

A WAVELET-BASED SPECTRAL METHOD FOR STEADY-STATE SIMULATION ANALYSIS

Emily K. Lada
 James R. Wilson

Graduate Program in Operations Research
 North Carolina State University
 Raleigh, NC 27695-7906, U.S.A.

Natalie M. Steiger

Maine Business School
 University of Maine
 Orono, ME 04469-5723, U.S.A.

ABSTRACT

We develop an automated wavelet-based spectral method for constructing an approximate confidence interval on the steady-state mean of a simulation output process. This procedure, called *WASSP*, determines a batch size and a warm-up period beyond which the computed batch means form an approximately stationary Gaussian process. Based on the log-smoothed-periodogram of the batch means, *WASSP* uses wavelets to estimate the batch means log-spectrum and ultimately the steady-state variance constant (SSVC) of the original (unbatched) process. *WASSP* combines the SSVC estimator with the grand average of the batch means in a sequential procedure for constructing a confidence-interval estimator of the steady-state mean that satisfies user-specified requirements on absolute or relative precision as well as coverage probability. An extensive performance evaluation provides evidence of *WASSP*'s robustness in comparison with some other output analysis methods.

1 INTRODUCTION

In a nonterminating simulation, we are interested in long-run (steady-state) average performance measures. Let $\{X_i : i = 1, 2, \dots\}$ denote a stochastic process representing the sequence of outputs generated by a single run of a nonterminating probabilistic simulation. If the simulation is in steady-state operation, then the random variables $\{X_i\}$ will have the same steady-state cumulative distribution function (c.d.f.) $F_X(x) = \Pr\{X_i \leq x\}$ for $i = 1, 2, \dots$, and for all real x .

Usually in a nonterminating simulation, we are interested in constructing point and confidence interval (CI) estimators for some parameter of the steady-state c.d.f. $F_X(x)$. In this work, we are primarily interested in estimating the steady-state mean, $\mu_X = E[X] = \int_{-\infty}^{\infty} x dF_X(x)$; and we limit the discussion to output processes for which $E[X_i^2] < \infty$ so that the process mean μ_X and process vari-

ance $\sigma_X^2 = \text{Var}[X_i] = E[(X_i - \mu_X)^2]$ are well defined. We let n denote the length of the time series $\{X_i\}$ of outputs generated by a single, long run of the simulation.

The sample mean, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, is an intuitively appealing point estimate of μ_X . Furthermore, if $\{X_i\}$ is weakly stationary, then the covariance of the process at lag l is $\gamma_X(l) = E[(X_i - \mu_X)(X_{i+l} - \mu_X)]$ for all $i \geq 1$ and $l = 0, \pm 1, \pm 2, \dots$; and the steady-state variance constant (SSVC) of the process is

$$\gamma_X = \sum_{l=-\infty}^{\infty} \gamma_X(l). \quad (1)$$

If $\sum_{l=-\infty}^{\infty} |\gamma_X(l)| < \infty$ and n is sufficiently large, then the variance of \bar{X} can be approximated by $\text{Var}[\bar{X}] \approx \gamma_X/n$; and for $0 < \beta < 1$, an asymptotically valid $100(1 - \beta)\%$ CI for μ_X is given by $\bar{X} \pm z_{1-\beta/2} \sqrt{\gamma_X/n}$, where $z_{1-\beta/2}$ is the $1 - \beta/2$ quantile of the standard normal distribution.

At the frequency ω expressed in cycles per time unit, the power spectrum $p_X(\omega)$ of the output process $\{X_i : i = 1, 2, \dots, n\}$ is defined as the cosine transform of the covariance function $\gamma_X(l)$,

$$p_X(\omega) = \sum_{l=-\infty}^{\infty} \gamma_X(l) \cos(2\pi\omega l) \text{ for } -\frac{1}{2} \leq \omega \leq \frac{1}{2} \quad (2)$$

(Heidelberger and Welch 1981). At frequencies of the form $\frac{l}{n}$ for $l = 0, 1, \dots, n-1$, an asymptotically unbiased estimate of the spectrum $p_X(\frac{l}{n})$ is given by the periodogram,

$$\begin{aligned} I\left(\frac{l}{n}\right) &= \frac{1}{n} \left\{ \left[\sum_{j=1}^n X_j \cos\left(\frac{2\pi(j-1)l}{n}\right) \right]^2 \right. \\ &\quad \left. + \left[\sum_{j=1}^n X_j \sin\left(\frac{2\pi(j-1)l}{n}\right) \right]^2 \right\} \\ &= |(\mathcal{F}X)_l|^2 / n \text{ for } l = 0, 1, \dots, n-1, \end{aligned} \quad (3)$$

where $\mathcal{F}X$ is the fast Fourier transform of the simulation-generated time series $X = \{X_1, \dots, X_n\}$. Letting $\{\chi_l^2(2) : l = 1, 2, \dots\}$ denote independent and identically distributed

(i.i.d.) chi-square random variates each with two degrees of freedom, we see the periodogram has the following asymptotic properties for large n :

$$\left. \begin{aligned} E[I(\frac{l}{n})] &\approx p_X(\frac{l}{n}), & \text{if } 0 < l < \frac{n}{2}, \\ \text{Var}[I(\frac{l}{n})] &\approx p_X^2(\frac{l}{n}), & \text{if } 0 < l < \frac{n}{2}, \\ I(\frac{l}{n}), I(\frac{j}{n}) &\text{ independent} & \text{if } 0 < l \neq j < \frac{n}{2}, \\ I(\frac{l}{n}) &\sim p_X(\frac{l}{n}) \chi_l^2(2)/2, & \text{if } 0 < l < \frac{n}{2} \end{aligned} \right\}. \quad (4)$$

Instead of working in the time domain with the original output process $\{X_i\}$, we are able to work in the frequency domain if we exploit a spectral analysis approach to steady-state simulation output analysis. At $\omega = 0$, we have

$$p_X(0) = \gamma_X = \sum_{l=-\infty}^{\infty} \gamma_X(l); \quad (5)$$

and consequently the goal of any spectral analysis method is to estimate $p_X(0)$ from the values of the periodogram in a neighborhood of zero frequency.

