

Gaussian Process-Based Bayesian Nonparametric Inference of Population Size Trajectories from Gene Genealogies

Julia A. Palacios and Vladimir N. Minin*

Department of Statistics, University of Washington, Seattle, Washington 98195-4322, U.S.A.

*email: vminin@uw.edu

SUMMARY. Changes in population size influence genetic diversity of the population and, as a result, leave a signature of these changes in individual genomes in the population. We are interested in the inverse problem of reconstructing past population dynamics from genomic data. We start with a standard framework based on the coalescent, a stochastic process that generates genealogies connecting randomly sampled individuals from the population of interest. These genealogies serve as a glue between the population demographic history and genomic sequences. It turns out that only the times of genealogical lineage coalescences contain information about population size dynamics. Viewing these coalescent times as a point process, estimating population size trajectories is equivalent to estimating a conditional intensity of this point process. Therefore, our inverse problem is similar to estimating an inhomogeneous Poisson process intensity function. We demonstrate how recent advances in Gaussian process-based nonparametric inference for Poisson processes can be extended to Bayesian nonparametric estimation of population size dynamics under the coalescent. We compare our Gaussian process (GP) approach to one of the state-of-the-art Gaussian Markov random field (GMRF) methods for estimating population trajectories. Using simulated data, we demonstrate that our method has better accuracy and precision. Next, we analyze two genealogies reconstructed from real sequences of hepatitis C and human Influenza A viruses. In both cases, we recover more believed aspects of the viral demographic histories than the GMRF approach. We also find that our GP method produces more reasonable uncertainty estimates than the GMRF method.

KEY WORDS: Coalescent; Data augmentation; Doubly intractable distribution; Point process; Thinning.

1. Introduction

Statistical inference in population genetics increasingly relies on the coalescent (Kingman, 1982), the probability model that describes the relationship between a gene genealogy of a random sample of molecular sequences and effective population size. This model provides a good approximation to the distribution of ancestral histories that arise from classical population genetics models (Rosenberg and Nordborg, 2002). More importantly, coalescent-based inference methods allow us to estimate population genetic parameters, including population size trajectories, directly from genomic sequences (Griffiths and Tavaré, 1994). Recent examples of coalescent-based population dynamics estimation include reconstructing demographic histories of musk ox (Campos et al., 2010) from fossil DNA samples and elucidating patterns of genetic diversity of the dengue virus (Bennett et al., 2010).

Here, we are interested in estimating effective population size trajectories from gene genealogies. The *effective population size* is an abstract parameter that for a real biological population is proportional to the rate at which genetic diversity is lost or gained. In the absence of natural selection, the effective population size can be used to approximate census population size by knowing the generation time in calendar units (e.g., years) and the population variability in number of offspring (Wakeley and Sargsyan, 2009). The latter quan-

tity might be difficult to know; however, sometimes it suffices to analyze an arbitrarily rescaled population size trajectory, assuming the variability in number of offspring remains constant. The effective population size is equal to the census population size in an idealized Wright–Fisher model. The Wright–Fisher model is a simple and established model of neutral reproduction in population genetics that assumes random mating and nonoverlapping generations. For some RNA viruses, for example human influenza A virus, the effective population size rescaled by generation time (3–4 days) cannot be interpreted directly as the effective number of infections because of the presence of strong natural selection. However, one can always adopt a more cautious interpretation of the effective population size as a measure of relative genetic diversity (Rambaut et al., 2008; Frost and Volz, 2010).

Coalescent-based methods for estimation of population size dynamics have evolved from stringent parametric assumptions, such as constant population size or exponential growth (Griffiths and Tavaré, 1994; Kuhner, Yamato, and Felsenstein, 1998; Drummond et al., 2002), to more flexible nonparametric approaches that assume piecewise linear population trajectories (Strimmer and Pybus, 2001; Drummond et al., 2005; Opgen-Rhein, Fahrmeir, and Strimmer, 2005; Heled and Drummond, 2008; Minin, Bloomquist, and Suchard, 2008). The latter class of methods is more appropriate in the

absence of prior knowledge about the underlying demographic dynamics, allowing researchers to infer shapes of population size trajectories rather than to impose parametric constraints on these shapes. These nonparametric methods, however, model population dynamics by imposing a priori piecewise continuous functions which require regularization either by smoothing or by controlling the number of change points, also a priori. The former regularization—which works better in practice (Minin et al., 2008)—is inherently difficult because these piecewise continuous functions are defined on intervals of varying size. The piecewise nature of these methods creates further modeling problems if one wishes to incorporate covariates into the model or impose constraints on population size dynamics (Minin et al., 2008). In this article, we propose to solve these problems by bringing the coalescent-based estimation of population dynamics up to speed with modern Bayesian nonparametric methods. Making this leap in statistical methodology will allow us to avoid artificial discretization of population trajectories, to perform regularization without making arbitrary scale choices, and, in the future, to extend our method into a multivariate setting.

Our key insight stems from the fact that the coalescent with variable population size is an inhomogeneous continuous-time Markov chain (Tavaré, 2004) and, therefore, can be viewed as an inhomogeneous point process (Andersen et al., 1995). In fact, all current Bayesian nonparametric methods of estimation of population size dynamics resemble early Bayesian approaches to nonparametric estimation of the Poisson intensity function via piecewise continuous functions (Arjas and Heikkinen, 1997). Estimation of the intensity function of an inhomogeneous Poisson process is a mature field that evolved from maximum likelihood estimation under parametric assumptions (Brillinger, 1979) to frequentist (Diggle, 1985) and, more recently, Bayesian nonparametric methods (Arjas and Heikkinen, 1997; Møller, Syversveen, and Waagepetersen 1998; Kottas and Sansó, 2007; Adams, Murray, and MacKay, 2009).

Following Adams et al. (2009), we a priori assume that population trajectories follow a transformed Gaussian process (GP), allowing us to model the population trajectory as a continuous function. This is a convenient way to specify prior beliefs without a particular functional form on the population trajectory. The drawback of such a flexible prior is that the likelihood function involves integration over an infinite-dimensional random object and, as a result, likelihood evaluation becomes intractable. Fortunately, we are able to avoid this intractability and perform inference exactly by adopting recent algorithmic developments proposed by Adams et al. (2009). We achieve tractability by a novel data augmentation for the coalescent process that relies on thinning algorithms for simulating the coalescent.

Thinning is an accept/reject algorithm that was first proposed by Lewis and Shedler (1979) for the simulation of inhomogeneous Poisson processes and was later extended to a more general class of point processes by Ogata (1981). In the spirit of Ogata (1981), we develop novel thinning algorithms for the simulation of the coalescent. These algorithms, interesting in their own right, open the door for latent variable representation of the coalescent. This representation leads to a new data augmentation that is computationally tractable and

amenable to standard Markov chain Monte Carlo (MCMC) sampling from the posterior distribution of model parameters and latent variables.

We test our method on simulated data and compare its performance with a representative piecewise linear approach, a Gaussian Markov random field (GMRF)-based method (Minin et al., 2008). We demonstrate that our method is more accurate and more precise than the GMRF method in all simulation scenarios. We also apply our method to two real data sets that have been previously analyzed in the literature: a hepatitis C virus (HCV) genealogy estimated from sequences sampled in 1993 in Egypt and a genealogy of the H3N2 human influenza A virus estimated from sequences sampled in New York state between 2002 and 2005. In the HCV analysis, we successfully recover all believed key aspects of the population size trajectory. Compared to the GMRF method, our GP method better reflects the uncertainty inherent in the HCV data. In our second real data example, our GP method successfully reconstructs a population trajectory of the human influenza A virus with an expected seasonal series of peaks followed by population bottlenecks, although the GMRF method’s reconstructed trajectory fails to recover some of the peaks and bottlenecks.

