Statistical Properties of a Modified Welch Method That Uses Sample Percentiles

Felix Schwock [1] Shima Abadi [1]

[1]Department of Electrical and Computer Engineering, University of Washington

Introduction

The **Welch's overlapped segment averaging (WOSA)** method (Welch, 1967) is a popular approach for estimating power spectral densities (PSDs) of stochastic signals. The estimation is performed by dividing the signal into K potentially overlapping segments of length N_s, computing the periodogram $\hat{P}_i$ for each segment, and taking the average over all K periodograms (see Figure 1)

- **Problem:** The WOSA method can suffer from strong outliers in the data caused by transients or other broad band interfering signals, which can limit the scope of the estimator (see Figure 2)
- **Solution:** Replace the arithmetic mean over multiple periodogram estimates in the WOSA estimator by a percentile estimation

In this work, we have derived simple expressions for bias, variance, and limiting distribution of the **Welch percentile (WP)** estimator for various percentiles and sample sizes.

Welch Percentile (WP) Estimator

Key idea: Replace averaging over periodograms in WOSA estimate by a frequency-wise quantile estimation.

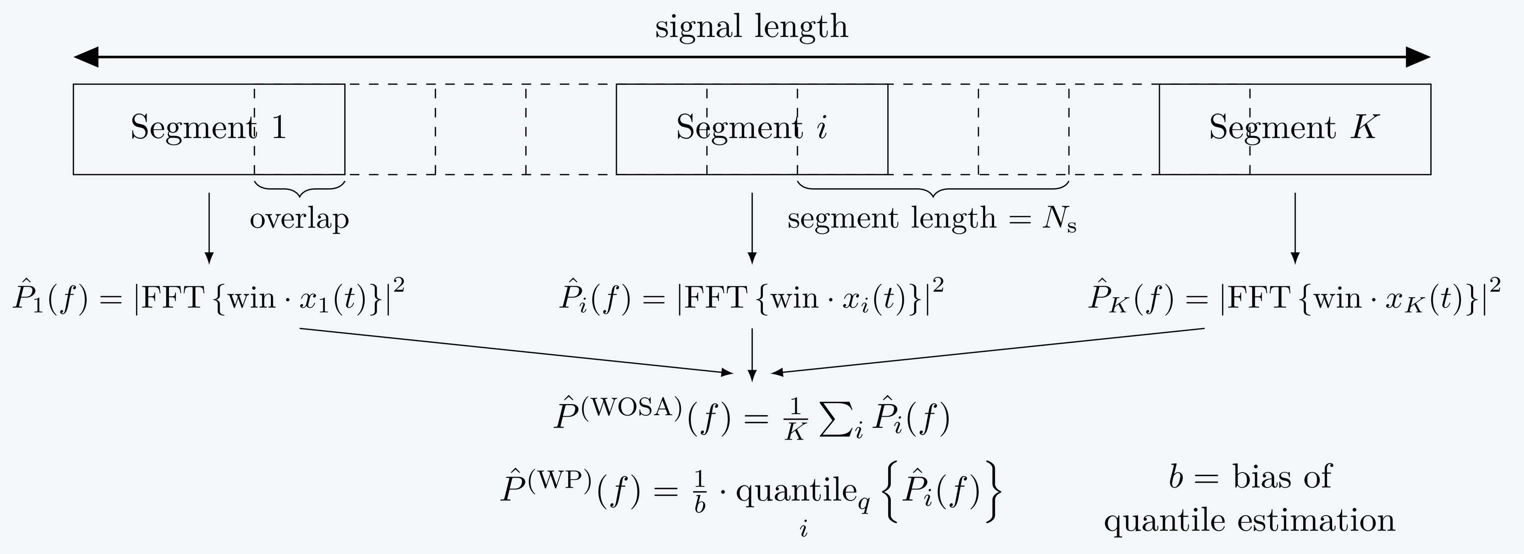


Figure 1. Procedure of computing the Welch's overlapped segment averaging (WOSA) and Welch percentile (WP) estimate.

