their mean [10] and the ranks of the residuals [11], respectively. These robustifications show better performance than [2] when having to deal with impulsive noise, but have not been designed to yield good results for outliers in non-stationary processes. In this work, we analyze the robustness of the SST and propose different approaches of robustification.

Our contributions are: a proof of the increase of the Type I error rate for the SST in case of additive outliers; four new robust SST-based stationarity tests and a validation of these presented methods via simulated data and a real-world data application.

The paper is organized as follows. Section 2 briefly reviews the SST and illustrates the increase of the Type I error rate for additive outliers. Section 3 presents the proposed robust stationarity tests. Section 4, compares the proposed methods to existing approaches via extensive simulations and a real-world data example. Section 5 concludes the paper.

2. ROBUSTNESS ANALYSIS OF THE SPHERICITY STATISTIC TEST (SST)

2.1. The Sphericity Statistic Test (SST)

The sphericity statistic test (SST) is a frequency domain stationarity test that is defined as [3]

$$S_M(\omega_i) = \frac{A_M(C_{XX}(\omega_i))}{G_M(C_{XX}(\omega_i))} \underset{H_1}{\overset{H_0}{\lessgtr}} \gamma_i, \qquad (1)$$

where H_0 is the hypothesis that the random process $X(t)$, $t \in \mathbb{Z}$ is stationary and H_1 represents non-stationarity. Here,

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

$$A_M(C_{XX}(\omega_i)) = \frac{1}{M}\sum_{k=1}^{M} C_{XX}^{(k)}(\omega_i) \tag{2}$$

and

$$G_M(C_{XX}(\omega_i)) = \left(\prod_{k=1}^{M} C_{XX}^{(k)}(\omega_i)\right)^{\frac{1}{M}} \tag{3}$$

are the arithmetic and geometric means; M is the number of segments and $C_{XX}^{(k)}(\omega_i)$ is the spectrum of the k-th segment of $X(t)$ evaluated at frequency ω_i. γ_i is the critical value of the test for frequency ω_i. The thresholds for the test are based on a 5 % level of significance that is estimated empirically [3]. To test Eq. (1) for all frequencies, a suitable projection of the statistics onto one dimension is the mean over all frequencies [3]. Estimates of the spectrum are obtained by kernel smoothing of the periodogram using a rectangular window.

2.2. Additive Outlier Model

Various kinds of contamination in real-world applications can be modeled as additive outliers [6, 12], i.e.,

$$Y(t) = X(t) + \xi^p(t)\varepsilon(t), \tag{4}$$

with $X(t)$ being the outlier free time series and $\varepsilon(t)$ the contamination process. $\xi^p(t) = 1$ with probability p and $\xi^p(t) = 0$ with probability $(1-p)$. We generate isolated and patchy outliers by letting the event $\xi^p(t) = 1$ occur either isolated or in patches. In case of patchy outliers, the patch length l is modeled by an exponential distribution with probability distribution $f(l;\beta) = \frac{1}{\beta}e^{-l/\beta}$. For the robustness evaluation of stationarity tests, $\varepsilon(t)$ must be non-stationary when $X(t)$ is stationary and vice-versa.

2.3. Increase of the Type 1 Error Rate for the SST for Additive Outliers

In the sequel, we show the non-robustness of the SST against even a single additive outlier under the following assumptions: (i) $X(t)$ and $\varepsilon(t)$ are statistically independent; (ii) $X(t)$ is a stationary process and (iii) $M > 1$. Let $\xi^p(t_0) = 1$, where t_0 is a sample in segment s and $\xi^p(t \neq t_0) = 0$ in Eq. (4) model the occurrence of a single outlier. Then, the following holds for the expectations of the spectra of $X^{(s)}(t), Y^{(s)}(t)$ and $\varepsilon^{(s)}(t)$:

$$E[C_{YY}^{(s)}(\omega_i)] = E[C_{XX}^{(s)}(\omega_i)] + E[C_{\varepsilon\varepsilon}^{(s)}(\omega_i)] \geq E[C_{XX}^{(s)}(\omega_i)] \tag{5}$$

