

Adaptive Minimum Confidence Region Rule for Multivariate Initialization Bias Truncation in Discrete-Event Simulations

Jianguo Wu, Honglun Xu, Feng Ju & Tzu-Liang (Bill) Tseng

To cite this article: Jianguo Wu, Honglun Xu, Feng Ju & Tzu-Liang (Bill) Tseng (2020) Adaptive Minimum Confidence Region Rule for Multivariate Initialization Bias Truncation in Discrete-Event Simulations, *Technometrics*, 62:4, 499-512, DOI: [10.1080/00401706.2019.1665590](https://doi.org/10.1080/00401706.2019.1665590)

To link to this article: <https://doi.org/10.1080/00401706.2019.1665590>

 View supplementary material 

 Published online: 15 Oct 2019.

 Submit your article to this journal 

 Article views: 167

 View related articles 

 View Crossmark data 

 Citing articles: 1 View citing articles 



Adaptive Minimum Confidence Region Rule for Multivariate Initialization Bias Truncation in Discrete-Event Simulations

Jianguo Wu^a, Honglun Xu^b, Feng Ju^c, and Tzu-Liang (Bill) Tseng^d

^aIndustrial Engineering and Management, College of Engineering, Peking University, Beijing, China; ^bComputational Science Program, University of Texas at El Paso (UTEP), El Paso, TX; ^cSchool of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ; ^dIndustrial, Manufacturing and Systems Engineering, University of Texas at El Paso (UTEP), El Paso, TX

ABSTRACT

Initialization bias truncation is critically important for system performance assessment and warm-up length estimation in discrete-event simulations. Most of the existing methods are for univariate signals, while multivariate truncation has been rarely studied. To fill such gap, this article proposes an efficient method, called adaptive minimum confidence region rule (AMCR) for multivariate initialization bias truncation. It determines the truncation point by minimizing the modified confidence volume with a tuning parameter for the mean estimate. An elbow method is developed for adaptive selection of the tuning parameter. Theoretical properties of the AMCR rule for both data with and without autocorrelations have been derived for justification and practical guidance. The effectiveness and superiority of the AMCR rule over other existing approaches have been demonstrated through thorough numerical studies and real application. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received March 2018
Accepted September 2019

KEYWORDS

Asymptotically unbiased estimator; Autocorrelation; Generalized variance; Minimum confidence region; Steady state

1. Introduction

In the practice of discrete-event simulations, the starting condition is often arbitrarily selected due to the lack of true system performance. As a result, the system steady state will not be achieved until the simulation is initiated for certain period of time (Schruben 1982). This initial situation is often referred to as a transient state, startup or warm-up period. Since the startup period does not represent the true system performance, the initialization bias may seriously contaminate the simulation output. One natural way is to allow the program to run for a certain amount of time before the output data are collected, which is equivalent to removing some early portion of the output. However, if too little of the early output is truncated, the remaining bias may still affect the performance evaluation; while ignorance of too much of the initial output will result in insufficient or unnecessary waste of observations. On the other hand, the length of the warm-up period itself may be an important performance measure, which captures the recovering speed of the system from transient state to steady state (Ju, Li, and Horst 2017). For example, a long transient period in production systems may result in substantial amount of production loss. In the transient analysis, the warm-up length needs to be identified. Therefore, an accurate method to find the truncation point is critically important for both steady state performance evaluation and warm-up period length estimation.

During the past few decades, tens of methods or rules have been developed for initialization bias truncation. According to Robinson (2014), these methods can be classified into five

categories: (1) graphic methods, which involve visual inspection of the time-series output and human judgment. The typical examples include various statistical process control (SPC) charts (Rossetti, Li, and Qu 2005; Robinson 2007), where the steady state observations are selected conservatively to construct the SPC chart and then use the constructed chart to determine the truncation point. (2) Heuristic approaches, which provide simple rules to determine when to truncate the data, with few underlying assumptions. One well-known heuristic approach is the marginal standard error rule (MSER) (White Jr. 1997; White Jr., Cobb, and Spratt 2000; Wu, Zhou, and Li 2013), which selects as the truncation point the one that minimizes the width of the marginal confidence interval about the estimated steady state mean. (3) Statistical methods, which are based on statistical principles. One typical example is the Kelton and Law's method (Linton and Harmonosky 2002), which tests the regression slope of the times series starting from the end. Once the slope is no longer zero with certain significance level, the point is selected as the truncation point. (4) Initialization bias tests, which test whether there is any initialization bias in the data. One representative method is the batch mean test (Goldsman, Schruben, and Swain 1994), where the output series is divided into successive b batches of equal length, and then these batches are grouped into two sets of b' and $b - b'$ successive batches. The mean of each batch is calculated, and the variances for the b' and $b - b'$ batches are estimated using the calculated means. The ratio of the two variances is used as test statistic to test if there is any initialization bias. (5) Hybrid methods (Pawlikowski 1990; Jackway and Desilva 1992), which combines the initialization

bias tests with truncation methods to determine the warm-up period.

However, all of these aforementioned methods are for univariate output. In many applications, multiple performance measures, which are often correlated, may be of interest. For example, in the simulation of serial production lines (Kang, Ju, and Zheng 2017), several performance measures are of particular interests, including production rate (PR), consumption rate (CR), scrap rate (SR), and work-in-process (WIP), which are all correlated. To get an accurate performance evaluation, all the performance measures or even their correlations have to be in a steady state. One possible approach is to select the truncation point based on an aggregate of all the output variables, for example, the mean time series. However, this may result in loss of important information regarding the initialization bias. Another natural method would be to apply a certain univariate truncation method to each univariate output individually, and then select the largest truncation point as the starting point of the steady state time series. However, similar to SPC charts (Montgomery 2009), truncating each dimension separately may not be effective when all dimensions have small biases. Considering all dimensions together would “aggregate” all the biases of each dimension and thus enhance the capability of truncating small biases. Besides, a univariate method is not able to detect the initialization bias in terms of dependencies or correlations. Therefore, the multivariate initialization bias truncation method is desirable.

There are a very limited number of multivariate methods in existing literature. To our best knowledge, there are only two multivariate truncation methods. One was developed by Shruben (1981), who proposed to select a truncation point based on Hotelling’s T^2 statistic. In this method, the observations are divided into small batches. It is assumed that the simulation run is long enough so that the last batch can be assumed to be at steady state. Then a Hotelling’s T^2 control chart is developed based on the observations of the last batch. The T^2 statistic for each batch is calculated, and the truncation point is selected as the first in-control batch with a certain significant level α . However, this method requires that there are sufficient observations in the steady state so that the mean and covariance matrix can be accurately estimated. When the dimension of the multivariate output is very large while the number of steady state observations is low, this method may not be effective. Another method was developed by Gallagher, Bauer, and Maybeck (1994), who proposed a special state space model to capture the evolution of the mean output and measured output, and then applied the multiple model adaptive estimation (MMAE) based Kalman filter to estimate the mean sequentially. Once the sequentially estimated mean is sufficiently close to the mean in the steady state, the time index is selected as the truncation point. However, in this method, the state-space transition model uses a rigid autoregressive model, which is not flexible or general. Besides, the prior selection in the MMAE is very subjective.

To fill the gap, an efficient method, called adaptive minimum confidence region rule (AMCR), is developed in this article for multivariate initialization bias truncation. It is an extension of the univariate MSER method to multivariate cases. More specifically, this method is to find a truncation point that minimizes the confidence region (volume) of the mean estimate of

the observations after truncation. To make the method more sensitive to initialization bias and more robust to signals of various initialization bias shapes and severity, an innovative tuning parameter is incorporated into the confidence volume calculation, and can be adaptively selected through a developed elbow method. Various theoretical properties regarding the average behavior of this method for both data with and without autocorrelations have been derived, which provides theoretical foundation and insightful guidance on methodology understanding and designing. Although the AMCR rule is mainly for mean bias truncation, it can also be effectively applied to remove initialization variance or covariance matrix biases, as theoretically and numerically shown in this article.

The rest of this article is organized as follows. Section 2 provides a detailed description of the AMCR rule. In Section 3, various theoretical properties have been derived to facilitate deep understanding of this method and provide insightful application guidance. Numerical illustrations, tuning parameters selection, performance evaluation, and real applications are given in Section 4. Section 5 presents the conclusions.

2. Adaptive Minimum Confidence Region Rule for Multivariate Initialization Bias Truncation

In the simulation output, the data are often autocorrelated and may not follow normal distributions. However, the iid Gaussian assumption is often the basic assumption in many initialization bias truncation methods. To address the nonnormality issue, many existing truncation methods (Shruben 1981; Robinson 2007) propose to perform a series of independent replications. Based on the central limit theorem, the sample mean at each time interval tends toward normality. To handle the autocorrelation issue, the batch means method is commonly used (Shruben 1981; Goldsman, Shruben, and Swain 1994; Runger and Willemain 1996; Robinson 2007). By dividing observations into batches, the batch means become approximately uncorrelated as the batch size increases. Since the nonnormality and autocorrelation issues can be solved through these data preprocessing steps, we assume that the multivariate output to be truncated already follow iid normal distributions. In Section 3, we will theoretically show that the proposed approach can also be effectively applied to the data with autocorrelations.

Suppose the steady state p -dimension multivariate output $X_1, X_2, \dots, X_n$ follow iid normal distributions, that is, $X_i \sim N(\mu, \Sigma)$ for $i = 1, \dots, n$. Denote the mean of the output by $\bar{X}$, that is, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, then

$$n(\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu) \sim \chi_p^2, \quad (1)$$

where χ_p^2 is the chi-squared distribution with p degree of freedom. A joint $100(1 - \alpha)\%$ confidence region for the mean satisfies

$$(\mu - \bar{X})^T \Sigma^{-1} (\mu - \bar{X}) \leq \frac{\chi_{p,\alpha}^2}{n} \quad (2)$$

where $\chi_{p,\alpha}^2$ is the upper $100\alpha\%$ percentile of χ_p^2 distribution. Clearly, the confidence region is a p -dimensional ellipsoid (ellip

for $p = 2$) with semi-axes $\sqrt{\frac{\chi_{p,\alpha}^2}{n}} \lambda_i, i = 1, 2, \dots, p$ where $\lambda_i, i = 1, 2, \dots, p$ are eigenvalues of the covariance matrix Σ satisfying $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$. The volume of the confidence region can be calculated as (Wilson 2010)

$$V(\alpha) = \frac{2\pi^{\frac{p}{2}}}{p\Gamma(\frac{p}{2})} \left(\frac{\chi_{p,\alpha}^2}{n} \right)^{p/2} \sqrt{\prod_{i=1}^p \lambda_i} \\ = \frac{2\pi^{\frac{p}{2}}}{p\Gamma(\frac{p}{2})} \left(\frac{\chi_{p,\alpha}^2}{n} \right)^{p/2} |\Sigma|^{1/2}. \quad (3)$$

