A Spline Function Method for Modelling and Generating a Nonhomogeneous Poisson Process

L. E. Morgan[a], Barry. L. Nelson[b], Andrew. C. Titman[c] and David. J. Worthington[a]

[a]Department of Management Science, Lancaster University, Lancaster, LA1 4YX, UK;
[b]Department of Industrial Engineering & Management Sciences, Northwestern University, Evanston, IL, 60208-3119; [c]Department of Mathematics and Statistics, Lancaster University, LA1 4YR.

ABSTRACT
This paper presents a spline-based input modelling method for inferring the rate function of a nonhomogeneous Poisson process (NHPP) given arrival-time observations and a simple method for generating arrivals from the resulting rate function. Splines are a natural choice for modelling rate functions as they are smooth by construction, and highly flexible. Although flexibility is an advantage in terms of reducing the bias with respect to the true rate function, it can lead to overfitting. Our method is therefore based on maximising the penalised NHPP log-likelihood, where the penalty is a measure of rapid changes in the spline-based representation. A controlled empirical comparison of the spline-based method against two recently developed input modelling techniques is presented considering: the recovery of the rate function; the propagation of input modelling error and the performance of methods given data that is under or over-dispersed in comparison to a Poisson process.

1. Introduction

This paper presents a spline-based method for modelling and generating non-homogeneous Poisson process (NHPP) arrivals within stochastic simulation models. Our motivation for the creation of a new input modelling method is twofold: to reduce the input modelling error passed from arrival processes to simulation output performance measures during simulation experiments, and because in reality rate functions are likely to be smooth. There are a number of interval-based (piecewise) input modelling methods for NHPPs, but these models assume that the arrival-rate function can jump instantaneously in time and this is often unrealistic and limits their usefulness for modelling general processes. Interval-based methods also require knowledge of the interval boundaries which, if unknown, are hard to determine, and if the true rate function is smooth do not exist.

We assume that we have a finite number of arrival-time observations of a real-world system from which to estimate the arrival process. The input model will thus

---

inevitably contain error. In stochastic simulation, error in the input models propagates through the simulation model to the output of interest. This error is known as input modelling error; for further information see Song, Nelson, and Pegden (2014), Lam (2016), Morgan, Titman, Worthington, and Nelson (2016), Morgan, Rhodes-Leader, and Barton (2022), and references therein. We aim to provide an input modelling method that propagates less input modelling error to the simulation output than its competitors.

Our focus is arrival processes, which are a requirement of many simulation models, and more specifically NHPPs which are commonly used in practice. Pritsker et al. (1996) shows the use of a NHPP for modelling donor and patient arrivals within a large scale simulation model developed for the United Network of Organ Sharing (UNOS). A NHPP is a generalisation of a Poisson process where the non-negative rate function, $\lambda(t)$, is allowed to change through time, t. The probabilistic behaviour of a NHPP can be completely characterised by $\lambda(t)$ or the cumulative rate $\Lambda(t) = \int_0^t \lambda(s)ds$; all NHPP input modelling approaches aim to estimate one of these functions. We focus on modelling the rate, $\lambda(t)$.

Our main contributions are a spline-based method for modelling the rate function, $\lambda(t)$, of a NHPP and a simple algorithm to generate arrivals from the estimated rate function. We define a spline function as a linear combination of n cubic B-spline basis functions. Spline functions are piecewise-polynomials that are, by design, smooth and satisfy continuity constraints at the knots joining their pieces; see de Boor (1978). Cubic splines are continuous, and twice continuously differentiable everywhere; rate functions known to have jumps or non-differentiable points therefore lie outside of the scope of this paper. Spline functions become increasingly flexible as the number of basis functions used in their construction is increased. We propose to use a large number of basis functions, n, to model $\lambda(t)$ to enable a reduction in the bias with respect to the true arrival process. We do however wish to avoid over-fitting the model to the observed data, and therefore opt for a penalised log-likelihood approach. The penalty acts to reduce the variability and stabilise the resulting representation.

The paper is organised as follows: In Section 2 we discuss the current literature for modelling the rate function of a NHPP. In Section 3 our spline-based input model is presented, and in Section 4 we introduce a thinning method for simulating arrivals from it. In Section 5 we evaluate our method in comparison to relevant competitors and in Section 6 we conclude.

A preliminary proposal of the ideas presented here appeared in Morgan, Nelson, Titman, and Worthington (2019b), but it did not consider the consistency of the spline based estimator, or provide practical guidance for how to choose n. It also did not consider arrival generation which is key to the use of the spline-based estimator in practical simulation experiments. The evaluation has also been extensively extended including the addition of an evaluation of the method when fitting data that is over- or under-dispersed in comparison to a Poisson process and a consideration of input model uncertainty when using the spline-based method.

2. Background

A number of methods exist for modelling arrival processes using NHPPs. Our focus is on modelling the rate function of a NHPP using arrival-time observations. An alternative approach is to model the cumulative rate function, $\Lambda(t)$; see Leemis (1991) and references therein. Also, in some contexts arrival counts over intervals are the only

available data instead of arrival times; see Nicol and Leemis (2014) and references therein for input modelling methods using arrival counts. When only a single realisation of a NHPP arrival process is available Nelson and Leemis (2020) discuss methods for input modelling.

A common approach to modelling the rate function of a NHPP using arrival-time observations is to use an exponential function, $\lambda(t) = \exp\{g(t, \boldsymbol{\phi})\}$ where $g(t, \boldsymbol{\phi})$ describes polynomial or trigonometric components. This exponential form ensures the rate function is non-negative at all time points as required. Amongst others, Lewis and Shedler (1976), Lewis (1971), Kuhl, Wilson, and Johnson (1995), Kuhl, Wilson, and Johnson (1997) and Lee, Wilson, and Crawford (1991) adopted this approach. But note that numerically optimising parameters $\boldsymbol{\phi}$ in $g(t, \boldsymbol{\phi})$ within these methods is computationally expensive and often requires a good starting point.

Alternative approaches to modelling the rate function assume that the rate is a piecewise polynomial of some degree. Chen and Schmeiser (2017) start from a piecewise-constant representation of the rate function and present the MNO-PQRS algorithm that results in a piecewise-quadratic representation. This algorithm is not restricted to Poisson arrival processes. Zheng and Glynn (2017) present a piecewise-linear approach to modelling the rate function, and develop a convex programming formulation to estimate the rate at the interval boundaries given either arrival-times or arrival-count data. Kao and Chang (1988) present a piecewise polynomial representation, where the interval boundaries and polynomial degree are selected subjectively within each interval. This is generally an unrealistic assumption in practice when considering real-world non-stationary processes. In Section 5 we compare the spline-based model to the piecewise-linear model presented by Zheng and Glynn (2017), and the piecewise-quadratic model by Chen and Schmeiser (2017).

Channouf (2008) also uses splines to represent the rate function of both NHPPs and doubly stochastic Poisson processes. However, they do not take advantage of the B-spline composition of a spline function which requires many fewer parameters to be estimated, and is key to our arrival-generation approach in Section 4.