In this paper we develop an automated wavelet-based spectral method for constructing an approximate CI on the steady-state mean μ_X of a simulation output process $\{X_i\}$. This procedure, called \mathcal{WASSP} , uses wavelets to obtain an estimator $\widehat{\gamma}_X$ of γ_X ; and then \mathcal{WASSP} combines $\widehat{\gamma}_X$ with a version of the overall sample mean \bar{X} that has been suitably truncated if necessary to eliminate initialization bias so as to deliver an approximate $100(1 - \beta)\%$ CI of the form

$$\bar{X} \pm t_{1-\beta/2, \nu} \sqrt{\widehat{\gamma}_X/n}, \quad (6)$$

where ν is the “effective” degrees of freedom associated with $\widehat{\gamma}_X$ and $t_{1-\beta/2, \nu}$ is the $1 - \beta/2$ quantile of Student’s t -distribution with ν degrees of freedom. \mathcal{WASSP} is a sequential procedure and may request additional data iteratively before it delivers a final CI of the form (6) that has approximate coverage probability $1 - \beta$ and that satisfies a user-specified absolute or relative precision requirement.

The rest of this paper is organized as follows. A brief overview of \mathcal{WASSP} is given in §2; and the major steps of \mathcal{WASSP} are elaborated in §§2.1–2.4. Some results from our performance evaluation of \mathcal{WASSP} are presented in §3. Finally in §4 we summarize the main findings of this research. Lada (2003) provides a complete development of the results summarized in this paper.

2 OVERVIEW OF \mathcal{WASSP}

\mathcal{WASSP} requires the following user-supplied inputs:

- a simulation-generated output process $\{X_i : i = 1, \dots, n\}$ from which the steady-state expected response μ_X is to be estimated;

- the desired CI coverage probability $1 - \beta$, where $0 < \beta < 1$; and
- an absolute or relative precision requirement specifying the final CI half-length in terms of (a) a maximum acceptable half-length h^* , or (b) a maximum acceptable fraction r^* of the magnitude of the midpoint of the final CI.

\mathcal{WASSP} delivers the following outputs:

- a nominal $100(1 - \beta)\%$ CI for μ_X that satisfies the specified precision requirement, provided no additional data are required; or
- a new, larger sample size n to be supplied to \mathcal{WASSP} when it is executed again.

Figure 1 depicts a high-level flowchart of the operation of \mathcal{WASSP} . The algorithm begins by dividing the initial simulation-generated output process into a fixed number of batches of uniform size. Batch means are computed for all batches, and a randomness test is applied to the set of batch means. The randomness test serves two purposes:

- It is used to construct a set of spaced batch means such that the interbatch spacer preceding each batch is sufficiently large to ensure all computed batch means are approximately i.i.d. so that subsequently the batch means can be tested for normality.
- It is used to determine an appropriate data-truncation point—that is, the interbatch spacer preceding the first batch—beyond which all computed batch means are approximately independent of the simulation model’s initial conditions.

Once the randomness test is passed, the set of approximately i.i.d. spaced batch means is tested for normality. Each time the normality test is failed, the following steps are executed: (a) the batch size is increased; (b) a new set of spaced batch means is computed using the final spacer size determined by the randomness test; and (c) the normality test is repeated for the new set of spaced batch means.

Once the normality test is passed, all simulation-generated data beyond the warm-up period are used to compute adjacent (nonspaced) batch means of the batch size determined by the normality test; then the periodogram of the approximately normal batch means is computed and smoothed by taking a moving average of A points on the periodogram of the batch means. \mathcal{WASSP} allows the user to specify the value of A in the set $\{5, 7, 9, 11\}$, with the default taken as $A = 7$.

To obtain an estimator of the SSVC of the original (unbatched) process, we compute a wavelet-based estimator of the batch means log-spectrum by taking the discrete wavelet transform of the log of the smoothed periodogram of the batch means over the frequency range $(-\frac{1}{2}, \frac{1}{2})$. The estimated wavelet coefficients are thresholded using a variant of the thresholding algorithm of Gao (1997). From the thresholded wavelet approximation to the log-spectrum of the batch means, we compute an estimate of the spectrum

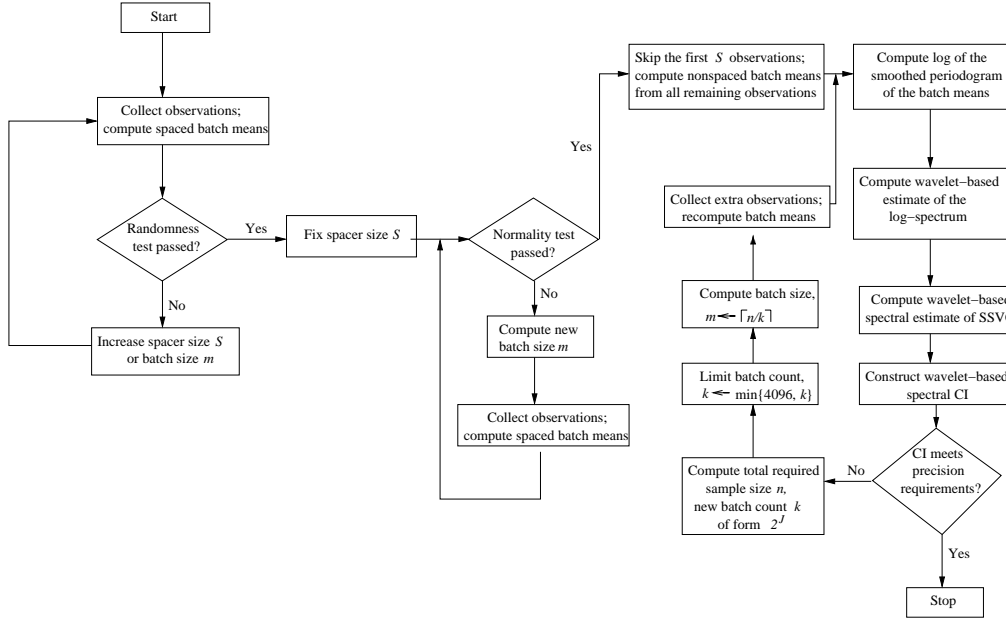


Figure 1: High-Level Flowchart of \mathcal{WASSP}

of the original (unbatched) process at zero frequency (that is, the SSVC); and finally we compute a CI of the form (6), where the midpoint of the CI is the grand average of all the adjacent (nonspaced) batch means that are computed after skipping the initial spacer.