When Σ is unknown, which is usually the case in practice, the sample covariance could be used instead. Denote the sample covariance matrix by S , which is calculated as

$$S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T. \quad (4)$$

Then

$$T^2 = n(\bar{X} - \mu)^T S^{-1} (\bar{X} - \mu) \sim T_{p,n-1}^2 = \frac{p(n-1)}{n-p} F_{p,n-p}, \quad (5)$$

where $T_{p,n-1}^2$ is a Hotelling T^2 distribution and $F_{p,n-p}$ is the F -distribution with p and $n-p$ degrees of freedom (Hotelling 1931). Similarly, the volume of the 100 $(1-\alpha)\%$ confidence region can be obtained as

$$V(\alpha) = \frac{2\pi^{\frac{p}{2}}}{p\Gamma(\frac{p}{2})} \left(\frac{p(n-1)}{n(n-p)} F_{p,n-p,\alpha} \right)^{p/2} |S|^{1/2}. \quad (6)$$

Here $|S|$ is the sample generalized variance used to measure the spread of observations. When $n \gg p$, $(n-1)/(n-p) \approx 1$. For the sake of simplicity, we approximate the volume by

$$V(\alpha) \approx \frac{2\pi^{\frac{p}{2}}}{p\Gamma(\frac{p}{2})} (pF_{p,n-p,\alpha})^{\frac{p}{2}} \frac{|S|^{1/2}}{n^{p/2}}. \quad (7)$$

Suppose the p -dimensional simulation output is $\{X_1, X_2, \dots, X_n\}$, which may have initialization bias. We determine the truncation point c^* by minimizing the volume of the confidence region for any confidence level $1-\alpha$:

$$c^* = \arg \min_{n \gg c \geq 0} \left(\frac{|S_{c+1:n}|^{1/2}}{(n-c)^{p/2}} \right) = \arg \min_{n \gg c \geq 0} \left(\frac{|S_{c+1:n}|}{(n-c)^p} \right), \quad (8)$$

where $S_{c+1:n}$ is the sample covariance matrix for output $\{X_{c+1}, X_{c+2}, \dots, X_n\}$. Note that $F_{p,n-p,\alpha}$ is excluded in Equation (8). The reason is that when n is large, according to the law of large numbers, $F_{p,n-p} = \frac{\chi_p^2/p}{\chi_{n-p}^2/(n-p)} \rightarrow \chi_p^2/p$, which does not depend on n . When $p = 1$, Equation (8) is exactly the MSER for univariate output. As observed in the experiment, the truncation rule given in Equation (8) often fails to truncate all the bias, especially when the process noise is large. Part of the initialization bias is often immersed by system noise and thus not detectable. To make the method more flexible, we use the following truncation rule by replacing the exponent of the denominator with a tuning parameter h :

$$c^* = \arg \min_{n \gg c \geq 0} \left(\frac{|S_{c+1:n}|}{(n-c)^h} \right). \quad (9)$$

Here h is a tuning parameter used to adjust the sensitivity of bias truncation, and it satisfies $p \geq h > 0$. Equation (9) can be better interpreted using the following equivalent form with a log-transformation:

$$c^* = \arg \min_{n \gg c \geq 0} (\log |S_{c+1:n}| - \lambda p \log(n-c)), \quad (10)$$

where $-\lambda p \log(n-c)$ can be treated as a penalty term favoring more observations, and $\lambda = h/p \in (0, 1]$ is a penalty coefficient. If there is no bias within $Y_{c+1:n}$, it is easy to show that $\log |S_{c+1:n}|$ will converge in probability to the constant $\log |\Sigma|$. In such case, more observations or smaller c would be more preferable. However, if there are initialization biases within $Y_{c+1:n}$, $\log |S_{c+1:n}|$ will increase as more biased observations are added. Therefore, λ controls the tradeoff between the number of observations and biases. In the following section, we will theoretically show why and how h influences the sensitivity of bias detection. The tuning parameter h may play a decisive role on the truncation accuracy. In Section 4, an adaptive tuning parameter selection method, called elbow method, will be developed and numerically illustrated for the tuning parameter selection. We call the proposed method adaptive minimum confidence region (AMCR) rule. In the following section, several theoretical properties for the proposed truncation rule are provided for justification and guidance.

3. Theoretical Foundation for AMCR

For notational convenience, we denote the AMCR statistic by R and define

$$R_n = \frac{|S_{1:n}|}{n^h} \text{ and } R_{i:n} = \frac{|S_{i:n}|}{(n-i+1)^h}. \quad (11)$$

In this section, several theoretical properties are derived for R , which lays the foundation for the AMCR rule. The first property is regarding the behavior of R if another observation following the same distribution is added to the existing iid observations. Specifically, we will show that for a set of iid observations with a sufficiently large sample size from a normal distribution, adding an extra independent observation from the same distribution will reduce the expectation and variance of R . This property is presented in Theorem 1 as follows.

Theorem 1. Suppose $X_1, X_2, \dots, X_n$ follow iid normal distributions of dimension p , that is, $X_i \sim N(\mu, \Sigma)$ for $i = 1, \dots, n$. Define $R_n = \frac{|S_{1:n}|}{n^h}$ where $0 < h \leq p \ll n$, then

1. When $n > g_1(p, h) = \frac{(p^2 + ph + h - p) + \sqrt{(p^2 + ph + h - p)^2 - 4ph(h - p + 1)}}{2h}$, $E(R_n)$ decreases monotonically with n and $\lim_{n \rightarrow \infty} E(R_n) = 0$;
2. When $n > g_2(p, h)$ where $g_2(p, h)$ is a function of p and h , $\text{var}(R_n)$ decreases monotonically with n and $\lim_{n \rightarrow \infty} \text{var}(R_n) = 0$.

To prove Theorem 1, it is essential to obtain $E(|S_{1:n}|)$ and $\text{var}(|S_{1:n}|)$. These two terms can be calculated as (Montgomery

2009; Muirhead 2009)

$$E(|S_{1:n}|) = \frac{\prod_{i=1}^p (n-i)}{(n-1)^p} |\Sigma|, \\ \text{var}(|S_{1:n}|) = \frac{\prod_{i=1}^p (n-i) [\prod_{j=1}^p (n-j+2) - \prod_{j=1}^p (n-j)]}{(n-1)^{2p}} \\ \times |\Sigma|^2. \quad (12)$$

The detailed proof of Theorem 1 is provided in Section 1 of the supplementary materials. In Theorem 1, $g_1(p, h)$ and $g_2(p, h)$ are two constants that depend on both p and h . $g_2(p, h)$ does not have analytical forms, and needs to be numerically calculated. These two constants are often small compared with the number of steady state observations in applications. It should be noted the actual lower ends of the decreasing intervals are often much smaller than these two constants. g_1 and g_2 are just upper bounds for the lower ends of the decreasing intervals. For the special case where $h = p$, $g_1(p, h)$ can be simplified as $g_1 = p + \sqrt{p^2 - 1}$. When $h = p = 1$, we can see that $E(R_n)$ decreases monotonically for all $n \geq 1$. It is also worth noting that if the exact volume of the confidence region in Equation (6) is used, the monotonicity property will be satisfied for all n , which can be shown numerically. The reason is that $F_{p, n-p, \alpha}$ decreases rapidly with n when n is small.

Theorem 1 tells us that the mean curve of R_n is monotonically decreasing and the confidence band of R_n becomes narrower and narrower as we add more observations from the same stationary process. Indeed, the variance of R_n converges so fast ($\sim O(\frac{1}{n^{1+2h}})$) that we can even safely treat R_n as a deterministic and monotonically decreasing function when n is sufficiently large. When applied to the proposed AMCR rule, this theorem tells us that the truncation point c^* is very likely to be on the left of the true steady state transition time c_0 , that is, $c^* \leq c_0$. The reason is that starting from the end to the beginning of the simulation output sequence, adding steady state observations tends to reduce the confidence volume, that is, $R_{i:n}$. As often observed in simulations, when there is no initialization bias, $c^* = 0$ is often selected as the truncation point using the AMCR rule.

The next step is to investigate the behavior of R if independent transient observations are added to the steady state observations. For simplicity, we assume that there is only mean shift for the observations in the transient period, while the covariance matrix is unchanged. First consider the simple case where only one transient observation is added to the steady state sequences. To investigate the mean behavior of R , the calculation of the expectation of the generalized variance is the main challenge, which is given in Theorem 2 as follows.

Theorem 2. Suppose $X_1, X_2, \dots, X_n$ follow iid normal distributions of dimension p , that is, $X_i \sim N(\mu_0, \Sigma)$ for $i = 1, \dots, n$, and $X_{n+1} \sim N(\mu, \Sigma)$ where $\mu \neq \mu_0$. Define the noncentrality matrix $T = \sum_{i=1}^{n+1} (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^T$ where $\mu_i = \mu_0$ for $i = 1, \dots, n$, $\mu_{n+1} = \mu$, and $\bar{\mu} = \frac{\sum_{i=1}^{n+1} \mu_i}{n+1}$. Let $k_j (j = 1, \dots, p)$ with $k_1 \geq k_2 \geq \dots \geq k_p$ be the real and nonnegative roots of the determinantal equation $|T - k^2 \Sigma| = 0$, then

$$1. \quad k_j = 0 \text{ for } j = 2, \dots, p \text{ and} \\ k_1^2 = \frac{n}{n+1} (\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0). \quad (13)$$

2.

$$E(|S_{1:n+1}|) = \frac{|\Sigma| (n + k_1^2)}{n^p (n - p)} \prod_{i=1}^p (n - i). \quad (14)$$

The proof of Theorem 2 is provided in Supplementary Section 2. Based on this theorem, we can conveniently derive the behavior of the expected AMCR statistic, which is summarized in Theorem 3 (see Supplementary Section 3 for proof).

Theorem 3. Suppose $X_1, X_2, \dots, X_n$ follow iid normal distributions of dimension p , that is, $X_i \sim N(\mu_0, \Sigma)$ for $i = 1, \dots, n$. If one independent observation $X_{n+1} \sim N(\mu, \Sigma)$ where $\mu \neq \mu_0$ is added, then $E(R_{n+1}) > E(R_n)$ iff

$$(\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0) > \frac{(n-p)(n+1)^{h+1}}{(n-1)^p n^{h-p+1}} - (n+1). \quad (15)$$

In the above inequality, $\frac{(n-p)(n+1)^{h+1}}{(n-1)^p n^{h-p+1}} - (n+1) \approx h$ (Supplementary Section 3), and $(\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0)$ is just the squared Mahalanobis distance or the squared generalized distance between μ and μ_0 , which measures the severity of the initialization bias. For notational convenience, we define

$$D_s^2 = (\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0). \quad (16)$$

Theorem 3 tells us that adding an independent observation with a different mean to the steady state sequence will increase the expected AMCR statistic as long as the mean shift is sufficiently large, for example, $D_s^2 > h$. We can also show that the difference $E(R_{n+1}) - E(R_n)$ increases as we reduce h . If we ignore the change of variance (when n is large, the variance is almost negligible), the lower the value h , the more likely the AMCR statistic will increase when adding an observation from a distribution with a different mean. Therefore, the tuning parameter h controls the sensitivity of the AMCR rule in detecting initialization bias. It should be noted that h also controls the decreasing rate of $E(R)$ for steady state observations, which can be easily seen from the proof of Theorem 1. The higher the parameter h , the higher the decreasing rate of $E(R)$ for steady state sequences and thus the less likely the occurrence of over-truncation, that is, some steady state observations are truncated as initialization bias. Therefore, the tradeoff between truncation sensitivity and over-truncation needs to be considered in the selection of the tuning parameter h .

