Welch Method Revisited: Nonparametric Power Spectrum Estimation Via Circular Overlap

Kurt Barbé, Rik Pintelon, and Johan Schoukens

Abstract—**The objective of this paper is twofold. The first part provides further insight in the statistical properties of the Welch power spectrum estimator. A major drawback of the Welch method reported in the literature is that the variance is not a monotonic decreasing function of the fraction of overlap. Selecting the optimal fraction of overlap, which minimizes the variance, is in general difficult since it depends on the window used. We show that the explanation for the nonmonotonic behavior of the variance, as reported in the literature, does not hold.**

In the second part, this extra insight allows one to eliminate the nonmonotonic behavior of the variance for the Welch power spectrum estimator (PSE) by introducing a small modification to the Welch method. The main contributions of this paper are providing extra insight in the statistical properties of the Welch PSE; modifying the Welch PSE to circular overlap—the variance is a monotonically decreasing function of the fraction of overlap, making the method more user friendly; and an extra reduction of variance with respect to the Welch PSE without introducing systematic errors—this reduction in variance is significant for a small number of data records only.

Index Terms—**Nonparametric, overlap, power spectrum, system identification, variance reduction, Welch's method, windowing, WOSA method.**

I. INTRODUCTION

AN important application area of digital signal processing (DSP) is the power spectral estimation of periodic and random signals. Speech recognition problems use spectrum analysis as a preliminary measurement to perform speech bandwidth reduction and further acoustic processing. Sonar systems use sophisticated spectrum analysis to locate submarines and surface vessels. Spectral measurements in radars are used to obtain target location and velocity information [1]. The vast variety of measurements that spectrum analysis encompasses is perhaps limitless, and thus, the subject received a lot of attention over the last five decades [2].

Another important area of application is system identification. Frequency-domain system identification offers a

tool to identify the transfer function $G(j\omega_k)$ of a linear dynamic time-invariant system measured at angular frequencies $\omega_k = (2\pi k)/(L), k = 0, \ldots, L-1$, where $j = \sqrt{-1}$. The user wants to identify the transfer function on a frequency band of interest such that a high-frequency resolution is desired. A high-frequency resolution implies long and costly experiments. When the noise characteristics are unknown, these are to be estimated from the input/output data. One can use nonparametric noise models to obtain this information, [3], which is closely related to nonparametric power spectrum estimation (PSE).

There are mainly two classes of power spectrum estimators: the parametric and the nonparametric estimators. The class of parametric PSE [4], [5] tries to fit a parametric model (AR, ARMA, MA, etc.) to the signal by minimizing a given cost function; for example: Burg's entropy method [6] and the Yule–Walker method [7]. In contrast to parametric methods, nonparametric methods do not make any assumptions on the data-generating process or model (e.g., autoregressive model, presence of a periodic component). There are five common nonparametric PSE available in the literature: the periodogram [8], the modified periodogram [2], Bartlett's method [9], Blackman–Tukey [7], and Welch's method [10]. However, all these nonparametric PSEs are modifications of the classical periodogram method introduced by Schuster [8].

The periodogram is defined as [2]

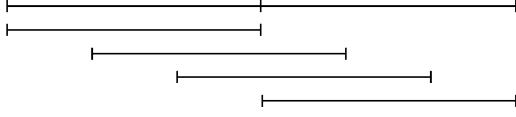$$\hat{S}_{xx}(k) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-\frac{2\pi jkn}{N}} \right|^2.$$

It is well known that the periodogram is asymptotically unbiased but inconsistent because the variance does not tend to zero for large record lengths. One can show [2], [11] that the variance on the periodogram $\hat{S}_{xx}(k)$ of an ergodic weakly stationary signal [12], $x(n)$ for $n = 0 : N - 1$ is asymptotically proportional to $S_{xx}^2(k)$, the square of the true power at frequency bin k. The periodogram uses a rectangular time-window, a weighting function to restrict the infinite time signal to a finite time horizon, [13]. The modified periodogram uses a nonrectangular time window [14].

A way to enforce a decrease of the variance is averaging. Bartlett's method [9] divides the signal of length N into K segments of length $L = N/K$ each. The periodogram method is then applied to each of the K segments. The average of the resulting estimated power spectra is taken as the estimated power spectrum. One can show that the variance is reduced by a factor K, but the spectral resolution is also decreased by a factor K, [2]. The Welch method [10] eliminates the tradeoff between spectral resolution and variance in the Bartlett method by allowing the segments to overlap; see Fig. 1. Furthermore,

Fig. 1.   Regular overlap for $K = 2$ and fraction of overlap $r = 2/3$.

Fig. 2.   Circular overlap for $K = 2$ and fraction of overlap $r = 2/3$.

the time window can also vary. Essentially, the modified periodogram method is applied to each of the overlapping segments and averaged out.

Like in system identification, if a high-frequency resolution is desired, one can only split the record in a small number K of segments of length L. In system identification, the number of segments K is typically 2,3,...,6; see [15] and [16]. Unfortunately, a small number of segments K implies a higher variance of the estimated power spectrum. Therefore, it is worth looking into methods using overlapping subrecords for reducing the variance as the main processing algorithm remains unchanged.

For more than three decades, many successful applications of the Welch PSE have been reported throughout the literature. However, there are still some open questions regarding the underlying theory of the Welch method. Indeed, only a handful of papers extend beyond the analysis of Welch [10]. In [10], the mean of the Welch PSE and the variance for a fraction of overlap $r = 0.5$ and $r = 0$ were studied. In [13], the variance of the Welch PSE was discussed for a fraction of overlap $r = 0.75$. In [17], the probability density function of the Welch PSE was investigated. A couple of papers [18], [19] reported a drawback connected to the Welch PSE: the variance is a nonmonotonically decreasing function of the fraction of overlap. The variance starts to grow when the fraction of overlap becomes too large. This makes the method difficult to apply, as it is not clear which fraction of overlap to use since the optimum also depends on the type of the window used.

It is explained intuitively [18], [19] that this is a result of the competing effects of more segments (lowering the variance) and more overlap among segments (increasing the variance). In this paper, we show that this intuitive explanation does not hold. In particular, we illustrate that by using circular overlap (Fig. 2), the variance of the PSE is monotonically decreasing as the fraction of overlap increases. We explain that Welch's method suffers from the nonmonotonicity due to the unequal weighting of the measured samples, increasing the uncertainty of the PSE considerably.

Furthermore, we show that the variance of the PSE can be reduced up to a factor $K/(K-r)$, with r the fraction of overlap and K the number of independent segments, with respect to the variance of the Welch PSE.

[1]isiknowledge.com.
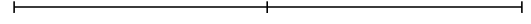
Fig. 3. $K = 2$ independent segments of the data record.

However, it is clear by using circular overlap that the concatenation of the end of the record with the beginning of the record creates a discontinuity in the signal. We need to examine the influence of this discontinuity on the statistical properties of the PSE with circular overlap. It is shown that no extra systematic errors are introduced by the discontinuity, implying that the PSE with circular overlap is asymptotically unbiased. However, the convergence rate is slightly slower. Furthermore, we show that the bias is of the same order of magnitude as the bias of Welch's method.

The user can gain these properties for free, as the algorithm for computing the PSE via circular overlap is essentially the same as computing the Welch PSE, taking into account some minor modifications.

In summary, the main contributions of this paper are:
- providing further insight in the statistical properties of the Welch PSE;
- improving the Welch PSE, which implies a decrease in uncertainty up to a factor $K/(K-r)$;
- guaranteeing a monotonic decrease in variance of the PSE via circular overlap as a function of the fraction of overlap r.

II. ASSUMPTIONS AND NOTATIONS

We assume that the signal $x(n)$ with $n = 0, 1, \ldots, N - 1$ is a weakly stationary Gaussian process. The signal $x(n)$ is divided into K segments, as in Fig. 3, such that every segment has length $L = N/K$. Further, we extract different overlapping subrecords by following the scheme of Fig. 2. The ith overlapping subrecord of the signal $x(n)$ satisfies

$$x^{[i]}(n) = x([n + L(1-r)(i-1)]_N) \qquad (1)$$

with $n = 0, \ldots, L-1$, $0 \le r < 1$, and $[n + L(1-r)(i-1)]_N$ denotes $n + L(1-r)(i-1)$ modulo N imposing circular overlap. We need to impose that circular overlap implies an integer number of overlapping subrecords; further, the different subrecords need to overlap an integer number of time samples. One can show that these two properties are, respectively, satisfied if the following two conditions hold simultaneously:

$$\begin{cases} r = 1 - \frac{K}{m} & \text{for } m \in \{K, K+1, \ldots\} \\ r = 1 - \frac{v}{L} & \text{for } v \in \{1, 2, \ldots, L\} \end{cases} . \qquad (2)$$

The first expression in (2) implies that the number of subrecords extracted via circular overlap is an integer. The last expression implies that only an integer number of points are overlapped.

In the beginning of the section, we assumed the signal $x(n)$ to be weakly stationary. Next, we apply Wold's decomposition theorem [12], which states that a weakly stationary Gaussian signal can be interpreted as filtered white Gaussian noise, where the filter has a square summable impulse response, but not necessarily stable in the bounded-input bounded-output
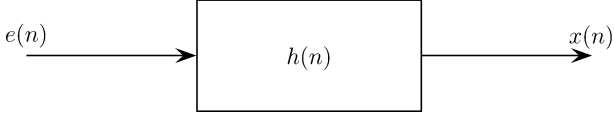
Fig. 4. The Wold decomposition of a weakly stationary signal.

sense, which requires an absolutely summable impulse response; see Fig. 4. In the sequel of this paper, we shall assume the following.