The CI (6) is then tested to determine if it satisfies a user-specified absolute or relative precision requirement. If the precision requirement is satisfied, then \mathcal{WASSP} delivers the latest CI and terminates. Otherwise, the following steps are executed:

- a) The total required sample size is estimated; and on the assumption that the current batch size is maintained, the estimated batch count is expressed as the largest power of two yielding a total delivered sample size not exceeding the required sample size.
- b) If the estimated batch count exceeds 4,096, then the batch count is reduced to 4,096. Given the batch count, we adjust the batch size if necessary so that the total delivered sample size is not less than the total required sample size.
- c) The required additional observations are obtained (by restarting the simulation if necessary); and the batch means are recomputed using the latest batch size after skipping the initial spacer.
- d) The log of the smoothed periodogram for the new set of batch means is computed.
- e) A new estimate of the SSVC is obtained from a wavelet-based estimate of the log of the smoothed periodogram for the latest set of batch means.
- f) The CI (6) is recomputed and the precision requirement (stopping condition) is retested.

Note that if the CI (6) in step f) above fails to satisfy the precision requirement, then it is not necessary to repeat the independence test or the normality test; instead steps a)–f) are repeated until the precision requirement is satisfied.

2.1 Eliminating Initialization Bias

\mathcal{WASSP} begins by dividing the initial sample $\{X_i : i = 1, \dots, n\}$ into $k = 256$ batches of size $m = 16$. Let

$$\bar{X}_j = \bar{X}_j(m) = \frac{1}{m} \sum_{i=m(j-1)+1}^{mj} X_i \quad (7)$$

denote the j th batch mean for $j = 1, \dots, k$; and let

$$\bar{\bar{X}}(m, k) = \frac{1}{k} \sum_{j=1}^k \bar{X}_j(m) \quad (8)$$

denote the grand average of the k batch means.

The von Neumann test for randomness is applied to the batch means $\{\bar{X}_1(m), \dots, \bar{X}_k(m)\}$ by computing the ratio of the mean square successive difference of the batch means to the sample variance of the batch means; see von Neumann (1941). At the level of significance $\alpha_{\text{ind}} = 0.2$, we test the null hypothesis of independent, identically distributed batch means,

$$\mathcal{H}_{\text{ind}} : \{\bar{X}_j(m) : j = 1, \dots, k\} \text{ are i.i.d.}, \quad (9)$$

by computing the test statistic,

$$C_k = 1 - \frac{\sum_{j=1}^{k-1} [\bar{X}_j(m) - \bar{X}_{j+1}(m)]^2}{2 \sum_{i=1}^k [\bar{X}_i(m) - \bar{\bar{X}}(m, k)]^2}, \quad (10)$$

which is a relocated and rescaled version of the ratio of the mean square successive difference to the variance of the batch means. Since \mathcal{WASSP} 's test for randomness always involves at least 25 batch means, we use a normal approximation to the null distribution of the test statistic (10). If

$$|C_k| \leq z_{1-\alpha_{\text{ind}}/2} \sqrt{\frac{k-2}{k^2-1}}, \quad (11)$$

then the hypothesis (9) is accepted; otherwise the hypothesis (9) is rejected so that \mathcal{WASSP} must increase the spacer size before retesting (9). Through extensive experimentation, we found that setting $\alpha_{\text{ind}} = 0.2$ works well in practice and provides an effective balance between errors of type I and II in testing the hypothesis (9).

If the $k = 256$ adjacent batch means defined by (7) pass the randomness test (9)–(12) at the level of significance α_{ind} , then we set the number of batch means retained for the normality test according to $k' \leftarrow 256$; and we proceed to perform the normality test as detailed in §2.2 below. On the other hand, if the $k = 256$ batch means fail the test for randomness, then we insert *spacers* consisting of one ignored batch between the $k' \leftarrow 128$ remaining batch means that are to be retested for randomness; and thus the initial spacer size is $\mathcal{S} \leftarrow m$ observations. That is, every other batch mean, beginning with the second batch mean, is retained and the alternate batch means are ignored.

The $k' = 128$ remaining spaced batch means $\{\bar{X}_2(m), \bar{X}_4(m), \dots, \bar{X}_{256}(m)\}$ are retested for randomness. If the randomness retest is passed, then we proceed to perform the normality test in §2.2 with the current value of k' ; otherwise we add another ignored batch to each spacer so that the spacer size and number of remaining batches are updated according to

$$\mathcal{S} \leftarrow \mathcal{S} + m \quad \text{and} \quad k' \leftarrow \left\lfloor \frac{n}{m + \mathcal{S}} \right\rfloor. \quad (12)$$

Thus we now have $k' = 85$ remaining spaced batch means $\{\bar{X}_3(m), \bar{X}_6(m), \dots, \bar{X}_{255}(m)\}$; and again the remaining batch means are retested for randomness. This process is continued until one of the following conditions occurs: (a) the randomness test is passed; or (b) the randomness test is failed and in the update step (12), the batch count k' drops below the lower limit of 25 batches. If condition (a) occurs, then we proceed to the normality test in §2.2 with

the current value of k' . On the other hand if condition (b) occurs, then the batch size m is increased according to

$$m \leftarrow \left\lceil \sqrt{2m} \right\rceil; \quad (13)$$

the initial sample $\{X_i : i = 1, \dots, n\}$ is rebatched into $k \leftarrow 256$ adjacent (nonspaced) batches of size m ; and the k batch means are recomputed. If $n < km$, then more data must be collected and n must be updated before the adjacent (nonspaced) batch means can be recomputed and retested for randomness.