The expectations of the arithmetic and geometric means are thus:

$$\begin{aligned} E[A_M(C_{YY}(\omega_i))] &= E\left[\frac{1}{M}\left(C_{YY}^{(s)}(\omega_i) + \sum_{k=1,k\neq s}^{M} C_{XX}^{(k)}(\omega_i)\right)\right] \\ E[G_M(C_{YY}(\omega_i))] &= E\left[\left(C_{YY}^{(s)}(\omega_i)\cdot \prod_{k=1,k\neq s}^{M} C_{XX}^{(k)}(\omega_i)\right)^{\frac{1}{M}}\right] \end{aligned} \tag{6}$$

Re-writing Eq. (6) in terms of Eq. (5) under the assumptions (i) and (ii) yields

$$\begin{aligned} E[A_M(C_{YY}(\omega_i))] &= E\left[\frac{1}{M}\sum_{k=1}^{M} C_{XX}(\omega_i)\right] + E\left[\frac{C_{\varepsilon\varepsilon}^{(s)}(\omega_i)}{M}\right] \\ E[G_M(C_{YY}(\omega_i))] &= E\left[\left(C_{XX}(\omega_i)^M + C_{\varepsilon\varepsilon}^{(s)}(\omega_i)C_{XX}(\omega_i)^{M-1}\right)^{\frac{1}{M}}\right] \end{aligned} \tag{7}$$

The Type 1 error rate is increased when

$$\frac{E[A_M(C_{YY}(\omega_i))]}{E[G_M(C_{YY}(\omega_i))]} > 1. \tag{8}$$

E.g., when $X(t)$ and $\varepsilon(t)$ are zero-mean white noise processes

$$\frac{E[A_M(C_{YY}(\omega_i))]}{E[G_M(C_{YY}(\omega_i))]} = \frac{\sigma_X^2 + \sigma_\varepsilon^2/M}{(\sigma_X^{2M} + \sigma_\varepsilon^2\sigma_X^{2(M-1)})^{\frac{1}{M}}}, \tag{9}$$

which for $\sigma_\varepsilon^2 = \sigma_X^2$, i.e., contamination of the same power as the uncontaminated process reduces to

$$\frac{E[A_M(C_{YY}(\omega_i))]}{E[G_M(C_{YY}(\omega_i))]} = \frac{1+\frac{1}{M}}{2^{\frac{1}{M}}} > 1.$$

Showing that Eq. (8) holds for $\sigma_\varepsilon^2 > \sigma_X^2$ can be derived from Eq. (9) analogously. More general statements can be derived from Eq. (7) and Eq. (8).

3. PROPOSED METHODS

3.1. Method 1: Filter-Cleaner (ACM-SST)

An intuitive way to deal with outliers in dependent data is to perform a data cleaning operation [6, 9]. Filter-cleaning algorithms [6, 9, 13, 14] "clean" $Y(t)$ from contamination by replacing corrupted samples with an estimate of $X(t)$ obtained from the prediction of a robust filter and leaving $Y(t)$ untouched, otherwise. The first proposed robustification of the SST is therefore given by combining a filter-cleaner with the classical SST. In this paper, we apply the approximate conditional mean (ACM), a state-space representation based filter-cleaner, which has been shown to be optimal in case of a Gaussian distributed state prediction density [13]. Obviously, the performance of Method 1 is directly dependent on the performance of the applied filter-cleaner.

3.2. Method 2: Robust Spectrum Estimation (ℓ_p-SST)

Eq. (1) dictates that the SST inherits the non-robustness of the spectrum estimates $\hat{C}_{XX}^{(k)}(\omega)$. A straight forward robustification is therefore to replace the non-robust spectrum estimates by their robust counterparts. In this paper, we adapt the ℓ_1-norm M-periodogram proposed in [15]. This estimate solves the ℓ_1-norm minimization of a (nonlinear) harmonic regression, which results in a spectrum estimate that is robust against additive outliers. For details, see [15]. The robustness of Method 2 is thus directly dependent on the chosen spectrum estimator.

