http://pubsonline.informs.org/journal/ijoc/

INFORMS JOURNAL ON COMPUTING
Vol. 31, No. 2, Spring 2019, pp. 347–366
ISSN 1091-9856 (print), ISSN 1526-5528 (online)

Modeling and Simulation of Nonstationary Non-Poisson Arrival Processes

Ran Liu,[a] Michael E. Kuhl,[b] Yunan Liu,[c] James R. Wilson[c]

[a] SAS Institute, Cary, North Carolina 27513; [b] Department of Industrial and Systems Engineering, Rochester Institute of Technology, Rochester, New York 14623; [c] Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, North Carolina 27695-7906

Contact: ran.liu@sas.com (RL); mekeie@rit.edu (MEK); yliu48@ncsu.edu (YL); jwilson@ncsu.edu,
http://orcid.org/0000-0002-6255-4485 (JRW)

Received: August 16, 2015
Revised: January 25, 2015; August 16, 2015; March 15, 2017; November 22, 2017; March 15, 2018
Accepted: March 23, 2018
Published Online in Articles in Advance: April 30, 2019

https://doi.org/10.1287/ijoc.2018.0828

Abstract. We develop CIATA, a combined inversion-and-thinning approach for modeling a nonstationary non-Poisson process (NNPP), where the target arrival process is described by a given rate function and its associated mean-value function together with a given asymptotic variance-to-mean (dispersion) ratio. CIATA is based on the following: (i) a piecewise-constant majorizing rate function that closely approximates the given rate function from above; (ii) the associated piecewise-linear majorizing mean-value function; and (iii) an equilibrium renewal process (ERP) whose noninitial interrenewal times have mean 1 and variance equal to the given dispersion ratio. Transforming the ERP by the inverse of the majorizing mean-value function yields a majorizing NNPP whose arrival epochs are then thinned to deliver an NNPP having the specified properties. CIATA-Ph is a simulation algorithm that implements this approach based on an ERP whose noninitial interrenewal times have a phase-type distribution. Supporting theorems establish that CIATA-Ph can generate an NNPP having the desired mean-value function and asymptotic dispersion ratio. Extensive simulation experiments substantiated the effectiveness of CIATA-Ph with various rate functions and dispersion ratios. In all cases, we found approximate convergence of the dispersion ratio to its asymptotic value beyond a relatively short warm-up period.

History: Accepted by Marvin Nakayama, Area Editor for Simulation.
Funding: Y. Liu received financial support from the National Science Foundation [Grant CMMI-1362310]. J. R. Wilson received financial support from the National Science Foundation [Grant CMMI-1232998].
Supplemental Material: The online supplement is available at https://doi.org/10.1287/ijoc.2018.0828.

Keywords: nonstationary arrival process • non-Poisson process • time-dependent arrival rate • dispersion ratio • index of dispersion for counts

1. Introduction

In the formulation of a high-fidelity stochastic simulation model of a complex system, special attention must often be given to the system's arrival processes. A stream of random arrivals with a constant arrival rate is usually modeled by a homogeneous Poisson process (HPP), which is characterized by interarrival times that are independent and identically distributed (i.i.d.) exponential random variables whose mean is the reciprocal of the arrival rate. Unfortunately, many arrival processes of interest have arrival rates that exhibit substantial variation over time. For instance, in many locales the rate of occurrence of storms exhibits a time-of-year effect within each year as well as a long-term trend over successive years (Lee et al. 1991). In call centers, the call arrival rate can exhibit intraday (within-day), daily, weekly, monthly, and yearly effects; and system performance depends strongly on these effects (Ibrahim et al. 2016).

Nonhomogeneous Poisson processes (NHPPs) have been used to model arrival processes with time-dependent arrival rates in a broad range of application domains (Lewis and Shedler 1976, Pritsker et al. 1995, Kim and Whitt 2014). For an NHPP with a given rate function and the associated mean-value function describing the expected buildup of arrivals over time, exact algorithms for simulating that process are based on the method of inversion or the method of thinning. Introduced by Çinlar (1975), the method of inversion requires using the inverse of the given mean-value function to transform the arrival epochs of a "base" HPP with arrival rate equal to 1, yielding the arrival epochs of the target process. However, in general, the mean-value function corresponding to an arbitrary rate function can be analytically intractable or difficult to invert numerically, so the method of inversion may be impractical or computationally inefficient at best. Consequently, this method is limited to special forms of

the rate function for which the mean-value function is readily invertible. In particular, inversion is used mostly with rate functions that are piecewise constant (Chen and Schmeiser 2015), piecewise linear (Nicol and Leemis 2014), piecewise quadratic (Chen and Schmeiser 2017), or trigonometric (Chen and Schmeiser 1992).

Proposed by Lewis and Shedler (1979), the method of thinning in its simplest form exploits an upper bound for the given rate function to generate a base HPP with arrival rate equal to that upper bound; then each arrival epoch in the base HPP is independently accepted for inclusion in the target NHPP with probability equal to the ratio of the given rate function evaluated at that epoch divided by the upper bound. If this ratio is much less than 1 over a substantial part of the simulation's time horizon, then thinning is computationally inefficient because a relatively large percentage of the arrival epochs in the base HPP are rejected.

1.1. The Need to Go Beyond NHPPs

Although an NHPP can accurately represent a given time-dependent arrival rate and the associated mean-value function, in many simulation studies an arrival process exhibits stochastic variability about its mean-value function that cannot be represented even approximately by an NHPP, while system performance depends strongly on this variability (Fendick and Whitt 1989, Ibrahim et al. 2016). For example, there is substantial empirical evidence of nonstationary non-Poisson arrival processes (NNPPs) for call centers and healthcare systems (Avramidis et al. 2004, Jongbloed and Koole 2001, Aldor-Noiman et al. 2009, Steckley et al. 2009). The variance-to-mean (dispersion) ratio measures such variability through the ratio of the variance of the arrival process divided by its mean-value function evaluated at each point in the relevant time horizon. (The dispersion ratio is also called the index of dispersion for counts.) An NHPP has a dispersion ratio that is exactly equal to 1 at each point in time, while many nonstationary arrival processes have dispersion ratios that differ substantially from 1 and thus exhibit non-Poisson behavior. In the rest of this subsection, we briefly discuss the following: (i) a numerical example of such non-Poisson behavior in a queueing system, and (ii) the importance of the dispersion ratio as a property of the arrival process that affects the principal performance measures for a broad class of queueing systems.

1.1.1. Numerical Example of Non-Poisson Behavior.

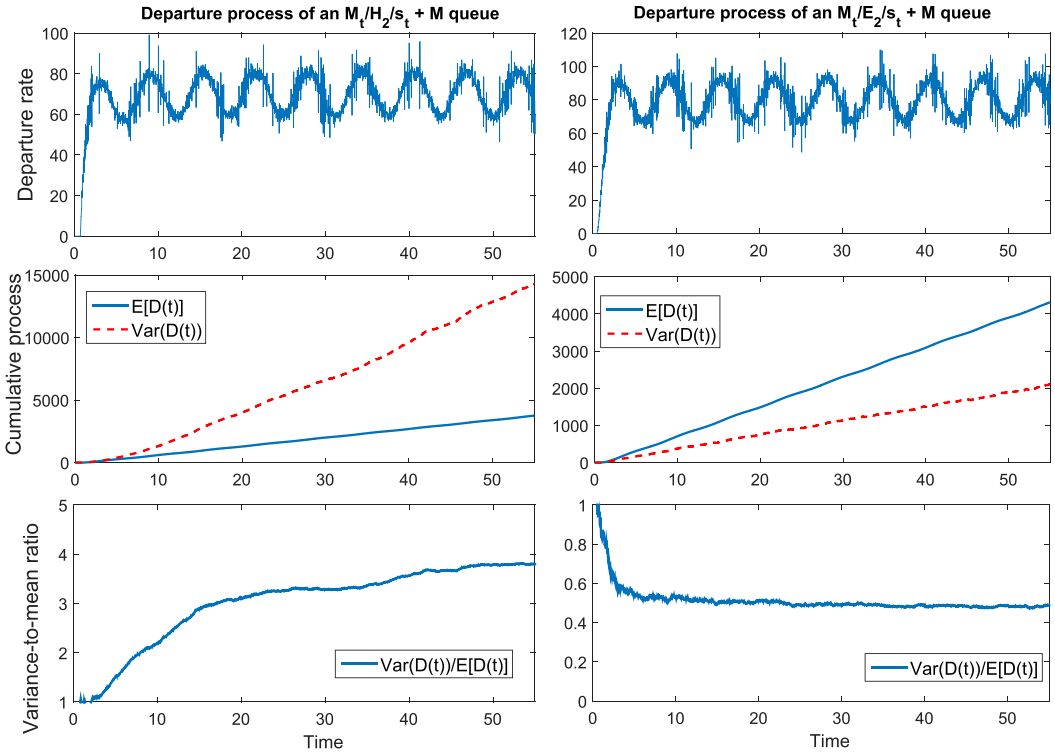
In many types of service networks (e.g., call centers, hospitals, and production systems), the arrival process for each downstream node is composed of the departure (service completion) processes from the associated upstream nodes (service centers) (Liu and Whitt 2012, 2014, 2017; Liu 2018). Hence it is important to model these departure processes carefully and to develop effective simulation procedures for performance forecasting. For simplicity, we consider the departure process from a single $M_t/GI/s_t + M$ queuing system having an arrival process that is an NHPP with sinusoidal arrival rate (the M_t), nonexponential service distribution (the GI), time-varying staffing level (the s_t), and exponential abandonment distribution (the M). In the $M_t/GI/s_t + M$ system, when the quality of service is low (e.g., the mean waiting time is high), the departure process $\{D(t) = \text{number of departures in } [0, t] \text{ for } t \geq 0\}$ exhibits pronounced non-Poisson behavior, having a dispersion ratio $\text{Var}[D(t)]/\text{E}[D(t)]$ far from 1 as well as a time-varying departure rate.

On the basis of data from Liu and Whitt (2014), Figure 1 shows that for the departure process from our $M_t/GI/s_t + M$ system, the dispersion ratio depends strongly on the squared coefficient of variation (CV^2) of the service distribution. In particular, the dispersion ratio of the departure process $\{D(t) : t \geq 0\}$ converges to the following: (i) the value 4 for a hyperexponential service distribution with $\text{CV}^2 = 4$ (bottom left panel of Figure 1), and (ii) the value 0.5 for an Erlang service distribution with $\text{CV}^2 = 0.5$ (bottom right panel of Figure 1). All these considerations motivate the need for effective and computationally efficient methods for modeling and simulating NNPPs.

1.1.2. Importance of the Dispersion Ratio in Modeling and Simulating NNPPs.

For a single-server queue in which the (general) service times are i.i.d. and independent of the (general) arrival process $\{N(t) : t \geq 0\}$, the time-dependent dispersion ratio $C(t) \equiv \text{Var}[N(t)]/\text{E}[N(t)]$ is one of the most important properties affecting key performance measures, and an estimate of the asymptotic dispersion ratio is often used in analytical approximations to those performance measures (Fendick and Whitt 1989). In recent work (He et al. 2016, Liu and Whitt 2017, Liu 2018), the arrival process for a $G_t/GI/s_t + GI$ system is characterized by two properties: the time-dependent rate function $\{\lambda(t) : t \geq 0\}$ and the asymptotic dispersion ratio C. One of the main conclusions is that inaccurate estimation of C results in significant underestimation or overestimation of the appropriate staffing levels $\{s_t : t \geq 0\}$ for the system. In the context of many-server queues, Liu and Whitt (2014, 2017) conclude that the variability in the arrival process chiefly affects system performance and staffing through the asymptotic dispersion ratio. Similarly, in a feedforward queueing network with nodes that are instances of an $M_t/GI/s_t + GI$ system, the key performance measures for a downstream node are affected mainly by the time-dependent rate function and the dispersion ratio of the aggregate departure process

Figure 1. (Color online) Estimated Departure Rate, Mean, Variance, and Variance-to-Mean (Dispersion) Ratio for the Number of Service Completions in Two $M_t/GI/s_t + GI$ Queueing Systems



Note. Based on data from Liu and Whitt (2014), the left panel depicts system performance with the waiting-time target $w = 0.2$, the rate function $\lambda(t) = 100 + 20\sin(t)$, the time-dependent staffing function s_t of Liu and Whitt (2012), the exponential distribution of patience times with rate 0.5, and a hyperexponential (H_2) distribution of service times. In the right panel, the service-time distribution is Erlang (E_2).

$\{D(t) : t \geq 0\}$ that is generated by the associated upstream nodes and is routed to the given downstream node (Liu and Whitt 2014). Thus the asymptotic dispersion ratio for the aggregate process $\{D(t) : t \geq 0\}$ is central to analytical and simulation-based approximations of the performance of the downstream node. Moreover, Liu and Whitt (2014) find that the dispersion ratio for $\{D(t) : t \geq 0\}$ is an effective tool for studying the deviation of this process from the NHPP property; see also Cox and Lewis (1966), Fendick and Whitt (1989), and Sriram and Whitt (1986).

1.2. CIATA: A Combined Inversion-and-Thinning Model of an NNPP

In this article, we develop methods for modeling and simulating an NNPP with a given asymptotic dispersion ratio as well as given rate and mean-value functions. First, we explain our approach to modeling such an NNPP, and we discuss the theoretical and practical basis for this approach. The corresponding algorithm for simulating our NNPP model is detailed in Section 3.

The nonstationary arrival process $\{N(t) = \text{number of arrivals in } [0,t] : t \geq 0\}$ has the given mean-value function

$$\mathrm{E}[N(t)] = \mu(t) \equiv \int_0^t \lambda(y)\mathrm{d}y \text{ for } t \geq 0, \tag{1}$$

where the associated rate function $\lambda(y)$ is assumed to be nonnegative, bounded, and continuous on $[0,\infty)$ such that $\mu(t) \to \infty$ as $t \to \infty$. Moreover, each finite time interval $[0,t]$ is assumed to have a (finite) partition such that $\lambda(y)$ is constant or quasiconcave on each subinterval of the partition. Finally, $\{N(t) : t \geq 0\}$ is assumed to have a given asymptotic dispersion ratio

$$C \equiv \lim_{t\to\infty} \frac{\mathrm{Var}[N(t)]}{\mathrm{E}[N(t)]}, \text{ where } 0 < C < \infty. \tag{2}$$

Because both HPPs and NHPPs have $\mathrm{Var}[N(t)]/\mathrm{E}[N(t)] = 1$ for all $t > 0$, an arrival process with a given time-dependent arrival rate and $C \neq 1$ must be an NNPP.

To model a given NNPP with the properties (1) and (2), we formulate CIATA, a combined inversion-and-thinning approach, in two steps as follows.

