Adaptive Minimal Confidence Region Rule for Multivariate Initialization Bias Truncation in Discrete-event Simulations

Jianguo Wu*1, Honglun Xu2, Feng Ju3, Bill Tseng4

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 paper proposes an efficient method, called adaptive minimal 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.

Keywords: Minimal confidence region, Auto-correlation, Steady state, Generalized variance, Asymptotically unbiased estimator

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, start-up or warm-up period. Since the start-up

---

*Corresponding author, Industrial Engineering and Management, College of Engineering, Peking University, Beijing, China 100871, Email: j.wu@pku.edu.cn

1 Computational Science Program, University of Texas at El Paso (UTEP), Texas 79968, USA, Email: hxu3@miners.utep.edu

2 School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, Arizona 85281, USA, Email: Feng.Ju@asu.edu

3 Industrial, Manufacturing and Systems Engineering, UTEP, Texas 79968, USA, Email: btseng@utep.edu
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 et al. 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 (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 et al. 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, Wu et al. 2013, White Jr et al. 2000), 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 et al. 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 et al. 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, e.g., 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 is 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 Schruben (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 $\alpha$. 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 et al.
(Gallagher et al. 1994), who proposed a special state space model to capture the evolution of the mean output and measured output, and then applied the MMAE (multiple model adaptive estimation) 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 minimal confidence region rule (AMCR), is developed in this paper 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 paper.

The rest of this paper 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 Minimal 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 i.i.d. Gaussian assumption is often the basic assumption in many initialization bias truncation methods. To address the non-normality issue, many existing
truncation methods (Robinson 2007, Shruben 1981) propose to perform a series of independent replications. Based on the central limit theorem, the sample mean at each time interval tends towards normality. To handle the autocorrelation issue, the batch means method is commonly used (Goldsman et al. 1994, Runger and Willemain 1996, Robinson 2007, Shruben 1981). By dividing observations into batches, the batch means become approximately uncorrelated as the batch size increases. Since the non-normality and autocorrelation issues can be solved through these data preprocessing steps, we assume that the multivariate output to be truncated already follow i.i.d. 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 $X_1, X_2, ..., X_n$ follow i.i.d. normal distributions, i.e., $X_i \sim N(\mu, \Sigma)$ for $i = 1, ..., n$. Denote the mean of the output by $\bar{X}$, i.e., $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, then

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

(1)

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

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

(2)

where $\chi_{p, \alpha}^2$ is the upper $100\alpha$-th percentile of $\chi_p^2$ distribution. Clearly, the confidence region is a $p$-dimensional ellipsoid (ellipse for $p = 2$) with semi-axes $\sqrt{\frac{\chi_{p, \alpha}^2}{n}} \lambda_i, i = 1, 2, ..., p$ where $\lambda_i, i = 1, 2, ..., p$ are eigenvalues of the covariance matrix $\Sigma$ satisfying $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$. The volume of the confidence region can be calculated as (Wilson 2010)

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

(3)

When $\Sigma$ 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

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

(4)

Then
\[ T^2 = n(\bar{X} - \mu)^T S^{-1}(\bar{X} - \mu) \rightarrow T_{p,n-1}^2 = \frac{p(n-1)}{n-p} F_{p,n-p} \]  

(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

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

(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

\[ V(\alpha) \approx \frac{2\pi^\frac{p}{2}}{p\Gamma\left(\frac{p}{2}\right)} \left( p F_{p,n-p,\alpha} \right)^{p/2} |S|^{1/2} \]  

(7)

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

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

(8)

where \( S_{c+1:n} \) is the sample covariance matrix for output \( \{X_{c+1}, X_{c+2}, \ldots, X_n\} \). Note that \( F_{p,n-p,\alpha} \) is excluded in Eq. (8). The reason is that when \( n \) is large, according to the law of large numbers, \( F_{p,n-p} = \frac{X_p^2/p}{X_{n-p}^2/(n-p)} \rightarrow \chi_p^2/p \), which does not depend on \( n \). When \( p = 1 \), Eq. (8) is exactly the MSER for univariate output. As observed in the experiment, the truncation rule given in Eq. (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 \( h \):

