Received: 13 August 2018 | Accepted: 9 July 2019

DOI: 10.1111/biom.13276

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Horseshoe-based Bayesian nonparametric estimation of effective population size trajectories

James R. Faulkner[1,2] | **Andrew F. Magee**[3] | **Beth Shapiro**[4,5] | **Vladimir N. Minin**[6]

[1]Quantitative Ecology and Resource Management, University of Washington, Seattle, Washington

[2]Fish Ecology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, Washington

[3]Department of Biology, University of Washington, Seattle, Washington

[4]Ecology and Evolutionary Biology Department and Genomics Institute, University of California Santa Cruz, Santa Cruz, California

[5]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, California

[6]Department of Statistics, University of California Irvine, Irvine, California

Correspondence
Vladimir N. Minin, Department of Statistics, University of California Irvine, Irvine, CA 92697.
Email: vminin@uci.edu

Funding information
National Institute of Allergy and Infectious Diseases, Grant/Award Number: R01-AI107034; National Institute of General Medical Sciences, Grant/Award Number: U54-GM111274; Division of Environmental Biology, Grant/Award Number: 1754461

Abstract

Phylodynamics is an area of population genetics that uses genetic sequence data to estimate past population dynamics. Modern state-of-the-art Bayesian nonparametric methods for recovering population size trajectories of unknown form use either change-point models or Gaussian process priors. Change-point models suffer from computational issues when the number of change-points is unknown and needs to be estimated. Gaussian process-based methods lack local adaptivity and cannot accurately recover trajectories that exhibit features such as abrupt changes in trend or varying levels of smoothness. We propose a novel, locally adaptive approach to Bayesian nonparametric phylodynamic inference that has the flexibility to accommodate a large class of functional behaviors. Local adaptivity results from modeling the log-transformed effective population size a priori as a horseshoe Markov random field, a recently proposed statistical model that blends together the best properties of the change-point and Gaussian process modeling paradigms. We use simulated data to assess model performance, and find that our proposed method results in reduced bias and increased precision when compared to contemporary methods. We also use our models to reconstruct past changes in genetic diversity of human hepatitis C virus in Egypt and to estimate population size changes of ancient and modern steppe bison. These analyses show that our new method captures features of the population size trajectories that were missed by the state-of-the-art methods.

KEYWORDS

coalescent, Gaussian Markov random field, phylodynamics, phylogenetics, shrinkage prior

1 | INTRODUCTION

Estimation of population sizes and population dynamics over time is an important task in ecology and epidemiology. Census population sizes can be difficult to estimate due to infeasible sampling requirements or study costs. Genetic sequences are a growing source of information that can be used to infer past population sizes from the signatures of genetic diversity. Phylodynamics is a discipline that uses genetic sequence data to estimate past population dynamics. Many phylodynamic models draw on coalescent theory (Kingman, 1982; Griffiths and Tavaré, 1994), which provides a probabilistic framework that connects the branching times of a genealogical tree with the effective population size and other demographic variables, such as migration rates, of the population from which the genealogy was drawn. Effective population size can be interpreted as a measure of genetic diversity in a population and is proportional to census population size if coalescent model assumptions are met. When genetic diversity is high, the effective population size approaches the census

 © 2020 The International Biometric Society

population size, given random mating and no inbreeding or genetic drift, but is otherwise smaller than the census size. In our work, we concentrate on estimation of effective population sizes over evolutionary time, which can be short for rapidly evolving virus populations and longer (but still estimable with preserved ancient molecular sequence samples) for more slowly evolving organisms. Some examples of successful application of phylodynamics include describing seasonal trends of influenza virus spread around the world (Rambaut *et al.*, 2008), quantifying dynamics of outbreaks like hepatitis C (Pybus *et al.*, 2003) and Ebola viruses (Alizon *et al.*, 2014), and assessing the effects of climate change on populations of large mammals during the ice ages using ancient DNA (Shapiro *et al.*, 2004; Lorenzen *et al.*, 2011).

Some approaches to phylodynamics use parametric functional relationships to describe effective population size trajectories (eg, Pybus *et al.*, 2003; Rasmussen *et al.*, 2014), but nonparametric methods offer a flexible alternative when an accurate estimate of a complex population size trajectory is needed and knowledge of the mechanisms driving population size changes is incomplete. Nonparametric models have a long history of use in inferring effective population size trajectories. Pybus *et al.* (2000) introduced a nonparametric method, called the skyline plot, that produced point-wise estimates of population size, where the number of estimates was equal to the number of sampled genetic sequences minus one. The estimates from this method were highly variable, so a modification, referred to as the generalized skyline plot, created a set of discrete time interval groups that shared a single effective population size (Strimmer and Pybus, 2001). These likelihood-based approaches were adapted to a Bayesian framework with the Bayesian skyline plot (Drummond *et al.*, 2005) and the variable-knot spline approach of Opgen-Rhein *et al.* (2005). Minin *et al.* (2008) provided an alternative to these change-point methods by introducing a Gaussian Markov random field (GMRF) smoothing prior that connected the piecewise-constant population size estimates between coalescent events without needing to specify or estimate knot locations. Palacios and Minin (2012) and Gill *et al.* (2013) extended the GMRF approach of Minin *et al.* (2008) by constructing a GMRF prior on a discrete uniform grid. A grid-free approach, introduced by Palacios and Minin (2013), allowed the population size trajectories to vary continuously by using a Gaussian process (GP) prior.

Modern nonparametric Bayesian methods offer the state-of-the-art for recovering effective population size trajectories of unknown form. However, current methods cannot accurately recover trajectories that exhibit challenging features such as abrupt changes or varying levels of smoothness. Such features may arise in populations in the form of bottlenecks, rapid population changes, or aperiodic fluctuations with varying amplitudes. Accurate estimation of features like these can be important for understanding the demographic history of a population. Outside of phylodynamics, various nonparametric statistical methods have been developed to deal with such nonstationary or locally varying behavior under more standard likelihoods. These methods include, but are not limited to, GPs with nonstationary covariance functions (Paciorek and Schervish, 2006), nonstationary process convolutions (Higdon, 1998; Fuentes, 2002), non-Gaussian Matérn fields (Wallin and Bolin, 2015), and adaptive smoothing splines (Yue *et al.*, 2012, 2014). Each of these methods has good qualities and could potentially be adapted for inferring effective population sizes, but methods based on continuous random fields or process convolutions can be computationally challenging for large data sets, and some spline methods require selection or modeling of the number and location of knots.

A recent method by Faulkner and Minin (2018) uses shrinkage priors in combination with Markov random fields to perform nonparametric smoothing with locally adaptive properties. This is a fully Bayesian method that does not require the use of knots and avoids the costly computations of inverting dense covariance matrices. Computations instead take advantage of the sparsity in the precision matrix of the Markov random field to avoid matrix inversion. Faulkner and Minin (2018) compared different specifications of their shrinkage prior Markov random field (SPMRF) models and found that putting a horseshoe prior on the kth order differences between successive function values had superior performance when applied to underlying functions with sharp breaks or varying levels of smoothness. We refer to the model with the horseshoe prior as a horseshoe Markov random field (HSMRF).

In this paper, we propose an adaptation of the HSMRF approach of Faulkner and Minin (2018) for use in phylodynamic inference with coalescent priors. We devise a new Markov chain Monte Carlo (MCMC) scheme for the model that uses efficient, tuning-parameter-free, high-dimensional block updates. We provide an implementation of this MCMC in the program `RevBayes`, which allows us to target the joint distribution of genealogy, evolutionary model parameters, and effective population size parameters. We also develop a method for setting the hyperparameter on the prior for the global shrinkage parameter for coalescent data. We use simulations to compare the performance of the HSMRF model to that of a GMRF model and show that our model has lower bias and higher precision across a set of population trajectories that are difficult to estimate. We then apply our model to two real data examples that are well known in the phylodynamics literature and compare its performance to other popular nonparametric methods. The first example reanalyzes epidemiological dynamics of hepatitis C virus in Egypt and the second looks at estimation of ancient bison population size changes from DNA data.

2 | METHODS

2.1 | Sequence data and substitution model

Suppose we have a set of n aligned RNA or DNA sequences for a set of L sites within a gene. We assume the sequences come from a random sample of n individuals from a well-mixed population, where samples were collected potentially at different times. Let $\mathbf{Y}$ be the $n \times L$ sequence alignment matrix. We assume the sites are fully linked with no recombination possible between the sequences. This allows us to assume the existence of a genealogy $\boldsymbol{g}$, which is a rooted bifurcating tree that describes the ancestral relationships among the sampled individuals.