The process of testing for randomness is repeated, starting with the $k = 256$ recomputed adjacent (nonspaced) batch means of the current batch size m . If the randomness test is passed, then we set $k' \leftarrow k$ and proceed to the normality test in §2.2; otherwise the steps outlined in the two immediately preceding paragraphs are repeated. Once the randomness test is passed, there will be a set of k' approximately i.i.d. batch means, where $25 \leq k' \leq 256$. At this point, the spacer size \mathcal{S} is finalized; and the first \mathcal{S} observations $\{X_1, \dots, X_{\mathcal{S}}\}$ will henceforth be regarded as the warm-up period to be ignored in all subsequent calculations.

2.2 Testing Batch Means for Normality

Because the Shapiro-Wilk test for normality (Shapiro and Wilk 1965) requires a data set consisting of i.i.d. observations, we apply this test to the k' batch means that were retained after passing the preceding test for randomness. To assess the normality of the sample $\{\bar{X}_1(m), \dots, \bar{X}_{k'}(m)\}$, we start by sorting the observations in ascending order to obtain the order statistics $\bar{X}_{(1)}(m) \leq \bar{X}_{(2)}(m) \leq \dots \leq \bar{X}_{(k')}(m)$. The Shapiro-Wilk test statistic is then computed as follows,

$$W = \frac{\left\{ \sum_{l=1}^{\lfloor k'/2 \rfloor} \delta_{k'-l+1} [\bar{X}_{(k'-l+1)}(m) - \bar{X}_{(l)}(m)] \right\}^2}{\sum_{l=1}^{k'} [\bar{X}_l(m) - \bar{\bar{X}}(m, k)]^2}, \quad (14)$$

where the coefficients $\{\delta_{k'-l+1} : l = 1, \dots, \lfloor k'/2 \rfloor\}$ are tabulated in Shapiro and Wilk (1965). The test statistic W is then compared to the appropriate lower $100\alpha_{\text{nor}}\%$ critical value $w_{\alpha_{\text{nor}}}$ of the distribution of W under the null hypothesis of i.i.d. normal batch means,

$$\mathcal{H}_{\text{nor}} : \{\bar{X}_j(m) : j = 1, \dots, k'\} \stackrel{\text{i.i.d.}}{\sim} N[\mu_X, \sigma_{\bar{X}(m)}^2]. \quad (15)$$

If $W \leq w_{\alpha_{\text{nor}}}$, then at the level of significance α_{nor} we reject the hypothesis \mathcal{H}_{nor} that the retained batch means $\{\bar{X}_j(m) : j = 1, \dots, k'\}$ are i.i.d. normal.

For the first iteration of the normality test, the iteration counter is set to $i \leftarrow 1$ and the level of significance for the

Shapiro-Wilk test is $\alpha_{\text{nor}}(1) \leftarrow 0.05$. In general if on the i th iteration of the normality test (14)–(15) the hypothesis (15) is accepted at the level of significance $\alpha_{\text{nor}}(i)$, then we proceed to estimate the SSVC as outlined in §2.3; otherwise, we perform the following steps:

- a) The iteration counter i is increased, $i \leftarrow i + 1$, and the batch size m is increased, $m \leftarrow \lfloor \sqrt{2m} \rfloor$.
- b) The overall data set $\{X_1, \dots, X_n\}$ is redivided into k' batches of size m so that each batch of size m is separated by \mathcal{S} observations, where the final spacer size \mathcal{S} was fixed in the preceding test for randomness; and if necessary, additional simulation-generated data is collected to allow computation of k' spaced batch means with the new batch size m and the fixed spacer size \mathcal{S} .
- c) The spaced batch means $\{\bar{X}_j(m) : j = 1, \dots, k'\}$ are recomputed.
- d) The level of significance $\alpha_{\text{nor}}(i)$ for the current iteration i of the Shapiro-Wilk test is set as

$$\alpha_{\text{nor}}(i) \leftarrow \alpha_{\text{nor}}(1) \exp[-\tau(i-1)^2] \quad (16)$$

for $i = 1, 2, \dots$, where $\tau = 0.184206$.

- e) The k' spaced batch means $\{\bar{X}_j(m) : j = 1, \dots, k'\}$ are retested for normality at the level of significance $\alpha_{\text{nor}}(i)$.

If the normality hypothesis (15) is rejected in step *e*), then steps *a*)–*e*) above are repeated until the hypothesis is finally accepted so that we can proceed to estimating the SSVC as outlined in §2.3.

2.3 Estimating the SSVC via a Wavelet-Based Spectral Method

Once the normality test is passed, independence of the batch means is no longer required. Therefore, the first spacer consisting of the observations $\{X_1, X_2, \dots, X_{\mathcal{S}}\}$ is skipped (to eliminate initialization bias), and the remaining $n' = n - \mathcal{S}$ observations are rebatched into k adjacent (nonspaced) batches of size m . To construct the wavelet-based estimate of the log-spectrum of the batch means in a neighborhood of zero frequency, we see that the number of points in the log-periodogram (that is, the number of batch means k) must be a power of two. Therefore, k is set to the largest power of two less than or equal to n'/m ,

$$k = 2^{\lfloor \log_2(n'/m) \rfloor}, \quad (17)$$

where m is the final batch size required for the batch means to pass the normality test. For $j = 1, \dots, k$, the j th adjacent (nonspaced) batch mean $\bar{X}_j(m)$ is computed. The next step in \mathcal{WASSP} is to do the following: (a) smooth the periodogram of the batch means $\{\bar{X}_1(m), \dots, \bar{X}_k(m)\}$ by

computing a moving average with a sufficient number of points on the batch means periodogram; and (b) apply the logarithmic transformation to the smoothed periodogram of the batch means so as to obtain a reasonably stable estimator of the batch means log-spectrum.