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

(9)

Here \( h \) is a tuning parameter used to adjust the sensitivity of bias truncation, and it satisfies \( p \geq h > 0 \). Eq. (9) can be better interpreted using the following equivalent form with a log-transformation:
\[ c^* = \arg \min_{n \gg c \geq 0} (\log |S_{c+1:n}| - \lambda p \log (n - c)), \]  

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, \(\lambda\) controls the trade-off 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 minimal 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

\[ R_n = \frac{|S_{1:n}|}{n^h} \text{ and } R_{i:n} = \frac{|S_{i:n}|}{(n-i+1)^h} \]  

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 i.i.d. observations. Specifically, we will show that for a set of i.i.d. 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 \(X_1, X_2, \ldots, X_n\) follow i.i.d. normal distributions of dimension \(p\), i.e., \(X_i \sim N(\mu, \Sigma)\) for \(i = 1, \ldots, n\). 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 \(\lim_{n \to \infty} E(R_n) = 0\);

2. When \(n > g_2(p, h)\) where \(g_2(p, h)\) is a function of \(p\) and \(h\), \(Var(R_n)\) decreases
monotonically with \( n \) and \( \lim_{n \to \infty} \text{Var}(R_n) = 0 \).

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)

\[
E(|S_{1:n}|) = \prod_{i=1}^{p} \frac{(n-i)(n-1)^p}{(n-1)^p} |\Sigma|,
\]

\[
\text{Var}(|S_{1:n}|) = \frac{\prod_{i=1}^{p} (n-i)}{(n-1)^{2p}} \left[ \prod_{j=1}^{p} (n-j+2) - \prod_{j=1}^{p} (n-j) \right] |\Sigma|^2
\]

The detailed proof of Theorem 1 is provided in Section 1 of the Supplementary Material. 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 Eq. (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 \( (~O\left(\frac{1}{n^{1+2\alpha}}\right)) \) 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 \( c_0 \), i.e., \( c^* \leq c_0 \). 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, i.e., \( R_{t:n} \). 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 $X_1, X_2, \ldots, X_n$ follow i.i.d. normal distributions of dimension $p$, i.e., $X_i \sim N(\mu_0, \Sigma)$ for $i = 1, \ldots, n$, and $X_{n+1} \sim N(\mu, \Sigma)$ where $\mu \neq \mu_0$. Define the non-centrality matrix $T = \sum_{i=1}^{n+1} (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^T$ where $\mu_i = \mu_0$ for $i = 1, \ldots, n, \mu_{n+1} = \mu$, and $\bar{\mu} = \frac{\sum_{i=1}^{n+1} \mu_i}{n+1}$. Let $k_j (j = 1, \ldots, p)$ with $k_1 \geq k_2 \geq \cdots \geq k_p$ be the real and non-negative roots of the determinantal equation $|T - k^2 \Sigma| = 0$, then

1. $k_j = 0$ for $j = 2, \ldots, p$ and

$$k_1^2 = \frac{n}{n+1} (\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0)$$

(13)

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

(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 $X_1, X_2, \ldots, X_n$ follow i.i.d. normal distributions of dimension $p$, i.e., $X_i \sim N(\mu_0, \Sigma)$ for $i = 1, \ldots, n$. 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

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

(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 $\mu$ and $\mu_0$, which measures the severity of the initialization bias. For notational convenience, we define
\[ D_s^2 = (\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0) \tag{16} \]

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, e.g., \( D_s^2 > h \). 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, i.e., 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. Firstly, 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. Secondly, 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 $X_1, X_2, \ldots, X_c$ in the transient period follow independent normal distributions with $X_i \sim N(\mu_i, \Sigma)$ for $i = 1, \ldots, c$, where

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

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

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

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

where

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

2. 