Step 1. Let $\{N^\circ(u) : u \geq 0\}$ be an equilibrium renewal process (ERP) with independent interrenewal times $\{X_n^\circ : n \geq 1\}$ such that (i) the $\{X_n^\circ : n \geq 2\}$ are i.i.d. continuous random variables having $\mathrm{E}[X_2^\circ] = 1$,

$\mathrm{Var}[X_2^\circ] = C$, and cumulative distribution function (c.d.f.) $G(x) = \Pr\{X_2^\circ \le x\}$ for $x \ge 0$ with $G(0) = 0$ and $G(x) < 1$ for $x > 0$; and (ii) X_1° has the associated equilibrium c.d.f. $G_e(x) = (1/\mathrm{E}[X_2^\circ])\int_0^x [1 - G(y)]\mathrm{d}y$ for $x \ge 0$. The renewal epochs of the ERP are denoted by $\{S_n^\circ : n \ge 0\}$, where $S_0^\circ = 0$. Let $\widetilde{\lambda}(t)$, $t \ge 0$, be a positive, piecewise-constant majorizing approximation to the given rate function $\lambda(t)$, $t \ge 0$ (i.e., $\widetilde{\lambda}(t) \ge \lambda(t)$ and $\widetilde{\lambda}(t) > 0$ for $t \ge 0$) so that the associated piecewise-linear function $\widetilde{\mu}(t) = \int_0^t \widetilde{\lambda}(y)\mathrm{d}y$, $t \ge 0$, is a majorizing approximation to the given mean-value function (1). By transforming the ERP's renewal epochs using the inverse of $\widetilde{\mu}(t)$, we obtain the arrival epochs $\{\widetilde{S}_n = \widetilde{\mu}^{-1}(S_n^\circ) : n \ge 0\}$ (with $\widetilde{S}_0 = 0$) of an NNPP $\{\widetilde{N}(t) : t \ge 0\}$ that is defined by $\widetilde{N}(t) = \max\{n : \widetilde{S}_n \le t\}$ for $t \ge 0$ and that has rate function $\widetilde{\lambda}(t)$, $t \ge 0$, and mean-value function $\mathrm{E}[\widetilde{N}(t)] = \widetilde{\mu}(t)$, $t \ge 0$.

Step 2. For $n \ge 1$, the nth arrival epoch $\widetilde{S}_n$ is independently accepted for inclusion in the final simulation-generated arrival process $\{\widehat{N}(t) : t \ge 0\}$ with probability $\lambda(\widetilde{S}_n)/\widetilde{\lambda}(\widetilde{S}_n)$ so that (i) the thinned arrival process $\{\widehat{N}(t) : t \ge 0\}$ is our model of the given NNPP $\{N(t) : t \ge 0\}$, and (ii) the thinned arrival process has the desired mean-value function (1) and asymptotic dispersion ratio (2).

Remark 1 (Basis for the CIATA Model). For the reasons discussed in Section 1.1, CIATA is based on the two simplest and most prominent properties of an NNPP (namely, $\{\lambda(t) : t \ge 0\}$ and C), which can be estimated from arrival data more easily and rapidly than other properties of an NNPP such as quantiles and the covariances between the arrival counts in nonoverlapping time intervals. The relative simplicity and ease of estimating the input parameters of CIATA will enable practitioners to obtain useful and timely information about the behavior of a complex NNPP. All the foregoing theoretical and practical considerations form the basis on which CIATA is proposed as a model of an NNPP with a given rate function and asymptotic dispersion ratio in a broad class of simulation applications. ◁

Remark 2 (Limitations of the CIATA Model). In modeling daily arrivals to a telephone call center, Avramidis et al. (2004) find that system performance is sensitive not only to a time-varying arrival rate and a non-Poisson dispersion ratio for the arrival process during each day but also to nonzero covariances between the arrival counts in different (nonoverlapping) time periods within a day and across successive days. Although CIATA provides a way to match the first and second properties closely under certain conditions, it lacks a method for matching the third property; thus CIATA may be unsuitable in applications that require accurate approximations for all three properties. ◁

1.3. Summary of Contributions

Our contributions are threefold. The novelty of CIATA is that it effectively combines the inversion and thinning techniques in a model of NNPPs, which accommodates the use of flexible and convenient distributions for the noninitial interrenewal c.d.f. $G(x)$ when CIATA is implemented in a simulation algorithm. As detailed in the literature review (Section 2), other recent methods for modeling and simulating a given NNPP can only match the given mean-value function and may not be able to match the given value of the asymptotic dispersion ratio C even approximately as the length of the simulation's time horizon increases. Moreover, some of the latter methods can be either difficult to implement or computationally inefficient if the inverse of the given mean-value function is difficult or expensive to compute. CIATA avoids both of these drawbacks. Therefore we believe that CIATA is an advance in modeling NNPPs.

To implement the CIATA model, we develop CIATA-Ph, a simulation algorithm that uses simple phase-type distributions for $G(x)$ to generate a realization of the NNPP $\{\widehat{N}(t)\}$ over a given finite time horizon $[0, S]$. We formulate conditions sufficient to ensure that CIATA-Ph can achieve the desired mean-value function exactly on $[0, S]$ and that it can approach the desired value of C as $S \to \infty$ and $t \to \infty$. The analysis establishing these results clearly reveals the precise role of each major step of CIATA-Ph, with special emphasis on how the inversion and thinning steps fit together in the operation of CIATA-Ph.

To evaluate the effectiveness of CIATA-Ph in practice, we conduct extensive simulation experiments on several types of arrival processes, including some arrival processes that do not satisfy the assumptions on which CIATA-Ph is based. We provide evidence that CIATA-Ph can closely achieve a wide range of desired values for the dispersion ratio (from $C = 0.2$ to 10) as well as the desired mean-value function. We quantify the associated errors using a variety of measures of the closeness of key empirical characteristics of the NNPP $\{\widehat{N}(t) : t \in [0, S]\}$ delivered by CIATA-Ph to the corresponding theoretical characteristics of the given NNPP $\{N(t) : t \in [0, S]\}$. We also conduct a sensitivity analysis with respect to various parameters of CIATA and CIATA-Ph.

The rest of this article is organized as follows. In Section 2, we review the related literature on methods for modeling, estimating, and simulating NNPPs. In Section 3, we develop CIATA-Ph. In Section 4, we summarize the results of our experimental performance evaluation for CIATA-Ph, and we close Section 4 with a comparison of the performance of CIATA and CIATA-Ph with other recent procedures for modeling and simulating NNPPs. Conclusions and recommendations for future work are summarized in Section 5.

The online supplement contains proofs of the main theoretical results supporting the CIATA model and its implementation in CIATA-Ph as well as additional experimental results for CIATA-Ph. This article is a follow-up to Kuhl and Wilson (2009), in which the idea for CIATA is first proposed. Liu (2013) provides the basis for this article, and Liu et al. (2015) summarize some additional preliminary simulation results for a predecessor of CIATA-Ph.

2. Literature Review

Gerhardt and Nelson (2009) develop an inversion method for modeling and simulating an NNPP that can achieve desired values of the following: (i) the mean-value function $\mu(t)$ at each time $t \geq 0$ and (ii) the asymptotic dispersion ratio C as $t \to \infty$. Originally proposed by Massey and Whitt (1994), the idea is a generalization of the conventional inversion method for simulating an NHPP in which the usual base process (i.e., a rate-1 HPP) is replaced by an ERP whose noninitial interrenewal times have mean equal to 1 and variance equal to C. This method is useful when the mean-value function is easily invertible.

Gerhardt and Nelson (2009) also develop a thinning method for modeling and simulating an NNPP. On the basis of an upper bound $\tilde{\lambda}$ for the given rate function, thinning is applied to a majorizing ERP with noninitial interrenewal times having mean $1/\tilde{\lambda}$ and variance $C/(\tilde{\lambda})^2$. The authors show that the resulting thinned arrival process has the desired mean-value function $\mu(t)$ for all $t \geq 0$; however, the thinned process does not, in general, have the desired asymptotic dispersion ratio C. The other disadvantage of this method is that it may be computationally inefficient if $\lambda(t) \ll \tilde{\lambda}$ over a substantial range of values for t, resulting in a relatively large number of rejections.

All the existing methods for estimating the rate function of an NNPP seem to be based directly on the nonparametric techniques of Henderson (2003) for estimating the rate function of an NHPP, where the latter rate function is assumed to be piecewise constant. These nonparametric techniques exploit K i.i.d. realizations of the given NHPP, and all the resulting arrival epochs are accumulated in adjacent observation intervals of a common length ζ that is independent of the data but may depend on K. Henderson (2003) establishes key asymptotic properties of the associated estimators of the rate and mean-value functions as $K \to \infty$, both when ζ is constant and when ζ tends to zero with increasing K. However, it is unclear whether these asymptotic properties also apply to estimation of the rate function of an NNPP because the proofs by Henderson (2003) depend critically on key properties of the Poisson distribution.

On the basis of the techniques of Henderson (2003), Gerhardt and Nelson (2009) develop a nonparametric estimator of C using weighted least-squares regression to fit the relation $\mathrm{Var}[N(t)] = C\mu(t) + \varepsilon_t$ for selected values of t in a finite observation interval $[0, T_E]$. As $K \to \infty$, the asymptotic properties of this estimator of C are unclear. In particular, from the statement and proof of Theorem 2 below, we see that in this situation the appropriate linear model is $\mathrm{Var}[N(t)] = C\mu(t) + \theta^* + o(1)$, where $o(1)$ denotes a function $\vartheta(t) \xrightarrow[t\to\infty]{} 0$; and, in general, the intercept θ^* can have a relatively large magnitude. Therefore the authors' estimator of C is generally biased by using regression through the origin or by taking observation epochs $t \in [0, T_E]$ for which $\mu(t)$ is small relative to $|\theta^*|$.

As mentioned in Remark 2, Avramidis et al. (2004) formulate three models for the stream of daily arrivals at a call center that provide approximations to the following features of the daily arrival process: (i) a time-dependent arrival rate, (ii) overdispersion compared with a Poisson arrival process, and (iii) dependence between arrival counts in different time periods within the day. However, the accuracy that can be achieved by these approximations is not entirely clear, and the authors do not provide conditions under which these models can match properties (i)–(iii) exactly.

3. CIATA-Ph: An Algorithm for Simulating CIATA-Based NNPPs

Extending the inversion and thinning methods of Gerhardt and Nelson (2009) for simulating an NNPP with given mean-value function and asymptotic dispersion ratio, CIATA-Ph is designed to have broader applicability and substantially improved computational efficiency compared with either of those methods. In particular, CIATA-Ph can handle a broader range of functional forms for $\lambda(t)$ and $\mu(t)$ than the inversion method, and unlike the thinning method, CIATA-Ph can achieve any value of C. At certain points in the rest of this article, one or both of the following assumptions will be required.

Assumption 1. *The given rate function $\lambda(t)$ for $t \geq 0$ has a finite upper bound λ^*. Moreover, each finite time interval $[0, t]$ has a (finite) partition such that $\lambda(y)$, $y \in [0, t]$, is constant or quasiconcave on each subinterval of the partition.*

Assumption 2. *The rate function $\lambda(t)$ is continuous at every $t \geq 0$, and the associated mean-value function has the property $\mu(t) \to \infty$ as $t \to \infty$.*

3.1. Overview of CIATA-Ph

To simulate the CIATA model of a given NNPP $\{N(t) : t \in [0, S]\}$ with $0 < S < \infty$ and with a given rate function $\lambda(t)$, $t \in [0, S]$, that satisfies Assumptions 1 and 2 as well as Equations (1) and (2), we perform the following five steps of CIATA-Ph:

Step 1. Construct a positive piecewise-constant majorizing rate function $\widetilde{\lambda}_Q(t)$ that closely bounds the given rate function $\lambda(t)$ for $t \in [0, S]$ based on a partition of $[0, S]$ with Q associated subintervals of $[0, S]$, where Q is taken sufficiently large to ensure that $\lambda(t)$ is constant or quasiconcave on each subinterval. This step is detailed in Section 3.2 and is stated formally in Algorithm 1 below.

Step 2. Construct the piecewise-linear mean-value function $\widetilde{\mu}_Q(t) = \int_0^t \widetilde{\lambda}_Q(y)\mathrm{d}y$ for $t \in [0, S]$.

Step 3. Generate an ERP $\{N^\circ(u) : u \in [0, \widetilde{\mu}_Q(S)]\}$ with a noninitial interrenewal distribution $G(x)$ having the properties specified in Section 1.2 so as to yield the ERP's renewal epochs $\{S_n^\circ : n \geq 1\}$. This step is elaborated in Section 3.3 below.

Step 4. Generate the arrival epochs $\{\widetilde{S}_n : n = 1, \ldots, \widetilde{N}_Q(S)\}$ for the majorizing NNPP $\{\widetilde{N}_Q(t) : t \in [0, S]\}$, where $\widetilde{N}_Q(S) = N^\circ[\widetilde{\mu}_Q(S)]$, by evaluating the inverse of the majorizing mean-value function at the ERP's renewal epochs so that $\widetilde{S}_n = \widetilde{\mu}_Q^{-1}(S_n^\circ)$ for $n = 1, \ldots, \widetilde{N}_Q(S)$.

Step 5. Apply the method of thinning to the $\{\widetilde{S}_n : n = 1, \ldots, \widetilde{N}_Q(S)\}$ so that for $n = 1, \ldots, \widetilde{N}_Q(S)$, the nth arrival epoch $\widetilde{S}_n$ of the majorizing NNPP is independently accepted for inclusion in the delivered NNPP with probability $\lambda(\widetilde{S}_n)/\widetilde{\lambda}_Q(\widetilde{S}_n)$. Let $\{S_\ell : \ell = 1, \ldots, N_Q(S)\}$ denote the resulting sequence of accepted arrival epochs, and let $N_Q(t) = \max\{\ell : S_\ell \leq t\}$ for $t \in [0, S]$. Then $\{\widehat{N}(t) = N_Q(t) : t \in [0, S]\}$ is the NNPP delivered by CIATA-Ph.

Algorithm 2 in Section 3.4 is a formal statement of CIATA-Ph. Theorem 1 below establishes that if Assumption 1 holds and Q is sufficiently large, then we have $\mathrm{E}[N_Q(t)] = \mu(t)$ for $t \in [0, S]$; thus the NNPP generated by CIATA-Ph achieves the given mean-value function (1) exactly. Theorem 2 below establishes that if Assumptions 1 and 2 hold, then $\lim_{t\to\infty}\lim_{S\to\infty}\lim_{Q\to\infty}\mathrm{Var}[N_Q(t)]/\mathrm{E}[N_Q(t)] = C$ so that the NNPP generated by CIATA-Ph approximately achieves the given asymptotic dispersion ratio (2) if t, S, and Q are sufficiently large.