3. Fitting a spline function via penalised log-likelihood

Suppose we observe a NHPP with true rate function $\lambda^c(t)$, on the interval $[0, T]$, m times. We assume that $\lambda^c(t)$ is twice continuously differentiable, $\lambda^c(t) \in \mathcal{C}^2([0, T])$, like the cubic spline function that we use to represent it. We also assume that $\lambda^c(t)$ is bounded away from 0 within $[0, T]$, i.e. that there is no unknown downtime where $\lambda(t) = 0$.

We let m denote a number of days, but in practice m could represent other units of time. We use a cubic, degree $d = 3$, spline function to represent the rate function. This is a linear combination of n cubic basis functions, otherwise known as cubic B-splines. Let $B_{k,\boldsymbol{s}_k}(t)$ denote the kth cubic B-spline at time t defined over the ordered knot sequence $\boldsymbol{s}_k = \{s_{k-(d+1)}, s_{k-d}, \ldots, s_k\}$ where the knot points can be interpreted as points on the time axis. B-splines are locally defined functions; for $t \in (s_{k-(d+1)}, s_k)$, a cubic B-spline is non-negative and twice continuously differentiable, otherwise it is equal to 0. For $d > 1$, B-splines are composed recursively from lower degree B-splines as follows

$$B_{k,d,\boldsymbol{s}_k}(t) = \frac{t - s_{k-(d+1)}}{s_{k-1} - s_{k-(d+1)}} B_{k,d-1,\boldsymbol{s}_k}(t) + \frac{s_k - t}{s_k - s_{k-d}} B_{k+1,d-1,\boldsymbol{s}_{k+1}}(t),$$

where d denotes the degree of the B-spline; see de Boor (1978). The spline-based model of the NHPP rate is

$$\lambda(t; \boldsymbol{c}) = \sum_{k=1}^{n} c_k B_{k, \boldsymbol{s}_k}(t), \qquad (1)$$

where $c_k \in \Re$ is the spline coefficient of the kth B-spline for $k = 1, 2, \ldots, n$ and $\boldsymbol{c} = \{c_1, c_2, \ldots, c_n\}$.

The knot sequence upon which the spline function (1) is built combines the n knot vectors of its component B-splines, we denote this as $\boldsymbol{s}$, where $\boldsymbol{s} = \{s_{-d}, s_{-d+1}, \ldots, s_0, s_1, \ldots, \ldots, s_n\}$. It may seem unconventional to start the knot sequence $\boldsymbol{s}$ with knot s_{-d}, but this ensures that for all $t \in [0, T]$, $d + 1$ B-splines are non-zero. We also choose to use cardinal B-splines which are spline functions defined on uniformly spaced local knot sequences. This is not a requirement for the methods presented, but can be advantageous for arrival generation as we show in Section 4. From here on, we drop the knot sequence subscript on the B-spline and let $B_k(t)$ denote the kth B-spline.

Once the knot sequence, $\boldsymbol{s}$, has been set $B_{k, \boldsymbol{s}_k}(t)$ is fixed for all t for all B-splines. The shape of the spline rate function is therefore completely determined by the spline coefficients, $\boldsymbol{c}$. It is thus $\boldsymbol{c}$ that we optimise, given arrival-time observations, to fit the spline-based rate function. To optimise the spline coefficients, $\boldsymbol{c}$, we maximise the penalised log-likelihood of the NHPP. Our choice to use penalised likelihood stems from the computationally convenient form of the NHPP likelihood function, and our desire to exploit the assumption of the process being Poisson. We propose fitting the spline function using a large number of basis functions n allowing for a highly flexible representation. Penalisation then acts to prevent over-fitting and stabilise the representation.

Let a_i denote the number of arrivals observed on the ith day, and $0 \leq t_{i1} < t_{i2} < \cdots < t_{ia_i} \leq T$, for $i = 1, 2, \ldots, m$, denote the observed arrival times. The likelihood and log-likelihood of our spline-based model of a NHPP conditional on m days of observations over the interval $[0, T]$ are

$$L(\lambda(t; \boldsymbol{c})) \propto \prod_{i=1}^{m} \prod_{j=1}^{a_i} (\lambda(t_{ij}; \boldsymbol{c})) \exp \left\{ - \int_0^T \lambda(y; \boldsymbol{c}) dy \right\}^m, \text{ and}$$

$$l(\lambda(t; \boldsymbol{c})) \propto \sum_{i=1}^{m} \sum_{j=1}^{a_i} \log(\lambda(t_{ij}; \boldsymbol{c})) - m \int_0^T \lambda(y; \boldsymbol{c}) dy,$$

respectively. Our chosen penalty penalises rapid changes in the spline function using its integrated second derivative, a standard penalty for cubic splines within the smoothing

spline literature; see (Reinsch, 1967). The penalised log-likelihood is

$$l_p(\lambda(t;\boldsymbol{c})) \propto \sum_{i=1}^{m}\sum_{j=1}^{a_i} \log\left(\lambda(t_{ij};\boldsymbol{c})\right) - m\int_0^T \lambda(y;\boldsymbol{c})dy - \frac{1}{2}\theta\int_0^T \{\lambda''(u;\boldsymbol{c})\}^2 du$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{a_i} \log\left(\sum_{k=1}^{n} c_k B_k(t_{ij})\right) - m\sum_{k=1}^{n} c_k \int_0^T B_k(y)dy$$

$$- \frac{1}{2}\theta\sum_{k=1}^{n}\sum_{h=1}^{n} c_k c_h \int_0^T B_k''(u)B_h''(u)du, \qquad (2)$$

where θ controls the penalty term. Let $\widehat{\boldsymbol{c}}_\theta$ denote the maximum likelihood estimates of the spline coefficients found by maximising (2) at a fixed penalty value θ. For large θ rapid changes in the spline function are reduced forcing the fitted rate function closer to the overall mean rate; in the limit $\lim_{\theta\to\infty}\{\lambda''(t;\widehat{\boldsymbol{c}}_\theta)\}^2 = 0$ almost surely, and the fitted rate function will become linear in t, (Eilers & Marx, 1996). When $\theta = 0$, (2) is proportional to the log-likelihood of the NHPP.

In introducing penalisation we transfer control of the regularisation of the rate function into the penalty parameter, θ. The choice of the number of knots n is therefore not critical, but should be large enough to allow for the flexibility needed to fit the underlying rate function.

Consider fitting a spline-based rate function with a fixed number of B-splines, n, at fixed locations, and assume that the true rate function $\lambda(t;\boldsymbol{c}^0)$ falls within this family of models, where $\boldsymbol{c}^0$ are the true spline coefficients. If the rate function is truly a spline function then there is no need to penalise the log-likelihood to fit the model, and the penalty θ can be fixed at 0. With the penalty fixed at $\theta = 0$, and under assumptions common in the spline estimation literature, as $m \to \infty$ $\widehat{\boldsymbol{c}}_0 \xrightarrow{P} \boldsymbol{c}^0$, and the spline-based rate function converges to the true rate function; see Stone (1986) and Xue and Liang (2010). Note that in reality the true rate function is unlikely to be a spline function. In Section 3.1 we discuss how to select the optimal combination of the penalty parameter and spline coefficients, $\{\theta,\widehat{\boldsymbol{c}}_\theta\}$. For now we assume θ is fixed, and use a trust region algorithm to optimise the penalised log-likelihood and provide estimates of the spline coefficients.