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

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

$$
t^* \approx \frac{\sqrt{r_s} - 1}{1 - \sqrt{r_s}}
$$

where $r_s$ is the mean shifting rate defined as
From Eq. (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 Eq. (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, Eq. (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 $X = (x_1, ..., x_p) \sim N_p(\mu, \Sigma)$ where $\Sigma = (\sigma_{ij}^2)$, and suppose $i_1, ..., i_r$ are indices from $\{1,2, ..., p\}$ and they need not all be distinct, then

$$E(x_{i_1}x_{i_2} \cdots x_{i_r}) = \sum_{l+2k=r} \sum_{\mu_{i_1} \cdots \mu_{i_k} \sigma_{j_1 j_2}^2 \cdots \sigma_{j_{2k-1} j_{2k}}^2} \sum_{m=1}^{r!} \frac{1}{l!2^k k!} \text{ permutaitons } \{i_1 \cdots i_k j_1 j_2 \cdots j_{2k}\} \text{ of } \{i_1, \cdots, i_r\} \text{ giving distinct terms allowing for the symmetry of } \Sigma.$$  

(23)

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 $\mu$ is the steady state mean vector, all the eigenvalues $\alpha_1, ..., \alpha_p$ of the matrix $A$ have modulus less than 1 (stability condition), i.e., $|\alpha_i| < 1$, and $u_t$ is the i.i.d. noise terms with $u_t \sim N(0, \Sigma_u)$. Suppose $X_0 \sim N(\mu_0, \Sigma)$ and the covariance matrix is stable, i.e., $\Sigma = A\Sigma A^T + \Sigma_u$ 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

\begin{equation}
E|S^*_{k+1:n}| = |\Sigma| \left( 1 + \frac{1}{n-k} \sum_{i=k+1}^{n} b_i^T \Sigma^{-1} b_i \right) + O\left( \frac{1}{(n-k)^2} \right) \tag{24}
\end{equation}

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

(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

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

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, i.e., $\lim_{n \to \infty} \sum_{i=k+1}^{n} \frac{b_i}{n-k} = 0$. 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 Theorem 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 \( X_1, X_2, \ldots, X_c \) follow independent normal distributions \( N(\mu, \Sigma) \) and \( X_{c+1}, \ldots, X_{c+n} \) follow independent normal distributions \( N(\mu, \Sigma_0) \). Then \( E(R_{1:n+c}) > E(R_{c+1:n+c}) \) if

\[
\frac{|\Sigma_1|}{|\Sigma_0|} > \left( \frac{1 + \frac{c}{n}}{1} \right)^h \left( \frac{n}{n+c-1} \right)^p \frac{\prod_{i=1}^p (n-i)}{(n-1)^p} \left( \frac{c}{c-1} \right)^p \prod_{i=1}^p \frac{(c-i)}{(c-1)^p}
\]

The derivation of the above result is provided in Supplementary Section 6. Note that the right-hand side of Eq. (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 Sec. 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 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 \( p = 10, n = 400, \mu_0 = (1,1,\ldots,1)^T, \mu_b = (0,0,\ldots,0)^T, \) and \( \Sigma = 0.1 I_p \) 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 Figure 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. Figure 2 (c) and (d) show the true and approximated
critical point (or delay time) $t^*$ using Eq. (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 Eq. (21) and Eq. (22) that $t^*$ is approximately linear with $\sqrt{h}$ and $c$.

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

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 $h = 2$.
4.2 Illustration of the AMCR Rule

To illustrate the AMCR rule and evaluate its performance, four types of mean bias functions, i.e., 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 (White Jr et al. 2000, Cash et al. 1992, Hoad et al. 2010, Wu, Chen, Zhou, et al. 2016, Wu, Chen, and Zhou 2016, Hou et al. 2016).