Assumption 1: The signal $x(n)$ with $n = 0, \ldots, N-1$ is a weakly stationary Gaussian process, such that $x(n) = h(n) * e(n)$, where $e(n)$ is zero-mean white Gaussian noise with unit variance and $h(n)$ denotes the impulse response of the corresponding filter. We assume that the sequence $h(n)$ satisfies $\sum_{n=0}^{\infty} n|h(n)| < \infty$.

Assumption 1 implies that the filter characteristic $H(\omega)$ is continuously differentiable. This condition can be relaxed to continuously filter characteristics, while the conclusions of this paper remain valid, except the convergence rate of the bias, which drops from $\mathcal{O}(1/L)$ to $\mathcal{O}(1/\sqrt{L})$; see [11] for more details.

The main objective is to estimate the power spectrum of the signal $x(n)$. The true power spectrum $S_{xx}(\omega)$ at frequency ω of a weakly stationary signal $x(n)$ is defined as [2]

$$
S_{xx}(\omega) = \sum_{n=-\infty}^{\infty} \mathbb{E}[x(n+m)x(m)]e^{-j\omega n}
$$
$$
= |H(\omega)|^2 \tag{3}
$$

where $\mathbb{E}[\cdot]$ denotes the expectation and the last equality is a direct consequence of Wold's decomposition [12].

Assumption 2: The time window $w(t)$ is assumed to be zero for $t \notin [0,1]$. The windowing function $w(t)$ is either a rectangular window $w(t) = 1, t \in [0,1]$ or nonnegative, unimodal, and symmetrical at $t = 1/2$, such that $w(1/2) = 1$.

Due to the fact that the time window is unimodal, Assumption 2 implies that the Riemann integral of the window function $w(t)$ exists [20]. Furthermore, it is important to note that Assumption 2 captures all windows analyzed in [13].

We define the windowed discrete Fourier transform (DFT) of the ith subrecord $x^{[i]}(n)$ at frequency bin k as

$$
X_w^{[i]}(k) = \left(\sum_{n=0}^{L-1} x^{[i]}(n) w\left(\frac{n}{L}\right) e^{-\frac{2\pi jkn}{L}} \right) e^{-\frac{2\pi jk}{L}\tau_i} \tag{4}
$$

with $\tau_i = (1-r)(i-1)L$. We corrected the phase, for the delay τ_i, to refer all subrecords to the same time origin.

III. DIRECT GENERALIZATION OF THE PERIODOGRAM TO CIRCULAR OVERLAP

In this section, we propose a direct generalization of the periodogram method and Welch method to circular overlapping subrecords and make a detailed analysis of both methods.

For every subrecord as in Fig. 2, the estimated power for subrecord i of the signal $x(n)$ is defined as

$$
\hat{S}_{xx}^{[i]}(k) = \left(\sum_{n=0}^{L-1} w\left(\frac{n}{L}\right)^2 \right)^{-1} \left| X_w^{[i]}(k) \right|^2. \tag{5}
$$

By averaging over the different overlapping subrecords in (5), an estimate of the power spectrum of the signal $x(n)$ is defined as

$$
\hat{S}_{xx}(k) = \frac{1-r}{K} \sum_{i=1}^{K/(1-r)} \hat{S}_{xx}^{[i]}(k). \tag{6}
$$

Before studying the bias and the variance of the estimator (6), we prove the following property.

Theorem 1: Under the conditions of Assumptions 1 and 2, the equation shown at the bottom of the page holds asymptotically $(L \to \infty)$. Further, $H(\omega_k)$ denotes the DFT of the filter $h(n)$, as explained in Section II and

$$
\Delta_1^{[m]}(k) \propto \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{n=0}^{m-1} w\left(\frac{n}{L}\right) e^{-j(\theta-\omega_k)n} \right|^2 \\
\times |H(\theta) - H(\omega_k)|^2 \, d\theta
$$

with $\omega_k = (2\pi k)/(L)$.

The proof can be found in Appendix A.

Remark 1:

- In Theorem 1, the first convergence rate holds for the noncircular segments and the second convergence rate holds for the circular segments.
- Note that since the function w is a windowing function, and if $m = L(1-r), Lr$, the function $|\sum_{n=0}^{m-1} w(n/L)e^{-j(\theta-\omega_k)n}| \to \delta(\theta - \omega_k)$ for $L \to \infty$. Furthermore, under Assumption 1, the filter characteristic $H(\omega)$ is continuously differentiable. Hence, one can show, by applying [21, Theorem 3.15], that $\mathcal{O}((\Delta^{[L(1-r)]}(k) + \Delta^{[Lr]}(k))) = \mathcal{O}(\log L)$.

Theorem 1 further implies the following.

Corollary 1: The first two moments of $\hat{S}_{xx}(k)$ equal

(i) $\mathbb{E}[\hat{S}_{xx}(k)]$
$= |H(\omega_k)|^2 \mathbb{E}[\hat{S}_{ee}(k)]$

$$
\mathbb{E}\left[\left| X_w^{[i]}(k) - H(\omega_k)E_w^{[i]}(k) \right|^4 \right] = \begin{cases} \mathcal{O}\left(\left(\Delta_1^{[L]}(k) \right)^2 \right), & \text{for } i \leq \frac{K-r}{1-r} + 1 \\ \mathcal{O}\left(\left(\Delta_1^{[L(1-r)]}(k) + \Delta_1^{[Lr]}(k) \right)^2 \right), & \text{elsewhere} \end{cases}
$$

$$+ \mathcal{O}\left(\frac{1}{L}\left(\Delta_2^{[L(1-r)]}(k) + \Delta_2^{[Lr]}(k)\right)\right)$$

$$\text{(ii)} \quad \mathbb{E}\left[\hat{S}_{xx}^2(k)\right]$$
$$= |H(\omega_k)|^4 \mathbb{E}\left[\hat{S}_{ee}^2(k)\right] + o(L^0)$$

with

$$\Delta_2^{[m]}(k) \propto \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{n=0}^{m-1} w\left(\frac{n}{L}\right)e^{-j(\theta-\omega_k)n}\right|^2$$
$$\times (|H(\theta)|^2 - |H(\omega_k)|^2)d\theta.$$

The proof can be found in Appendix B.

We can analyze the bias and the variance of the PSE (6). The expected value of the PSE (6) is given by

$$\mathbb{E}[\hat{S}_{xx}(k)]$$
$$= |H(\omega_k)|^2 \mathbb{E}[\hat{S}_{ee}(k)]$$
$$+ \mathcal{O}\left(\frac{1}{L}\left(\Delta_2^{[L(1-r)]}(k) + \Delta_2^{[Lr]}(k)\right)\right)$$
$$= |H(\omega_k)|^2 + \mathcal{O}\left(\frac{1}{L}\left(\Delta_2^{[L(1-r)]}(k) + \Delta_2^{[Lr]}(k)\right)\right)$$

where the last equality comes from the fact that, for every i

$$\mathbb{E}[\hat{S}_{ee}^{[i]}(k)] = \frac{1}{\sum_n w\left(\frac{n}{L}\right)^2}\sum_{n_1,n_2=0}^{L-1}\mathbb{E}\left[e^{[i]}(n_1)e^{[i]}(n_2)\right]$$
$$\times w\left(\frac{n_1}{L}\right)w\left(\frac{n_2}{L}\right)e^{-j\omega_k(n_2-n_1)}$$
$$= \frac{1}{\sum_n w\left(\frac{n}{L}\right)^2}\sum_{n_1,n_2=0}^{L-1}\delta(n_2-n_1)$$
$$\times w\left(\frac{n_1}{L}\right)w\left(\frac{n_2}{L}\right)e^{-j\omega_k(n_2-n_1)}$$
$$= 1.$$

This indicates that the bias is given by

$$\text{Bias}(\hat{S}_{xx}(k)) = \mathcal{O}\left(\frac{1}{L}\left(\Delta_2^{[L(1-r)]}(k) + \Delta_2^{[Lr]}(k)\right)\right) \quad (7)$$

and the variance is given by

$$\text{Var}(\hat{S}_{xx}(k)) = \mathbb{E}[\hat{S}_{xx}(k)^2] - \mathbb{E}[\hat{S}_{xx}(k)]^2$$
$$\approx |H(\omega_k)|^4 \text{Var}(\hat{S}_{ee}(k)) \quad (8)$$

where the last approximate equality holds asymptotically by application of Corollary 1. Since nonparametric PSEs are inconsistent, the variance does not vanish asymptotically but is dominated by (8).

A. Collating the MSE of the PSE (5) With the MSE of the Welch Method

To study the differences between the MSE of (6) and the Welch method, we discuss first the bias contribution $\Delta_2^{[Lr]}(k) + \Delta_2^{[L(1-r)]}(k)$ and secondly the variance contribution $|H(\omega_k)|^4 \text{Var}(\hat{S}_{ee}(k))$ separately.