3.2. Constructing a Majorizing Rate Function

To enable efficient generation of arrival epochs, we seek to construct a positive piecewise-constant majorizing rate function $\widetilde{\lambda}_Q(t)$ that converges uniformly to the given rate function $\lambda(t)$ for $t \in [0, S]$ as $Q \to \infty$ provided Assumptions 1 and 2 hold. We let

$$\mathscr{L} \equiv \bigcup_{\ell=1}^{L} (\zeta_\ell^\dagger, \zeta_\ell^\ddagger) \tag{3}$$

represent the set of nonoverlapping intervals within $[0, S]$ on which $\lambda(t)$ is constant, and we assume that the associated endpoints $\{\zeta_\ell^\dagger, \zeta_\ell^\ddagger : \ell = 1, \ldots, L\}$ are known. Let $\varepsilon = 10^{-6}$. For each $Q \geq 1$ and $t \in \mathscr{L}$, we take

$$\widetilde{\lambda}_Q(t) \equiv \max\{\lambda[(\zeta_\ell^\dagger + \zeta_\ell^\ddagger)/2], \varepsilon/Q\} \text{ if } t \in (\zeta_\ell^\dagger, \zeta_\ell^\ddagger). \tag{4}$$

In $[0, S] \setminus \mathscr{L}$, we define $\widetilde{\lambda}_Q(t)$ separately on each of the associated nonoverlapping intervals $\{[\xi_j^\dagger, \xi_j^\ddagger] : j = 1, \ldots, M\}$ so that we have

$$\mathscr{M} \equiv \bigcup_{j=1}^{M} [\xi_j^\dagger, \xi_j^\ddagger] = [0, S] \setminus \mathscr{L}. \tag{5}$$

Assumption 1 ensures that we can find a sufficiently large integer Q^* such that for $Q \geq Q^*$ and for each interval $[\xi_j^\dagger, \xi_j^\ddagger]$ (where $j = 1, \ldots, M$), there is a partition $\mathbb{P}_{j,Q} = \{z_{i,j} : 0 \leq i \leq Q_j^\#\}$ of that interval with the associated subintervals $\{[z_{i-1,j}, z_{i,j}] : i = 1, \ldots, Q_j^\#\}$ such that we have the following properties: (i) $Q_j^\# \geq Q$; (ii) $\lambda(t)$ is quasiconcave on each subinterval of the partition; (iii) the mesh of the partition (i.e., the maximum length of the partition's subintervals) does not exceed $(\xi_j^\ddagger - \xi_j^\dagger)/Q$; and (iv) the partition $\mathbb{P}_{j,Q} \subset \mathbb{P}_{j,Q-1}$ if $Q > Q^*$. Moreover, Assumptions 1 and 2, the extreme value theorem (Royden and Fitzpatrick 2010), and proposition 3.8 of Avriel et al. (2010) ensure the following additional property: (v) on each subinterval associated with the partition $\mathbb{P}_{j,Q}$ of $[\xi_j^\dagger, \xi_j^\ddagger]$, the restriction of $\lambda(t)$ to that subinterval takes its maximum value and is unimodal; therefore we can set $\widetilde{\lambda}_Q(t)$ equal to the maximum of $\lambda(t)$ over that subinterval. In practice, we assume that the regularly spaced points $\{z_{i,j} \equiv \xi_j^\dagger + i(\xi_j^\ddagger - \xi_j^\dagger)/Q : i = 0, 1, \ldots, Q\}$ constitute a such a partition of $[\xi_j^\dagger, \xi_j^\ddagger]$.

Remark 3. In definition (3) of $\mathscr{L}$, we chose to take its associated subintervals as open rather than closed or "half open, half closed" (i.e., including one endpoint and excluding the other endpoint). This choice was made merely for convenience in the complementary definition (5) of $\mathscr{M}$; as demonstrated in the subsequent discussion, it has no bearing on the validity of the fundamental properties we establish for the majorizing rate function, the majorizing mean-value function, or the NNPPs generated by CIATA-Ph. ◁

Remark 4. In the following development, including the proofs of Theorems 1 and 2, the function $\lambda(t)/\widetilde{\lambda}_Q(t)$ for $t \in [0, S]$ must be measurable and bounded above by 1. As detailed below, the method for constructing $\widetilde{\lambda}_Q(t)$ ensures that $\widetilde{\lambda}_Q(t) > 0$ and $\lambda(t) \leq \widetilde{\lambda}_Q(t)$ for $t \in [0, S]$ so the required properties are guaranteed. ◁

A formal statement of the scheme to compute $\widetilde{\lambda}_Q(t)$ for $t \in [0, S]$ is given in Algorithm 1. For $1 \leq j \leq M$ and $1 \leq i \leq Q_j^\#$, we construct $\widetilde{\lambda}_Q(t)$ on the subinterval $[z_{i-1,j}, z_{i,j}]$ using the golden section search procedure to find the maximum value $\lambda_{i,j}^*$ of $\lambda(t)$ on that subinterval, where $\lambda_{i,j}^* > 0$. Theorem 5.1 and section 5.4 of Simmons (1975) provide the basis for this approach to computing

Algorithm 1 (Constructing the Piecewise-Constant Majorizing Rate Function)

Initialization: Set $\phi = (1+\sqrt{5})/2$, $\delta \leftarrow 10^{-4}$, and set Q and $\{Q_j^{\#} : 1 \le j \le M\}$ depending on the application (see Remark 5).
for $j = 1, \ldots, M$ **do**
  Partition $[\xi_j^{\dagger}, \xi_j^{\ddagger}]$ into $Q_j^{\#}$ equal-length subintervals $\{[z_{i-1,j}, z_{i,j}] : i = 1, \ldots, Q_j^{\#}\}$:
    Set $z_{i,j} \leftarrow \xi_j^{\dagger} + i(\xi_j^{\ddagger} - \xi_j^{\dagger})/Q_j^{\#}$ for $0 \le i \le Q_j^{\#}$.
  for $i = 1, \ldots, Q_j^{\#}$ **do**
    Set $a \leftarrow z_{i-1,j}, b \leftarrow z_{i,j}, y_1 \leftarrow a + (2-\phi)(b-a), y_2 \leftarrow a + (\phi-1)(b-a)$.
    while $(b-a) \ge \delta$ **do**
      if $\lambda(y_1) < \lambda(y_2)$ **then**
        Set $\lambda_{i,j}^{*} \leftarrow \lambda(y_2)$, $a \leftarrow y_1, y_1 \leftarrow y_2$, $y_2 \leftarrow a + (\phi-1)(b-a)$;
      else
        Set $\lambda_{i,j}^{*} \leftarrow \lambda(y_1)$, $b \leftarrow y_2$, $y_2 \leftarrow y_1$, $y_1 \leftarrow a + (2-\phi)(b-a)$.
      end if
    end while
  end for
end for
Assign $\widetilde{\lambda}_Q(t)$ on $\mathcal{L}$ and $\mathcal{M}$ according to Equations (4) and (6), respectively.

$\lambda_{i,j}^{*}$ in practice. Therefore the majorizing rate function is defined on $\mathcal{M}$ as follows:

$$\widetilde{\lambda}_Q(t) = \begin{cases} \sum_{i=1}^{Q_j^{\#}} \sum_{j=1}^{M} \lambda_{i,j}^{*} \mathbb{I}_{(z_{i-1,j}, z_{i,j}]}(t) & \text{for } t \in \mathcal{M} \setminus \{\xi_j^{\dagger} : \\ & j = 1, \ldots, M\}, \\ \lambda_{1,j}^{*} & \text{if } t = \xi_j^{\dagger} \text{ for some} \\ & j \in \{1, \ldots, M\}, \end{cases} \tag{6}$$

where $\mathbb{I}_{(z_{i-1,j}, z_{i,j}]}(t)$ is the indicator function for $(z_{i-1,j}, z_{i,j}]$ so that we have $\mathbb{I}_{(z_{i-1,j}, z_{i,j}]}(t) \equiv 1$ if $z_{i-1,j} < t \le z_{i,j}$ and $\mathbb{I}_{(z_{i-1,j}, z_{i,j}]}(t) \equiv 0$ otherwise. Using the piecewise-constant majorizing rate function defined on $[0, S]$ by (4) and (6), we obtain the easily inverted piecewise-linear majorizing mean-value function as

$$\widetilde{\mu}_Q(t) \equiv \int_0^t \widetilde{\lambda}_Q(y)\mathrm{d}y \text{ for } t \in [0, S]. \tag{7}$$

Remark 5. Assigning a suitable value to Q (and hence assigning suitable values to $\{Q_j^{\#} : 1 \le j \le M\}$) in Algorithm 1 depends on the following: (i) the behavior of the rate function $\lambda(t)$ over the time horizon $[0, S]$, and (ii) the execution time of CIATA-Ph as a function of Q. In Section 4.4, we explain our assignment procedure as applied to the test processes used in the experimental performance evaluation of CIATA-Ph. ◁

3.3. Generating an ERP Yielding an NNPP with the Desired Dispersion Ratio

To construct an ERP that ultimately yields an NNPP with the given mean-value function and asymptotic dispersion ratio, CIATA-Ph uses phase-type distributions for the associated noninitial interrenewal times. Specifically, CIATA-Ph uses the hyperexponential distribution for the case $C \ge 1$ and the hyper-Erlang distribution for the case $0 < C < 1$ as suggested by the example in Section 1.1.1 and recommended by Gerhardt and Nelson (2009).

In previous simulation implementations of CIATA, generation of the ERP's interrenewal times was based on the lognormal distribution (Liu 2013) or the Weibull distribution (Liu et al. 2015). This setup was motivated primarily by the convenience of generating the ERP based on a single type of distribution for all values of C. Ultimately, however, we found that in all the NNPPs tested, the phase-type distributions detailed here consistently yielded faster convergence of $\mathrm{Var}[N_Q(t)]/\mathrm{E}[N_Q(t)]$ to C as $t \to \infty$. If general distributions are used for the noninitial interrenewal times, then they can be approximated by phase-type distributions with simple parameters; for example, Asmussen (1996) developed an effective EM algorithm for fitting phase-type distributions to general distributions.

3.3.1. High Dispersion Ratio ($C \ge 1$).

Hyperexponential distributions have larger coefficients of variation than the exponential distribution. For the *2-phase hyperexponential* (H_2) distribution, with probability $p \in (0,1)$ the upper exponential phase with parameter μ_1 is sampled, and with probability $(1-p)$ the lower exponential phase with parameter μ_2 is sampled. The resulting random variable X has c.d.f. $F_{H_2}(x; p, \mu_1, \mu_2) = 1 - pe^{-\mu_1 x} - (1-p)e^{-\mu_2 x}$ for $x \ge 0$. The first two noncentral moments of X are given by $\mathrm{E}[X] = p/\mu_1 + (1-p)/\mu_2$ and $\mathrm{E}[X^2] = 2p/\mu_1^2 + 2(1-p)/\mu_2^2$, respectively. The squared coefficient of variation of X is $\mathrm{CV}^2[X] = \mathrm{Var}(X)/\mathrm{E}^2[X] = [2p/\mu_1^2 + 2(1-p)/\mu_2^2]/[p/\mu_1 + (1-p)/\mu_2]^2 - 1$. The following proposition characterizes

the desired ERP based on an H_2 distribution. The proof of this result is given in Section S2 of the online supplement.

Proposition 1 (Parameters of the ERP with H_2 interrenewal distribution for $C \geq 1$). *If $C \geq 1$ and*

$$p = \frac{1 + C \pm \sqrt{C^2 - 1}}{2(1 + C)},$$

then with the interrenewal times $X_1^\circ \sim G_e(x) \equiv F_{H_2}[x; 1/2, 2p, 2(1-p)]$ and $\{X_n^\circ : n \geq 2\} \overset{\text{i.i.d.}}{\sim} G(x) \equiv F_{H_2}[x; p, 2p, 2(1-p)]$ for $x \geq 0$, we have $\mathrm{E}[X_2^\circ] = 1$, $\mathrm{Var}[X_2^\circ] = C$, $G(0) = 0$, and $G(x) < 1$ for $x > 0$.

Remark 6. If $C = 1$, then we have $p = 1/2$ and $\mu_1 = \mu_2 = 1$ so that in Proposition 1 the resulting ERP is a rate-1 HPP. ◁

Remark 7. In the case that $C \geq 1$, CIATA-Ph uses the two-phase balanced-means hyperexponential distribution for $\{X_n^\circ : n \geq 2\}$ so that $\mu_1 = 2p$ and $\mu_2 = 2(1-p)$. With this setup, the two solutions to the equations $\mathrm{E}[X_2^\circ] = 1$, $\mathrm{Var}[X_2^\circ] = C$ yield the same interrenewal distribution. Moreover, in our computational experience this setup leads to relatively short warm-up periods beyond which the associated NNPP approximately achieves its asymptotic dispersion ratio. ◁

3.3.2. Low Dispersion Ratio ($C < 1$). Hyper-Erlang distributions have smaller coefficients of variation than the exponential distribution. For the hyper-Erlang distribution, with probability $p \in [0, 1)$ CIATA-Ph samples the upper Erlang distribution with shape parameter $k - 1$ (k is an integer with $k \geq 2$) and scale parameter β; and with probability $(1 - p)$, CIATA-Ph samples the lower Erlang distribution with shape parameter k and scale parameter β. The distribution with shape parameter $k \geq 1$ and scale parameter $\beta > 0$ has c.d.f. $F_{\mathrm{Er}}(x; k, \beta) = \int_0^x \tau^{k-1} \exp(-\tau/\beta)/[(k-1)!\beta^k] \mathrm{d}\tau$ for $x \geq 0$. For a random variable $X \sim F_{\mathrm{Er}}(x; k, \beta)$, the first two noncentral moments are given by $\mathrm{E}[X] = k\beta$ and $\mathrm{E}[Y^2] = k(k+1)\beta^2$, respectively. For any real number r, let $\lceil r \rceil$ denote the ceiling of r. The following proposition characterizes the desired ERP based on a hyper-Erlang distribution. The proof of this result is given in Section S2 of the online supplement.

Proposition 2 (Parameters of the ERP with hyper-Erlang interrenewal distribution for $0 < C < 1$). *If $0 < C < 1$ and we take*

$$k = \lceil 1/C \rceil,\ p = \frac{kC - \sqrt{k(1+C) - k^2 C}}{1 + C},\ \text{and}\ \beta = 1/(k - p),$$

then with the interrenewal times

$$X_1^\circ \sim G_e(x) \equiv x[1 - F_{\mathrm{Er}}(x; k-1, \beta)] + F_{\mathrm{Er}}(x; k, \beta) \text{ for } x \geq 0$$

and

$$\{X_n^\circ : n \geq 2\} \overset{\text{i.i.d.}}{\sim} G(x) \equiv pF_{\mathrm{Er}}(x; k-1, \beta) + (1-p)F_{\mathrm{Er}}(x; k, \beta) \text{ for } x \geq 0,$$

we have $\mathrm{E}[X_2^\circ] = 1$, $\mathrm{Var}[X_2^\circ] = C$, $G(0) = 0$, and $G(x) < 1$ for $x > 0$.