We assume that $\mathbf{Y}$ is generated by a continuous time Markov chain (CTMC) substitution model that models the evolution of the discrete states (eg, A,C,T,G for DNA) along the genealogy $\boldsymbol{g}$ for each alignment site. A variety of substitution models are available and are typically differentiated by the form of the transition matrix $M(\boldsymbol{\Omega})$, which controls the substitution rates in the CTMC for the nucleotide bases with a set of parameters $\boldsymbol{\Omega}$ (see Yang (2014) for examples). Let the likelihood of the sequence data given the genealogy and substitution parameters be denoted by $p(\mathbf{Y} \mid \boldsymbol{g}, \boldsymbol{\Omega})$.

2.2 | Coalescent

Suppose that we now have a genealogy $\boldsymbol{g}$, where branch lengths of the genealogical tree are measured in units of clock time (eg, years). To build a Bayesian hierarchical model, we need a prior density for $\boldsymbol{g}$. The times at which two lineages merge into a common ancestor on the tree are called coalescent times. The coalescent model provides a probabilistic framework for relating the coalescent times in the sample to the effective size of the population. Kingman (1982) developed the coalescent model for a constant effective population size and Griffiths and Tavaré (1994) extended it for varying effective population sizes.

Let the $n-1$ coalescent times arising from genealogy $\boldsymbol{g}$ be denoted by $0 < t_{n-1} < \cdots < t_1$, where 0 is the present and time is measured backward from there. We will assume the general case where sampling of the genetic sequences occurs at different times (*heterochronous sampling*), which will include the special case where all sampling occurs at time 0 (*isochronous sampling*). We denote the set of unique sampling times as $s_m = 0 < s_{m-1} < \cdots < s_1 < t_1$ for samples of size $n_m, \ldots, n_1$, respectively, where $n = \sum_{j=1}^{m} n_j$ and we assume no sample times are equal to coalescent times (Figure 1). We let $\boldsymbol{s}$ denote the vector of sampling times. Further, we let the intervals that end with a coalescent event be denoted $I_{0,k} = (\max\{t_{k+1}, s_j\}, t_k]$, for $s_j < t_k$ and $k = 1, \ldots, n-1$, and let the intervals that end with a sampling event be denoted $I_{i,k} = (\max\{t_{k+1}, s_{j+i}\}, s_{j+i-1}]$, for $s_{j+i-1} > t_{k+1}$ and $s_{j+i} < t_k$, $k = 1, \ldots, n-1$. For $k = n-1$, we substitute $t_{k+1} = 0$. We let $n_{i,k}$ be the number of lineages present in interval $I_{i,k}$ and let the vector of number of lineages be denoted $\boldsymbol{n}$. Further, we denote the number of unique sampling times in interval $(t_{k+1}, t_k]$ as m_k, where $m = 1 + \sum_{k=1}^{n-1} m_k$. The joint density of the coalescent times given $\boldsymbol{s}$ and the effective population size trajectory $N_e(t)$ can then be written as

$$\begin{aligned} p(t_1, \ldots, t_{n-1} \mid \boldsymbol{s}, \boldsymbol{n}, N_e(t)) &= \prod_{k=1}^{n-1} p(t_k \mid t_{k+1}, \boldsymbol{s}, \boldsymbol{n}, N_e(t)) \\ &= \prod_{k=1}^{n-1} \frac{C_{0,k}}{N_e(t_k)} e^{-\sum_{i=0}^{m_k} \int_{I_{i,k}} \frac{C_{i,k}}{N_e(t)} dt}, \end{aligned} \tag{1}$$

where $C_{i,k} = \binom{n_{i,k}}{2}$ is the coalescent factor (Felsenstein and Rodrigo, 1999). This model can be seen as an inhomogeneous Markov point process where the conditional intensity is $C_{i,k}[N_e(t)]^{-1}$ (Palacios and Minin, 2013).

Here we assume $N_e(t)$ is an unknown continuous function, so the integrals in Equation (1) must be computed with numerical approximation techniques. We follow Palacios and Minin (2012), Gill *et al.* (2013), and Lan *et al.* (2015) and use discrete approximations of the integrals over a finite grid. We construct a regular grid, $\boldsymbol{x} = \{x_h\}_{h=1}^{H+1}$, and set the end points of the grid $\boldsymbol{x}$ such that $x_1 = 0$ and $x_{H+1} = t_1$ (Figure 1). This results in H grid cells and $H+1$ cell boundaries. Now for $t \in (x_h, x_{h+1}]$, we have $N_e(t) \approx \exp[\theta_h]$, where θ_h is an unknown model parameter. This implies that $\boldsymbol{\theta} = \{\theta_h\}_{h=1}^{H}$ is a piecewise-constant approximation to $f(t) = \ln[N_e(t)]$ for $t \in [s_m, t_1]$. The piecewise constant population size can be integrated analytically, leading to a discrete approximation to the likelihood in Equation (1). The details of this approximation are provided in Web Appendix A.

2.3 | Prior for effective population size trajectory

Next we develop a prior for the unknown function $N_e(t)$ that describes the effective population size trajectory over time. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_H)$ be a vector of parameters that govern the effective population size trajectory $N_e(t)$. We propose using a SPMRF model (Faulkner and Minin, 2018) for $\boldsymbol{\theta}$, which is a type of Markov model where the pth-order differences in the forward-time evolution of the sequence $\{\theta_h\}_{h=1}^{H}$ are independent and follow a shrinkage prior distribution. We define the pth-order forward difference as $\Delta^p \theta_l \equiv (-1)^p \sum_{j=0}^{p} (-1)^j \binom{p}{j} \theta_{l+j-p+1}$, for $l = p, \ldots, H-1$, which is a discrete approximation to the pth derivative of $f(t)$ evaluated at t. If we assume a horseshoe distribution (Carvalho *et al.*,

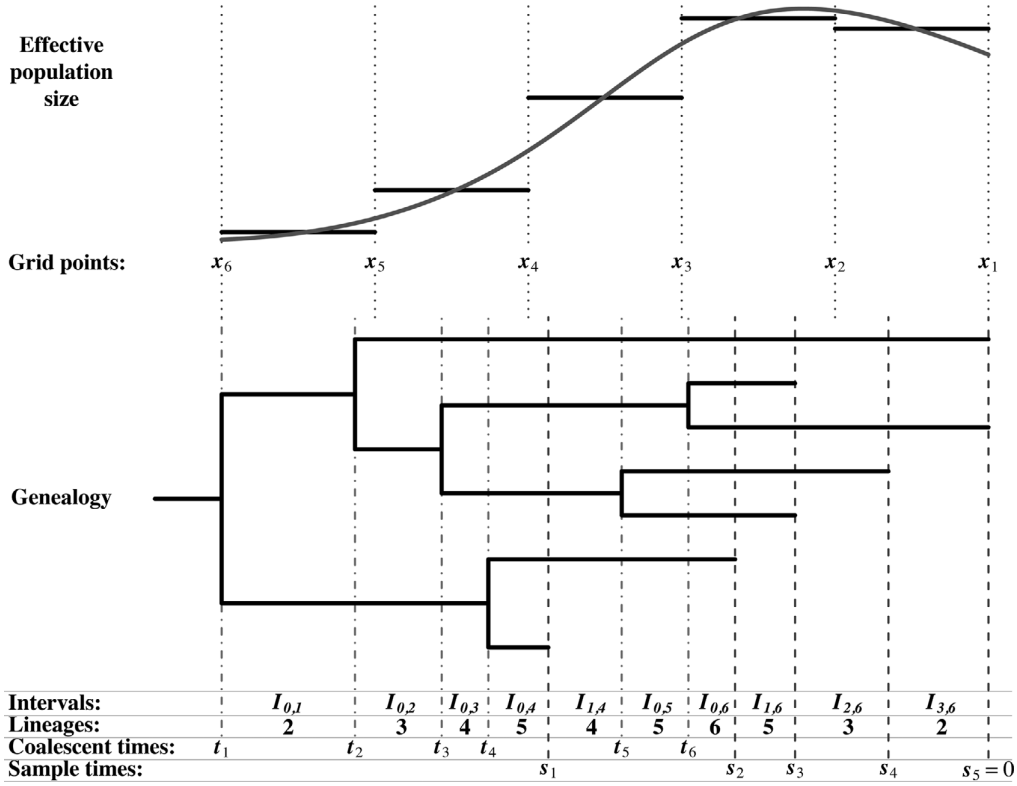

FIGURE 1 Effective population size trajectory and associated genealogical tree under heterochronous sampling. The top panel shows a continuous effective population size trajectory (gray) and an associated piecewise constant approximation to it. Also shown are the relationships between the genealogy and sampling times s_i, coalescent times t_i, intervals $I_{i,k}$, number of lineages $n_{i,k}$, and the uniform grid points, x_h, used for approximating coalescent densities

2010) as our shrinkage prior on the order-p differences in $\boldsymbol{\theta}$, then

$$\Delta^p \theta_l \mid \gamma \sim \mathcal{HS}(\gamma), \tag{2}$$

where the location parameter of the horseshoe distribution is zero and γ is the scale parameter and controls how much $f(t)$ is allowed to vary a priori. Following Carvalho *et al.* (2010), we put a half-Cauchy prior on γ with scale hyperparameter ζ, so that $\gamma \sim \mathcal{C}^+(0, \zeta)$. We chose the half-Cauchy here because it has desirable properties as a prior on a scale parameter (Gelman *et al.*, 2006; Polson and Scott, 2012) and its single hyperparameter simplifies implementation. Depending on the order p of the model, we also place proper priors on $\theta_1, \ldots, \theta_p$. To do this, we start by setting $\theta_1 \sim \mathcal{N}(\mu, \sigma^2)$, where μ and σ are hyperparameters typically set to create a diffuse prior. Then for $p \geq 2$ and $q = 1, \ldots, p-1$, we let $\Delta^q \theta_q \mid \gamma \sim \mathcal{HS}(a_q \gamma)$, where $a_q = 2^{-(p-q)/2}$, which follows from the recursive property and independence of the order-p differences. For example, for $p = 2$, $a_1 = 2^{-1/2}$, and for $p = 3$, $a_2 = 2^{-1/2}$ and $a_1 = 4^{-1/2}$. We will refer to this specific model formulation as a state-space formulation of a HSMRF.