2. Methods

2.1 Coalescent Background

The coalescent model allows us to trace the ancestry of a random sample of n genomic sequences. These ancestral relationships are represented by a genealogy or tree; the times at which two sequences or lineages merge into a common ancestor are called coalescent times. The coalescent with variable population size can be viewed as a nonhomogeneous Markov death process that starts with n lineages at present time $t_n = 0$ and decreases by one, with time running backwards, until reaching one lineage at t_1 , at which point the samples have been traced to their most recent common ancestor (Griffiths and Tavaré, 1994).

Here, we assume that a genealogy with time measured in units of generations is observed. The shape of the genealogy depends on the effective population size trajectory, $N_e(t)$, and the number of samples accumulated through time: the larger the effective population size, the longer two lineages need to wait to meet a common ancestor and the larger the sample size, the faster two lineages coalesce. Formally, let $t_n = 0$ denote the present time when all the n available sequences are sampled (*isochronous coalescent*) and let $t_n = 0 < t_{n-1} < \dots < t_1$ denote the coalescent times of lineages in the genealogy with time going backwards. Then, the conditional density of the coalescent time t_{k-1} takes the following form:

$$P[t_{k-1}|t_k, N_e(t)] = \frac{C_k}{N_e(t_{k-1})} \exp \left\{ - \int_{t_k}^{t_{k-1}} \frac{C_k}{N_e(t)} dt \right\}, \quad (1)$$

where $C_k = \binom{k}{2}$ is the coalescent factor that depends on the number of lineages $k = 2, \dots, n$.

The *heterochronous coalescent* arises when samples of sequences are collected at different times (Figure 1). Such serially sampled data are common in studies of rapidly evolving viruses and analyses of ancient DNA (Campos et al., 2010).

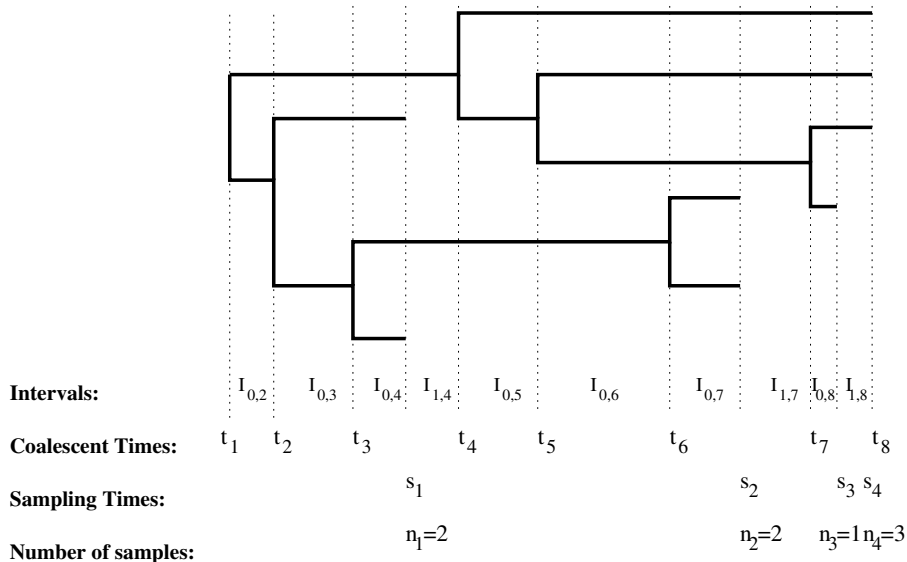


Figure 1. Example of a genealogy relating serially sampled sequences (heterochronous sampling). The number of lineages changes every time we move between intervals ($I_{i,k}$). Each endpoint of an interval is a coalescent time ($\{t_k\}_{k=1}^n$) or a sampling time ($\{s_j\}_{j=1}^m$). The number of sequences sampled at time s_j is denoted by n_j .

Let $t_n = 0 < t_{n-1} < \dots < t_1$ denote the coalescent times as before, but now let $s_m = 0 < s_{m-1} < \dots < s_1 < s_0 = t_1$ denote sampling times of n_m, \dots, n_1 sequences, respectively, $\sum_{j=1}^m n_j = n$. Further, let \mathbf{s} and \mathbf{n} denote the vectors of sampling times and numbers of sequences sampled at these times, respectively (Figure 1). Now, the coalescent factor changes not only at the coalescent events but also at the sampling times. Let

$$I_{0,k} = (\max\{t_k, s_j\}, t_{k-1}], \text{ for } s_j < t_{k-1} \text{ and } k = 2, \dots, n, \quad (2)$$

be the intervals that end with a coalescent event and

$$I_{i,k} = (\max\{t_k, s_{j+i}\}, s_{j+i-1}], \text{ for } s_{j+i-1} > t_k \text{ and } s_j < t_{k-1}, k = 2, \dots, n, \quad (3)$$

be the intervals that end with a sampling event. We denote the number of lineages in $I_{i,k}$ with $n_{i,k}$. Then, for $k = 2, \dots, n$,

$$P[t_{k-1}|t_k, \mathbf{s}, \mathbf{n}, N_e(t)] = \frac{C_{0,k}}{N_e(t_{k-1})} - \exp \left\{ \int_{I_{0,k}} \frac{C_{0,k}}{N_e(t)} dt + \sum_{i=1}^m \int_{I_{i,k}} \frac{C_{i,k}}{N_e(t)} dt \right\}, \quad (4)$$

where $C_{i,k} = \binom{n_{i,k}}{2}$. That is, the density for the next coalescent time t_{k-1} is the product of the density of the coalescent time $t_{k-1} \in I_{0,k}$ and the probability of not having a coalescent event during the period of time spanned by intervals $I_{1,k}, \dots, I_{m,k}$ (Felsenstein and Rodrigo, 1999).

2.2 GP Prior for Population Size Trajectories

For both isochronous or heterochronous data, we place the same prior on $N_e(t)$:

$$N_e(t) = \left[\frac{\lambda}{1 + \exp\{-f(t)\}} \right]^{-1}, \quad (5)$$

where

$$f(t) \sim \mathcal{GP}(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta})) \quad (6)$$

and $\mathcal{GP}(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta}))$ denotes a GP with mean function $\mathbf{0}$ and covariance function $\mathbf{C}(\boldsymbol{\theta})$ with hyperparameters $\boldsymbol{\theta}$. A priori, $1/N_e(t)$ is a sigmoidal GP, a scaled logistic function of a GP whose range is restricted to lie in $[0, \lambda]$; λ is a positive constant hyperparameter, the inverse of which serves as a lower bound of $N_e(t)$ (Adams et al., 2009).

A GP is a stochastic process such that any finite sample from the process has a joint Gaussian distribution. The process is completely specified by its mean and covariance functions (Rasmussen and Williams, 2005). For computational convenience we use Brownian motion as our GP prior. Generating a finite sample from a GPs requires $\mathcal{O}(n^3)$ computations because of the inversion of the covariance matrix. However, when the precision matrix, the inverse of the covariance, is sparse, such simulations can be accomplished much faster (Rue and Held, 2005). For example, when we choose to work with a Brownian motion with covariance matrix elements $C(t, t') = \frac{1}{\theta}(\min(t, t'))$ and precision parameter θ , then the inverse of this matrix is tri-diagonal, which reduces the computational complexity of simulations from $\mathcal{O}(n^3)$ to $\mathcal{O}(n)$. In our MCMC algorithm, we need to generate realizations from the GPs at thousands of points, so the speed-up afforded by the Brownian motion becomes almost a necessity, prompting us to use this process as our prior in all our examples.