Table 1. Four mean bias functions and their shapes

<table>
<thead>
<tr>
<th>Bias Type</th>
<th>Functional Form</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>( y(i) = \begin{cases} \frac{i - 1}{c_0}H, &amp; i = 1, \ldots, c_0 \ H, &amp; i = c_0 + 1, \ldots, N \end{cases} )</td>
<td></td>
</tr>
<tr>
<td>Quadratic</td>
<td>( y(i) = \begin{cases} \frac{(i - c_0 - 1)^2}{c_0^2}H, &amp; i = 1, \ldots, c_0 \ H, &amp; i = c_0 + 1, \ldots, N \end{cases} )</td>
<td></td>
</tr>
<tr>
<td>Exponential</td>
<td>( y(i) = \begin{cases} H \left[ 1 - \left( \frac{1 - i}{c_0} \right)^2 \right], &amp; i = 1, \ldots, c_0 \ 0.9H, &amp; i = c_0 + 1, \ldots, N \end{cases} )</td>
<td></td>
</tr>
<tr>
<td>Oscillating</td>
<td>( y(i) = \begin{cases} \frac{c_0 - i + 1}{c_0} \sin \left( \frac{\pi i}{f} \right), &amp; i = 1, \ldots, c_0 \ 0, &amp; i = c_0 + 1, \ldots, N \end{cases} )</td>
<td></td>
</tr>
</tbody>
</table>

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.

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 $i = 1, \ldots, c_0$ and $\Sigma_0 = 0.01I_2$, optimal truncation point $c^* = 271$. The vertical dashed lines denote the optimal truncation points obtained by AMCR rule.
We also illustrate the truncation process with only covariance matrix bias, though such scenario is not the major focus of this paper. 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 $|\Sigma_1 \Sigma_0| \approx 1.6$. For the second case, the parameters are set to $p = 2, \mu_0, c_0 = 300, N = 600, \Sigma_{1i} = 30(c_{0-\text{i+1}}/c_0) \times 0.01I_2$ for $i = 1, ..., c_0$, 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.

### 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, e.g., 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 paper, 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 $H = 1, N = 600, \Sigma = \sigma^2I_{10}$. Figure 5 shows the truncation point as a function of $h$ with different $c_0$ (top row: 200; bottom row: 300) and noise $\sigma$ (left column: 0.06; middle column:
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 $c^*$.

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, e.g., the point of the last sharp decrease. In this paper, 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 end point, 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 et al. 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

$$RMSE_c = \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} (c_i^* - c_{i0})^2}, MSE_{\mu} = \frac{1}{N_s} \sum_{i=1}^{N_s} (\bar{X}_i - \mu_i)^T (\bar{X}_i - \mu_i)$$

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 $\mu_i$ 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.

<table>
<thead>
<tr>
<th>Type</th>
<th>Equation</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(0)</td>
<td>$\psi_t = \epsilon_t$</td>
<td>$\epsilon_t \sim N(0, \sigma^2 I_p)$</td>
</tr>
<tr>
<td>AR(1)</td>
<td>$\psi_t = \phi_1 \psi_{t-1} + \epsilon_t$</td>
<td>$\epsilon_t \sim N(0, \sigma^2 I_p)$, $\phi_1 = 0.4$</td>
</tr>
<tr>
<td>AR(2)</td>
<td>$\psi_t = \phi_2 \psi_{t-1} + \phi_3 \psi_{t-2} + \epsilon_t$</td>
<td>$\epsilon_t \sim N(0, \sigma^2 I_p)$, $\phi_2 = -0.25$, $\phi_3 = 0.5$</td>
</tr>
<tr>
<td>CR</td>
<td>$\psi_t = e_t$</td>
<td>$\epsilon_t \sim N(0, \sigma^2 C_r)$</td>
</tr>
<tr>
<td>VAR(1)</td>
<td>$\psi_t = A \psi_{t-1} + \epsilon_t$</td>
<td>$\psi_0 \sim N(0, \Sigma)$, $\Sigma = \sigma^2 C_r$, $\epsilon_t \sim N(0, \Sigma_0)$, $\Sigma = A \Sigma A^T + \Sigma_0$</td>
</tr>
</tbody>
</table>