3.3. Method 3: Robust Means (Mean-SST)

A different view on the SST is given by interpreting the geometric and arithmetic means as a source of non-robustness. We suggest to replace the $A_M(\omega_i)$ and $G_M(\omega_i)$ by their robust counterparts. In particular, $A_M(\omega_i)$ is replaced by an M-estimator of location [6] and a robustification of $G_M(\omega_i)$ is obtained via $e^{\hat{\mu}_{\text{rob},i}(\log(\hat{C}_{XX}(\omega_i)))}$. The robustified SST is therewith

$$\hat{S}_{M,rob}(\omega_i) = \frac{\hat{\mu}_{\text{rob},i}(\hat{C}_{XX}(\omega_i))}{e^{\hat{\mu}_{\text{rob},i}(\log(\hat{C}_{XX}(\omega_i)))}} \underset{H_1}{\overset{H_0}{\lessgtr}} \gamma_i, \qquad (10)$$

Method 4 requires that the spectrum estimate $\hat{C}_{XX}(\omega_i)$ is non-zero $\forall i$.

3.4. Method 4: Weighting of Segments (Seg-SST)

When dealing with patchy outliers, the degree of contamination of the segments in the observed time series can strongly vary [9, 16]. We assess the quality of the segments by the following measure

$$\Delta_\sigma^{(k)} = \hat{\sigma}_{\text{non-rob}}^{(k)} - \hat{\sigma}_{\text{rob}}^{(k)}, \quad k = 1, \dots, M \qquad (11)$$

where $\hat{\sigma}_{\text{non-rob}}^{(k)}$ is, e.g., the sample standard deviation of segment k and $\hat{\sigma}_{\text{rob}}^{(k)}$ is its robust counterpart, e.g., the normalized median absolute deviation [6]. With $\boldsymbol{\Delta}_\sigma = (\Delta_\sigma^{(1)}, \dots, \Delta_\sigma^{(M)})^T$, an M-type weighting function

$$W^{(k)}\left(\frac{\Delta_\sigma^{(k)} - \hat{\mu}_{\text{rob}}(\boldsymbol{\Delta}_\sigma)}{\hat{\sigma}_{\text{rob}}(\boldsymbol{\Delta}_\sigma)}\right)$$

that assigns weights to the segments depending on the deviation of the quality measure from the median absolute deviation is constructed. $W(\cdot)$ can be chosen, e.g., of the Huber-type with parameters as described in [6]. Method 4 requires the signal to have a rate of at least 50 % of clean or only negligibly contaminated segments.

3.5. Method 5: Combination of Method 2 and Method 4

As Methods 2 and 4 are independent of each other, it is possible to down-weight the influence of contaminated segments by Method 4 in addition to robust spectrum estimation. This is useful when outlier patches contaminate more than 50 % of a given segment, which yields non-robustness of this segments spectrum estimate, even when using a robust estimator.

4. RESULTS

4.1. Simulations

$X(t)$ was generated for five stationary ARMA-models from [3]: (i) iid Gaussian: $X(t) = Z(t)$; (ii) AR(1): $a_1 = 0.5$; $X(t) = 0.5X(t1) + Z(t)$; (iii) AR(1): $a_1 = -0.5$; (iv) MA(1): $b_1 = 1, X(t) = Z(t) + Z(t1)$; (v) AR(5): $X(t) = 0.5X(t1)0.6X(t2) + 0.3X(t3)0.4X(t4) + 0.2X(t5) + Z(t)$ and eight non-stationary processes from [3]: (i) ARIMA(0, 1, 0), $X(t) = X(t1) + Z(t)$; (ii) ARIMA(0, 2, 0); (iii) ARIMA(1, 1, 0) with $a_1 = 0.5$; (iv) ARIMA(0, 1, 1) with $b_1 = 1$; (v) ARIMA(1, 1, 1) with $a_1 = 0.5, b_1 = 1$; (vi) TVAR(1) with a_1 varying linearly between -0.5 and 0.5; (vii) GARCH: AR(1) model with $a_1 = 0.5$ where the variance of the innovations varies linearly from 0.5 to 2; (viii) GARCH: AR(1) model with $a_1 = 0.5$ where the variance of the innovations varies linearly from 0.1 to 1.

