Proceedings of the 2004 Winter Simulation Conference
R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds.

PERFORMANCE EVALUATION OF A WAVELET-BASED SPECTRAL METHOD FOR STEADY-STATE SIMULATION ANALYSIS

Emily K. Lada

Statistical and Applied Mathematical Sciences Institute
Research Triangle Park, NC 27709-4006, U.S.A.

James R. Wilson

Department of Industrial Engineering
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.

Jeffrey A. Joines

Department of Textile Engineering, Chemistry, & Science
North Carolina State University
Raleigh, NC 27695-7913, U.S.A.

ABSTRACT

We summarize the results of an experimental performance evaluation of WASSP, an automated wavelet-based spectral method for constructing an approximate confidence interval on the steady-state mean of a simulation output process so that the delivered confidence interval satisfies user-specified requirements on absolute or relative precision as well as coverage probability. We applied WASSP to test problems designed specifically to explore its efficiency and robustness in comparison with ASAP3 and the Heidelberger-Welch algorithm, two sequential procedures based respectively on the methods of nonoverlapping batch means and spectral analysis. Concerning efficiency, WASSP compared favorably with its competitors, often requiring smaller sample sizes to deliver confidence intervals with the same nominal levels of precision and coverage probability. Concerning robustness against the statistical anomalies commonly encountered in simulation studies, WASSP outperformed its competitors, delivering confidence intervals whose actual half-lengths and coverage probabilities were frequently closer to the corresponding user-specified nominal levels.

1 INTRODUCTION

A nonterminating simulation is one in which we are interested in long-run (steady-state) average performance measures. Usually in a nonterminating probabilistic simulation, we seek to compute point and confidence-interval estimators for some parameter, or characteristic, of the steady-state cumulative distribution function (c.d.f.) of a particular simulation-generated response. In Lada and Wilson (2004), 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 first a batch size and a warm-up period beyond which the computed batch means form an approximately stationary Gaussian process. For this purpose we use the randomness test of von Neumann (1941) to determine an interbatch spacer preceding each batch that is sufficiently large to ensure the resulting spaced batch means are approximately independent and identically distributed (i.i.d.). We then take the spacer preceding the first batch to be the warm-up period, and we use the univariate normality test of Shapiro and Wilk (1965) to determine a batch size that is sufficiently large to ensure the spaced batch means are approximately normal.

Next WASSP computes the discrete wavelet transform of the bias-corrected log-smoothed-periodogram of the batch means; and the resulting wavelet coefficients are denoised by applying a soft-thresholding scheme. Then by computing the inverse discrete wavelet transform of the thresholded wavelet coefficients, WASSP delivers an estimator of the batch means log-spectrum and ultimately the steady-state variance constant (SSVC) of the original (unbatched) process—that is, the sum of the covariances at all lags for the original process. Finally WASSP combines the estimator of the SSVC 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.

Given the output process $\{X_u : u = 1, 2, \ldots, n\}$, WASSP is designed to deliver a $100(1-\beta)\%$ confidence interval for the steady-state mean μ_X having the form

$$\overline{\overline{X}} \pm t_{1-\beta/2,\nu}\sqrt{\frac{\widehat{\gamma}_X}{n'}}, \qquad (1)$$

where: (*i*) n' is the length of the truncated output process after deleting (if necessary) a warm-up period containing initialization bias; (*ii*) the grand mean $\overline{\overline{X}}$ and the SSVC estimator $\widehat{\gamma}_X$ are computed from the truncated output process; (*iii*) ν denotes the "effective" degrees of freedom associated with $\widehat{\gamma}_X$; and (*iv*) $t_{1-\beta/2,\nu}$ denotes the $1-\beta/2$ quantile of Student's t-distribution with ν degrees of freedom, provided $0 < \beta < 1$.

In this paper, we summarize some experimental results that are representative of the performance we observed in applying WASSP and other selected procedures for steady-state simulation output analysis to a suite of five particularly difficult test problems. The experimental performance evaluation reported in this paper is focused on the following test problems:

a) an $M/M/1$ queue waiting time process for which the underlying system has long-run server utilization equal to 0.90 and an empty-and-idle initial condition;

b) the first-order autoregressive (AR(1)) process that has lag-one correlation equal to 0.995, white noise variance equal to one, steady-state mean equal to 100, and initial condition equal to zero; and

c) the "AR(1)-to-Pareto" (ARTOP) process that has marginals given by a Pareto distribution with lower limit and shape parameter equal to 1 and 2.1, respectively (implying the marginal mean and variance are both finite while the marginal skewness and kurtosis are both infinite), and that is obtained by applying to process *b*) above the composite of the inverse of the specified Pareto c.d.f. and the standard normal c.d.f.

For each of the five test problems (including the test problems *a*)–*c*) discussed in this paper), we used the following measures to evaluate the performance of WASSP and its competitors: (*i*) the empirical coverage probability of the delivered confidence intervals; (*ii*) the mean and variance of the half-lengths of the delivered confidence intervals; and (*iii*) the mean and maximum of the total required sample sizes. We performed independent replications of each simulation analysis procedure to construct nominal 90% and 95% confidence intervals that satisfy a given precision requirement, specified as either a maximum percentage of the magnitude of the batch means grand average (for a relative precision requirement), or as a maximum absolute half-length (for an absolute precision requirement). We used the following three precision requirements:

- no precision—that is, we set the maximum confidence interval half-length equal to infinity so the final confidence interval delivered by WASSP was based on the batch count and batch size required to pass the randomness and normality tests;
- ± 15% precision—that is, the half-length of the final confidence interval delivered by WASSP was less than or equal to 15% of the magnitude of the midpoint of the confidence interval; and
- ± 7.5% precision—that is, the half-length of the final confidence interval delivered by WASSP was less than or equal to 7.5% of the magnitude of the midpoint of the confidence interval.

For the sake of comparison, we also applied ASAP3, the batch means algorithm of Steiger et al. (2004), and the spectral method of Heidelberger and Welch (1983) to the five selected test problems. Lada et al. (2004) provide a full discussion of the experimental results summarized in this paper.

The rest of this paper is organized as follows. In §§2–4 we present the results of applying WASSP to the $M/M/1$ queue waiting time process, the AR(1) process, and the ARTOP process. The results for ASAP3 and the method of Heidelberger and Welch (H&W) for all three test problems are also presented in these sections. Conclusions and directions for future research are given in §5.