1) The Bias Contribution: In the case of circular overlap, the bias contribution (7) is dominated by $\Delta_2^{[L(1-r)]}(k)$ and $\Delta_2^{[Lr]}(k)$, which is not the magnitude of the DFT of the whole windowing function, where for the Welch PSE the bias contribution is dominated by $\Delta_2^{[L]}(k)$. This observation indicates that an extra leakage contribution is introduced by the circular segments. Therefore, by breaking the window, the good properties of the windowing function are ruined in the circular segments.

2) The Variance Contribution: Next, we study the variance contribution in the MSE of (6). Expression (8) reveals that the variance can be approximated by

$$\text{Var}(\hat{S}_{xx}(k)) \approx |H(\omega_k)|^4 \text{Var}(\hat{S}_{ee}(k)).$$

By taking circular overlap (2) into account, it can easily be seen that

$$\text{Var}(\hat{S}_{xx}(k))$$
$$\approx |H(\omega_k)|^4 \text{Var}(\hat{S}_{ee}(k))$$
$$= |H(\omega_k)|^4 \frac{1-r}{K}\left\{1 + 2\sum_{i=1}^{\frac{1}{1-r}-1}\rho(i+1)\right\} \quad (9)$$

where $\rho(i) = \text{correlation}(\hat{S}_{ee}^{[1]}(k), \hat{S}_{ee}^{[i]}(k))$. Welch [10] showed that the variance of the Welch PSE equals

$$\text{Var}\left(\hat{S}_{xx}^{\text{Welch}}(k)\right)$$
$$\approx \frac{|H(\omega_k)|^4}{M}\left\{1 + 2\sum_{i=1}^{\frac{1}{1-r}-1}\frac{M-i}{M}\rho(i+1)\right\} \quad (10)$$

with $M = (K-1)/(1-r) + 1$ the number of overlapping subrecords for regular overlap. Note that the expression for the variance of the Welch PSE uses an unequal weighting for the correlation between two overlapping subrecords. Circular overlap weights the correlation between two overlapping subrecords equally. Although the correlation between two overlapping subrecords does not depend on the method of overlap (circular overlap or regular overlap), the nonmonotonicity of the variance reported in [18] and [19] is not observed for circular overlap.

Although we are not able to formally prove this claim, we can analytically show that the variance is a monotonically decreasing function of the fraction of overlap when a rectangular window is used. For a rectangular window and $k \neq 0, L/2$, the variance (9) equals

$$\text{Var}(\hat{S}_{xx}(k)) = |H(\omega_k)|^4 \frac{1}{3K}((1-r)^2 + 2). \quad (11)$$

The computation of (11) is found in Appendix C.

Hence, the variance of the PSE via circular overlap is a parabola with minimum for $r = 1$. Extensive simulations, without a formal proof, revealed that for the windows described by Assumption 2, the following conjectured claim holds if the window function $w(x)$ is m-times continuously differentiable:

$$\text{Var}(\hat{S}_{xx}(k)) \approx \alpha + \mathcal{O}((1-r)^{3+m}) \quad (12)$$

with α the lower bound for r tending to one.

Next, we compare the variance expression (9) for circular overlap with (10), the Welch method, as follows:

$$
\begin{aligned}
\operatorname{Var}&\left(\hat{S}_{xx}^{\text{Welch}}(k)\right) \\
&= \frac{|H(\omega_k)|^4}{M}\left\{1 + 2\sum_{i=1}^{\frac{1}{1-r}-1}\frac{M-i}{M}\rho(i+1)\right\} \\
&= \frac{K}{M(1-r)}\operatorname{Var}(\hat{S}_{xx}(k)) - 2\frac{|H(\omega_k)|^4}{M^2} \\
&\quad \times \sum_{i=1}^{\frac{1}{1-r}-1}i\rho(i+1).
\end{aligned}
$$

Since the correlation $\rho(i)$ is positive, the following inequality holds:

$$
\operatorname{Var}\left(\hat{S}_{xx}^{\text{Welch}}(k)\right) \le \frac{K}{K-r}\operatorname{Var}(\hat{S}_{xx}(k)). \tag{13}
$$

An important consequence of (13) is that the variance $\operatorname{Var}(\hat{S}_{xx}(k))$ can be reduced as low as $(K-r)/(K)\operatorname{Var}(\hat{S}_{xx}^{\text{Welch}}(k))$. Hence, circular overlap can reduce the variance of the Welch PSE up to factor $1-r/K$.

It is known that the variance of the Welch PSE is not monotonically decreasing as a function of overlap [18], [19]. A priori comparing (9) and (10) suggests that the reason the PSE via circular overlap outperforms the Welch PSE is due to the fact that more segments are averaged out. However, the fact that the variance of the Welch PSE is not monotonically decreasing implies that increasing the fraction of overlap will not provide a reduction in variance.

Next, we illustrate these results with a numerical example. In Section IV, we address the question of whether it is possible to eliminate this bias–variance tradeoff.

B. Numerical Example

In this section, we illustrate the method of circular overlap as described in the last section. In particular, we shall illustrate the variance reduction (13) and the MSE. In a second example, we illustrate that the variance of the PSE via circular overlap is monotonically decreasing where the Welch PSE suffers from nonmonotonicity as reported in [19] and [18].

1) Example: Keeping in mind the requirement of a maximal frequency resolution (e.g., system identification), we illustrate the method for $K = 2$ (two data records only). We used a type 1 digital Chebychev filter of order two, a stopband ripple of 20 dB, and a cutoff frequency at $0.15 \times f_s$. A zero-mean white Gaussian noise sequence with unit variance of $N = 2000$ points was generated. The sequence was partitioned in two records of $L = 1000$ points, and a Hanning window was applied. The light gray curve in the first plot shows one realization of the PSE via circular overlap, and the dark gray curve shows one realization of the PSE of Welch. The second plot shows the corresponding root mean squared error (RMSE) of the PSE via circular overlap and the PSE of Welch, respectively, both estimated on $10\,000$ Monte Carlo simulations. The third and fourth plots zoom in at the resonance. In Fig. 5, we see that we gain approximately 2
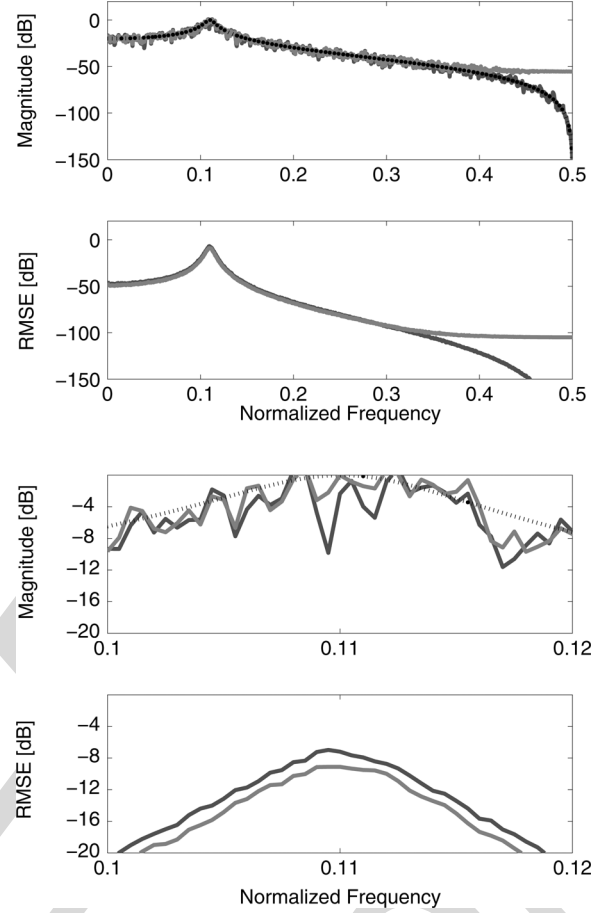

Fig. 5. PSE via circular overlap ($r = 0.8$) for a type–1 digital Chebychev filter of order two. The first plot shows the estimated power spectrum; the second plot shows the root mean squared error of the estimated PSE of the first plot. The third and fourth plots show a zoom-in at the resonance of the first and second plots, respectively. The dashed curve is the true power spectrum, the light gray curve is the PSE via circular overlap, and the dark gray curve is the Welch PSE.

dB in RMSE in the passband of the estimator, which is in agreement with a factor $K/(K-r)$ of (13). However, for the higher frequencies, it is clear that the PSE with circular overlap has difficulties following the slope towards infinity, where the Welch PSE does not suffer from this effect. Due to the fact that the variance of the PSE is proportional to the true power, the MSE for the higher frequencies is dominated by the bias contribution. We conclude that the variance of the PSE via circular overlap is reduced, but at the expense of an increase in bias. By introducing a small modification for the PSE via circular overlap, we show in Section IV that this extra bias can be suppressed.

2) Example: In this example, we use the same setup as in the previous example. We used a type–1 digital Chebychev filter of order two, a stopband ripple of 20 dB, and a cutoff frequency at $0.15 \times f_s$. A zero-mean white Gaussian noise sequence with unit variance of $N = 2400$ points was generated via Matlab. The sequence was partitioned in $K = 4$ records of $L = 600$ points, and a Hanning window was applied.