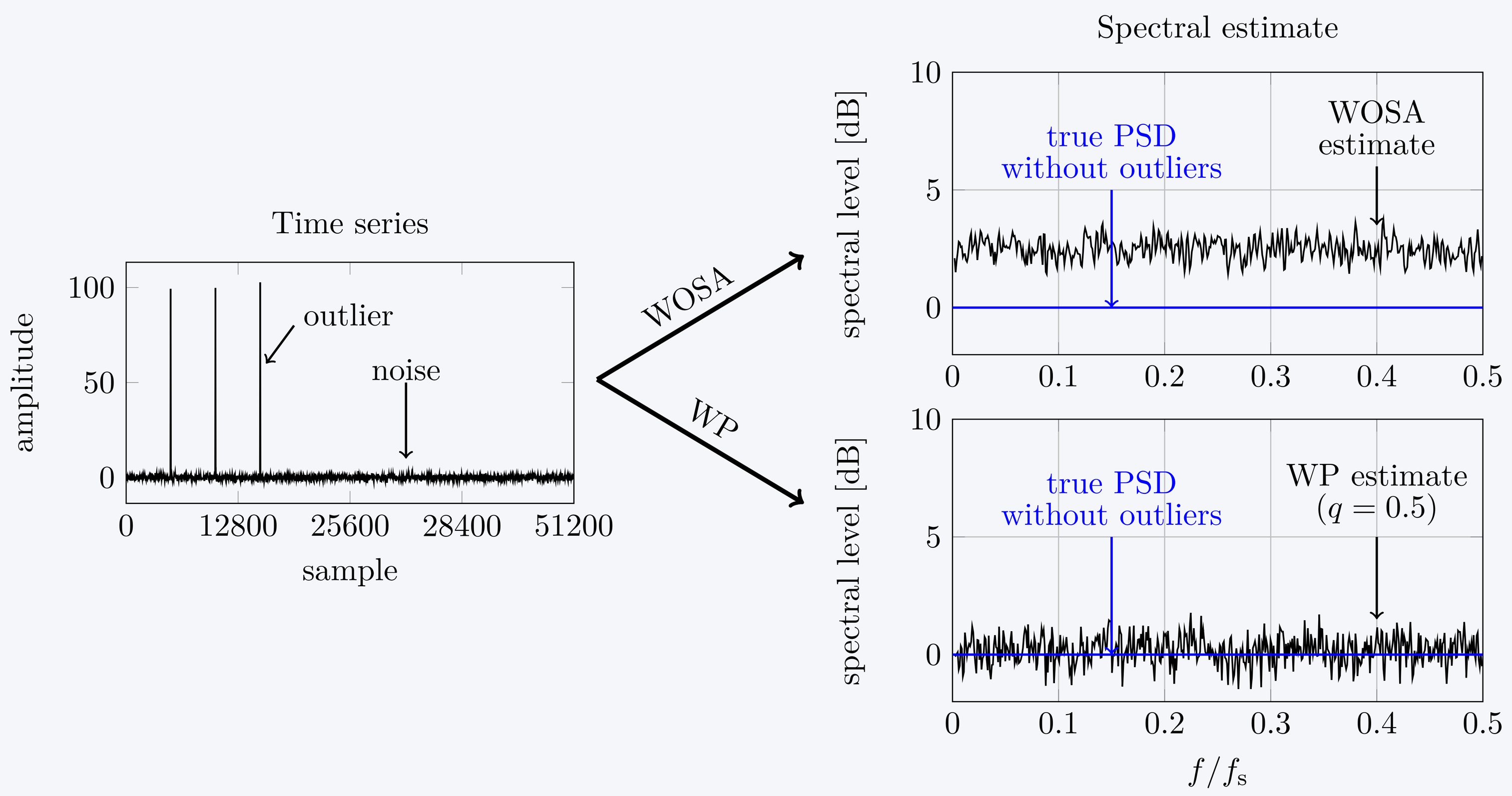


Figure 2. Welch's overlapped segment averaging (WOSA) and Welch percentile (WP) estimate in the presence of outliers.

Statistical Properties of WP Estimator

Statistical properties of WP estimator can be derived from order statistics and spectral estimation theory.

