Proceedings of the 2008 Winter Simulation Conference
S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds.

SMOOTH FLEXIBLE MODELS OF NONHOMOGENEOUS POISSON PROCESSES USING ONE OR MORE PROCESS REALIZATIONS

Michael E. Kuhl
Shalaka C. Deo

Industrial & Systems Engineering Department
Rochester Institute of Technology
Rochester, NY 14623, U.S.A.

James R. Wilson

Edward P. Fitts Department of Industrial and Systems Engineering
North Carolina State University
Raleigh, NC 27695, U.S.A.

ABSTRACT

We develop and evaluate a semiparametric method to estimate the mean-value function of a nonhomogeneous Poisson process (NHPP) using one or more process realizations observed over a fixed time interval. To approximate the mean-value function, the method exploits a specially formulated polynomial that is constrained in least-squares estimation to be nondecreasing so the corresponding rate function is nonnegative and smooth (continuously differentiable). An experimental performance evaluation for two typical test problems demonstrates the method's ability to yield an accurate fit to an NHPP based on a single process realization. A third test problem shows how the method can estimate an NHPP based on multiple realizations of the process.

1 INTRODUCTION

Nonstationary (time-dependent) processes are commonly encountered in simulation studies, including applications to manufacturing, health care, and telecommunications systems. Time-dependent point processes are often modeled as Nonhomogeneous Poisson Processes (NHPPs)—for example, the stream of patients arriving at a hospital, the arrivals of customer-service requests at a call center, and the pattern of occurrences over time of device failures in a telecommunications network.

Let $N(t)$ denote the number of arrivals (events) in the time interval $(0, t]$ for each $t \in (0, S]$, a fixed observation interval of interest. If $\{N(t) : t \in (0, S]\}$ is an NHPP, then it is completely defined by its rate function, $\lambda(t)$, or by its mean-value function,

$$\mu(t) = \mathrm{E}[N(t)] = \int_0^t \lambda(u)\, du,$$

for all $t \in (0, S]$ (Çinlar 1975). In this paper we propose a specially formulated polynomial approximation to the mean-value function of an NHPP that can be readily estimated by constrained least squares from one or more independent realizations of the process; and the resulting estimator of the rate function is smooth (continuously differentiable) over the entire observation interval.

The rest of this paper is organized as follows. Section 2 gives the background and motivation for this research. Section 3 details the NHPP-fitting method for the situation in which we have a single realization of the target arrival process. Section 4 presents an experimental performance evaluation of the method for two typical test processes when only a single realization of each process is available. Section 5 extends the NHPP-fitting method to handle multiple realizations of the fitted process, and Section 5 also provides an illustrative example of the extended procedure. Section 6 contains conclusions and a description of future work.

2 BACKGROUND AND RELATED WORK

The NHPP literature includes a number of methods to model the mean-value function and the associated rate function for a selected point process so as to estimate these functions accurately from sample data and subsequently to generate independent realizations of the fitted process efficiently. The current work is motivated by the multiresolution procedure of Kuhl and Wilson (2001) and Kuhl, Sumant, and Wilson (2006) for modeling, estimating, and simulating NHPPs that exhibit one or more nested cyclic effects (for example, time-of-day effects and day-of-the-week effects) and that may also exhibit a long-term trend. The multiresolution procedure estimates the mean-value function in a semiparametric manner from a single realization of the process. However, this method cannot handle NHPPs that lack cyclic rate components. This paper addresses this issue as well as the case of multiple realizations of the target arrival process.

Some closely related work on NHPPs includes the following. MacLean (1974) approximates the rate function of an NHPP using an exponential-polynomial function. Leemis (1991) estimates a piecewise-constant approximation to the rate function of an NHPP from multiple realizations of the process. Johnson, Lee, and Wilson (1994) use an exponential-polynomial-trigonometric rate function for processes having a single cyclic effect or a long-term trend (or both). Kuhl, Wilson, and Johnson (1997) employ an "exponential-polynomial-trigonometric rate function with multiple periodicities" to model processes having a long-term trend or one or more rate components exhibiting cyclic behavior. In this paper, we propose a semiparametric method for modeling the mean-value function of an NHPP that will complement this set of modeling tools.

3 METHODOLOGY

In this section, we seek to fit the mean-value function of an NHPP to a single realization of observed arrivals over the observation interval (0, *S*]. We propose a semiparametric model of the form

$$\mu(t) = \mu(S)R(t) \quad \text{for all } t \in (0, S], \tag{1}$$

where $R(t)$ is a nondecreasing function representing the cumulative proportion of arrivals up to time *t*. In principle, a uniformly accurate approximation to the function $R(t)$ can always be achieved using a polynomial of sufficiently high degree *r* with the special form

$$R(t) = \begin{cases} t/S, & \text{if } r = 1, \\ \displaystyle\sum_{k=1}^{r-1} \beta_k (t/S)^k + \left(1 - \sum_{k=1}^{r-1} \beta_k\right)(t/S)^r, & \text{if } r > 1, \end{cases} \tag{2}$$

where the coefficient vector $\mathbf{B}_r = (\beta_1, \ldots, \beta_{r-1})$ is constrained to yield $R'(t) \geq 0$ for all $t \in (0, S]$. Note that the form of Equation (2) ensures the initial value is $R(0) = 0$, and the final value is $R(S) = 1$ for all values of $\mathbf{B}_r$.