2.3 Priors for Hyperparameters

The precision parameter θ controls the degree of autocorrelation of our Brownian motion prior and influences the “smoothness” of the reconstructed population size trajectories. We place on θ a Gamma prior distribution with parameters α and β . The other hyperparameter in our model is the upper bound of $1/N_e(t)$, λ . When this upper bound λ is unknown,

the model is unidentifiable (see equation (5)). However, in many circumstances it is possible to obtain an upper bound λ (or equivalently, a lower bound on $N(t)$) from previous studies and use this value to define the prior distribution of λ . We use the following strategy to construct an informative prior for λ . Let $\hat{\lambda}$ denote our best guess of the upper bound, possibly obtained from previous studies. Then, the prior on λ is a mixture of a uniform distribution for values to the left of $\hat{\lambda}$ and an exponential distribution to the right:

$$P(\lambda) = \epsilon \frac{1}{\hat{\lambda}} I_{\{\lambda < \hat{\lambda}\}} + (1 - \epsilon) \frac{1}{\hat{\lambda}} e^{-\frac{1}{\hat{\lambda}}(\lambda - \hat{\lambda})} I_{\{\lambda \geq \hat{\lambda}\}}, \quad (7)$$

where $\epsilon > 0$ is a mixing proportion. When $\hat{\lambda}$ is considerably smaller than the unknown λ , the recovered curve will be visibly truncated around $\hat{\lambda}$, indicating that one needs to try higher values of λ .

2.4 Doubly Intractable Posterior

Coalescent times $\mathcal{T} = \{t_n, t_{n-1}, \dots, t_1\}$ of a given genealogy contain information needed to estimate $N_e(t)$ (see equations (1) and (4)). Given that $N_e(t)$ is a one-to-one function of $f(t)$ (equation (5)), we will focus the discussion on the inference of $f(t)$. The posterior distribution of $f(t)$ and hyperparameters θ and λ becomes

$$P(f(t), \theta, \lambda | \mathcal{T}) \propto P(\mathcal{T} | \lambda, f(t)) P(f(t) | \theta) P(\theta) P(\lambda), \quad (8)$$

where $P(f(t) | \theta)$ is a GP prior with hyperparameter θ and

$$P(\mathcal{T} | \lambda, f(t)) = \prod_{k=2}^n \frac{C_k \lambda}{1 + \exp\{-f(t_{k-1})\}} \times \exp \left[-C_k \int_{t_k}^{t_{k-1}} \frac{\lambda}{1 + \exp\{-f(t)\}} dt \right] \quad (9)$$

is the likelihood function for the isochronous data (heterochronous data likelihood has a similar form). The integral in the exponent of equation (9) and the normalizing constant of equation (8) are computationally intractable, making the posterior doubly intractable (Murray, Ghahramani, and MacKay 2006).

Adams et al. (2009) faced a similar doubly intractable posterior distribution in the context of nonparametric estimation of intensity of the inhomogeneous Poisson process. These authors propose an introduction of latent variables so that the augmented data likelihood becomes tractable. This tractability makes the posterior distribution of latent variables and model parameters amenable to standard MCMC algorithms. Because Adams et al. (2009) based their data augmentation on the thinning algorithm for simulating inhomogeneous Poisson processes, we would like to devise a similar data augmentation based on a thinning algorithm for simulation of the coalescent with variable population size. In this simulation, we envision generating coalescent times assuming a constant population size and then thinning these times so that the distribution of the remaining (nonrejected) coalescent times follows the coalescent with variable population size. As no thinning algorithm for simulating the coalescent process exists, we develop a series of such algorithms. In developing these algorithms, we find it useful to view the coalescent as a point process, a representation that we discuss below.

2.5 The Coalescent as a Point Process

The joint density of coalescent times is obtained by multiplying the conditional densities defined in equations (1) or (4). This density can be expressed as

$$P(t_1, \dots, t_{n-1} | N_e(t)) = \prod_{k=2}^n \lambda^*(t_{k-1} | t_k) \exp \left\{ - \int_{t_k}^{t_{k-1}} \lambda^*(t | t_k) dt \right\}, \quad (10)$$

where $\lambda^*(t | t_k)$ denotes the conditional intensity function of a point process on the real line (Daley and Vere-Jones, 2002). For isochronous coalescent, the conditional intensity is defined by the step function:

$$\lambda^*(t | t_k) = \binom{k}{2} N_e(t)^{-1} \mathbf{1}_{\{t \in (t_k, t_{k-1}]\}}, \quad \text{for } k = 2, \dots, n, \quad (11)$$

and the conditional intensity of the heterochronous coalescent point process is:

$$\lambda^*(t | \mathbf{n}, \mathbf{s}, t_k) = \sum_{i=1}^m \binom{n_{i,k}}{2} N_e(t)^{-1} \mathbf{1}_{\{t \in I_{i,k}\}}, \quad \text{for } k = 2, \dots, n. \quad (12)$$

This novel representation allows us to reduce the task of estimating $N_e(t)$ to the estimation of the inhomogeneous intensity of the coalescent point process and to develop simulation algorithms based on thinning.

2.6 Coalescent Simulation via Thinning

To the best of our knowledge, the only method available for simulating the coalescent under the deterministic variable population size model is a time transformation method (Slatkin and Hudson, 1991; Hein, Schierup, and Wiuf 2005). This method is based on the random time-change theorem due to Papangelou (1972). Under the time transformation method, to simulate coalescent times, we proceed sequentially starting with $k = n$ and $t_n = 0$, generating t from an exponential distribution with unit mean, solving

$$t = \int_{t_k}^{t_{k-1}} \lambda^*(u | t_k) du \quad (13)$$

for t_{k-1} analytically or numerically and repeating the procedure until $k = 2$. For isochronous coalescent, $\lambda^*(u | t_k)$ is defined in equation (11) and for the heterochronous coalescent, $\lambda^*(u | t_k) = \lambda^*(u | \mathbf{n}, \mathbf{s}, t_k)$ is the piecewise function defined in equation (12). When $N_e(t)$ is stochastic, the integral in equation (13) becomes intractable and the time transformation method is no longer practical. Instead, we propose to use *thinning*, a rejection-based method that does not require calculation of the integral in equation (13).

Lewis and Shedler (1979) proposed thinning a homogeneous Poisson process for the simulation of an inhomogeneous Poisson process with intensity $\lambda(t)$. The idea is to start with a realization of points from a homogeneous Poisson process with intensity λ and accept/reject each point with acceptance probability $\lambda(t)/\lambda$, where $\lambda(t) \leq \lambda$. The collection of accepted points forms a realization of the inhomogeneous Poisson process with conditional intensity $\lambda(t)$. Ogata (1981) extended Lewis and Shedler's thinning for the simulation of any point process that is absolutely continuous with respect

to the standard Poisson process. We develop a series of thinning algorithms for the coalescent process that are similar to Ogata's algorithms, but not identical to them. Algorithm 1 outlines the simulation of n coalescent times under the isochronous sampling. Given t_k , we start generating and accumulating exponential random numbers E_i with rate $C_k \lambda$, until $t_{k-1} = t_k + E_1 + E_2 + \dots$ is accepted with probability $1/N_e(t_{k-1})\lambda$. (See Web Appendix A for details and simulation algorithms of coalescent times for heterochronous sampling.) To ensure convergence of the algorithm, we require $\int_0^\infty \frac{du}{N_e(u)} = \infty$ a.s., which is equivalent to requiring that all sampled lineages can be traced back to their single common ancestor with probability 1. Note that $N_e(t)$ can be either deterministic or stochastic. The latter case is considered in Web Supplementary Algorithm 2, where we work with $f(t)$ instead of $N_e(t)$ for notational convenience.

