Area variance estimators for simulation using folded standardized time series

CLAUDIA ANTONINI
Departamento de Matemáticas Puras y Aplicadas,
Universidad Simón Bolívar,
Sartenejas, 1080, Venezuela

CHRISTOS ALEXOPOULOS
H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332, USA

DAVID GOLDSMAN
H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332, USA

JAMES R. WILSON
Edward P. Fitts Department of Industrial and Systems Engineering,
North Carolina State University,
Campus Box 7906,
Raleigh, NC 27695-7906, USA

1E-mail: <cfmantonini@usb.ve>. Telephone: +58-212-906-3294.

2E-mail: <christos@isye.gatech.edu>. Telephone: (404) 894-2361. Fax: (404) 894-2301.

3Member, Institute of Industrial Engineers, and corresponding author. E-mail: <sman@gatech.edu>. Telephone: (404) 894-2365. Fax: (404) 894-2301.

4Fellow, Institute of Industrial Engineers. E-mail: <jwilson@ncsu.edu>. Telephone: (919) 515-6415. Fax: (919) 515-5281.
Area variance estimators for simulation using folded standardized time series

We estimate the variance parameter of a stationary simulation-generated process using “folded” versions of standardized time series area estimators. Asymptotically as the sample size increases, different folding levels yield unbiased estimators that are independent scaled chi-squared variates, each with one degree of freedom. We exploit this result to formulate improved variance estimators based on the combination of multiple levels as well as the use of batching. The improved estimators preserve the asymptotic bias properties of their predecessors, but have substantially lower asymptotic variances. The performance of the new variance estimators is demonstrated in a first-order autoregressive process with autoregressive parameter 0.9 and in the queue-waiting-time process for an $M/M/1$ queue with server utilization 0.8.

1. Introduction

One of the most important problems in simulation output analysis is the estimation of the mean $\mu$ of a steady-state (stationary) simulation-generated process $\{Y_i : i = 1, 2, \ldots\}$. For instance, we may be interested in determining the steady-state mean transit time in a job shop or the long-run expected profit per period arising from a certain inventory policy. If the simulation is indeed operating in steady state, then the estimation of $\mu$ is not itself a particularly difficult problem—simply use the sample mean of the simulation outputs, $\bar{Y}_n \equiv \frac{1}{n} \sum_{i=1}^{n} Y_i$, as the point estimator of $\mu$, where $n$ is the sample size.

Because any serious statistical analysis should also include a measure of the variability of the sample mean, point estimation of the mean is usually not enough. One of the most commonly used measures of this variability is the variance parameter $\sigma^2$, which is defined as the sum of covariances of the process $\{Y_i : i = 1, 2, \ldots\}$ at all lags, and which can often be written in the intuitively appealing form $\sigma^2 = \lim_{n \to \infty} n \text{Var}(\bar{Y}_n)$. With knowledge of such a measure in hand, we could provide, among other benefits, confidence intervals for $\mu$—typically of the form

$$\bar{Y}_n \pm t\sqrt{\hat{\sigma}^2/n},$$

where $t$ is a quantile from the appropriate pivot distribution and $\hat{\sigma}^2$ is an estimator for $\sigma^2$. 
Unfortunately, the problem of estimating the variance of the sample mean is not so straightforward. The trouble is caused by discrete-event simulation data almost always being serially correlated, nonnormal, and contaminated by initialization bias, e.g., consecutive waiting times in a queueing system starting in an empty and idle state. These characteristics invalidate traditional statistical analysis methods, which usually rely on the assumption of independent and identically distributed (i.i.d.) normal observations. With these problems in mind, this article is concerned with providing underlying theory for estimating the variance parameter $\sigma^2$ of a stationary simulation-generated process—that is, the output from a simulation in steady-state operation so that any effects arising from initialization bias have been largely eliminated, for example, by using an appropriate procedure to identify and delete the warm-up period (Lada et al., 2006).

Over the years, a number of methodologies for estimating $\sigma^2$ have been proposed in the literature (see Law, 2007), e.g., the techniques referred to as independent replications (IR), nonoverlapping batch means (NBM), overlapping batch means (OBM), standardized time series (STS), spectral analysis, and regenerative analysis. IR works quite well when the output process starts in steady state, but it can have severe drawbacks in the presence of an initial transient phase (see Alexopoulos and Goldsman, 2004), which is usually the situation in large-scale practical applications. NBM—conceptually the simplest of these methodologies—divides the data $\{Y_i : i = 1, 2, \ldots, n\}$ into nonoverlapping batches, and uses the sample variance of the sample means derived from the batches (i.e., the batch means) as a foundation to estimate $\sigma^2$. OBM (Meketon and Schmeiser, 1984), on the other hand, effectively reuses the data by forming overlapping batches, and then invokes an appropriately scaled sample variance of the resulting sample means derived from the batches to estimate $\sigma^2$. The result is an OBM variance estimator having about the same bias as, but significantly lower variance than, the benchmark NBM estimator employing the same batch size and total sample size. STS (Schruben, 1983) uses a functional central limit theorem to standardize a stationary time series, such as output from a steady-state discrete-event simulation, into a process that converges in distribution to a limiting Brownian bridge process as the batch size or total sample size becomes large. Known properties of the Brownian bridge process are then used to obtain estimators for $\sigma^2$, e.g., the basic area and Cramér–von Mises (CvM) STS estimators.

Alexopoulos et al. (2007b,c) show that similar to OBM, overlapping batched versions of the area and CvM STS estimators have the same bias as, but substantially lower variance than, their nonoverlapping counterparts. Aktaran-Kalayci et al. (2007) form minimum-variance linear combinations of overlapping variance estimators of the same type (for example, an area estimator with a certain weighting function) but with different batch sizes.
for different constituent estimators. Additional variance-reducing tricks in which STS reuses data are given by Goldsman et al. (2007), who form linear combinations of two different types of variance estimators (specifically, the area and CvM estimators), and Foley and Goldsman (1999), who orthonormalize various area estimators.

A recurring theme that emerges in the development of new estimators for $\sigma^2$ is the effective reuse of data. In the current article, we study the consequences of a new “folding” operation on the original STS process (and its limiting Brownian bridge process). The folding operation produces multiple standardized time series processes, which in turn will ultimately allow us to use the original data to produce multiple area estimators for $\sigma^2$ that use a common batch size—estimators that are often asymptotically independent as the sample size grows. These folded estimators will lead to combined estimators having smaller variance than existing estimators not based on the folding operation.

The article is organized as follows. Section 2 gives some background material on STS. In Section 3, we introduce the notion of folding a Brownian bridge, and we show that each application of folding yields a new Brownian bridge process. We also derive useful expressions for these folded processes in terms of the original Brownian bridge and in terms of the original underlying Brownian motion. Section 4 is concerned with derivations of the expected values, variances, and covariances of certain functionals related to the area under a folded Brownian bridge. In Section 5, we finally show how to apply these results to the problem of estimating the variance parameter of a steady-state simulation output process. The idea is to start with a single STS; form folded versions of that STS (which converge to the corresponding folded versions of a Brownian bridge process); calculate an estimator for $\sigma^2$ from each folded STS; and then combine the estimators into one low-variance estimator. We also show how to use batching to further improve the estimators; and we illustrate the performance of the resulting variance estimators when they are applied to a first-order autoregressive process with autoregressive parameter 0.9 and to the queue-waiting-time process for an $M/M/1$ queue with server utilization 0.8. We find that the new estimators indeed reduce estimator variance at little cost in bias. Section 6 presents conclusions and suggestions for future research, while the technical details of some of the proofs are relegated to the Appendix. Antonini et al. (2007) give an abridged version of some of the theoretical and experimental results that are fully detailed in this article.

2. Background

This section lays out preliminaries on the STS methodology. We begin with some standard assumptions that we shall invoke whenever needed in the sequel. These assumptions
ensure that the proposed variance estimators work properly on a wide variety of stationary stochastic processes.

**Assumptions A**

1. The process \{Y_i : i = 1, 2, \ldots\} is stationary and satisfies the following Functional Central Limit Theorem. For \( n = 1, 2, \ldots \), the process

   \[
   X_n(t) \equiv \frac{|nt| (\bar{Y}_{|nt|} - \mu)}{\sigma \sqrt{n}} \quad \text{for} \quad t \in [0, 1]
   \]  

   satisfies \( X_n \xrightarrow{n \to \infty} W \), where: \( \mu \) is the steady-state mean; \( \sigma^2 \) is the variance parameter; \( \lfloor \cdot \rfloor \) denotes the greatest integer function; \( W \) is a standard Brownian motion process on \([0, 1]\); and \( \xrightarrow{n \to \infty} \) denotes convergence in distribution on \( D[0, 1] \), the space of functions on \([0, 1]\) that are right-continuous with left-hand limits, as \( n \to \infty \). (See also Billingsley, 1968 and Glynn and Iglehart, 1990.)

2. The series of covariances at all lags \( \sum_{i=-\infty}^{\infty} R_i = \sigma^2 \in (0, \infty) \), where \( R_i \equiv \text{Cov}(Y_1, Y_{1+i}) \), \( i = 0, \pm 1, \pm 2, \ldots \).