The horseshoe distribution is leptokurtic with an infinite spike in density at zero and Cauchy-like tails. In our setting, this combination results in small θ differences being shrunk toward zero and larger differences being maintained, which corresponds to smoothing over smaller noisy signals while retaining the ability to adapt to rapid functional changes. This is in contrast to the normal distribution, which has higher density around medium-sized values and normal tails. These attributes result in noisier estimates and reduced ability to capture abrupt functional changes. Different shrinkage priors will result in different levels of shrinkage and therefore different smoothing behavior. Faulkner and Minin (2018) found that the horseshoe prior performed better than the Laplace prior in terms of bias and precision for nonparametric smoothing with SPMRFs, but we do not investigate the effect of different shrinkage priors here.

The horseshoe density does not have a closed form (although see Faulkner and Minin (2018) for an approximation in closed form). However, a horseshoe distribution can be represented hierarchically as a scale mixture of normal distributions by introducing a latent scale parameter that follows a half-Cauchy distribution (Carvalho *et al.*, 2010). That is, if

$\tau_l \sim C^+(0,\gamma)$ and $\Delta^p\theta_l \mid \tau_l \sim \mathcal{N}(0,\tau_l^2)$, then integrating over τ_l results in the marginal relationship in Equation (2).

The hierarchical HSMRF models are a type of pth-order normal random walk with separate variance parameters for each increment. The inherent Markov properties and properties of the normal distribution allow the joint distribution of $\boldsymbol{\theta}$ conditional on the vector of scale parameters $\boldsymbol{\tau}$ to be expressed $p(\boldsymbol{\theta} \mid \boldsymbol{\tau}, \mu, \sigma^2) = p(\theta_1 \mid \mu, \sigma^2)p(\Delta^1\theta_1, \ldots, \Delta^p\theta_p, \Delta^p\theta_{p+1}, \ldots \Delta^p\theta_{H-1} \mid \boldsymbol{\tau})$, which results in a multivariate normal distribution with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q}(\boldsymbol{\tau})$. Specifically, $\boldsymbol{\theta}$ follows a Gaussian Markov random field (Rue and Held, 2005) conditional on $\boldsymbol{\tau}$, where the order p of the differencing in θ determines the structure of the sparse $\mathbf{Q}(\boldsymbol{\tau})$. For the models presented here, $\boldsymbol{\mu} = \mu\mathbf{1}$, where μ is a constant and $\mathbf{1}$ is a vector of ones. We specify $p(\boldsymbol{\tau})$ by assuming that the τ's are independent $C^+(0,\gamma)$-distributed random variables, where $\tau_l \sim C^+(0,\gamma)$ for $l = p, \ldots, H-1$ and $\tau_l \sim C^+(0, a_l\gamma)$ for $l = 1, \ldots, p-1$ and $p \geq 2$. The marginal joint distribution of $\boldsymbol{\theta}$ that results from integrating over $\boldsymbol{\tau}$ is an HSMRF. Note that a GMRF model results when a single scale parameter τ is used for all order-p differences in $\boldsymbol{\theta}$. For our GMRF models, we use $\tau \sim C^+(0,\zeta)$, where ζ is a fixed hyperparameter. The order of the HSMRF will determine the amount of smoothing, with higher orders resulting in more smoothing. We only consider first-order and second-order models here. In practice, we use the state-space formulation described previously but with the independent hierarchical representations of the horseshoe distributions for the individual order-p differences, which improves computational efficiency over the conditional multivariate normal representation.

2.4 | Posterior inference

For the case where we have a fixed genealogical tree, $\boldsymbol{g}$, which consists of sampling times $\boldsymbol{s}$ and coalescent times $\boldsymbol{t}$, the posterior distribution of the parameters $\{\boldsymbol{\theta}, \boldsymbol{\tau}, \gamma\}$ can be written as

$$p(\boldsymbol{\theta}, \boldsymbol{\tau}, \gamma \mid \boldsymbol{g}) \propto p(\boldsymbol{g} \mid \boldsymbol{\theta})p(\boldsymbol{\theta} \mid \boldsymbol{\tau})p(\boldsymbol{\tau} \mid \gamma)p(\gamma). \tag{3}$$

Here $\boldsymbol{g}$ is considered data and we assume the coalescent times are known. Then $p(\boldsymbol{g} \mid \boldsymbol{\theta})$ is the coalescent likelihood and $p(\boldsymbol{\theta} \mid \boldsymbol{\tau})p(\boldsymbol{\tau} \mid \gamma)p(\gamma)$ is the HSMRF prior described in Section 2.3. For our GMRF models, the right-hand side of Equation (3) becomes $p(\boldsymbol{g} \mid \boldsymbol{\theta})p(\boldsymbol{\theta} \mid \tau)p(\tau)$.

For our analyses with fixed genealogical trees, we follow Faulkner and Minin (2018) and Lan *et al.* (2015) and use Hamiltonian Monte Carlo (HMC; Neal, 2011) for posterior inference. HMC performs joint proposals for the parameters that are typically far from the current parameter state and have high acceptance rates, resulting in efficient posterior sampling. We used the `Stan` computing environment (Carpenter *et al.*, 2016) for implementing HMC. Specifically, we used the open source package `rstan` (Stan Development Team, 2017), which provides a platform for fitting models using HMC in the `R` computing environment (R Core Team, 2017). Our `R` package titled `spmrf` allows for easy implementation of our models for use on fixed genealogical trees via a wrapper to the `rstan` tools. A link to the package code is provided in the Supporting Information section. We present a method for objectively setting the scale hyperparameter ζ of the prior distribution of the global smoothing parameter γ in Web Appendix B.

When there are genetic sequence data available and we want to jointly estimate evolutionary parameters, coalescent times, and population size trajectories, our posterior can be written as

$$p(\boldsymbol{g}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \boldsymbol{\tau}, \gamma \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \boldsymbol{g}, \boldsymbol{\Omega})p(\boldsymbol{g} \mid \boldsymbol{\theta})p(\boldsymbol{\Omega}) \cdot p(\boldsymbol{\theta} \mid \boldsymbol{\tau})p(\boldsymbol{\tau} \mid \gamma)p(\gamma), \tag{4}$$

where $\mathbf{Y}$ are the sequence data and $\boldsymbol{\Omega}$ are the parameters related to the DNA substitution model. The likelihood of the sequence data given the parameters is $p(\mathbf{Y} \mid \boldsymbol{g}, \boldsymbol{\Omega})$, and now $p(\boldsymbol{g} \mid \boldsymbol{\theta})$ is a prior for the genealogy given the population sizes and is proportional to $p(\boldsymbol{g} \mid \boldsymbol{\theta})$ in Equation (3). The remaining components are the prior for the evolution parameters $p(\boldsymbol{\Omega})$ and the HSMRF prior as in Equation (3).

HMC requires the calculation of gradients over continuous parameter space and therefore cannot be used for inference on discrete parameters. Therefore, we developed a custom MCMC algorithm that uses a combination of Gibbs sampling, elliptical slice sampling, and the Metropolis-Hastings (MH) algorithm to sample from the joint posterior of the evolution parameters and the effective population size parameters. In particular, elliptical slice sampling (Murray *et al.*, 2010) was used to sample from the joint field of log effective population sizes conditional on the latent scale parameters, a Gibbs sampler based on an approach developed by Makalic and Schmidt (2016) for horseshoe random variables was used to sample the latent scale parameters conditional on the field parameters, and standard phylogenetic MH steps were used to update the genealogy and substitution model parameters. We implemented our custom MCMC in `RevBayes`—a statistical computing environment geared primarily for phylogenetic inference (Höhna *et al.*, 2016). The standard phylogenetic MH updates mentioned above were already implemented in `RevBayes`, so we contributed a heterochronous coalescent likelihood calculator, elliptical slice sampling, and Gibbs updates of our model parameters to the `RevBayes` source code. The details of the sampling scheme are provided in Web Appendix C and a link to the code for implementing our methods for analyzing sequence data is provided in the Supporting Information section.