The trust region algorithm is a well known optimisation approach that moves towards an optimum by taking steps within a region in which it trusts a local model of the function to be optimised. We point the reader to Conn, Gould, and Toint (2000) for more details. In practice we set the initial radius of the trust region to be large so we tend to get a globally good solution. We use a second-order Taylor series as a local model of the penalised log-likelihood. Note that our choice of penalty enables easy calculation of the gradient and the Hessian for this local model. At each step, the second-order model leads to a convex, quadratic trust region sub-problem with a unique solution. We add an additional constraint in the trust region sub-problem, forcing all spline coefficients to be non-negative, $\boldsymbol{c} \geq 0$. This ensures the rate function, $\lambda(t;\boldsymbol{c})$, is non-negative, but we acknowledge that this constraint is stronger than necessary since negative spline coefficients are possible whilst still maintaining a positive rate function. Another implication of non-negative spline coefficients is that all spline coefficients can themselves be treated as rates which is advantageous for arrival generation; see Section 4. When the rate function is cyclic in nature we add constraints of the form: $\lambda(0;\boldsymbol{c}) = \lambda(T;\boldsymbol{c})$ $\lambda'(0;\boldsymbol{c}) = \lambda'(T;\boldsymbol{c})$, $\lambda''(0;\boldsymbol{c}) = \lambda''(T;\boldsymbol{c})$. The trust re-

gion algorithm has a number of control parameters and threshold values; we used the values suggested by Wright and Nocedal (1999). Morgan (2019) discusses the possibility of slow or poor convergence of the trust region algorithm as the number of knots used in the spline-based rate function is increased. Increasing the number of knots increases the flexibility of the spline-based function over small intervals and thus we must take smaller steps in the trust region algorithm to ensure the local second order model of the penalised likelihood is a valid approximation. In some cases this could cause the algorithm to terminate before optimum spline coefficients have been found. Note that alternative optimisation algorithms could be used in place of the trust region algorithm, but the trust region algorithm performed well within the limits of our experiments.

3.1. Selecting $\{\theta, \widehat{\boldsymbol{c}}_\theta\}$

Information criterion estimate prediction error. They can be used as a measure of relative model quality, for a given set of data, in the process of model selection. Our approach for choosing the combination $\{\theta, \widehat{\boldsymbol{c}}_\theta\}$ utilises a modification of the AIC score of Cavanaugh and Neath (2011), known as the regularisation information criterion (RIC); see Dixon and Ward (2018) and Shibata (1989). This score is based on Kullback-Leibler (KL) information, a measure of the distance between two distributions (Kullback, 1997). RIC trades off the goodness-of-fit of a proposed model and its complexity, whilst also taking into account the penalisation of the log-likelihood. If the combination $\{\theta, \widehat{\boldsymbol{c}}_\theta\}$ was selected by maximising the penalised log-likelihood alone, then a penalty parameter of $\theta = 0$ would always be chosen, as the unpenalised log-likelihood is more able to adapt to the characteristics in the observed data. The RIC score is defined as follows:

$$\begin{aligned} \text{RIC} &= -2\, l(\lambda(\boldsymbol{t}; \widehat{\boldsymbol{c}}_\theta)) + 2\, e, \\ &= -2\, l(\lambda(\boldsymbol{t}; \widehat{\boldsymbol{c}}_\theta)) + 2\, \text{tr}(I_p(\widehat{\boldsymbol{c}}_\theta) J_p(\widehat{\boldsymbol{c}}_\theta)^{-1}), \end{aligned}$$

where e can be considered a measure of the effective degrees of freedom and $I_p(\widehat{\boldsymbol{c}}_\theta)$ and $J_p(\widehat{\boldsymbol{c}}_\theta)$ are defined as

$$I_p(\widehat{\boldsymbol{c}}_\theta) = \sum_{i=1}^{m} \left[\frac{\partial}{\partial \boldsymbol{c}} l_p(\lambda(\boldsymbol{t}_i; \widehat{\boldsymbol{c}}_\theta))\, \frac{\partial}{\partial \boldsymbol{c}^T} l_p(\lambda(\boldsymbol{t}_i; \widehat{\boldsymbol{c}}_\theta)) \right]$$

$$J_p(\widehat{\boldsymbol{c}}_\theta) = -\frac{\partial^2}{\partial \boldsymbol{c} \partial \boldsymbol{c}^{\mathrm{T}}} l_p(\lambda(\boldsymbol{t}; \widehat{\boldsymbol{c}}_\theta)).$$

The combination, $\{\theta, \widehat{\boldsymbol{c}}_\theta\}$, that minimises the RIC score is chosen.

Let $\theta^\star$ denote the optimal penalty value. Given a fixed penalty value θ, the optimal spline coefficients, $\widehat{\boldsymbol{c}}_\theta$, are found by trust region optimisation. The search for the combination $\{\theta^\star, \widehat{\boldsymbol{c}}_{\theta^\star}\}$ that minimises the RIC score therefore reduces to a one-dimensional line search for $\theta^\star \in [0, \infty)$. For speed, we propose starting the search with a high initial penalty value, $\theta_0 = \eta$, then jumping backwards towards 0 by halving the penalty at each step, $\theta_0 = \eta, \theta_1 = \eta/2, \theta_2 = \eta/4, \ldots$; allowing us to take larger steps initially. By evaluating the RIC at each step we can identify an interval of penalty values, O, in which at least a local minimum of RIC lies. We must check that we are moving towards the minimum RIC in the first step of the search by checking that

$RIC_{\theta_0} > RIC_{\theta_1}$. When this is not the case a new, higher, η should be chosen and the procedure restarted. Assuming the initial penalty passes this check; the following algorithm describes the kth step of the search:

(1) Fix $\theta_k = \frac{1}{2^k}\eta$
    (a) Evaluate $\widehat{\boldsymbol{c}}_{\theta_k}$ using the trust region algorithm.
    (b) Evaluate RIC_{θ_k}
(2) If $RIC_{\theta_k} > RIC_{\theta_{k-1}}$, then stop the search and set $O = (\theta_k, \theta_{k-1})$.
(3) Else $k = k + 1$. Return to Step 1.

When the algorithm terminates at step k, we then complete a more intensive search for $\theta^\star \in O = \{\theta_k, \theta_{k-1}\}$. In practice we use the R function 'optimise' (R Core Team. 2018) for the search of the narrower interval O. There is no guarantee that the RIC score function is convex, however in practice lack of convexity only appeared to be an issue around low penalty values, which is not concerning as the final spline rate function does not change much for small changes in a small penalty. In practice the search procedure outlined above worked well. An alternative approach is to do a simple grid search to study the RIC over a large interval.

Returning to the choice of the number of B-spline basis functions, n. As of yet there is not theoretical guidance for the choice of n but we follow the practical guidelines provided by Wood (2003). We know that in using penalisation the choice of n is not critical but it should be large enough to allow for the flexibility needed to fit the underlying rate function. To monitor whether we have enough flexibility we might observe the value of the optimal penalty suggested by our algorithm, $\theta^\star$. If $\theta^\star$ is very small this suggests we are not penalising the fit of the model much and the choice of n is too small. It is however hard to give a general value for a 'small' penalty; this is likely to be context specific. As an alternative Wood (2003) suggests comparing the effective degrees of freedom of the model with and without penalisation. By introducing a penalty to (2) we reduce the degrees of freedom in the model, so if the effective degrees of freedom after penalisation is close to the effective degrees of freedom without any penalisation (when $\theta = 0$) this suggests that penalisation was not required, and thus that n was too small. In practice Wood (2003) suggests increasing n if the effective degrees of freedom of the penalised model exceed some proportion between 0.8 to 0.9 of n. In all of the experiments in Section 5 this was either checked or used as a guideline to choose n.