3.1 Estimation of the Function *R*(*t*)

Given the observed arrival times from a single realization of the target process sorted in ascending order $t_{(1)} \leq t_{(2)} \leq \ldots \leq t_{(N(S))}$, we let $W_i = i/N(S)$ denote the corresponding cumulative proportion of arrivals up to the time $t_{(i)}$ of the i th arrival for $i = 1, 2, \ldots, N(S)$. We seek to fit the polynomial function $R(t)$ to the points $[t_{(i)}, W_i]$ for $i = 1, 2, \ldots, N(S)$ via the following steps:

- Transform the data using a variance-stabilizing transformation;
- For the transformed data, estimate the degree *r* of the best-fitting polynomial of the form (4) below (which is a suitably rescaled version of (2) above) using a modified likelihood ratio test; and
- Given the polynomial degree r, estimate the vector $\mathbf{B}_r$ in Equation (2) by applying the method of least squares to the original arrival data.

The procedure for determining the degree of the fitted polynomial and then estimating the polynomial coefficients is based on a forward-selection type of regression analysis. However, conventional regression analysis requires responses that are independent and normally distributed with a constant variance. The observations $\{W_i\}$ from an NHPP do not satisfy this requirement. Therefore, following Kuhl, Sumant, and Wilson (2006), we employ a variance-stabilizing transformation of the form

$$Y_i = \sin^{-1}\left(\sqrt{W_i}\right) \tag{3}$$

for $i = 1, \ldots, N(S)$. For convenience, the arrival times are scaled to the unit interval so that we take

$$Z_i = t_{(i)}/S,$$

resulting in the transformed points $[Z_i, Y_i]$ for $i = 1, \ldots, N(S)$. Equation (3) yields responses $\{Y_i\}$ that are approximately normal with a constant variance σ^2.

To find the appropriate degree r of the polynomial $R(t)$, we apply the method of constrained least squares to the transformed data set of size $m = N(S)$. We fit a statistical model of the form $\mathrm{E}[Y_i] = f_r(Z_i; \mathbf{C}_r)$ such that

$$f_r(u; \mathbf{C}_r) = \begin{cases} \left(\dfrac{\pi}{2}\right)u, & \text{if } r = 1, \\ \displaystyle\sum_{k=1}^{r-1} C_k u^k + \left(\frac{\pi}{2} - \sum_{k=1}^{r-1} C_k\right)u^r, & \text{if } r > 1, \end{cases} \tag{4}$$

for $u \in [0,1]$, where $\mathbf{C}_r$ is the coefficient vector

$$\mathbf{C}_r \equiv \begin{cases} \pi/2, & \text{if} \quad r=1, \\ \left[C_1,\ldots,C_{r-1}\right]^{\mathrm{T}}, & \text{if} \quad r>1, \end{cases}$$

subject to the constraint $f_r'(u;\mathbf{C}_r)\ge 0$ for all $u \in [0,1]$. Note that for $r\ge 2$, the nonnegativity constraint is equivalent to requiring that the zeros of the degree-$(r-1)$ polynomial $f_r'(u;\mathbf{C}_r)$ lie outside the interval (0, 1).

To determine the appropriate degree for the statistical model (4), a modified likelihood ratio test is used. For successive values of *r*, the vector $\mathbf{C}_r$ is estimated via constrained least squares, yielding

$$\tilde{\mathbf{C}}_r = \underset{\hat{\mathbf{C}}_r : f_r'(u;\hat{\mathbf{C}}_r)\ge 0}{\arg\min} \sum_{i=1}^{m}\left[Y_i - f_r(Z_i;\hat{\mathbf{C}}_r)\right]^2$$

for $r \ge 2$. The corresponding error sum of squares for the degree-r fit is

$$\mathrm{SSE}_r = \sum_{i=1}^{m}\left[Y_i - f_r(Z_i;\tilde{\mathbf{C}}_r)\right]^2$$

for $r\ge 2$; and for $r=1$, we take

$$\mathrm{SSE}_1 = \sum_{i=1}^{m}\left[Y_i - \frac{i\pi}{2(m+1)}\right]^2.$$

The corresponding total sum of squares,

$$\mathrm{SST} = \sum_{i=1}^{m}\left(Y_i - \bar{Y}\right)^2, \quad \text{where} \quad \bar{Y} = m^{-1}\sum_{i=1}^{m} Y_i,$$

is used in the first step of the modified likelihood-ratio procedure to yield a constant rate function if such a model is appropriate.

In terms of the maximum likelihood estimator of the response variance σ^2 for each postulated value of r,

$$\tilde{\sigma}_r^2 = \mathrm{SSE}_r / m \quad \text{for} \quad r = 1,\ 2,\ \ldots,$$

we see that the associated likelihood function is given as

$$L_r(\tilde{\mathbf{C}}_r;\mathbf{Y}) = \prod_{i=1}^{m}\frac{1}{\tilde{\sigma}_r\sqrt{2\pi}}\exp\left\{-\frac{\left[Y_i - f_r(Z_i;\tilde{\mathbf{C}}_r)\right]^2}{2\tilde{\sigma}_r^2}\right\}$$
$$=\left(2\pi e\tilde{\sigma}_r^2\right)^{-m/2};$$

and the resulting log-likelihood function for degree *r* is

$$\Psi_r(\tilde{\mathbf{C}}_r;\mathbf{Y}) = -\frac{m}{2}\left[\ln(2\pi)+1+\ln(\tilde{\sigma}_r^2)\right].$$

The degree *r* is determined using the following likelihood ratio test (Kuhl, Sumant and Wilson 2006) at the level of significance α (where $0 < \alpha < 1$). The final estimate of *r* is

$$\tilde{r} = \begin{cases} 1, & \text{if} \quad \mathrm{SSE}_1/\mathrm{SST} < 0.01 \text{ or } m \le 2, \\ \min\left\{r : r\ge 2; -m\ln\left(\dfrac{\tilde{\sigma}_r^2}{\tilde{\sigma}_{r-1}^2}\right) \le \chi^2_{1-\alpha}(1)\right\} - 1, & \text{otherwise}, \end{cases}$$

where $\chi^2_{1-\alpha}(1)$ denotes the $1-\alpha$ quantile of the chi-squared distribution with 1 degree of freedom.