We compute the variances of the PSE via circular overlap and the PSE of Welch as a function of the fraction of overlap. In particular, we choose all fractions of overlap between zero and .9 such that, for the Welch PSE, the number of segments is
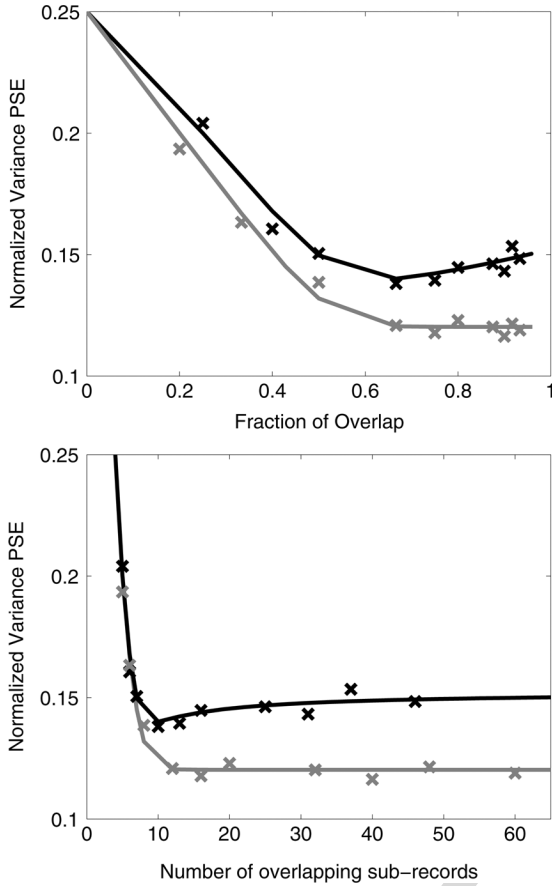
Fig. 6. The black curve indicates the theoretical normalized variance of the Welch PSE, and the gray curve indicates the theoretical variance of the PSE via circular overlap as a function of the fraction r in the top plot and as a function of the number of overlapping subrecords in the bottom plot. The "x"-markers are the simulated normalized variance at the resonance frequency on 6000 Monte Carlo simulations of the PSE of a type–1 digital Chebychev filter of order two.

an integer $((K - r)/(1 - r) \in \mathbb{N})$. For the PSE via circular overlap, we choose all the possible fractions satisfying (2). This is illustrated in the top plot of Fig. 6.

Secondly, the top plot of Fig. 6 was redrawn, such that the x-axis is a function of the number of overlapping segments. If the Welch PSE was computed for the fractions $r(1), r(2), \ldots, r(n)$, we plotted the variance of the Welch PSE as a function of $(K - r(1))/(1 - r(1)), \ldots, (K - r(n))/(1 - r(n))$ and similarly for the PSE via circular overlap. This is illustrated in the bottom plot of Fig. 6. In both figures, we normalized the computed variances by dividing these by the true power.

It is clear that in both plots of Fig. 6, the variance of the Welch PSE (indicated in black) is not a decreasing function of both the fraction of overlap and the number of overlapping subrecords. The variance of the Welch PSE has a minimum that depends on the window used. In general, the question arises of which fraction of overlap to use in order to minimize the variance. In the case of the PSE via circular overlap, the variance (indicated by the gray curve) does not suffer from this effect. The variance is a monotonically decreasing function of both the fraction of

overlap and the number of overlapping subrecords. It is important to stress that Fig. 6 illustrates the conjectured claim (12) as variance for the PSE with circular overlap decreases monotonically to a lower bound for r tending to one.

Note that the "x"-markers coincide with the theoretic curves within the accuracy of the Monte Carlo simulation ($0.97 < (\hat{\sigma}^2)/(\sigma^2) < 1.03$ with 95% confidence).

Furthermore, this example shows that the common explanation for the nondecreasing behavior of the variance of the Welch PSE—that a tradeoff needs to be made between increasing the fraction of overlap (more correlation between two overlapping subrecords) and more subrecords to average—does not hold. The correlation, in the first plot of Fig. 6, between two subrecords is the same for both the Welch PSE and the PSE via circular overlap. However, the variance of the PSE via circular overlap is monotonically decreasing.

In the next section, we discuss a way to obtain the reduction in variance without increasing the bias.

IV. IMPROVED POWER SPECTRUM ESTIMATION VIA CIRCULAR OVERLAP

In Section III, it was shown that the bias of the PSE via circular overlap is a function of $\Delta^{[L(1-r)]}(k)$ and $\Delta^{[Lr]}(k)$. We observed that the good properties of the windowing function were ruined by using circular overlap since, for instance, $\Delta^{[L(1-r)]}(k)$ is a function of $|\sum_{n=0}^{L(1-r)-1} w(n/L)e^{-j\omega n}|$, which is not the magnitude of the DFT of the whole windowing function. This led to significant bias for the frequency lines where the true power $S_{xx}(k) \approx 0$, as illustrated in example B.1. The following modification solves this problem without losing the variance reduction.

We introduce the following notation, which is needed in the modification. Let $w(n/L)$ be the windowing function used; then we define

$$w^{[v]}\left(\frac{n}{L}\right) = \begin{cases} w\left(\frac{n}{v}\right) & \text{for } 0 \le n < v \\ w\left(\frac{n-v}{L-v}\right) & \text{for } v \le n < L \end{cases}.$$

For the noncircular subrecords $i \le (K - 1)/(1 - r) + 1$, no modification of the periodogram is needed. For the circular subrecords $(K - 1)/(1 - r) + 1 < i \le (K/1 - r)$, we define

$$\tilde{S}_{xx}^{[i]}(k) = \left(\sum_{n=0}^{L-1} w^{[v_i]}\left(\frac{n}{L}\right)^2\right)^{-1} \left|X_{w^{[v_i]}}^{[i]}(k)\right|^2 \qquad (14)$$

where $X_{w^{[v_i]}}^{[i]}(k)$ is defined as in (4), where the window $w^{[v_i]}$ with $v_i = (i - (K - 1)/(1 - r) + 1)L(1 - r)$ is used. The only difference between (5) and (14) is the use of the window $w^{[v]}$ for the circular segments, as illustrated in Fig. 7. The improved PSE via circular overlap becomes

$$\tilde{S}_{xx}(k) = \frac{1 - r}{K} \left\{ \sum_{i=1}^{\frac{K-1}{1-r}+1} \hat{S}_{xx}^{[i]}(k) + \sum_{i=\frac{K-1}{1-r}+2}^{\frac{K}{1-r}} \tilde{S}_{xx}^{[i]}(k) \right\}. \qquad (15)$$
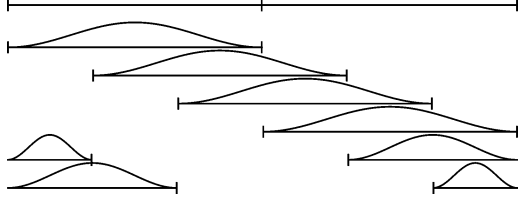
Fig. 7. Illustration of the applied windows for each subrecord for $K = 2$ and $r = 2/3$.

A. Properties of (15)

Following the scheme of Fig. 7, it is clear that the window applied to the circular segments will not lose its good properties. Applying Theorem 1 and repeating the same discussion as in Section III, the bias and the variance of (15) can be approximated by

$$\text{Bias}(\tilde{S}_{xx}(k)) = \mathcal{O}\left(\frac{1}{L}\left(\tilde{\Delta}_2^{[Lr]}(k) + \tilde{\Delta}_2^{[L(1-r)]}(k)\right)\right) \tag{16}$$

$$\text{Var}(\tilde{S}_{xx}(k)) = |H(\omega_k)|^4 \, \text{Var}(\tilde{S}_{ee}(k)) + o(L^0) \tag{17}$$

where

$$\tilde{\Delta}_2^{[m]}(k) \propto \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{n=0}^{m-1} w^{[m]}\left(\frac{n}{L}\right)e^{-j(\theta-\omega_k)n}\right|^2 \\ \times(|H(\theta)|^2 - |H(\omega_k)|^2)d\theta.$$

Fortunately, due to the fact that, for instance, $\left|\sum_{n=0}^{L(1-r)-1} w^{[L(1-r)]}(n/L)e^{-j(\theta-\omega_k)n}\right|$ is the magnitude of the DFT of the whole window function, the good properties of the windowing function remain valid.

For the variance analysis of $\text{Var}(\tilde{S}_{ee}(k))$ in (17), it was not possible to find an analytical expression. However, extensive simulations showed that the following inequality holds asymptotically:

$$\text{Var}\left(\hat{S}_{xx}^{\text{Welch}}(k)\right) \le \frac{K}{K-r}\,\text{Var}(\tilde{S}_{xx}(k)). \tag{18}$$

This is what we expect due to the fact that we actually only changed the resolution of the window in the scheme in Fig. 7 but not the type of window.

In conclusion, the improved estimator $\tilde{S}_{xx}(k)$ (15) has approximately the same variance as the estimator $\hat{S}_{xx}(k)$ (6). However, the bias of the estimator $\tilde{S}_{xx}(k)$ is significantly reduced with respect to the bias of $\hat{S}_{xx}(k)$. This is mainly due to fact that the good properties of windowing for leakage reduction are ruined in $\hat{S}_{xx}(k)$, whereas this is not the case for the estimator $\tilde{S}_{xx}(k)$.

B. Numerical Example

In this section, we illustrate the modified PSE via circular overlap by repeating example III.B. In particular, we illustrate that the bias due to the modification is of the same magnitude

as the bias of the Welch PSE and that the variance reduction (18) is approximately achieved. In the second example, we show that the variance of the modified PSE via circular overlap remains a monotonically decreasing function of both the fraction of overlap and the number of overlapping subrecords.