Theorem 3 provides some insight into the behavior of the expected AMCR statistic when adding only one observation from the transient period. In practice, however, the multivariate output often evolves slowly from transient period to steady state period. In other words, the initialization bias becomes less and less severe in the transient period and eventually diminishes after entering the steady state period. That means, the bias of observations right before the true steady state time is so small that these observations may not cause the AMCR statistic to increase. Therefore, it is necessary to investigate the cumulative influence of successive biased observations on the AMCR statistic. In other words, the problem of how many biased observations are needed for $E(R)$ to increase needs to be investigated,

which is very important since it directly determines the bias truncation capability or sensitivity.

In practice, the mean paths of the multivariate output in transient period often vary across different applications. For simplicity, we assume that the mean of the observations changes linearly while the covariance matrix is fixed in transient period. Besides simplicity, the linear-form assumption of the initialization bias has another two advantages. First, in many cases, the functional form within a short time period just before the true steady state time can be well approximated by a linear function. Second, by specifying a linear form, we can conveniently set the mean shifting rate, or the severity of the initialization bias, and thus are able to investigate how the severity of initialization bias affects the truncation accuracy. The theoretical results for the case of linear initialization bias are summarized in [Theorem 4](#) as follows (see Supplementary Section 4 for proof).

Theorem 4. Suppose p -dimensional vector $X_1, X_2, \dots, X_c$ in the transient period follow independent normal distributions with $X_i \sim N(\mu_i, \Sigma)$ for $i = 1, \dots, c$, where

$$\mu_i = \mu_b + \frac{(i-1)(\mu_0 - \mu_b)}{c}, i = 1, 2, \dots, c, \quad (17)$$

and μ_b is the mean of the first observation. Suppose p -dimensional vector $X_{c+1}, X_{c+2}, \dots, X_{c+n}$ in the steady state period follow iid normal distributions with $X_i \sim N(\mu_0, \Sigma)$, $i = c+1, \dots, c+n$. Assume t biased observations right before steady state time are added. Define the noncentrality matrix $T = \sum_{i=c-t+1}^{c+n} (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^T$ where $\bar{\mu} = \frac{\sum_{i=c-t+1}^{c+n} \mu_i}{n+t}$. Let $k_{j,c-t+1:n+c}$ ($j = 1, \dots, p$) with $k_{1,c-t+1:n+c} \geq \dots \geq k_{p,c-t+1:n+c}$ be the real and nonnegative roots of the following determinantal equation $|T - k^2 \Sigma| = 0$, then

1. $k_{j,c-t+1:n+c} = 0$ for $j = 2, \dots, p$ and

$$k_{1,c-t+1:n+c}^2 = \frac{D_s^2 f(n, c, t)}{4c^2 (n+t)^2}, \quad (18)$$

where

$$f(n, c, t) = \left\{ nt^2 (t+1)^2 + \sum_{i=c-t+1}^c [-2nc + 2(i-1)(n+t) - (2c-t-1)t^2] \right\}. \quad (19)$$

- 2.

$$E(|S_{c-t+1:n+c}|) = \frac{|\Sigma| (n+t-1 + k_{1,c-t+1:n+c}^2)}{(n+t-1)^p (n+t-1-p)} \times \prod_{i=1}^p (n+t-1-i). \quad (20)$$

3. If $E(R_{c-t+1:n+c})$ starts to increase with t when $t > t^*$, then t^* can be approximated by

$$t^* \approx \frac{\frac{\sqrt{h}}{r_s} - 1}{1 - \frac{\sqrt{h}}{nr_s}}, \quad (21)$$

where r_s is the mean shifting rate defined as

$$r_s = \frac{D_s}{c} = \frac{\sqrt{(\mu_0 - \mu_b)^T \Sigma^{-1} (\mu_0 - \mu_b)}}{c}. \quad (22)$$

From Equation (21) we can clearly see that the number of biased observations needed for $E(R_{c-t+1:n+c})$ to increase depends on both the tuning parameter h and the mean shifting rate r_s . t^* is an increasing function of h , which is consistent with [Theorem 3](#) that increasing h will reduce the bias truncation sensitivity. Besides, t^* monotonically decreases with the mean shifting rate r_s , which is expected since increasing r_s will enlarge the bias severity, and consequently reduce the number of biased observations for $E(R)$ to increase. Note that in Equation (21), when $r_s > \sqrt{h}$, the expected AMCR statistic will increase immediately when a biased observation is added to the steady state sequences, which is exactly the result obtained in [Theorem 3](#). Therefore, Equation (21) could provide us insightful guidance in understanding the behavior and designing of the AMCR rule.

Although the proposed AMCR is based on the assumption that all observations are independent, it can also be effectively applied to initialization bias truncation problems with autocorrelations. Due to the introduction of autocorrelation, however, the expectation of the determinant of the sample covariance matrix $E(|S_{k:n}|)$ is extremely difficult to obtain analytically, if not impossible. Instead, we derive the asymptotic form of $E(|S_{k:n}|)$, which is given in [Theorem 5](#). For simplicity yet without loss of generality, we assume that the stochastic process is a first-order vector autoregressive process, denoted by VAR(1). To prove [Theorem 5](#), [Lemma 1](#) is first given as follows.

Lemma 1. Suppose $X = (x_1, \dots, x_p) \sim N_p(\mu, \Sigma)$ where $\Sigma = (\sigma_{ij}^2)$, and suppose $i_1, \dots, i_r$ are indices from $\{1, 2, \dots, p\}$ and they need not all be distinct, then

$$E(x_{i_1} x_{i_2} \dots x_{i_r}) = \sum_{l+2k=r}^m \sum \mu_{i_1} \dots \mu_{i_l} \sigma_{j_1 j_2}^2 \dots \sigma_{j_{2k-1} j_{2k}}^2, \quad (23)$$

where Σ^m sums over all $m = \frac{r!}{l! 2^k k!}$ permutations $\{i_1 \dots i_l j_1 j_2 \dots j_{2k}\}$ of $\{i_1, \dots, i_r\}$ giving distinct terms allowing for the symmetry of Σ .

The proof of [Lemma 1](#) can be found in [Withers \(1985\)](#). Based on [Lemma 1](#), we can get [Theorem 5](#) as follows.

Theorem 5. Suppose X_t is a stable p -dimensional VAR(1) process with $X_t = \mu + A(X_{t-1} - \mu) + u_t$, where μ is the steady state mean vector, all the eigenvalues $\alpha_1, \dots, \alpha_p$ of the matrix A have modulus less than 1 (stability condition), that is, $|\alpha_i| < 1$, and u_t is the iid noise terms with $u_t \sim N(0, \Sigma_u)$. Suppose $X_0 \sim N(\mu_0, \Sigma)$ and the covariance matrix is stable, that is, $\Sigma = A \Sigma A^T + \Sigma_u$ or $\text{vec}(\Sigma) = (I - A \otimes A)^{-1} \text{vec}(\Sigma_u)$ where $\otimes$ is the Kronecker product. Define $b_i = \mu_i - \mu$ as the mean bias at the time step i , then we have

- 1.

$$E|S_{k+1:n}^*| = |\Sigma| \left(1 + \frac{1}{n-k} \sum_{i=k+1}^n b_i^T \Sigma^{-1} b_i + \frac{C}{n-k} + O\left(\frac{1}{(n-k)^2}\right) \right), \quad (24)$$

where $S_{k+1:n}^* = \frac{n-k-1}{n-k} S_{k+1:n}$, C is a constant independent of k and n .

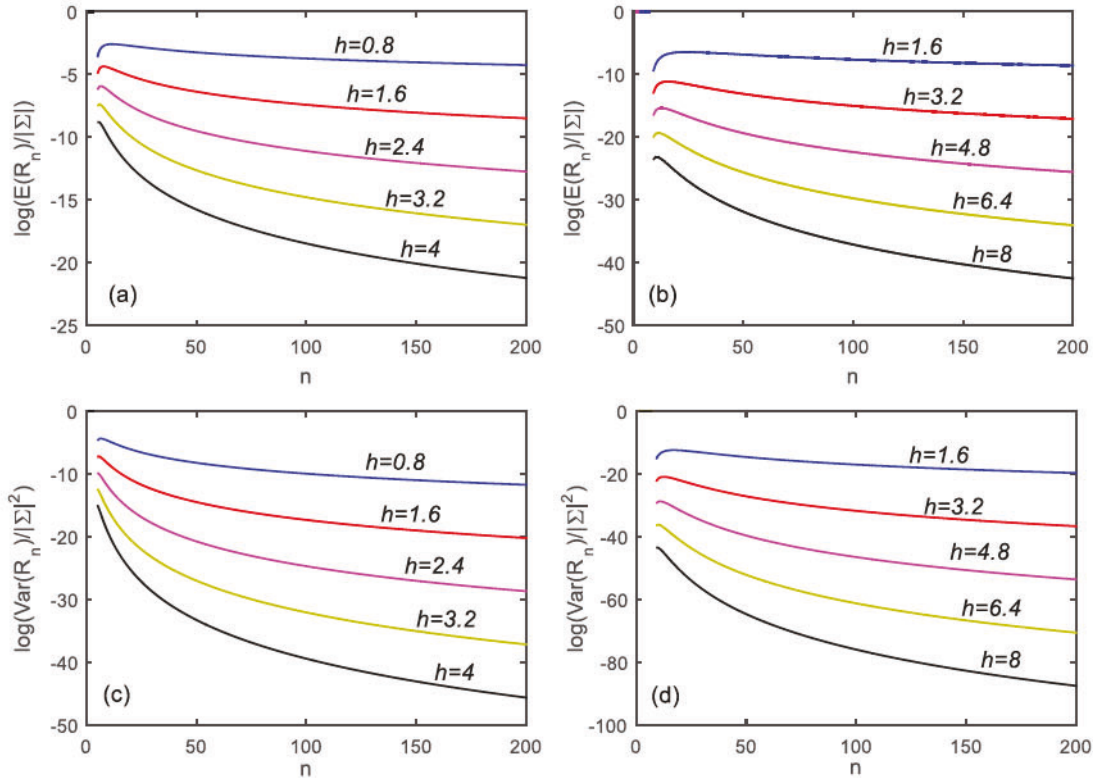


Figure 1. The monotonicity of $E(R_n)$ and $\text{var}(R_n)$ for steady state observations with different p and h : (a) $E(R_n)$ for $p = 4$; (b) $E(R_n)$ for $p = 8$; (c) $\text{var}(R_n)$ for $p = 4$; and (d) $\text{var}(R_n)$ for $p = 8$.