3. The series \( \sum_{i=1}^{\infty} i^2 |R_i| < \infty \).

4. The function \( f(\cdot) \), defined on \([0, 1]\), is twice continuously differentiable. Further, \( f(t) \) satisfies the normalizing condition \( \int_0^1 \int_0^1 f(s)f(t)|\min\{s,t\} - st| \, ds \, dt = 1 \).

As elaborated in Glynn and Iglehart (1990) and in Remark 1 of Aktaran-Kalaycı et al. (2007), Assumptions A.1–A.3 are mild conditions that hold for a variety of processes encountered in practice. On the other hand, these assumptions appear to be violated in computer and telecommunications systems, where some target processes have distributions and autocorrelation functions with heavy tails (Crovella and Lipsky, 1997). Assumption A.4 gives conditions on the normalized weight function \( f(\cdot) \) that will be used in our estimators for \( \sigma^2 \).

Of fundamental importance to the rest of the paper is the standardized time series of the underlying stochastic process \( \{Y_i : i = 1, 2, \ldots\} \). It is the STS that will form the basis of most of the estimators studied herein.

**Definition 1.** As in Schruben (1983), for \( n = 1, 2, \ldots \), the (level-0) standardized time series based on a sample of size \( n \) from the process \( \{Y_i : i = 1, 2, \ldots\} \) is

\[
T_{0,n}(t) \equiv \frac{|nt| (\bar{Y}_n - \bar{Y}_{|nt|})}{\sigma \sqrt{n}} \quad \text{for} \quad t \in [0, 1].
\]
In the next section, we discuss how the STS defined by Equation (2) is related to a Brownian bridge process; and in Section 5.2, we show how to use this process to derive estimators for $\sigma^2$.

3. Folded Brownian bridges

Our development requires some additional nomenclature. First of all, we define a Brownian bridge, which will turn out to characterize the asymptotic probability law governing a standardized time series, much as the standard normal distribution characterizes the asymptotic probability law governing the properly standardized sample mean of i.i.d. data.

**Definition 2.** If $W(\cdot)$ is a standard Brownian motion process, then the associated level-0 Brownian bridge process is

$$B_0(t) \equiv B(t) \equiv tW(1) - W(t) \text{ for } t \in [0, 1]. \ \triangleright$$

In fact, a Brownian bridge $\{B(t) : t \in [0, 1]\}$ is a Gaussian process with $E[B(t)] = 0$ and $\text{Cov}(B(s), B(t)) = \min\{s, t\} - st$, for $s, t \in [0, 1]$. Brownian bridges are important for our purposes because under Assumptions A.1–A.3, Schruben (1983) shows that $T_{0,n}(\cdot) \Rightarrow B(\cdot)$ and that $\sqrt{n}(\bar{Y}_n - \mu)$ and $T_{0,n}(\cdot)$ are asymptotically independent as $n \to \infty$.

The contribution of the current paper is the development and evaluation of folded estimators for $\sigma^2$. We now define precisely what we mean by the folding operation, a map that can be applied either to an STS or a Brownian bridge.

**Definition 3.** The folding map $\Psi : Y \in D[0, 1] \to \Psi_Y \in D[0, 1]$ is defined by

$$\Psi_Y(t) \equiv Y\left(\frac{t}{2}\right) - Y\left(1 - \frac{t}{2}\right) \text{ for } t \in [0, 1]. \ \triangleright$$

Moreover for each nonnegative integer $k$, we define $\Psi^k : Y \in D[0, 1] \to \Psi_Y^k \in D[0, 1]$, the $k$-fold composition of the folding map $\Psi$, so that for every $t \in [0, 1]$,

$$\Psi_Y^k(t) \equiv \begin{cases} Y(t), & \text{if } k = 0, \\ \Psi \circ \Psi_Y^{k-1}(t), & \text{if } k = 1, 2, \ldots \end{cases} \ \triangleright$$

The folding operation can be performed multiple times on a Brownian bridge process, as demonstrated by the following definition. The motivation comes from Exercise 5 on p. 32 of Shorack and Wellner (1986), who only consider single-level ($k = 1$) folding.
**Definition 4.** For $k = 1, 2, \ldots$, the level-$k$ folded version of a Brownian bridge process $B_0(\cdot)$ is defined recursively by

$$B_k(t) \equiv \Psi_{B_{k-1}}(t) = B_{k-1}(\frac{1}{2}) - B_{k-1}(1 - \frac{t}{2}),$$

so that $B_k(t) = \Psi_{B_0}^k(t)$ for $t \in [0, 1]$. ▷

Intuitively speaking, when the folding operator $\Psi$ is applied to a Brownian bridge process $\{B_0(t) : t \in [0, 1]\}$, it does the following: (i) $\Psi$ reflects (folds) the portion of the original process defined on the subinterval $[\frac{1}{2}, 1]$ (shown in the upper right-hand portion of Fig. 1(a)) about the vertical line $t = \frac{1}{2}$ (yielding the subprocess shown in the upper left-hand portion of Fig. 1(a)); and (ii) $\Psi$ takes the difference between these two subprocesses defined on $[0, \frac{1}{2}]$ and stretches that difference over the unit interval $[0, 1]$ (yielding the new process shown in Fig. 1(b)).

![Diagram](image)

(a) Reflecting about $t = 1/2$.  
(b) Differencing and stretching.

**Fig. 1.** Geometric illustration of folded Brownian bridges.

Lemma 1 shows that as long as we start with a Brownian bridge, repeated folding will produce Brownian bridges as well. The proof simply requires that we verify the necessary covariance structure (see Antonini, 2005).

**Lemma 1.** For $k = 1, 2, \ldots$, the process $\{B_k(t) : t \in [0, 1]\}$ is a Brownian bridge.

The next lemma gives an equation relating the level-$k$ Brownian bridge with the original (level-0) Brownian bridge and the initial Brownian motion process. These results will be useful later on when we derive properties of certain functionals of $B_k(\cdot)$. 

**Refer to the original text for further details.**
Lemma 2. For \( k = 1, 2, \ldots \) and \( t \in [0, 1] \),

\[
\mathcal{B}_k(t) = \sum_{i=1}^{2^{k-1}} \left[ \mathcal{B}(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}) - \mathcal{B}(\frac{i-1}{2^{k-1}} - \frac{t}{2^k}) \right]
\]

(3)

\[
= \sum_{i=1}^{2^{k-1}} \left[ \mathcal{W}(\frac{i}{2^{k-1}} - \frac{t}{2^k}) - \mathcal{W}(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}) \right] - (1-t)\mathcal{W}(1).
\]

(4)

It is easy to show that Equation (3) is a direct consequence of Definition 4; and then Equation (4) follows from Equation (3) and Definition 2. See Antonini (2005) for the details.

4. Some functionals of folded Brownian bridges

The purpose of this section is to highlight results on the weighted areas under successively higher levels of folded Brownian bridges. Such functionals will be used in Section 5 to construct estimators for the variance parameter \( \sigma^2 \) arising from a stationary stochastic process.

Definition 5. For \( k = 0, 1, \ldots \), the weighted area under the level-\( k \) folded Brownian bridge is

\[
N_k(f) \equiv \int_0^1 f(t)\mathcal{B}_k(t) \, dt.
\]

Under simple conditions, Theorem 4.1 shows that \( N_k(f) \) has a standard normal distribution; its proof is in the Appendix.

Theorem 4.1. For any weight function \( f(\cdot) \) satisfying Assumption A.4 and for \( k = 0, 1, \ldots \), we have \( N_k(f) \sim \text{Nor}(0, 1) \).

Corollary 1. Under the conditions of Theorem 4.1, we have \( A_k(f) \equiv \sigma^2 N_k(f) \sim \sigma^2 \chi_1^2 \).

Of course, the corollary is an immediate consequence of Theorem 4.1. Besides the distributional result, it follows that \( \mathbb{E}[A_k(f)] = \sigma^2 \), a finding that we will revisit in Theorem 5.3 when we develop estimators for \( \sigma^2 \).

Meanwhile, we proceed with several results concerning the joint distribution of the \( \{N_k(f) : k = 0, 1, \ldots \} \). Our first such result, the proof of which is in the Appendix, gives an explicit expression for the covariance between folded area functionals from different levels. Before stating the theorem, for any weight function \( f(\cdot) \), we define

\[
F(t) \equiv \int_0^t f(s) \, ds \text{ for } t \in [0, 1], \quad F \equiv F(1),
\]

\[
\bar{F}(t) \equiv \int_0^t F(s) \, ds \text{ for } t \in [0, 1], \quad \bar{F} \equiv \bar{F}(1).
\]
Theorem 4.2. If \( f_1(t) \) and \( f_2(t) \) are weight functions satisfying Assumption A.4, then for \( \ell = 0, 1, \ldots \) and \( k = 1, 2, \ldots \), we have

\[
\text{Cov}[N_\ell(f_1), N_{\ell+k}(f_2)] = -\bar{F}_1 \bar{F}_2 + 2 \sum_{i=1}^{k} \int_0^1 f_2(t) \left[ \bar{F}_1 \left( \frac{i}{2^k-1} - \frac{t}{2^k} \right) - \bar{F}_1 \left( \frac{i-1}{2^k-1} + \frac{t}{2^k} \right) \right] dt.
\] (5)

Lemmas 3–6 give results on the covariance between functionals of Brownian motion from different levels; these will be used later on to establish asymptotic covariances of estimators for \( \sigma^2 \) from different levels. In particular, Lemmas 5 and 6 give simple conditions under which these functionals are uncorrelated.