The contaminating signal as described in Eq. (4) was varied as follows: For stationary $X(t)$, $\varepsilon(t)$ is an ARIMA(0,1,0) and for non-stationary $X(t)$, $\varepsilon(t)$ is iid. In both cases, the distribution of $\varepsilon(t)$ is Gaussian, where $[\mu; \sigma]$ are varied from 0 to 10 with a step-size of 0.1. The outlier occurrence probability p was varied from 0 to 0.25 with a step-size of 0.05. Patchy outliers of different lengths were created by varying the patch length's scale parameter $\beta \in \{0.2; 2\}$. Isolated outliers were obtained by letting the event $\xi^p(t) = 1$ follow an iid Bernoulli distribution with outlier occurrence probability p.

To evaluate the performance, the receiver operating characteristic (ROC) is calculated, where the average is taken over 1000 MC realizations of all 13 processes described above. The results for the clean data case ($p = 0$) are displayed in Fig. 1. It can be seen that all methods provide a correct decision on the stationarity of $X(t)$, i.e., $P_{\text{CD}} \approx 1$ given a false alarm probability $P_{\text{FA}} > 0.05$, except for Method 3 (Mean-SST). The reason for the poor performance of this method is that the robust M-estimator of location falsely down-weights spectra in case of non-stationary $X(t)$.

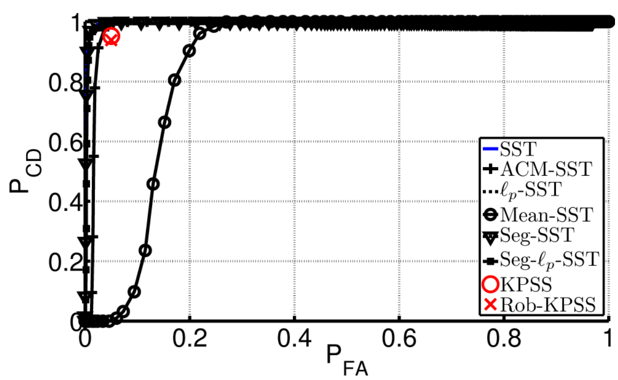

Fig. 1. Average ROC of the stationarity tests for the clean data case.

As shown in Fig. 2, the power of the non-robust tests, i.e.,the SST and KPSSS, degrades drastically in the presence of patchy outliers and even the rank based KPPS test [11], denoted as Rob-KPSS, does not provide robustness, with only a slightly better performance, compared to the KPSS. Fig. 3 displays the effect of isolated outliers on the stationarity tests. These are most demanding for all tests, as the outliers are distributed evenly over the entire signal length. This causes, e.g., Method 4 (Seg-SST), which is based on down-weighting contaminated data segments, to break down. Best performance across all scenarios is experienced by Method 5 (Seg-ℓ_p-SST), which combines robust spectrum estimation with segment quality information (see Eq. (11)).

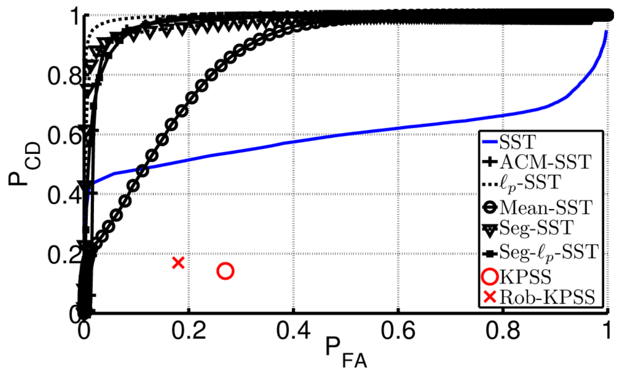

Fig. 2. Average ROC of the stationarity tests for patchy additive outliers.