To emphasize the dependence of the polynomial in Equation (2) on its degree r and its coefficient vector $\mathbf{B}_r$, in the rest of this section we write this function as $R(t;r,\mathbf{B}_r)$ for $t\in(0,S]$. Given the estimated degree $\tilde{r}$ to be used in Equation (2), the coefficient vector $\mathbf{B}_{\tilde{r}}$ is estimated by applying constrained least squares to the original (untransformed) data, yielding

$$\tilde{\mathbf{B}}_{\tilde{r}} = \underset{\hat{\mathbf{B}}_{\tilde{r}} : R'(u;\tilde{r},\hat{\mathbf{B}}_{\tilde{r}})\ge 0}{\arg\min} \sum_{i=1}^{m}\left[W_i - R(t_{(i)};\tilde{r},\hat{\mathbf{B}}_{\tilde{r}})\right]^2, \quad (5)$$

provided $\tilde{r}\ge 2$; and if $\tilde{r}=1$, then we take $\tilde{\mathbf{B}}_1 = \tilde{\beta}_1 = 1$ in Equation (2) so that the fitted rate function is constant. Thus the final estimator of the mean-value function (1) is

$$\tilde{\mu}(t) = N(S)R(t;\tilde{r},\tilde{\mathbf{B}}_{\tilde{r}}) \quad \text{for all} \quad t\in(0,S], \qquad (6)$$

and the estimated rate function is $\tilde{\lambda}(t) = \tilde{\mu}'(t)$ for all $t\in(0,S]$.

4 EXPERIMENTAL PERFORMANCE EVALUATION

To evaluate the effectiveness of the semiparametric method for modeling the mean-value function of an NHPP, we conduct an experimental performance evaluation in which we present two cases for evaluation. For each case, we generate one realization of the NHPP and use the semiparametric method to fit the mean-value function. This procedure is carried out for *K*=100 replications of each case. Numerical and graphical results of the study are presented.

4.1 Statistical Performance Measures

To evaluate the ability of the semiparametric method to fit either the underlying NHPP or the observed data with sufficient accuracy, we use several statistical performance measures previously formulated by other researchers—specifically, Johnson, Lee, Wilson (1994); Kuhl, Wilson, and Johnson (1997); Kuhl and Wilson (2001); and Kuhl, Sumant, and Wilson (2006). For completeness the definitions of these performance measures are included here.

In the rest of this section, $\tilde{\lambda}_k(t)$ denotes the estimated rate function and $\tilde{\mu}_k(t)$ denotes the estimated mean-value function for the k^{th} replication of a test problem (case). Two sets of statistical performance measures are used. The first set of measures allows us to compare the estimated rate or mean-value function with its theoretical counterpart, while the second set of statistics allows us to compare the fitted rate or mean-value function with its empirical counterpart defined by the observed set of arrivals.

In estimating the rate function $\lambda(t)$ on the k^{th} replication of a test process, the *average absolute error* δ_k and *maximum absolute error* δ_k^* are, respectively,

$$\delta_k \equiv \frac{1}{S}\int_0^S \left|\tilde{\lambda}_k(t)-\lambda(t)\right| dt,$$

$$\delta_k^* = \max\left\{\left|\tilde{\lambda}_k(t)-\lambda(t)\right| : 0 \le t \le S\right\}$$

for $k = 1, \ldots, K$. In estimating the mean-value function $\mu(t)$, similar measures for the *average absolute error* Δ_k and the *maximum absolute error* Δ_k^* are defined by

$$\Delta_k = \frac{1}{S}\int_0^S \left|\tilde{\mu}_k(t)-\mu(t)\right| dt,$$

$$\Delta_k^* = \max\left\{\left|\tilde{\mu}_k(t)-\mu(t)\right| : 0 \le t \le S\right\},$$

respectively, for $k = 1, \ldots, K$.

Aggregate performance measures summarize the error in estimating the rate function over all replications of the test process. The sample mean of the $\{\delta_k : k = 1, \ldots, K\}$ is denoted by

$$\bar{\delta} = \frac{1}{K}\sum_{k-1}^{K}\delta_k;$$

and the sample coefficient of variation over all replications of the test process is

$$V_\delta = \left[\frac{1}{K-1}\sum_{k=1}^{K}(\delta_k-\bar{\delta})^2\right]^{1/2} \Big/ \bar{\delta}.$$

Maximum-absolute-error statistics for the theoretical rate function are computed similarly so that the sample mean and coefficient of variation of the $\{\delta_k^* : k = 1, \ldots, K\}$ are denoted by $\bar{\delta}^*$ and V_{δ^*}, respectively.