At this point we have provided a spline-based representation, $\lambda(t; \widehat{\boldsymbol{c}}_{\theta^\star})$, of the rate function of a NHPP.

4. Generating arrivals from the spline rate function

Arrival generation is a requirement for the spline-based rate function to be used as an input process in stochastic simulation. As we directly model the rate function, thinning (or rejection) is arguably the most appropriate method for arrival generation. For thinning some majorising function is required; see Devroye (2006). The overall maximum of a spline made of component B-splines is not straightforward to calculate, but we do know the maximum of each B-spline function. We therefore propose generating arrivals from $\lambda(t; \widehat{\boldsymbol{c}}_{\theta^\star})$ by using a thinning algorithm on each of the scaled B-spline components, $\lambda_k(t) = c_k B_k(t)$, for $k = 1, 2, \ldots, n$. By the superposition property of a Poisson process, the superposition of arrivals generated from each of the n scaled B-splines is equivalent to arrivals generated from the spline function $\lambda(t) = \sum_{k=1}^{n} c_k B_k(t)$; see

(Kingman, 1992). In essence we are treating each of the scaled B-splines as the rate function of a NHPP. The kth cubic B-spline has a maximum at its central knot, s_{k-2}; thus the maximum of the kth scaled B-spline is $c_k B_k(s_{k-2})$.

Let us first consider the generation of arrivals from a single scaled B-spline, $\lambda_{d+1}(t) = \lambda_4(t) = c_4 B_4(t)$, with maximum at knot $s_{d+1-2} = s_2$, as $B_4(t)$ is the first cubic B-spline with fully positive support $[0, s_4]$. Let $B_4^\star(t)$ denote a function that majorises the cubic B-spline such that $B_4^\star(t) \geq B_4(t)$ for all $t \in [0, s_4]$; a simple choice of majorising function is $B_4^\star(t) = \max_t B_4(t) = B_4(s_2)$ for $t \in [0, s_4]$. The scaled majorising function, $c_4 B_4^\star$, therefore majorises the scaled B-spline, $c_4 B_4(t)$. Using thinning, an arrival at time $q^\star$ generated from $c_4 B_4^\star(t)$ has probability of rejection $1 - B_4(q^\star)/B_4^\star(q^\star)$. Thinning can be used in the same way to generate arrivals from the remaining $k = 2, \ldots, n$ scaled B-splines.

By using cardinal B-splines, arrival generation can be simplified further as each of the n spline components, $\lambda_k(t; \boldsymbol{c})$, are a scaled translation of any other. Let h denote the uniform difference between two successive knots, then $\lambda_k(t; \boldsymbol{c}) = c_k B_4(t - h(k-4))$ for all k. This means we can use the first B-spline, scaled by each of the spline coefficients in turn, to generate all arrivals and then translate the arrivals to the correct interval. The following algorithm provides an overview of how to generate a sequence of arrivals, $q_1, q_2, \ldots$, from a spline-based rate function constructed from cardinal B-splines. To avoid excessive rejections arrival generation for B-splines $B_1(t), B_2(t), B_3(t), B_{n-2}(t), B_{n-1}(t)$ and $B_n(t)$ is restricted to the part of their support that falls within the interval of interest $[0, T]$.

(0) Preliminary Step. Let h be the difference between any two knots in the uniform knot sequence of the spline function.
(1) For j in 1 to n:
    (a) Generate arrivals $q_{1j}^\star, q_{2j}^\star, \ldots$ from the scaled majorising function $c_j B_4^\star(t)$. If $1 \leq j \leq 3$ then restrict arrival generation to the interval $[s_{4-j}, s_4]$. If $n - 2 \leq j \leq n$ then restrict arrival generation to the interval $[s_0, s_{n-j+1}]$.
    (b) Thin arrivals, $q_{1j}^\star, q_{2j}^\star, \ldots$, with probability of thinning $1 - \frac{B_4(\cdot)}{B_4^\star(\cdot)}$, leaving arrivals $q_{1j}, q_{2j}, \ldots$ from NHPP with rate function $c_j B_4(t)$.
    (c) Translate the arrivals to the jth B-spline knot sequence by adding $h \times (j-4)$ to each of the arrival times in turn.
(2) Superpose and sort the arrivals from the n B-splines.

This algorithm can be repeated if more than a single sequence is required. To deploy the algorithm the user will require storage capacity of the order of the expected number of arrivals. If the arrival rate is extremely high this may impose a demand on memory resources, but even tens of thousands of arrivals would not tax modest memory resources. Another consideration is the potential expense of sorting the arrivals that takes place in the last step of the algorithm. By generating the arrivals in an ordered way (from the first B-spline to the last) the arrivals will be pretty well ordered before sorting. We therefore do not expect this sort to be computationally prohibitive even when large numbers of arrivals are generated. An R package with the functionality to fit and generate arrivals from a spline-based rate function was developed alongside this project. This package is available on GitHub at www.github.com/morganle/NHPPspline (Morgan, Titman, Worthington, & Nelson, 2020).

Reducing the number of arrivals rejected within the thinning algorithm will increase its efficiency. This is particularly important in simulation experiments requiring the generation of a large number of arrivals. To this end we propose using our knowledge

of the shape and maximum of each B-spline to create a tighter majorising function. For the algorithm above this simply amounts to creating a single majorising function for $B_4(t)$. To retain the advantages of thinning, arrival generation from the proposed majorising function should be simple and efficient. Klein and Roberts (1984) propose a highly efficient and simple method for generating arrivals from a piecewise-linear rate function based on inversion. We therefore propose a piecewise-linear majorising function for $B_4(t)$. The ratio of the area under the majorising function to the area under the B-spline is an indicator of how tight the fit of a majorising function is with 1 being perfect agreement. This ratio can be interpreted as the expected number of possible arrivals needed to generate one arrival from the function of interest.

For illustration consider a majorising function with 5 piecewise-linear segments. Due to the symmetry of the B-spline we need only focus on the first and second segments; the third segment is a horizontal line at the B-spline maximum $B_4(s_2)$. Once the tangent point at which the second segment touches the B-spline has been selected, optimising the majorising function amounts to a one-dimensional search for the optimal position of the point joining the first and second piecewise-linear segments. Figure 1 illustrates both a constant majorising function (red), and our optimised 5 segment piecewise-linear majorising function (blue) for B-spline $B_4(t)$ (black). The ratio of the area under the B-spline to the area under the constant majorising function is 2.67, and is 1.07 for the piecewise-linear majorising function; a large improvement. Using more segments in the construction of the piecewise-linear majorising function could reduce this ratio further, but a ratio of 1.07 is already a very efficient algorithm.
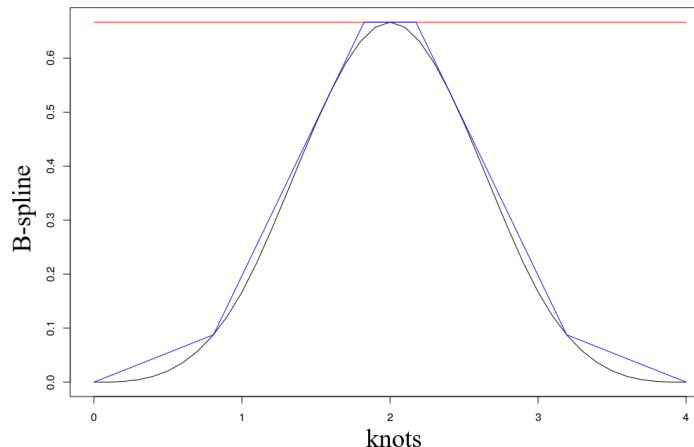

Figure 1. B-spline $B_4(t)$ (black) plus, a piecewise-constant majorising function (red) and a piecewise-linear majorising function (blue).