2 THE *M/M/*1 QUEUE WAITING TIME PROCESS

For the first test problem, we let X_u denote the waiting time for the uth customer, $u = 1, 2, \ldots,$ 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 (so that $X_1 = 0$). The theoretical mean for this waiting time process is $\mu_X = 9.0$.

In WASSP the batch means periodogram is smoothed by computing a moving average of $A = 2a + 1$ points, where $2 \le a \le 5$. As explained in Lada and Wilson (2004), at zero frequency the resulting smoothed periodogram is approximately a chi-squared random variable with $2a$ degrees of freedom that has been scaled by the multiplier $\gamma_X/(2a)$. As a consequence, the final $100(1-\beta)$% confidence interval delivered by WASSP is based on the $1-\beta/2$ percentile of Student's t-distribution with $2a$ degrees of freedom. The user selects the value of the smoothing parameter A from the set of values $\{5, 7, 9, 11\}$, with the default being $A = 7$. Table 1 shows the performance of WASSP for the $M/M/1$ queue waiting time process using the smoothing parameter values $A = 5, 7, 9,$ and 11. The results are based on 1,000 independent replications of nominal 90% confidence intervals (CIs). Table 2 shows the corresponding results for nominal 95% CIs. The standard error is less than 1% for each coverage estimator in Tables 1 and 2.

From these tables, it is evident that the coverage probability decreases in general as the smoothing parameter

Table 1: Performance of WASSP Using Different Values of A for the $M/M/1$ Queue Waiting Time Process with 90% Server Utilization and Empty-and-Idle Initial Condition; Results Are Based on 1,000 Independent Replications of Nominal 90% CIs

Prec. Req.	Performance Measure	Smoothing Parameter $A=5$	$A=7$	$A=9$	$A=11$
None	CI coverage	88.8%	87.7%	86.1%	84.2%
	Avg. sample size	18,369	18,090	17,696	18,369
	Avg. CI half-length	3.3957	3.0715	2.9116	2.6684
	Var. CI half-length	2.6495	2.0026	1.6165	1.2476
±15%	CI coverage	89.6%	87.2%	83.5%	82.8%
	Avg. sample size	114,710	92,049	79,824	68,533
	Avg. CI half-length	1.1000	1.1103	1.1223	1.1451
	Var. CI half-length	0.0414	0.0387	0.0381	0.0340
±7.5%	CI coverage	93.6%	90.4%	88.5%	91.5%
	Avg. sample size	467,370	388,000	341,380	322,990
	Avg. CI half-length	0.5846	0.5866	0.5855	0.5911
	Var. CI half-length	0.0072	0.0072	0.0067	0.0060

Table 2: Performance of WASSP Using Different Values of A for the $M/M/1$ Queue Waiting Time Process with 90% Server Utilization and Empty-and-Idle Initial Condition; Results Are Based on 1,000 Independent Replications of Nominal 95% CIs

Prec. Req.	Performance Measure	Smoothing Parameter $A=5$	$A=7$	$A=9$	$A=11$
None	CI coverage	94.4%	93.4%	92%	89.2%
	Avg. sample size	18,816	17,971	19,296	18,116
	Avg. CI half-length	4.6864	3.9987	3.6049	3.4274
	Var. CI half-length	5.1663	3.6999	2.3715	2.1640
±15%	CI coverage	96%	93%	91.5%	89%
	Avg. sample size	196,060	143,920	123,220	113,160
	Avg. CI half-length	1.1364	1.1342	1.1351	1.1440
	Var. CI half-length	0.0341	0.0314	0.0305	0.0301
±7.5%	CI coverage	98%	97%	96.3%	95.5%
	Avg. sample size	809,840	598,020	532,600	480,820
	Avg. CI half-length	0.5897	0.5950	0.5960	0.5962
	Var. CI half-length	0.0064	0.0056	0.0061	0.0057

increases. This is due to the target process having a power spectrum with a sharp peak in the neighborhood of zero frequency. As the smoothing parameter A is increased, WASSP's estimate of the power spectrum near zero frequency becomes flatter, resulting in an estimate of the SSVC that is biased low. For the no precision case, clearly $A = 5$ yields the best results in terms of coverage probability. If one were only interested in generating an initial, or pilot, CI for the steady-state mean of this process without imposing a precision requirement, then it might be desirable to change the default smoothing parameter from $A = 7$ to $A = 5$. Similarly, for the ±15% precision case, the results for $A = 5$ appear to be better than those for $A = 7$. However, for the ±7.5% case, there is significant overcoverage for $A = 5$; and it appears that asymptotically, the default smoothing parameter $A = 7$ produces better results than $A = 5$. For $A = 9$ and $A = 11$, the coverage probabilities for the ±7.5% precision case are also excellent. However, the small sample results for $A = 9$ and $A = 11$ are not as good as those for $A = 5$ and $A = 7$.

In summary, it is evident from Tables 1 and 2 that while there may be slight differences in the results for the allowable values of A, setting $A = 5, 7, 9$, or 11 yields acceptable results for this system and WASSP appears to be robust in terms of the smoothing parameter A.

2.1 Summary of Experimental Results for the *M/M/*1 Queue Waiting Time Process

Table 3 shows a comparison of the performance of WASSP (using $A = 7$) and ASAP3 for the $M/M/1$ queue waiting time process. The coverage probabilities for ASAP3 have a standard error of approximately 1.5% for nominal 90% CIs and a standard error of approximately 1% for nominal 95% CIs since only 400 replications of ASAP3 were performed. The coverage probabilities for WASSP have a standard error of 0.95% for nominal 90% CIs and a standard error of 0.69% for nominal 95% CIs since we performed 1,000 replications of WASSP. From this table, it appears that in the no precision case, WASSP and ASAP3 yield similar results in terms of CI coverage. However, WASSP requires nearly half as many observations as ASAP3.