2. $E(R_{k+1:n})$ starts to increase once $b_k^T \Sigma^{-1} b_k > h$ as we decrease k . The optimal truncation point can be approximated by

$$t^* \approx \max_k \{\arg(b_k^T \Sigma^{-1} b_k > h)\}. \quad (25)$$

The proof is provided in Supplementary Section 5. From Theorem 5, we can see that $|S_{k+1:n}|$ is actually an asymptotically unbiased estimator of the generalized variance $|\Sigma|$, even if there exist initialization biases and autocorrelations. It is worth noting that although VAR(1) is assumed in Theorem 5, the results can be applicable to the general cases, as long as (i) the autocorrelations decay to 0 exponentially or even faster as the lag increases, and (ii) the mean of the bias asymptotically approaches zero, that is, $\lim_{n \rightarrow \infty} \frac{\sum_{i=k+1}^n b_i}{n-k} = 0$. This generalization can be easily verified based on the proof in Supplementary Section 5. These two conditions are often satisfied for almost all the autoregressive processes, and Theorems 1–4 are actually special cases of this generalization.

It should be pointed out that the proposed AMCR rule may also be able to handle the initial-transient covariance matrix problem, though such case has not been considered in almost all the existing literature. Based on the definition of the AMCR statistic, it is intuitive that for multivariate output of fixed mean, if the determinant of covariance matrix in the transient period is much larger than the steady-state period, then adding transient observations will increase the expectation of the R statistic. However, this problem is extremely complicated because of the intractable expectation. Here, we roughly derive the following result through approximation for the purpose of completeness.

Proposition 1. Suppose the p -dimensional vector $X_1, X_2, \dots, X_c$ follow independent normal distributions $N(\mu, \Sigma_1)$ and $X_{c+1}, \dots, X_{c+n}$ follow independent normal distributions $N(\mu, \Sigma_0)$. Then $E(R_{1:n+c}) > E(R_{c+1:n+c})$ if

$$\frac{|\Sigma_1|}{|\Sigma_0|} > \frac{\left(1 + \frac{c}{n}\right)^h - \left(\frac{n}{n+c-1}\right)^p}{\left(\frac{c}{n+c-1}\right)^p} \frac{\prod_{i=1}^p (n-i)}{(n-1)^p}. \quad (26)$$

The derivation of the above result is provided in Supplementary Section 6. Note that the right-hand side of Equation (26) is only an upper bound for the lower end of the increasing interval. This upper bound increases as h increases. It should also be noted that if both mean and covariance matrix have initialization bias, it would be easier for the AMCR rule to find an accurate truncation point than with only one single type of bias. In the following section, we will present the numerical studies for illustration, performance evaluation, and real application.

4. Numerical Studies

4.1. Illustration of Theoretical Properties

In this section, we numerically illustrate the theoretical properties provided in Section 3. Since the monotonicity of $E(R_n)$ or $\text{var}(R_n)$ is independent of $|\Sigma|$, we use $\log\left(\frac{E(R_n)}{|\Sigma|}\right)$ and $\log\left(\frac{\text{var}(R_n)}{|\Sigma|^2}\right)$ instead for illustration. Figure 1 shows the monotonicity of these two functions with different p and h . Clearly, both $E(R_n)$ and $\text{var}(R_n)$ are monotonically decreasing when n is larger than a small constant. As we enlarge the tuning

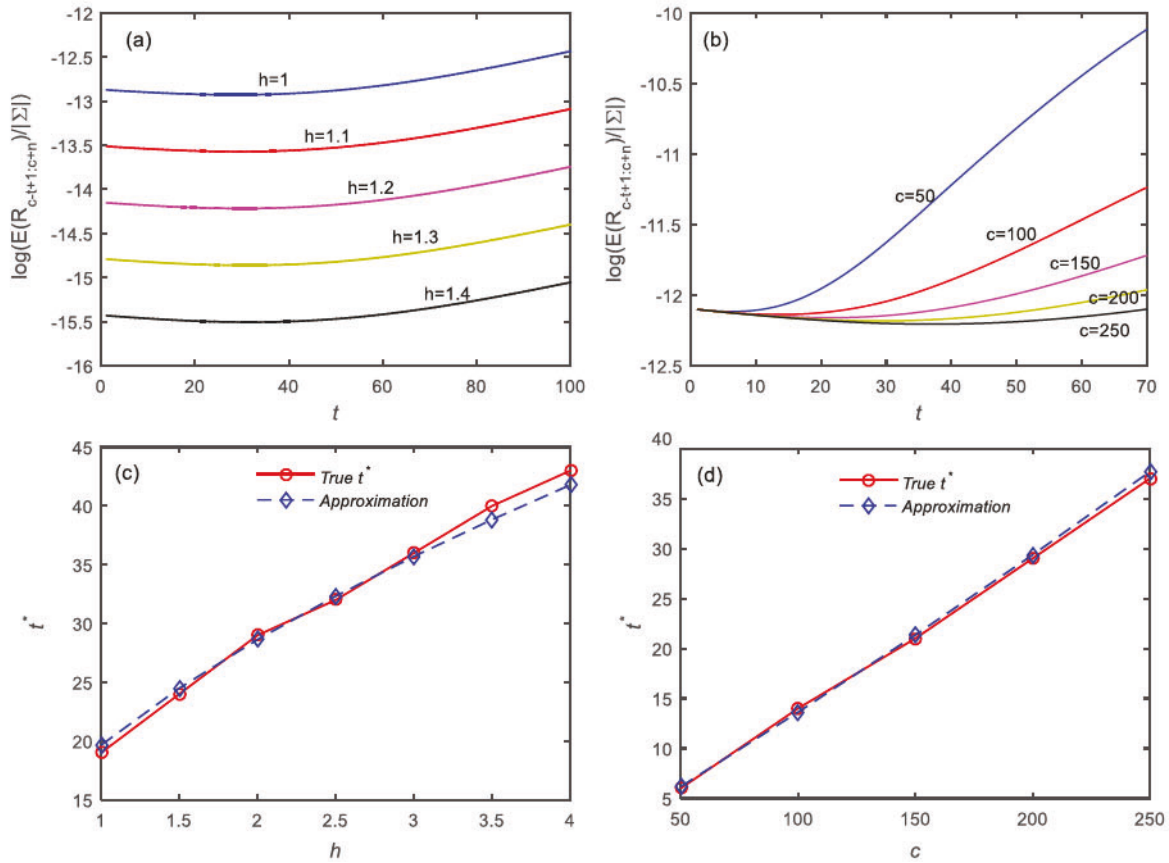


Figure 2. The behavior of $E(R_{c-t+1:c+n})$ with different h and c for observations containing initialization bias: (a) $E(R_{c-t+1:c+n})$ for $c = 200$; (b) $E(R_{c-t+1:c+n})$ for $h = 2$; (c) the increasing point t^* for $c = 200$; and (d) the increasing point t^* for $h = 2$.

parameter h , the decreasing rate is increased, and the lower end of the decreasing interval is reduced. All the results are consistent with what we get from Theorem 1.

To illustrate the behavior of $E(R)$ for observations with linear initialization bias, we set $p = 10$, $n = 400$, $\mu_0 = (1, 1, \dots, 1)^T$, $\mu_b = (0, 0, \dots, 0)^T$, and $\Sigma = 0.1I_p$ where I_p is the p -dimensional identity matrix. The shifting rate for the mean can be calculated as $r_s = 10/c$. Figure 2 shows the behavior of $E(R_{c-t+1:c+n})$ with different c and h . From Figures 2(a) and (b), we see that $E(R_{c-t+1:c+n})$ continues to decrease at the beginning, and then monotonically increases as more and more biased observations are added. Figures 2(c) and (d) show the true and approximated critical point (or delay time) t^* using Equation (21). As we can see, the approximated t^* is very close to the true values. Besides, t^* increases with both c and h , which is consistent with Equations (21) and (22) that t^* is approximately linear with $\sqrt{h}$ and c .

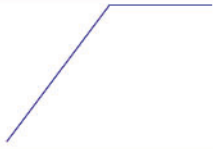
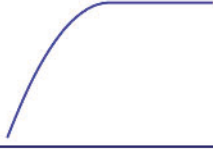
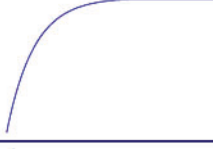
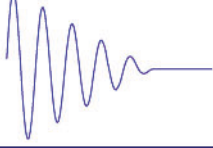
4.2. Illustration of the AMCR Rule

To illustrate the AMCR rule and evaluate its performance, four types of mean bias functions, that is, linear, quadratic, exponential, and oscillating functions are used, as shown in Table 1. These bias functions consist of an initial transient period and steady-state period, and have been widely used to test the initialization bias truncation algorithms in discrete-event simulations (Cash et al. 1992; White Jr, Cobb, and Spratt 2000; Hoad, Robinson, and Davies 2010; Hou, Wu, and Chen 2016; Wu et al. 2016; Wu, Chen, and Zhou 2016).

To illustrate how the algorithm works, four-dimensional ($p = 4$) signals are simulated with each bias function simulating one dimension. The parameters in Table 1 are set to $H = 1$, $N = 600$, $f = 30$. Two initialization bias changing rates are considered, $c_0 = 200$ and $c_0 = 400$. The signal noise is set to $\Sigma = 0.01I_4$. The tuning parameter for the AMCR rule is set to $h = 1$. Figure 3 shows the initialization bias truncation process for the two randomly generated four-dimensional output. As we can see, if we decrease i , the statistic $R_{i:N}$ decreases first for steady state observations, and then increases when biased observations are added. Note that due to signal noises, the statistic $R_{i:N}$ is not monotonically decreasing or increasing. The optimal truncation points are $c^* = 192$ and $c^* = 374$, respectively, which are very close to the true values.

We also illustrate the truncation process with only covariance matrix bias, though such scenario is not the major focus of this article. Two cases are considered, as shown in Figure 4. In the first case, the signal parameters are set to $p = 2$, $\mu = 0$, $c_0 = 300$, $N = 600$, $\Sigma_1 = 0.01I_2$, $\Sigma_0 = 0.01 \times \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$. In this case, the determinant ratio is $|\frac{\Sigma_1}{\Sigma_0}| \approx 1.6$. For the second case, the parameters are set to $p = 2$, $\mu_0, c_0 = 300$, $N = 600$, $\Sigma_{1i} = 30^{(c_0-i+1)/c_0} \times 0.01I_2$ for $i = 1, \dots, c_0$, and $\Sigma_0 = 0.01I_2$. In this case, the noise variance exponentially decreases before c_0 and then becomes steady. Clearly, the AMCR rule can effectively truncate the biased observations. Note that for the first case with correlation change, all the univariate truncation methods will not work at all.

Table 1. Four mean bias functions and their shapes.