5. Evaluation

In this section we evaluate our spline-based input modelling method by comparing it to two input modelling methods that have recently been presented in the literature in a controlled experiment where all true intensities are known. We also compare the methods in terms of how much input modelling error they pass to the output of a controlled simulation experiment. Finally we investigate the robustness of the methods for fitting the rate function of an input process when the observations are under- or

overdispersed in comparison to a Poisson process.

The chosen methods for comparison also allow estimation of the rate function of a Poisson process given arrival-time observations. These methods are the piecewise-quadratic input model presented by Chen and Schmeiser (2017), known as MNO-PQRS, and the piecewise-linear approach by Zheng and Glynn (2017).

A practical example of using the spline-based method for modelling arrivals to an A&E department is provided in Morgan (2019).

5.1. Recovering the rate function

Our first comparison considers metrics that indicate how well the true rate function, $\lambda^c(t)$, is recovered. We use the average percentage discrepancy (APD), δ, and the maximum percentage discrepancy (MPD), ζ,

$$\delta = \frac{1}{T} \int_0^T \left| \frac{\lambda(q; \widehat{\boldsymbol{c}}) - \lambda^c(q)}{\lambda^c(q)} \right| \, dq$$
$$\zeta = \max_{0 \leq q \leq T} \left| \frac{\lambda(q; \widehat{\boldsymbol{c}}) - \lambda^c(q)}{\lambda^c(q)} \right|$$

which are also used by Liu, Kuhl, Liu, and Wilson (2019) for evaluating their method for modelling nonhomogeneous non-Poisson processes. We also record the coefficient of variation of the APD, denoted ι, as an indicator of the dispersion of the fit of each method. For each level of input data, m, we fit the rate function $G = 500$ times. The same m days of arrival observations were used for all three methods for each replication. Average APD and average MPD over the $G = 500$ replications, denoted $\bar{\delta}$ and $\bar{\zeta}$ respectively, were recorded.

Both methods to which we compare assume that the number and position of the intervals from which the piecewise intensities are built are known. Morgan et al. (2019b) compared the three methods for a piecewise rate function with known interval boundaries in an experiment that was designed to be as advantageous to the competing methods as possible. Nevertheless when the true rate function was piecewise-linear the spline-based method performed well. In reality it is unlikely that the number and position of the intervals are known, even if the function is truly piecewise-linear, which is also unlikely. In fact, unless the rate function is truly piecewise it is impossible to say what the optimal placement of the intervals is. We choose to utilise the data-driven method by Chen and Schmeiser (2019) to pre-process each of the $G = 500$ data sets. The method aims to identify the optimal number of equal length intervals $k^\star$ from which to start the MNO-PQRS algorithm presented by Chen and Schmeiser (2017).

During the experiment the interval placement was the same for the piecewise-linear and piecewise-quadratic methods in each replication. For the spline-based method, 50 equally spaced knots were used, equating to $n = 46$ B-spline basis functions. In the majority of replications this led to effective degrees of freedom less than 0.8 times the effective degrees of freedom of the unpenalised model indicating that $n = 46$ was 'large enough' to represent the rate function. To look at this in more detail we follow this experiment with an investigation of the penalty and effective degrees of freedom in each experiment. In total 9 rate functions were considered for three levels of input data totalling 27 experiments for each input modelling method. The rate functions followed a general sinusoidal shape with the addition of a peak, see Figures 2 and 3 which show $G = 500$ estimated rate functions for each of the three input modelling methods for

two of the nine considered rate functions, one with a short sharp peak and the other with a long duration peak. Within the following figures and tables the three input modelling methods are denoted "SPL" (spline), "PQ" (piecewise-quadratic) and "PL" (piecewise-linear). The results of all 27 experiments are listed in the supplementary material available with this paper.

In all but one of the 27 experiments the spline-based method out-performed the piecewise quadratic and the piecewise-linear approaches attaining the lowest average MPD, $\bar{\zeta}$. The spline-based method also achieved the smallest average APD, $\bar{\delta}$, in all but a few experiments where the rate function had a short duration peak. In these cases MNO-PQRS achieved a smaller average APD by a small margin. Increasing the number of basis functions would perhaps lead to a decrease in the APD for the spline-based model given a rate function with an abrupt but brief change in behaviour. All three methods struggled to estimate the short duration peak, as seen in Figure 2. As peak duration increased the average MPD decreased for all methods for all levels of data. In all experiments the location of the MPD tended to be at the centre of the peak, time $t = 15$. In around 75% of cases the spline-based method achieved the lowest coefficient of variation of the APD; this indicates the method has improved stability compared to the piecewise methods.

As the number of sets of observations, m, increased all methods improved for both metrics. In Table 1 we present the results of the two most extreme cases: one rate function with a high magnitude but short duration peak, and one with a low magnitude but long duration peak. Figures 2 and 3 show the fit of the $G = 500$ rate functions for these two experiments when $m = 30$.

In Table 2 we compare metrics for the spline-based rate function for $n = 23$ and $n = 46$ basis functions for the same rate functions used in Table 1. We expected to

Table 1. The average MPD, $\bar{\zeta}$, the average APD, $\bar{\delta}$, and the coefficient of variation of the APD, ι, for two of the 27 rate functions.

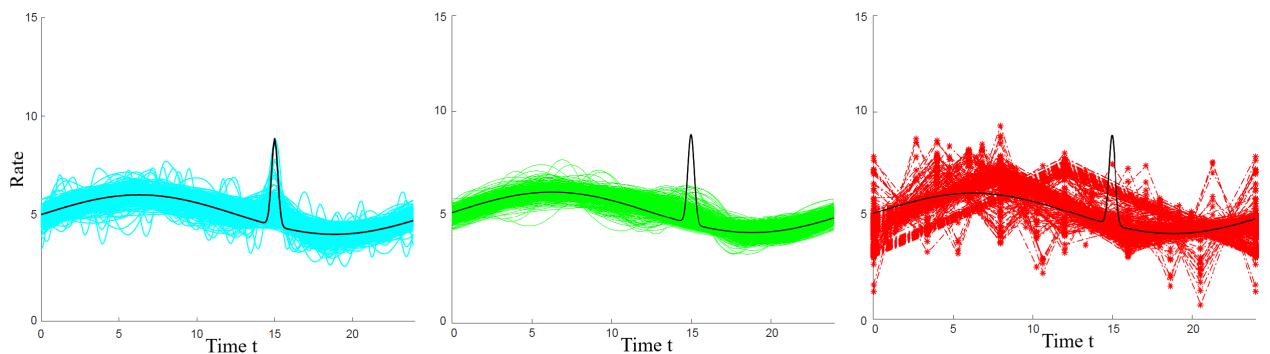|	$m = 15$, short sharp peak			$m = 100$, long duration peak		
	ζ (se)	δ (se)	ι	ζ (se)	δ (se)	ι
SPL	41.77% (0.23)	6.61% (0.09)	0.26	5.58% (0.07)	1.98% (0.03)	0.29
PQ	45.47% (0.13)	5.92% (0.09)	0.33	30.64% (0.13)	2.26% (0.03)	0.25
PL	48.97% (0.27)	13.15% (0.30)	0.51	28.29% (0.35)	10.86% (0.17)	0.36