Remark 8. In the case that $C < 1$, for noninitial interrenewal times $\{X_n^\circ : n \geq 2\}$, CIATA-Ph uses the mixture c.d.f. of the form $pF_{\mathrm{Er}}(x; k-1, \beta) + (1-p) \times F_{\mathrm{Er}}(x; k, \beta)$ for two reasons. With this setup and with the choice of the integer $k \geq 2$ such that $k = \lceil 1/C \rceil$, there is a unique feasible solution (β, p) to the equations $\mathrm{E}[X_2^\circ] = 1$, $\mathrm{Var}[X_2^\circ] = C$. Moreover, in our computational experience this setup leads to relatively short warm-up periods beyond which the associated NNPP approximately achieves its asymptotic dispersion ratio. ◁

Remark 9. When C is close to zero so that k is large, it is inefficient to generate samples from the Erlang distributions $F_{\mathrm{Er}}(x; k-1, \beta)$ and $F_{\mathrm{Er}}(x; k, \beta)$ as sums of exponential random variables with scale parameter β. In this case, other algorithms (e.g., ratio-of-uniforms or acceptance–rejection algorithms) can be used to generate samples from these distributions efficiently (Law 2015, sections 8.2.5 and 8.3.4). ◁

3.4. Using CIATA-Ph to Generate NNPPs

A formal statement of CIATA-Ph is given in Algorithm 2 below. Theorem 1 below specifies conditions under which CIATA-Ph delivers an arrival process with the desired mean-value function. Section S1 of the online supplement contains the proof of this result. For clarity and simplicity in the rest of this article, the notation $\{N_Q(t) : 0 \leq t \leq S\}$ is always used to denote the NNPP generated by CIATA-Ph.

Theorem 1. *If $\lambda(t)$ satisfies Assumption 1 and Q is sufficiently large, then the NNPP $\{N_Q(t) : t \in [0, S]\}$ generated by CIATA-Ph has the given mean-value function,*

$$\mathrm{E}[N_Q(t)] = \mu(t) \text{ for } t \in [0, S]. \tag{8}$$

For every $t \in [0, S]$, we let $C_Q(t) \equiv \mathrm{Var}[N_Q(t)]/\mathrm{E}[N_Q(t)] = \mathrm{Var}[N_Q(t)]/\mu(t)$ denote the dispersion ratio at time t for the NNPP generated by CIATA-Ph. In terms of the noncentral moments $\theta_\ell \equiv \mathrm{E}[(X_2^\circ)^\ell]$ for $\ell \geq 1$, we have $C = (\theta_2 - \theta_1^2)/\theta_1^2$. Because $\theta_1 = 1$ and $\theta_3 < \infty$ for both of

Algorithm 2 (Using CIATA-Ph to Generate an NNPP)

1: Construct the majorizing rate function $\widetilde{\lambda}_Q(t)$ given by Equations (4) and (6) using Algorithm 1.
2: **if** $C \geq 1$, **then**
3: Choose the c.d.f.'s $G_e(x)$ and $G(x)$ as in Proposition 1,
4: **else**
5: Choose the c.d.f.'s $G_e(x)$ and $G(x)$ as in Proposition 2.
6: **end if**
7: Set $n \leftarrow 1$, $\ell \leftarrow 0$, $S_0^\circ \leftarrow 0$, $\widetilde{S}_0 \leftarrow 0$, and $S_0 \leftarrow 0$. Generate $X_n^\circ \sim G_e$ and set $S_n^\circ \leftarrow S_{n-1}^\circ + X_n^\circ$.
Set $\widetilde{S}_n \leftarrow \widetilde{\mu}_Q^{-1}(S_n^\circ)$.
8: **while** $\widetilde{S}_n \leq S$ **do**
9: Generate $U_n \sim$ Uniform[0,1].
10: **if** $U_n \leq \lambda(\widetilde{S}_n)/\widetilde{\lambda}_Q(\widetilde{S}_n)$, **then**
11: Set $\ell \leftarrow \ell + 1$ and $S_\ell \leftarrow \widetilde{S}_n$.
12: **end if**
13: Set $n \leftarrow n + 1$. Generate $X_n^\circ \sim G$. Set $S_n^\circ \leftarrow S_{n-1}^\circ + X_n^\circ$ and $\widetilde{S}_n \leftarrow \widetilde{\mu}_Q^{-1}(S_n^\circ)$.
14: **end while**

the interrenewal distributions used in CIATA-Ph, we can define the constant

$$\theta^* \equiv \frac{1}{6} + \frac{C^2}{2} - \frac{\theta_3}{3} \tag{9}$$

that appears in the statement of Theorem 2. Section S1 of the online supplement contains the proof of this theorem.

Theorem 2. *If the given rate and mean-value functions $\lambda(t)$ and $\mu(t)$ satisfy Assumptions 1 and 2, then*

$$\lim_{S\to\infty} \mu(S) = \infty; \tag{10}$$

and the NNPP $\{N_Q(t) : t \in [0, S]\}$ generated by CIATA-Ph has the following properties:

$$\lim_{Q\to\infty} C_Q(t) = C + \theta^*/\mu(t) + o(1)/\mu(t) \text{ for } t \in [0, S], \tag{11}$$

where $o(1)$ denotes a function $\vartheta(t) \xrightarrow[t\to\infty]{} 0$; and we have

$$\lim_{t\to\infty} \lim_{S\to\infty} \lim_{Q\to\infty} C_Q(t) = C. \tag{12}$$

4. Experimental Performance Evaluation

In this section, we discuss a comprehensive performance evaluation of CIATA-Ph. In Section 4.1, we specify the rate function and the asymptotic dispersion ratio for each NNPP used in the experimentation. In Section 4.2, we detail our performance-estimation methods. In Section 4.3, we summarize all the simulation results for CIATA-Ph. In Section 4.4, we explain our method for setting Q. Finally, in Section 4.5, we compare the performance of CIATA-Ph with that of the NNPP-simulation procedures of Gerhardt and Nelson (2009).

4.1. Experimental Setup

The test processes used in the performance evaluation are derived from our earlier work on modeling and simulation of arrival streams exhibiting strong dependencies on the time of day, the day of the week, or the season of the year as well as long-term trends over successive years (Lee et al. 1991, Pritsker et al. 1995). For each NNPP used in the experimental performance evaluation, we choose rate functions of the type *exponential-polynomial-trigonometric with multiple periodicities* (EPTMP), which means they have the form

$$\lambda(t) = \exp\{h(t; m, p, \Theta)\} \text{ with}$$
$$h(t; m, p, \Theta) = \sum_{i=0}^{m} \alpha_i t^i + \sum_{j=1}^{p} \gamma_j \sin(\omega_j t + \phi_j), \quad \text{and} \tag{13}$$
$$\Theta = [\alpha_0, \alpha_1, \ldots, \alpha_m, \gamma_1, \ldots, \gamma_p, \omega_1, \ldots, \omega_p, \phi_1, \ldots, \phi_p,]$$

is the vector of continuous parameters of the designated rate function (Kuhl et al. 1997). The first $m + 1$ terms in Equation (13) define a degree-m polynomial representing a possible long-term evolutionary trend in the arrival rate over time. The next p terms in Equation (13) define the trigonometric functions representing possible periodic effects exhibited by the arrival process. The use of an exponential rate function is a convenient means of ensuring that the instantaneous arrival rate is always positive. In a specific application, we can assign the appropriate degree m for the polynomial rate component as well as the appropriate oscillation amplitude (γ_j), oscillation frequency (ω_j), and phase delay (ϕ_j) for each of the cyclic rate components (if applicable). For more details on estimation and simulation of EPTMP-type rate functions, see Kuhl et al. (1997).

Table 1 displays the parameters for the five test cases of EPTMP-type rate functions that are used in our performance evaluation of CIATA-Ph. All five test cases have a time horizon of $S = 8$ time units and contain two

Table 1. Parameters of the NNPPs with EPTMP-Type Rate Functions That Are Used in the Experimental Performance Evaluation of CIATA-Ph

	Parameter								
Test case	α_0	α_1	α_2	γ_1	ϕ_1	ω_1	γ_2	ϕ_2	ω_2
1	3.6269	–	–	1.0592	−0.6193	6.2831	0.5000	0.5000	12.5664
2	3.6269	0.1000	–	1.0592	−0.6193	6.2831	0.5000	0.5000	12.5664
3	3.6269	−0.1000	0.0200	1.0592	−0.6193	6.2831	0.5000	0.5000	12.5664
4	3.6269	–	–	1.0592	−0.6193	0.3927	0.5000	0.5000	0.7854
5	3.6269	−0.1000	0.0200	1.0592	−0.6193	0.3927	0.5000	0.5000	0.7854

nested cyclic effects. In cases 1–3, the first cyclic effect has a period of 1 time unit ($\omega_1 = 2\pi$) and the second cyclic effect has a period of 0.5 time units ($\omega_2 = 4\pi$). In cases 4 and 5, the first cyclic effect has a period of 16 time units ($\omega_1 = \pi/8$); and the second cyclic effect has a period of 8 time units ($\omega_2 = \pi/4$). Cases 1 and 4 do not contain a general trend over time. Cases 2, 3, and 5 contain general trends that are represented by polynomials of degrees 1, 2, and 2, respectively. In the performance evaluation of CIATA-Ph for cases 1–3, the asymptotic dispersion ratio is assigned the values $C = 0.2, 0.8, 1.5$, and 10.0. Cases 4 and 5 demonstrate the performance of CIATA-Ph in test processes for which the rate function changes relatively slowly over the time horizon. In cases 4 and 5, the asymptotic dispersion ratio is assigned the values $C = 0.2$ and 1.5.

To evaluate the performance of CIATA-Ph, we carry out a *metaexperiment* for each procedure consisting of $R = 100$ independent basic experiments, and in each basic experiment, we execute $K = 200$ independent replications of CIATA-Ph to generate 200 realizations of the NNPP $\{N_Q(t) : t \in [0, S]\}$ defined by each of the relevant test cases in Table 1 with each of the associated values of the dispersion ratio. For each case, this experimental setup is designed to yield valid point and confidence interval (CI) estimators of the mean-value function $\mu(t)$ and the dispersion-ratio function $C_Q(t)$ for selected values of $t \in (0, S]$, enabling quantitative and visual assessment of the extent to which the simulation-generated NNPP satisfies the desired conditions (1) and (2). In the next section, we detail the statistical methods used in the experimental performance evaluation of CIATA-Ph.

4.2. Performance Estimation Methods

On the basis of 100 replications of each basic experiment, for each case, we compute point and 95% CI estimators of $\mu(t_i)$ and $C_Q(t_i)$ at the observation times $\{t_i \equiv i\zeta : i = 1, \ldots, \mathbb{T}\}$, where the spacing ζ between successive observation times is chosen so that $\mathbb{T} = S/\zeta$ is a positive integer. For all the experiments in this section, we set $\zeta = 0.2$ so that $\mathbb{T} = 40$.

The following methods are used to compute the point and CI estimators required for the experimental performance evaluation. On the kth realization of a given NNPP (case) in the rth basic experiment, we let $N_Q(t; r, k)$ denote the number of accepted arrivals up to time $t \in [0, S]$, where $r = 1, \ldots, R$, and $k = 1, \ldots, K$. In the rth basic experiment, the estimators of the mean-value and variance functions for the given NNPP are the sample statistics

$$\left.\begin{aligned} &\overline{N}_Q(t; r) = \frac{1}{K}\sum_{k=1}^{K} N_Q(t; r, k) \\ &\widehat{\mathrm{Var}}[N_Q(t; r, 1)] \\ &\quad = \frac{1}{K-1}\sum_{k=1}^{K}[N_Q(t; r, k) - \overline{N}_Q(t; r)]^2 \end{aligned}\right\} \text{for } t \in (0, S] \text{ and } r = 1, \ldots, R.$$

Note that the statistic $\widehat{\mathrm{Var}}[N_Q(t; r, 1)]$ is computed from the results generated in the rth basic experiment, and it is an estimator of the variance of the random variable $N_Q(t)$ observed on one replication of the given arrival process $\{N_Q(y) : y \in [0, S]\}$.

The estimated dispersion-ratio function at time t is

$$\widehat{C}_Q(t; r) = \frac{\widehat{\mathrm{Var}}[N_Q(t; r, 1)]}{\overline{N}_Q(t; r)} \text{ for } t \in (0, S] \text{ and } r = 1, \ldots, R.$$

On the basis of the entire metaexperiment, the overall estimators of $\mu(t)$, $\mathrm{Var}[N_Q(t)]$, $C_Q(t)$, and $\mathrm{Var}[\widehat{C}_Q(t; 1)]$ are, respectively,

$$\begin{aligned} \widehat{\mu}_Q(t) &= \frac{1}{RK}\sum_{r=1}^{R}\sum_{k=1}^{K} N_Q(t; r, k), \\ \widehat{\mathrm{Var}}[N_Q(t)] &= \frac{1}{R}\sum_{r=1}^{R}\widehat{\mathrm{Var}}[N_Q(t; r, 1)], \qquad (14) \\ \widehat{C}_Q(t) &= \frac{1}{R}\sum_{r=1}^{R}\widehat{C}_Q(t; r), \text{ and} \\ \widehat{\mathrm{Var}}[\widehat{C}_Q(t; 1)] &= \frac{1}{R-1}\sum_{r=1}^{R}[\widehat{C}_Q(t; r) - \widehat{C}_Q(t)]^2. \qquad (15) \end{aligned}$$

Therefore, for $u \in (0,1)$, the approximate $100(1-u)\%$ CI estimators of $u(t)$ and $C_Q(t)$ are, respectively,

$$\widehat{\mu}_Q(t) \pm z_{1-u/2}\sqrt{\widehat{\mathrm{Var}}[N_Q(t)]/R} \quad \text{and}$$

$$\widehat{C}_Q(t) \pm z_{1-u/2}\sqrt{\widehat{\mathrm{Var}}[\widehat{C}_Q(t;1)]/R} \quad \text{for } t \in (0,S], \tag{16}$$

where $z_{1-u/2}$ is the $1-u/2$ quantile of the standard normal distribution.