1) Example: In the first example, we show that the modification works where a Hanning window was applied. In particular, we show on the same simulation example as in Section III-B that the approximately the same RMSE reduction is achieved and this without paying the price in bias. The light gray curve in the first plot shows one realization of the PSE via circular overlap, and the dark gray curve shows one realization of the PSE of Welch. The second plot shows the corresponding RMSE of the PSE via circular overlap and the PSE of Welch, respectively, both estimated on 10 000 Monte Carlo simulations. The third and fourth plots zoom in on the resonance frequency. In Fig. 5, we see that we gain approximately 2 dB in RMSE in the passband of the estimator, which is in agreement with the factor $K/(K-r)$ of (18).

Fortunately, we observe that the bias of the Modified PSE via circular overlap is of the same magnitude as the bias of the Welch PSE, as indicated by the previous discussion. In the next example, we shall look into the approximating expression for the variance (18).

2) Example: In this example, we use the same setup as in the example as in Section III-B2. A zero-mean white Gaussian noise sequence with unit variance of $N = 2400$ points was generated via Matlab. The sequence was partitioned in four records of $L = 600$ points, and a Hanning window was applied.

In this example, we want to illustrate that the approximate expression for the variance (18) holds. We compute the variances of the modified PSE via circular overlap and the PSE of Welch as a function of the fraction of overlap. In particular, we choose all fractions of overlap between zero and .9 such that, for the Welch PSE, the number of segments is an integer $((K-r)/(1-r) \in \mathbb{N})$. For the modified PSE via circular overlap, we choose all the possible fractions satisfying (2). This is illustrated in the top plot of Fig. 9.

Secondly, the top plot of Fig. 9 was redrawn, such that the x-axis is a function of the number of overlapping segments. If the Welch PSE was computed for the fractions $r(1), r(2), \ldots, r(n)$, we plotted the variance of the Welch PSE as a function of $(K-r(1))/(1-r(1)), \ldots, (K-r(n))/(1-r(n))$ and similarly for the modified PSE via circular overlap. This is illustrated in the bottom plot of Fig. 9. In both figures, we normalized the computed variances by dividing these by the true power.

We see from both plots in Fig. 9 that the simulated variance computed from 1000 Monte Carlo simulations (indicated with the cross-markers) follows the theoretical curves (indicated by the solid lines) within the uncertainty of the simulation ($0.97 < (\hat{\sigma}^2)/(\sigma^2) < 1.03$ with 95% confidence). This illustrates the claim that the variance of the modified PSE via circular overlap follows approximately (18). Therefore, we conclude that we approximately can achieve a reduction in variance up to a factor $K/(K-r)$, and this without increasing the bias.
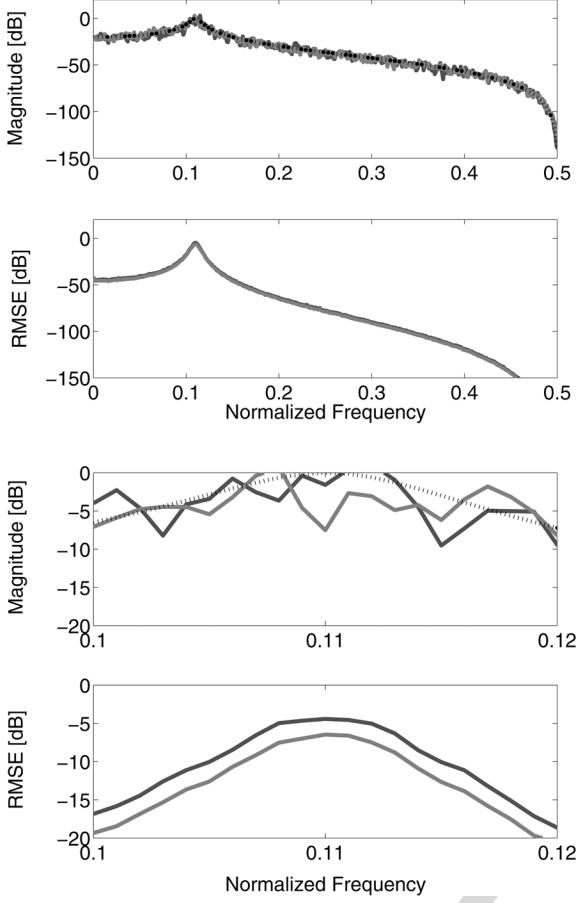
Fig. 8. PSE via circular overlap ($r = 0.8$) for a type–1 digital Chebychev filter of order two. The first plot shows the estimated power spectrum; the second plot shows the root mean squared error of the estimated PSE of the first plot. The third and fourth plots show a zoom-in at the resonance of the first and second plots, respectively. The dashed curve is the true power spectrum, the light gray curve is the PSE via circular overlap, and the dark gray curve is the Welch PSE.
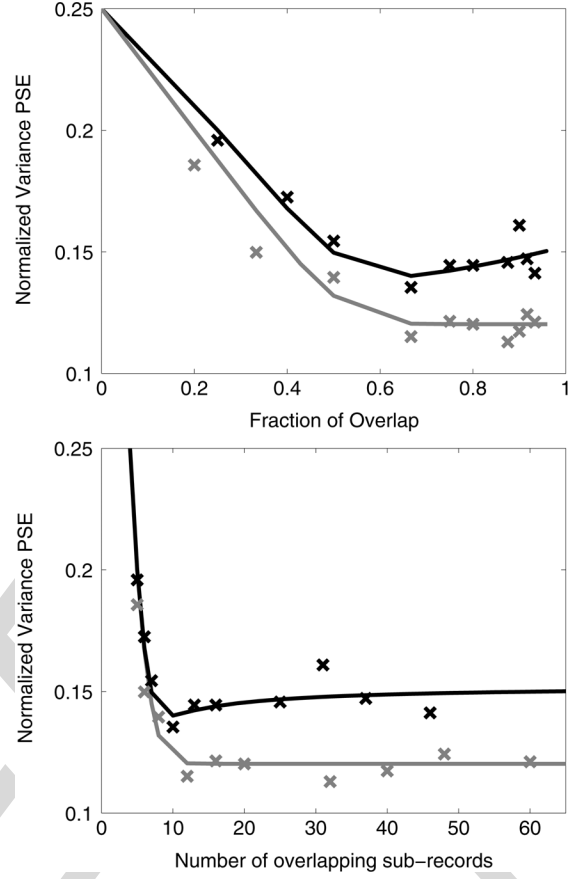

Fig. 9. The black curve indicates the theoretical normalized variance of the Welch PSE, and the gray curve indicates the theoretical variance of the PSE via circular overlap as a function of the fraction r in the top plot and as a function of the number of overlapping subrecords in the bottom plot. The cross-markers are the simulated normalized variance at the resonance frequency on 1000 Monte Carlo simulations of the PSE of a type 1 digital Chebychev filter of order 2.

V. Conclusion

The main goal of this paper was to investigate some aspects of the Welch PSE in more detail. In particular, we tried to find a theoretical explanation of the drawback (the nonmonotonicity of the variance of the Welch PSE) reported in the literature. Our analysis revealed the following.

- Circular overlap revealed that the nonmonotonicity of the variance of the Welch PSE is not due to the following tradeoff: increasing the fraction of overlap results in more subrecords to average but increases the correlation between two subrecords.
- The nonmonotonicity of the variance of the Welch PSE is due to the unequal weighting of the data, which implies that the covariance between two overlapping subrecords is weighted with a triangular window.
- The variance of the improved PSE via circular overlap (15) is monotonically decreasing without introducing extra systematic errors.
- The improved PSE via circular overlap is user-friendly since choosing an optimal percentage of overlap is no

longer an issue for the improved PSE via circular overlap. As a rule of thumb, we propose a fraction of overlap of $r = 0.8$.
- The PSE via circular overlap provides the user with a reduction of the variance with respect to the variance of the Welch PSE up to a factor $K/(K - r)$. However, this reduction in variance is significant for small values of K only.

The improved PSE via circular overlap is a small modification of the Welch PSE but has some better properties than the classical Welch PSE; however, theoretically, the bias vanishes slightly slower than the bias for the Welch PSE. The user can receive these extra properties for free as the main processing algorithm remains unchanged.

Appendix

Throughout the Appendix, we shall use the notation $w(n)$ instead of $w(n/L)$ for simplicity if there is no confusion possible.

A. Proof of Theorem 1

We shall only prove Theorem 1 for the subrecords $i = K/(1 - r)$ (see Fig. 10), but the proof for the other
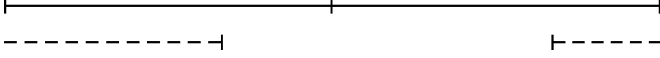
Fig. 10. The subrecord used in the proof is denoted by the dashed line.

subrecords is similar. Before proving the main result, we formulate two lemmas.