Table 3: Performance of WASSP (Using $A = 7$) and ASAP3 for the $M/M/1$ Queue Waiting Time Process with 90% Server Utilization and Empty-and-Idle Initial Condition; Results Are Based on Independent Replications of Nominal 90% and 95% CIs

Prec. Req.	Performance Measure	90% CIs WASSP	90% CIs ASAP3	95% CIs WASSP	95% CIs ASAP3
None	# replications	1,000	400	1,000	400
	CI coverage	87.7%	87.5%	93.4%	91.5%
	Avg. sample size	18,090	31,181	17,971	31,181
	Max. sample size	241,152	185,344	171,456	185,344
	Avg. CI half-length	3.0715	2.0719	3.9987	2.5209
	Var. CI half-length	2.0026	0.3478	3.6999	0.5350
±15%	# replications	1,000	400	1,000	400
	CI coverage	87.2%	91%	93%	95.5%
	Avg. sample size	92,049	103,742	143,920	140,052
	Max. sample size	688,256	424,536	953,424	418,263
	Avg. CI half-length	1.1103	1.1820	1.1342	1.2059
	Var. CI half-length	0.0387	0.0259	0.0314	0.0205
±7.5%	# replications	1,000	400	1,000	400
	CI coverage	90.4%	89.5%	97%	94%
	Avg. sample size	388,000	287,568	598,020	382,958
	Max. sample size	2,614,458	700,700	3,408,016	956,610
	Avg. CI half-length	0.5866	0.6273	0.5950	0.6324
	Var. CI half-length	0.0072	0.0023	0.0056	0.0020

For the case of ±7.5% precision, Table 3 indicates that WASSP and ASAP3 perform essentially the same, suggesting that as the relative precision requirement goes to zero, WASSP and ASAP3 will produce comparable results for this test process in terms of coverage probability, average

CI half-length, and variance of the CI half-length. The average sample size is higher for WASSP than for ASAP3, however.

Table 4 shows the results for WASSP and the H&W spectral method for nominal 90% and 95% CIs. Since it is possible that the H&W algorithm could run out of data before the precision requirement is satisfied, we included in Table 4 the overall coverage (for all 1,000 replications, whether or not the precision requirement was met) as well as the coverage for those CIs that satisfied the precision requirement (Satisfied coverage). Furthermore, Table 4 reports the estimated mean squared error of the grand mean,

$$\widehat{\mathrm{MSE}}\big[\,\overline{\overline{X}}(m,k)\,\big] = \frac{1}{1000}\sum_{u=1}^{1000}\big[\,\overline{\overline{X}}_u(m_u,k_u) - \mu_X\,\big]^2, \qquad (2)$$

where on replication u of WASSP or the H&W procedure, $\overline{\overline{X}}_u(m_u,k_u)$ denotes the delivered grand mean based on k_u batches of size m_u for $u = 1,\ldots,1{,}000$; and the corresponding estimated standard error of (2),

$$\widehat{\mathrm{SE}}\Big\{\widehat{\mathrm{MSE}}\big[\,\overline{\overline{X}}(m,k)\big]\Big\} = \frac{1}{(1000\times 999)^{1/2}} \qquad (3)$$
$$\times\left(\sum_{u=1}^{1000}\Big\{\big[\,\overline{\overline{X}}_u(m_u,k_u)-\mu_X\big]^2 - \widehat{\mathrm{MSE}}\big[\,\overline{\overline{X}}(m,k)\big]\Big\}^2\right)^{1/2}.$$

The statistics (2) and (3) provide an indication of the amount of bias associated with the final point estimator $\overline{\overline{X}}(m,k)$ delivered by WASSP or the H&W procedure. This bias is the result of a combination of two different effects. First, the grand average of the truncated batch means $\overline{\overline{X}}(m,k)$ is influenced in general by residual initialization bias—after all, there is no unique, well-defined end of the warm-up period for the $M/M/1$ waiting times or more generally for the output responses generated by many types of discrete-event stochastic systems. Second, the truncated simulation run length $n' = mk$ is random. Moreover, the truncation point $\mathcal{S}$ determined by WASSP or the H&W procedure (that is, the end of the warm-up period) is also a random variable. It follows that $\overline{\overline{X}}(m,k) = \overline{X}(n') = \sum_{u=\mathcal{S}+1}^{\mathcal{S}+n'} X_u\big/n'$ is a ratio of two random variables; and for such a ratio estimator, in general we have

$$\mathrm{E}\Bigg[\sum_{u=\mathcal{S}+1}^{\mathcal{S}+n'} X_u\Big/n'\Bigg] \neq \mathrm{E}\Bigg[\sum_{u=\mathcal{S}+1}^{\mathcal{S}+n'} X_u\Bigg]\Big/\mathrm{E}[n']; \qquad (4)$$

see §6.3 and §6.8 of Cochran (1977). It is clear that when WASSP or the H&W procedure is applied to the process $\{X_u : u = 1,2,\ldots\}$, the truncation point $\mathcal{S}$ and the total sample size $n = \mathcal{S} + n'$ are both *stopping times* for the process (Ross 1983); and thus, for example, if the $\{X_u\}$ were i.i.d., then we would have

$$\begin{aligned}\mathrm{E}\Bigg[\sum_{u=\mathcal{S}+1}^{\mathcal{S}+n'} X_u\Bigg] &= \mathrm{E}\Bigg[\sum_{u=1}^{\mathcal{S}+n'} X_u\Bigg] - \mathrm{E}\Bigg[\sum_{u=1}^{\mathcal{S}} X_u\Bigg]\\ &= \mathrm{E}[\mathcal{S}+n']\mu_X - \mathrm{E}[\mathcal{S}]\mu_X \qquad (5)\\ &= \mathrm{E}[n']\mu_X, \qquad (6)\end{aligned}$$

where (5) follows from two applications of Wald's equation (Ross 1983). Combining (4) and (6), we see that in general the grand average of the truncated batch means $\overline{\overline{X}}(m,k)$ delivered by WASSP or the H&W procedure is a biased estimator of the steady-state mean,

$$\mathrm{E}\big[\,\overline{\overline{X}}(m,k)\,\big] \neq \mu_X, \qquad (7)$$

not only because of initialization effects but also because of ratio-estimator bias due to randomness of the truncation point and the final total sample size.