Bias type	Functional form	Shape
Linear	$y(i) = \begin{cases} \frac{i-1}{c_0}H, & i = 1, \dots, c_0 \\ H, & i = c_0 + 1, \dots, N \end{cases}$	
Quadratic	$y(i) = \begin{cases} H \left[1 - \frac{(i-c_0-1)^2}{c_0^2} \right], & i = 1, \dots, c_0 \\ H, & i = c_0 + 1, \dots, N \end{cases}$	
Exponential	$y(i) = \begin{cases} H \left[1 - 10^{\frac{1-i}{c_0}} \right], & i = 1, \dots, c_0 \\ 0.9H, & i = c_0 + 1, \dots, N \end{cases}$	
Oscillating	$y(i) = \begin{cases} H \frac{c_0-i+1}{c_0} \sin\left(\frac{\pi i}{T}\right), & i = 1, \dots, c_0 \\ 0, & i = c_0 + 1, \dots, N \end{cases}$	

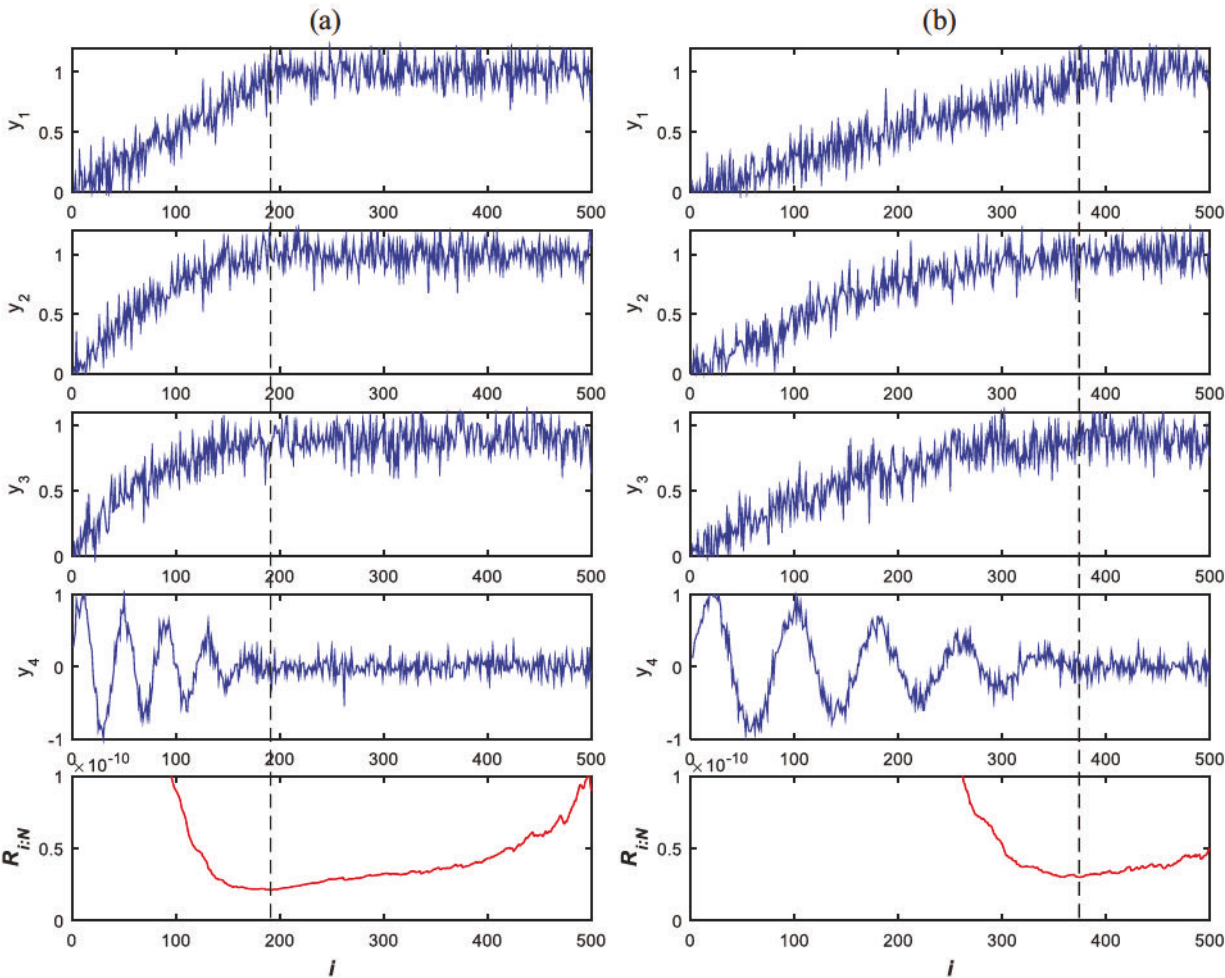


Figure 3. Illustration of initialization bias truncation through AMCR rule: (a) $c_0 = 200$, optimal truncation point $c^* = 192$; (b) $c_0 = 400$, optimal truncation point $c^* = 374$. The vertical dashed lines denote the optimal truncation points obtained by AMCR rule.

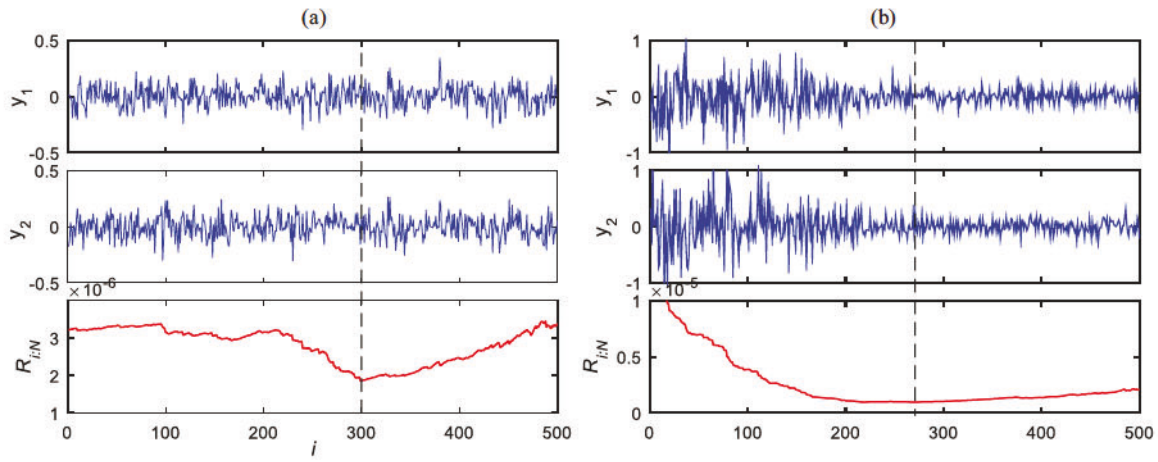


Figure 4. Illustration of initialization bias truncation for covariance matrix change: (a) $c_0 = 300$, $\Sigma_1 = 0.01I_2$, $\Sigma_0 = 0.01 \times \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$, optimal truncation point $c^* = 298$; (b) $c_0 = 300$, $\Sigma_1 = 30^{(c_0-i+1)/c_0} \times 0.01I_2$ for $i = 1, \dots, c_0$ and $\Sigma_0 = 0.01I_2$, optimal truncation point $c^* = 271$. The vertical dashed lines denote the optimal truncation points obtained by AMCR rule.

4.3. Elbow Method for Tuning Parameter Selection

In practical applications, the tuning parameter h needs to be selected appropriately. Section 3 tells us that the higher the h , the more rapidly the R will decrease as the length of steady state observations increases, which is preferred to reduce over-truncation. On the other hand, too large h will reduce the sensitivity of bias detection, which may result in under-truncation or insufficient truncation. Therefore, an appropriate h may play a decisive role on the truncation accuracy of certain signals.

One approach is that we roughly estimate the mean and noise covariance matrix using the steady state observations (e.g., second half of the sequence), and then generate a training database through simulation using various initialization bias functions, for example, the ones shown in Table 1, with different steady state times c_0 . The optimal h can be selected by minimizing the overall truncation error. This method may be effective in selecting an h that has an optimal overall performance, but may not be uniformly optimal for each multivariate signal. Besides, this method is time consuming. In this article, we propose a simpler and more effective method, named the elbow method for its similarity to the one widely used in cluster analysis (Thorndike 1953).

To illustrate the elbow method, we use simulated signals of dimension $p = 10$ where each dimension is simulated by a bias function randomly selected from Table 1. The other parameters are set to $H = 1$, $N = 600$, $\Sigma = \sigma^2 I_{10}$. Figure 5 shows the truncation point as a function of h with different c_0 (top row: 200; bottom row: 300) and noise σ (left column: 0.06; middle column: 0.1; right column: 0.14).

As we can see, all the curves decrease sharply first and then decrease slowly when h is increased from 0 to p . The reason is that for small h , $E(R_{i:N})$ decreases very slowly as we reduce i in the steady state period. Due to the existence of signal noise, the optimal truncation point by AMCR rule may be any point within the steady state period, and this point is very sensitive to the value of h . As h is increased, there will occur one or more consecutive sharp decreases and then c^* becomes very close to the true value. After that, the sensitivity of c^* to h is significantly

reduced due to the occurrence of initialization bias, which can be seen from Figure 5 that c^* is unchanged or decreases with a small value when h continues to increase. The higher the changing rate of the bias, the less the sensitivity of c^* to h . Therefore, the “elbow point” can be selected as the optimal h and the corresponding truncation point c^* . The “elbow point” can be selected visually, for example, the point of the last sharp decrease. In this article, it is automatically selected by fitting two connected line segments through total least squares, with the first line passing through the first point and the second line passing through the endpoint, and the connection point is on the curve. The “elbow point” is the connection point that minimizes the total sum of squared distances between the points and the fitted line. The identified elbow points with this method are shown in Figure 5. Clearly, this method is very effective and the identified c^* is very close to the true value c_0 .