Figure 2. $G = 500$ fits of the rate function with a short sharp peak (black) with $m = 30$ days of input data. Left shows SPL (cyan), ecntre shows PQ (green) and right shows PL (red).
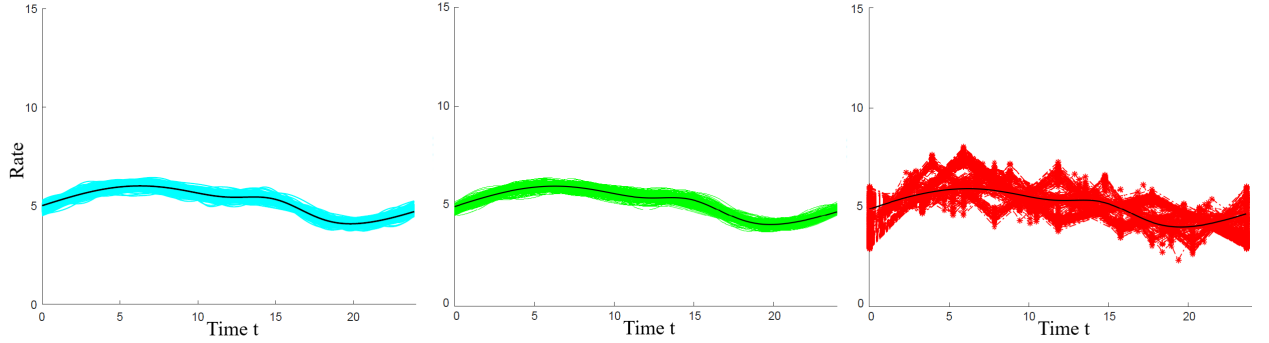
11

Figure 3. $G = 500$ fits of the rate function with a long duration peak (black) with $m = 30$ days of input data. Left shows SPL (cyan), Centre shows PQ (green), and right shows PL (red).

see a decrease in average penalty for lower n which would also be reflected in the additional metric $\overline{e_\theta/e_0}$ describing the average proportion of the penalised effective dof to the unpenalised effective dof. The intensities constructed using $n = 23$ were slightly better able to recover the true rate function. They achieved slightly smaller average APD, and in all cases but the ones with short high peaks the average MPD was also smaller. When $n = 46$ the spline-based representation was more flexible and therefore better able to represent the peak in the rate function. The improvement in MPD from $n = 23$ to $n = 46$ became more evident as the number of days of data used to fit the rate function, m, increased.

The impact of the short sharp peak in the experiments above brings into the question the use of a single uniform penalty for the entire spline-based representation. Due to the local nature of the B-spline basis functions it would be possible to penalise regions of the rate function unequally allowing additional flexibility in local areas of the function. We leave this as an idea for future work.

5.2. Propagation of input modelling error

At this point we have shown that the spline-based method recovers the true rate function well. We now consider the propagation of input modelling error, caused by the estimation of the rate function of a NHPP, to the output of a $M(t)/M/\infty$ simulation model. The output performance measure studied is the expected number of people in the system, $E(\bar{N})$, where $\bar{N} = \frac{1}{T} \int_0^T N(t)dt$. The service rate μ is treated as known so all input modelling error can be assumed to come from the arrival process. Input modelling error can be broken down into bias caused by input modelling, b, and the

Table 2. The average MPD, $\bar{\zeta}$, the average APD, $\bar{\delta}$, the average penalty, $\bar{\theta}$, and the average proportion of the penalised edof to the unpenalised edof, $\overline{e_\theta/e_0}$, used with our check on n, for two rate functions constructed from $n = 23$ and $n = 46$ B-spline basis functions.

	$m = 15$ short sharp peak				$m = 100$ long duration peak			
	$\bar{\zeta}$	$\bar{\delta}$	$\bar{\theta}$	$\overline{e_\theta/e_0}$	$\bar{\zeta}$	$\bar{\delta}$	$\bar{\theta}$	$\overline{e_\theta/e_0}$
$n = 23$	42.66%	6.27%	12.44	0.34	5.52%	1.97%	20.13	0.35
$n = 46$	41.77%	6.61%	54.68	0.22	5.58%	1.98%	91.36	0.18

12

variance caused by input modelling, or input uncertainty, σ_I^2 (Morgan, Nelson, Titman, & Worthington, 2019a). We shall consider both components.

All 9 arrival processes introduced in Section 5.1 were considered for $m = 15, 30, 100$. This gave a total of 27 experiments for us to compare the spline-based method to the piecewise input modelling methods. The $G = 500$ fits of the rate function created for the experiment in Section 5.1 were used as input to the simulation model. In each simulation experiment $r = 100$ replications of the simulation were run, and the results averaged to reduce the intrinsic error in the estimate of the expected number in the system caused by random variation within the simulation.

Knowledge of the true rate function allowed us to estimate the bias caused by input modelling, b, by comparing the output of the simulation run using the true rate function to the output of the simulation run using $\widehat{\lambda}(t)$, an estimated rate function created using each of the three methods. Input uncertainty, σ_I^2, was approximated by subtracting the stochastic estimation error from the total simulation error given the simulation output from the $G = 500$ estimated rate functions; see Nelson and Pei (2021). Table 3 shows an average estimate of bias caused by input modelling, and an estimate of the square root of input uncertainty for the two rate functions from Table 1. For the spline we use $n = 46$ basis functions. In the supplementary material available with this paper the results of all 27 experiments are presented.

In all 27 experiments the spline-based rate function achieved the lowest input uncertainty variance, and in $> 80\%$ of experiments it also achieved the lowest average bias caused by input modelling. When $m = 100$, the highest level of input data we tested, the spline-based rate function propagated the least bias and variance caused by input modelling to the simulation output. These results are a promising indication that the spline-based input modelling method passes less input modelling error to the output of a simulation model than rival methods.

5.3. Under and over-dispersed data

Our final consideration is to test the robustness of the methods to non-Poisson observations. Specifically, we consider both underdispersed and overdispersed data. Arrival times were generated from a Markov-MECO process using the Markov-MECO-based tool for generating nonhomogeneous non-renewal arrival processes presented by Nelson and Gerhardt (2011). The Markov-MECO-based tool allows the user to select a target squared coefficient of variation, cv^2, of the process. For Poisson distributed data the squared coefficient of variation, cv^2, equals 1, by definition of a Poisson process. We consider underdispersed, $cv^2 = 0.5$, and overdispersed, $cv^2 = 1.5$ data for the 9 intensities investigated in Section 5.1; the averaged metrics from fitting $G = 500$ rate

Table 3. The estimate average bias caused by input modelling $\bar{b}$ and an estimate of the square root of the input uncertainty, $\widehat{\sigma}_I$, for the fit of two rate functions. The expected number in the system having run the simulation with the true rate function was 0.512 for both systems.