Table 4: Performance of WASSP (Using $A = 7$) and the H&W Spectral Method for the $M/M/1$ Queue Waiting Time Process with 90% Server Utilization and Empty-and-Idle Initial Condition; Results Are Based on 1,000 Independent Replications of Nominal 90% and 95% CIs

Prec. Req.	Performance Measure	90% CIs WASSP	90% CIs H&W	95% CIs WASSP	95% CIs H&W
None	# reps.	1,000	1,000	1,000	1,000
	Overall coverage	87.7%	67.8%	93.4%	76.2%
	Avg. sample size	18,090	2,714	17,971	2,696
	Max. sample size	241,152	36,173	171,456	25,719
	Avg. CI half-length	3.0715	4.0535	3.9987	5.1817
	Var. CI half-length	2.0026	4.4582	3.6999	7.9996
	$\widehat{\mathrm{MSE}}[\overline{\overline{X}}(m,k)]$	2.8688	11.3983	3.4439	12.0800
	$\widehat{\mathrm{SE}}\{\widehat{\mathrm{MSE}}[\overline{\overline{X}}(m,k)]\}$	0.1767	0.5860	0.2423	0.6447
	# reps. satisfying	1,000	1,000	1,000	1,000
	Satisfied coverage	87.7%	67.8%	93.4%	76.2%
±15%	# reps.	1,000	1,000	1,000	1,000
	Overall coverage	87.2%	81.3%	93%	88.6%
	Avg. sample size	92,049	62,112	143,920	98,838
	Max. sample size	688,256	348,434	953,424	482,673
	Avg. CI half-length	1.1103	1.1486	1.1342	1.1550
	Var. CI half-length	0.0387	0.0406	0.0314	0.0347
	$\widehat{\mathrm{MSE}}[\overline{\overline{X}}(m,k)]$	0.5588	0.7596	0.3545	0.4972
	$\widehat{\mathrm{SE}}\{\widehat{\mathrm{MSE}}[\overline{\overline{X}}(m,k)]\}$	0.0362	0.0409	0.0211	0.0273
	# reps. satisfying	1,000	939	1,000	944
	Satisfied coverage	87.2%	80.9%	93%	88.4%
±7.5%	# reps.	1,000	1,000	1,000	1,000
	Overall coverage	90.4%	85%	97%	91.8%
	Avg. sample size	388,000	275,610	598,020	431,590
	Max. sample size	2,614,458	1,323,572	3,408,016	1,517,616
	Avg. CI half-length	0.5866	0.5899	0.5950	0.5903
	Var. CI half-length	0.0072	0.0072	0.0056	0.0078
	$\widehat{\mathrm{MSE}}[\overline{\overline{X}}(m,k)]$	0.1151	0.1692	0.0680	0.1016
	$\widehat{\mathrm{SE}}\{\widehat{\mathrm{MSE}}[\overline{\overline{X}}(m,k)]\}$	0.0053	0.0087	0.0034	0.0053
	# reps. satisfying	1,000	918	1,000	917
	Satisfied coverage	90.4%	83.9%	97%	91.1%

When WASSP is applied to a given steady-state simulation model, asymptotically as the relative precision specification tends to zero, we see that the final batch size $m \to \infty$ while the final batch count k remains in the range $256 \le k \le 4{,}096$ so that the truncation point $\mathcal{S} \to \infty$ and the final truncated sample size $n' \to \infty$. A similar conclusion applies to the operation of WASSP in the case that the absolute precision specification tends to zero. These properties ensure that WASSP's point estimator $\overline{\overline{X}}(m,k) = \overline{X}(n')$ converges to μ_X with probability one as the relevant precision specification tends to zero. Although a similar statement can be made about the asymptotic unbiasedness of the point estimator delivered by the H&W procedure as the relative precision specification tends to zero and the user-specified maximum sample size $t_{\max} \to \infty$, it should be noted that the H&W procedure provides no mechanism for estimating the value of $t_{\max}$ necessary to deliver a CI with the required relative precision.

In the no precision case for the $M/M/1$ waiting times, we see from Table 4 that the mean squared error for the final point estimator $\overline{\overline{X}}(m,k)$ delivered by the H&W method is nearly four times that for WASSP. Beyond the usual ratio-estimator bias due to the randomness of its truncation point and its final sample size, the H&W procedure exhibits significant bias due to initialization effects because its built-in Cramér–von Mises test for such bias is not effective in detecting and eliminating those effects. Table 5 summarizes the results obtained from versions of the H&W algorithm that were implemented with and without the Cramér–von Mises test. Notice that in Table 5, we let HW−CV denote the original Heidelberger-Welch (1981) procedure without the Cramér–von Mises test.

From Table 5, it is clear that using the Cramér–von Mises statistic to detect and eliminate initialization bias has a negligible effect on overall CI coverage, CI half-length, final sample size, and the estimate of the mean squared error of the truncated grand mean. There does appear to be some improvement in the precision of the CIs delivered by the H&W algorithm that incorporates the Cramér–von Mises test, as evidenced by the variance of the CI half-length and the number of replications satisfying the precision requirement. We found that the results in Table 5 are indicative of the results obtained for the AR(1) and ARTOP processes as well. That is, we did not see a significant improvement in general in the performance of the H&W algorithm by incorporating the Cramér–von Mises test into the procedure. For a complete summary of the results obtained by applying the HW−CV method, see Lada (2003). Note that throughout this paper, all references to the H&W procedure mean the Heidelberger-Welch spectral procedure that incorporates the Cramér–von Mises test for initialization bias as detailed in Heidelberger and Welch (1983).

Table 5: Performance of the H&W Procedure versus That of HW−CV, the Version of H&W without the Cramér–von Mises Test for Initialization Bias; Results for the $M/M/1$ Queue Waiting Time Process with 90% Server Utilization and Empty-and-Idle Initial Condition, Based on 1,000 Independent Replications of Nominal 90% and 95% CIs

Prec. Req.	Performance Measure	90% CIs HW−CV	90% CIs H&W	95% CIs HW−CV	95% CIs H&W
±15%	# reps.	1,000	1,000	1,000	1,000
	Overall coverage	79.6%	81.3%	87%	88.6%
	Avg. sample size	65,282	62,112	104,290	98,838
	Max. sample size	314,464	348,434	482,306	482,673
	Avg. CI half-length	1.3154	1.1486	1.3521	1.1550
	Var. CI half-length	0.3765	0.0406	0.3791	0.0347
	$\widehat{\text{MSE}}[\overline{\overline{X}}(m,k)]$	0.894	0.7596	0.6325	0.4972
	$\widehat{\text{SE}}\{\widehat{\text{MSE}}[\overline{\overline{X}}(m,k)]\}$	0.0442	0.0409	0.0313	0.0273
	# reps. satisfying	767	939	717	944
	Satisfied coverage	75%	80.9%	83.7%	88.4%
±7.5%	# reps.	1,000	1,000	1,000	1,000
	Overall coverage	84.10%	85%	92.6%	91.8%
	Avg. sample size	298,860	275,610	458,310	431,590
	Max. sample size	1,216,420	1,323,572	1,841,421	1,517,616
	Avg. CI half-length	0.6852	0.5899	0.6910	0.5903
	Var. CI half-length	0.0599	0.0072	0.0523	0.0078
	$\widehat{\text{MSE}}[\overline{\overline{X}}(m,k)]$	0.2186	0.1692	0.1378	0.1016
	$\widehat{\text{SE}}\{\widehat{\text{MSE}}[\overline{\overline{X}}(m,k)]\}$	0.0117	0.0087	0.0077	0.0053
	# reps. satisfying	673	918	673	917
	Satisfied coverage	79.35%	83.9%	89.6%	91.1%