If $N_e(t)$ is deterministic and equation (13) can be solved analytically, the time transformation method is likely to be more efficient than thinning because the thinning algorithm is an accept/reject algorithm with the acceptance probability highly dependent on the definition of λ . However, efficiency of the thinning algorithm can be improved by replacing the constant upper bound λ on $1/N_e(t)$, by a piecewise constant or a piecewise linear function of local upper bounds to achieve higher acceptance probabilities, similarly to the adaptive rejection sampling of Gilks and Wild (1992).

2.7 Data Augmentation and Inference

As mentioned in the previous section, our thinning algorithm for the coalescent is motivated by our desire to construct a data augmentation scheme. We imagine that observed coalescent times \mathcal{T} were generated by the thinning procedure described in Algorithm 1, so we augment \mathcal{T} with rejected (thinned) points \mathcal{N} . If we keep track of the rejected points resulting from Algorithm 1, then, given t_k , $f(t_k)$, $\mathbf{f}_{\mathcal{N}_k} = \{f(t_{k,i})\}_{i=1}^{m_k}$ and λ , we start proposing new time points $\mathcal{N}_k = \{t_{k,1}, \dots, t_{k,m_k}\}$ until t_{k-1} is accepted, so that

Algorithm 1

Simulation of isochronous coalescent times by thinning - $N_e(t)$ is a deterministic function.

Input: $k = n$, $t_n = 0$, $t = 0$, $1/N_e(t) \leq \lambda$, $N_e(t)$

Output: $\mathcal{T} = \{t_k\}_{k=1}^n$

- 1: **while** $k > 1$ **do**
- 2: Sample $E \sim \text{Exponential}(C_k \lambda)$ and $U \sim U(0, 1)$
- 3: $t = t + E$
- 4: **if** $U \leq \frac{1}{N_e(t)\lambda}$ **then**
- 5: $k \leftarrow k - 1$, $t_k \leftarrow t$
- 6: **end if**
- 7: **end while**

$$P(t_{k-1}, \mathcal{N}_k | t_k, f(t_{k-1}), \mathbf{f}_{\mathcal{N}_k}, \lambda) = (C_k \lambda)^{m_k+1} \exp\{-C_k \lambda(t_k - t_{k-1})\} \left[\frac{1}{1 + \exp\{-f(t_{k-1})\}} \right] \times \prod_{i=1}^{m_k} \left[1 - \frac{1}{1 + \exp\{-f(t_{k,i})\}} \right]. \quad (14)$$

For the heterochronous coalescent (see Algorithm 3 in Web Supplementary Materials), equation (14) is modified in the following way:

$$P(t_{k-1}, \mathcal{N}_k | t_k, f(t_{k-1}), \mathbf{f}_{\mathcal{N}_k}, \lambda, \mathbf{s}, \mathbf{n}) = (\lambda C_{0,k})^{1+m_{0,k}} \exp\{-\lambda C_{0,k} l(I_{0,k})\} \times \left[\frac{1}{1 + \exp\{-f(t_{k-1})\}} \right] \times \prod_{i=1}^{m_k} \frac{1}{1 + \exp\{f(t_{k,i})\}} \times \prod_{i=1}^m [(\lambda C_{i,k})^{m_{i,k}} \exp\{-\lambda C_{i,k} l(I_{i,k})\}], \quad (15)$$

where $l(I_{i,k})$ denotes the length of the interval $I_{i,k}$ and $m_{i,k} = \sum_{l=1}^{m_k} 1\{t_{k,l} \in I_{i,k}\}$ denotes the number of latent points in interval $I_{i,k}$. Let $\mathbf{f}_{\mathcal{T}, \mathcal{N}} = \{\{f(t_k)\}_{k=1}^n, \{\{f(t_{k,i})\}_{i=1}^{m_k}\}_{k=2}^n\}$, then the augmented data likelihood of \mathcal{T} and \mathcal{N} becomes

$$P(\mathcal{T}, \mathcal{N} | \mathbf{f}_{\mathcal{T}, \mathcal{N}}, \lambda) = \prod_{k=2}^n P(t_{k-1}, \mathcal{N}_k | t_k, f(t_{k-1}), \mathbf{f}_{\mathcal{N}_k}, \lambda). \quad (16)$$

Then, the posterior distribution of $f(t)$ and hyperparameters evaluated at the observed \mathcal{T} and latent \mathcal{N} time points is

$$P(\mathbf{f}_{\mathcal{T}, \mathcal{N}}, \lambda, \theta | \mathcal{T}, \mathcal{N}) \propto P(\mathcal{T}, \mathcal{N} | \mathbf{f}_{\mathcal{T}, \mathcal{N}}, \lambda) P(\mathbf{f}_{\mathcal{T}, \mathcal{N}} | \theta) P(\lambda) P(\theta). \quad (17)$$

The augmented posterior can now be easily evaluated because it does not involve integration of infinite-dimensional random functions. We follow Adams et al. (2009) and develop a MCMC algorithm to sample from the posterior distribution (16). At each iteration of our MCMC, we update the following variables: (1) number of "rejected" points $\#\mathcal{N}$; (2) the locations of the rejected points \mathcal{N} ; (3) the function values $\mathbf{f}_{\mathcal{T}, \mathcal{N}}$, and (4) the hyperparameters θ and λ . We use a Metropolis-Hastings algorithm to sample the number and locations of latent points and the hyperparameter λ ; we Gibbs sample the hyperparameter θ . Updating the function values $\mathbf{f}_{\mathcal{T}, \mathcal{N}}$ is non-trivial, because this high-dimensional vector has correlated components. In such cases, single-site updating is inefficient and block updating is preferred (Rue and Held 2005). We use elliptical slice sampling, proposed by Murray, Adams, and MacKay (2010), to sample $\mathbf{f}_{\mathcal{T}, \mathcal{N}}$. The advantage of using the elliptical slice sampling proposal is that it does not require the specification of tuning parameters and works well in high dimensions. The details of our MCMC algorithm can be found in Web Appendix B.

We summarize the posterior distribution of $N_e(t)$ by its empirical median and 95% Bayesian credible intervals (BCIs) evaluated at a grid of points. This grid can be made as fine as necessary after the MCMC is finished. That is, given the function values $\mathbf{f}_{\mathcal{T}, \mathcal{N}}$ at coalescent and latent time points, and the value of the precision parameter θ at each iteration, we sample the function values at a grid of points $\mathbf{g} = \{g_1, \dots, g_B\}$ from its predictive distribution $\mathbf{f}_{\mathbf{g}} \sim P(\mathbf{f}_{\mathbf{g}} | \mathbf{f}_{\mathcal{T}, \mathcal{N}}, \theta)$, and evaluate $\{N_e(g_i)\}_{i=1}^B$.