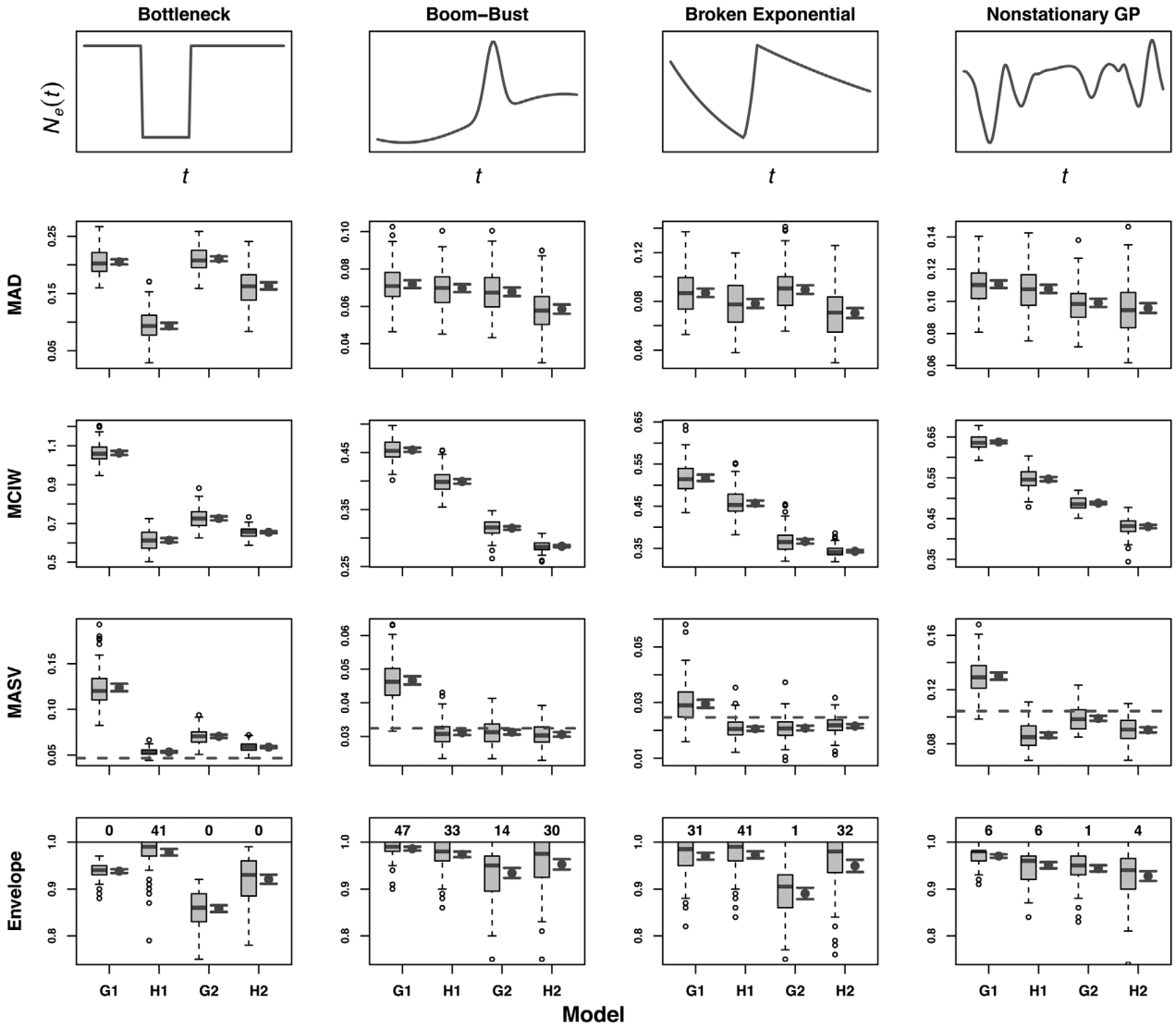

FIGURE 2 Effective population size trajectories used in simulations and simulation results by model and scenario. Models are GMRF of order 1 (G1) and order 2 (G2) and HSMRF of order 1 (H1) and order 2 (H2). Top row shows true effective population size trajectories used to simulate coalescent data. Remaining rows show mean absolute deviation (MAD), mean credible interval width (MCIW), mean absolute sequential variation (MASV), and credible interval Envelope. Horizontal dashed lines in the third row plots indicate the true mean absolute sequential variation (TMASV) values. Shown for each model are standard boxplots of the performance metrics (left) and mean values with 95% frequentist confidence intervals (right). Also shown for Envelope are the number of simulations with Envelope equal to 1.0

3 | RESULTS

3.1 | Simulated data

We used simulated data to assess the performance of the HSMRF model relative to the GMRF model. We investigated four scenarios with different trajectories for $N_e(t)$: (a) Bottleneck (BN), (b) Boom-Bust (BB), (c) Broken Exponential (BE), and (d) Nonstationary Gaussian Process (NGP) realization. The trajectory shapes are shown at the top of Figure 2. For each scenario, we generated 100 data sets of coalescent times and fit GMRF and HSMRF models of first and second order using the fixed-tree approach. The scenario descriptions and further methodological details of the simulations are provided in Web Appendix D.

We assessed the relative performance of the models using a set of summary statistics. As a measure of bias, we used the mean absolute deviation (MAD) to compare the posterior medians of the trend parameters ($\hat{\theta}_i$) to the true trend values (θ_i): $\text{MAD} = \frac{1}{H}\sum_{i=1}^{H}|\hat{\theta}_i - \theta_i|$. We assessed the width of the 95% Bayesian credible intervals (BCIs) using the mean credible interval width (MCIW):

MCIW $= \frac{1}{H}\sum_{i=1}^{H}(\hat{\theta}_{97.5,i} - \hat{\theta}_{2.5,i})$, where $\hat{\theta}_{97.5,i}$ and $\hat{\theta}_{2.5,i}$ are the 97.5% and 2.5% quantiles of the posterior distribution for θ_i. We assessed the coverage of BCIs using Envelope $= \frac{1}{H}\sum_{i=1}^{H} I(\theta_i \in [\hat{\theta}_{97.5,i}, \hat{\theta}_{2.5,i}])$, where $I(\cdot)$ is the indicator function. To measure local variability in the estimated population trend, we used the mean absolute sequential variation (MASV) of $\hat{\boldsymbol{\theta}}$, which was computed as MASV $= \frac{1}{H-1}\sum_{i=1}^{H-1}|\hat{\theta}_{i+1} - \hat{\theta}_i|$. We compared the observed MASV to the true MASV (TMASV) in the underlying trend function, which is calculated by substituting true θ's into the equation for MASV. For a measure of model complexity, we estimated the effective number of parameters p_{eff} using an approach suggested by Raftery *et al.* (2006): $p_{\text{eff}} = \frac{2}{R-1}\sum_{r=1}^{R}(\mathcal{L}_r - \bar{\mathcal{L}})^2$, where $\mathcal{L}_r$ is the log-likelihood evaluated at the parameter values for the rth of R samples from the posterior, and $\bar{\mathcal{L}}$ is the mean value of $\mathcal{L}$ across the R samples. We used the Watanabe-Akaike information criterion (WAIC; Watanabe, 2010) to calculate model weights and rank model performance. The weight for model m was calculated as $w_m = \exp(-0.5\Delta\text{W}_m)/\sum_{j=1}^{M}\exp(-0.5\Delta\text{W}_j)$ for a set of M models, where $\Delta\text{W}_m = \text{WAIC}_m - \min_{j\in M}\text{WAIC}_j$. We utilized the `loo` package (Vehtari *et al.*, 2017) to calculate WAIC. For a measure of computational efficiency, we calculated the mean effective sample size (ESS) of the posterior samples across parameters for each model and simulated data set and used those with the total sampling times to calculate the mean ESS per second of sampling time.