Returning to the overall results obtained by applying WASSP and the H&W procedure to the $M/M/1$ queue waiting times as summarized in Table 4, we see that once a precision requirement is imposed on either procedure and the sample size begins to increase, the mean squared error $\widehat{\text{MSE}}\big[\overline{\overline{X}}(m,k)\big]$ begins to decrease with the diminishing effects of initialization bias and ratio-estimator bias on the final point estimator $\overline{\overline{X}}(m,k)$. Moreover, Table 4 clearly reveals that the final point estimator delivered by WASSP is less influenced by initialization bias or ratio-estimator bias than is the final point estimator delivered by the H&W procedure.

From Table 4, we also see that the H&W method consistently requires smaller average sample sizes than those required by WASSP. It is also evident that the coverages of the CIs delivered by the H&W method are consistently much less than the coverages of the CIs generated by WASSP. The unacceptably low coverage probabilities for CIs produced by the H&W method in the no precision case provide further evidence of the ineffectiveness of the Cramér–von Mises test in detecting and eliminating initialization bias. Essentially, the Cramér–von Mises test is being passed at very small sample sizes t_i; and as a result, the sample $\{X_u : u = 1, \ldots, t_i\}$ used to construct the H&W confidence interval of the form (1) is simply not large enough to estimate accurately the steady-state mean of the process.

3 THE FIRST-ORDER AUTOREGRESSIVE (AR(1)) PROCESS

The next test process used in the performance evaluation of WASSP is the first-order autoregressive process. Let $\{\delta_u : u = 1, 2, \ldots\}$ be a white noise process that is randomly sampled from $N(0, \sigma_\delta^2)$. We define an autoregressive process of order one (that is, an AR(1) process) as follows,

$$X_u = \mu_X + \rho(X_{u-1} - \mu_X) + \delta_u \quad \text{for} \quad u = 1, 2, \ldots, \tag{8}$$

where μ_X is the steady-state mean of the process and ρ is the lag-one correlation of the process in steady-state operation. To generate the AR(1) process $\{X_u : u = 1, \ldots, n\}$ from (8), we first set $X_0 = 0$, corresponding to an empty-and-idle initial condition. We then set the autoregressive parameter $\rho = 0.995$, the mean $\mu_X = 100$, and the variance of the white noise process $\sigma_\delta^2 = 1$. One of the most difficult aspects of this test process is its exceptionally long initial transient period.

Table 6 shows a comparison of the performance of WASSP (using $A = 7$) and ASAP3 for the AR(1) process (8) . Notice that for many of the cases contained in Table 6, the actual precision of the confidence intervals delivered by ASAP3 is significantly lower (i.e., better) than the requested precision level. For example, the average confidence interval half-length for the no precision case for nominal 90% confidence intervals is 2.325. This indicates that at the no precision level, ASAP3 is delivering CIs whose half-lengths are within about ±2% of the mean. Furthermore, WASSP is delivering CIs at the no precision level that are within about ±6% of the mean. Consequently, in Table 6 we have included the results for the ±3.75%, the ±1.875%, and the ±0.9375% precision cases to provide an indication of the asymptotic behavior of the two methods. For all five precision levels, ASAP3-generated confidence intervals are exhibiting significant overcoverage and the sample sizes for ASAP3 in the no precision and the ±7.5% precision cases are about four times higher than for WASSP. From the ±0.9375% results, however, it appears that WASSP and ASAP3 are yielding comparable results asymptotically.

Table 7 shows the results of applying WASSP and the H&W spectral method to the AR(1) process. From Table 7, examination of the statistic $\widehat{\text{MSE}}\big[\,\overline{\overline{X}}(m, k)\,\big]$ reveals significant bias in the estimate of the mean for the H&W method. As mentioned in §2.1, the Cramér–von Mises test is not sensitive enough to detect and eliminate initialization bias. For the no precision and the ±15% precision cases, H&W-based nominal 90% confidence intervals exhibit significant undercoverage; however, for the ±7.5% precision case the confidence intervals exhibit slight overcoverage. Overall, the H&W spectral method completely breaks down in the small-sample cases; and clearly WASSP outperforms the H&W method for the AR(1) process.

Table 6: Performance of WASSP (Using $A = 7$) and ASAP3 for the AR(1) Process (8) with $\mu_X = 100$, $X_0 = 0$, $\rho = 0.995$, and $\sigma_\delta^2 = 1$ Based on Independent Replications of Nominal 90% and 95% CIs

Prec. Req.	Performance Measure	90% CIs WASSP	90% CIs ASAP3	95% CIs WASSP	95% CIs ASAP3
None	# replications	1,000	400	1,000	400
	CI coverage	90.9%	95.5%	94.5%	98.8%
	Avg. sample size	9,866	41,076	9,824	41,076
	Max. sample size	30,208	68,864	22,592	68,864
	Avg. CI half-length	5.3048	2.325	6.7331	2.825
	Var. CI half-length	1.8280	0.170	2.8826	0.270
±7.5%	# replications	1,000	400	1,000	400
	CI coverage	91%	95.5%	95%	98.8%
	Avg. sample size	9,975	41,076	9,820	41,076
	Max. sample size	24,612	68,864	21,504	68,864
	Avg. CI half-length	5.2512	2.325	6.0977	2.825
	Var. CI half-length	1.3046	0.170	1.0834	0.270
±3.75%	# replications	1,000	400	1,000	400
	CI coverage	87%	95.5%	95%	98.8%
	Avg. sample size	13,535	41,076	21,099	41,208
	Max. sample size	48,288	68,864	59,360	68,864
	Avg. CI half-length	3.2133	2.325	3.2800	2.817
	Var. CI half-length	0.1420	0.170	0.1529	0.257
±1.875%	# replications	1,000	400	1,000	400
	CI coverage	93.5%	93.5%	97.7%	99.25%
	Avg. sample size	57,449	68,474	90,371	101,526
	Max. sample size	166,308	147,877	328,160	223,173
	Avg. CI half-length	1.6481	1.7603	1.6584	1.7703
	Var. CI half-length	0.0423	0.0134	0.0429	0.0120
±0.9375%	# replications	1,000	400	1,000	400
	CI coverage	94%	94.25%	98%	97.25%
	Avg. sample size	229,730	213,826	333,050	254,920
	Max. sample size	717,388	381,184	967,136	384,512
	Avg. CI half-length	0.8297	0.8941	0.8667	0.8959
	Var. CI half-length	0.0105	0.0026	0.0115	0.0021