Analogous performance measures for average absolute errors in estimating the theoretical mean-value function are denoted by $\bar{\Delta}$ and V_Δ. For maximum absolute errors in estimating the theoretical mean-value function, the corresponding summary statistics are denoted by $\bar{\Delta}^*$ and V_{Δ^*}, respectively.

Normalized statistics are also computed to facilitate comparison of results for different rate and mean-value functions—that is, for different test processes:

$$Q_\delta = \frac{\bar{\delta}}{\mu(S)/S}, \quad Q_{\delta^*} = \frac{\bar{\delta}^*}{\mu(S)/S},$$

$$Q_\Delta = \bar{\Delta}\Big/\left[\frac{1}{S}\int_0^S \mu(t)\,dt\right],$$

$$Q_{\Delta^*} = \bar{\Delta}^*\Big/\left[\frac{1}{S}\int_0^S \mu(t)\,dt\right].$$

In addition to the quantities mentioned above, statistics are developed to measure the ability of the proposed procedure to approximate each observed arrival process. On the k^{th} replication of a given NHPP ($k = 1, 2, \ldots, K$), we let $\{t_{(i),k} : i = 1, 2, \ldots, N_k(S)\}$ denote the ordered arrival epochs observed in the time interval $(0, S]$. Averages are reported over all K replications of the following: the sum of squared errors (SS_E) and the mean-squared error (MS_E) along with the associated coefficients of variation. Moreover in estimating the empirical mean-value function on the k^{th} replication of a test process, the average absolute error and maximum absolute error are, respectively,

$$D_k \equiv \frac{1}{N_k(S)}\sum_{i=1}^{N_k(S)}\left|\tilde{\mu}_k(t_{(i),k})-i\right|,$$

$$D_k^* \equiv \max\left\{\left|\tilde{\mu}_k(t_{(i),k})-i\right| : 1 \le i \le N_k(S)\right\}$$

for $k = 1, 2, \ldots, K$. To compare the averages $\bar{D}$ and $\bar{D}^*$ across test cases, we use the normalized statistics

$$Q_D = \frac{\bar{D}}{(1/K)\sum_{k=1}^{K}(1/S)\int_0^S N_k(t)\,dt},$$

$$Q_{D^*} = \frac{\bar{D}^*}{(1/K)\sum_{k=1}^{K}(1/S)\int_0^S N_k(t)\,dt}.$$

The second type of aggregate performance measure is estimated by expressing each performance measure D_k and D_k^* observed on the k^{th} replication as a percentage of the average level of the empirical mean-value function on that replication and then calculating the average across replications as follows:

$$H_D = \frac{1}{K}\sum_{k=1}^{K}\frac{D_k}{(1/S)\int_0^S N_k(t)\,dt},$$

$$H_{D^*} = \frac{1}{K}\sum_{k=1}^{K}\frac{D_k^*}{(1/S)\int_0^S N_k(t)\,dt}.$$

4.2 Experimental Cases

In this section, we present results for two test problems whose mean-value functions have the semiparametric form of Equations (1)–(2). In Case 1, the underlying NHPP has $S = 4$, $\mu(S) = 2000$, and degree $r = 5$. In Case 2, the NHPP has $S = 2$, $\mu(S) = 1500$, and degree $r = 3$. Table 1 summarizes the configuration for each test case. In the experimentation, we generated $K = 100$ independent replications of each test process; and each of the resulting data sets was supplied to the NHPP-fitting procedure.

Table 1: Polynomial coefficients for Cases 1 and 2.

	Polynomial Coefficients				
Case	β_1	β_2	β_3	β_4	β_5
1	0.89262	–1.19061	0.78378	–0.21507	0.02807
2	0.81499	–0.67499	0.25875	-	-

4.3 Results and Analysis

The fitting procedure is used to estimate the mean-value function of the NHPP for Cases 1 and 2. For Case 1, the procedure is illustrated for fitting an NHPP to a single realization of the process. Graphs are plotted at every step of the estimation process. Then, numerical and graphical performance measures are presented for both cases for K=100 replications of the fitting procedure when it is applied to a single realization of the process.

To illustrate the fitting procedure for Case 1, Figure 1 shows a histogram for the number of arrivals over equal subintervals of length 0.1 time units. Figure 2 displays the step function for the observed cumulative arrivals from the single realization. Figure 3 shows the transformed data and the estimated function (4) for the transformed response. For this realization, the fitting procedure yielded a degree-6 polynomial. Given the degree of the polynomial, the procedure fit a degree-6 polynomial to the original data. The resulting polynomial coefficients are shown in Table 2, and the estimated mean-value function is plotted along with the empirical mean-value function in Figure 4. The estimated rate function calculated as the derivative of the mean-value function is shown in Figure 5.

The numerical and statistical results for the experimental performance evaluation using Cases 1 and 2 are shown in Tables 3 and 4 and Figures 6–9, respectively. The statistical performance measures are collected for K=100 replications of the fitting procedure for each case. Table 3 contains the performance measures for evaluating the fit of the estimated rate and mean-value functions to the underlying rate and mean-value functions for each case. Table 4 compares the fits for each case with the empirical mean-value function. Figures 6 and 8 show 90% tolerance intervals (bands) for the fitted mean-value function around the underlying mean-value function in Cases 1 and 2. Figures 7 and 9 show 90% tolerance intervals (bands) for the fitted rate function for Cases 1 and 2, respectively.

Table 2: Estimated polynomial coefficients for the mean-value function fit to one realization of Case 1.