Lemma 3. For \( \ell, k = 0, 1, \ldots \) and \( s, t \in [0, 1] \), \( \text{Cov}[B_\ell(s), B_{\ell+k}(t)] = \text{Cov}[B_0(s), B_k(t)] \).

Proof. Follows by induction on \( k \). \( \square \)

Lemma 4. For \( \ell, k = 0, 1, \ldots \), \( \text{Cov}[N_\ell(f_1), N_{\ell+k}(f_2)] = \text{Cov}[N_0(f_1), N_k(f_2)] \).

Proof. Follows immediately from Definition 5 and Lemma 3. \( \square \)

Lemma 5. If the weight function \( f(\cdot) \) satisfies the symmetry condition \( f(t) = f(1-t) \) for all \( t \in [0, 1] \) as well as Assumption A.4, then \( \text{Cov}(N_0(f), N_k(f)) = 0 \) for all \( k = 1, 2, \ldots \).

Proof. Applying integration by parts to Equation (5) with \( f_1 = f_2 = f \) and reversing the order of terms in the first resulting summation, we obtain

\[
\text{Cov}[N_0(f), N_k(f)] = \frac{1}{2^k} \sum_{i=1}^{2^k-1} \int_0^1 F(t) \left[ F \left( \frac{i}{2^k-1} - \frac{t}{2^k} \right) + F \left( \frac{i-1}{2^k-1} + \frac{t}{2^k} \right) \right] dt - \bar{F}^2
\]

\[
= \frac{1}{2^k} \int_0^1 F(t) \left[ \sum_{i=1}^{2^k-1} \left[ F \left( 1 - \left( \frac{i-1}{2^k-1} + \frac{t}{2^k} \right) \right) + F \left( \frac{i-1}{2^k-1} + \frac{t}{2^k} \right) \right] \right] dt - \bar{F}^2
\]

\[
= \frac{F \bar{F}}{2} - \bar{F}^2,
\]

which follows since

\[
F(1-x) = \int_x^1 f(1-z) \, dz = \int_x^1 f(z) \, dz,
\]
so that $F(1-x) + F(x) = F$ for all $x \in [0,1]$. The proof is completed by noting that

$$
\bar{F} = \int_0^1 f(x)(1-x) \, dx \\
= \int_0^{1/2} f(x)(1-x) \, dx + \int_0^{1/2} f(y) y \, dy \\
= \int_0^{1/2} f(x) \, dx = \frac{F}{2}. \quad \square
$$

**Lemma 6.** If the weight function $f(\cdot)$ satisfies the symmetry condition $f(t) = f(1-t)$ for all $t \in [0,1]$ as well as Assumption A.4, then for $\ell = 0,1,\ldots$ and $k = 1,2,\ldots$, $\text{Cov}[N_\ell(f), N_{\ell+k}(f)] = \text{Cov}[N_0(f), N_k(f)] = 0$.

**Proof.** Immediate from Lemmas 4 and 5. \quad \square

The following lemma, proven in the Appendix, establishes the multivariate normality of the random vector $N(f) \equiv [N_0(f), \ldots, N_k(f)]$. It will be used in Theorem 4.3 to obtain the remarkable result that, under relatively simple conditions, the folded functionals $\{N_k(f) : k = 0,1,\ldots\}$ are i.i.d. $\text{Nor}(0,1)$.

**Lemma 7.** If the weight function $f(\cdot)$ satisfies the symmetry condition $f(t) = f(1-t)$ for all $t \in [0,1]$ as well as Assumption A.4, then for each positive integer $k$ the random vector $N(f)$ has a nonsingular multivariate normal distribution.

**Theorem 4.3.** If the weight function $f(\cdot)$ satisfies the symmetry condition $f(t) = f(1-t)$ for all $t \in [0,1]$ as well as Assumption A.4, then the random variables $\{N_k(f) : k = 0,1,\ldots\}$ are i.i.d. $\text{Nor}(0,1)$ random variables.

**Proof.** Immediate from Lemmas 6 and 7. \quad \square

**Corollary 2.** Under the conditions of Theorem 4.3, the random variables $\{A_k(f) : k = 0,1,\ldots\}$ are i.i.d. $\sigma^2 \chi_1^2$.

**Remark 1.** To the best of our knowledge, the symmetry condition $f(t) = f(1-t)$ for $t \in [0,1]$ has not been used previously to derive independent variance estimators based on the STS method. \quad ◄

**Example 1.** The following weight functions satisfy both Assumption A.4 and the symmetry condition $f(t) = f(1-t)$ for $t \in [0,1]$; and these functions naturally arise in simulation output analysis applications (see Foley and Goldsman, 1999 and Section 5 herein): $f_0(t) \equiv \sqrt{12}$, $f_2(t) \equiv \sqrt{840}(3t^2 - 3t + 1/2)$, and $f_{\cos,j}(t) \equiv \sqrt{8} \pi j \cos(2\pi j t)$, $j = 1,2,\ldots$, all for $t \in [0,1]$. By Corollary 2, $\{A_k(f), k \geq 0\}$ are i.i.d. $\sigma^2 \chi_1^2$ for $f = f_0$, $f_2$, or $f_{\cos,j}$, $j = 1,2,\ldots$. \quad ◄
5. Application to variance estimation

We finally show how our work on properties of area functionals of folded Brownian bridges can be used in simulation output analysis. With this application in mind, we apply the folding transformation to Schruben’s level-0 STS (Schruben, 1983) in Section 5.1, thereby obtaining several new versions of the STS. These new series are used in Section 5.2 to produce new estimators for $\sigma^2$. Section 5.3 gives obvious methods to improve the estimators, Section 5.4 summarizes certain asymptotic properties, and Section 5.5 presents two Monte Carlo examples showing that the estimators work as intended.

5.1. Folded standardized time series

Analogous to the level-$k$ folded Brownian bridge from Definition 4, we define the level-$k$ folded STS.

**Definition 6.** For $k=1,2,\ldots$, the level-$k$ folded STS is

$$T_{k,n}(t) \equiv \Psi_{T_{k-1,n}}(t) = T_{k-1,n}(\frac{t}{2}) - T_{k-1,n}(1 - \frac{t}{2})$$

so that $T_{k,n}(t) = \Psi_{T_{0,n}}(t)$ for $t \in [0,1]$. 

The next goal is to examine the convergence of the level-$k$ folded STS to the analogous level-$k$ folded Brownian bridge process. The following result is an immediate consequence of the almost-sure continuity of $\Psi^k$ on $D[0,1]$ for $k=0,1,\ldots$, and the Continuous Mapping Theorem (CMT) (Billingsley, 1968).

**Theorem 5.1.** If Assumptions A.1–A.3 hold, then for $k=0,1,\ldots$, we have

$$T_{k,n}(\cdot) \equiv [T_{0,n}(\cdot), \ldots, T_{k,n}(\cdot)] \Rightarrow B_k(\cdot) \equiv [B_0(\cdot), \ldots, B_k(\cdot)],$$

where $\{B_k(t): t \in [0,1]\}$ is a multivariate Brownian bridge process whose component univariate Brownian bridge processes have the following cross-covariances: for $s,t \in [0,1]$ and $i,j \in \{0,1,2,\ldots,k\}$,

$$\text{Cov}[B_i(s), B_j(t)] = \mathbb{E}[B_0(s), B_{i-j}(t)].$$

Moreover, $\sqrt{n} \left( \bar{Y}_n - \mu \right)$ and $T_{k,n}(\cdot)$ are asymptotically independent as $n \to \infty$.

**Remark 2.** The univariate Brownian bridge processes $B_0(\cdot), \ldots, B_k(\cdot)$ that constitute the components of $B_k(\cdot)$ are not in general independent, but their cross-covariances can be evaluated via Equation (4). Because $B_k(\cdot)$ is a multivariate Gaussian process with mean $\mathbb{E}[B_k(t)] = 0_{k+1}$, the $(k+1)$-dimensional vector of zeros, for all $t \in [0,1]$, Theorem 5.1 completely characterizes the asymptotic probability law governing the behavior of the multivariate STS process $\{T_{k,n}(t): t \in [0,1]\}$ as the simulation run length $n \to \infty$. 

\[\triangleright\]
5.2. Folded area estimators

We introduce folded versions of the STS area estimator for $\sigma^2$, along with their asymptotic distributions, expected values, and variances. To begin, we define the new estimators, along with their limiting Brownian bridge functionals.