Lemma 1: Under Assumption 1, the following holds for $n_1 \in \{0, 1, \dots, Lr - 1\}$ and $n_2 \in \{(K-1)L + Lr, \dots, KL - 1\}$ and any arbitrary $\epsilon > 0$:

$$(i) \quad \mathbb{E}[x(n_1)x(n_2)] = \mathcal{O}\left(\frac{1}{L^{2+\epsilon}}\right)$$

$$(ii) \quad \mathbb{E}[e(n_1)x(n_2)] = \mathcal{O}\left(\frac{1}{L^{2+\epsilon}}\right).$$

Proof: The proof follows immediately from the fact that $\sum_{n=0}^{\infty} |n h(n)| < \infty$, implying that $h(n) = \mathcal{O}(n^{-2-\epsilon})$. We give the proof for claim i), but the proof for claim ii) is completely similar. We denote $n_1 \in \{0, 1, \dots, Lr - 1\}$ as a time sample in the beginning of the record and $n_2 \in \{(K-1)L + Lr, \dots, KL - 1\}$ at the end of the measured record. As indicated by Fig. 10, the points $x(n_1), x(n_2)$ are separated by at least L time samples. We compute

$$\begin{aligned}
&|\mathbb{E}[x(n_1)x(n_2)]| \\
&\leq \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} |h(m_1)h(m_2)\mathbb{E}[e(n_1 - m_1) \\
&\qquad \times e(n_2 - m_2)]| \\
&= \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} |h(m_1)h(m_2)| \\
&\qquad \times \delta(n_2 - m_2 - n_1 + m_1) \\
&= \sum_{m_1=0}^{\infty} |h(m_1)h(n_2 - n_1 + m_1)|.
\end{aligned}$$

Using the fact that $n_2 - n_1 + m_1 \geq L$ and $|h(n)| = \mathcal{O}(n^{-2-\epsilon})$ implies that

$$\begin{aligned}
|\mathbb{E}[x(n_1)x(n_2)]| &\leq \sum_{m_1=0}^{\infty} |h(m_1)h(n_2 - n_1 + m_1)| \\
&= \mathcal{O}(L^{-2-\epsilon})| \sum_{m_1=0}^{\infty} |h(m_1)|.
\end{aligned}$$

Since $|\sum_{m_1=0}^{\infty} |h(m_1)| < \infty$, the proof is complete. ∎

Lemma 2: Under Assumptions 1 and 2, the following holds asymptotically ($L \to \infty$):

$$\begin{aligned}
&\mathbb{E}\left[\left|X_w^{[i]}(\omega) - H(\omega)E_w^{[i]}(\omega)\right|^2\right] \\
&\qquad = \Delta^{[L(1-r)]}(\omega) + \Delta^{[Lr]}(\omega) + \mathcal{O}\left(\frac{1}{L^\epsilon}\right).
\end{aligned}$$

Proof: We compute

$$\begin{aligned}
&\mathbb{E}\left[\left|X_w^{[i]}(\omega) - H(\omega)E_w^{[i]}(\omega)\right|^2\right] \\
&= \sum_{n_1, n_2} \mathbb{E}\left[\left(x^{[i]}(n_1) - H(\omega)e^{[i]}(n_1)\right)\right.
\end{aligned}$$

$$\begin{aligned}
&\times \left(x^{[i]}(n_2) - \overline{H(\omega)}e^{[i]}(n_2)\right) \\
&\times w(n_1)w(n_2)e^{-j\omega(n_1 - n_2)}\bigg]. \tag{19}
\end{aligned}$$

Further, using (1), we compute the different terms in the expected value (19). Separately, for $i = K/(1-r)$

$$\begin{aligned}
(i) \quad & \sum_{n_1, n_2 = 0}^{L-1} \mathbb{E}\left[x^{[i]}(n_1)x^{[i]}(n_2)\right] w(n_1)w(n_2)e^{-j\omega(n_1-n_2)} \\
&= \sum_{n_1, n_2 = (K-1)L+Lr}^{KL-1} w(n_1 - (K-1)L - Lr) \\
&\quad \times w(n_2 - (K-1)L - Lr) \\
&\quad \times \mathbb{E}\left[x(n_1)x(n_2)\right]e^{-j\omega(n_1-n_2)} \\
&\quad + \sum_{n_1, n_2 = 0}^{Lr-1} w(n_1 + L(1-r))w(n_2 + L(1-r)) \\
&\quad \times \mathbb{E}[x(n_1)x(n_2)]e^{-j\omega(n_1-n_2)} \\
&\quad + 2\text{Re}\Bigg\{ \sum_{n_1=0}^{Lr-1} \sum_{n_2=(K-1)L+Lr}^{KL-1} w(n_1 + L(1-r)) \\
&\qquad\qquad \times w(n_2 - (K-1)L - Lr)\mathbb{E} \\
&\qquad\qquad \times [x(n_1)x(n_2)]e^{-j\omega(n_1-n_2)} \Bigg\}
\end{aligned}$$

where $\text{Re}(x)$ denotes the real part of x. Applying Lemma 1, the last term becomes $\mathcal{O}(1/L^\epsilon)$. Furthermore, if we denote $R_{xx}(n_1 - n_2) = \mathbb{E}[x(n_1)x(n_2)]$, then we find after renumbering the first sum with $\tilde{n}_l = n_l - (K-1)L - Lr$, where $l = 1, 2$, and the second sum with $\tilde{n}_l = n_l + L(1-r)$

$$\begin{aligned}
&\sum_{n_1, n_2 = 0}^{L-1} \mathbb{E}\left[x^{[i]}(n_1)x^{[i]}(n_2)\right] w(n_1)w(n_2)e^{-j\omega(n_1-n_2)} \\
&= \sum_{n_1, n_2 = 0}^{L(1-r)-1} R_{xx}(n_1 - n_2)w(n_1)w(n_2)e^{-j\omega(n_1-n_2)} \\
&\quad + \sum_{n_1, n_2 = L(1-r)}^{L-1} R_{xx}(n_1 - n_2)w(n_1)w(n_2)e^{-j\omega(n_1-n_2)} \\
&\quad + \mathcal{O}\left(\frac{1}{L^\epsilon}\right) \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx}(\theta) \left| \sum_{n=0}^{L(1-r)-1} w(n)e^{-j(\omega-\theta)n} \right|^2 d\theta \\
&\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx}(\theta) \left| \sum_{n=L(1-r)}^{L-1} w(n)e^{-j(\omega-\theta)n} \right|^2 d\theta \\
&\quad + \mathcal{O}\left(\frac{1}{L^\epsilon}\right)
\end{aligned}$$

where we used the Wiener–Khintchine theorem [2] to obtain the second equality.

In a completely similar way, we obtain for the other terms in the expected value (19)

$$\sum_{n_1,n_2=0}^{L-1} \mathbb{E}\left[x^{[i]}(n_1)e^{[i]}(n_2)\right] \overline{H(\omega)}w(n_1)w(n_2)e^{-j\omega(n_1-n_2)}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\overline{H(\omega)}H(\theta)\left|\sum_{n=0}^{L(1-r)-1}w(n)e^{-j(\omega-\theta)n}\right|^2 d\theta$$

$$+ \frac{1}{2\pi}\int_{-\pi}^{\pi}\overline{H(\omega)}H(\theta)\left|\sum_{n=L(1-r)}^{L-1}w(n)e^{-j(\omega-\theta)n}\right|^2 d\theta$$

$$+ \mathcal{O}\left(\frac{1}{L^\epsilon}\right).$$

The third term becomes

$$(iii)\quad \sum_{n_1,n_2=0}^{L-1}\mathbb{E}\left[e^{[i]}(n_1)x^{[i]}(n_2)\right]$$

$$\times H(\omega)w(n_1)w(n_2)e^{-j\omega(n_1-n_2)}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}H(\omega)\overline{H(\theta)}\left|\sum_{n=0}^{L(1-r)-1}w(n)e^{-j(\omega-\theta)n}\right|^2 d\theta$$

$$+ \frac{1}{2\pi}\int_{-\pi}^{\pi}H(\omega)\overline{H(\theta)}\left|\sum_{n=L(1-r)}^{L-1}w(n)e^{-j(\omega-\theta)n}\right|^2 d\theta$$

$$+ \mathcal{O}\left(\frac{1}{L^\epsilon}\right).$$

The fourth term becomes

$$(iv)\quad |H(\omega)|^2\sum_{n_1,n_2=0}^{L-1}\mathbb{E}\left[e^{[i]}(n_1)e^{[i]}(n_2)\right]w(n_1)$$

$$\times w(n_2)e^{-j\omega(n_1-n_2)}$$

$$= \frac{|H(\omega)|^2}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{n=0}^{L(1-r)-1}w(n)e^{-j(\omega-\theta)n}\right|^2 d\theta$$

$$+ \frac{|H(\omega)|^2}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{n=L(1-r)}^{L-1}w(n)e^{-j(\omega-\theta)n}\right|^2 d\theta$$

$$+ \mathcal{O}\left(\frac{1}{L^\epsilon}\right).$$

Combining results i)–iv) completes the proof. ∎

Next, we compute

$$\mathbb{E}\left[\left|X_w^{[i]}(k) - H(\omega_k)E_w^{[i]}(k)\right|^4\right]$$

$$= 2\left(\mathbb{E}\left[\left|X_w^{[i]}(k) - H(\omega_k)E_w^{[i]}(k)\right|^2\right]\right)^2$$

$$+ \left|\mathbb{E}\left[\left(X_w^{[i]}(k) - H(\omega_k)E_w^{[i]}(k)\right)^2\right]\right|^2 \quad (20)$$

$$= 2\left(\mathbb{E}\left[\left|X_w^{[i]}(k) - H(\omega_k)E_w^{[i]}(k)\right|^2\right]\right)^2 \quad (21)$$

$$= \mathcal{O}\left(\left(\Delta^{[L(1-r)]}(k) + \Delta^{[Lr]}(k)\right)^2\right) \quad (22)$$

where in (20) and (21), we used the fact that the Fourier coefficients $X_w^{[i]}(k)$, $E_w^{[i]}(k)$ are circular complex Gaussian distributed. See [22], which implies that the second term in (21) equals zero; and we used Lemma 2 for (22).