4.4. Performance Evaluation and Comparison

The performance measures for initialization bias truncation often include the closeness of the estimated truncation point to the actual one, the percentage of bias removed by truncation, and the steady-state mean estimation accuracy (Hoad, Robinson, and Davies 2010). If the focus is to estimate the length of the transient period, then the closeness measure alone is adequate for truncation performance evaluation. While if the steady state performance is of interest, the estimation accuracy of performance parameters needs to be considered. Here we use both the detection closeness and parameter estimation accuracy to evaluate and compare the performance. Specifically, the root mean squared errors for closeness and the mean squared error for parameter accuracy are used and defined as

$$\begin{aligned} \text{RMSE}_c &= \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} (c_i^* - c_{i0})^2}, \\ \text{MSE}_\mu &= \frac{1}{N_s} \sum_{i=1}^{N_s} (\bar{X}_i - \mu_i)^T (\bar{X}_i - \mu_i), \end{aligned} \quad (27)$$

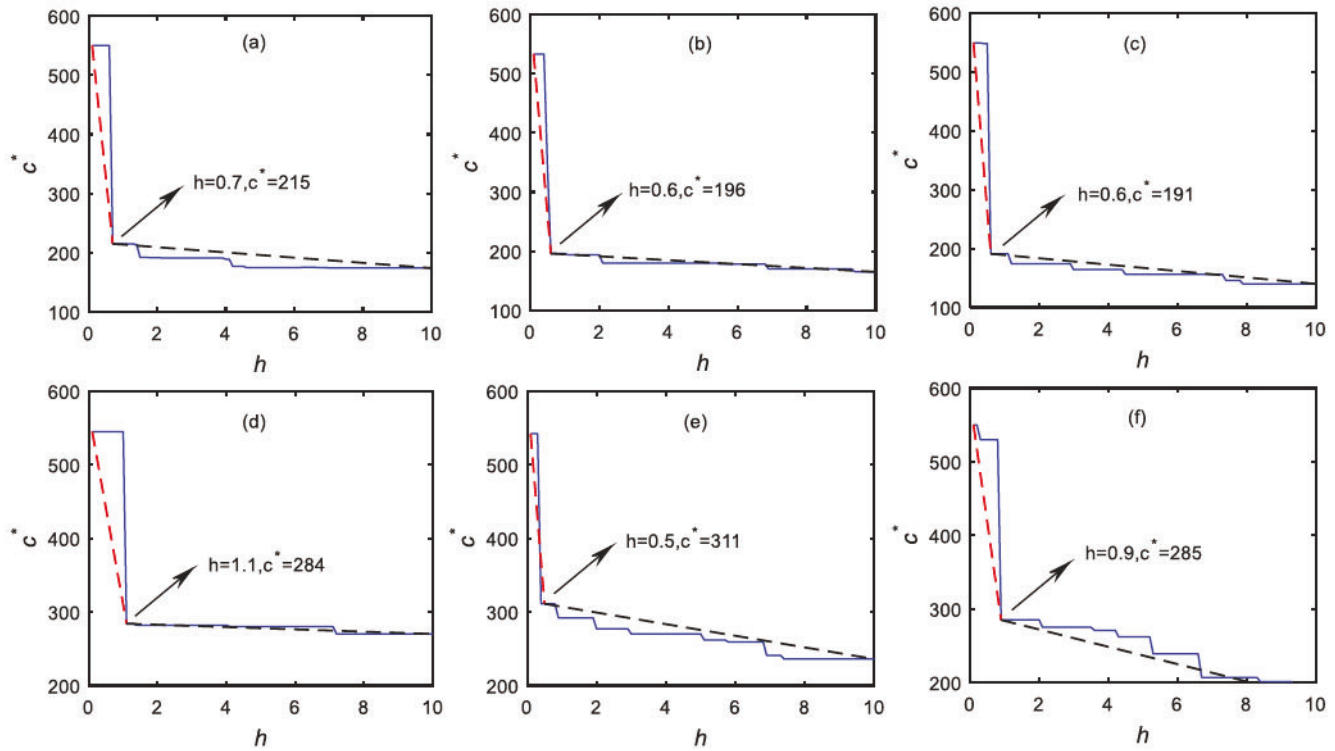


Figure 5. Tuning parameter selection through elbow method: (a) $c_0 = 200, \sigma = 0.06$; (b) $c_0 = 200, \sigma = 0.1$; (c) $c_0 = 200, \sigma = 0.14$; (d) $c_0 = 300, \sigma = 0.06$; (e) $c_0 = 300, \sigma = 0.1$; and (f) $c_0 = 300, \sigma = 0.14$. The intersection of two dashed lines is identified as the optimal h and corresponding truncation point c^* .

where N_s is the total number of multivariate output sequences, c_i^* and c_{i0} are the truncation point and true value, respectively, $\bar{X}_i$ and μ_i are the estimated and true steady state mean, respectively, for the i th multivariate output. To show how many signals are under-truncated, another metric called the under-truncation rate (UTR) is used as an ancillary metric, which is the proportion of signals under-truncated out of all multivariate signals.

In the performance evaluation and comparison, $p = 4$ and 10 are selected and each dimension is simulated using a bias function randomly selected from the four given in Table 1. The height of the signal is set to $H = 1$. To simulate different bias severity, two levels for c_0 which controls the changing rate of the initialization bias are selected: $c_0 = 200$ and 300. It is assumed that all dimensions reach steady state at the same time. The length of the signal is set to be $N = 600$. Five types of Gaussian noise are used to test the algorithm: (1) no autocorrelation and no correlation or dependence among variables, denoted by AR(0); (2) first-order autoregressive correlation and no correlation, denoted by AR(1); (3) second-order autoregressive correlation and no correlation, denoted by AR(2); (4) no autocorrelation but with correlation among variables, denoted by CR; and (5) first-order vector autoregressive model, denoted by VAR(1). The noises types and their parameters are listed in Table 2. Different noise amplitudes are also considered: for AR(0), CR, and VAR(1), $\sigma = 0.06, 0.1, 0.14$; for AR(1) and AR(2), $\sigma = 0.055, 0.089, 0.13$ and $\sigma = 0.045, 0.071, 0.106$, respectively, to match the noise variances of AR(0) and CR. For CR and VAR(1), all the off-diagonal elements (correlation coefficients) of C_r are set to 0.6. For VAR(1), A is randomly generated with modulus of eigenvalues less than 1. Each signal is replicated 100 times in the simulation.

The proposed AMCR rule is compared with the well-known MSER method (White Jr 1997; White Jr, Cobb, and Spratt 2000; Wu, Zhou, and Li 2013), Hotelling T^2 method (Shruben 1981) and batch means test (BMT) (Goldsman, Shruben, and Swain 1994; White Jr, Cobb, and Spratt 2000). For both AMCR and MSER methods, when the truncation point is close to the end of the time series, due to large sampling error, the truncation statistics may get very low values. Typically the truncation point of MSER is restricted to the first half of the time series (Pasupathy and Schmeiser 2010). Here we ignore the last 50 observations when finding the minimum of the truncation statistics for both AMCR and MSER. Since MSER and BMT are univariate methods, they are applied to each dimension of the multivariate signal and then the largest truncation point is selected as the optimal truncation point. The batching parameters for BMT are set at $b = 16$ and $b' = 8$, as recommended by Cash et al. (1992). For the T^2 method, the second half of the multivariate signal is used to construct the T^2 control chart. The first in-control observation from the beginning of the output with a certain significant level α is selected as the truncation point. In Shruben (1981), α is set to 0.05. However, as observed in our simulation, $\alpha = 0.05$ would result in a very large RMSE and a UTR of almost 100%. In fact, the T^2 based initialization bias truncation is just the reverse process of T^2 based statistical quality control. For statistical quality control, T^2 should be sufficiently large to reduce the alpha error. However, for initialization bias truncation, the T^2 statistic should be sufficiently close to 0 to reduce WSE and UTR. Therefore, we select $\alpha = 0.95$, which is very close to the optimal value in terms of overall RMSE_c. The other multivariate truncation method by Gallagher, Bauer, and Maybeck (1994) is not used for comparison since it is only

Table 2. Noise types and their parameters.

Type	Equation	Parameter
AR(0)	$\psi_t = \epsilon_t$	$\epsilon_t \sim N(0, \sigma^2 I_p)$
AR(1)	$\psi_t = \phi_1 \psi_{t-1} + \epsilon_t$	$\epsilon_t \sim N(0, \sigma^2 I_p), \phi_1 = 0.4$
AR(2)	$\psi_t = \phi_2 \psi_{t-1} + \phi_3 \psi_{t-2} + \epsilon_t$	$\epsilon_t \sim N(0, \sigma^2 I_p), \phi_2 = -0.25, \phi_3 = 0.5$
CR	$\psi_t = \epsilon_t$	$\epsilon_t \sim N(0, \sigma^2 C_r)$
VAR(1)	$\psi_t = A\psi_{t-1} + \epsilon_t$	$\psi_0 \sim N(0, \Sigma), \Sigma = \sigma^2 C_r, \epsilon_t \sim N(0, \Sigma_u), \Sigma = A\Sigma A^T + \Sigma_u$

Table 3. Detailed comparison of AMCR, MSER, T^2 , and BMT for the case of $p = 4$.

Noise	Signal		RMSE _c				MSE _μ				UTR			
	c ₀	σ _x	AMCR	MSER	T ²	BMT	AMCR	MSER	T ²	BMT	AMCR	MSER	T ²	BMT
AR(0)	200	0.06	12.0	18.9	22.6	25.2	0.0107	0.0109	0.0108	0.0108	0.84	0.99	0.58	0.85
		0.10	17.2	28.1	28.4	42.6	0.0118	0.0121	0.0120	0.0020	0.81	1.00	0.76	0.92
		0.14	21.4	36.3	37.1	37.2	0.0119	0.0122	0.0128	0.0125	0.82	1.00	0.82	0.98
	300	0.06	17.9	30.9	31.3	30.4	0.0095	0.0098	0.0096	0.0097	0.84	1.00	0.75	0.94
		0.10	24.6	41.8	43.8	42.1	0.0109	0.0113	0.0113	0.0112	0.85	0.99	0.83	0.93
		0.14	34.7	56.9	50.7	49.7	0.0112	0.0119	0.0119	0.0119	0.89	1.00	0.89	0.98
AR(1)	200	0.06	12.5	17.5	21.3	27.6	0.0107	0.0108	0.0108	0.0096	0.76	0.98	0.56	0.83
		0.10	18.3	27.1	27.4	36.2	0.0119	0.0121	0.0119	0.0103	0.90	0.99	0.70	0.94
		0.14	20.7	34.9	32.1	33.5	0.0119	0.0126	0.0120	0.0115	0.82	1.00	0.76	0.96
	300	0.06	16.5	27.0	33.0	28.8	0.0095	0.0097	0.0097	0.0097	0.82	0.99	0.71	0.86
		0.10	24.6	44.1	44.3	37.5	0.0109	0.0114	0.0114	0.0117	0.84	1.00	0.85	0.92
		0.14	30.6	54.5	55.4	49.4	0.0120	0.0116	0.0119	0.0127	0.88	1.00	0.97	0.97
AR(2)	200	0.06	12.2	17.2	23.8	19.2	0.0107	0.0108	0.0108	0.0104	0.65	0.98	0.52	0.85
		0.10	16.6	24.5	29.3	30.3	0.0118	0.0120	0.0119	0.0107	0.70	0.99	0.67	0.93
		0.14	20.8	37.9	36.2	44.3	0.0119	0.0122	0.0124	0.0127	0.84	1.00	0.74	0.92
	300	0.06	13.9	25.4	26.7	21.4	0.0095	0.0096	0.0095	0.0105	0.82	0.99	0.60	0.95
		0.10	21.1	37.3	37.7	32.7	0.0108	0.0111	0.0113	0.0094	0.77	1.00	0.82	0.94
		0.14	31.3	54.5	52.6	49.4	0.0110	0.0121	0.0118	0.0112	0.82	1.00	0.92	0.92
CR	200	0.06	17.6	18.5	23.0	27.1	0.0107	0.0108	0.0107	0.0129	0.70	0.99	0.63	0.87
		0.10	19.0	27.6	28.5	28.1	0.0118	0.0120	0.0119	0.0110	0.74	1.00	0.75	0.94
		0.14	20.2	37.5	38.1	37.4	0.0120	0.0128	0.0122	0.0112	0.80	1.00	0.83	0.94
	300	0.06	18.3	27.8	28.8	32.2	0.0095	0.0097	0.0097	0.0103	0.76	0.98	0.70	0.87
		0.10	25.7	41.3	49.0	47.2	0.0109	0.0113	0.0118	0.0104	0.85	1.00	0.88	0.93
		0.14	27.9	51.8	46.6	51.4	0.0120	0.0115	0.0125	0.0124	0.86	1.00	0.85	0.97
VAR(1)	200	0.06	18.4	21.8	34.3	30.8	0.0121	0.0119	0.0122	0.0122	0.81	0.95	0.48	0.90
		0.10	19.5	29.5	35.8	35.9	0.0292	0.0285	0.0290	0.0284	0.68	1.00	0.69	0.95
		0.14	25.4	38.9	36.0	41.6	0.0594	0.0596	0.0601	0.0594	0.73	1.00	0.73	0.97
	300	0.06	21.1	29.6	30.4	33.3	0.0111	0.0118	0.0122	0.0119	0.72	1.00	0.59	0.98
		0.10	25.3	42.0	39.5	50.8	0.0283	0.0291	0.0296	0.0284	0.76	0.99	0.84	0.97
		0.14	35.0	58.5	55.9	64.6	0.0522	0.0534	0.0541	0.0524	0.84	1.00	0.84	0.97
Overall			22.2	36.6	37.3	38.6	0.0153	0.0156	0.0157	0.0153	0.80	0.99	0.74	0.93

applicable to certain autoregressive data and is unnecessarily complex.