2.3.1 Computing the Log-Smoothed-Periodogram of the Batch Means

The periodogram of the batch means process is computed by taking the fast Fourier transform of the adjacent (nonspaced) batch means $\bar{X} = \{\bar{X}_1(m), \dots, \bar{X}_k(m)\}$,

$$(\mathcal{F}\bar{X})_l = \sum_{j=1}^k \bar{X}_j(m) \exp[-2\pi(\sqrt{-1})(j-1)l/k] \quad \text{for } l = 1, 2, \dots, k-1. \quad (18)$$

Since we will be interested in obtaining an estimate of the log-spectrum of the batch means in a neighborhood of zero frequency using the values of the log-smoothed-periodogram of the batch means in a neighborhood of zero, we will use a full set of points of the periodogram on both sides of zero. The periodogram is symmetric about the origin, so that for $l = 1, 2, \dots, \frac{k}{2} - 1$,

$$I_{\bar{X}(m)}\left(\frac{l}{k}\right) = I_{\bar{X}(m)}\left(-\frac{l}{k}\right) = |(\mathcal{F}\bar{X})_l|^2 / k. \quad (19)$$

To compute the smoothed periodogram of the batch means based on a moving average of $A = 2a + 1$ periodogram values, we first must determine appropriate values for the periodogram at $l = 0$ and $l = \frac{k}{2}$. Using the definition (19), we see that the value of the periodogram at $l = 0$ is simply a scaled sum of the batch means and provides no information about the spectrum of the batch means at zero frequency. As an alternative, we take the value of the periodogram at $l = 0$ as follows,

$$I_{\bar{X}(m)}(0) \equiv \frac{1}{a} \sum_{l=1}^a I_{\bar{X}(m)}\left(\frac{l}{k}\right). \quad (20)$$

Moreover, we assume that for $l \neq 0$ and for a sufficiently small relative to k , the periodogram values $\{I_{\bar{X}(m)}\left(\frac{l+u}{k}\right) : u = 0, \pm 1, \dots, \pm a\}$ have expected values approximately equal to $p_{\bar{X}(m)}\left(\frac{l}{k}\right)$.

By the same reasoning that led to (20), we make the following definition of the periodogram at frequency $\frac{1}{2}$:

$$I_{\bar{X}(m)}\left(\frac{1}{2}\right) \equiv \frac{1}{a} \sum_{l=1}^a I_{\bar{X}(m)}\left(\frac{\lfloor k/2 \rfloor - l}{k}\right). \quad (21)$$

We define the periodogram at frequency $\frac{1}{2}$ for the sole purpose of facilitating wavelet-based estimation of the log-spectrum of the batch means, as described in Section 2.3.2 below.

We see that for $l = 1, 2, \dots, (\frac{k}{2} - 1)$,

$$I_{\bar{X}(m)}\left(\frac{\lfloor k/2 \rfloor + l}{k}\right) = I_{\bar{X}(m)}\left(\frac{\lfloor k/2 \rfloor - l}{k}\right); \quad (22)$$

and therefore for the frequency index set defined by

$$\mathcal{G}_k \equiv \{0, \pm 1, \dots, \pm (\frac{k}{2} - 1), \frac{k}{2}\}, \quad (23)$$

the smoothed periodogram of the batch means $\{\tilde{I}_{\bar{X}(m)}(\frac{l}{k}) : l \in \mathcal{G}_k\}$ can be computed as a moving average of $A = 2a + 1$ points,

$$\tilde{I}_{\bar{X}(m)}\left(\frac{l}{k}\right) = \frac{1}{A} \sum_{u=-a}^a I_{\bar{X}(m)}\left(\frac{l+u}{k}\right) \text{ for } l \in \mathcal{G}_k. \quad (24)$$

WASSP allows the user to select the value of the smoothing parameter A from the set of values $\{5, 7, 9, 11\}$. We selected $A = 7$ as the default since we found through extensive experimentation that setting $A = 7$ works well for a variety of types of simulation applications.

The natural log of the smoothed periodogram of the batch means,

$$\tilde{\mathcal{L}}_{\bar{X}(m)}\left(\frac{l}{k}\right) \equiv \ln\left[\tilde{I}_{\bar{X}(m)}\left(\frac{l}{k}\right)\right] \text{ for } l \in \mathcal{G}_k, \quad (25)$$

is used as an estimator of the log-spectrum of the corresponding batch means process,

$$\zeta_{\bar{X}(m)}(\omega) \equiv \ln\left[p_{\bar{X}(m)}(\omega)\right] \text{ for } \omega \in \left[-\frac{1}{2}, \frac{1}{2}\right]. \quad (26)$$

However, upon applying the properties of the periodogram in (4), we find that the expected value of our estimator (25) is

$$\begin{aligned} & \mathbb{E}\left\{\ln\left[\tilde{I}_{\bar{X}(m)}\left(\frac{l}{k}\right)\right]\right\} \\ & \approx \mathbb{E}\left\{\ln\left[\sum_{u=-a}^a p_{\bar{X}(m)}\left(\frac{l}{k}\right) w_u \frac{\chi_u^2(2)}{2}\right]\right\} \\ & = \zeta_{\bar{X}(m)}\left(\frac{l}{k}\right) + \mathbb{E}\left\{\ln\left[\frac{1}{2} \sum_{u=-a}^a w_u \chi_u^2(2)\right]\right\}, \end{aligned} \quad (27)$$

where $\{\chi_u^2(2) : u = 0, \pm 1, \dots, \pm a\}$ are i.i.d. chi-square random variables with two degrees of freedom and the $\{w_u : u = 0, \pm 1, \dots, \pm a\}$ are nonnegative deterministic weights such that $\sum_{u=-a}^a w_u = 1$. Therefore, when we take the log of the smoothed periodogram, bias is introduced and must be removed.

Table 1 shows the bias $\eta_l \equiv \mathbb{E}\left[\tilde{\mathcal{L}}_{\bar{X}(m)}\left(\frac{l}{k}\right)\right] - \zeta_{\bar{X}(m)}\left(\frac{l}{k}\right)$ at frequency $\frac{l}{k}$, where $l \in \mathcal{G}_k$. For the ‘‘effective’’ degrees of freedom ν_j that are inherent in the estimator $\tilde{\mathcal{L}}_{\bar{X}(m)}\left(\frac{l}{k}\right)$, we use the following result,

$$\nu_j^\# = \frac{2aA^2}{4a^2 - 2aj + 4a - 2j + 1} \text{ and } \nu_j = \lfloor \nu_j^\# \rfloor. \quad (28)$$

Note that in Table 1, the digamma function $\Psi(z)$ is defined in terms of the gamma function $\Gamma(z)$ as follows, $\Psi(z) \equiv \frac{d}{dz} \ln[\Gamma(z)] = \frac{\Gamma'(z)}{\Gamma(z)}$ for all z with $\text{Re}(z) > 0$. For a complete derivation of the bias terms in Table 1, see Lada (2003).