For the BN scenario, the HSMRF model clearly had better performance than the GMRF model for the main performance metrics for both model orders (Figure 2, Table 1, and Table 1 in Web Appendix D). Example model fits from each scenario provide some intuition for the simulation results (Figure 3). First-order models did better than second-order models within model types for the BN scenario. Differences between model types were not as strong for the other scenarios. The second-order HSMRF performed the best in terms of MAD, MCIW, and WAIC for the remaining scenarios. Among second-order models, the HSMRF was clearly favored over the GMRF in terms of WAIC across all scenarios. However, the HSMRF models were not noticeably different from the second-order GMRF in terms of MASV for the BB and BE scenarios. The second-order GMRF had mean MASV closer to TMASV than did the second-order HSMRF for the NGP scenario. Although the GMRF tended to estimate excess variation in the middle section of the trend for the NGP scenario, it did capture the peaks and troughs a little better than the HSMRF in other parts of the trend (see Figure 3 for an example). In all scenarios, the HSMRF had lower p_{eff} compared to the GMRF of the same order. The GMRF was consistently more computationally efficient than the HSMRF, with mean ESS/second approximately 1.5 to 6 times higher for models of the same order. These differences are due to the additional parameters in the HSMRF models. The second-order models were relatively slow for both model types, but the HSMRF was always slower. As we show in the following data examples, however, the differences in computational speed between the HSMRF and GMRF models is negligible when genealogies and effective population size trajectories are jointly estimated.

TABLE 1 Summary of model selection criteria across 100 simulations by scenario and model set

Metric	Model Set	Model	BN	BB	BE	NGP
Best model (%)	All models	G1	1	9	13	1
		H1	**93**	14	34	9
		G2	0	3	1	24
		H2	6	**74**	**52**	**66**
	Order 1	G1	1	**51**	29	50
		H1	**99**	49	**71**	50
	Order 2	G2	9	9	5	27
		H2	**91**	**91**	**95**	**73**
Mean weight	All models	G1	0.03	0.11	0.14	0.04
		H1	0.89	0.15	0.35	0.09
		G2	0.01	0.10	0.07	0.26
		H2	0.08	0.63	0.44	0.61
	Order 1	G1	0.03	0.48	0.24	0.46
		H1	0.97	0.52	0.76	0.54
	Order 2	G2	0.12	0.11	0.11	0.43
		H2	0.88	0.89	0.89	0.57

Note. WAIC weights were calculated and the best model (greatest WAIC weight) was determined for each simulated data set within each scenario and model set. Metrics shown are the percentage of simulations each model was determined best and the mean model weight across simulations. Values for each metric are compared among models within each scenario and model set. Highest percentage of best models is in bold within each scenario and model set. Scenarios are Bottleneck (BN), Boom-Bust (BB), Broken Exponential (BE), and Nonstationary Gaussian Process (NGP). Models are GMRF of order 1 (G1) and order 2 (G2) and HSMRF of order 1 (H1) and order 2 (H2).

3.2 | Egyptian hepatitis C virus

The hepatitis C virus (HCV) is a blood-borne RNA virus that exclusively infects humans. HCV infection is often asymptomatic, but can lead to liver disease and liver failure. HCV infections have historically had high prevalence in Egypt (Miller and Abu-Raddad, 2010). This is thought to be due to past widespread use of unsanitary medical practices in the region. Of particular interest is a treatment for the parasite disease schistosomiasis known as parenteral antischistosomal therapy (PAT), which uses intravenous injections. PAT was practiced from the 1920s to 1980s in Egypt and is thought to have contributed to the spread of HCV during that period due to unsterilized injection equipment (Frank *et al.*, 2000).

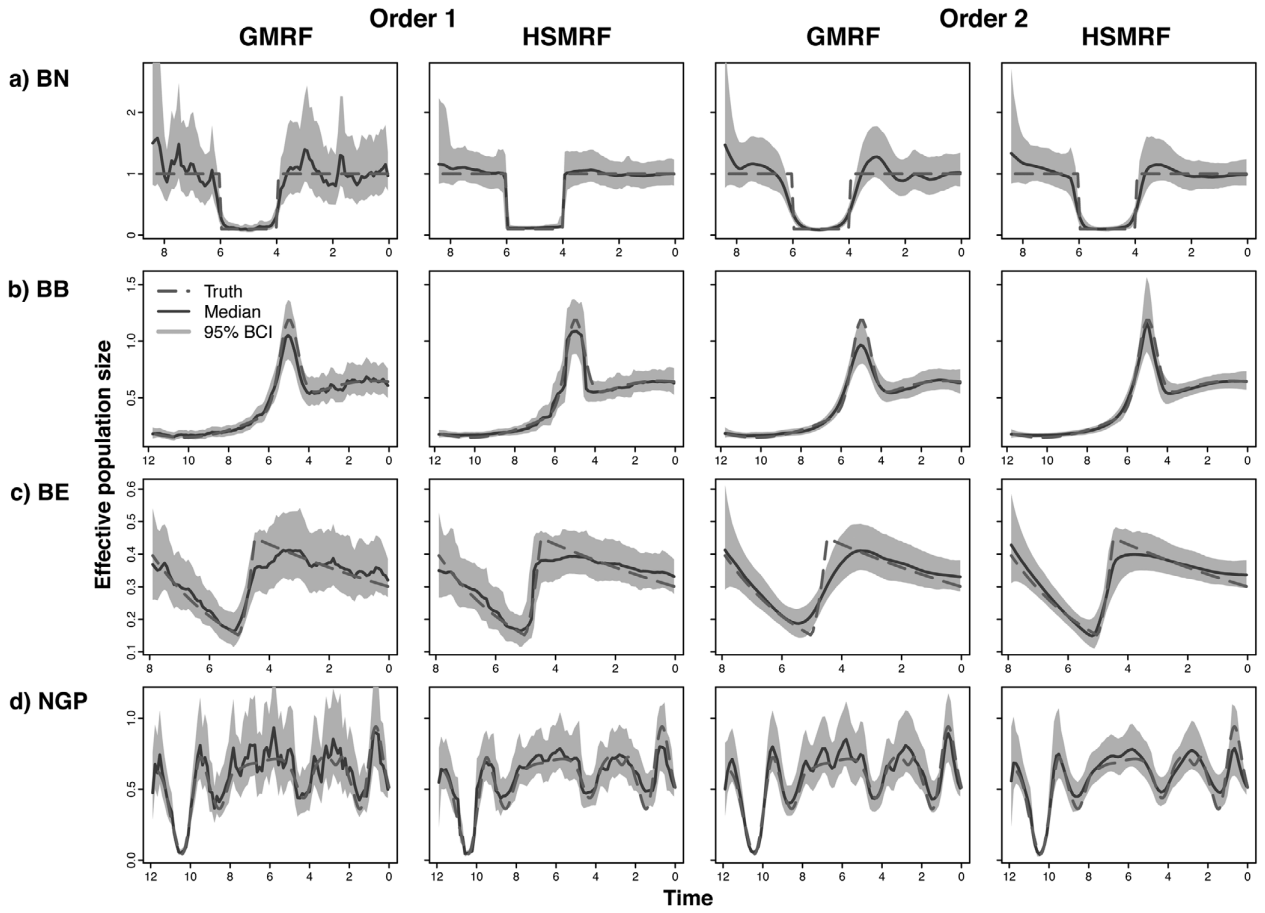

FIGURE 3 Example fits of first- and second-order Gaussian Markov random field and horseshoe Markov random field models for four different simulation scenarios. Scenarios are (a) Bottleneck, (b) Boom-Bust, (c) Broken Exponential, and (d) Nonstationary Gaussian Process. Results for all models within a particular scenario are for the same set of simulated data. Given are the true effective population size trajectories that generated the data (dashed line), posterior medians of estimated trajectories (solid line), and associated 95% Bayesian credible intervals (shaded band)

We analyze 63 RNA sequences of type 4 with 411 base pairs from the E1 region of the HCV genome that were collected in 1993 in Egypt (Ray *et al.*, 2000). Pybus *et al.* (2003) used a piecewise demographic model for effective population size with a period of exponential growth between two periods of constant population size and concluded that the HCV population grew exponentially during the period of PAT treatment. Other authors have applied nonparametric methods to estimate the effective population size trajectory for these data (eg, Drummond *et al.*, 2005; Minin *et al.*, 2008; Palacios and Minin, 2013). Different nonparametric methods lead to different estimated trajectories and different levels of uncertainty. We are interested in estimating the rapid change of HCV effective population size during the epidemic.

We fit six different nonparametric models to these data: (a) Bayesian Skyline—a piecewise constant/linear model with estimable locations of change-points (SkyLine; Drummond *et al.*, 2005), (b) Bayesian Skyride (SkyRide; Minin *et al.*, 2008) (c) GMRF-1 (similar to Bayesian Skygrid, Gill *et al.* (2013)), (d) GMRF-2, (e) HSMRF-1, and (f) HSMRF-2. We note that the SkyRide model is also a type of GMRF model where the nonuniform grid cell boundaries are determined by coalescent events. For all six models we jointly estimated the evolutionary model parameters, genealogies, and effective population size parameters. We used the program `BEAST` implementation of the SkyLine and SkyRide models (Drummond *et al.*, 2012), and used our own `RevBayes` implementation of the GMRF and HSMRF models. Although the Skygrid implementation of the GMRF-1 model is available in `BEAST`, the GMRF-2 and the HSMRF models are not, so we decided to use common software for the GMRF and HSMRF models. For the GMRF and HSMRF models, we used 100 equally spaced grid cells where the first 99 ended at a fixed boundary of 227 years before 1993, and the final cell captured any coalescent events beyond the boundary (see Web Appendix E for discussion on setting grids). The SkyLine model requires specification of the number of discrete population intervals, where each interval

describes a piecewise constant population size between two coalescent events. We used 20 population intervals to allow fair flexibility to capture sharp features in the population trajectory. Further details about the MCMC implementation and computation times are provided in Web Appendix F. For model comparison, we calculated posterior model probabilities using marginal likelihood estimates calculated with steppingstone sampling (Xie *et al.*, 2011). See Web Appendix G for details on calculation of posterior model probabilities.