	$m = 15$, short sharp peak			$m = 100$, long duration peak		
	$\mathrm{E}(\bar{N})$	$\bar{b}$	$\widehat{\sigma}_I$	$\mathrm{E}(\bar{N})$	$\bar{b}$	$\widehat{\sigma}_I$
SPL	0.502	0.010	0.049	0.515	0.004	0.045
PQ	0.479	0.033	0.074	0.482	0.030	0.076
PL	0.478	0.034	0.120	0.485	0.026	0.125

functions are presented in Tables 4 and 5, respectively. We use $n = 46$ B-spline basis functions to fit the spline intensities, and the pre-processing method of Chen and Schmeiser (2019) was again utilised to select the number of equal length intervals on which to construct the intensities for the competing methods.

In Tables 4 and 5 the spline-based method gives the smallest average MPD for the two experiments for both under and overdispersed data. This held for all bar one of the 27 experiments given underdispersed data and all of the experiments for overdispersed data. When $m = 100$ in Tables 4 and 5 we see the spline-based method provides a considerable improvement in average MPD.

The spline-based rate function also achieved the smallest APD in the majority of experiments given underdispersed data, and in all but three of the experiments with overdispersed data. Just as in the original experiment from Section 5.1, the spline-based method appeared to perform worse in cases where the duration of the peak in the rate function was shortest which reiterates that rate functions with short abrupt peaks are a challenge for the spline-based method. This problem might be alleviated by increasing the number of basis functions, n, in the spline function construction.

These results are promising as in most of the 27 experiments the spline-based method outperforms the piecewise-quadratic method, MNO-PQRS, which is designed for general, not specifically Poisson, input models. In all 27 experiments the piecewise-linear approach performed worse than the spline-based rate function in terms of both metrics. This may be partially due to our use of the MNO-PQRS pre-processing technique for the selection of the number of intervals from which to run the algorithm. A bespoke algorithm to find the optimal choice of starting point for the piecewise-linear method may have led to better results.

By considering the same rate functions presented in Section 5.1 we can directly compare the results of the fit of the rate function for underdispersed, Poisson and overdispersed data. Looking at the results in Tables 1, 4 and 5, it appears that all methods perform worst when the arrivals are overdispersed; this held for all experiments.

Within the parameters of this experiment we have demonstrated that when the arrival data departs from Poisson, by being over- or underdispersed, the spline-based

Table 4. The average MPD, $\bar{\zeta}$, the average APD, $\bar{\bar{\delta}}$, and the coefficient of variation of the integrated absolute difference, ι, for the fit of two rate functions given underdispersed data.

$cv^2 = 0.5$	$m = 15, \kappa = 5, \xi = 1$			$m = 100, \kappa = 1, \xi = 10$		
	$\bar{\zeta}$ (se)	$\bar{\bar{\delta}}$ (se)	ι	$\bar{\zeta}$ (se)	$\bar{\bar{\delta}}$ (se)	ι
SPL	40.16% (0.25)	5.82% (0.07)	0.27	6.69% (0.09)	2.40% (0.04)	0.39
PQ	45.61% (0.11)	5.25% (0.07)	0.28	29.76% (0.11)	2.48% (0.04)	0.36
PL	48.97% (0.27)	13.14% (0.29)	0.51	28.29% (0.34)	10.86% (0.17)	0.36

Table 5. The average MPD, $\bar{\zeta}$, the average APD, $\bar{\bar{\delta}}$, and the coefficient of variation of the integrated absolute difference, ι, for the fit of two rate functions given overdispersed data.

$cv^2 = 1.5$	$m = 15, \kappa = 5, \xi = 1$			$m = 100, \kappa = 1, \xi = 10$		
	$\bar{\zeta}$ (se)	$\bar{\bar{\delta}}$ (se)	ι	$\bar{\zeta}$ (se)	$\bar{\bar{\delta}}$ (se)	ι
SPL	43.69% (0.23)	8.00% (0.13)	0.36	9.32% (0.14)	4.20% (0.10)	0.53
PQ	45.28% (0.18)	7.94% (0.14)	0.39	31.47% (0.18)	4.37% (0.10)	0.50
PL	50.10% (0.42)	15.81% (0.29)	0.42	28.74% (0.36)	10.53% (0.17)	0.36

method performs well in terms of the metrics of average APD and average MPD relative to the piecewise-quadratic and piecewise-linear methods. This indicates that the spline-based input modelling approach is robust to arrivals that are under- or overdispersed in comparison to a Poisson distribution. In the supplementary material we report the proportion of times that the spline-based input model achieved the smallest MPD, ζ, or APD, δ, compared to the piecewise-linear and piecewise-quadratic input models across the $G = 500$ macro replications. With proportions exceeding 0.8 for the vast majority of cases, the spline-based method clearly performs very well.

6. Conclusion

In this paper we presented a spline-based method for modelling and generating arrivals from a NHPP. In a controlled comparison to two recent methods in the literature, the spline-based method was seen to perform very well in terms of recovering the true rate function of a range of NHPPs. The spline-based rate function was also seen to dominate its competitors in terms of how much input modelling error it passed to the output of a simple $M(t)/M/\infty$ queueing model.

Given both under and overdispersed arrivals compared to a Poisson process the spline-based method performed very well at recovering the underlying rate function. Within the constraints of the experiment the spline-based method for estimating $\lambda(t)$ appeared to be robust to non-Poisson data.

The other key contribution of the paper was a thinning-based method for simulating arrivals from the resulting spline-based rate function. The method takes advantage of the composition of the spline function as a linear combination of B-spline basis functions with known maxima. We presented this method alongside an algorithm for its implementation, and provide a link to an R package for its practical implementation.

In practice, arrival counts are sometimes recorded instead of arrival times. Our method could be extended to work for arrival count data through a simple modification of the log-likelihood. In the same way, provided an appropriate likelihood can be derived, the penalised log-likelihood method could be extended for use with other non-stationary non-Poisson arrival processes. We leave these extensions for future work.

References

Cavanaugh, J. E., & Neath, A. A. (2011). Akaike's Information Criterion: Background, Derivation, Properties, and Refinements. In *International encyclopedia of statistical science* (pp. 26–29). Springer.

Channouf, N. (2008). *Modélisation et Optimisation d'un Centre d'appels Téléphoniques: étude du Processus d'arrivée* (Unpublished doctoral dissertation). Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Montréal.

Chen, H., & Schmeiser, B. W. (2017). MNO–PQRS: Max Nonnegativity Ordering Piecewise-Quadratic Rate Smoothing. *ACM Transactions on Modeling and Computer Simulation*, *27*(3), 1–19.

Chen, H., & Schmeiser, B. W. (2019). Mise-optimal intervals for mno–pqrs estimators of poisson rate functions. In N. Mustafee, K.-H. Bae, S. Lazarova-Molnar, C. S. M. Rabe, P. Haas, & Y.-J. Son (Eds.), *Proceedings of the 2019 Winter Simulation Conference* (pp. 368–379).