**Definition 7.** For $k = 0, 1, \ldots$, the STS level-$k$ folded area estimator for $\sigma^2$ is

$$A_k(f; n) \equiv N_k^2(f; n),$$

where

$$N_k(f; n) \equiv \frac{1}{n} \sum_{j=1}^{n} f\left(\frac{j}{n}\right) \sigma T_{k,n}\left(\frac{j}{n}\right)$$

and $f(\cdot)$ is a weight function satisfying Assumption A.4. The case $k = 0, f = f_0$ corresponds to Schruben’s original area estimator (Schruben, 1983).

**Definition 8.** Let $A(f; n) \equiv [A_0(f; n), \ldots, A_k(f; n)]$ and $A(f) \equiv [A_0(f), \ldots, A_k(f)]$.

The following definitions provide the necessary set-up to establish in Theorem 5.2 below the asymptotic distribution of the random vector $A(f; n)$ as $n \to \infty$.

**Definition 9.** Let $\Lambda$ denote the class of strictly increasing, continuous mappings of $[0, 1]$ onto itself such that for every $\lambda \in \Lambda$, we have $\lambda(0) = 0$ and $\lambda(1) = 1$. If $X,Y \in D[0,1]$, then the Skorohod metric $\rho(X,Y)$ defining the “distance” between $X$ and $Y$ in $D[0,1]$ is the infimum of those positive $\xi$ for which there exists a $\lambda \in \Lambda$ such that $\sup_{t \in [0,1]} |\lambda(t) - t| \leq \xi$ and $\sup_{t \in [0,1]} |X(t) - Y[\lambda(t)]| \leq \xi$. (See Billingsley, 1968 for further details.)

**Definition 10.** For each positive integer $n$, let $\Omega^n : Y \in D[0,1] \to \Omega^n_Y \in D[0,1]$ be the approximate (discrete) STS map

$$\Omega^n_Y(t) \equiv \frac{\lfloor nt \rfloor Y(1)}{n} - Y(t)$$

for $t \in [0,1]$. Moreover, let $\Omega : Y \in D[0,1] \to \Omega_Y \in D[0,1]$ denote the corresponding asymptotic STS map

$$\Omega_Y(t) \equiv \lim_{n \to \infty} \Omega^n_Y(t) = tY(1) - Y(t) \text{ for } t \in [0,1].$$

Note that $\Omega^n$ maps the process (1) into the corresponding standardized time series (2) so that we have $\Omega^n_{X_n}(t) = T_{0,n}(t)$ for $t \in [0,1]$ and $n = 1, 2, \ldots$; moreover, $\Omega$ maps a standard Brownian motion process into a standard Brownian bridge process, $\Omega_W(t) = tW(1) - W(t) \sim \mathcal{B}_0(t)$ for $t \in [0,1]$. 
Definition 11. For a given normalized weight function, for every nonnegative integer $k$, and for every positive integer $n$, the approximate (discrete) level-$k$ folded area map $\Theta_n^k : Y \in D[0,1] \to \Theta_n^k(Y) \in \mathbb{R}$ is defined by

$$\Theta_n^k(Y) \equiv \left[ \frac{1}{n} \sum_{i=1}^{n} f\left(\frac{i}{n}\right) \sigma \psi^k \circ \Omega_n^Y \left(\frac{i}{n}\right) \right]^2.$$ 

Moreover, the corresponding asymptotic folded area map $\Theta^k : Y \in D[0,1] \to \Theta^k(Y) \in \mathbb{R}$ is defined by

$$\Theta^k(Y) \equiv \left[ \int_0^1 f(t) \sigma \psi^k \circ \Omega_Y(t) \, dt \right]^2.$$

In terms of Equation (1), the definition of $A_k(f)$ from Corollary 1, and Definitions 8–11, we see that $\Theta_n^k(X_n) = A_k(f;n)$ and $\Theta^k(W) = A_k(f)$ for every nonnegative integer $k$. We are now ready to proceed with the main convergence theorem, which shows that the folded area estimators converge jointly to their asymptotic counterparts.

**Theorem 5.2.** If Assumptions A hold, then

$$A(f;n) \xrightarrow{n \to \infty} A(f). \quad (6)$$

**Sketch of Proof.** Although the proof of Theorem 5.2 is detailed in the Appendix, it can be summarized as follows. Our goal is to apply the generalized CMT—that is, Theorem 5.5 of Billingsley (1968)—to prove that the $(k+1) \times 1$ random vector with $j$th element $\Theta_n^j(X_n)$ converges in distribution to the $(k+1) \times 1$ random vector with $j$th element $\Theta^j(W)$ for $j = 0, 1, \ldots, k$. To establish the hypotheses of the generalized CMT, we show that if $\{x_n\} \subset D[0,1]$ is any sequence of functions converging to a realization $W$ of a standard Brownian motion process in the Skorohod metric on $D[0,1]$, then the real-valued sequence $\Theta_n^j(x_n)$ converges to $\Theta^j(W)$ almost surely. First we exploit the almost-sure continuity of $W(u)$ at every $u \in [0,1]$ and the convergence of $\{x_n\}$ to $W$ in $D[0,1]$ to show that for every nonnegative integer $j$, with probability one, we have $|\psi^j \circ \Omega_{x_n}^Y(t) - \psi^j \circ \Omega_W^Y(t)| \to 0$ uniformly for $t \in [0,1]$ as $n \to \infty$; and it follows that

$$\lim_{n \to \infty} |\Theta_n^j(x_n) - \Theta_n^j(W)| = 0 \quad \text{with probability one.} \quad (7)$$

Next we exploit the almost-sure convergence $|\psi^j \circ \Omega_{x_n}^W(t) - \psi^j \circ \Omega_W^Y(t)| \to 0$ for all $t \in [0,1]$ as $n \to \infty$, together with the almost-sure continuity and Riemann integrability of $f(t)\psi^j \circ \Omega_W^Y(t)$ for $t \in [0,1]$, to show that

$$\lim_{n \to \infty} |\Theta_n^j(W) - \Theta^j(W)| = 0 \quad \text{with probability one.} \quad (8)$$
Combining (7) and (8) and applying the triangle inequality, we see that the corresponding vector-valued sequence \( \{ [\Theta^0_n(x_n), \ldots, \Theta^k_n(x_n)] : n = 1, 2, \ldots \} \) converges to \( [\Theta^0(W), \ldots, \Theta^k(W)] \) in \( \mathbb{R}^{k+1} \) with probability one; and thus the desired result follows directly from the generalized CMT. \( \square \)

**Remark 3.** Under Assumptions A, and for the weight functions in Example 1, Theorem 5.2 and Corollary 2 imply that \( A_0(f; n), \ldots, A_k(f; n) \) are asymptotically (as \( n \to \infty \)) i.i.d. \( \sigma^2 \chi_1^2 \) random variables. \( \triangleq \)

Under relatively modest conditions, Theorem 5.3 gives asymptotic expressions for the expected values and variances of the level-\( k \) area estimators.

**Theorem 5.3.** Suppose that Assumptions A hold. Further, for fixed \( k \geq 0 \), suppose that the family of random variables \( \{ A^2_k(f; n) : n \geq 1 \} \) is uniformly integrable (see Billingsley, 1968 for a definition and sufficient conditions). Then we have

\[
\begin{align*}
\lim_{n \to \infty} E[A_k(f; n)] &= E[A_k(f)] = \sigma^2, \\
\lim_{n \to \infty} \text{Var}[A_k(f; n)] &= \text{Var}[A_k(f)] = 2\sigma^4.
\end{align*}
\]

**Remark 4.** One can obtain finer-tuned results for \( E[A_0(f; n)] \) and \( E[A_1(f; n)] \). In particular, under Assumptions A, Foley and Goldsman (1999) and Goldsman et al. (1990) show that

\[
E[A_0(f; n)] = \sigma^2 + \frac{[(F - F)^2 + F^2] \gamma}{2n} + o(1/n),
\]

where

\[
\gamma \equiv -2 \sum_{i=1}^{\infty} iR_i
\]

is well defined when Assumption A.3 holds; see also Song and Schmeiser (1995). Antonini (2005) shows that if Assumptions A hold and \( n \) is even, then

\[
E[A_1(f; n)] = \sigma^2 + \frac{F^2 \gamma}{n} + o(1/n).
\]

**5.3. Enhanced estimators**

The individual estimators whose properties are given in Theorem 5.3 are all based on a single long run of \( n \) observations, and all involve a single level of folding so that \( k = 1 \). This section discusses some obvious extensions of the estimators that have improved asymptotic properties—batching and combining levels.

**Batching:** In actual applications, we often organize the data by breaking the \( n \) observations into \( b \) contiguous, nonoverlapping batches, each of size \( m \) so that \( n = bm \); and
then we can compute the folded variance estimators from each batch separately. As the batch size \( m \to \infty \), the variance estimators computed from different batches are asymptotically independent under broadly applicable conditions on the original (unbatched) process \( \{ Y_i : i = 1, 2, \ldots \} \); and thus more stable (i.e., more precise) variance estimators can be obtained by combining the folded variance estimators computed from all available batches.