Table 1: Bias $\eta_l \equiv \mathbb{E}\left[\tilde{\mathcal{L}}_{\bar{X}(m)}\left(\frac{l}{k}\right)\right] - \zeta_{\bar{X}(m)}\left(\frac{l}{k}\right)$ of the Log-Smoothed-Periodogram at Frequency $\frac{l}{k}$, Where $l \in \mathcal{G}_k$

Frequency Index l	Bias η_l
$l = 0, \frac{k}{2}$	$\Psi(a) - \ln(a)$
$1 \leq l \leq a$	$\Psi\left(\frac{\nu_{ l }}{2}\right) - \ln\left(\frac{\nu_{ l }}{2}\right)$
$a < l < \frac{k}{2} - a$	$\Psi(A) - \ln(A)$
$\frac{k}{2} - a \leq l \leq \frac{k}{2} - 1$	$\Psi\left(\frac{\nu_{k/2- l }}{2}\right) - \ln\left(\frac{\nu_{k/2- l }}{2}\right)$

2.3.2 Using Wavelets to Estimate the Spectrum of the Batch Means

The next step in **WASSP** is to expand $\{\tilde{\mathcal{L}}_{\bar{X}(m)}(\frac{l}{k}) : l \in \mathcal{G}_k\}$ as a wavelet series to obtain a wavelet-based estimate of the log-spectrum of the batch means (26). To compute the discrete wavelet transform (DWT) of the log-smoothed-periodogram (25) of the batch means, we must have a power of two for the total number of points in (25) (Vidakovic 1999). To obtain such a sample size, we include the endpoint at frequency $\frac{1}{2}$ as discussed in (21) above.

To compute the DWT of the batch means log-smoothed-periodogram (25) using k data points, we first correct for the bias in each of the components of (25) so that we can compute the DWT

$$\tilde{\mathbf{W}} = \Theta \tilde{\mathcal{L}}, \quad (29)$$

where the vector $\tilde{\mathcal{L}}$ has elements $\{\tilde{\mathcal{L}}_{\bar{X}(m)}(\frac{l}{k}) - \eta_l : l \in \mathcal{G}_k\}$; and the matrix Θ defines the DWT associated with the $s8$ symmlet (Bruce and Gao 1996).

Since the total number of points in the bias-corrected log of the smoothed periodogram of the batch means has the form $k = 2^J$, we found in practice that an excellent wavelet approximation to $\{\tilde{\mathcal{L}}_{\bar{X}(m)}(\frac{l}{k}) - \eta_l : l \in \mathcal{G}_k\}$ can be

obtained by setting the number of resolution levels L in the wavelet approximation as follows,

$$L \equiv \left\lfloor \frac{J}{2} \right\rfloor = \left\lfloor \frac{\log_2(k)}{2} \right\rfloor; \quad (30)$$

and the coarsest level of resolution j_0 is then given by $j_0 = J - L$. This will give a total of 2^{j_0} coefficients at the coarsest level of resolution.

After computing the DWT of the bias-corrected log-smoothed-periodogram of the batch means $\tilde{\mathcal{L}}$ as given by (29), we threshold the resulting wavelet coefficients $\{\hat{d}_{j,l}^* : j = j_0, \dots, J-1; l = 0, \dots, 2^j - 1\}$ using Gao's (1997) thresholding scheme with the soft threshold

$$\lambda_j = \max \left[\frac{\pi}{\sqrt{6k}} \sqrt{2 \ln(k)}, 2^{-(J-j-1)/4} \ln(2k) \right] \quad (31)$$

on the magnitude of the retained wavelet coefficients at level j for $j = j_0, \dots, J-1$ to obtain the coefficients $\{\hat{d}_{j,l}^*\}$. The empirical scaling coefficients $\{\hat{c}_{j_0,l}^* : l = 0, \dots, 2^{j_0} - 1\}$ are not thresholded since it is presumed they contain information about the coarse features of the log-spectrum of the batch means.

The next step is to compute the inverse DWT,

$$\tilde{\mathcal{L}}^* = \Theta^T \tilde{\mathcal{W}}^*, \quad (32)$$

where $\tilde{\mathcal{W}}^*$ is the vector containing the scaling coefficients $\{\hat{c}_{j_0,l}^* : l = 0, \dots, 2^{j_0} - 1\}$ and thresholded wavelet coefficients $\{\hat{d}_{j,l}^* : j = j_0, \dots, J-1; l = 0, \dots, 2^j - 1\}$. Thus (32) yields the thresholded wavelet approximation

$$\begin{aligned} \tilde{\mathcal{L}}^* &= [\tilde{\mathcal{L}}_1^*, \dots, \tilde{\mathcal{L}}_k^*]^T \\ &= \left[\tilde{\mathcal{L}}_{\bar{X}(m)}^* \left(\frac{-(\frac{k}{2} - 1)}{k} \right), \dots, \tilde{\mathcal{L}}_{\bar{X}(m)}^* \left(\frac{1}{2} \right) \right]^T \end{aligned}$$

to the vector $\tilde{\mathcal{L}}$ of the bias-corrected log of the smoothed periodogram of the batch means. Therefore, our wavelet-based estimate of the log-spectrum of the batch means is

$$\hat{\zeta}_{\bar{X}(m)} \left(\frac{l}{k} \right) = \tilde{\mathcal{L}}_{\bar{X}(m)}^* \left(\frac{l}{k} \right) \text{ for } l \in \mathcal{G}_k. \quad (33)$$

The wavelet-based estimate of the spectrum of the batch means process can now be computed from (33) as

$$\hat{p}_{\bar{X}(m)} \left(\frac{l}{k} \right) = \exp \left[\hat{\zeta}_{\bar{X}(m)} \left(\frac{l}{k} \right) \right] \text{ for } l \in \mathcal{G}_k; \quad (34)$$

and a wavelet-based estimate of the SSVC for the original (unbatched) process is recovered from (34) as follows:

$$\hat{\gamma}_X = m \cdot \hat{p}_{\bar{X}(m)}(0). \quad (35)$$

An approximate $100(1 - \beta)\%$ confidence interval for μ_X is then given by

$$\overline{\bar{X}}(m, k) \pm t_{1-\beta/2, 2a} \sqrt{\frac{\hat{\gamma}_X}{n'}}, \quad (36)$$

where $n' = mk$ and $\overline{\bar{X}}(m, k)$ is the grand average of the k batch means $\{\bar{X}_1(m), \dots, \bar{X}_k(m)\}$.