Conn, A. R., Gould, N. I., & Toint, P. L. (2000). *Trust Region Methods* (Vol. 1). Philadelphia: Society for Industrial and Applied Mathematics (SIAM).

de Boor, C. (1978). *A Practical Guide to Splines* (Vol. 27). New York: Springer-Verlag.

Devroye, L. (2006). Nonuniform random variate generation. *Handbooks in operations research and management science*, *13*, 83–121.

Dixon, M., & Ward, T. (2018). *Takeuchi's Information Criteria as a Form of Regularization.* (https://arxiv.org/pdf/1803.04947.pdf)

Eilers, P. H., & Marx, B. D. (1996). Flexible Smoothing with B-Splines and Penalties. *Statistical science*, 89–102.

Kao, E. P., & Chang, S.-L. (1988). Modeling Time-Dependent Arrivals to Service Systems: A Case in using a Piecewise-Polynomial Rate Function in a Nonhomogeneous Poisson Process. *Management Science*, *34*(11), 1367–1379.

Kingman, J. F. C. (1992). *Poisson Processes* (Vol. 3). Clarendon Press.

Klein, R. W., & Roberts, S. D. (1984). A Time-Varying Poisson Arrival Process Generator. *Simulation*, *43*(4), 193–195.

Kuhl, M. E., Wilson, J. R., & Johnson, M. A. (1995). Estimation and Simulation of Nonhomogeneous Poisson Processes having Multiple Periodicities. In C. Alexopoulos, K. Kang, W. R. Lilegdon, & D. Goldsman (Eds.), *Proceedings of the 1995 Winter Simulation Conference* (pp. 374–383). Piscataway, New Jersey.

Kuhl, M. E., Wilson, J. R., & Johnson, M. A. (1997). Estimating and Simulating Poisson Processes Having Trends or Multiple Periodicities. *IIE Transactions*, *29*(3), 201–211.

Kullback, S. (1997). *Information Theory and Statistics*. Courier Corporation.

Lam, H. (2016). Advanced Tutorial: Input Uncertainty and Robust Analysis in Stochastic Simulation. In T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, & S. E. Chick (Eds.), *Proceedings of the 2016 Winter Simulation Conference* (pp. 178–192). Piscataway, New Jersey.

Lee, S., Wilson, J. R., & Crawford, M. M. (1991). Modeling and Simulation of a Nonhomogeneous Poisson Process having Cyclic Behavior. *Communications in Statistics-Simulation and Computation*, *20*(2-3), 777–809.

Leemis, L. M. (1991). Nonparametric Estimation of the Cumulative Intensity Function for a Nonhomogeneous Poisson Process. *Management Science*, *37*(7), 886–900.

Lewis, P. A. (1971). *Recent Results in the Statistical Analysis of Univariate Point Processes* (Tech. Rep.). Naval Postgraduate School. Monterey, California.

Lewis, P. A., & Shedler, G. S. (1976). *Statistical Analysis of Non-stationary Series of Events in a Data Base System* (Tech. Rep.). Naval Postgraduate School. Monterey, California.

Liu, R., Kuhl, M. E., Liu, Y., & Wilson, J. R. (2019). Modeling and simulation of nonstationary non-poisson arrival processes. *INFORMS Journal on Computing*, *31*(2), 347–366.

Morgan, L. E. (2019). *Quantifying and reducing error caused by input modelling in simulation* (Unpublished doctoral dissertation). Lancaster University, UK.

Morgan, L. E., Nelson, B. L., Titman, A. C., & Worthington, D. J. (2019a). Detecting Bias due to Input Modelling in Computer Simulation. *European Journal of Operational Research*. (doi: https://doi.org/10.1016/j.ejor.2019.06.003)

Morgan, L. E., Nelson, B. L., Titman, A. C., & Worthington, D. J. (2019b). A spline-based method for modelling and generating a nonhomogeneous poisson process. In N. Mustafee, K.-H. Bae, S. Lazarova-Molnar, C. S. M. Rabe, P. Haas, & Y.-J. Son (Eds.), *Proceedings of the 2019 Winter Simulation Conference* (pp. 356–367). Piscataway, New Jersey.

Morgan, L. E., Rhodes-Leader, L., & Barton, R. R. (2022). Reducing and calibrating for input model bias in computer simulation. *INFORMS Journal on Computing*, *0(0)*.

Morgan, L. E., Titman, A. C., Worthington, D. J., & Nelson, B. L. (2016). Input Uncertainty Quantification for Simulation Models with Piecewise-constant Non-stationary Poisson Arrival Processes. In T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, & S. E. Chick (Eds.), *Proceedings of the 2016 Winter Simulation Conference* (pp. 370–381). Piscataway, New Jersey.

Morgan, L. E., Titman, A. C., Worthington, D. J., & Nelson, B. L. (2020). *NHPP spline.* (https://github.com/morganle/NHPPspline)

Nelson, B. L., & Gerhardt, I. (2011). Modelling and Simulating Non-stationary Arrival Processes to Facilitate Analysis. *Journal of Simulation*, *5*(1), 3–8.

Nelson, B. L., & Leemis, L. M. (2020). The ease of fitting but futility of testing a nonstationary poisson processes from one sample path. In K. H. Bae et al. (Eds.), *Proceedings of the 2020 Winter Simulation Conference* (pp. 266–276). Piscatway New Jersey.

Nelson, B. L., & Pei, L. (2021). *Foundations and methods of stochastic simulation* (Tech. Rep.). Springer.

Nicol, D. M., & Leemis, L. M. (2014). A Continuous Piecewise-Linear NHPP Intensity Function Estimator. In A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, & J. A. Miller (Eds.), *Proceedings of the 2014 Winter Simulation Conference* (pp. 498–509). Piscataway, New Jersey.

Pritsker, A., Martin, D., Reust, J., Wagner, M., Wilson, J., Kuhl, M., ... Allen, M. (1996). Organ Transplantation Modeling and Analysis. In A. J.G. & M. Katzper (Eds.), *Proceedings of the 1996 Western Multiconference: Simulation in the Medical Sciences* (pp. 29–35).

Reinsch, C. H. (1967). Smoothing by spline functions. *Numerische mathematik*, *10*(3), 177–183.

Shibata, R. (1989). Statistical Aspects of Model Selection. In *From Data to Model* (pp. 215–240). Berlin, Heidelberg: Springer.

Song, E., Nelson, B. L., & Pegden, C. D. (2014). Advanced Tutorial: Input Uncertainty Quantification. In A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, & J. A. Miller (Eds.), *Proceedings of the 2014 Winter Simulation Conference* (pp. 162–176). Piscataway, New Jersey.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *The Annals of Statistics*, 590–606.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95–114.

Wright, S., & Nocedal, J. (1999). *Numerical Optimization* (Vol. 35). Springer.

Xue, L., & Liang, H. (2010). Polynomial spline estimation for a generalized additive coefficient model. *Scandinavian Journal of Statistics*, *37*(1), 26–46.

Zheng, Z., & Glynn, P. W. (2017). Fitting Continuous Piecewise Linear Poisson Intensities via Maximum Likelihood and Least Squares. In W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, & E. Page (Eds.), *Proceedings of the 2017 Winter Simulation Conference* (pp. 1740–1749). Piscataway, New Jersey.