		Polynomial Coefficients			
$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$	$\tilde{\beta}_4$	$\tilde{\beta}_5$	$\tilde{\beta}_6$
0.9043	–1.2233	0.8133	–0.2245	.02189	1.03E–5

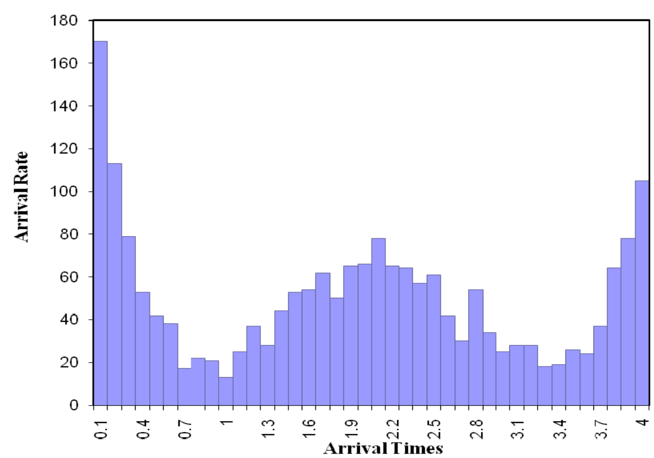

Figure 1: Histogram of arrival data for $t \in [0,4]$, in Case 1.

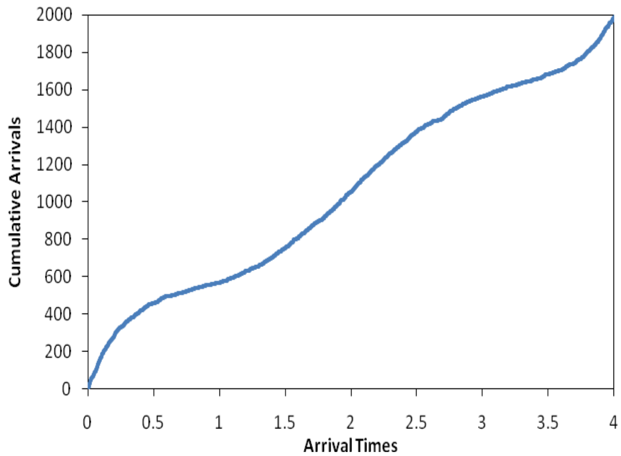

Figure 2: Cumulative arrivals for $t \in [0,4]$, in Case 1.

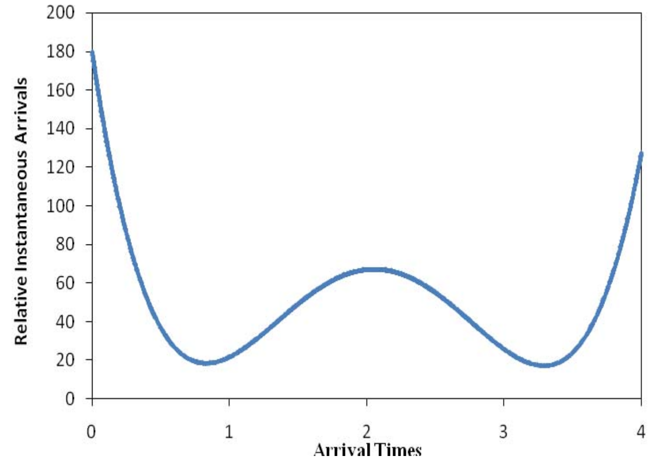

Figure 5: Estimated rate function for $t \in [0,4]$, in Case 1.

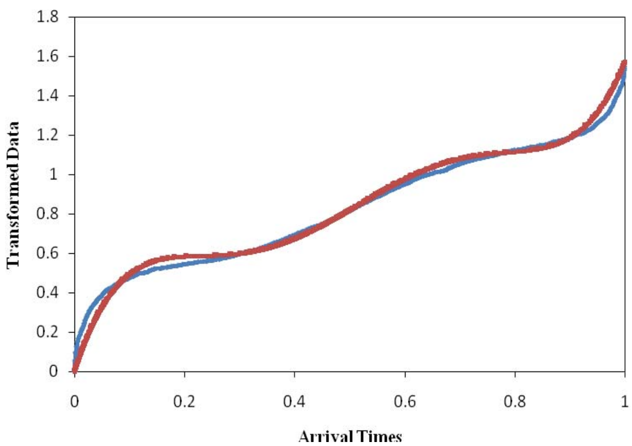

Figure 3: Transformed fit vs transformed data for $t \in [0,4]$, in Case 1.

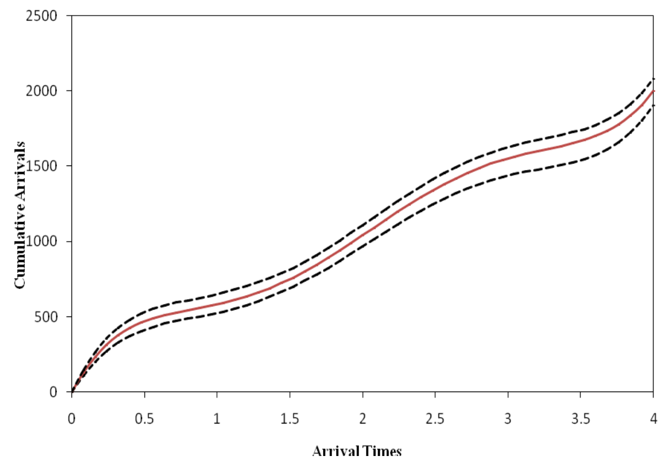

Figure 6: 90% tolerance intervals for $\mu(t)$, $t \in [0,4]$, in Case 1.