The detailed truncation results and comparison in terms of RMSE_c, MSE_μ, and UTR are shown in Tables 3 and 4 for $p = 4$ and $p = 10$, respectively. The main findings and discussions are listed as follows:

1. The proposed AMCR outperforms MSER, T^2 , and BMT significantly for all cases in terms of RMSE_c. MSER is slightly better than T^2 and BMT in terms of overall RMSE_c. For the mean estimate accuracy MSE_μ, the AMCR is also better than the other three methods. However, the advantage is not significant. The reason is that the biases shortly before the steady state period are so small that including these biased observations may not influence the mean estimation much. Among these four methods, MSER has the largest UTR, which is consistent with White et al.'s finding that

MSER often fails to truncate all biases (White Jr, Cobb, and Spratt 2000). Compared with MSER, the proposed AMCR has much lower UTR, though slightly higher than T^2 .

2. Increasing c_0 or noise σ will increase RMSE_c, MSE_μ, and UTR for all methods. The reason is that the higher noise or smaller mean shifting rate, the more initialization bias will be immersed into signal noise and become undetectable. The performances are almost the same under different noise types for all four methods. Although these methods are designed under the assumption of Gaussian noise without autocorrelation, they are still applicable for autoregressive noises.
3. Increasing the dimension from $p = 4$ to $p = 10$, the performances (RMSE_c, MSE_μ/p, UTR) of all methods are improved. It is a key advantage of multivariate truncation method over univariate method. For AMCR, increasing p

Table 4. Detailed comparison of AMCR, MSER, T^2 , and BMT for the case of $p = 10$.

Signal			RMSE _c				MSE _μ				UTR			
Noise	c ₀	σ _x	AMCR	MSER	T ²	BMT	AMCR	MSER	T ²	BMT	AMCR	MSER	T ²	BMT
AR(0)	200	0.06	8.5	10.9	20.5	27.9	0.0259	0.0260	0.0259	0.0259	0.74	1.00	0.44	0.73
		0.10	10.7	18.8	23.0	36.8	0.0276	0.0279	0.0277	0.0277	0.64	1.00	0.49	0.78
		0.14	11.1	26.0	22.0	25.8	0.0283	0.0282	0.028	0.0286	0.76	1.00	0.67	0.87
	300	0.06	13.9	17.8	20.2	21.0	0.0246	0.0247	0.0245	0.0246	0.74	1.00	0.50	0.82
		0.10	20.9	28.6	27.9	26.9	0.0261	0.0264	0.0261	0.0261	0.79	0.99	0.70	0.83
		0.14	27.4	41.5	34.8	30.2	0.0262	0.0272	0.0264	0.0274	0.88	1.00	0.79	0.93
AR(1)	200	0.06	9.5	11.0	19.4	23.8	0.0259	0.0260	0.0259	0.0249	0.69	0.96	0.44	0.67
		0.10	14.0	18.5	26.3	29.1	0.0277	0.0278	0.0276	0.0269	0.76	0.98	0.50	0.83
		0.14	17.6	26.6	22.5	24.3	0.0279	0.0280	0.0282	0.0280	0.69	1.00	0.65	0.88
	300	0.06	13.3	18.3	28.3	27.1	0.0246	0.0247	0.0245	0.0256	0.73	0.99	0.57	0.66
		0.10	18.8	30.0	25.1	23.8	0.0261	0.0264	0.0262	0.0286	0.76	1.00	0.75	0.84
		0.14	24.1	40.2	31.1	29.8	0.0275	0.0269	0.0272	0.0292	0.80	1.00	0.84	0.95
AR(2)	200	0.06	8.7	9.6	25.5	20.5	0.0259	0.0260	0.0259	0.0222	0.68	0.96	0.48	0.65
		0.10	13.5	17.0	24.3	26.2	0.0277	0.0278	0.0276	0.0231	0.73	0.99	0.50	0.84
		0.14	17.6	28.6	28.5	38.5	0.0281	0.0280	0.0279	0.0278	0.83	1.00	0.65	0.88
	300	0.06	10.7	16.1	23.3	21.9	0.0246	0.0246	0.0246	0.0263	0.77	0.96	0.51	0.82
		0.10	19.7	28.7	23.3	24.5	0.026	0.0264	0.0261	0.0254	0.78	1.00	0.74	0.83
		0.14	24.0	42.2	35.8	36.8	0.0265	0.0267	0.0271	0.0261	0.84	1.00	0.76	0.86
CR	200	0.06	9.8	12.6	23.3	22.3	0.0259	0.0260	0.0259	0.0290	0.67	0.97	0.41	0.78
		0.10	12.4	18.8	22.3	21.5	0.0277	0.0278	0.0277	0.0275	0.75	1.00	0.62	0.8
		0.14	14.7	27.9	22.8	22.2	0.0280	0.0281	0.0281	0.0281	0.78	1.00	0.70	0.88
	300	0.06	14.2	18.8	23.8	23.8	0.0246	0.0247	0.0245	0.0253	0.81	0.98	0.49	0.77
		0.10	18.2	29.9	31.5	30.8	0.0260	0.0264	0.0264	0.0255	0.84	1.00	0.66	0.89
		0.14	25.5	43.2	37.5	33.3	0.0262	0.0269	0.0268	0.0265	0.82	1.00	0.82	0.92
VAR(1)	200	0.06	8.2	12.9	31.1	18.1	0.0307	0.0304	0.0308	0.0304	0.75	0.92	0.34	0.90
		0.10	13.3	19.9	27.4	27.2	0.0732	0.0722	0.0736	0.0726	0.77	0.99	0.47	0.83
		0.14	15.9	28.4	24.7	29.5	0.1454	0.1456	0.1481	0.1461	0.79	1.00	0.57	0.93
	300	0.06	12.7	20.1	20.2	25.7	0.0279	0.0293	0.0300	0.0295	0.80	0.97	0.51	0.84
		0.10	20.1	30.9	29.0	33.5	0.0754	0.0759	0.0778	0.0767	0.69	1.00	0.66	0.88
		0.14	27.2	44.2	35.7	42.6	0.1446	0.1450	0.1474	0.1463	0.82	1.00	0.81	0.88
Overall			16.8	26.6	26.8	28.1	0.0378	0.0379	0.0382	0.0379	0.76	0.99	0.60	0.83

will increase the mean shifting rate r_s as defined in Equation (22) in the transient period, and thus will result in more accurate estimation. For T^2 method, increasing p leads to an increased truncation closeness and also reduced UTR. It can be explained by the fact that for T^2 control charts, if there are comparable mean shifts in all dimensions, the detection power ($1 - \beta$ where β is the beta error) of the full T^2 control will always be higher than the one with reduced dimensionality, that is, using partial dimensions. For MSER with multiple univariate truncation, the truncation point is getting closer and closer to the true value from the left as the dimension is increased, while the UTR is almost unchanged (near 100%). It can be explained as follows: suppose the truncation for i th dimension is c_i^* , $i = 1, \dots, p$, then $c^*(p) = \max(c_1^*, \dots, c_p^*)$. Since $c^* < c_0$ for almost all the cases, $\max(c_1^*, \dots, c_p^*) \leq \max(c_1^*, \dots, c_{p+1}^*) \leq c_0$. Therefore, c^* will get closer to c_0 from the left as p increases. BMT also has similar behavior as MSER.

4.5. Application to Transient Analysis of Serial Production Lines With Perishable Products

The AMCR rule is applied to the transient analysis of serial production lines with perishable products (Ju, Li, and Horst 2017).

In many production systems, the products, for example, yogurt, have a fixed maximum allowable waiting time, exceeding which the item will be scrapped due to quality deterioration. These products are referred to as perishable products. Due to dynamic changes and frequent disruptions in the manufacturing process, the production system often operates partially or even entirely in the transient regime. The traditional steady state performance measures ignore the substantial amount of production loss due to the transient behavior, and thus are insufficient to capture the system performance. Therefore, transient analysis of production lines with perishable products is critically important.

In the transient analysis, one critical characteristic is the settling time, or the length of the transient period. Accurate estimation of the settling time is essential in performance evaluation, real-time production optimization and control. Suppose there is a serial production line with two Bernoulli reliability machines (m_1 and m_2), a finite buffer B_1 and perishable products, as shown in Figure 6. Both machines have constant and identical processing time, which is slotted with cycle time. The two machines follow the Bernoulli reliability model. In each cycle, the machine m_i , $i = 1, 2$ is capable of producing a part with probability p_i and fails to do so with probability $1 - p_i$. The buffer B_1 has a finite capacity of N_b with first-in-first-out outflow process. At the beginning of operation, there are n_0 parts in the buffer with waiting time of the first one being τ_0 . The

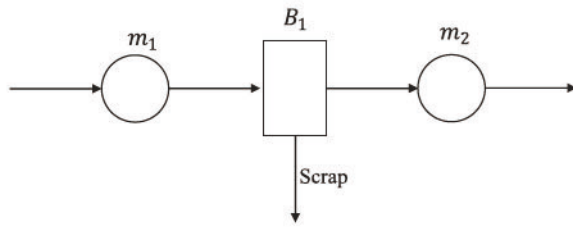


Figure 6. Bernoulli line with perishable products.