4 THE AR(1)-TO-PARETO (ARTOP) PROCESS

The next system used to test the performance of WASSP is the "AR(1)-to-Pareto," or ARTOP process. Let $\{Z_u : u = 1, 2, \ldots\}$ be a stationary AR(1) process with $N(0, 1)$ marginals and lag-one correlation ρ. The process $\{Z_u : u = 1, 2, \ldots\}$ can be generated as follows:

$$Z_u = \rho Z_{u-1} + \delta_u \quad \text{for} \quad u = 1, 2, \ldots, \tag{9}$$

where $Z_0 \sim N(0, 1)$ and $\{\delta_u : u = 1, 2, \ldots\} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_\delta^2)$ is a white noise process with variance $\sigma_\delta^2 = \sigma_Z^2(1 - \rho^2) = 1 - \rho^2$. If $\{X_u : u = 1, 2, \ldots\}$ is an ARTOP process with marginal c.d.f.

$$F_X(x) \equiv \Pr\{X \le x\} = \begin{cases} 1 - (\xi/x)^{\vartheta}, & x \ge \xi, \\ 0, & x < \xi, \end{cases} \tag{10}$$

where $\xi > 0$ is a location parameter and $\vartheta > 0$ is a shape parameter, then $\{X_u\}$ is generated from the "base process" (9) as follows. First, the base process $\{Z_u : u = 1, 2, \ldots\}$ is fed into the standard normal c.d.f. to get a sequence of correlated, uniform(0,1) random numbers $\{R_u : u = 1, 2, \ldots\}$; that is,

Table 7: Performance of WASSP (Using $A = 7$) and the H&W Spectral Method for the AR(1) Process (8) with $\mu_X = 100$, $X_0 = 0$, $\rho = 0.995$, and $\sigma_\delta^2 = 1$; Results Are Based on 1,000 Independent Replications of Nominal 90% and 95% CIs

Prec. Req.	Performance Measure	90% CIs WASSP	90% CIs H&W	95% CIs WASSP	95% CIs H&W
None	# reps.	1,000	1,000	1,000	1,000
	Overall coverage	90.9%	46.9%	94.5%	65.9%
	Avg. sample size	9,866	1,480	9,824	1,474
	Max. sample size	30,208	4,532	22,592	3,389
	Avg. CI half-length	5.3048	13.4345	6.7331	16.7078
	Var. CI half-length	1.8280	3.0321	2.8826	4.5517
	$\widehat{\text{MSE}}[\overline{\overline{X}}(m,k)]$	8.6470	218.5032	9.4243	226.7135
	$\widehat{\text{SE}}\{\widehat{\text{MSE}}[\overline{\overline{X}}(m,k)]\}$	0.4211	4.4678	0.5423	4.5781
	# reps. satisfying	1,000	1,000	1,000	1,000
	Satisfied coverage	90.9%	46.9%	94.5%	65.9%
±15%	# reps.	1,000	1,000	1,000	1,000
	Overall coverage	89%	64.3%	95%	92.9%
	Avg. sample size	10,007	2,187	9,882	3,711
	Max. sample size	21,504	6,033	21,504	8,364
	Avg. CI half-length	5.3104	11.9604	6.7608	12.4407
	Var. CI half-length	1.7314	1.3785	2.7774	1.5171
	$\widehat{\text{MSE}}[\overline{\overline{X}}(m,k)]$	8.9913	121.5678	9.2352	50.1782
	$\widehat{\text{SE}}\{\widehat{\text{MSE}}[\overline{\overline{X}}(m,k)]\}$	0.3878	3.2300	0.4417	1.9333
	# reps. satisfying	1,000	1,000	1,000	1,000
	Satisfied coverage	89%	64.3%	95%	92.9%
±7.5%	# reps.	1,000	1,000	1,000	1,000
	Overall coverage	91%	93.4%	95%	98.1%
	Avg. sample size	9,975	8,047	9,820	9,311
	Max. sample size	24,612	16,334	21,504	21,504
	Avg. CI half-length	5.2512	6.7223	6.0997	7.8967
	Var. CI half-length	1.3046	1.0769	1.0834	2.2671
	$\widehat{\text{MSE}}[\overline{\overline{X}}(m,k)]$	8.1948	12.4996	8.1725	10.1088
	$\widehat{\text{SE}}\{\widehat{\text{MSE}}[\overline{\overline{X}}(m,k)]\}$	0.3857	0.5063	0.3846	0.4368
	# reps. satisfying	1,000	807	1,000	415
	Satisfied coverage	91%	92.7%	95%	96.6%

$R_u = \Phi(Z_u)$ for $u = 1, 2, \ldots$, where $\varphi(z) \equiv e^{-z^2/2}\big/\sqrt{2\pi}$ and $\Phi(z) \equiv \int_{-\infty}^{z} \varphi(w)\,dw$ respectively denote the $N(0,1)$ p.d.f. and c.d.f. Finally, the process $\{R_u : u = 1, 2, \ldots\}$ is fed into the inverse of the Pareto c.d.f. (10) to generate the process $\{X_u\}$ as follows,

$$X_u = F_X^{-1}(R_u) = \xi\big/[1 - \Phi(Z_u)]^{1/\vartheta} \tag{11}$$

for $u = 1, 2, \ldots$. The mean and the variance of the ARTOP process (11) are given by $\mu_X = \text{E}[X_u] = \vartheta\xi(\vartheta - 1)^{-1}$, for $\vartheta > 1$ and $\sigma_X^2 = \xi^2\vartheta(\vartheta - 1)^{-2}(\vartheta - 2)^{-1}$, for $\vartheta > 2$, respectively (Johnson, Kotz, and Balakrishnan 1994).