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 auto-correlation 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 auto-correlation 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, Wu et al. 2013, White Jr et al. 2000), Hotelling $T^2$ method (Shruben 1981) and batch means test (BMT) (Goldsmn et al. 1994, White Jr et al. 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 (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 $\alpha$ is selected as the truncation point. In (Shruben 1981), $\alpha$ 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 $RMSE_e$. The other multivariate
truncation method by Gallagher et al. (Gallagher et al. 1994) is not used for comparison since it is only applicable to certain autoregressive data and is unnecessarily complex.

The detailed truncation results and comparison in terms of $RMSE_c$, $MSE_\mu$ and UTR are shown in Table 3 and Table 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 $MSE_\mu$, 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 et al. 2000). Compared with MSER, the proposed AMCR has much lower UTR, though slightly higher than $T^2$.

(2) Increasing $c_0$ or noise $\sigma$ will increase $RMSE_c, MSE_\mu$ 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 ($RMSE_c, MSE_\mu/p$, UTR) of all methods are improved. It is a key advantage of multivariate truncation method over univariate method. For AMCR, increasing $p$ will increase the mean shifting rate $r_s$ as defined in Eq. (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 $\beta$ is the beta error) of the full $T^2$ control will always be higher than the one with reduced dimensionality, i.e., 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 $c_i^*$, $i = 1, ..., p$, then $c^*(p) = \max(c_1^*, ..., c_p^*)$. Since $c^* < c_0$ for almost all the cases, $\max(c_1^*, ..., c_p^*) \leq \max(c_1^*, ..., c_{p+1}^*) \leq c_0$. 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 et al. 2017). In many production systems, the products, e.g., 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.