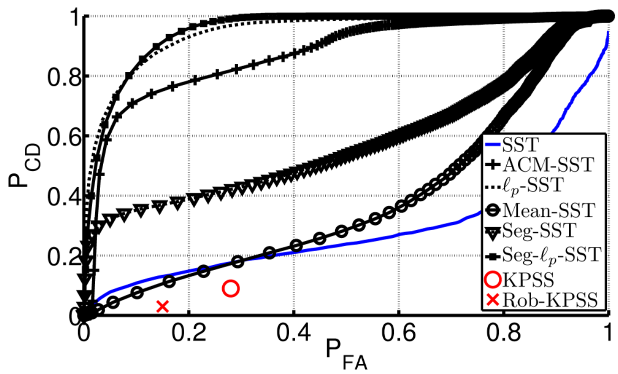

Fig. 3. Average ROC of the stationarity tests for isolated additive outliers.

Fig. 4 displays the probability of correct decision P_C w.r.t. the outlier occurrence probability p. In this setup, the ratio of the block length of samples containing iid outliers successively increased for $0 \leq p \leq 0.4$. The highest breakdown point (≈ 0.25) was obtained by ℓ_p-SST and Seg-ℓ_p-SST. This is explained by the fact that Eq. (1) contains a ratio of two spectra, each of which are estimated by a robust estimator with breakdown point 0.5 [15], thus limiting the achievable breakdown point to 0.5.

4.2. Real-Data Example: Window Regulator Motor

Window regulator motors are designed to produce as little noise as possible. Flaws in the mechanics of the motor can

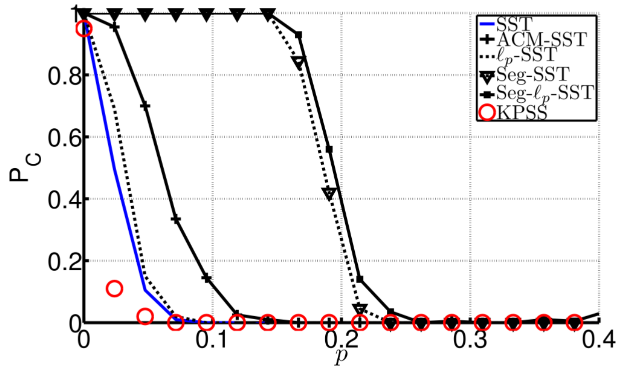

Fig. 4. Probability of correct decision P_C on the stationarity of $X(t)$ w.r.t. the outlier occurence probability p.

be contributed to a modulating character of dominant frequencies. Whereas the noise floor in the acoustic chambers is quite low, often measurements must be made in a noisier environment, e.g., in manufacturing halls. Figure 5 shows the correct rejection rate of the stationary null P_C, in dependence of the SNR for the above presented tests. The measurements were take in an anechoic chamber with a sampling frequency of 48 kHz and AWGN of different powers was added.

It can be seen that all robustifications of the SST outperform the classical SST in this example, except for Method 3, which again falsely interprets non-stationarity as outliers. Due to the constant noise level, Method 4 shows performance only similar to Rob-KPSS. Methods 1, 2 and 5 (ACM-SST, ℓ_p-SST and Seg-ℓ_p-SST) achieve $P_\mathrm{C} \approx 1$ for all SNR values.

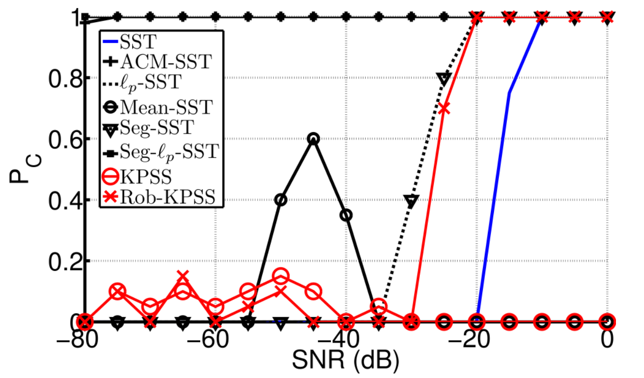

Fig. 5. P_C w.r.t. SNR for a malfunctioning window regulator motor, for which the dominant frequencies show a modulating character that results in a non-stationary signal.