- **Bias of Quantile:**

$$b = \psi(K+2) - \psi(K(1-q)+1) \quad (1)$$

$$\psi(n) \approx \ln(n) - \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} - \frac{1}{252n^6}$$

- **Variance of WP Estimator:**

$$\mathrm{var}\left\{\hat{P}_q^{(\mathrm{WP})}\right\} = \frac{P^2}{b^2}\left[\psi_1(K(1-q)+1) - \psi_1(K+2)\right] \quad (2)$$

$$\psi_1(n) = \frac{\mathrm{d}\psi(n)}{\mathrm{d}n} \approx \frac{1}{n} + \frac{1}{2n^2} + \frac{1}{6n^3} - \frac{1}{30n^5} + \frac{1}{42n^7}$$

Variables
- b – bias of quantile estimate
- K – number of periodograms
- q – q^{th} quantile
- ψ – digamma function
- ψ_1 – trigamma function
- P – true PSD

- **Limiting Properties** ($K \to \infty$):

$$b = -\ln(1-q) \quad (3) \qquad \mathrm{var}\left\{\hat{P}_q^{(\mathrm{WP})}\right\} = \left(\frac{P}{b}\right)^2 \cdot \frac{q}{K(1-q)} \quad (4)$$

Simulations

Expressions for bias and variance in Equation (1) - (4) for various q and $K \geq 3$ are compared with simulations using a white Gaussian noise process. To reduce the variability in the estimates, 51 100 independent trials are averaged for each K and q. All data segments have a length of $N_s = 1024$ and a Hann data taper with 50 % overlap is used.

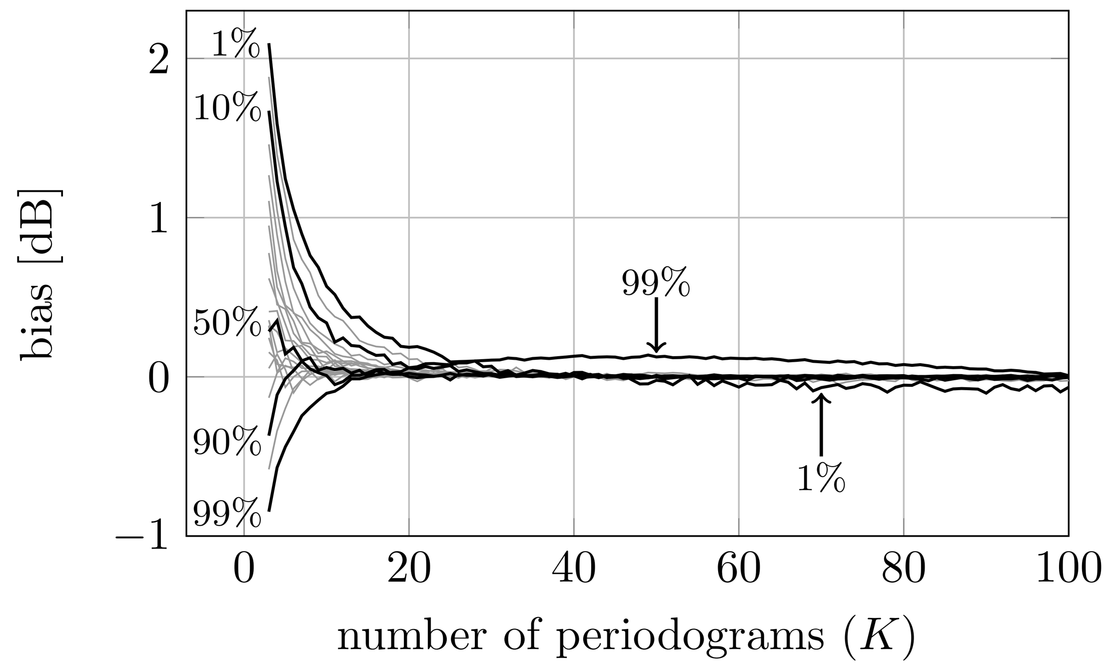


Figure 3. Remaining bias of Welch percentile (WP) estimate applied to white Gaussian noise sequence after correcting the quantile according to Equation (1). **The bias correction performs very well for most percentiles, and K ≥ 20.**