We decided to set the parameters of the Pareto distribution (10) according to $\vartheta = 2.1$ and $\xi = 1$; and we set the lag-one correlation in the base process (9) to $\rho = 0.995$. This yields an ARTOP process $\{X_u : u = 1, 2, \ldots\}$ whose marginal distribution has mean, variance, skewness, and kurtosis respectively given by $\mu_X = 1.9091$, $\sigma_X^2 = 17.3554$, $\text{E}\big\{[(X_u - \mu_X)/\sigma_X]^3\big\} = \infty$ and $\text{E}\big\{[(X_u - \mu_X)/\sigma_X]^4\big\} = \infty$ The most difficult aspect of this system is that the marginals are highly nonnormal, and their distribution has a very heavy tail. We sampled Z_0 from the $N(0,1)$ distribution when generating the process $\{X_u\}$ so that the process was started in steady-state operation. Therefore, there is no warm-up problem for this process.

Table 8 shows the performance of WASSP for the ARTOP process (11) described above using the smoothing parameter values $A = 5$, 7, and 9. The results are based on 400 independent replications of nominal 90% CIs. From this table, we see that the coverage decreases in general as the smoothing parameter increases. For nominal 90% CIs with $A = 7$ and $A = 9$, the resulting coverage probabilities are unacceptable at all three precision levels. In this application of WASSP, the smoothing parameter value $A = 5$ appears to yield the best results for both nominal 90% and 95% confidence intervals, especially in the ±7.5% precision case. While it is true that for $A = 5$ there is significant undercoverage in the small-sample cases, clearly as the sample size increases the coverage probabilities approach the nominal level. It is unclear at this point why $A = 5$ produces the best results for this process. Nonetheless, it is recommended that the default smoothing parameter be changed from $A = 7$ to $A = 5$ for this ARTOP process.

Table 8: Performance of WASSP Using Different Values of A for the ARTOP Process (11) with $\xi = 1.0$, $\vartheta = 2.1$, $\rho = 0.995$ and $Z_0 \sim N(0,1)$; Results Are Based on 400 Independent Replications of Nominal 90% CIs

Prec. Req.	Performance Measure	Smoothing Parameter $A = 5$	$A = 7$	$A = 9$
None	CI coverage	84.2%	79%	78%
	Avg. sample size	19,880	22,512	22,512
	Avg. CI half-length	0.5183	0.4475	0.4155
	Var. CI half-length	0.0774	0.0544	0.0441
±15%	CI coverage	77.3%	71.5%	72.25%
	Avg. sample size	79,095	66,158	54,551
	Avg. CI half-length	0.2198	0.2231	0.2296
	Var. CI half-length	0.0022	0.0018	0.0018
±7.5%	CI coverage	89%	85.3%	82.5%
	Avg. sample size	430,430	345,870	272,670
	Avg. CI half-length	0.1152	0.1159	0.1177
	Var. CI half-length	0.0005	0.0005	0.0005

4.1 Summary of Experimental Results for the ARTOP Process

Table 9 shows a comparison of the performance of WASSP (using $A = 5$) and ASAP3 for the ARTOP process (11). For the no precision case, WASSP and ASAP3 yield similar results in terms of confidence interval coverage. However, ASAP3 is requiring significantly larger sample sizes than WASSP is for the no precision case. For nominal 90% CIs in the ±15% precision case, ASAP3 clearly outperforms WASSP; and the two methods produce comparable results for nominal 95% CIs with ±15% precision. Asymptotically, WASSP appears to outperform ASAP3 in the ARTOP process, however. For nominal 90% and 95% confidence intervals with ±7.5% precision, the coverage probability for WASSP is right at the nominal level, while the coverage probability for ASAP3 is significantly below the nominal level. In fact, the coverage probabilities for ASAP3 are about the same for both nominal 90% and nominal 95% confidence intervals at all three levels of precision.

Table 9: Performance of WASSP (Using $A = 5$) and ASAP3 for the ARTOP Process (11) with $\xi = 1.0$, $\vartheta = 2.1$, $\rho = 0.995$, and $Z_0 \sim N(0, 1)$; Results Are Based on 400 Independent Replications of Nominal 90% and 95% CIs

Prec. Req.	Performance Measure	90% CIs WASSP	90% CIs ASAP3	95% CIs WASSP	95% CIs ASAP3
None	# reps.	400	400	400	400
	CI coverage	84.2%	85.5%	89%	90.75%
	Avg. sample size	19,880	114,053	17,028	114,053
	Max. sample size	955,008	524,288	454,410	524,288
	Avg. CI half-length	0.5183	0.0909	0.7042	0.1089
	Var. CI half-length	0.0774	0.0019	0.1802	0.0029
±15%	# reps.	400	400	400	400
	CI coverage	77.3%	85.5%	87%	90.75%
	Avg. sample size	79,095	117,092	151,190	120,660
	Max. sample size	1,674,368	722,944	3,059,792	1,028,096
	Avg. CI half-length	0.2198	0.0867	0.2232	0.1008
	Var. CI half-length	0.0020	0.0006	0.0021	0.0006
±7.5%	# reps.	400	400	400	400
	CI coverage	89%	84%	95%	90.25%
	Avg. sample size	430,430	186,517	747,640	255,512
	Max. sample size	8,466,630	2,873,344	6,975,686	2,097,152
	Avg. CI half-length	0.1152	0.0676	0.1190	0.0696
	Var. CI half-length	0.0005	0.00005	0.0005	0.00003

Table 10 shows the results of comparing the performance of WASSP (using $A = 5$) with the H&W sequential spectral method. We first notice from this table that the values of $\widehat{\mathrm{MSE}}\big[\,\overline{\overline{X}}(m,k)\,\big]$ for WASSP and the H&W method are comparable. Since this ARTOP process is started in steady-state operation, $\widehat{\mathrm{MSE}}\big[\,\overline{\overline{X}}(m,k)\,\big]$ for the H&W method will not be inflated by system warm-up bias, as we have seen in some of the other test processes. We also notice from this table that WASSP-generated CIs satisfying the precision requirement have much better coverage probabilities than the H&W-generated CIs. Even though WASSP produces significant undercoverage in the small-sample case (especially for nominal 90% CIs), asymptotically it produces coverages that are at the nominal level. Even at the ±7.5% precision level, the coverage for the H&W method is significantly below the nominal level. Overall we concluded that WASSP outperformed the H&W method in the ARTOP process.