2.4 Fulfilling the Precision Requirement

The half-length of the CI (36) is given by

$$H = t_{1-\beta/2, 2a} \sqrt{\frac{\hat{\gamma}_X}{n'}} = t_{1-\beta/2, 2a} \sqrt{\frac{\hat{p}_{\bar{X}(m)}(0)}{k}}.$$

If the CI (36) satisfies the precision requirement

$$H \leq H^*, \quad (37)$$

where H^* is given by

$$H^* = \begin{cases} \infty, & \text{for no prec. reqmt.,} \\ r^* |\overline{\bar{X}}(m, k)|, & \text{for rel. prec. level } r^*, \\ h^*, & \text{for abs. prec. level } h^*, \end{cases} \quad (38)$$

then \mathcal{WASSP} terminates and the CI (36) is delivered.

If the precision requirement (37) is not satisfied, then the total number of batches required to satisfy the precision requirement is computed as follows,

$$k^* \leftarrow \left\lceil \left(\frac{H}{H^*} \right)^2 k \right\rceil;$$

and thus the total sample size required to meet the precision requirement is k^*m . However, since the number of batches must be a power of two, the batch count k is set for the next iteration of \mathcal{WASSP} as follows:

$$k \leftarrow \min \left\{ 2^{\lceil \log_2(k^*) \rceil}, 4096 \right\}, \quad (39)$$

where 4,096 is the upper bound on the number of batch means used in \mathcal{WASSP} .

The new batch size m for the next iteration of \mathcal{WASSP} is assigned according to

$$m \leftarrow \left\lceil \left(\frac{k^*}{k} \right) m \right\rceil, \quad (40)$$

so that the total sample size n is increased approximately by a factor of $(H/H^*)^2$.

On the next iteration of \mathcal{WASSP} , the total sample size including the warm-up period is thus given by $n \leftarrow km + \mathcal{S}$, where the corresponding batch count k and batch size m are given by (39) and (40), respectively. The additional simulation-generated observations are obtained by restarting the simulation or by retrieving the extra data from storage; and then the next iteration of \mathcal{WASSP} is performed.

3 PERFORMANCE EVALUATION OF \mathcal{WASSP}

Lada (2003) details an extensive performance evaluation of \mathcal{WASSP} . In this section we summarize the results of applying \mathcal{WASSP} to the $M/M/1$ queue waiting time process. Here X_i is the waiting time for the i th customer in a single-server queueing system with i.i.d. exponential interarrival times having mean $10/9$, i.i.d. exponential service times having mean 1, a steady-state server utilization of 90%, and an empty-and-idle initial condition. The steady-state mean waiting time is $\mu_X = 9.0$.

The $M/M/1$ queue waiting time process with 90% server utilization and empty-and-idle initial condition is a particularly difficult test process for the following reasons: (a) the initial transient is pronounced and persists over an extended period of time; (b) the correlation function decays slowly with increasing lags once the system has reached steady-state operation; (c) the marginal distribution of waiting times is markedly nonnormal; and (d) the spectrum of the batch means process, $p_{\bar{X}(m)}(\omega)$, is sharply peaked in the neighborhood of zero frequency. This test process will allow thorough evaluation of the robustness of \mathcal{WASSP} 's procedure for eliminating initialization bias as well the robustness of \mathcal{WASSP} 's wavelet-based technique for estimating the SSVC of the original waiting time process $\{X_i\}$.

The parameters used to evaluate the performance of \mathcal{WASSP} are the coverage probability of its CIs, the mean and half-length of its CIs, and the total required sample size. We performed independent replications of \mathcal{WASSP} to construct nominal 90% CIs that satisfy a specified precision requirement. The following three precision requirements were used:

- no precision—that is, we set $h^* = \infty$ in (37) and (38) so \mathcal{WASSP} delivers the CI (6) using the batch count and batch size required to pass the randomness and normality tests;
- $\pm 15\%$ precision—that is, \mathcal{WASSP} delivers the CI (6) satisfying the relative precision requirement given by (37) and (38) with $r^* = 0.15$; and
- $\pm 7.5\%$ precision—that is, \mathcal{WASSP} delivers the CI (6) satisfying the relative precision requirement given by (37) and (38) with $r^* = 0.075$.

For the sake of comparison, we also applied ASAP2 (Steiger et al. 2002) and Heidelberg and Welch's spectral method (Heidelberg and Welch 1981a, 1981b, 1983) to the same $M/M/1$ queue waiting time process. To make a

fair comparison of the performance of \mathcal{WASSP} with that of Heidelberg and Welch's sequential spectral method, we first applied \mathcal{WASSP} to a realization of a particular process so as to obtain not only the corresponding \mathcal{WASSP} -based CI but also a complete data set to which we may apply the Heidelberg-Welch procedure. In particular on each replication of \mathcal{WASSP} and Heidelberg and Welch's procedure for the $\pm 15\%$ and the $\pm 7.5\%$ precision cases, we set the maximum run length t_{\max} for Heidelberg and Welch's method to the final sample size required by \mathcal{WASSP} for that replication. For the no precision case, we applied the Heidelberg and Welch method to the first 4,096 observations of the data set used by \mathcal{WASSP} for each replication.