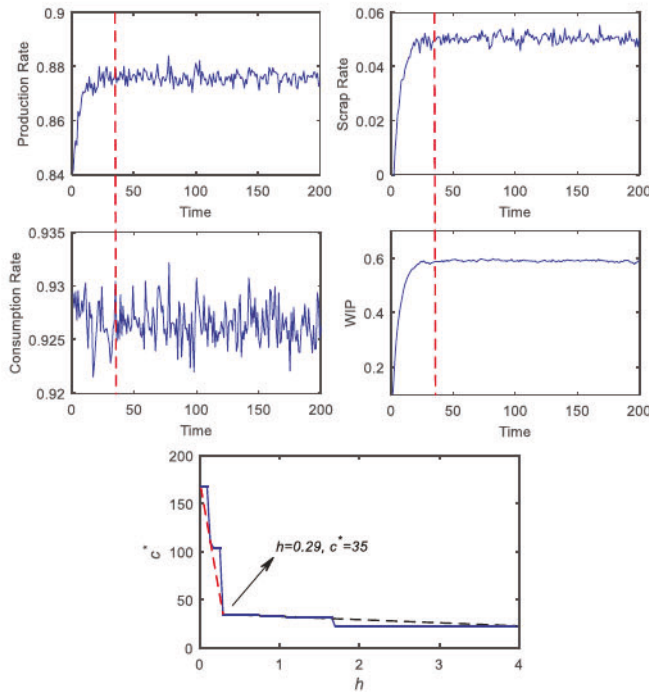


Figure 7. Simulation output and initialization bias truncation for $p_1 = 0.9256$, $p_2 = 0.9070$, $N_b = 2$, $T_{\max} = 2$, $n_0 = 0$, $\tau_0 = 0$. The vertical dashed lines denote the truncation point $c^* = 35$.

maximum allowable waiting cycle in buffer B_1 is $T_{\max}$. When a part's waiting time exceeds $T_{\max}$, it is scrapped from the buffer. For other details, please refer to (Ju, Li, and Horst 2017).

The performance measures that are of particular interests are: (1) production rate $PR(t)$ —the expected number of parts produced by machine m_2 in the t th cycle; (2) consumption rate $CR(t)$ —the expected number of parts consumed by machine m_1 in the t th cycle; (3) scrap rate $SR(t)$ —the expected number of scrapped parts in the t th cycle; and (4) work-in-process $WIP(t)$ —the expected number of parts in buffer B_1 at the end of the t th cycle.

Figure 7 shows the four performance measures through discrete-event simulation with $p_1 = 0.9256$, $p_2 = 0.9070$, $N_b = 2$, $T_{\max} = 2$, $n_0 = 0$, $\tau_0 = 0$. The initialization bias truncation through AMCR rule is also provided in the bottom row, where

the identified truncation point is 35. To compare AMCR with the other three methods, another two sets of simulation parameters used in Ju, Li, and Horst (2017) are also included to generate the data. The three sets of simulation parameters and the corresponding steady state performance measures are given in Table 5. The AMCR identified truncation times are 35, 110, and 140, respectively. As the true steady state transition times are unknown, we use the MSE_{μ} as the evaluation metric. The MSE_{μ} for AMCR, MSER, T^2 , and BMT are 0.004309, 0.004317, 0.004312, and 0.00431, respectively. Therefore, the proposed AMCR has higher estimation accuracy than the other three methods.

5. Conclusion

In this article, an efficient method named AMCR has been developed for initialization bias truncation of multivariate output in discrete-event simulations. The basic idea of this method is to determine the optimal truncation point that minimizes the confidence volume of the mean estimate using the observations after truncation. To adjust the sensitivity of bias truncation and enhance the robustness, a tuning parameter is innovatively incorporated, which can be adaptively selected through elbow method. Several theoretical properties in terms of the expected behavior of the truncation statistic on both autocorrelated data and data without autocorrelation have been derived, which provide theoretical foundations and insightful guidance for understanding and methodology designing in practical applications. The performance of the proposed method has been thoroughly evaluated through intensive simulation and real case studies. The evaluation and comparison results indicate that the proposed method can accurately estimate the truncation point for various multivariate output of different bias shapes and severity, and is superior to the existing methods in terms of truncation accuracy.

Along this research direction, there are several issues worthy of further investigation. First, the bias truncation accuracy is significantly influenced by the measurement noise levels. As the noise increases, small initialization bias before the steady state period may not be detectable. To further increase the sensitivity of the proposed method to small bias, a moving average scheme, such as EWMA could be incorporated. Secondly, we only considered the confidence region of the mean parameter in the current work. In fact, we can also take into account the confidence region of the covariance matrix parameter, or the joint confidence region of both the mean and the covariance matrix parameters. Last but not least, in some applications, not all the dimensions of the output reach the steady state at the same time. Similar to the sparse mean-shift issue in SPC charts, using the joint confidence region of all the means may not be the optimal choice.

Table 5. Simulation parameters and steady state performance measures.

Setting	Simulation parameters						Steady state performance			
	p_1	p_2	N_b	$T_{\max}$	n_0	τ_0	$PR(\infty)$	$SR(\infty)$	$CR(\infty)$	$WIP(\infty)$
1	0.9256	0.9070	2	2	0	0	0.9102	0.093	0.9324	0.6102
2	0.8305	0.7623	6	6	0	0	0.7631	0.071	0.8219	0.3806
3	0.8554	0.7273	5	6	1	1	0.7318	0.0801	0.8321	0.3997

Supplementary Materials

The proofs for Theorems 1–5 and Proposition 1 are provided in online supplementary materials.

Acknowledgments

The authors thank the editor, associate editor, and two anonymous reviewers for their constructive comments and suggestions that have considerably improved the article.

Funding

Jianguo Wu was partially supported by National Natural Science Foundation of China grant NSFC-51875003 and key program NSFC-71932006. Feng Ju was partially supported by the National Science Foundation grant NSF CMMI-1829238.

ORCID

Jianguo Wu  <https://orcid.org/0000-0002-2885-8725>

References

- Cash, C. R., Nelson, B. L., Dippold, D. G., Long, J. M., and Pollard, W. P. (1992), "Evaluation of Tests for Initial-Condition Bias," in *Proceedings of the 24th Conference on Winter Simulation*. [505,508]
- Gallagher, M. A., Bauer, K. W., and Maybeck, P. S. (1994), "Initial Data Truncation for Multivariate Output of Discrete-Event Simulations Using the Kalman Filter," *Annals of Operations Research*, 53, 419–441. [500,508]
- Goldsmann, D., Schruben, L. W., and Swain, J. J. (1994), "Tests for Transient Means in Simulated Time Series," *Naval Research Logistics*, 41, 171–187. [499,500,508]
- Hoad, K., Robinson, S., and Davies, R. (2010), "Automating Warm-Up Length Estimation," *Journal of the Operational Research Society*, 61, 1389–1403. [505,507]
- Hotelling, H. (1931), "The Generalization of Student's Ratio," *The Annals of Mathematical Statistics*, 2, 360–378. [501]
- Hou, Y., Wu, J., and Chen, Y. (2016), "Online Steady State Detection Based on Rao–Blackwellized Sequential Monte Carlo," *Quality and Reliability Engineering International*, 32, 2667–2683. [505]
- Jackway, P., and Desilva, B. M. (1992), "A Methodology for Initialisation Bias Reduction in Computer Simulation Output," *Asia-Pacific Journal of Operational Research*, 9, 87–100. [499]
- Ju, F., Li, J., and Horst, J. A. (2017), "Transient Analysis of Serial Production Lines With Perishable Products: Bernoulli Reliability Model," *IEEE Transactions on Automatic Control*, 62, 694–707. [499,510,511]
- Kang, N., Ju, F., and Zheng, L. (2017), "Transient Analysis of Geometric Serial Lines With Perishable Intermediate Products," *IEEE Robotics and Automation Letters*, 2, 149–156. [500]
- Linton, J. R., and Harmonosky, C. M. (2002), "A Comparison of Selective Initialization Bias Elimination Methods," in *Proceedings of the 2002 Winter Simulation Conference: 1951–1957*. [499]
- Montgomery, D. C. (2009), *Introduction to Statistical Quality Control*, New York: Wiley. [500,502]
- Muirhead, R. J. (2009), *Aspects of Multivariate Statistical Theory* (Vol. 197), New York: Wiley. [502]
- Pasupathy, R., and Schmeiser, B. (2010), "The Initial Transient in Steady-State Point Estimation: Contexts, a Bibliography, the MSE Criterion, and the MSER Statistic," in *Winter Simulation Conference*. [508]
- Pawlikowski, K. (1990), "Steady-State Simulation of Queueing Processes: Survey of Problems and Solutions," *ACM Computing Surveys*, 22, 123–170. [499]
- Robinson, S. (2007), "A Statistical Process Control Approach to Selecting a Warm-Up Period for a Discrete-Event Simulation," *European Journal of Operational Research*, 176, 332–346. [499,500]
- (2014), *Simulation: The Practice of Model Development and Use*, London: Palgrave Macmillan. [499]
- Rossetti, M. D., Li, Z., and Qu, P. (2005), "Exploring Exponentially Weighted Moving Average Control Charts to Determine the Warm-Up Period," in *Proceedings of the 37th Conference on Winter Simulation*. [499]
- Runger, G. C., and Willemain, T. R. (1996), "Batch-Means Control Charts for Autocorrelated Data," *IIE Transactions*, 28, 483–487. [500]
- Schruben, L. W. (1982), "Detecting Initialization Bias in Simulation Output," *Operations Research*, 30, 569–590. [499]
- Schruben, L. W. (1981), "Control of Initialization Bias in Multivariate Simulation Response," *Communications of the ACM*, 24, 246–252. [500,508]
- Thorndike, R. L. (1953), "Who Belongs in the Family?," *Psychometrika*, 18, 267–276. [507]
- White, K. P., Jr. (1997), "An Effective Truncation Heuristic for Bias Reduction in Simulation Output," *Simulation*, 69, 323–334. [499,508]
- White, K. P., Jr., Cobb, M. J., and Spratt, S. C. (2000), "A Comparison of Five Steady-State Truncation Heuristics for Simulation," in *Proceedings of the 32nd Conference on Winter Simulation*. [499,505,508,509]
- Wilson, A. J. (2010), "Volume-of-n-Dimensional-Ellipsoid," *Scientia Acta Xaveriana*, 1, 101–106. [501]
- Withers, C. S. (1985), "The Moments of the Multivariate Normal," *Bulletin of the Australian Mathematical Society*, 32, 103–107. [503]
- Wu, J., Chen, Y., and Zhou, S. (2016), "Online Detection of Steady-State Operation Using a Multiple-Change-Point Model and Exact Bayesian Inference," *IIE Transactions*, 48, 599–613. [505]
- Wu, J., Chen, Y., Zhou, S., and Li, X. (2016), "Online Steady-State Detection for Process Control Using Multiple Change-Point Models and Particle Filters," *IEEE Transactions on Automation Science and Engineering*, 13, 688–700. [505]
- Wu, J., Zhou, S., and Li, X. (2013), "Acoustic Emission Monitoring for Ultrasonic Cavitation Based Dispersion Process," *Journal of Manufacturing Science and Engineering*, 135, 031015. [499,508]