To provide additional criteria for evaluating the performance of CIATA-Ph, we examine the relative errors of the point estimators $\widehat{\mu}_Q(t)$ and $\widehat{C}_Q(t)$ (i.e., their respective percentage deviations from the associated target values $\mu(t)$ and C), which we call *closeness* measures, and we study the sensitivity of these closeness measures to the dispersion ratio C and to the number of majorizing intervals Q. We define the *average percentage discrepancy* (APD) $\Delta_{\mathbb{T}}(\cdot)$ and *maximum percentage discrepancy* (MPD) $\Delta^*_{\mathbb{T}}(\cdot)$ as two measures of closeness:

$$\left.\begin{aligned}\Delta_{\mathbb{T}}(\widehat{\mu}_Q) &= \frac{1}{\mathbb{T}}\sum_{i=1}^{\mathbb{T}} D_i(\widehat{\mu}_Q), \quad \Delta^*_{\mathbb{T}}(\widehat{\mu}_Q) = \max_{1\le i\le \mathbb{T}}\{D_i(\widehat{\mu}_Q)\},\\ \Delta_{\mathbb{T}}(\widehat{C}_Q) &= \frac{1}{\mathbb{T}}\sum_{i=1}^{\mathbb{T}} D_i(\widehat{C}_Q), \quad \Delta^*_{\mathbb{T}}(\widehat{C}_Q) = \max_{1\le i\le \mathbb{T}}\{D_i(\widehat{C}_Q)\},\end{aligned}\right\} \tag{17}$$

where the percentage discrepancy $D_i(\cdot)$ is the percentage difference of the estimators $\widehat{\mu}_Q(t_i)$ and $\widehat{C}_Q(t_i)$ from their respective target values $\mu(t_i)$ and C at each observation time t_i:

$$D_i(\widehat{\mu}_Q) = \left|\frac{\widehat{\mu}_Q(t_i)-\mu(t_i)}{\mu(t_i)}\right| \times 100\% \quad \text{and}$$

$$D_i(\widehat{C}_Q) = \left|\frac{\widehat{C}_Q(t_i)-C}{C}\right| \times 100\% \quad \text{for } i = 1,\ldots,\mathbb{T}.$$

Finally, we compute the approximate 95% CI estimator for the expected value of $\Delta_{\mathbb{T}}(\widehat{C}_Q)$,

$$\Delta_{\mathbb{T}}(\widehat{C}_Q) \pm z_{1-\xi/2}\frac{S_\Delta(\widehat{C}_Q)}{\sqrt{\mathbb{T}}}, \quad \text{where}$$

$$S_\Delta(\widehat{C}_Q) = \left\{\frac{1}{\mathbb{T}-1}\sum_{i=1}^{\mathbb{T}}[D_i(\widehat{C}_Q)-\Delta_{\mathbb{T}}(\widehat{C}_Q)]^2\right\}^{1/2}, \tag{18}$$

and $\xi = 0.05$. For the APD estimator $\Delta_{\mathbb{T}}(\widehat{\mu}_Q)$, the standard-deviation estimator $S_\Delta(\widehat{\mu}_Q)$, and the CI estimator of $\mathrm{E}[\Delta_{\mathbb{T}}(\widehat{\mu}_Q)]$ are defined similarly.

4.3. Performance of CIATA-Ph

On the basis of Equations (14)–(18), in this section, we report results for CIATA-Ph in cases 1, 3, and 4, with the target dispersion ratios $C = 0.2$ and 1.5. In the online supplement, we summarize the results for CIATA-Ph in cases 2 and 5 and with other dispersion ratios (i.e., $C = 0.8$ and 10). For case 1 and $t \in [0,8]$, we plot the following in Figure 2: (i) the given rate function $\lambda(t)$ (dashed curve) and the corresponding majorizing step rate function $\widetilde{\lambda}_Q(t)$ (solid step function); (ii) the 95% CI estimators of the mean-value function $\mu(t)$ (vertical bars) superimposed on $\mu(t)$ (solid curve); and (iii) the 95% CI estimators for the dispersion-ratio function $C_Q(t)$ (vertical bars) superimposed on the value of C (horizontal line). Here $Q = 160$ so the majorizing step size is $S/Q = 0.05$.

Figures 3 and 4 depict the CIATA-Ph–generated results for cases 3 and 4, respectively, based on the same layout used in Figure 2. Also, see Tables S7–S10 and Tables S23–S26 in the online supplement for the values of the CI estimators for $\mu(t)$ and $C_Q(t)$ depicted in Figures 2–4.

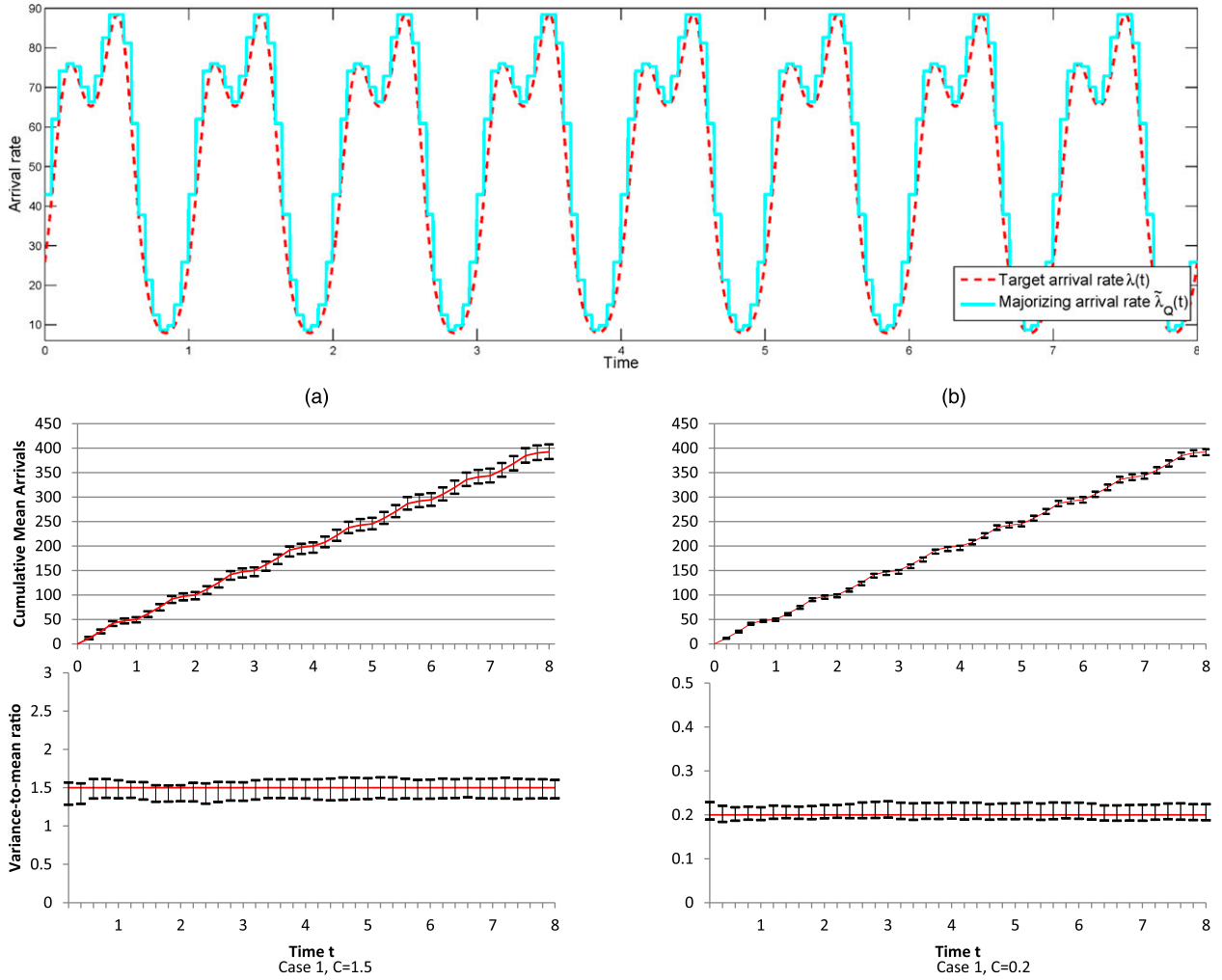
The bottom panels of Figures 2–4 exemplify the warm-up periods required for a CIATA-Ph–generated arrival process to achieve approximate convergence to the desired asymptotic dispersion ratio. Except for "extreme" situations such as $C = 0.2$ or $C = 10$, we found that the time-dependent dispersion ratio usually achieved approximate convergence to C within a relatively short warm-up period; consequently, we concluded that CIATA-Ph could deliver reasonably fast convergence to a given nonextreme dispersion ratio. See Table S5 in the online supplement for a comparison of the warm-up times across different dispersion ratios.

Table 2 provides the closeness measures $\Delta_{\mathbb{T}}(\widehat{\mu}_Q)$, $\Delta^*_{\mathbb{T}}(\widehat{\mu}_Q)$, $\Delta_{\mathbb{T}}(\widehat{C}_Q)$, and $\Delta^*_{\mathbb{T}}(\widehat{C}_Q)$ and the approximate 95% CIs for $\mathrm{E}[\Delta_{\mathbb{T}}(\widehat{\mu}_Q)]$ and $\mathrm{E}[\Delta_{\mathbb{T}}(\widehat{C}_Q)]$ for the cases 1, 3, and 4 with $C = 0.2$ and 1.5. For the situations in which $C = 1.5$, Table 2 shows that the APD and MPD for the mean-value function of the CIATA-Ph–generated NNPP are less than 1% and 3%, respectively; and the APD and MPD for the dispersion ratio of the CIATA-Ph–generated NNPP are less than 2% and 8%, respectively. For the more extreme situations in which $C = 0.2$, we see that the APD and MPD for the mean-value function of the CIATA-Ph–generated NNPP are less

Table 2. CIATA-Ph–Based Closeness Measures for Cases 1, 3, and 4

C	$\Delta_{\mathbb{T}}(\widehat{\mu}_Q)$	$\Delta^*_{\mathbb{T}}(\widehat{\mu}_Q)$	$\Delta_{\mathbb{T}}(\widehat{C}_Q)$	$\Delta^*_{\mathbb{T}}(\widehat{C}_Q)$
Case 1				
0.2	0.918% ± 0.197%	2.538%	3.530% ± 0.266%	6.150%
1.5	0.819% ± 0.176%	2.463%	1.901% ± 0.35%	5.233%
Case 3				
0.2	0.056% ± 0.009%	0.22%	1.304% ± 0.202%	5.00%
1.5	0.163% ± 0.038%	0.56%	1.196% ± 0.196%	3.89%
Case 4				
0.2	0.037% ± 0.025%	0.70%	3.297% ± 0.900%	20.93%
1.5	0.124% ± 0.053%	1.21%	0.913% ± 0.293%	7.97%

Figure 2. (Color online) Performance of CIATA-Ph for Case 1 with Dispersion Ratio $C = 1.5, 0.2$ and with $Q = 160$



Note. The top panel shows $\widetilde{\lambda}_Q(t)$ versus $\lambda(t)$; the middle panel shows 95% CIs for $\mu(t)$ when $C = 1.5$ (left side) and $C = 0.2$ (right side); and the bottom panel shows comparable 95% CIs for $C_Q(t)$.

than 1% and 3%, respectively; and the APD and MPD for the dispersion ratio of the CIATA-Ph–generated NNPP are less than 4% and 21%, respectively. We judged these results to provide good evidence of the effectiveness of the CIATA-Ph–based majorizing approximations of the given rate and mean-value functions together with the resulting approximation of the given dispersion ratio. See pages S-18 and S-19 in the online supplement for case-1 results with $C = 20$, 15, 10, and 0.8 and for a discussion of these results.

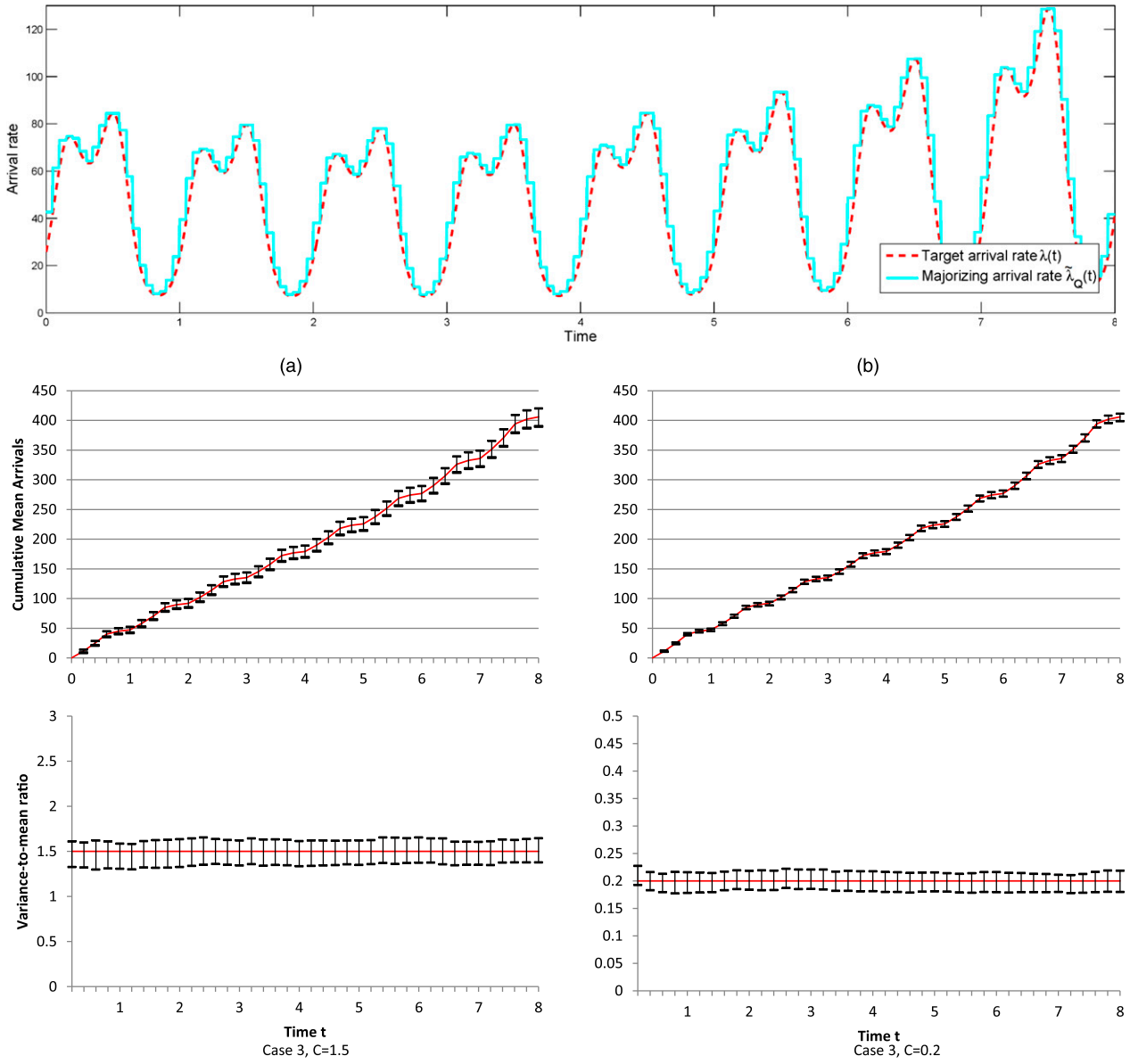
4.4. Assigning a Value to Q When Using CIATA-Ph

Assigning a suitable value to Q in an application of CIATA-Ph depends on the behavior of the given rate function $\lambda(t)$ over the time horizon $[0, S]$ and on the execution time of CIATA-Ph as a function of Q. This assignment often requires a pilot study to carry out the following steps for trial values of Q: (i) visual inspection of a graph of $\widetilde{\lambda}_Q(t)$ superimposed on $\lambda(t)$ for $t \in [0, S]$; (ii) sensitivity analysis of selected CIATA-Ph–generated performance measures as they depend on Q; and (iii) sensitivity analysis of the execution time of CIATA-Ph as it depends on Q. For concreteness, we discuss how these steps were carried out for case 1 with $C = 1.5$ and $C = 0.2$.