Although the broad pattern of the demographic trajectory was similar among the six models, they differed in the estimated rate of change in effective population size and in the uncertainty around the effective population size estimates (Figure 4). The SkyLine and HSMRF-1 models had the highest posterior model probabilities, with the SkyLine favored a little over the HSMRF-1 (Figure 4). The shape of the median trajectory from the HSMRF-1 model was similar to that of the SkyLine model, yet the HSMRF-1 model showed a very rapid increase in population between 1925 and 1945, while the SkyLine and other models showed more gradual increases that started earlier and ended later. The increase estimated by the SkyRide model lasted the longest, starting near 1900 and ending near 1970. The HSMRF and the SkyLine also showed relatively constant population size following the increase in the mid 20th century, whereas the SkyRide and GMRF-1 models showed a decrease after 1970.

In addition to differing in the rate of population growth after the epidemic began, the models differed in their estimates of when the epidemic began. The posterior mean densities of frequencies of coalescent times provide an indication of when the HCV epidemic started (Figure 4). The results of the HSMRF-1 support the idea that HCV epidemic started after PAT was introduced and suggest that early PAT campaigns may have used less sanitary practices and contributed more to the spread of HCV than the major PAT campaigns started in the 1950s. Plots of the effective population trajectories covering the entire span of the coalescent times are provided with further discussion in Web Appendix H.

3.3 | Beringian steppe bison

Modern molecular methods have allowed the recovery of DNA samples from specimens that lived hundreds to hundreds of thousands of years ago (Pääbo *et al.*, 2004; Shapiro and Hofreiter, 2014). Large mammals that lived in the Northern Hemisphere during the Pleistocene and Holocene epochs have been a valuable source of this ancient DNA due to conditions favorable for specimen preservation in the northern latitudes (eg, Shapiro *et al.*, 2004; Lorenzen *et al.*, 2011). We focus on bison (*Bison* spp.) that lived on the steppe-tundra of Northern Asia and Europe and crossed into North America over the Bering land bridge during the middle to late Pleistocene (Froese *et al.*, 2017). Interest has been in determining whether human impact or climate and related habitat change instigated the decline of bison across their range during the late Pleistocene. Shapiro *et al.* (2004) used a parametric piecewise-exponential model for the bison effective population size and estimated that the time of transition from population growth to decline was 37 000 years ago (kya). Drummond *et al.* (2005) used the more flexible SkyLine model, which indicated a more rounded and prolonged peak in population size followed by a rapid decline and bottleneck around 10 kya. Here we use a modified version of the bison data described by Shapiro *et al.* (2004) and fit coalescent models directly to the sequence data as with the HCV data. We make qualitative comparisons among the resulting estimated population trajectories and in relation to some benchmark times describing the arrival of humans and the period of the Last Glacial Maximum (LGM).

We analyze 152 sequences (135 ancient and 17 modern) of mitrochondrial DNA with 602 base pairs from the mitochondrial control region. DNA was extracted from bison fossils from Alaska (68), Canada (46), Siberia (13), the lower 48 United States (6), and China (2). Sample dates were estimated for the ancient samples using radiocarbon dating, with dates ranging up to 59k years. We treat the calibrated radiocarbon dates as known in the following analyses. These data are the same as those used by Gill *et al.* (2013), and are slightly modified from the data first described by Shapiro *et al.* (2004) to remove sequences identified as potentially contaminated with young radiocarbon (Shapiro *et al.*, 2010) and include additional sequences generated since generation of the initial data set. In this data set, radiocarbon dates are calibrated to calendar time using the IntCal09 calibration curve (Reimer *et al.*, 2009).

The LGM in the Northern Hemisphere is estimated to have occurred between 26.5 and 19 kya (Clark *et al.*, 2009). A small, isolated population of humans existed in central Beringia, including, potentially, the land bridge that connected the continents during the LGM (Llamas *et al.*, 2016). Humans may have ventured into eastern Beringia (Alaska and Yukon) as early as 26 kya (Bourgeon *et al.*, 2017), but there is as yet no evidence of continuous occupation until 14 kya (Easton *et al.*, 2011; Holmes, 2011). Humans probably first entered continental North America via a western coastal route that became available close to 16 kya (Heintzman *et al.*, 2016; Llamas *et al.*, 2016), where they would have encountered the population of steppe bison that were isolated in the south with the coalescence of the Laurentide and Cordilleran glaciers (Shapiro *et al.*, 2004; Heintzman *et al.*, 2016). Because the majority of our bison samples were collected in North America, we used 16-14 kya as the time of first human occupation.

We used methods similar to those used in the HCV example. We also calculated posterior distributions for the time of

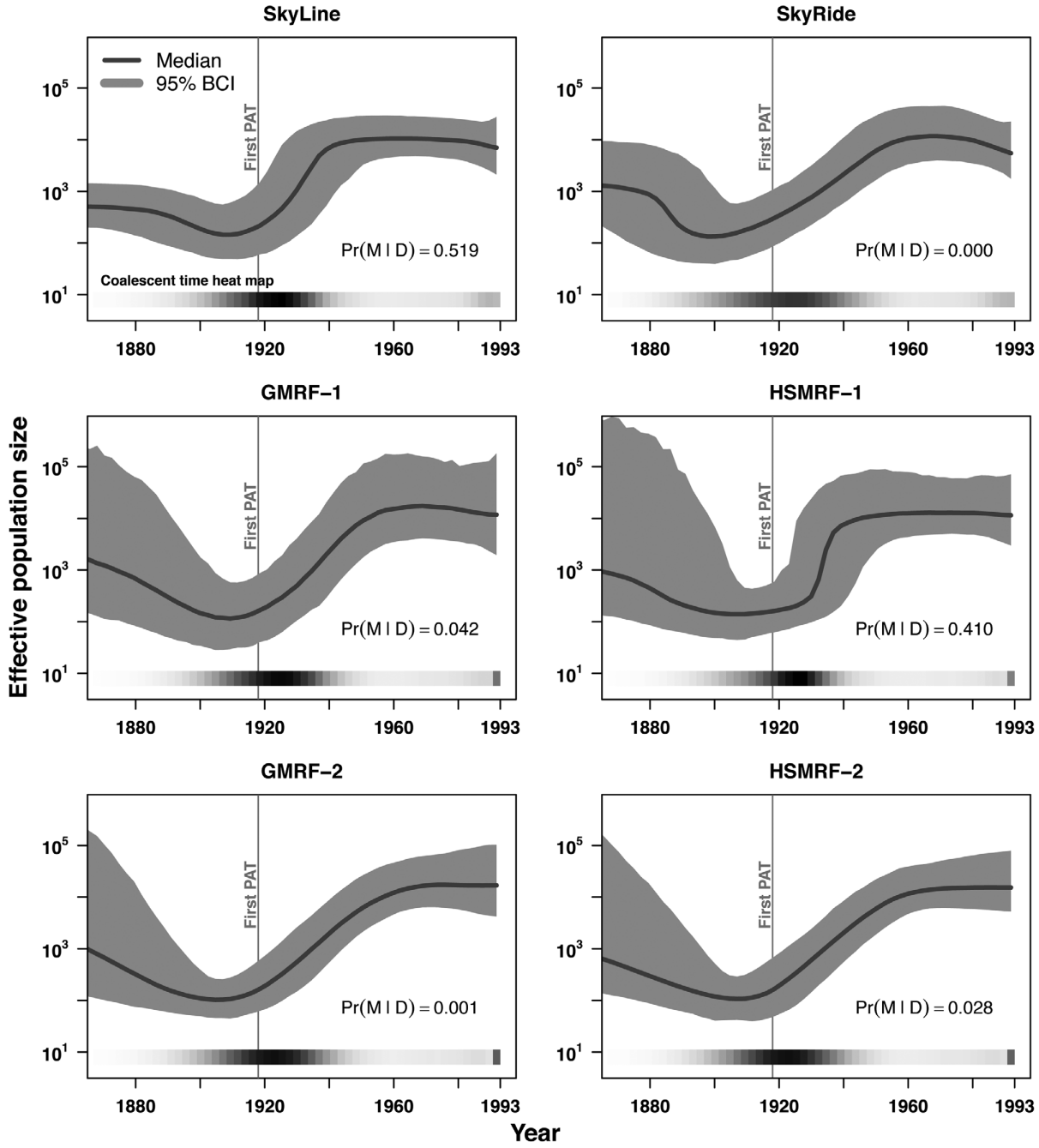

FIGURE 4 Posterior medians (solid black lines) of effective population sizes and associated 95% credible intervals (gray shaded areas) for the HCV data for the Bayesian Skyline (SkyLine), Bayesian Skyride (SkyRide), Gaussian Markov random field of order 1 (GMRF-1) and order 2 (GMRF-2), and horseshoe Markov random field of order 1 (HSMRF-1) and order 2 (HSMRF-2). Also shown for each model are posterior model probabilities (Pr(M | D)) and heat maps of mean posterior frequencies of coalescent times. A vertical reference line is shown at year 1918, which is the year PAT was introduced

the peak in population size. Method details can be found in Web Appendices F and G.