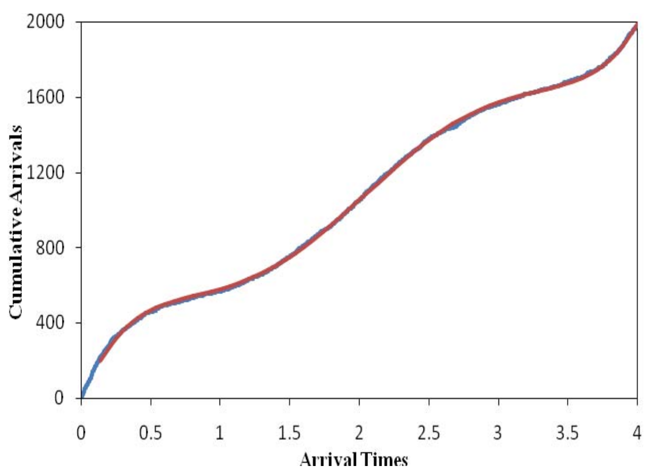

Figure 4: Estimated mean-value function for $t \in [0,4]$, in Case 1.

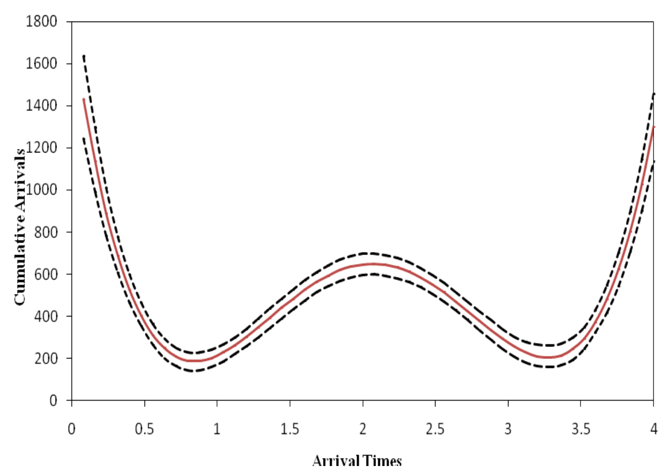

Figure 7: 90% tolerance intervals for $\lambda(t)$, $t \in [0,4]$, in Case 1.

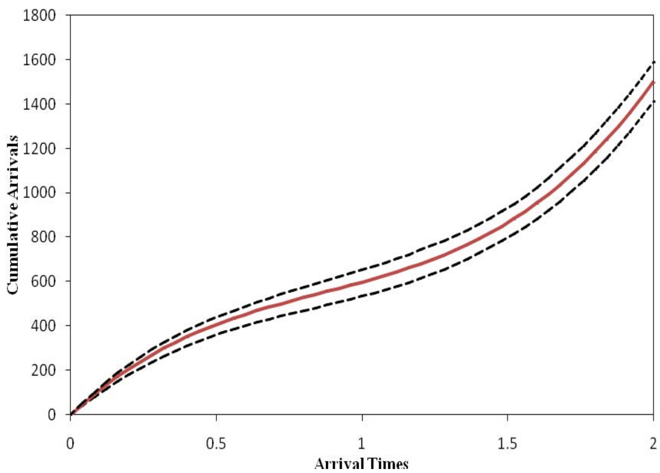

Figure 8: 90% tolerance intervals for $\mu(t)$, $t\in[0,2]$, in Case 2.

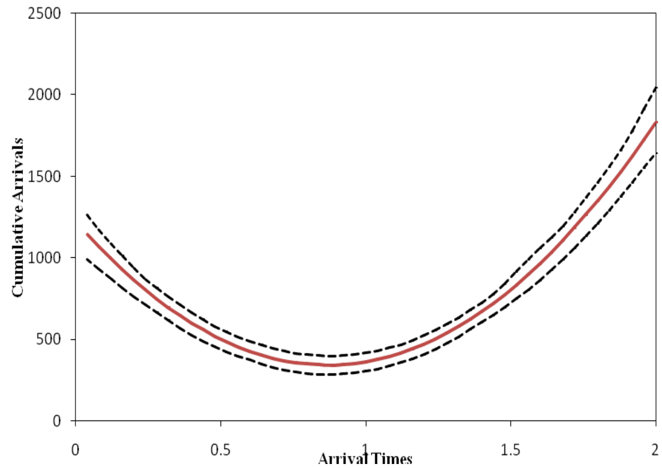

Figure 9: 90% tolerance intervals for $\lambda(t)$, $t\in[0,2]$, in Case 2.

Table 3: Goodness-of-fit statistics for estimating $\lambda(t)$ and $\mu(t)$.

Performance Measures	Case 1 ($S = 4$)	Case 2 ($S = 2$)
δ	21.05100	29.42100
V_δ	0.30193	0.39835
Q_δ	0.04210	0.03923
$\overline{\delta}^*$	83.82000	85.85000
V_{δ^*}	0.50184	0.56484
Q_{δ^*}	0.16764	0.11447
$\overline{\Delta}$	25.88600	21.10200
V_Δ	0.55332	0.54295
Q_Δ	0.09620	0.06389
$\overline{\Delta}^*$	47.51000	40.24200
V_{Δ^*}	0.45866	0.51260
Q_{Δ^*}	0.17662	0.12184

Table 4: Goodness-of-fit statistics for estimating $N(t)$.