4.4.1. Accuracy of CIATA-Ph.

To perform step (i), we examined the top panel of Figure 5, which shows $\widetilde{\lambda}_Q(t)$ superimposed on $\lambda(t)$ for $0 \le t \le S = 8$ and $Q = 80$. In our experience, if $\lambda(t)$ is smooth (differentiable) except at a finite set of times in $[0, S]$, then a good starting point for step (i) is to take $Q \in [80, 320]$. Such an initial value of Q ensures that Algorithm 1 will deliver a partition of $[0, S]$ for which Assumption 1 is satisfied except on relatively small subintervals whose interior contains an isolated local minimum of $\lambda(t)$. Because $\lambda(t)$ is not quasiconcave over such a subinterval, Algorithm 1 is not guaranteed to

Figure 3. (Color online) CIATA-Ph Performance for Case 3 with Dispersion Ratio $C = 1.5, 0.2$ and with $Q = 160$



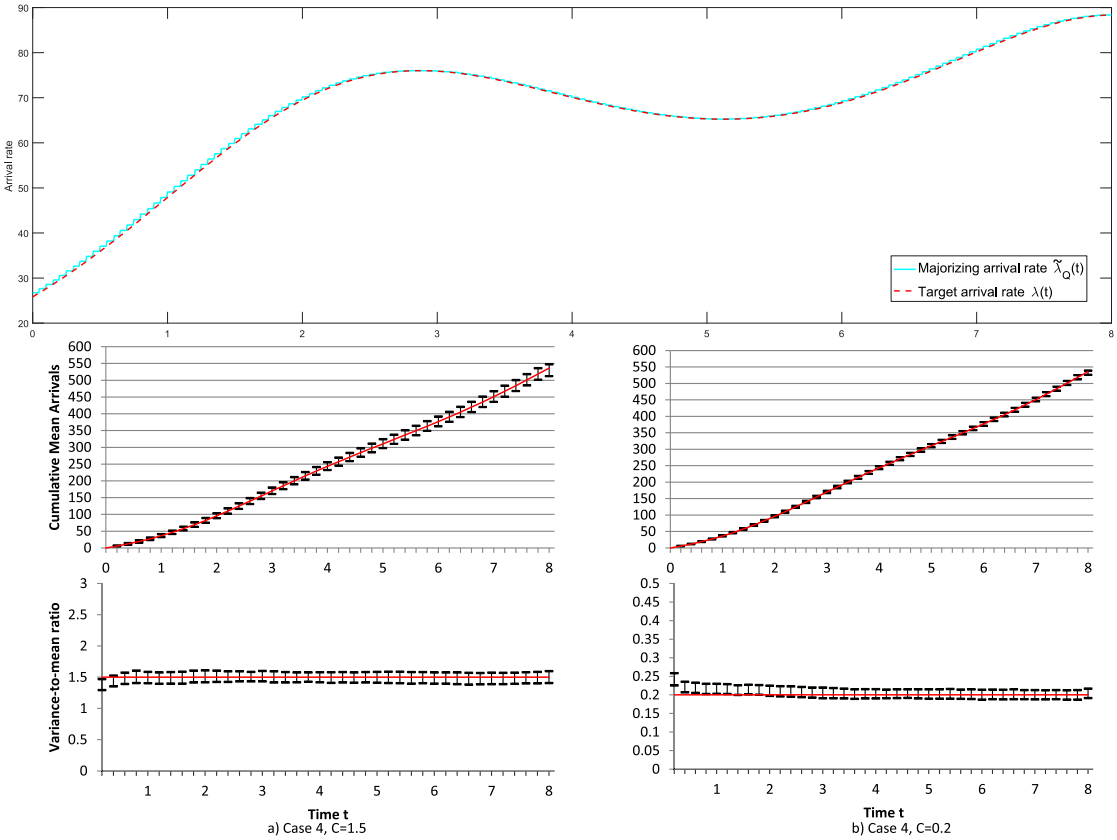
Note. The top panel shows $\widetilde{\lambda}_Q(t)$ versus $\lambda(t)$; the middle panel shows 95% CIs for $\mu(t)$ when $C = 1.5$ (left side) and $C = 0.2$ (right side); and the bottom panel shows comparable 95% CIs for $C_Q(t)$.

deliver a true majorizing rate function satisfying the required conditions $\widetilde{\lambda}_Q(t) \geq \lambda(t)$ and $\widetilde{\lambda}_Q(t) > 0$ over that subinterval. However, in our experience with Algorithm 1, these conditions have always been satisfied everywhere in $[0, S]$ provided $Q \geq 40$.

Visual inspection of Figure 5 reveals that $\widetilde{\lambda}_Q(t)$ is a true majorizing approximation to $\lambda(t)$ for all $t \in [0, S]$ when $Q = 80$, and Figure S1 in the online supplement shows that this conclusion also holds when $Q = 40$. However, Figure S1 shows that setting $Q = 40$ results in the undesirable condition $\widetilde{\lambda}_Q(t) \gg \lambda(t)$ over a substantial part of the time horizon. In this situation, the dispersion ratio of the CIATA-Ph–generated process $\{N_Q(t) : t \in [0, S]\}$ can differ significantly from C. This point is elaborated in the next paragraph. Thus we judged that we should take $Q \geq 80$ in performing steps (i) and (ii).

To carry out step (ii) for $Q = 80$ and $Q = 160$, we examined the middle and bottom panels of Figures 5 and 2, respectively. For both values of Q, clearly $\widehat{\mu}_Q(t)$ is an accurate estimator of $\mu(t)$ for all $t \in [0, S]$ and for both $C = 1.5$ and $C = 0.2$, which suggests that taking $Q \geq 80$ is sufficient to ensure the conclusion of Theorem 1. Similarly, when $Q = 160$, we see that $\widehat{C}_Q(t)$ is a reasonably accurate estimator of C for all $t \in [0, S]$; and this result agrees with the conclusion of Theorem 2. On the other hand, when $Q = 80$, we see that $\widehat{C}_Q(t)$

Figure 4. (Color online) CIATA-Ph Performance for Case 4 with Dispersion Ratio $C = 1.5, 0.2$ and with $Q = 160$



Note. The top panel shows $\tilde{\lambda}_Q(t)$ versus $\lambda(t)$; the middle panel shows 95% CIs for $\mu(t)$ when $C = 1.5$ (left side) and $C = 0.2$ (right side); and the bottom panel shows comparable 95% CIs for $C_Q(t)$.

consistently underestimates C for $C = 1.5$, and $\widehat{C}_Q(t)$ substantially overestimates C for $C = 0.2$. When $Q = 40$, the bottom panel of Figure S1 shows similarly poor behavior of $\widehat{C}_Q(t)$. On the basis of the empirical evidence from steps (i) and (ii), we decided to take $Q \geq 160$ so as to ensure that for the CIATA-Ph–generated arrival process $\{N_Q(t) : t \in [0, S]\}$, the statistic $\widehat{C}_Q(t), t \in [0, S]$, is a reasonably close approximation to C.

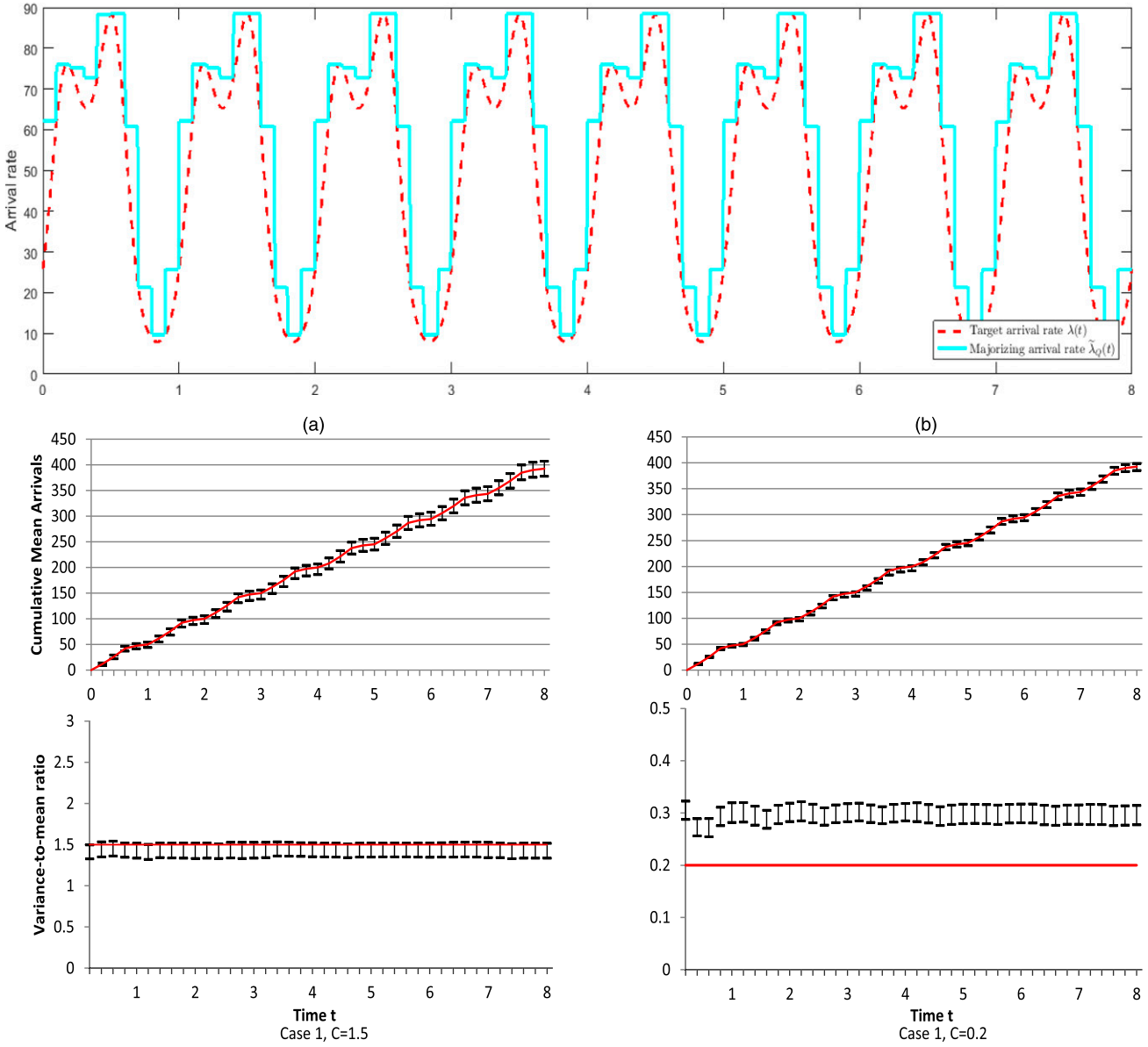
4.4.2. Computational Complexity of CIATA-Ph. By performing step (iii) of the pilot study for case 1, we finalized the assignment of Q. Before discussing the specifics of that assignment, first we formulate a general "big O" upper bound on $\mathcal{T}(Q, S, \lambda)$, the expected computation time for CIATA-Ph to generate a single realization of the NNPP $\{N_Q(t) : t \in [0, S]\}$ as Q or S increases. Recall that λ^* is a finite upper bound on $\lambda(t)$ for $t \geq 0$ as specified by Assumption 1. For the purpose of deriving the desired upper bound on $\mathcal{T}(Q, S, \lambda)$, we may assume that the points $\{z_i = iS/Q : i = 0, 1, \ldots, Q\}$ constitute a partition of $[0, S]$ satisfying Assumption 1.

To account for each step of Algorithm 2, we define the following upper bounds: (i) w_{ERP} bounds the computation time to generate one new renewal epoch S_n° in the ERP; (ii) w_{check} bounds the time to check the condition $S_n^\circ \leq \tilde{\mu}_Q(z_i)$ for one transformed subinterval $[\tilde{\mu}_Q(z_{i-1}), \tilde{\mu}_Q(z_i)] \subset [0, \tilde{\mu}_Q(S)]$; (iii) w_{invert} bounds the time to perform one inversion $\tilde{S}_n \leftarrow \tilde{\mu}_Q^{-1}(S_n^\circ)$; and (iv) w_{thin} bounds the time to accept or reject one arrival time $\tilde{S}_n$. Lemma 1 in the online supplement ensures that $\tilde{\mu}_Q(S) \leq \lambda^* S$ so we have

$$\mathcal{T}(Q, S, \lambda) \leq (w_{\text{ERP}} + w_{\text{invert}} + w_{\text{thin}})\lambda^* S + w_{\text{check}}\lambda^* SQ = O(SQ) \text{ as } Q \to \infty \text{ or } S \to \infty. \tag{19}$$

On the basis of Equation (19) with a given value of S, we see that a relatively large value of Q (e.g., $Q > 1{,}000$ in case 1) will make CIATA-Ph computationally inefficient because such a value of Q will require checking an excessive number of transformed subintervals of $[0, \tilde{\mu}_Q(S)]$ in order to locate the transformed subinterval $[\tilde{\mu}_Q(z_{i-1}), \tilde{\mu}_Q(z_i)]$ in which to perform the inversion $\tilde{S}_n \leftarrow \tilde{\mu}_Q^{-1}(S_n^\circ)$ required in Steps 7 and 13 of Algorithm 2. On the other hand, a relatively small value of Q (e.g., $Q < 10$ in case 1) will be at best computationally inefficient if $\tilde{\lambda}_Q(t) \gg \lambda(t)$ over much of the time horizon so that in many iterations of Step 10 in Algorithm 2, the

Figure 5. (Color online) CIATA-Ph Performance for Case 1 with Dispersion Ratio $C = 1.5, 0.2$ and with $Q = 80$



Note. The top panel shows $\tilde{\lambda}_Q(t)$ versus $\lambda(t)$; the middle panel shows 95% CIs for $\mu(t)$ when $C = 1.5$ (left side) and $C = 0.2$ (right side); and the bottom panel shows comparable 95% CIs for $C_Q(t)$.