Although the broad pattern of an increase followed by a decrease in effective population size was recovered by all six models, the timing and nature of the population size change differed considerably between them (Figure 5). The HSMRF-1 model had the highest posterior model probability among the six models. The posterior median trajectory from the HSMRF-1 model was most similar to the SkyLine model, but the credible intervals for the HSMRF-1 model were most

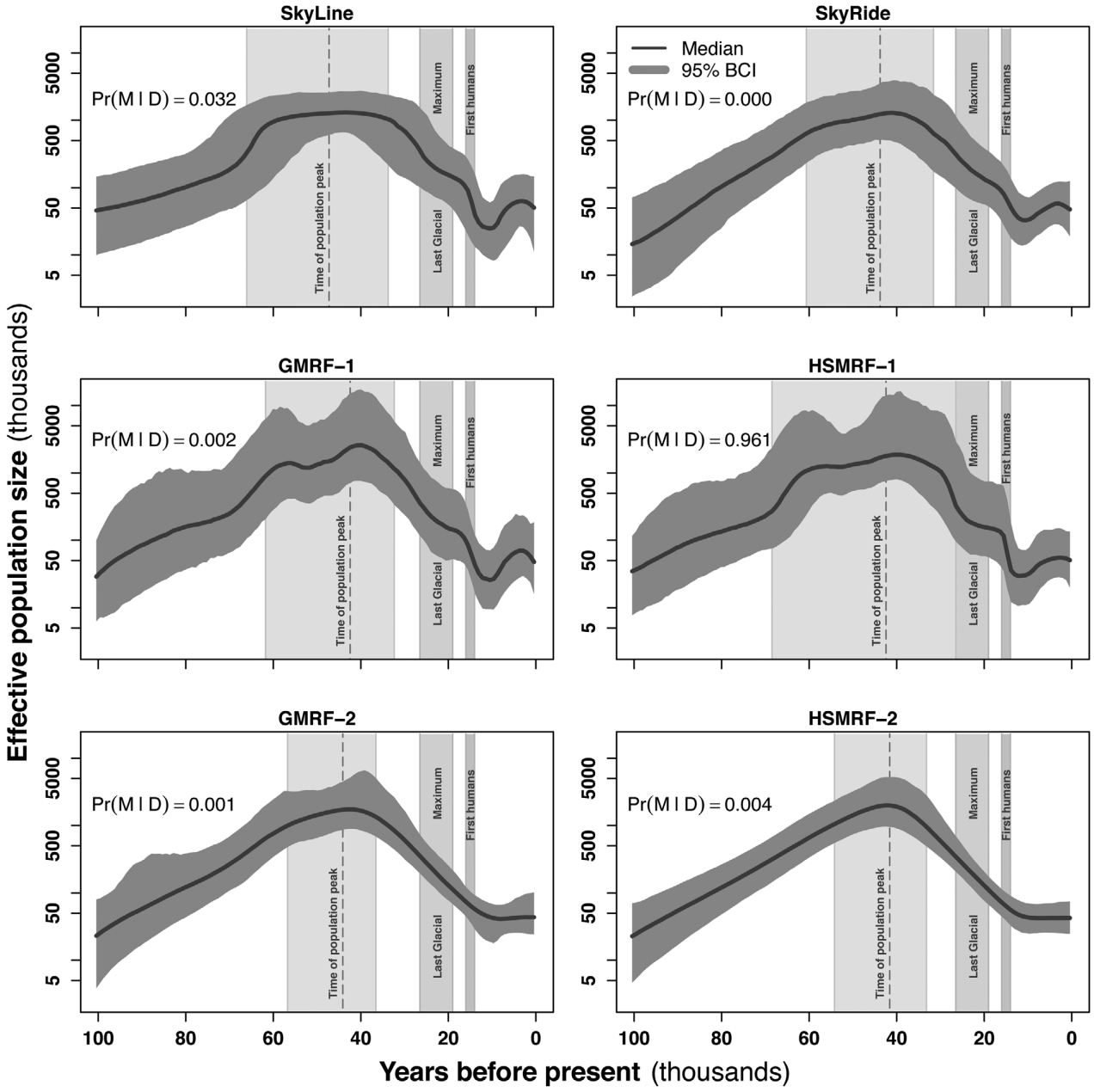

FIGURE 5 Posterior medians of effective population sizes and associated 95% credible intervals obtained from the bison DNA sequence data using the Bayesian Skyline (SkyLine), Bayesian Skyride (SkyRide), and GMRF and HSMRF models of order 1 and order 2. Also shown for each model are posterior model probabilities (Pr(M | D)) and posterior median and 95% credible intervals for the time of peak effective population size. The period of the Last Glacial Maximum and timing of first human settlement in North America are shown for reference

similar to the GMRF-1 model. The second-order models both produced strongly piecewise-linear trajectories with relatively narrow credible intervals, but had low posterior probability and smoothed over some of the local features displayed by other models. The HSMRF-1 model displayed a more complex descent from the peak size to the present in comparison to the other models, and the areas of rapid descent are coincident with the arrival of humans in eastern Beringia and ice-free North America and the initial retreat of the glaciers, both of which are coincident with changes in habitat. All models suggested that the overall decline in population size started before the LGM, and all had median time of population peak between 41.6 and 47.3 kya, but uncertainty in the time of peak population size varied widely across the models.

4 | DISCUSSION

We introduced a novel and fully Bayesian method for nonparametric inference of changes in effective population size that we call the HSMRF. This method utilizes a shrinkage prior known as the horseshoe distribution, which allows more flexibility to respond to rapid changes in effective population size trajectories, yet also generates smoother trajectories in comparison to standard GMRF methods. Our simulations demonstrated that the HSMRF had lower bias and higher precision than the GMRF and was able to recover the underlying true trajectories better in most cases.

There are many situations where the local adaptivity of the HSMRF models would provide advantages over the GMRF and other models. In infectious disease dynamics, examples that could lead to rapid changes in effective population sizes include sudden changes in contact rates due to behavioral changes or quarantine, or sudden changes in the infection rate due to introduction of treatment or vaccine. At a macroevolutionary scale, sudden changes in effective population size could be brought on by sudden population collapse (eg, extinction) or rapid expansions due to dispersals or ecological release. As we have demonstrated, in situations like these the GMRF and other models tend to smooth over the sharp changes that the HSMRF can capture.

Our results from both data examples indicated that the properties of the population size trajectories estimated by the HSMRF-1 model were somewhere between those from the GMRF-1 model and the SkyLine model. The SkyLine model is a type of change-point model, which suggests the HSMRF-1 can produce behavior of change-point models without explicitly needing to specify number or location of change points.

We demonstrated in our simulations that second-order models for either the HSMRF or GMRF formulations can perform better than first-order models in many cases. Although the second-order models did not perform as well as the first-order models in our particular data examples, they would likely do well in other examples with smoother trajectories. Among the second-order models, the HSMRF did as well or better than the GMRF for the simulated examples and had higher posterior model probabilities for both of the data examples.

Second-order models have not been used much for estimating effective population sizes previously. Palacios and Minin (2013), whose method assumes a fixed and known genealogy, tested an integrated Brownian motion (IBM) prior for their GP model for the purpose of testing prior sensitivity but did not use the prior beyond that. The IBM prior is equivalent to the second-order GMRF in continuous time. Our use of second-order GMRF model for jointly estimating genealogy and effective population size trajectory is the first we are aware of in the literature. The second-order GMRF and HSMRF can have similar performance in many cases, but HSMRF has the advantage of added flexibility when needed, so it is a reasonable default choice over the GMRF. We suggest that researchers fit both orders and use a metric such as Bayes factors to select the best order of model for the data.

ACKNOWLEDGMENTS

J.R.F. and V.N.M. were supported by the NIH Grant U54 GM111274. V.N.M. was supported by the NIH Grant R01 AI107034. A.F.M. was supported by ARCS Foundation Fellowship. B.S. was supported by the NSF Grant DEB-1754461. We thank the associate editor and reviewers for their constructive criticism and helpful suggestions.