In view of this motivation, suppose that the \( i \)th nonoverlapping batch of size \( m \) consists of the observations \( Y_{(i-1)m+1}, Y_{(i-1)m+2}, \ldots, Y_{im} \), for \( i = 1, 2, \ldots, b \). Using the obvious minor changes to the appropriate definitions, one can construct the level-\( k \) STS from the \( i \)th batch of observations, say \( T_{k,i,m}(t) \); and from there, one can obtain the resulting level-\( k \) area estimator from the \( i \)th batch, say \( A_{k,i}(f; m) \). Finally, we define the level-\( k \) batched folded area estimator for \( \sigma^2 \) by averaging the estimators from the individual batches,

\[
\tilde{A}_k(f; b, m) \equiv \frac{1}{b} \sum_{i=1}^{b} A_{k,i}(f; m).
\]

Under the conditions of Theorem 5.3, we have

\[
\begin{align*}
\lim_{m \to \infty} \mathbb{E}[\tilde{A}_k(f; b, m)] &= \sigma^2, \\
\lim_{m \to \infty} \text{Var}[\tilde{A}_k(f; b, m)] &= 2\sigma^4/b,
\end{align*}
\]

where the latter result follows from the fact that the \( A_{k,i}(f; m) \)'s, \( i = 1, 2, \ldots, b \), are asymptotically independent as \( m \to \infty \). Thus, we obtain batched estimators with approximately the same expected value \( \sigma^2 \) as a single folded estimator arising from one long run, yet with substantially smaller variance.

Combining levels of folding: Theorem 5.3 shows that, for any weight function \( f(\cdot) \) satisfying Assumption A.4, all of the area estimators from different levels of folding behave the same asymptotically in terms of their expected value and variance. If the weight function also satisfies the symmetry condition \( f(t) = f(1-t) \) for \( t \in [0, 1] \), then we can improve upon these individual estimators by averaging the estimators from different levels. To this end, denote the average of the batched folded area estimators from levels 0, 1, \ldots, \( k \) by

\[
A_k(f; b, m) \equiv \frac{1}{k+1} \sum_{j=0}^{k} \tilde{A}_j(f; b, m).
\]

If the symmetry condition \( f(t) = f(1-t) \) for \( t \in [0, 1] \) holds as well as the assumptions of Theorem 5.3, then we have

\[
\begin{align*}
\lim_{m \to \infty} \mathbb{E}[A_k(f; b, m)] &= \sigma^2 \\
\lim_{m \to \infty} \text{Var}[A_k(f; b, m)] &= 2\sigma^4/[b(k+1)].
\end{align*}
\]

Thus, we obtain estimators with approximately the same expected value \( \sigma^2 \) as that of a single batched folded estimator arising from one level, yet with significantly smaller variance.
5.4. Summary of asymptotic properties

This section summarizes the asymptotic (large \( b \) and \( m \)) values for the bias, variance, and mean squared error (MSE) of the batched folded area estimators we have considered as well as the NBM estimator. To make the discussion self-contained, we first review relevant facts about the NBM estimator,

\[
\mathcal{N}(b, m) \equiv \frac{m}{b - 1} \sum_{i=1}^{b} (\bar{Y}_{i,m} - \bar{Y}_n)^2,
\]

where \( \bar{Y}_{i,m} \equiv \frac{1}{m} \sum_{j=(i-1)m+1}^{im} Y_j \) is the sample average (batch mean) of the \( i \)th nonoverlapping batch, for \( i = 1, \ldots, b \). It turns out that for fixed \( b \),

\[
\mathcal{N}(b, m) \xrightarrow{n \to \infty} \frac{\sigma^2 \chi^2_{b-1}}{b - 1}
\]

(Glynn and Whitt, 1991; Steiger and Wilson, 2001). Further, under relatively mild conditions, one has

\[
\begin{align*}
\mathbb{E} \left[ \mathcal{N}(b, m) \right] &= \sigma^2 + \gamma (b + 1) / (bm) + o(1/m) \\
\lim_{n \to \infty} (b - 1) \text{Var} \left[ \mathcal{N}(b, m) \right] &= 2\sigma^4;
\end{align*}
\]

(11)

see Chien, Goldsman, and Melamed (1997).

Table 1 lists the key results of Equations (9) and (10) and Remark 4 for the folded STS area variance estimators together with the NBM properties (11). The entries in the table are valid for any stochastic process \( \{Y_i : i = 1, 2, \ldots\} \) and weight function \( f(\cdot) \) satisfying the symmetry condition \( f(t) = f(1-t) \) for \( t \in [0,1] \) as well as Assumptions A. Clearly, the combined estimator \( A_1(f_2; b, m) \) has smaller asymptotic MSE than the other estimators.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>((m/\gamma))Bias</th>
<th>((b/\sigma^4))Var</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\tilde{A}_0(f_2; b, m))</td>
<td>0</td>
<td>2</td>
<td>(2\sigma^4/b)</td>
</tr>
<tr>
<td>(\tilde{A}_1(f_2; b, m))</td>
<td>0</td>
<td>2</td>
<td>(2\sigma^4/b)</td>
</tr>
<tr>
<td>(A_1(f_2; b, m))</td>
<td>0</td>
<td>1</td>
<td>(\sigma^4/b)</td>
</tr>
<tr>
<td>(\mathcal{N}(b; m))</td>
<td>1</td>
<td>2</td>
<td>(\gamma^2/m^2 + 2\sigma^4/b)</td>
</tr>
</tbody>
</table>
5.5. Monte Carlo examples

We illustrate the performance of the proposed folded estimators with simple Monte Carlo experiments involving a stationary first-order autoregressive [AR(1)] process and a stationary M/M/1 queue-waiting-time process. In both cases, the experimentation is based on 4096 independent replications using \( b = 32 \) batches and various batch sizes (with the associated various sample sizes); and all STS area estimators are based on the quadratic weight function \( f_2(\cdot) \) from Example 1. We used common random numbers across all variance estimation methods based on the combined generator given in Figure 1 of L’Ecuyer (1999). We also checked the performance of the combined batched folded estimator

\[
\tilde{A}_1(f_2; 32, m) = \frac{1}{2}[\tilde{A}_0(f_2; 32, m) + \tilde{A}_1(f_2; 32, m)].
\]

5.5.1. AR(1) process

A Gaussian AR(1) process is constructed by setting \( Y_i = \phi Y_{i-1} + \epsilon_i, \ i = 1, 2, \ldots \), where the \( \epsilon_i \)'s are i.i.d. \( \text{Nor}(0, 1 - \phi^2) \), \( Y_0 \) is a standard normal random variable that is independent of the \( \epsilon_i \)'s, and \(-1 < \phi < 1\) (to preserve stationarity). It is well known that, for the AR(1) process, \( R_k = \phi^k, \ k = 0, 1, 2, \ldots \), and \( \sigma^2 = (1 + \phi)/(1 - \phi) \).

In the current example, we set the parameter \( \phi = 0.9 \) (so that \( \sigma^2 = 19 \)). Table 2 displays the averages of the batched folded area estimators \( \tilde{A}_0(f_2; 32, m) \) and \( \tilde{A}_1(f_2; 32, m) \) (from levels 0 and 1, respectively), the combined estimator \( A_1(f_2; 32, m) \), and the NBM estimator \( \mathcal{N}(32, m) \). Table 3 displays the standard errors of all of the estimators as well as estimates of the correlation between the level-0 and level-1 batched folded area estimators (in column 5). The standard errors—which range from about 0.05 for the combined estimators to about 0.07 for all of the other estimators—also serve as a comparative measure of estimator variation.

<table>
<thead>
<tr>
<th>( m )</th>
<th>( \tilde{A}_0(f_2; 32, m) )</th>
<th>( \tilde{A}_1(f_2; 32, m) )</th>
<th>( A_1(f_2; 32, m) )</th>
<th>( \mathcal{N}(32, m) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>18.04</td>
<td>16.48</td>
<td>17.26</td>
<td>18.12</td>
</tr>
<tr>
<td>512</td>
<td>18.66</td>
<td>18.06</td>
<td>18.36</td>
<td>18.60</td>
</tr>
<tr>
<td>1024</td>
<td>18.90</td>
<td>18.70</td>
<td>18.80</td>
<td>18.87</td>
</tr>
<tr>
<td>2048</td>
<td>19.01</td>
<td>18.85</td>
<td>18.93</td>
<td>18.85</td>
</tr>
<tr>
<td>4096</td>
<td>18.91</td>
<td>19.00</td>
<td>18.95</td>
<td>18.83</td>
</tr>
</tbody>
</table>

We summarize our conclusions from Tables 2 and 3 as follows:
Table 3. Estimated standard errors for the estimators in Table 2 and estimated correlations between the level-0 and level-1 batched folded area estimators for an AR(1) process with $\phi = 0.9$, $\sigma^2 = 19$, and $b = 32$