3. Results

3.1 Simulated Data

We simulate three genealogies relating 100 individuals, sampled at the same time $t = 0$ (isochronous sampling) under the following demographic scenarios: (1) constant population size trajectory: $N_e(t) = 1$; (2) exponential growth: $N_e(t) = 25e^{-5t}$; and (3) population expansion followed by a crash: $N_e(t) = e^{4t} \mathbf{1}_{\{t \in [0, 0.5]\}} + e^{-2t+3} \mathbf{1}_{\{t \in (0.5, \infty)\}}$. We compare the posterior median with the truth by the sum of relative errors (SRE):

$$SRE = \sum_{i=1}^K \frac{|\hat{N}_e(s_i) - N_e(s_i)|}{N_e(s_i)}, \quad (18)$$

where $\hat{N}_e(s_i)$ is the estimated trajectory at time s_i with $s_1 = t_1$, the time to the most recent common ancestor and $s_K = t_n = 0$ for any finite K . Similarly, we compute the mean relative width (MRW) of the 95% BCIs defined in the following way:

$$MRW = \sum_{i=1}^K \frac{|\hat{N}_{97.5}(s_i) - \hat{N}_{2.5}(s_i)|}{KN_e(s_i)}. \quad (19)$$

We also compute the percentage of time, the 95% BCIs cover the truth (envelope) in the following way:

$$envelope = \frac{\sum_{i=1}^K I(\hat{N}_{2.5}(s_i) \leq N(s_i) \leq \hat{N}_{97.5}(s_i))}{K}. \quad (20)$$

As a measure of the frequentist coverage, we calculate the percentage of times the truth is completely covered by the 95% BCIs (envelope = 1), by simulating each demographic scenario and performing Bayesian estimation of each such simulation 100 times.

We compute the three statistics for the three simulation scenarios for $K = 150$ at equally spaced time points (Table 1). These statistics do not change significantly when we use higher values of K . In addition, we compute the variation of $\hat{N}_e(t)$ over a regular grid of $K = 150$ points as follows:

$$variation = \sum_{i=1}^{K-1} |\hat{N}_e(s_{i+1}) - \hat{N}_e(s_i)|. \quad (21)$$

For all simulations, we set the mixing parameter ϵ of the prior density for λ (equation (7)) to $\epsilon = 0.01$. The parameters of the Gamma prior on the GP precision parameter θ were set to $\alpha = \beta = 0.001$. We summarize our posterior inference in Figure 2 and compare our GP method to the GMRF smoothing method (Minin et al., 2008). The effective population trajectory is log transformed and time is measured in units of generations.

For the constant population scenario (first row in Figure 2), the truth (dashed lines) is almost perfectly recovered by the GP method (solid black line) and the 95% BCIs shown as gray shaded areas are remarkably tight. For the exponential growth simulation (second row), the GMRF method recovers the truth better in the right tail, although our GP method recovers it much better in the left tail. The higher variation of the GP reconstruction in the right tail makes this measure higher than for the GMRF reconstruction. Overall, our GP

method better recovers the truth in the exponential growth scenario, as evidenced by SREs and MRWs in Table 1. The last row in Figure 2 shows the results for a population that experiences expansion followed by a crash in effective population size. In this case, 95% BCIs of the two methods do not completely cover the true trajectory. Although an area near the bottleneck is particularly problematic, the GP method's envelope is much higher (92%) than the envelope produced by the GMRF method (77.3%), the variation recovered by the GP method is closer to the true variation in all simulation scenarios and in general, in terms of the four statistics employed here, the GP method shows better performance. Results for the GMRF method were obtained using the **BEAST** software (Drummond et al., 2012) with running times ranging from 25 to 40 minutes, although results for the GP method were obtained using **R** with running times ranging from 60 to 180 minutes. Although our GP implementation takes longer, we obtain better performance in a still reasonable amount of time.

Next, we simulate each of the three scenarios 100 times and compute the four statistics described before for both methods. The distributions of these statistics are represented by the boxplots depicted in Figure 3. In general, our GP method has smaller SREs, except in the constant case, where the distributions look very similar; smaller MRWs, larger envelopes, and variation closer to the truth. In addition, we calculate the percentage of times, the envelope is 1 as a proxy for frequentist coverage of the 95% BCIs. Because the 95% BCIs are calculated point-wise at 150 equally spaced points, we do not necessarily expect frequentist coverage to be close to 95%. The results are shown as the numbers at the top of the right plot in Figure 3. The coverage levels obtained using the GP method are larger than those obtained using the GMRF method.

3.2 Egyptian HCV

HCV was first identified in 1989. By 1992, when HCV antibody testing became widely available, the prevalence of HCV in Egypt was about 10.8%. Today, Egypt is the country with the highest HCV prevalence (Miller and Abu-Raddad, 2010). A widely held hypothesis that can explain the epidemic emphasizes the role of a parenteral antischistosomal therapy (PAT) campaign, that started in the 1920s, combined with lack of sanitary practices. The campaign was discontinued in the 1970s when the intravenous treatment was gradually replaced by oral administration of the treatment (Ray et al., 2000). Coalescent demographic methods developed over the last 10 years demonstrated evidence in favor of this hypothesis (Pybus et al., 2003; Drummond et al., 2005; Minin et al., 2008). Therefore, this example is well suited for testing our method. We analyze the genealogy estimated by Minin et al. (2008), based on 63 HCV sequences sampled in Egypt in 1993, and compare our method to the GMRF smoothing method (Minin et al., 2008). The results are depicted in Figure 4, with time scaled in units of years. In line with previous results (Pybus et al., 2003; Drummond et al., 2005; Minin et al., 2008), our estimation recovers the exponential growth of the HCV population size starting from the 1920s when the intravenously administered PAT was introduced. Both Pybus et al. (2003) and Minin et al. (2008) hypothesize that the population trajectory remained constant before the start of

Table 1

Summary of simulation results depicted in Figure 2. SRE is the sum of relative errors as defined in equation (19), MRW is the mean relative width of the 95% BCI as defined in equation (20), envelope is calculated as in equation (21), and variation is calculated as in equation (22). The values in bold indicate the best performing method.

Simulations	SRE		MRW		Envelope		Variation		TRUTH
	GMRF	GP	GMRF	GP	GMRF	GP	GMRF	GP	
Constant	50.41	4.15	4.21	0.72	100.0%	100.0%	2.27	0.08	0.00
Exp. growth	47.65	33.60	2.55	2.35	100.0%	100.0%	30.19	52.41	24.80
Expansion/crash	181.88	140.88	10.7	7.26	77.33%	92.0%	5.69	6.94	13.46