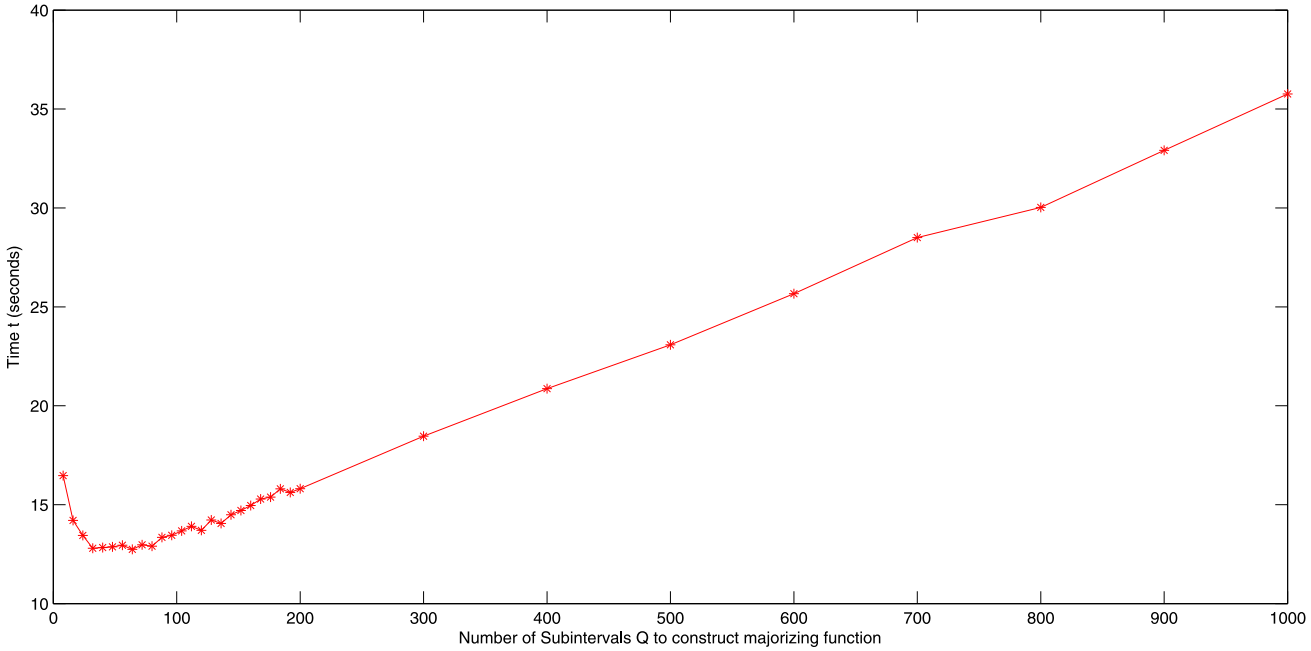
acceptance probability $\lambda(\tilde{S}_n)/\tilde{\lambda}_Q(\tilde{S}_n) \ll 1$. In this situation, the **while** loop in steps 8–14 of Algorithm 2 must be repeated excessively often just to generate one new arrival in the process $\{N_Q(t) : t \in [0, S]\}$. Consequently, there must be some values of Q, neither too small nor too large, that achieve the minimum computational complexity for CIATA-Ph.

To carry out step (iii) of the pilot study for case 1, we recorded and plotted in Figure 6 the execution time of CIATA-Ph for $Q = 8, 16, 24, \ldots, 200, 300, \ldots, 1{,}000$ in case 1 with $C = 1.5$. In Figure 6, we see that the execution time for small values of Q (e.g., 8 and 16) is relatively large, and it initially decreases with increasing values of Q, achieving a local minimum for Q in the approximate range $30 \leq Q \leq 80$. For values of $Q > 80$, the execution time increases as an approximately linear function of Q, which is consistent with Equation (19). Moreover, we see that the execution times are nearly equal for $Q = 16$ and $Q = 160$. Evaluating the trade-off between the accuracy and computational complexity of CIATA-Ph, we concluded that the assignment $Q = 160$ was appropriate for our test processes.

4.5. Comparison of CIATA-Ph with the Simulation Procedures of Gerhardt and Nelson (2009)

In Section 2, we summarized the procedures of Gerhardt and Nelson (2009) for simulating an NNPP

Figure 6. (Color online) Computation Time (in Seconds) for CIATA-Ph in Case 1 with $C = 1.5$



with given rate function, mean-value function, and asymptotic dispersion ratio by inversion of the mean-value function and by thinning. In this section, we compare the performance of CIATA-Ph with both of the procedures of Gerhardt and Nelson (2009).

4.5.1. Thinning Procedure of Gerhardt and Nelson (2009). In Section 2, we mentioned that although the thinning procedure of Gerhardt and Nelson (2009) is guaranteed to achieve the desired mean-value function $\mu(t)$ for all t, it is not guaranteed to achieve the asymptotic dispersion ratio C even for large t. We illustrate this phenomenon in the case 1 test process for scenarios in which $C = 10$ and $C = 1.5$. In Figure 7, we plot the estimated dispersion-ratio function $\widehat{C}(t)$ (see Section 4.2) for the thinning procedure of Gerhardt and Nelson (labeled GNTA) and for CIATA-Ph. Figure 7 shows that GNTA does not accurately yield the desired dispersion ratio C; instead $\widehat{C}(t)$ quickly settles down to a value that is significantly lower than the target. On the basis of result 2.3 of Gerhardt and Nelson (2009), this phenomenon appears to be caused at least in part by the relatively small value for the ratio $[\mu(S)/S]/\widetilde{\lambda} \approx [390/8]/88 \approx 0.55$. On the other hand, the results for CIATA-Ph are much more accurate because the majorizing rate function $\widetilde{\lambda}_Q(t)$ is a much closer approximation to the target rate function $\lambda(t)$ for $t \in [0, S]$ than is achieved by GNTA with the upper bound $\widetilde{\lambda}$.

To gain more understanding of the advantages of CIATA-Ph and its sensitivity with respect to $[\mu(S)/S]/\widetilde{\lambda}$, we consider sinusoidal arrival rates

$$\lambda(t) = \overline{\lambda}[1 + \gamma\sin(t)] \text{ for } t \in [0, S] \tag{20}$$

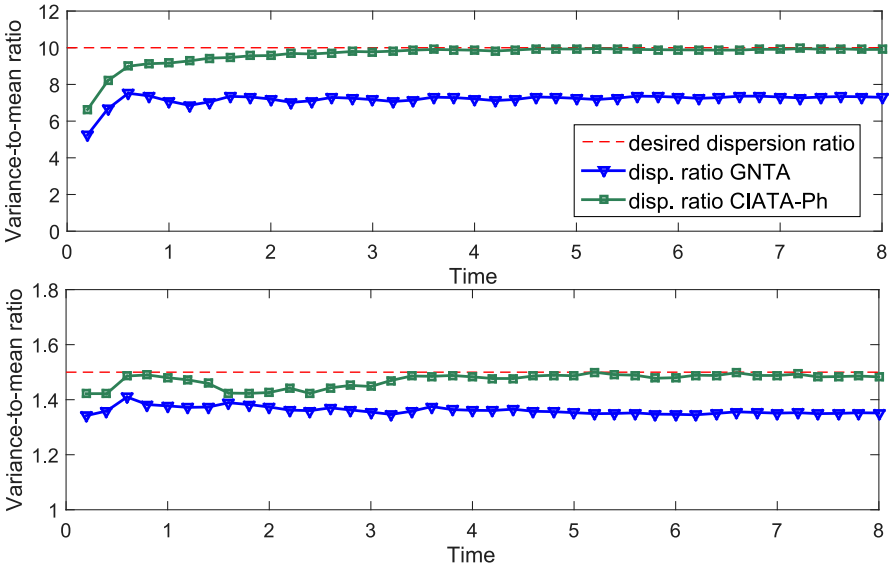
with average arrival rate $\overline{\lambda}$ and relative oscillation amplitude $\gamma \in (0, 1)$. Let the upper bound rate be $\widetilde{\lambda} = \overline{\lambda}(1 + \gamma)$. Thus we have

$$\eta \equiv \frac{\mu(S)/S}{\widetilde{\lambda}} = \frac{\left\{\int_0^S \overline{\lambda}[1 + \gamma\sin(t)]\mathrm{d}t\right\}\Big/S}{\overline{\lambda}(1 + \gamma)} = 1 - \frac{\gamma}{1 + \gamma}\left[\frac{S + \cos(S) - 1}{S}\right] < 1 \text{ for all } S > 0, \tag{21}$$

where the final inequality in Equation (21) follows easily by checking that the inequality $S + \cos(S) > 1$ holds not only for all $S \in (0, 2]$ but also for all $S > 2$.

Figure 8 depicts the results of comparing the performance of CIATA-Ph with that of GNTA for the following parameter settings: average arrival rate $\overline{\lambda} = 50$; time horizon $S = 8$; dispersion ratio $C = 0.2, 1.5, 4, 10$; and relative oscillation amplitude $\gamma = 0.2, 0.8$ (so that $\eta = 0.84, 0.56$). Figure 8 shows that CIATA-Ph delivers a dispersion ratio (dashed-and-dotted line connected by squares) that is substantially closer to the target value (dashed line) than the dispersion ratio delivered by GNTA (solid line connected by triangles). Moreover, the dispersion ratio delivered by CIATA-Ph is nearly constant while that delivered by GNTA is significantly time unstable, exhibiting a highly time-varying pattern following the time variability of the arrival rate.

Figure 7. (Color online) Performance Comparison of GNTA and CIATA-Ph with Case-1 Arrival Rates



Note. The top panel shows the sample mean estimators of $C_Q(t)$ for selected values of $t \in (0, 8]$ when $C = 10$; and the bottom panel shows the corresponding statistics when $C = 1.5$.

In Table 3 (top half), we give the time average of the dispersion function in $[0, S]$ ($S = 8$) for both CIATA-Ph and GNTA; and we compute the average relative errors of both methods (i.e., the average over time of the ratio $|\text{performance} - \text{target}|/\text{target}$). In this half of Table 3, we consider the following parameter settings: $\overline{\lambda} = 50$; $\gamma = 0.2, 0.5, 0.8$ (so that $\eta = 0.84, 0.67, 0.56$); and $C = 0.2, 0.8, 1.1, 1.2, 1.5, 4, 10$. Table 3 (bottom half) repeats the same experiment with the smaller average arrival rate $\overline{\lambda} = 10$. For the smaller arrival rate, we observe performance degradation for both CIATA-Ph and GNTA. However, CIATA-Ph continues to perform significantly better than GNTA.

To provide a picture of when and how CIATA-Ph performs better than GNTA, we summarize below:

- Sign of error in estimating C: GNTA's dispersion ratio tends to be lower (higher) than the target dispersion ratio when $C > 1$ ($C < 1$), while CIATA-Ph gives much more accurate performance.
- Sensitivity to η: The performance of GNTA degrades when the arrival rate exhibits a bigger fluctuation around the mean arrival rate. Specifically, GNTA performs poorly as the relative oscillation amplitude γ increases (or equivalently, the ratio η in Equation (21) decreases). On the other hand, CIATA-Ph continues to achieve good performance for all values of η.
- Sensitivity to C: CIATA-Ph gives more benefits than GNTA when C is far from 1 (e.g., when $C > 1.5$ or $C < 0.8$). In practice, we recommend using CIATA-Ph rather than GNTA when the target arrival process is highly overdispersed or underdispersed.

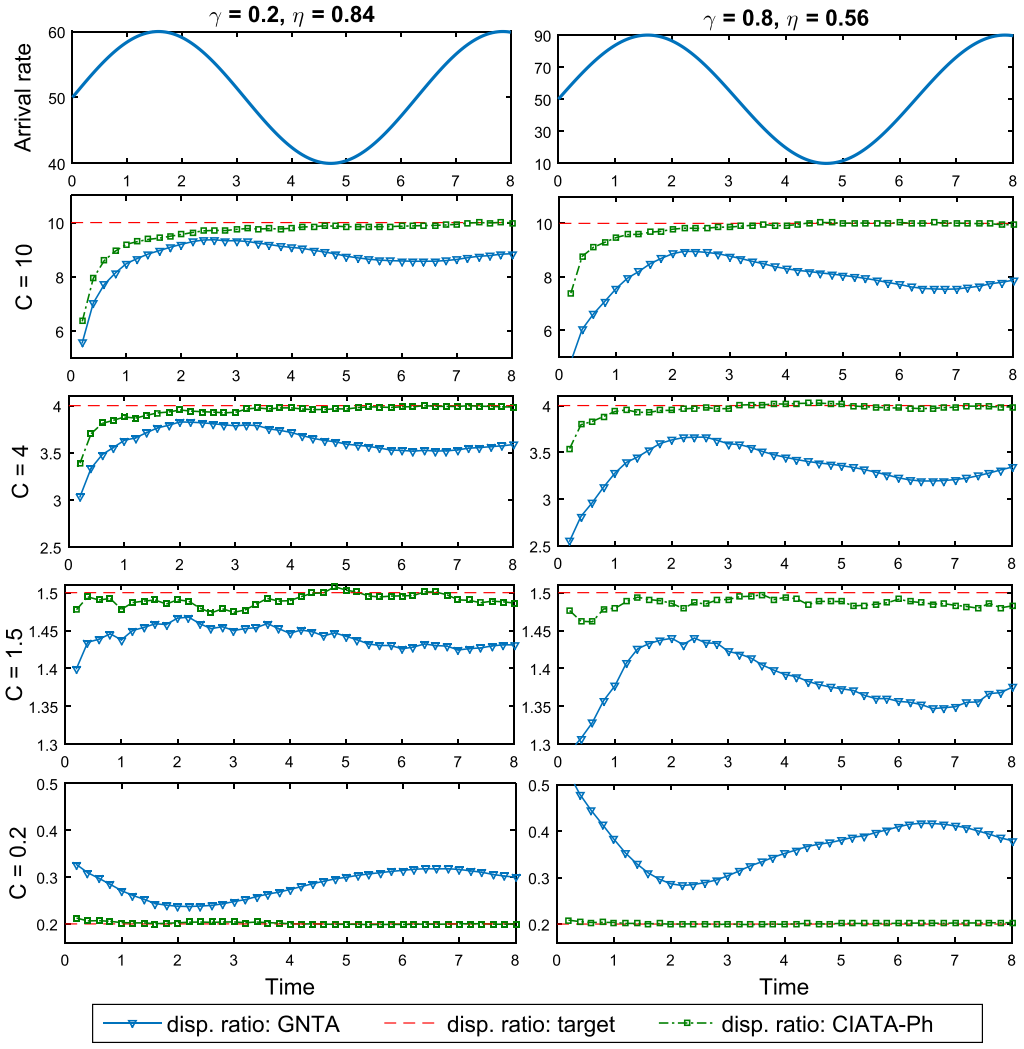
4.5.2. Inversion Procedure of Gerhardt and Nelson (2009).