5. CONCLUSIONS

The need for robustification of the SST was demonstrated and new approaches for robust stationarity testing were presented. Performance was evaluated and compared to existing stationarity tests for a large range of signals and contaminating processes. A real-data example has illustrated the applicability of the tests in industrial applications.

6. REFERENCES

[1] D. A. Dickey and W. A. Fuller. "Distribution of the Estimators for Autoregressive Time Series With a Unit Root." *J. Amer. Statist. Assoc.*, vol. 74, no. 366, pp. 427–431, June 1979.

[2] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin. "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root." *J. Econometrics*, vol. 54, no. 1-3, pp. 159–178, 1992.

[3] R. F. Brcich and D. R. Iskander. "Testing for Stationarity in the Frequency Domain Using a Sphericity Statistic." In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2006)*, pp. III-464–III-467.

[4] T. K. Blankenship, D. M. Kritzman, and T. S. Rappaport. "Measurements and Simulation of Radio Frequency Impulsive Noise in Measurements and Hospitals." In *Proc. IEEE 47th Vehicular Technology Conf.*, vol. 3, pp. 1942–1946, 1997.

[5] D. Middleton "Non-Gaussian Noise Models in Signal Processing for Telecommunications: New Methods and Results for Class A and Class B Noise Models" *IEEE Trans. Inform. Theory*, vol. 45, no. 4, pp. 1129–1149, 1999.

[6] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma. "Robust Estimation in Signal Processing: a Tutorial-Style Treatment of Fundamental Concepts." *IEEE Signal Proc. Mag.*, vol. 4, no. 29, pp. 61–80, 2012.

[7] M. Muma and A. M. Zoubir. "Robust Model Selection for Corneal Height Data Based on τ-Estimation." In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2011)*, pp. 4096 - 4099.

[8] M. Muma, Y. Cheng, F. Roemer, M. Haardt, and A. M. Zoubir. "Robust Source Number Enumeration for R-dimensional Arrays in Case of Brief Sensor Failures." In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2012)*, pp. 3709–3712.

[9] F. Strasser, M. Muma, and A. M. Zoubir. "Motion Artifact Removal in ECG Signals Using Multi-Resolution Thresholding." In *Proc. European Signal Processing Conference (EUSIPCO 2012)*, pp. 899–903.

[10] R. M. Yang, C. Amsler, and P. Schmidt. "A Robust Version of the KPSS Test Based on Indicators." *Journal of Econometrics*, vol. 137, no. 2, pp. 311–333, April 2007.

[11] M. Pelagatti and P. Sen. "A Robust Version of the KPSS Test Based on Ranks." in *Working Papers, Universita Degli Studi di Milano-Bicocca, Dipartimento di Statistica*, no. 20090701, July 2009.

[12] R. A. Maronna, R. D. Martin, and V. J. Yohai. *Robust Statistics: Theory and Methods*. Wiley Series in Probability and Statistics, 2006.

[13] R. D. Martin, and D. J. Thomson. "Robust-Resistant Spectrum Estimation." *Proc. IEEE*, vol. 70, no. 9, pp. 1097–1115, 1982.

[14] L. G. Tatum. "High Breakdown Methods of Time Series Analysis." *J. Roy. Statist. Soc. Ser.*, B 55, No 4, pp 881–896, 1993.

[15] V. Katkovnik. "Robust M-periodogram." *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 3104–3109, 1998.

[16] B. Han, M. Muma, and A. M. Zoubir. "An Online Approach for ICP Forecasting Based on Signal Decomposition and Robust Statistics." In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2013)*, pp. 6239–6243.