Table 10: Performance of WASSP (using $A = 5$) and the H&W Spectral Method for the ARTOP Process (11) with $\xi = 1.0$, $\vartheta = 2.1$, $\rho = 0.995$, and $Z_0 \sim N(0, 1)$; Results Are Based on 400 Independent Replications of Nominal 90% and 95% CIs

Prec. Req.	Performance Measure	90% CIs WASSP	90% CIs H&W	95% CIs WASSP	95% CIs H&W
None	# reps.	400	400	400	400
	Overall coverage	84.2%	67%	89%	75.5%
	Avg. sample size	19,880	2,982	17,028	2,555
	Max. sample size	955,008	143,252	454,410	10,741
	Avg. CI half-length	0.5183	0.7124	0.7042	0.9391
	Var. CI half-length	0.0774	0.6841	0.1802	1.8882
	$\widehat{\mathrm{MSE}}[\overline{\overline{X}}(m,k)]$	0.0752	0.3239	0.0919	0.5049
	$\widehat{\mathrm{SE}}\{\widehat{\mathrm{MSE}}[\overline{\overline{X}}(m,k)]\}$	0.0039	0.0266	0.0097	0.1118
	# reps. satisfying	400	400	400	400
	Satisfied coverage	84.2%	67%	89%	75.5%
±15%	# reps.	400	400	400	400
	Overall coverage	77.3%	72.8%	87%	84%
	Avg. sample size	79.095	39,781	151,190	72,093
	Max. sample size	1,674,368	673,442	3,059,792	688,454
	Avg. CI half-length	0.2198	0.2341	0.2232	0.2406
	Var. CI half-length	0.0022	0.0024	0.0021	0.0187
	$\widehat{\mathrm{MSE}}[\overline{\overline{X}}(m,k)]$	0.0365	0.0495	0.0259	0.0350
	$\widehat{\mathrm{SE}}\{\widehat{\mathrm{MSE}}[\overline{\overline{X}}(m,k)]\}$	0.0019	0.0026	0.0030	0.0036
	# reps. satisfying	400	386	400	394
	Satisfied coverage	77.3%	71.9%	87%	84%
±7.5%	# reps.	400	400	400	400
	Overall coverage	89%	81.8%	95%	90%
	Avg. sample size	430,430	208,570	747,640	348,470
	Max. sample size	8,466,630	1,311,863	6,975,686	2,354,295
	Avg. CI half-length	0.1152	0.1194	0.1190	0.1200
	Var. CI half-length	0.0005	0.0003	0.0005	0.0003
	$\widehat{\mathrm{MSE}}[\overline{\overline{X}}(m,k)]$	0.0070	0.0099	0.0028	0.0054
	$\widehat{\mathrm{SE}}\{\widehat{\mathrm{MSE}}[\overline{\overline{X}}(m,k)]\}$	0.0015	0.0013	0.0008	0.0006
	# reps. satisfying	400	395	400	334
	Satisfied coverage	89%	82%	95%	88.1%

5 CONCLUSIONS AND RECOMMENDATIONS

We presented three test processes that were specifically designed to explore the robustness of WASSP and its competitors against the statistical anomalies commonly encountered in the analysis of outputs generated from large-scale, steady-state simulation experiments. From the experimental results presented, it is evident that WASSP outperforms the H&W method; and we believe WASSP represents an advance in spectral methods for simulation output analysis. Furthermore, we can conclude that while WASSP and ASAP3 produce comparable results in some cases, WASSP is in general a more robust procedure than ASAP3.

It would be desirable in the future to continue the experimental work that has been designed to explore the robustness of WASSP. This extended performance evaluation should include applying WASSP to processes with long-range dependence (Suárez-González et al. 2002). It would also be informative to apply WASSP and its competitors to more realistic applications, such as the $M/M/1$ queue waiting time process with 70% traffic intensity, or the white noise process (that is, a random $N(0, 1)$ process).

An essential complement to our experimental performance evaluation of WASSP must be a theoretical characterization of key asymptotic properties of the procedure as the relative or absolute precision requirement tends to zero. This is the subject of ongoing work.

REFERENCES

Cochran, W. G. 1977. *Sampling techniques*. 3rd ed. New York: Wiley.

Heidelberger, P., and P. D. Welch. 1981. A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM* 24 (4): 233–245.

Heidelberger, P., and P. D. Welch. 1983. Simulation run length control in the presence of an initial transient. *Operations Research* 31 (6): 1109–1144.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. *Continuous univariate distributions*. 2nd ed. Vol. 1. New York: Wiley.

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> [accessed June 18, 2004].

Lada, E. K., and J. R. Wilson. 2004. A wavelet-based spectral procedure for steady-state simulation analysis. *European Journal of Operational Research* in review. Available online via <ftp.ncsu.edu/pub/eos/pub/jwilson/wassp-ejorv30.pdf> [accessed April 11, 2004].

Lada, E. K., J. R. Wilson, N. M. Steiger, and J. A. Joines. 2004. Performance of a wavelet-based spectral procedure for steady-state simulation analysis. *INFORMS Journal on Computing* in review. Available online via <ftp.ncsu.edu/pub/eos/pub/jwilson/ljswtr12.pdf> [accessed June 18, 2004].

Ross, S. M. 1983. *Stochastic processes*. New York: Wiley.

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, J. A. Joines, C. Alexopoulos, and D. Goldsman. 2004. ASAP3: A batch means procedure for steady-state simulation analysis. *ACM Transactions on Modeling and Computer Simulation* to appear. Available online via <ftp.eos.ncsu.edu/pub/jwilson/2003-34steiger.pdf> [accessed July 31, 2004].

Suárez-González, A., J. C. López-Ardao, C. López-García, M. Rodríguez-Pérez, M. Fernández-Veiga, and M. E. Sousa-Vieira. 2002. A batch means procedure for mean value estimation of processes exhibiting long range dependence. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 456–464. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via <www.informs-cs.org/wsc02papers/057.pdf> [accessed June 18, 2004].

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 postdoctoral fellow at the Statistical and Applied Mathematical Sciences Institute (SAMSI). She is a member of IIE and INFORMS. Her e-mail address is <eklada@eos.ncsu.edu>.

JAMES R. WILSON is professor and head of the Department of Industrial Engineering at North Carolina State University. He is a member of AAUW, ACM, ASA, and INFORMS; and he is a Fellow of IIE. His e-mail address is <jwilson@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>.

JEFFREY A. JOINES is an assistant professor of textile engineering at NC State University. His e-mail address is <JeffJoines@ncsu.edu>.