Measures	Case 1 ($S = 4$)	Case 2 ($S = 2$)
SS_E	64.11863	59.82660
$V_{\mathrm{SS}_\mathrm{E}}$	0.05753	0.03913
MS_E	0.03200	0.03986
$V_{\mathrm{MS}_\mathrm{E}}$	0.05986	0.03814
$\overline{D}$	5.60218	6.00775
$\overline{D}^*$	20.05102	20.25724
Q_D	0.00532	0.00925
Q_{D^*}	0.01904	0.03117
H_D	4.660E-5	8.529E-5
H_{D^*}	1.691E-4	2.632E-4

5 ESTIMATING AN NHPP FROM MULTIPLE REALIZATIONS OF THE PROCESS

If we were able to obtain multiple realizations of the point process under study (say, $P \geq 2$ realizations), and we wanted to fit an NHPP to the process, then the semiparametric procedure outlined in Section 3 can be applied with some minor modifications. Suppose that on observed realization ℓ of the process, we have a total of $N_\ell(S)$ arrivals $\{t_{i,\ell} : i = 1, \ldots, N_\ell(S)\}$; and let

$$\overline{N}(S) = \frac{1}{P}\sum_{\ell=1}^{P} N_\ell(S)$$

denote the average number of arrivals in $(0, S]$ taken over all P realizations of the process.

To obtain an estimate for the constrained polynomial (2), we form the consolidated set of $m = P \times \overline{N}(S)$ arrival times $\{t_{i,\ell} : i = 1, \ldots, N_\ell(S) \text{ and } \ell = 1, \ldots, P\}$ taken over all P realizations; and we sort this overall set in ascending order to yield the ordered arrival times $t_{(1)} \leq t_{(2)} \leq \cdots$

$\leq t_{(m)}$. To adapt the approach of Section 3.1, we let $W_i = i/m$ denote the cumulative fraction of the overall set of arrivals that occur up to the time $t_{(i)}$ of the ith earliest arrival for i = 1, …, m. The pairs $\{[t_{(i)}, W_i]\}$ are then transformed using the variance-stabilizing transformation detailed in Section 3.1. The transformed data set $\{[Z_i, Y_i]\}$ is used to estimate the appropriate degree $\tilde{r}$ of the fitted polynomial (2). The least-squares estimate $\tilde{\mathbf{B}}_{\tilde{r}}$ of the coefficient vector is obtained from (5) using the original (untransformed) data $\{[t_{(i)}, W_i]\}$. The final estimator of the mean-value function then has the form

$$\tilde{\mu}(t) = \bar{N}(S)R(t; \tilde{r}, \tilde{\mathbf{B}}_{\tilde{r}}) \quad \text{for all} \quad t \in (0, S]. \qquad (7)$$

5.1 Illustrative Example for an NHPP with Multiple Process Realizations

To illustrate the fitting procedure for fitting the mean value function given multiple realizations of the NHPP, we utilize a common example of arrivals to a lunch wagon which appears in Leemis and Park (2006) and elsewhere. In this example, $S = 4.5$. Here we utilized the piecewise linear mean-value function of Leemis and Park (2006) to generate $P = 3$ realizations of the arrival process. The arrival rate during each subinterval is illustrated in Figure 10. The generated data resulted in realizations of $N_1(S)$ = 60, $N_2(S)$ = 48, and $N_3(S)$ = 45 arrivals, for a total of $m = 153$ arrivals in the overall data set.

Given 3 realizations of the arrival process, the semiparametric procedure was employed, resulting in a fitted degree-5 polynomial having the coefficients shown in Table 5. Figure 11 shows the fitted mean-value function plotted along with the empirical mean-value function created from the overall set of 153 arrival events from the 3 realizations. Finally, Figure 12 shows the estimated rate function over the interval (0,4.5].

Table 5: Polynomial coefficients for the fitted polynomial in the lunch wagon example (multiple realizations).

Polynomial Coefficients				
$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$	$\tilde{\beta}_4$	$\tilde{\beta}_5$
2.57046	7.92849	19.58847	–19.88498	7.22533

6 CONCLUSION AND FUTURE WORK

In this paper, we have presented a semiparametric method for fitting the mean-valuefunction of a nonhomogeneous

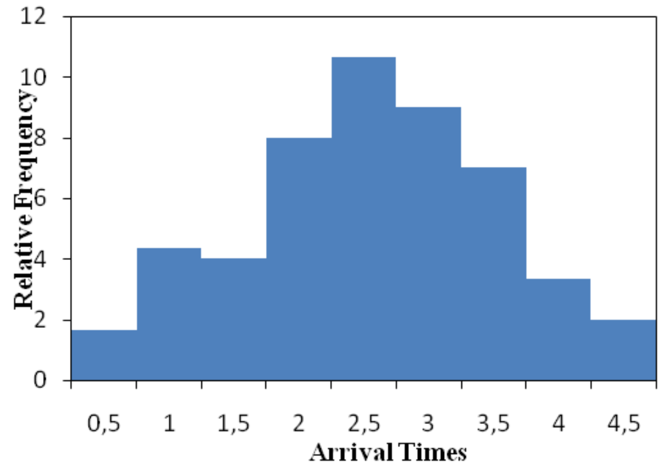

Figure 10: Arrival rate over the interval (0, 4.5] for the lunch wagon example (Leemis and Park 2006).