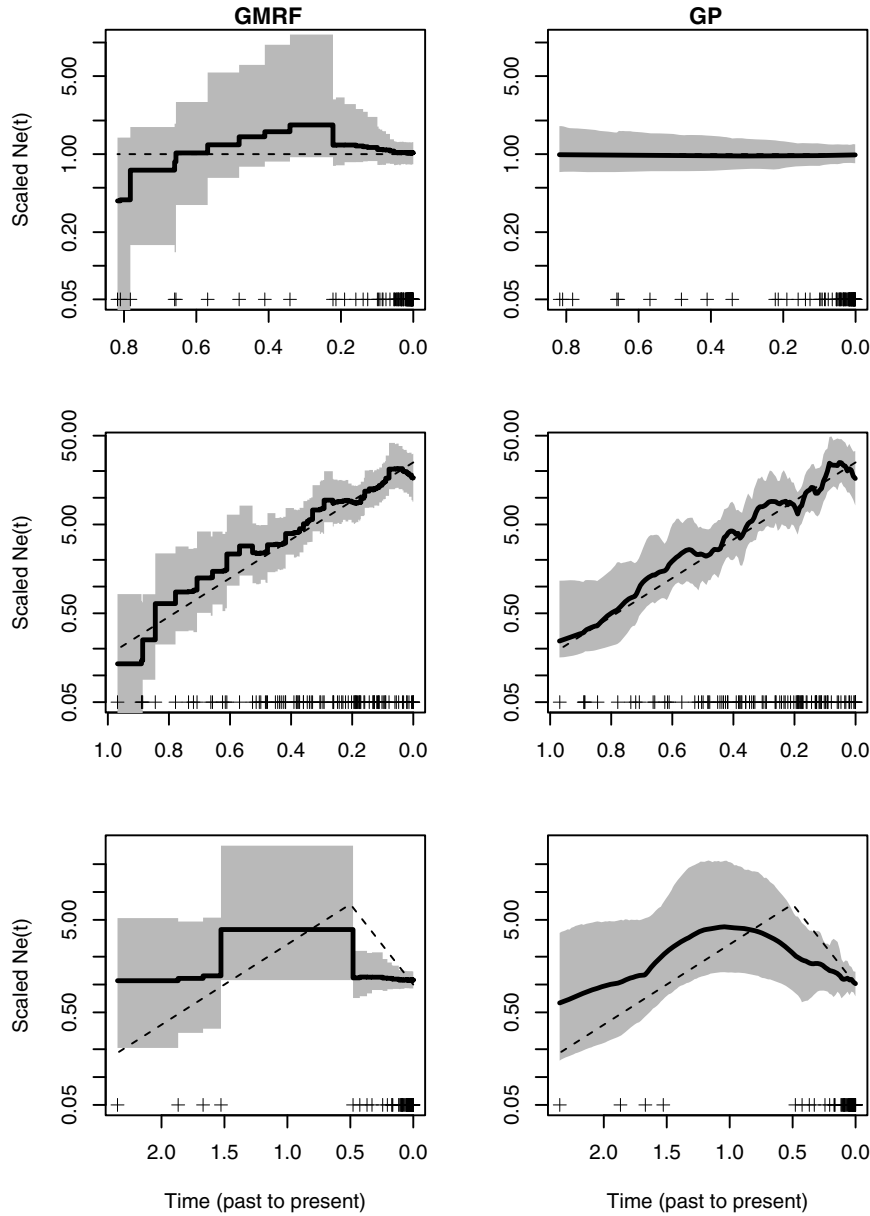


Figure 2. Simulated data under the constant population size (first row), exponential growth (second row), and expansion followed by a crash (third row). The simulated points are represented by the points at the bottom of each plot. We show the log of the effective population size trajectory estimated under the Gaussian Markov random field smoothing (GMRF) method and our method: Gaussian process-based nonparametric inference of effective population size (GP). We show the true trajectories as dashed lines, posterior medians as solid black lines, and 95% BCIs by gray shaded areas.

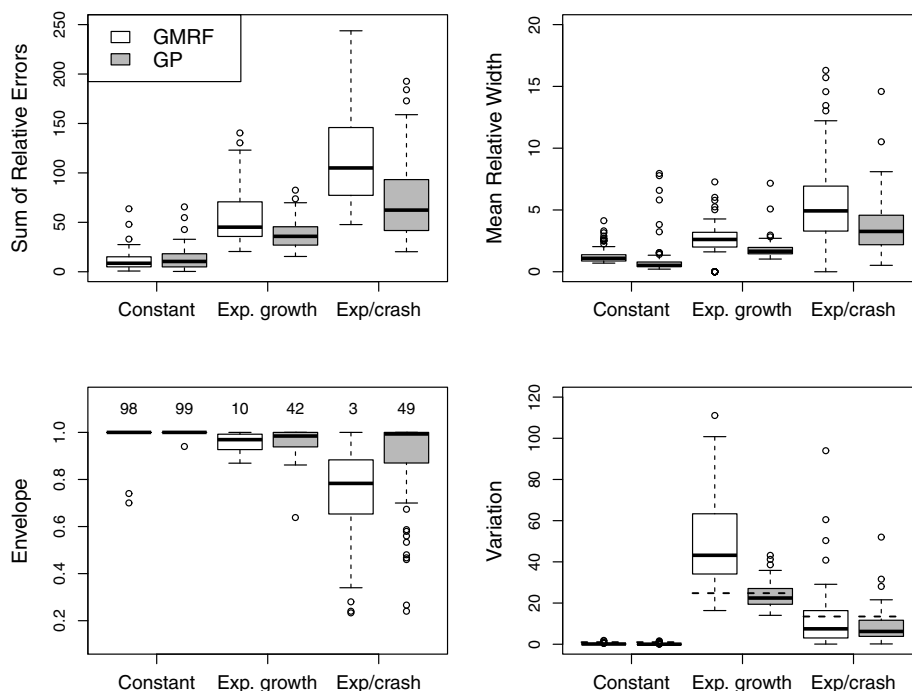


Figure 3. Boxplots of SRE (top left), MRW (top right), envelope (bottom left), and variation (bottom right) based on 100 simulations for a constant trajectory, exponential growth, and expansion followed by crash. The numbers above the boxplots of the bottom-left plot represent the estimated frequentist coverage of the 95% BCIs, and the dashed lines in the bottom-right plot indicate variations of the true simulated trajectories.

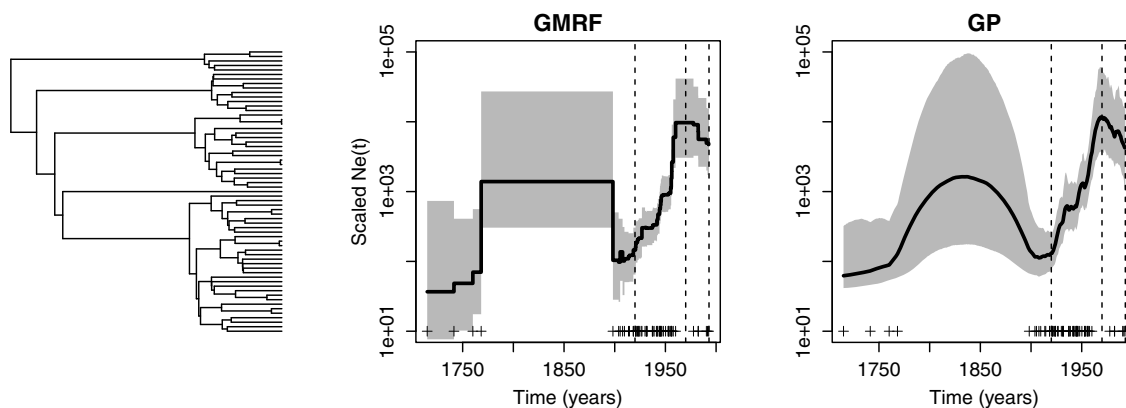


Figure 4. Egyptian HCV. The first plot (left to right) is one possible genealogy reconstructed by Minin et al. (2008). The next two plots represent the log of scaled effective population trajectory estimated using the GMRF smoothing method and our GP method. The posterior medians for the last two plots are represented by solid black lines and the 95% BCIs are represented by the gray shaded areas. The vertical dashed lines mark the years 1920 (the start of intravenous PAT), 1970 (the end of intravenous PAT), and 1993 (sampling time of sequences).

the exponential growth. The GMRF and GP approaches disagree the most on the HCV population size reconstruction before 1920s. The GP method produces narrower BCIs near the root of the genealogy (1710–1770) than the GMRF approach. In contrast, GP BCIs are inflated in the time period from 1770 to 1900 in comparison to the GMRF results. We believe that the uncertainty estimates produced by the GP approach are more reasonable than the GMRF result, be-

cause there are multiple coalescent events during 1710–1770, providing information about the population size, although the time interval 1770–1900 has no coalescent events, a data pattern that should result in substantial uncertainty about the HCV population size. Another notable difference between the GMRF and GP methods is in estimation of the HCV population trajectory after 1970. The GP approach suggests a sharper decline in population size during this time interval.