ORCID

James R. Faulkner https://orcid.org/0000-0003-4478-5113
Andrew F. Magee https://orcid.org/0000-0002-7403-5455
Beth Shapiro https://orcid.org/0000-0002-2733-7776
Vladimir N. Minin https://orcid.org/0000-0002-1917-9288

REFERENCES

Alizon, S., Lion, S., Murall, C.L. and Abbate, J.L. (2014) Quantifying the epidemic spread of Ebola virus (EBOV) in Sierra Leone using phylodynamics. *Virulence*, 5, 825–827.

Bourgeon, L., Burke, A. and Higham, T. (2017) Earliest human presence in North America dated to the last glacial maximum: new radiocarbon dates from Bluefish Caves, Canada. *PloS ONE*, 12, e0169486.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P. and Riddell, A. (2016) Stan: a probabilistic programming language. *Journal of Statistical Software*, 20, 1–37.

Carvalho, C.M., Polson, N.G. and Scott, J.G. (2010) The horseshoe estimator for sparse signals. *Biometrika*, 97, 465–480.

Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W. and McCabe, A.M. (2009) The last glacial maximum. *Science*, 325, 710–714.

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.

Easton, N.A., Mackay, G.R., Young, P.B., Schnurr, P. and Yesner, D.R. (2011) Chindadn in Canada? Emergent evidence of the Pleistocene transition in southeast Beringia as revealed by the Little John Site, Yukon. In: Goebel, T. and Buvit, I (Eds) *From the Yenisei to the Yukon: Interpreting Lithic Assemblage Variability in Late Pleistocene/Early Holocene Beringia*. College Station, TX: Texas A&M University Press, pp. 289–307.

Faulkner, J.R. and Minin, V.N. (2018) Locally adaptive smoothing with Markov random fields and shrinkage priors. *Bayesian Analysis*, 13, 225–252.

Felsenstein, J. and Rodrigo, A.G. (1999) Coalescent approaches to HIV population genetics. In: Crandall, K.A. (Ed.), *The Evolution of HIV*. Baltimore, MD: Johns Hopkins University Press, pp. 233–272.

Frank, C., Mohamed, M.K., Strickland, G.T., Lavanchy, D., Arthur, R.R., Magder, L.S., El Khoby, T., Abdel-Wahab, Y., Anwar, W. and Sallam, I. (2000) The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *The Lancet*, 355, 887–891.

Froese, D., Stiller, M., Heintzman, P.D., Reyes, A.V., Zazula, G.D., Soares, A.E., Meyer, M., Hall, E., Jensen, B.J., Arnold, L.J., MacPhee, R.D.E. and Shapiro, B. (2017) Fossil and genomic evidence constrains the timing of bison arrival in North America. *Proceedings of the National Academy of Sciences*, 114, 3457–3462.

Fuentes, M. (2002) Spectral methods for nonstationary spatial processes. *Biometrika*, 89, 197–210.

Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1, 515–534.

Gill, M.S., Lemey, P., Faria, N.R., Rambaut, A., Shapiro, B. and Suchard, M.A. (2013) Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Molecular Biology and Evolution*, 30, 713–724.

Griffiths, R.C. and Tavaré, S. (1994) Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 344, 403–410.

Heintzman, P.D., Froese, D., Ives, J.W., Soares, A.E., Zazula, G.D., Letts, B., Andrews, T.D., Driver, J.C., Hall, E., Hare, P.G., Jass, C.N., Mackay, G., Southon, P.R., Stiller, M., Woywitka, R., Suchard, M.A. and Shapiro, B. (2016) Bison phylogeography constrains dispersal and viability of the Ice Free Corridor in western Canada. *Proceedings of the National Academy of Sciences*, 113, 8057–8063.

Higdon, D. (1998) A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*, 5, 173–190.

Höhna, S., Landis, M.J., Heath, T.A., Boussau, B., Lartillot, N., Moore, B.R., Huelsenbeck, J.P. and Ronquist, F. (2016) RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, 65, 726–736.

Holmes, C.E. (2011) The Beringian and Transitional periods in Alaska. In: Goebel, T. and Buvit, I. (Eds) *From the Yenisei to the Yukon: Interpreting Lithic Assemblage Variability in Late Pleistocene/Early Holocene Beringia*. College Station, TX: Texas A&M University Press, pp. 179–191.

Kingman, J.F.C. (1982) The coalescent. *Stochastic Processes and Their Applications*, 13, 235–248.

Lan, S., Palacios, J.A., Karcher, M., Minin, V.N. and Shahbaba, B. (2015) An efficient Bayesian inference framework for coalescent-based nonparametric phylodynamics. *Bioinformatics*, 31, 3282–3289.

Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S. *et al.* (2016) Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*, 2, e1501385.

Lorenzen, E.D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K.A., Ugan, A. *et al.* (2011) Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, 479, 359–364.

Makalic, E. and Schmidt, D.F. (2016) A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23, 179–182.

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*, 107, 14757–14762.

Minin, V.N., Bloomquist, E.W. and Suchard, M.A. (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, 25, 1459–1471.

Murray, I., Adams, R.P. and Mackay, D. (2010) Elliptical slice sampling. *Journal of Machine Learning Research*, 9, 541–548.

Neal, R. (2011) MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2, 113–162.

Opgen-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, 6.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L. and Hofreiter, M. (2004) Genetic analyses from ancient DNA. *Annual Review of Genetics*, 38, 645–679.

Paciorek, C.J. and Schervish, M.J. (2006) Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17, 483–506.

Palacios, J.A. and Minin, V.N. (2012) Integrated nested Laplace approximation for Bayesian nonparametric phylodynamics. *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*. AUAI Press, pp. 726–735.

Palacios, J.A. and Minin, V.N. (2013) Gaussian process-based Bayesian nonparametric inference of population size trajectories from gene genealogies. *Biometrics*, 69, 8–18.

Polson, N.G. and Scott, J.G. (2012) On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, 7, 887–902.

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.

Pybus, O.G., Rambaut, A. and Harvey, P.H. (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*, 155, 1429–1437.

R Core Team. (2017) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raftery, A.E., Newton, M.A., Satagopan, J.M. and Krivitsky, P.N. (2006) Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. Technical Report 499, University of Washington.

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.

Rasmussen, D.A., Volz, E.M. and Koelle, K. (2014) Phylodynamic inference for structured epidemiological models. *PLoS Computational Biology*, 10, e1003570.

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.

Reimer, P.J., Baillie, M.G., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Ramsey, C.B. *et al.* (2009) IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon*, 51, 1111–1150.

Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, FL: CRC Press.

Shapiro, B., Drummond, A.J., Rambaut, A., Wilson, M.C., Matheus, P.E., Sher, A.V., Pybus, O.G. *et al.* (2004) Rise and fall of the Beringian steppe bison. *Science*, 306, 1561–1565.

Shapiro, B., Ho, S.Y., Drummond, A.J., Suchard, M.A., Pybus, O.G. and Rambaut, A. (2010) A Bayesian phylogenetic method to estimate unknown sequence ages. *Molecular Biology and Evolution*, 28, 879–887.

Shapiro, B. and Hofreiter, M. (2014) A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. *Science*, 343, 1236573.

Stan Development Team. (2017) RStan: the R interface to Stan, Version 2.14.2.

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.

Vehtari, A., Gelman, A. and Gabry, J. (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27, 1413–1432.

Wallin, J. and Bolin, D. (2015) Geostatistical modelling using non-Gaussian Matérn fields. *Scandinavian Journal of Statistics*, 42, 872–890.

Watanabe, S. (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.

Xie, W., Lewis, P.O., Fan, Y., Kuo, L. and Chen, M.-H. (2011) Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, 60, 150–160.

Yang, Z. (2014) *Molecular Evolution: A Statistical Approach*. Oxford: Oxford University Press.

Yue, Y.R., Simpson, D., Lindgren, F. and Rue, H. (2014) Bayesian adaptive smoothing splines using stochastic differential equations. *Bayesian Analysis*, 9, 397–424.

Yue, Y.R., Speckman, P.L. and Sun, D. (2012) Priors for Bayesian adaptive spline smoothing. *Annals of the Institute of Statistical Mathematics*, 64, 577–613.

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 2.2, 2.4, 3.1, 3.2, and 3.3 are available with this paper at the Biometrics website on Wiley Online Library. Our `R` package titled `spmrf` can be used to fit our models to fixed genealogical trees and is available at `https://github.com/jrfaulkner/spmrf`. The data and `RevBayes` code for fitting our models to the molecular sequence data described in Sections 3.2 and 3.3 are available at `https://github.com/jrfaulkner/phylocode`.

How to cite this article: Faulkner JR, Magee AF, Shapiro B, Minin VN. Horseshoe-based Bayesian nonparametric estimation of effective population size trajectories. *Biometrics*. 2020;76:677–690. https://doi.org/10.1111/biom.13276