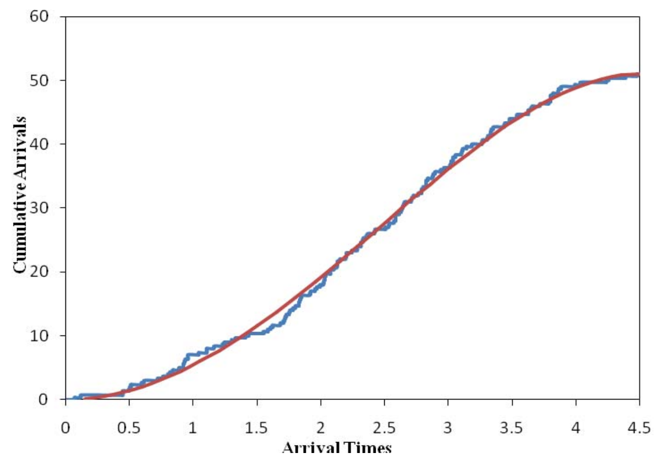

Figure 11: Fitted mean-value function for the lunch wagon example.

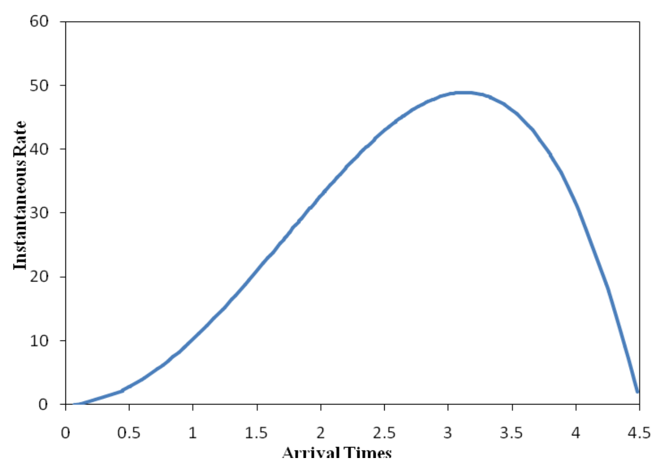

Figure 12: Fitted rate function for the lunch wagon example.

Poisson process when one or more realizations of the process are available. The experimental performance evaluation demonstrates that the proposed method is capable of consistently estimating the underlying mean-value function of the NHPP in some cases of interest. Although not discussed here, the method also results in an efficient method for generating realizations from the fitted NHPP in simulation experiments. We believe that the proposed procedure for modeling and simulation of arrival processes has distinct advantages in applications for which the "physics" of the problem require a smooth rate function rather than a rate function that is piecewise constant or exhibits other types of nonsmooth behavior.

Future research includes conducting an extended experimental performance evaluation for a wide variety of point processes that are frequently encountered in certain application domains—for example, modeling and analysis of call centers. In addition, we plan to "stress-test" the procedure to determine the level of complexity of the underlying arrival process that the method will be able to handle reliably.

REFERENCES

Çinlar, E. 1975. *Introduction to stochastic processes.* Englewood Cliffs, New Jersey: Prentice Hall.

Johnson, M. A., S. Lee, and J. R. Wilson. 1994. Experimental evaluation of a procedure for estimating nonhomogeneous Poisson processes having cyclic behavior. *ORSA Journal on Computing* 6 (4): 356–368.

Kuhl, M.E., and J.R. Wilson. 2001. Modeling and simulating Poisson processes having trends or nontrigonometric cyclic effects. *European Journal of Operational Research* 133:566–582.

Kuhl, M. E., J. R. Wilson, and M. A. Johnson. 1997. Estimating and simulating Poisson processes having trends or multiple periodicities. *IIE Transactions* 29 (3): 201–211.

Kuhl, M. E., S. G. Sumant, and J. R. Wilson. 2006. An automated multiresolution procedure for modeling complex arrival processes. *INFORMS Journal on Computing* 18 (1): 3–18.

Leemis, L. M. 1991. Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process. *Management Science* 27 (7): 886–900.

Leemis, L. M., and S. K. Park. 2006. *Discrete event simulation.* Englewood Cliffs, New Jersey: Pearson Prentince Hall.

MacLean, C. J. 1974. Estimating and tesing of an exponential polynomial rate function within the nonstationary Poisson process. *Biometrika* 61 (1): 81–85.

AUTHOR BIOGRAPHIES

MICHAEL E. KUHL is an Associate Professor in the Industrial and Systems Engineering Department at Rochester Institute of Technology. He has a Ph.D. in Industrial Engineering from North Carolina State University (1997). His research interests include simulation modeling methodologies with application to cyber security, healthcare, and semiconductor manufacturing, and simulation analysis procedures for input modeling and output analysis. He served as *Proceedings* Editor for the 2005 Winter Simulation Conference. He is president of the INFORMS Simulation Society, and a member of IIE and ASEE. His e-mail address is <Michael.Kuhl@rit.edu> and his Web address is <people.rit.edu/mekeie>.

SHALAKA C. DEO is a student pursuing a Master's degree in Industrial and Systems Engineering at Rochester Institute of Technology. Her research interest is in field of systems modeling.

JAMES R. WILSON is a professor of the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. He is a member of AAUW, ACM, and ASEE; and he is a Fellow of IIE and INFORMS. He has held several positions in WSC, including *Proceedings* Editor (1986); Program Chair (1992); and member of the Board of Directors (1997–2004). His e-mail address is <jwilson@ncsu.edu>, and his Web address is <www.ise.ncsu.edu/jwilson>.