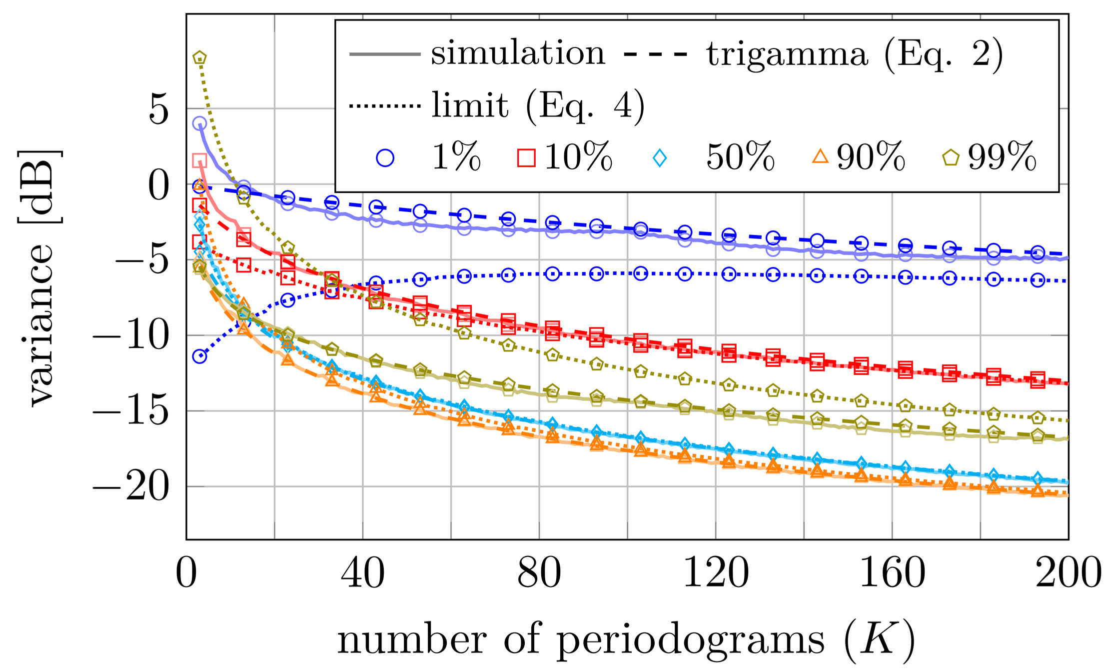
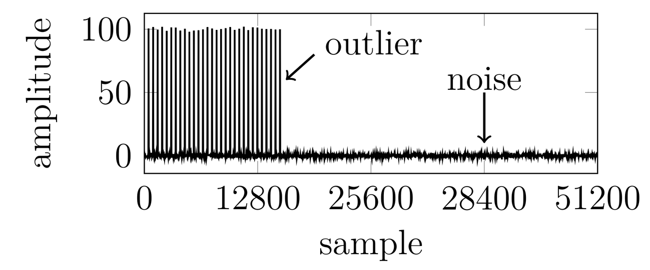


Figure 4. Variance of Welch percentile (WP) estimate applied to white Gaussian noise sequence and theoretical variance according to Equation (2) and (4). **Equation (2) deviates by less than 0.5 dB from the simulations for K ≥ 16 and percentiles between 10 % and 90 %.**

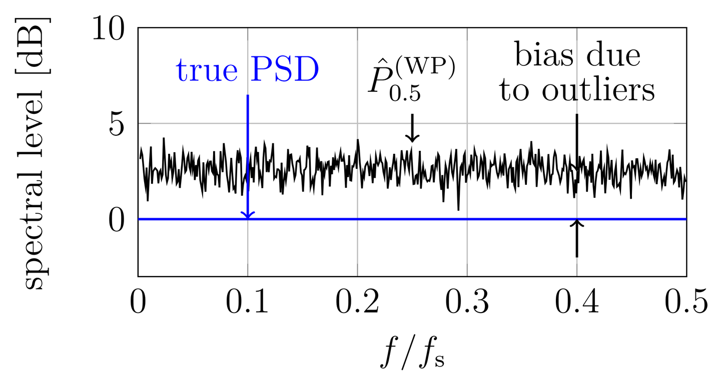
Extension: WP Estimator in Presence of Outliers