<table>
<thead>
<tr>
<th>$m$</th>
<th>$\hat{A}_0(f_2; 32, m)$</th>
<th>$\hat{A}_1(f_2; 32, m)$</th>
<th>$\hat{A}_1(f_2; 32, m)$</th>
<th>$\text{Corr}(\hat{A}_0, \hat{A}_1; 32, m)$</th>
<th>$\mathcal{N}(32; m)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.0711</td>
<td>0.0643</td>
<td>0.0479</td>
<td>-0.0031</td>
<td>0.0716</td>
</tr>
<tr>
<td>512</td>
<td>0.0735</td>
<td>0.0709</td>
<td>0.0509</td>
<td>-0.0075</td>
<td>0.0750</td>
</tr>
<tr>
<td>1024</td>
<td>0.0743</td>
<td>0.0740</td>
<td>0.0519</td>
<td>-0.0206</td>
<td>0.0739</td>
</tr>
<tr>
<td>2048</td>
<td>0.0757</td>
<td>0.0736</td>
<td>0.0528</td>
<td>0.0018</td>
<td>0.0743</td>
</tr>
<tr>
<td>4096</td>
<td>0.0757</td>
<td>0.0743</td>
<td>0.0531</td>
<td>0.0035</td>
<td>0.0748</td>
</tr>
<tr>
<td>$\rightarrow \infty$</td>
<td>0.0742</td>
<td>0.0742</td>
<td>0.0525</td>
<td>0.0000</td>
<td>0.0754</td>
</tr>
</tbody>
</table>

- The estimated expected values of all variance estimators converge to $\sigma^2$ as $m$ increases, in accordance with our theoretical results.

- For small values of $m$, the NBM variance estimator has expected values that are closer to $\sigma^2$ than those of the batched folded area estimators; but the latter estimators quickly catch up to (and pass) the NBM estimator as $m$ gets larger. This is a manifestation of the bias results summarized in Table 1, which show that the area estimators under study do better than NBM in terms of asymptotic bias.

- For small values of $m$, the level-1 folded estimators are more biased for $\sigma^2$ than their level-0 counterparts. This issue is the subject of ongoing research.

- As the batch size $m$ becomes large, the combined estimator $\hat{A}_1(f_2; 32, m)$ attains expected values that are comparable to those of its constituent estimators from levels 0 and 1, but with the bonus of substantially reduced variance compared to its constituents and the NBM estimator. Recall that the asymptotic variance of a batched folded area estimator is $2\sigma^4/b$, which translates to an asymptotic variance of approximately 22.56 for $b = 32$ and an approximate standard error of 0.0742 based on the 4096 independent experiments. The asymptotic variance of the NBM variance estimator is $2\sigma^4/(b - 1)$, which is close to that of a batched folded area estimator. On the other hand, the asymptotic variance for the combined estimator $\hat{A}_1(f_2; 32, m)$ is half of that of its constituent estimators (see Section 5.3), resulting in a reduction of the standard error by a factor of $\sqrt{2}$; see columns 3 and 4 of Tables 1 and 3, respectively.

- The low correlation estimates in column 5 of Table 3 confirm the result in Lemma 6.
5.5.2. \( M/M/1 \) queue-waiting-time process

We also consider the stationary waiting-time process for an \( M/M/1 \) queue with arrival rate \( \lambda \) and traffic intensity \( \rho < 1 \), i.e., a queueing system with Poisson arrivals and first-in-first-out i.i.d. exponential service times at a single server. For this process, we have \( \sigma^2 = \rho^3(2 + 5\rho - 4\rho^2 + \rho^3)/[\lambda^2(1 - \rho)^4] \); see, for example, Steiger and Wilson (2001).

In this particular example, we set the arrival rate at 0.8 and the service rate at 1.0, so that the server utilization parameter is \( \rho = 0.8 \), corresponding to a highly positive autocorrelation structure and variance parameter \( \sigma^2 = 1976 \). The set-ups of Tables 4 (for expected values) and 5 (for standard errors) for the \( M/M/1 \) process correspond to those of Tables 2 and 3, respectively, for the AR(1) process.

| \( m \) | \( \hat{A}_0(f_2; 32, m) \) | \( \hat{A}_1(f_2; 32, m) \) | \( \hat{A}_1(f_2; 32, m) \ | \( \hat{N}(32; m) \) |
|---|---|---|---|
| 1024 | 1827 | 1584 | 1705 | 1893 |
| 2048 | 1917 | 1831 | 1874 | 1930 |
| 4096 | 1955 | 1925 | 1940 | 1964 |
| 8192 | 1984 | 1966 | 1975 | 1963 |

The conclusions from Tables 4 and 5 are similar to those drawn from Tables 2 and 3. The convergence of all moments to their limiting values is generally slower due to the nonnormal nature of the marginal distribution of the waiting times and the long tail of their autocorrelation function. These effects are also evident in the slower convergence (to zero) of the correlation estimates in column 5 of Table 5.

6. Conclusions

The main purpose of this article is to introduce “folded” versions of the standardized time series area estimator for the variance parameter arising from a steady-state simulation output process. We provided theoretical results showing that the folded estimators converge to appropriate functionals of Brownian motion; and these convergence results allow us to
Table 5. Estimated standard errors for the estimators in Table 4 and estimated correlations between the level-0 and level-1 batched folded area estimators for an M/M/1 queue-waiting-time process with $\rho = 0.8$, $\sigma^2 = 1976$, and $b = 32$

<table>
<thead>
<tr>
<th>$m$</th>
<th>$A_0(f_2; 32, m)$</th>
<th>$A_1(f_2; 32, m)$</th>
<th>$A_1(f_2; 32, m)$</th>
<th>$\text{Corr}(\hat{A}_0, \hat{A}_1; 32, m)$</th>
<th>$N(32; m)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1024</td>
<td>17.15</td>
<td>12.49</td>
<td>12.66</td>
<td>0.4438</td>
<td>18.12</td>
</tr>
<tr>
<td>2048</td>
<td>15.14</td>
<td>13.34</td>
<td>12.10</td>
<td>0.4428</td>
<td>14.66</td>
</tr>
<tr>
<td>4096</td>
<td>13.13</td>
<td>12.36</td>
<td>10.66</td>
<td>0.3988</td>
<td>12.07</td>
</tr>
<tr>
<td>8192</td>
<td>11.03</td>
<td>10.65</td>
<td>8.76</td>
<td>0.3057</td>
<td>10.24</td>
</tr>
<tr>
<td>16384</td>
<td>9.60</td>
<td>9.51</td>
<td>7.45</td>
<td>0.2145</td>
<td>9.08</td>
</tr>
<tr>
<td>32768</td>
<td>8.41</td>
<td>8.76</td>
<td>6.47</td>
<td>0.1362</td>
<td>8.30</td>
</tr>
<tr>
<td>65536</td>
<td>7.98</td>
<td>8.18</td>
<td>5.83</td>
<td>0.0397</td>
<td>8.15</td>
</tr>
<tr>
<td>$\to \infty$</td>
<td>7.72</td>
<td>7.72</td>
<td>5.46</td>
<td>0.0000</td>
<td>7.84</td>
</tr>
</tbody>
</table>

produce asymptotically unbiased and low-variance estimators using multiple folding levels in conjunction with standard batching techniques.

Ongoing work includes the following. As in Remark 4, we can derive more-precise expressions for the expected values of the folded estimators—expressions that show just how quickly any estimator bias dies off as the batch size increases. We can also produce analogous folding results for other “primitive” STS variance estimators, e.g., for Cramér–von Mises estimators, as described in Antonini (2005). In addition, whatever type of primitive estimator we choose to use, there is interest in finding the best ways to combine batching and multiple folding levels in order to produce even-better estimators for $\sigma^2$, and subsequently, good confidence intervals for the underlying steady-state mean $\mu$. In any case, we have conducted a more-substantial Monte Carlo analysis to examine estimator performance over a variety of benchmark processes, the results of which will be given in Alexopoulos et al. (2007a). Future work includes the development of overlapping versions of the folded estimators, as in Alexopoulos et al. (2007b,c) and Meketon and Schmeiser (1984).