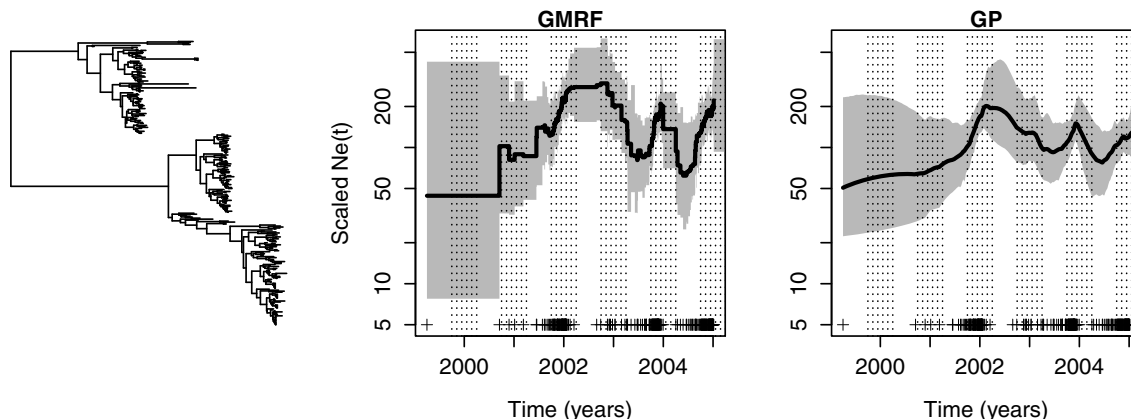


Figure 5. H3N2 Influenza A virus in New York state. The first plot (left) is the estimated genealogy. The second and third plots are the GMRF and GP estimations of log-scaled effective population trajectories. Winter seasons are represented by the dotted shaded areas. Posterior medians are represented by solid black lines and 95% BCIs are represented by gray shaded areas.

3.3 Seasonal Human Influenza

Here, we estimate population dynamics of human influenza A, based on 288 H3N2 sequences sampled in New York state from January 2001 to March 2005. Sequences from the coding region of the influenza hemagglutinin gene of H3N2 influenza A virus from New York state were collected from the National Center for Biotechnology Information (NCBI) Influenza Database (Influenza Genome Sequencing Project, 2011), incorporating the exact dates of viral sampling in weeks (heterochronous sampling) and aligned using the software package MUSCLE (Edgar, 2004). These sequences form a subset of sequences analyzed in Rambaut et al. (2008). We carried out a phylogenetic analysis using the software package BEAST (Drummond et al., 2012) to generate a majority clade support genealogy with median node heights as our genealogical reconstruction. The reconstructed genealogy is depicted in the left plot of Figure 5. Demography of H3N2 influenza A virus in temperate regions, such as New York, is characterized by epidemic peaks during winters followed by strong bottlenecks at the end of epidemic seasons. As expected, our method recovers the peaks in the effective number of infections during all seasons starting from the 2001–2002 flu season (flu seasons are represented as dotted rectangles in Figure 5). The GMRF method fails to recover the peak during the 2002–2003 season. The large uncertainty in population size estimation during the 1999–2000, 2000–2001, and at the beginning of 2005–2006 seasons is explained by the small number of coalescent events during those time periods, however, this uncertainty is larger in the GMRF recovered trajectory. During the 2001–2002 flu season, the GMRF method fails to recover the expected trajectory of a peak followed by a bottleneck and instead, this method recovers an epidemic that started during the end of 2001, increased and remained “at peak” until the end of the following winter. The GMRF recovered trajectory during the winter season of 2003 exhibits a steep decrease. In contrast, the GP method detects a late peak during the 2001–2002 season, followed by a decline in the number of infections. There is a small bump in the effective population size of influenza in the winter of 2003, which

is more realistic than a steady decline in the number of infections estimated by the GMRF method. Overall, we believe that the GP reconstructed trajectory is more feasible from an epidemiological point of view than the GMRF population size reconstruction.

4. Discussion

We propose a new nonparametric method for estimating population size dynamics from gene genealogies. To the best of our knowledge, we are the first to solve this inferential problem using modern tools from Bayesian nonparametrics. In our approach, we assume that the population size trajectory a priori follows a transformed GP. This flexible prior allows us to model population size trajectory as a continuous function without specifying its parametric form and without resorting to artificial discretization methods. We tested our method on simulated and real data and compared it with the competing GMRF method. On simulated data, our method recovers the truth with better accuracy and precision. On real data, where true population trajectories are unknown, our method recovers known epidemiological aspects of the population dynamics and produces more reasonable estimates of uncertainty than the competing GMRF method.

We bring Bayesian nonparametrics into the coalescent framework by viewing the coalescent as a point process. This representation allows us to adapt Bayesian nonparametric methods originally developed for Poisson processes to the coalescent modeling. In particular, it allows us to adapt the thinning-based data augmentation for Poisson processes developed by Adams et al. (2009). We devise an analogous data augmentation for the coalescent by developing a series of new thinning algorithms for the coalescent. Although we use these algorithms in a very narrow context, our novel coalescent simulation protocols should be of interest to a wide range of users of the coalescent. For example, we are not aware of any competitors of our Web Supplementary Algorithms 2 and 4 that allow one to simulate coalescent times with a continuously and *stochastically* varying population size.

Our method works with any GP with mean $\mathbf{0}$ and covariance matrix \mathbf{C} , where the latter controls the level of smoothness and autocorrelation. For computational tractability however, sparsity in the precision matrix (inverse covariance matrix) may be necessary for complex trajectories with a high number of latent points. One way to achieve sparse matrix computations and computational tractability is to use GP that is also Markov. In all our examples, we use Brownian motion with precision parameter θ ; however, the nondifferentiability characteristic of the Brownian motion is compensated by the fact that our estimate of effective population trajectory is the posterior median evaluated point-wise, which is smoother than any of the sampled posterior curves. In addition, we compared Brownian motion, Ornstein–Uhlenbeck and a higher order integrated Brownian motion for one of our examples and obtained very similar results under all three priors (see Web Supplementary Materials). Finite sample distributions under these three priors enjoy sparse precision matrices that yield computational tractability comparable to the GMRF method. In our Brownian motion prior, the precision parameter controls the level of smoothness of the estimated population size trajectory. We find that this important parameter shows little sensitivity to prior perturbations.

Our method assumes that a genealogy or tree is given to the researcher. However, genealogies are themselves inferred from molecular sequences, so we need to incorporate genealogical uncertainty into our estimation. Our framework can be extended to inference from molecular sequences instead of genealogies by introducing another level of hierarchical modeling into our Bayesian framework, similar to the work of Drummond et al. (2005) and Minin et al. (2008). Further, we plan to extend our method to handle molecular sequence data from multiple loci as in (Heled and Drummond, 2008). Finally, we would like to extend our nonparametric estimation into a multivariate setting, so that we can estimate cross-correlations between population size trajectories and external time series. Estimating such correlations is a critical problem in molecular epidemiology.