Although the inversion procedure of Gerhardt and Nelson (2009) is guaranteed to achieve the target mean-value function for all t as well as convergence to the target dispersion ratio for large t, it can be computationally inefficient if the target mean-value function is difficult to invert analytically or numerically. We illustrate this phenomenon in the case 1 test process for the scenario in which $C = 1.5$. In Table 4, we give the execution times of the inversion algorithm (labeled GNIA) and of CIATA-Ph based on $Q = 50$, 100, 150, 200, 250, and 300 subintervals. All execution times are based on running MATLAB on a 2.66-GHz Intel 2 Quad processor with 3.25-GB RAM, running Windows XP Professional. Although the results in Table 4 will vary depending on the software and hardware used to perform the comparison, it is clear that CIATA-Ph significantly outperforms inversion in this test process.

5. Conclusions and Recommendations

The main finding of this work is that CIATA and CIATA-Ph can be effective for modeling and simulating an NNPP that is specified solely in terms of given rate and mean-value functions as well as a given asymptotic dispersion ratio. In many application domains, we have often found that these characteristics represent the only available information, at least in initial simulation studies. In such situations, we believe that the proposed techniques based on CIATA and CIATA-Ph have definite advantages over competing procedures, in particular with respect to general applicability, the

Figure 8. (Color online) Performance Comparison of GNTA and CIATA-Ph with Sinusoidal Arrival Rates in Equation (20) with $\bar{\lambda} = 50$, $\gamma = 0.2, 0.8$ ($\eta = 0.84, 0.56$), and $C = 0.2, 1.5, 4, 10$



speed of convergence to the desired asymptotic dispersion ratio, and computational efficiency.

There are several promising directions for future work on modeling and simulation of NNPPs. With respect to issues raised in this article that require further investigation, we are particularly focused on the formulation of a robust rule of thumb for setting Q given the asymptotic dispersion ratio C, the length S of the time horizon, and the behavior of the mean-value value function $\mu(t)$ over that time horizon so as to minimize the average warm-up period required to achieve approximate convergence to C. We are also pursuing a comprehensive experimental comparison of CIATA-Ph to determine more precisely the types of applications in which one procedure is preferred over the other.

Beyond the specific issues related to further work on CIATA-Ph, several directions for future work merit special attention. With respect to estimation of NNPPs, there is a clear need for the following: (i) optimal nonparametric smoothing methods for estimating the rate function; and (ii) efficient, nearly unbiased estimators of the asymptotic dispersion ratio. With respect to improved methods for simulating an NNPP with a given dispersion ratio, Equation (11) seems to be a key result. In particular, substantially faster convergence to the asymptotic dispersion ratio may be achieved by selecting an interrenewal distribution $G(\cdot)$ for the underlying ERP whose noncentral moments $\{\theta_\ell : \ell = 1, 2, 3\}$ (nearly) minimize the magnitude of the constant θ^* defined in Equation (9). For this purpose, it may be desirable to exploit a distribution family on $\mathbb{R}_+$ that is capable of matching a broad range of feasible values for the noncentral moments $\{\theta_\ell : \ell = 1, 2, 3\}$. Closely related to this possibility is the adaptation of CIATA-Ph to generate an NNPP whose dispersion ratio is piecewise constant over the relevant time horizon. All these topics are the subject of ongoing research.

Table 3. Performance Comparison of GNTA and CIATA-Ph with Sinusoidal Arrival Rates in Equation (20)

		$\gamma = 0.2, \eta = 0.84$		$\gamma = 0.5, \eta = 0.67$		$\gamma = 0.8, \eta = 0.56$	
	Performance	CIATA-Ph	GNTA	CIATA-Ph	GNTA	CIATA-Ph	GNTA
$\overline{\lambda} = 50$	$C = 10$	9.590	8.723	9.684	8.122	9.780	7.943
	Rel. err.	4.1%	12.8%	3.2%	18.8%	2.2%	20.6%
	$C = 4$	3.934	3.625	3.930	3.387	3.961	3.347
	Rel. err.	1.7%	9.4%	1.8%	15.3%	1.0%	16.3%
	$C = 1.5$	1.490	1.442	1.480	1.394	1.486	1.382
	Rel. err.	0.7%	3.9%	1.3%	7.1%	0.9%	7.9%
	$C = 1.2$	1.201	1.188	1.186	1.152	1.180	1.149
	Rel. err.	0.1%	1.0%	1.2%	4.0%	1.7%	4.3%
	$C = 1.1$	1.098	1.084	1.092	1.080	1.102	1.081
	Rel. err.	0.2%	1.5%	0.7%	1.8%	0.2%	1.7%
	$C = 0.8$	0.802	0.836	0.800	0.834	0.801	0.841
	Rel. err.	0.3%	4.5%	0.0%	4.3%	0.1%	5.1%
	$C = 0.2$	0.201	0.284	0.199	0.346	0.201	0.371
	Rel. err.	0.5%	42.0%	0.5%	73.0%	0.5%	85.5%
$\overline{\lambda} = 10$	$C = 10$	8.306	7.553	8.564	7.217	8.791	7.204
	Rel. err.	16.9%	24.5%	14.4%	27.8%	12.1%	28.0%
	$C = 4$	3.729	3.459	3.738	3.231	3.781	3.208
	Rel. err.	6.8%	13.5%	6.6%	19.2%	5.5%	19.8%
	$C = 1.5$	1.472	1.424	1.485	1.404	1.487	1.389
	Rel. err.	1.9%	5.1%	1.0%	6.3%	0.9%	7.4%
	$C = 1.2$	1.179	1.162	1.199	1.165	1.209	1.162
	Rel. err.	1.8%	3.2%	0.1%	2.9%	0.7%	3.2%
	$C = 0.8$	0.808	0.827	0.808	0.842	0.809	0.846
	Rel. err.	1.0%	3.4%	1.0%	5.3%	1.1%	5.7%
	$C = 0.2$	0.210	0.291	0.206	0.351	0.206	0.370
	Rel. err.	4.9%	45.5%	2.8%	75.4%	2.8%	84.8%
	$C = 0.1$	0.107	0.199	0.104	0.265	1.104	0.292
	Rel. err.	6.9%	98.9%	3.6%	64.9%	4.4%	192.3%

Notes. For each combination $\overline{\lambda}, \gamma, C$, and simulation procedure, the table shows the resulting dispersion-ratio estimate $\widehat{C}$ and the magnitude of the associated relative error $100(|\widehat{C} - C|/C)\%$, which has the label "Rel. err."

Table 4. Running Time Comparison of GNIA and CIATA-Ph with Case-1 Arrival Rates and $C = 10, 1.5, 0.8, 0.2$

		Number of subintervals Q used in CIATA-Ph					
	GNIA	50	100	150	200	250	300
$C = 10$	134.83	8.53	9.51	11.73	16.14	17.64	20.58
$C = 1.5$	137.38	8.47	9.21	11.43	16.37	17.51	21.04
$C = 0.8$	139.50	8.77	9.57	11.72	16.42	17.80	23.14
$C = 0.2$	160.73	8.47	9.41	11.61	16.43	17.69	21.78

Acknowledgments

The authors thank the area editor, the associate editor, and the referees for their numerous constructive suggestions that substantially improved this article.

References

Aldor-Noiman S, Feigin PD, Mandelbaum A (2009) Workload forecasting for a call center: Methodology and a case study. *Ann. Appl. Statist.* 3(4):1403–1447.

Asmussen S (1996) Fitting phase-type distributions via the EM algorithm. *Scand. J. Statist.* 23(4):419–441.

Avramidis AN, Deslauriers A, L'Ecuyer P (2004) Modeling daily arrivals to a telephone call center. *Management Sci.* 50(7): 896–908.

Avriel M, Diewert WE, Schaible S, Zang I (2010) *Generalized Concavity*, vol. 63, Classics in Applied Mathematics (Society for Industrial and Applied Mathematics, Philadelphia).

Çinlar E (1975) *Introduction to Stochastic Processes* (Prentice-Hall, Englewood Cliffs, NJ).

Chen H, Schmeiser BW (1992) Simulation of Poisson processes with trigonometric rates. Swain JJ, Goldsman D, Crain RC, Wilson JR, eds. *Proc. Winter Simulation Conf.* (Institute of Electrical and Electronics Engineers, Piscataway, NJ), 609–617.

Chen H, Schmeiser BW (2015) I-SMOOTH: Iteratively smoothing mean-constrained and nonnegative piecewise-constant functions. *INFORMS J. Comput.* 25(3):432–445.

Chen H, Schmeiser BW (2017) MNO-PQRS: Max nonnegativity ordering—piecewise-quadratic rate smoothing. *ACM Trans. Model. Comput. Simul.* 27(3):1–19.

Cox DR, Lewis PAW (1966) *The Statistical Analysis of Series of Events* (Chapman and Hall, London).

Fendick KW, Whitt W (1989) Measurements and approximations to describe the offered traffic and predict the average workload in a single-server queue. *Proc. IEEE* 77(1):171–194.

Gerhardt I, Nelson BL (2009) Transforming renewal processes for simulation of nonstationary arrival processes. *INFORMS J. Comput*. 21(4):630–640.

He B, Liu Y, Whitt W (2016) Staffing a service system with non-Poisson non-stationary arrivals. *Probab. Engrg. Inform. Sci.* 30(4):593–621.

Henderson SG (2003) Estimation for nonhomogeneous Poisson processes from aggregated data. *Oper. Res. Lett*. 31(5):375–382.

Ibrahim R, Ye H, L'Ecuyer P, Shen H (2016) Modeling and forecasting call center arrivals: A literature survey and a case study. *Internat. J. Forecast*. 32(3):865–874.

Jongbloed G, Koole G (2001) Managing uncertainty in call centers using Poisson mixtures. *Appl. Stochastic Model. Bus. Indust*. 17(4): 307–318.

Kim S-H, Whitt W (2014) Are call center and hospital arrivals well modeled by nonhomogeneous Poisson processes? *Manufacturing Service Oper. Management* 16(3):464–480.

Kuhl ME, Wilson JR (2009) Advances in modeling and simulation of nonstationary arrival processes. Lee LH, Kuhl ME, Fowler JW, Robinson S, eds. *Proc. 2009 INFORMS Simulation Soc. Res. Workshop* (Institute for Operations Research and the Management Sciences, Catonsville, MD), 1–5.

Kuhl ME, Wilson JR, Johnson MA (1997) Estimating and simulating Poisson processes having trends or multiple periodicities. *IIE Trans*. 29(3):201–211.

Law AM (2015) *Simulation Modeling and Analysis*, 5th ed. (McGraw-Hill, New York).

Lee S, Wilson JR, Crawford MM (1991) Modeling and simulation of a nonhomogeneous Poisson process having cyclic behavior. *Comm. Statist. Simul. Comput*. 20(2–3):777–809.

Lewis PAW, Shedler GS (1976) Statistical analysis of non-stationary series of events in a data base system. *IBM J. Res. Development* 20(5):465–482.

Lewis PAW, Shedler GS (1979) Simulation of nonhomogeneous Poisson processes by thinning. *Naval Res. Logist. Quart*. 26(3):403–413.

Liu R (2013) Modeling and simulation of nonstationary non-Poisson processes. Ph.D. thesis, North Carolina State University, Edward P. Fitts Department of Industrial and Systems Engineering, Raleigh, NC. Accessed August 4, 2015, http://www.lib.ncsu.edu/resolver/1840.16/8661.

Liu R, Kuhl ME, Liu Y, Wilson JR (2015) Combined inversion and thinning methods for simulating nonstationary non-Poisson arrival processes. Yilmaz L, Chan WKV, Moon I, Roeder TMK, Macal C, Rossetti MD, eds. *Proc. 2015 Winter Simulation Conf*. (Institute of Electrical and Electronics Engineers, Piscataway, NJ), 586–597.

Liu Y (2018) Staffing to stabilize the tail probability of delay in service systems with time-varying demand. *Oper. Res*. 66(2):514–534.

Liu Y, Whitt W (2012) Stabilizing customer abandonment in many-server queues with time-varying arrivals. *Oper. Res.* 60(6):1551–1564.

Liu Y, Whitt W (2014) Stabilizing performance in networks of queues with time-varying arrival rates. *Probab. Engrg Inform. Sci*. 28(4): 419–449.

Liu Y, Whitt W (2017) Stabilizing performance in a service system with time-varying arrivals and customer feedback. *Eur. J. Oper. Res*. 256(2):473–486.

Massey WA, Whitt W (1994) Unstable asymptotics for nonstationary queues. *Math. Oper. Res*. 19(2):267–291.

Nicol DM, Leemis LM (2014) A continuous piecewise-linear NHPP intensity-function estimator. Tolk A, Diallo SD, Ryzhov IO, Yilmaz L, Buckley S, Miller JA, eds. *Proc. 2014 Winter Simulation Conf*. (Institute of Electrical and Electronics Engineers, Piscataway, NJ), 498–509.

Pritsker AAB, Martin DL, Reust JS, Wagner MA, Daily OP, Harper AM, Edwards EB, et al. (1995) Organ transplantation policy evaluation. Alexopoulos C, Kang K, Lilegdon WR, Goldsman D, eds. *Proc. 1995 Winter Simulation Conf*. (Institute of Electrical and Electronics Engineers, Piscataway, NJ), 1314–1323.

Royden HL, Fitzpatrick PM (2010) *Real Analysis*, 4th ed. (Prentice-Hall, Englewood Cliffs, NJ).

Simmons DM (1975) *Nonlinear Programming for Operations Research* (Prentice-Hall International Series in Management, Prentice-Hall, Englewood Cliffs, NJ).

Sriram K, Whitt W (1986) Characterizing superposition arrival processes in packet multiplexers for voice and data. *IEEE J. Sel. Areas Comm*. 4(6):833–846.

Steckley SG, Henderson SG, Mehrotra V (2009) Forecast errors in service systems. *Probab. Engrg. Inform. Sci*. 23(2):305–332.