B. Proof of Corollary 1

We start with a small technical lemma, which indicates that the sum of the squares of the windowing function is proportional to the number of data points.

Lemma 3: Under Assumption 2, the following holds for large record lengths L and a nonzero constant $C > 0$,

$$\sum_{n=0}^{L-1}w\left(\frac{n}{L}\right)^2 > LC.$$

Proof: Under Assumption 2 the window function $w(t)$ is Riemann-integrable [20]. Due to the fact that $w(t)$ is uniformly bounded by the constant one, it is easy to see that the window function $w(t)$ is quadratic integrable. It follows that

$$\lim_{L\to\infty}\frac{1}{L}\sum w\left(\frac{n}{L}\right)^2 = \int_0^1 w(t)^2 dt.$$

Let $M = \int_0^1 w(t)^2 dt$ and choose $\delta > 0$ such that, for $L > L_0$, the following holds:

$$\frac{1}{L}\sum_{n=0}^{L-1}w\left(\frac{n}{L}\right)^2 \geq M - \delta$$

which completes the proof. ∎

Now, we are able to prove claim (i) of Corollary 1. For the first claim, it is sufficient to study $\mathbb{E}[\hat{S}_{xx}^{[i]}(k)]$. We compute

$$\left|\mathbb{E}[\hat{S}_{xx}^{[i]}(k)] - |H(\omega_k)|^2\,\mathbb{E}\left[\hat{S}_{ee}^{[i]}(k)\right]\right|$$

$$= \frac{1}{2\pi\left(\sum_n w(n)^2\right)}\left|\int_{-\pi}^{\pi}\left(S_{xx}(\theta) - |H(\omega_k)|^2\right)\right.$$

$$\times \left|\sum_{n=0}^{L(1-r)-1}w(n)e^{-j(\omega_k-\theta)n}\right|^2 d\theta$$

$$+ \int_{-\pi}^{\pi}\left(S_{xx}(\theta) - |H(\omega_k)|^2\right)$$

$$\times \left|\sum_{n=L(1-r)}^{L-1}w(n)e^{-j(\omega_k-\theta)n}\right|^2 d\theta$$

$$\left.+ \mathcal{O}\left(\frac{1}{L^\epsilon}\right)\right|$$

with $\omega_k = (2\pi k)/(L)$. This follows immediately from results (i) and (iv) in the proof of Lemma 2. Using the triangular inequality together with Lemma 3 results in

$$\left|\mathbb{E}\left[\hat{S}_{xx}^{[i]}(k)\right] - |H(\omega_k)|^2\,\mathbb{E}\left[\hat{S}_{ee}^{[i]}(k)\right]\right|$$

$$\leq \frac{1}{2\pi CL}\left(\int_{-\pi}^{\pi}\left|S_{xx}(\theta) - |H(\omega_k)|^2\right|\right.$$

$$
\times \left| \sum_{n=0}^{L(1-r)-1} w(n)e^{-j(\omega_k-\theta)n} \right|^2 d\theta
$$

$$
+ \int_{-\pi}^{\pi} \left| S_{xx}(\theta) - |H(\omega_k)|^2 \right|
$$

$$
\times \left| \sum_{n=L(1-r)}^{L-1} w(n)e^{-j(\omega_k-\theta)n} \right|^2 d\theta
$$

$$
+ \mathcal{O}\left(\frac{1}{L^\epsilon} \right) \bigg)
$$

which completes the proof of the first claim.

To show the second claim, it is sufficient to study $\mathbb{E}[\hat{S}_{xx}^{[i]}(k)\hat{S}_{xx}^{[j]}(k)]$. The proof of the second claim of Corollary 1 is based on a very important corollary of mean square convergence. The corollary is found in [23, p. 41]. However, we shall slightly generalize the result.

Lemma 4: Let $X_n, \tilde{X}_n, Y_n, \tilde{Y}_n$ four sequences of (complex) random variables, such that $X_n, \tilde{X}_n$ are equivalent in mean square ($\mathbb{E}[|X_n - \tilde{X}_n|^2] \to 0$ if $n \to \infty$) and $Y_n, \tilde{Y}_n$ are equivalent in mean square. Then the following holds: if the second-order moments of the random variables $X_n, \tilde{X}_n, Y_n, \tilde{Y}_n$ are uniformly bounded, then

$$
|\mathbb{E}[X_nY_n] - \mathbb{E}[\tilde{X}_n\tilde{Y}_n]| = \mathcal{O}(\mathbb{E}[|X_n - \tilde{X}_n|^2]^{1/2})
$$
$$
+ \mathcal{O}(\mathbb{E}[|Y_n - \tilde{Y}_n|^2]^{1/2}).
$$

Remark 2: In [23], the result is formulated for $\tilde{X}_n = \tilde{X}$ and $\tilde{Y}_n = \tilde{Y}$ for every n. However, this makes no difference for the proof.

Proof: It is easy to verify that $X_nY_n - \tilde{X}_n\tilde{Y}_n = (X_n - \tilde{X}_n)(Y_n - \tilde{Y}_n) - (\tilde{X}_n - X_n)\tilde{Y}_n - (\tilde{Y}_n - Y_n)\tilde{X}_n$. Using the triangular inequality in combination with the Cauchy–Schwartz (or Hölder's) inequality, we compute

$$
|\mathbb{E}[X_nY_n] - \mathbb{E}[\tilde{X}_n\tilde{Y}_n]|
$$
$$
= |\mathbb{E}[(X_n - \tilde{X}_n)(Y_n - \tilde{Y}_n)]
$$
$$
- \mathbb{E}[(\tilde{X}_n - X_n)\tilde{Y}_n]
$$
$$
- \mathbb{E}[(\tilde{Y}_n - Y_n)\tilde{X}_n]|
$$
$$
\leq \mathbb{E}[|X_n - \tilde{X}_n||Y_n - \tilde{Y}_n|] + \mathbb{E}[|\tilde{X}_n - X_n||\tilde{Y}_n|]
$$
$$
+ \mathbb{E}[|\tilde{Y}_n - Y_n||\tilde{X}_n|]
$$
$$
\leq \mathbb{E}[|X_n - \tilde{X}_n|^2]^{1/2}\mathbb{E}[|Y_n - \tilde{Y}_n|^2]^{1/2}
$$
$$
+ \mathbb{E}[|\tilde{X}_n - X_n|^2]^{1/2}
$$
$$
\times \mathbb{E}[|\tilde{Y}_n|^2]^{1/2} + \mathbb{E}[|\tilde{Y}_n - Y_n|^2]^{1/2}\mathbb{E}[|\tilde{X}_n|^2]^{1/2}.
$$

This completes the proof of this lemma since $\mathbb{E}[|\tilde{X}_n|^2], \mathbb{E}[|\tilde{Y}_n|^2]$ are assumed uniformly bounded. ∎

To prove claim (ii) of Corollary 1, we compute

$$
\mathbb{E}\left[\hat{S}_{xx}^{[i]}(k)\hat{S}_{xx}^{[j]}(k)\right]
$$
$$
= \frac{1}{(\sum_n w(n)^2)^2}\mathbb{E}\left[X_w^{[i]}(k)\bar{X}_w^{[i]}(k)X_w^{[j]}(k)\bar{X}_w^{[j]}(k)\right]
$$
$$
= \mathbb{E}\left[\hat{S}_{xx}^{[i]}(k)\right]\mathbb{E}\left[\hat{S}_{xx}^{[j]}(k)\right] + \left|\mathbb{E}\left[\frac{X_w^{[i]}(k)\bar{X}_w^{[j]}(k)}{\sum_n w(n)^2}\right]\right|^2 \tag{23}
$$

where the last equality follows from the following rule [22, p. 123]: for Z_1, Z_2 complex circular Gaussian random variables, the following rule applies:

$$
\mathbb{E}[Z_1\bar{Z}_1Z_2\bar{Z}_2] = \mathbb{E}[|Z_1|^2]\mathbb{E}[|Z_2|^2] + |\mathbb{E}[Z_1\bar{Z}_2]|^2.
$$

To complete the proof, we apply the previously shown claim (i) for the first term in (23) such that

$$
\mathbb{E}\left[\hat{S}_{xx}^{[i]}(k)\right]\mathbb{E}\left[\hat{S}_{xx}^{[j]}(k)\right]
$$
$$
\approx |H(\omega_k)|^4\mathbb{E}\left[\hat{S}_{ee}^{[i]}(k)\right]\mathbb{E}\left[\hat{S}_{ee}^{[j]}(k)\right].
$$

For the second term in (23), we apply Lemma 4 with $X_n = X_w^{[i]}(k), Y_n = \bar{X}_w^{[j]}(k)$ and $\tilde{X}_n = H(\omega_k)E_w^{[i]}(k), \tilde{Y}_n = \bar{H}(k)\bar{E}_w^{[j]}(k)$, and where $n = L$. This immediately proves the result since Theorem 1 together with Remark 1 implies that $(1)/(\sum_n w(n)^2)\mathbb{E}[|X_w^{[i]}(k) - H(\omega_k)E_w^{[i]}(k)|^2] = \mathcal{O}(1/L)$ for L large enough.