We deliberately adapted the work of Adams et al. (2009) on estimating the intensity function of an inhomogeneous Poisson process, as opposed to alternative ways to attack this estimation problem (Møller et al., 1998; Kottas and Sansó, 2007), to the coalescent. We believe that among the state-of-the-art Bayesian nonparametric methods, our adopted GP-based framework is the most suitable for developing the aforementioned extensions. First, it is straightforward to incorporate external time series data into our method by replacing a univariate GP prior with a multivariate process that evolves the population size trajectory and another variable of interest in a correlated fashion (Teh, Seeger, and Jordan 2005). Second, the fact that our method does not operate on a fixed grid is critical for relaxing the assumption of a fixed genealogy, because fixing the grid a priori is problematic when one starts sampling genealogies, including coalescent times, during MCMC iterations.

Finally, as the coalescent model with varying population size can be viewed as a particular example of an inhomogeneous continuous-time Markov chain, all our mathematical and computational developments are almost directly transferable to this larger class of models. Therefore, our devel-

opments potentially have implications for nonparametric estimation of inhomogeneous continuous-time Markov chains with numerous applications.

5. Supplementary Materials

Web Appendices referenced in Sections 2.6, 2.7, and 4 are available with this paper at the Biometrics website on Wiley Online Library. The R code implementing the thinning algorithm and our MCMC algorithm is available at <http://www.stat.washington.edu/people/jpalacio>.

ACKNOWLEDGEMENTS

The authors thank Peter Guttorp, Joe Felsenstein, and Elizabeth Thompson for helpful discussions. VNM was supported by the NSF grant No. DMS-0856099. JAP acknowledges scholarship from CONACyT Mexico to pursue her doctoral work.

REFERENCES

- Adams, R., Murray, I., and MacKay, D. (2009). Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proceedings of the 26th International Conference on Machine Learning*, 9–16. Montreal, Canada.
- Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1995). *Statistical Models Based on Counting Processes*, 2nd edition. New York, NY: Springer-Verlag.
- Arjas, E. and Heikkinen, J. (1997). An algorithm for nonparametric Bayesian estimation of a Poisson intensity. *Computational Statistics* **12**, 385–402.
- Bennett, S., Drummond, A., Kapan, D., Suchard, M., Muñoz-Jordán, J., Pybus, O., Holmes, E., and Gubler, D. (2010). Epidemic dynamics revealed in dengue evolution. *Molecular Biology and Evolution* **27**, 811–818.
- Brillinger, D. R. (1979). Analyzing point processes subjected to random deletions. *Canadian Journal of Statistics* **7**, 21–27.
- Campos, P. F., Willerslev, E., Sher, A., Orlando, L., Axelsson, E., Tikhonov, A., Aaris Sørensen, K., Greenwood, A. D., Ralf-Dietrich, K., Kosintsev, P., Krakhmalnaya, T., Kuznetsova, T., Lemey, P., MacPhee, R., Norris, C. A., Shepherd, K., and Suchard, M. A. (2010). Ancient DNA analyses exclude humans as the driving force behind late Pleistocene usk ox (*Ovibos moschatus*) population dynamics. *Proceedings of the National Academy of Sciences* **107**, 5675–5680.
- Daley, D. and Vere-Jones, D. (2002). *An Introduction to the Theory of Point Processes*, Volume 1, 2nd edition. New York, NY: Springer-Verlag.
- Diggle, P. (1985). A kernel method for smoothing point process data. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **34**, 138–147.
- Drummond, A. J., Nicholls, G. K., Rodrigo, A. G., and Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**, 1307–1320.
- Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**, 1185–1192.
- Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797.

- Felsenstein, J. and Rodrigo, A. G. (1999). Coalescent approaches to HIV population genetics. In *The Evolution of HIV*, K. A. Crandall (ed), 233–272. Baltimore, MD: Johns Hopkins University Press.
- Frost, S. D. W. and Volz, E. M. (2010). Viral phylogenetics and the search for an ‘effective number of infections’. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **365**, 1879–1890.
- Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **41**, 337–348.
- Griffiths, R. C. and Tavaré, S. (1994). Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **344**, 403–410.
- Hein, J., Schierup, M. H., and Wiuf, C. (2005). *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*, 1st edition. New York, NY: Oxford University Press.
- Heled, J. and Drummond, A. (2008). Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology* **8**, 1–289.
- Influenza Genome Sequencing Project (2011). Available at: <http://www.niaid.nih.gov/labsandresources/resources/dmid/gsc/influenza/Pages/default.aspx>, accessed April, 2011.
- Kingman, J. (1982). The coalescent. *Stochastic Processes and their Applications* **13**, 235–248.
- Kottas, A. and Sansó, B. (2007). Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *Journal of Statistical Planning and Inference* **137**, 3151–3163.
- Kuhner, M. K., Yamato, J., and Felsenstein, J. (1998). Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149**, 429–434.
- Lewis, P. and Shedler, G. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly* **26**, 403–413.
- Miller, F. D. and Abu-Raddad, L. J. (2010). Evidence of intense ongoing endemic transmission of hepatitis C virus in Egypt. *Proceedings of the National Academy of Sciences, USA* **107**, 14757–14762.
- Minin, V. N., Bloomquist, E. W., and Suchard, M. A. (2008). Smooth skyline through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution* **25**, 1459–1471.
- Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25**, 451–482.
- Murray, I., Adams, R. P., and MacKay, D. J. (2010). Elliptical slice sampling. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, Volume 6 of JMLR: W&CP **6**, 541–548. Sardinia, Italy: Chia Laguna Resort.
- Murray, I., Ghahramani, Z., and MacKay, D. (2006). MCMC for doubly-intractable distributions. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*, 359–366. Arlington, VA: AUAI Press.
- Ogata, Y. (1981). On Lewis’ Simulation Method for Point Processes. *IEEE Transactions on Information Theory* **27**, 23–31.
- Oppen-Rhein, R., Fahrmeir, L., and Strimmer, K. (2005). Inference of demographic history from genealogical trees using reversible jump Markov chain Monte Carlo. *BMC Evolutionary Biology* **5**, 1–6.
- Papangelou, F. (1972). Integrability of expected increments of point processes and a related random change of scale. *Transactions of the American Mathematical Society* **165**, 483–506.
- Pybus, O. G., Drummond, A. J., Nakano, T., Robertson, B. H., and Rambaut, A. (2003). The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: A Bayesian coalescent approach. *Molecular Biology and Evolution* **20**, 381–387.
- Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K., and Holmes, E. C. (2008). The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453**, 615–619.
- Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press.
- Ray, S. C., Arthur, R. R., Carella, A., Bukh, J., and Thomas, D. L. (2000). Genetic epidemiology of hepatitis C virus throughout Egypt. *The Journal of Infectious Diseases* **182**, 698–707.
- Rosenberg, N. A. and Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* **3**, 380–390.
- Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, FL: Chapman and Hall.
- Slatkin, M. and Hudson, R. R. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555–562.
- Strimmer, K. and Pybus, O. G. (2001). Exploring the demographic history of DNA sequences using the generalized skyline plot. *Molecular Biology and Evolution* **18**, 2298–2305.
- Tavaré, S. (2004). Part I: Ancestral inference in population genetics. In *Lectures on Probability Theory and Statistics*, volume 1837 of *Lecture Notes in Mathematics*, J. Picard (ed), 1–188. New York, NY: Springer Verlag.
- Teh, Y. W., Seeger, M., and Jordan, M. I. (2005). Semiparametric latent factor models. In *Proceedings of the International Workshop on Artificial Intelligence and Statistics*, R. Cowell and Z. Ghahramani (eds), volume **10**, 333–340. Barbados: AISTATS.
- Wakeley, J. and Sargsyan, O. (2009). Extensions of the coalescent effective population size. *Genetics* **181**, 341–345.

Received December 2011. Revised October 2012.

Accepted October 2012.