The WP estimator performs well if the percentage of outliers is small (≲5 %). For a large percentage of outliers, the estimator becomes increasingly biased compared to the true PSD (see Figure 5). To avoid this bias the original values for q and K in Equation (1) - (4) have to be modified by a factor e, which represents the fraction of outliers in the data:

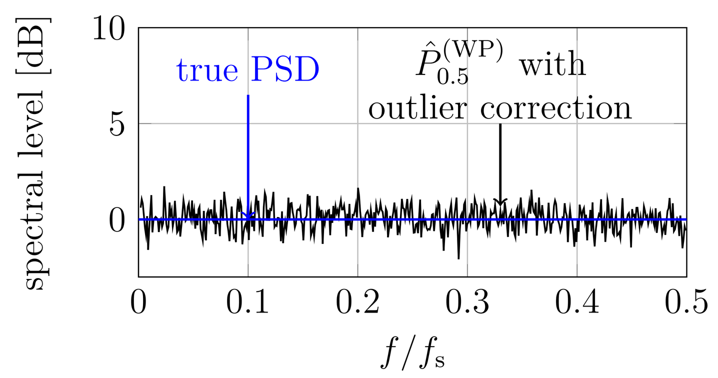
$$q \longrightarrow \frac{q}{1-e} \quad (5a) \qquad K \longrightarrow K(1-e), \quad (5b)$$



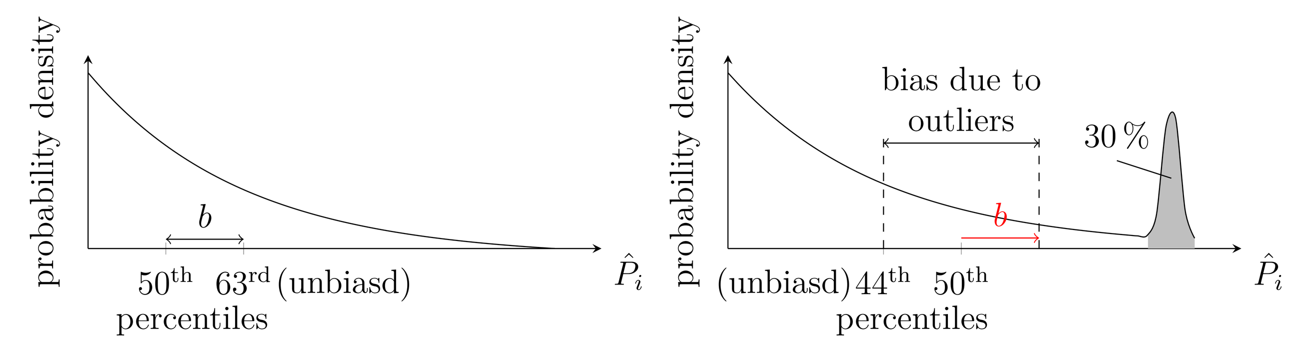
(a) White noise sequence in time domain



(b) WP estimate without outlier correction



(c) WP estimate with outlier correction



(d) PDF of periodograms without outliers (left) and with outliers (right)

Figure 5. (a) White noise time series with outliers such that 30 % of the periodgrams are effected. (b) WP estimate of signal in (a) without outlier correction. (c) WP estimate after applying the outlier correction according to Equation (5). (d) The origin of the bias in the WP estimate in (b) can be illustrated by comparing the percentiles in the probability density function (PDF) of the outlier-free periodograms (left) and in a PDF where 30 % of the periodograms are affected by outliers (right).

Conclusion

- the WP estimator is an outlier robust version of the standard Welch estimator
- simple expressions for bias, variance, and limiting distribution of the WP estimator can be derived from order statistics
- theoretical expressions show excellent agreement with simulations
- bias and variance expressions can easily be adapted if percentage of outliers is large

Acknowledgement

This research as supported by the Office of Naval Research grant number N00014-19-1-2644