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

<table>
<thead>
<tr>
<th>Signal</th>
<th>$c_0$</th>
<th>$\sigma_x$</th>
<th>AMCR</th>
<th>MSER</th>
<th>$T^2$</th>
<th>BMT</th>
<th>$MSE_e$</th>
<th>$MSE_{\mu}$</th>
<th>BMT</th>
<th>AMCR</th>
<th>MSER</th>
<th>$T^2$</th>
<th>BMT</th>
<th>AMCR</th>
<th>MSER</th>
<th>$T^2$</th>
<th>BMT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noise</td>
<td>AR(0)</td>
<td>0.06</td>
<td>8.5</td>
<td>10.9</td>
<td>20.5</td>
<td>27.9</td>
<td>0.0259</td>
<td>0.0260</td>
<td>0.0259</td>
<td>0.0259</td>
<td>0.74</td>
<td>1.00</td>
<td>0.44</td>
<td>0.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.14</td>
<td>11.1</td>
<td>26.0</td>
<td>22.0</td>
<td>25.8</td>
<td>0.0283</td>
<td>0.0282</td>
<td>0.0282</td>
<td>0.0286</td>
<td>0.76</td>
<td>1.00</td>
<td>0.67</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(1)</td>
<td>0.06</td>
<td>13.9</td>
<td>17.8</td>
<td>20.2</td>
<td>21.0</td>
<td>0.0246</td>
<td>0.0247</td>
<td>0.0245</td>
<td>0.0246</td>
<td>0.74</td>
<td>1.00</td>
<td>0.50</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.14</td>
<td>27.4</td>
<td>41.5</td>
<td>34.8</td>
<td>30.2</td>
<td>0.0262</td>
<td>0.0272</td>
<td>0.0264</td>
<td>0.0274</td>
<td>0.88</td>
<td>1.00</td>
<td>0.79</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR(2)</td>
<td>0.06</td>
<td>8.7</td>
<td>9.6</td>
<td>25.5</td>
<td>20.5</td>
<td>0.0259</td>
<td>0.0260</td>
<td>0.0259</td>
<td>0.0249</td>
<td>0.69</td>
<td>0.96</td>
<td>0.44</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.14</td>
<td>17.6</td>
<td>26.6</td>
<td>22.5</td>
<td>24.3</td>
<td>0.0279</td>
<td>0.0280</td>
<td>0.0282</td>
<td>0.0280</td>
<td>0.69</td>
<td>1.00</td>
<td>0.65</td>
<td>0.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CR</td>
<td>0.06</td>
<td>10.7</td>
<td>16.1</td>
<td>23.3</td>
<td>21.9</td>
<td>0.0246</td>
<td>0.0246</td>
<td>0.0246</td>
<td>0.0263</td>
<td>0.77</td>
<td>0.96</td>
<td>0.51</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.14</td>
<td>24.0</td>
<td>42.2</td>
<td>35.8</td>
<td>36.8</td>
<td>0.0265</td>
<td>0.0267</td>
<td>0.0271</td>
<td>0.0261</td>
<td>0.78</td>
<td>1.00</td>
<td>0.74</td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VAR(1)</td>
<td>0.06</td>
<td>9.8</td>
<td>12.6</td>
<td>23.3</td>
<td>22.3</td>
<td>0.0259</td>
<td>0.0260</td>
<td>0.0259</td>
<td>0.0290</td>
<td>0.67</td>
<td>0.97</td>
<td>0.41</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.14</td>
<td>14.7</td>
<td>24.7</td>
<td>35.8</td>
<td>39.5</td>
<td>0.0256</td>
<td>0.0267</td>
<td>0.0271</td>
<td>0.0281</td>
<td>0.84</td>
<td>1.00</td>
<td>0.76</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>0.06</td>
<td>8.2</td>
<td>12.9</td>
<td>31.1</td>
<td>18.1</td>
<td>0.0307</td>
<td>0.0304</td>
<td>0.0308</td>
<td>0.0304</td>
<td>0.75</td>
<td>0.92</td>
<td>0.34</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.14</td>
<td>15.9</td>
<td>28.0</td>
<td>24.2</td>
<td>31.4</td>
<td>0.1454</td>
<td>0.1456</td>
<td>0.1481</td>
<td>0.1461</td>
<td>0.79</td>
<td>1.00</td>
<td>0.57</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 $\tau_0$. The 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 et al. 2017).

Figure 6. Bernoulli line with perishable products.

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. 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 $c^* = 35$. 
Figure 7 shows the four performance measures through discrete-event simulation with $p_1 = 0.9265$, $p_2 = 0.9070$, $N_b = 2$, $T_{\text{max}} = 2$, $n_0 = 0$, $\tau_0 = 0$. 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 et al. 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 $MSE_\mu$ as the evaluation metric. The $MSE_\mu$ 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.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Simulation parameters</th>
<th>Steady state performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p_1$</td>
<td>$p_2$</td>
</tr>
<tr>
<td>1</td>
<td>0.9256</td>
<td>0.9070</td>
</tr>
<tr>
<td>2</td>
<td>0.8305</td>
<td>0.7623</td>
</tr>
<tr>
<td>3</td>
<td>0.8554</td>
<td>0.7273</td>
</tr>
</tbody>
</table>

5 Conclusion

In this paper, an efficient method named adaptive minimal confidence region rule (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 statistical process control charts, using the joint confidence region of all the means may not be the optimal choice.

**Supplementary Materials**

The proofs for Theorem 1-5 and Proposition 1 are provided in online supplementary material (Supplementary.pdf).

**Acknowledgement**

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

**References**


28


Pasupathy, Raghu, and Bruce Schmeiser. 2010. The initial transient in steady-state point estimation: Contexts, a bibliography, the mse criterion, and the MSER statistic. Paper read at Winter Simulation Conference.