Based on all our computational experience with \mathcal{WASSP} , we believe that the results given below are typical of the performance of \mathcal{WASSP} in many practical applications. Since each CI with a nominal coverage probability of 90% was replicated 1,000 times for \mathcal{WASSP} and Heidelberg and Welch's method, the standard error of each coverage estimator is approximately 0.95%. The coverage probabilities for ASAP2 have a standard error of approximately 1.5% since only 400 replications of ASAP2 were performed.

As can be seen from Table 2, \mathcal{WASSP} outperforms Heidelberg and Welch's method with respect to CI coverage for all three precision requirements. Furthermore, since Heidelberg and Welch's method terminates once t_{\max} is reached, it is possible that the Heidelberg and Welch algorithm could run out of data before the precision requirement is satisfied. Of the 1,000 CIs delivered by Heidelberg and Welch's method, only 767 actually satisfied the precision requirement for the $\pm 15\%$ precision level and only 673 satisfied the precision requirement for the $\pm 7.5\%$ precision level.

From Table 2 it is also evident that the variance of the CI half-length is markedly higher for ASAP2 than for \mathcal{WASSP} in the no precision case. This suggests that \mathcal{WASSP} produces much more stable CIs than ASAP2 in the no precision case. Once a precision requirement is imposed, ASAP2 produces CIs that exhibit the same stability as the CIs produced by \mathcal{WASSP} , as can be seen from the variance of the CI half-length for the $\pm 15\%$ and $\pm 7.5\%$ precision cases.

4 CONCLUSIONS

\mathcal{WASSP} is a completely automated wavelet-based spectral procedure for constructing an approximate confidence interval for the steady-state mean of a simulation output process. Our extensive performance evaluation of \mathcal{WASSP} indicates that \mathcal{WASSP} outperforms Heidelberg and Welch's method; and we believe \mathcal{WASSP} represents an advance in spectral methods for simulation output analysis. Furthermore, we can conclude that while \mathcal{WASSP} and ASAP2 produce comparable results in some cases, \mathcal{WASSP} is in general a more robust procedure than ASAP2. In particular, in the absence

Table 2: Performance of \mathcal{WASSP} (Using $A = 7$), Heidelberg and Welch's Spectral Method (H&W), and ASAP2 for the $M/M/1$ Queue Waiting Time Process with 90% Server Utilization and Empty-and-Idle Initial Condition, Where Results Are Based on Independent Replications of Nominal 90% CIs

Precision Requirement	Performance Measure	Procedure		
		\mathcal{WASSP}	H&W	ASAP2
None	# replications	1,000	1,000	400
	CI coverage	83.2%	78.6%	88%
	Avg. sample size	12,956	4,096	22,554
	Max. sample size	123,102	4,096	131,072
	Avg. CI half-length	3.1776	4.0415	6.44
	Var. CI half-length	2.5342	4.6892	167.0
$\pm 15\%$	# replications	1,000	1,000	400
	CI coverage	83.6%	79.6%	90%
	Avg. sample size	88,782	65,282	93,374
	Max. sample size	819,248	314,464	260,624
	Avg. CI half-length	1.106	1.3154	1.18
	Var. CI half-length	0.0368	0.3765	0.025
$\pm 7.5\%$	# replications	1,000	1,000	400
	CI coverage	90.8%	84.1%	92%
	Avg. sample size	371,380	298,860	281,022
	Max. sample size	1,871,936	1,216,420	796,076
	Avg. CI half-length	0.5914	0.6852	0.630
	Var. CI half-length	0.006	0.059	0.002

of a precision requirement, ASAP2-generated CIs can be highly variable in their half-lengths. We are continuing the experimental evaluation of \mathcal{WASSP} ; and future developments concerning \mathcal{WASSP} will be available on the website www.ie.ncsu.edu/jwilson.

REFERENCES

- Bruce, A., and H.-Y. Gao. 1996. *Applied wavelet analysis with S-PLUS*. New York: Springer.
- Gao, H.-Y. 1997. Choice of thresholds for wavelet shrinkage estimate of the spectrum. *Journal of Time Series Analysis* 18 (3): 231–251.
- Heidelberg, P., and P. D. Welch. 1981a. A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM* 24 (4): 233–245.
- Heidelberg, P., and P. D. Welch. 1981b. Adaptive spectral methods for simulation output analysis. *IBM Journal of Research and Development* 25 (6): 860–876.
- Heidelberg, P., and P. D. Welch. 1983. Simulation run length control in the presence of an initial transient. *Operations Research* 31 (6): 1109–1144.
- Lada, E. K. 2003. A wavelet-based procedure for steady-state simulation output analysis. Doctoral dissertation, Graduate Program in Operations Research, North Carolina State University, Raleigh, North Carolina. Available online via www.lib.ncsu.edu/theses/available/etd-04032003-141616/unrestricted/etd.pdf.
- Shapiro, S. S., and M. B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52 (3–4): 591–611.
- Steiger, N. M., E. K. Lada, J. R. Wilson, C. Alexopoulos, D. Goldsman, and F. Zouaoui. 2002. ASAP2: An improved batch means procedure for simulation output analysis. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 336–344. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via <http://www.informs-cs.org/wsc02papers/043.pdf> [accessed December 16, 2002].
- Vidakovic, B. 1999. *Statistical modeling by wavelets*. New York: Wiley.
- von Neumann, J. 1941. Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics* 12:367–395.

AUTHOR BIOGRAPHIES

EMILY K. LADA is a Senior Research Scientist at Old Dominion University's Virginia Modeling, Analysis & Simulation Center. She is a member of IIE and INFORMS. Her e-mail address is elada@odu.edu.

JAMES R. WILSON is Professor and Head of the Industrial Engineering Department at North Carolina State University. He is a member of ASA and INFORMS, and he is a Fellow of IIE. His e-mail address is jjwilson@eos.ncsu.edu, and his web page is www.ie.ncsu.edu/jwilson.

NATALIE M. STEIGER is an Assistant Professor of Production and Operations Management in the University of Maine Business School. She is a member of IIE and INFORMS. Her e-mail address is nsteiger@maine.edu.