Acknowledgements

We thank the referees for their substantive and thoughtful comments and suggestions. Partial support for our research was provided by National Science Foundation Grant DMI-0400260.
References


Appendix
Proofs of Main Theorems

1.A Proof of Theorem 4.1

Every realization of a standard Brownian motion process \( \{\mathcal{W}(t) : t \in [0,1]\} \) is continuous at every \( t \in [0,1] \) (see p. 64 of Billingsley, 1968); and thus Definitions 2 and 4 imply that every realization of \( B_k(t) \) is also continuous at every \( t \in [0,1] \) for \( k = 0, 1, \ldots \). It follows that \( N_k(f) \) is the integral of a continuous function over the closed interval \([0,1]\) so that its Riemann sum satisfies (see Theorem 7.27 of Apostol, 1974)

\[
N_{k,m}(f) \equiv \frac{1}{m} \sum_{i=1}^{m} f\left(\frac{i}{m}\right) B_k\left(\frac{i}{m}\right) \xrightarrow{\text{a.s.}} \frac{N_k(f)}{m \to \infty}, \tag{A1}
\]

where “ \( \xrightarrow{\text{a.s.}} \) \( m \to \infty \)” denotes almost-sure convergence as \( m \to \infty \). For fixed \( k \) and \( m \), \( N_{k,m}(f) \) is normal since it is a finite linear combination of jointly normal random variables. Furthermore, \( N_{k,m}(f) \) has expected value 0 and variance

\[
\text{Var}[N_{k,m}(f)] = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} f\left(\frac{i}{m}\right) f\left(\frac{j}{m}\right) \text{Cov}\left[B_k\left(\frac{i}{m}\right), B_k\left(\frac{j}{m}\right)\right] = \int_0^1 \int_0^1 f(s)f(t) [\min\{s,t\} - st] \, ds \, dt + O(1/m) = 1 + O(1/m).
\]

Equation (A1) and the continuity theorem for characteristic functions imply that the characteristic function of \( N_{k,m}(f) \) converges to the characteristic function of \( N_k(f) \) as \( m \to \infty \).
(see p. 172 of Grimmett and Stirzaker, 1992). Since \( N_{k,m}(f) \) is normal with mean 0 and variance \( 1 + O(1/m) \), its characteristic function is given by

\[
\varphi_m(t) = \mathbb{E}\{\exp[\sqrt{-1} t N_{k,m}(f)]\} = \exp \left\{ -\frac{t^2}{2} [1 + O(1/m)] \right\}.
\]

(A2)

It follows immediately that \( \lim_{m \to \infty} \varphi_m(t) = \exp(-t^2/2) \) for all \( t \), the characteristic function of the standard normal distribution; and thus \( N_k(f) \sim \text{Nor}(0,1) \) by the continuity theorem for characteristic functions. \( \Box \)

2.A Proof of Theorem 4.2

By Lemma 4 and Equation (3),

\[
\text{Cov}[N_k(f_1), N_{e+k}(f_2)] = \text{Cov}[N_0(f_1), N_k(f_2)]
\]

\[
= \sum_{i=1}^{2^k-1} \int_0^1 \int_0^1 f_1(s) f_2(t) \text{Cov}[B(s), B(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}) - B(\frac{i}{2^{k-1}} - \frac{t}{2^k})] \, ds \, dt
\]

\[
= \sum_{i=1}^{2^k-1} \int_0^1 \int_0^1 f_1(s) f_2(t) \left[ \min \left\{ s, \frac{i-1}{2^{k-1}} + \frac{t}{2^k} \right\} - \min \left\{ s, \frac{i}{2^{k-1}} - \frac{t}{2^k} \right\} + \frac{s(1-t)}{2^{k-1}} \right] \, ds \, dt
\]

\[
= \sum_{i=1}^{2^k-1} \int_0^1 f_2(t) \left[ - \left( \frac{i}{2^{k-1}} - \frac{t}{2^k} \right) F_1(\frac{i}{2^{k-1}} - \frac{t}{2^k}) + \left( \frac{i-1}{2^{k-1}} + \frac{t}{2^k} \right) F_1(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}) + \hat{F}_1(\frac{i}{2^{k-1}} + \frac{t}{2^k}) F_1(1) - F_1(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}) - \hat{F}_1(\frac{i}{2^{k-1}} - \frac{t}{2^k}) F_1(1) + \hat{F}_1(\frac{i-1}{2^{k-1}} - \frac{t}{2^k}) \right] \, dt
\]

(since \( \int_a^b s f_1(s) \, ds = bF_1(b) - aF_1(a) - \hat{F}_1(b) + \hat{F}_1(a) \))

\[
= \sum_{i=1}^{2^k-1} \int_0^1 f_2(t) \left\{ \hat{F}_1(\frac{i}{2^{k-1}} - \frac{t}{2^k}) - \hat{F}_1(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}) - \frac{(1-t)F_1(\frac{i}{2^{k-1}} - \frac{t}{2^k})}{2^{k-1}} \right\} \, dt + (F_1 - \hat{F}_1)\hat{F}_2,
\]

from which the result follows. \( \Box \)

3.A Proof of Lemma 7
First, we show that every linear combination $\sum_{j=0}^{k} a_j N_j(f)$ has a normal distribution, and hence, $\mathbf{N}(f)$ has a multivariate normal distribution by virtue of Theorem 2.6.2 of Anderson (1984). Indeed,

$$\sum_{j=0}^{k} a_j N_j(f) = \int_0^1 \left[ \sum_{j=0}^{k} a_j f(t) B_j(t) \right] dt.$$  

Further, by Definition 2 and Equation (4), for each $B_j(t)$,

$$Z(t) \equiv \sum_{j=0}^{k} a_j f(t) B_j(t) = a_0 f(t) (tW(1) - W(t)) + \sum_{j=1}^{k} \sum_{i=1}^{2^{j-1}} f(t) \left[ W\left(\frac{i-1}{2^j} - \frac{t}{2^j}\right) - W\left(\frac{i-1}{2^j} + \frac{t}{2^j}\right)\right] - \left( \sum_{j=1}^{k} a_j \right) (1 - t) f(t) W(1).$$

Now, let $c_1, c_2, \ldots, c_m$ be real constants and $0 \leq t_1 < \cdots < t_m \leq 1$. Then

$$\sum_{\ell=1}^{m} c_{\ell} Z(t_\ell) = \sum_{\ell=1}^{m} c_{\ell} a_0 f(t_\ell) (W(t_\ell) - t_\ell W(1)) + \sum_{\ell=1}^{m} c_{\ell} \sum_{j=1}^{k} \sum_{i=1}^{2^{j-1}} f(t_\ell) \left[ W\left(\frac{i-1}{2^j} + \frac{t_\ell}{2^j}\right) - W\left(\frac{i-1}{2^j} - \frac{t_\ell}{2^j}\right)\right] + \sum_{\ell=1}^{m} c_{\ell} \left( \sum_{j=1}^{k} a_j \right) (1 - t_\ell) f(t_\ell) W(1).$$

Let $\mathcal{T}$ be the set of all times of the form $\frac{i-1}{2^j} + \frac{t_\ell}{2^j}$ or $\frac{i}{2^j} - \frac{t_\ell}{2^j}$, for some $\ell = 1, \ldots, m$, $j = 1, \ldots, k$, and $i = 1, \ldots, 2^{j-1}$. Let $\{\tau_1, \ldots, \tau_L\}$ be an increasing ordering of $\mathcal{T} \cup \{1\}$. Clearly, we can write $\sum_{\ell=1}^{m} c_{\ell} Z(t_\ell)$ as $\sum_{\ell=1}^{L} d_\ell W(\tau_\ell)$, for some real constants $d_1, \ldots, d_L$ since the function $f(\cdot)$ is deterministic. Since $W$ is a Gaussian process, the latter summation is Gaussian and thus, $Z$ is a Gaussian process. Notice also that $Z$ has continuous paths because $W$ has continuous paths. Finally recall that $\sum_{j=0}^{k} a_j N_j(f) = \int_0^1 Z(t) dt$; and the same methodology used in Theorem 4.1 can be used to show that the latter integral is a normal random variable.

To prove that $\mathbf{N}(f) = [N_0(f), \ldots, N_k(f)]$ has a nonsingular multivariate normal distribution, we show that the variance-covariance matrix $\Sigma_{\mathbf{N}(f)}$ is positive definite. This follows
immediately from Lemma 6 since

\[ a \Sigma_{N(f)} a^T = \text{Var} \left( \sum_{j=0}^{k} a_j N_j(f) \right) \]

\[ = \sum_{j=0}^{k} a_j^2 > 0, \]

for all \( a = (a_0, \ldots, a_k) \in \mathbb{R}^{k+1} - \{0_{k+1}\} \). \( \Box \)

4.A \hspace{1em} \textbf{Proof of Theorem 5.2}

In terms of the definition (1) and the Definitions 8–11, we see that

\[ \Theta^k_n(X_n) = A_k(f; n) \quad \text{for} \quad k = 0, 1, \ldots; \]

and we seek to apply the generalized CMT—that is, Theorem 5.5 of Billingsley (1968)—to prove that the \((k+1) \times 1\) random vector with \(j\)th element \(\Theta^j_n(X_n), j = 0, 1, \ldots, k\), converges in distribution to the \((k+1) \times 1\) random vector with \(j\)th element \(\Theta^j(W)\). To verify the hypotheses of the generalized CMT, we establish the following result. In terms of the set of discontinuities

\[ D_j \equiv \left\{ x \in D[0, 1] : \text{for some sequence} \{x_n\} \subset D[0, 1] \text{converging to} \ x, \text{the sequence} \right. \]

\[ \left\{ \Theta^j_n(x_n) \right\} \text{does not converge to} \ \Theta^j(x) \right\} \]

for \( j = 0, 1, \ldots, k \), we will show that

\[ \Pr \left\{ \mathcal{W}(\cdot) \in D[0, 1] - \bigcup_{j=0}^{k} D_j \right\} = 1. \quad (A3) \]

To prove (A3), we will exploit the almost-sure continuity of sample paths of \(\mathcal{W}(\cdot)\):

\[ \text{With probability 1, the function } \mathcal{W}(t) \text{ is continuous at every } t \geq 0; \quad (A4) \]

see p. 64 of Billingsley (1968). Thus we may assume without loss of generality that we are restricting our attention to an event \( \mathcal{H} \subset D[0, 1] \) for which (A4) holds so that

\[ \Pr\{\mathcal{W} \in \mathcal{H}\} = 1. \quad (A5) \]
Suppose \( \{x_n\} \subset D[0,1] \) converges to \( W \in \mathcal{H} \) and that \( j \in \{0,1,\ldots,k\} \) is a fixed integer. Next we seek to prove the key intermediate result,

\[
\lim_{n \to \infty} |\Psi^j \circ \Omega^n_{x_n}(t) - \Psi^j \circ \Omega^n_{W}(t)| = 0
\]

uniformly for \( t \in [0,1] \).