C. Proof of (11)

In this section, we shall prove that the variance expression (9) for a rectangular window equals

$$
\text{Var}(\hat{S}_{xx}(k)) = \frac{1}{3K}((1-r)^2 + 2).
$$

First, we study $\rho(i)$ in more detai,l where we used the following fact [3] for Gaussian processes $e(n)$: $\mathbb{E}[e(n_1)e(n_2)e(n_3)e(n_4)] = \mathbb{E}[e(n_1)e(n_2)]\mathbb{E}[e(n_3)e(n_4)] + \mathbb{E}[e(n_1)e(n_3)]\mathbb{E}[e(n_2)e(n_4)] + \mathbb{E}[e(n_1)e(n_4)]\mathbb{E}[e(n_2)e(n_3)]$

$$
\rho(i) = \mathbb{E}\left[\hat{S}_{ee}^{[1]}(k)\hat{S}_{ee}^{[i]}(k)\right] - 1
$$
$$
= \frac{1}{(\sum_n w(n)^2)^2}\sum_{n_1,\ldots,n_4=0}^{L-1}\mathbb{E}\left[e^{[1]}(n_1)e^{[1]}(n_2)e^{[i]}(n_3)\right.
$$
$$
\left.\times e^{[i]}(n_4)\right]w(n_1)w(n_2)w(n_3)
$$
$$
\times w(n_4)e^{-j\omega_k(n_2-n_1+n_3-n_4)} - 1
$$
$$
= \frac{1}{(\sum_n w(n)^2)^2}
$$
$$
\times \left\{ \left| \sum_{n_1,n_3=0}^{L-1} \mathbb{E}\left[e^{[1]}(n_1)e^{[i]}(n_3)\right]w(n_1)\right.\right.
$$
$$
\left.\times w(n_3)e^{j\omega_k(n_1+n_3)}\right|^2
$$
$$
+ \left| \sum_{n_1,n_4=0}^{L-1} \mathbb{E}\left[e^{[1]}(n_1)e^{[i]}(n_4)\right]\right.
$$
$$
\left.\left.\times w(n_1)w(n_4)e^{j\omega_k(n_1-n_4)}\right|^2 \right\}
$$
$$
= \frac{1}{(\sum_n w(n)^2)^2}
$$
$$
\times \left\{ \left| \sum_{n=(i-1)L(1-r)}^{L-1} w(n-(i-1)L(1-r))\right.\right.
$$

$$\times \left. w(n)e^{2j\omega_k n} \right|^2$$

$$+ \left| \sum_{n=(i-1)L(1-r)}^{L-1} w(n)w(n-(i-1)L(1-r)) \right|^2 \Bigg\}. \tag{24}$$

Now, we use the fact that the windowing function $w(x) = 1$ and $k \neq 0, L/2$ such that we obtain

$$\rho(i)$$

$$= \frac{1}{L^2} \left\{ \left| \sum_{n=(i-1)L(1-r)}^{L-1} e^{2j\omega_k n} \right|^2 \right.$$

$$\left. + \left(L(1-(i-1)(1-r))\right)^2 \right\}$$

$$= \frac{1}{L^2} \left| \frac{e^{2j\omega_k(i-1)L(1-r)} - e^{2j\omega_k L}}{1 - e^{2j\omega_k}} \right|^2$$

$$+ (1-(i-1)(1-r))^2$$

$$= \frac{1}{L^2} \left| \frac{e^{2j\omega_k(i-1)L(1-r)} - 1}{1 - e^{2j\omega_k}} \right|^2$$

$$+ (1-(i-1)(1-r))^2$$

$$= \frac{1}{L^2} \left| \frac{\sin(\omega_k(i-1)(1-r)L)}{\sin(\omega_k)} \right|^2$$

$$+ (1-(i-1)(1-r))^2$$

$$\approx (1-(i-1)(1-r))^2 \tag{25}$$

where the last equality (25) holds asymptotically $(L \rightarrow \infty)$ since $(1/L^2)|(\sin(\omega_k(i-1)(1-r)L))/(\sin(\omega_k))|^2 \rightarrow 0$ for $\omega_k \neq 0, \pi$, [11]. Hence, we can compute the variance explicitly

$$\mathrm{Var}(\hat{S}_{xx}(k))$$

$$= |H(\omega_k)|^4 \frac{1-r}{K} \left\{ 1 + 2 \sum_{i=1}^{\frac{1}{1-r}-1} \rho(i+1) \right\}$$

$$\approx |H(\omega_k)|^4 \frac{1-r}{K} \left\{ 1 + 2 \sum_{i=1}^{\frac{1}{1-r}-1} (1-i(1-r))^2 \right\} \tag{26}$$

where we used (25) to obtain (26). Next, we use following computation rules [24]:

$$\sum_{n=1}^{N-1} n = \frac{N(N-1)}{2}$$

$$\sum_{n=1}^{N-1} n^2 = \frac{1}{6}N(N-1)(2N-1).$$

Finally, we obtain

$$\frac{\mathrm{Var}(\hat{S}_{xx}(k))}{|H(\omega_k)|^4}$$

$$= \frac{1-r}{K} \left\{ 1 + 2 \sum_{i=1}^{\frac{1}{1-r}-1} (1-2(1-r)i + (1-r)^2 i^2) \right\}$$

$$= \frac{1-r}{K} \left\{ 1 + \frac{2}{1-r} - 2 - 2\frac{r}{1-r} + \frac{1}{3}\frac{r(1+r)}{1-r} \right\}$$

$$= \frac{1}{3K}(r^2 - 2r + 3)$$

$$= \frac{1}{3K}((1-r)^2 + 2)$$

which completes the proof.

REFERENCES

[1] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*. San Francisco, CA: Holden-Day, 1968.

[2] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 3rd ed. Upper Saddle River, N.J.: Prentice-Hall, 1996.

[3] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*. New York: IEEE, 2001.

[4] P. Broersen, *Automatic Autocorrelation and Spectral Analysis*. Berlin, Germany: Springer, 2006.

[5] G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day, 1970.

[6] J. P. Burg, "The relationship between maximum entropy spectra and maximum likelihood spectra," *Geophysics*, vol. 37, pp. 375–376, Apr. 1972.

[7] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.

[8] A. Schuster, "On the investigation of hidden periodicities," *Terrestr. Magn.*, vol. 3, pp. 13–41, 1898.

[9] M. S. Bartlett, "Smoothing periodograms from time series with continuous spectra," *Nature*, vol. 161, pp. 686–687, 1948.

[10] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. AE-15, pp. 70–73, Jun. 1967.

[11] D. Brillinger, *Time Series: Data Analysis and Theory*. London, U.K.: Holt, Rinehart and Winston, 1975.

[12] B. Porat, *Digital Processing of Random Signals: Theory and Methods*. Upper Saddle River, NJ: Prentice-Hall, 1994.

[13] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, 1978.

[14] D. J. Thompson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[15] J. Schoukens, R. Pintelon, G. Vandersteen, and P. Guillaume, "Frequency-domain system identification using nonparametric noise models estimated from a small number of data sets," *Automatica*, vol. 33, no. 6, pp. 1073–1086, 1997.

[16] K. Barbé, J. Schoukens, and R. Pintelon, "Frequency domain errors-in-variables estimation of linear dynamic systems using data from overlapping sub-records," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 8, pp. 1529–1536, 2008.

[17] P. Johnson and D. Long, "The probability density of spectral estimates based on modified periodogram averages," *IEEE Trans. Signal Process.*, pp. 1255–1261, 1999.

[18] T. P. Bronez, "On the performance advantage of multitaper spectral analysis," *IEEE Trans. Signal Process.*, vol. 40, no. 12, pp. 2941–2946, 1992.

[19] H. Jokinen, J. Ollila, and O. Aumala, "On windowing effects in estimating averaged periodograms of noisy signals," *Measurement*, vol. 28, no. 3, pp. 197–207, 2000.

[20] G. E. Shilov and B. L. Gurevich, *Integral, Measure and Derivative: A Unified Approach*. Englewood Cliffs: Prentice-Hall, 1966.

[21] A. Zygmund, *Trigonometric Series*. Cambridge, U.K.: Univ. Press, 1968–1979.

[22] B. Pincinbono, *Random Signals and Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[23] E. Lukacs, *Stochastic Convergence*. London, U.K.: Academic, 1975.

[24] R. Grimaldi, *Discrete and Combinatorial Mathematics*. Reading, MA: Addison-Wesley, 1999.

Kurt Barbé received the degree in mathematics (option statistics) from the Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 2005.

He is currently a Ph.D. Researcher with the ELEC Department, VUB. His main interests are in the field of system identification, time series analysis, and statistics.

Johan Schoukens received the engineer degree and the doctoral degree in applied sciences from Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 1980 and 1985, respectively.

He is presently a Professor at VUB. The prime factors of his research are in the field of system identification for linear and nonlinear systems.

Rik Pintelon received the engineer degree and the doctoral degree in applied science from Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 1982 and 1988, respectively.

He is presently a Professor in the ELEC Department, VUB. His main research interests are in the field of parameter estimation/system identification and signal processing.