We prove (A6) by induction on \( j \), starting with \( j = 0 \). Choose \( \varepsilon > 0 \) arbitrarily. Throughout the following discussion, \( W \in \mathcal{H} \) and \( \{x_n\} \) are fixed; and thus virtually all the quantities introduced in the rest of the proof depend on \( W \) and \( \{x_n\} \). The sample-path continuity property (A4) and Theorem 4.47 of Apostol (1974) imply that \( W(t) \) is uniformly continuous on \( [0,1] \); and thus we can find \( \zeta > 0 \) such that

\[
\text{For all } t, t' \in [0,1] \text{ with } |t - t'| < \zeta, \text{ we have } |W(t) - W(t')| < \varepsilon/4.
\]

(A7)

Because \( \{x_n\} \) converges to \( W \) in \( D[0,1] \), there is a sufficiently large integer \( N \) such that for each \( n \geq N \), there exists \( \lambda_n(\cdot) \in \Lambda \) satisfying

\[
\sup_{t \in [0,1]} |\lambda_n(t) - t| < \min\{\zeta,\varepsilon/4\}
\]

(A8)

and

\[
\sup_{t \in [0,1]} |x_n(t) - W[\lambda_n(t)]| < \min\{\zeta,\varepsilon/4\}.
\]

(A9)

When \( j = 0 \), the map \( \Psi^j \) is the identity; and in this case for each \( n \geq N \) we have

\[
|\Psi^j \circ \Omega^n_{x_n}(t) - \Psi^j \circ \Omega^n_{W}(t)| = |\Omega^n_{x_n}(t) - \Omega^n_{W}(t)|
\]

\[
= \left| \left[ \frac{nt}{n} x_n(1) - x_n(t) \right] - \left[ \frac{nt}{n} W(1) - W(t) \right] \right|
\]

\[
\leq \frac{|nt|}{n} |x_n(1) - W(1)| + |x_n(t) - W(t)|
\]

(A10)

\[
\leq |x_n(1) - W[\lambda_n(1)]| + |W[\lambda_n(1)] - W(1)|
\]

\[
+ |x_n(t) - W[\lambda_n(t)]| + |W[\lambda_n(t)] - W(t)|
\]

(A11)

\[
\leq \varepsilon/4 + \varepsilon/4 + \varepsilon/4 + \varepsilon/4 = \varepsilon \text{ for } t \in [0,1],
\]

(A12)

where (A10) and (A11) follow from the triangle inequality and (A12) follows from (A7), (A8) and (A9). This establishes (A6) for \( j = 0 \).
Now suppose that (A6) holds for some \( j \geq 0 \). Again we choose \( \varepsilon > 0 \) arbitrarily. The induction hypothesis ensures that there exists \( N' \) sufficiently large such that for each \( n \geq N' \), we have
\[
|\Psi^j \circ \Omega^n_x(t) - \Psi^j \circ \Omega^n_W(t)| < \varepsilon/2 \text{ for all } t \in [0, 1]. \tag{A13}
\]
We have
\[
|\Psi^{j+1} \circ \Omega^n_x(t) - \Psi^{j+1} \circ \Omega^n_W(t)|
= \left| \left[ \Psi^j \circ \Omega^n_x \left( \frac{t}{2} \right) - \Psi^j \circ \Omega^n_x(1 - \frac{t}{2}) \right] - \left[ \Psi^j \circ \Omega^n_W \left( \frac{t}{2} \right) - \Psi^j \circ \Omega^n_W(1 - \frac{t}{2}) \right] \right| \tag{A14}
\leq \left| \Psi^j \circ \Omega^n_x \left( \frac{t}{2} \right) - \Psi^j \circ \Omega^n_W \left( \frac{t}{2} \right) \right|
+ \left| \Psi^j \circ \Omega^n_x \left( 1 - \frac{t}{2} \right) - \Psi^j \circ \Omega^n_W \left( 1 - \frac{t}{2} \right) \right| \tag{A15}
< \varepsilon/2 + \varepsilon/2 = \varepsilon \text{ for } t \in [0, 1] \text{ and } n \geq N', \tag{A16}
\]
where: (A14) follows from Definition 3 of the folding map; (A15) follows from the triangle inequality; and (A16) follows from (A13). This establishes (A6) for \( j = 0, 1, \ldots \).

Since \( f(\cdot) \) is continuous on \([0, 1]\) by Assumption A.4, we have
\[
f^* \equiv \max_{t \in [0, 1]} |f(t)| < \infty; \tag{A17}
\]
thus (A5), (A6), and (A17) imply that
\[
\lim_{n \to \infty} \left| \Theta^j_n(x_n) - \Theta^j_n(W) \right|
\leq f^* \sigma \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \left| \Psi^j \circ \Omega^n_x \left( \frac{i}{n} \right) - \Psi^j \circ \Omega^n_W \left( \frac{i}{n} \right) \right|
= 0 \text{ with probability 1.} \tag{A18}
\]

An argument similar to that justifying (A6) proves that

With probability 1, \( \lim_{n \to \infty} \left| \Psi^j \circ \Omega^n_W(t) - \Psi^j \circ \Omega_W(t) \right| = 0. \)

In view of the almost-sure continuity of sample paths of \( W(\cdot) \) and the continuity of \( f(\cdot) \), it is straightforward to show that, with probability 1, the function \( f(t) \Psi^j \circ \Omega_W(t) \) is continuous at every \( t \in [0, 1] \); and thus it follows that \( f(t) \Psi^j \circ \Omega_W(t) \) is Riemann integrable with probability 1 and that
\[
\lim_{n \to \infty} \left| \Theta^j_n(W) - \Theta^j(W) \right|
= \lim_{n \to \infty} \sigma \left| \frac{1}{n} \sum_{i=1}^n f \left( \frac{i}{n} \right) \Psi^j \circ \Omega_W \left( \frac{i}{n} \right) \right|
- \int_0^1 f(t) \Psi^j \circ \Omega_W(t) \, dt \right|
= 0 \text{ with probability 1.} \tag{A19}
\]
Combining (A18) and (A19) and applying the triangle inequality, we see that for each

\[ j \in \{0, 1, \ldots, k\}, \]

\[
\lim_{n \to \infty} |\Theta_j^n(x_n) - \Theta_j(W)| \\
\leq \lim_{n \to \infty} |\Theta_j^n(x_n) - \Theta_j^n(W)| + \lim_{n \to \infty} |\Theta_j^n(W) - \Theta_j(W)| \\
= 0 \text{ with probability 1.}
\]

It follows that the corresponding vector-valued sequence \( \{[\Theta_0^n(x_n), \ldots, \Theta_k^n(x_n)] : n = 1, 2, \ldots \} \) converges to \( [\Theta^0(W), \ldots, \Theta^k(W)] \) in \( \mathbb{R}^{k+1} \) with probability one; and thus the desired result (6) follows directly from the generalized CMT. \( \square \)

**Biographies**

Claudia Antonini has a tenure track position in the Department of Pure and Applied Mathematics at Simón Bolívar University in Caracas, Venezuela. Her research interests include simulation output analysis, quality control, stochastic processes, and statistics. Partially on the basis of her original work documented in this article, she received the 2006 Pritsker Doctoral Dissertation Award (Second Place) from IIE. She is a member of the AMV (Asociación Matemática Venezolana).

Christos Alexopoulos is an Associate Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. His research interests are in the areas of simulation, statistics, and optimization of stochastic systems. He is a member of INFORMS and an active participant in the Winter Simulation Conference, having been Proceedings Co-Editor in 1995, Associate Program Chair in 2006, and a member of the Board of Directors since 2007. He is also the Simulation Department Editor of *IIE Transactions*.

David Goldsman is a professor in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. His research interests include simulation output analysis and ranking and selection. He is an active participant in the Winter Simulation Conference, having been Program Chair in 1995, and having served on the WSC Board of Directors since 2002.

James R. Wilson is a professor in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. Currently he is Editor-in-Chief of *ACM Transactions on Modeling and Computer Simulation*. He has also held various offices in the INFORMS Simulation Society; and during the period 1997–2004, he was a carepresentative of that organization to the Board of Directors of the Winter Simulation Conference. He is a Fellow of IIE and INFORMS, and he is also a member of AAUW, ACM, ASA, and ASEE.