

Locally-adaptive Bayesian nonparametric inference for phylodynamics

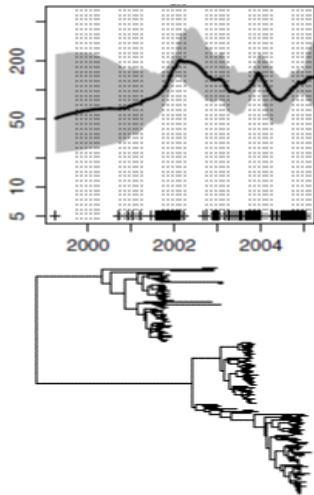
James Faulkner Vladimir Minin

University of Washington

JSM 2016

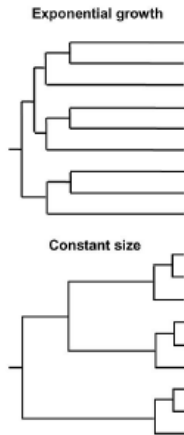
Motivation

- Suppose we have some unknown effective population size trajectory $N_e(t)$ we want to estimate.
- Have phylogenetic data from gene sequences that give estimates of coalescent times, which can be used to estimate effective population size.
- Want method to estimate population size that can adapt to local sharp features and still smooth over noisy regions



Effective population size

- Gene genealogies provide information about population size
- Larger population size results in longer times until common ancestor (coalescent time)
- Nonparametric methods useful when little prior knowledge of parametric form for dynamics



(Grenfell 2004)

From Karcher et al. (2015)

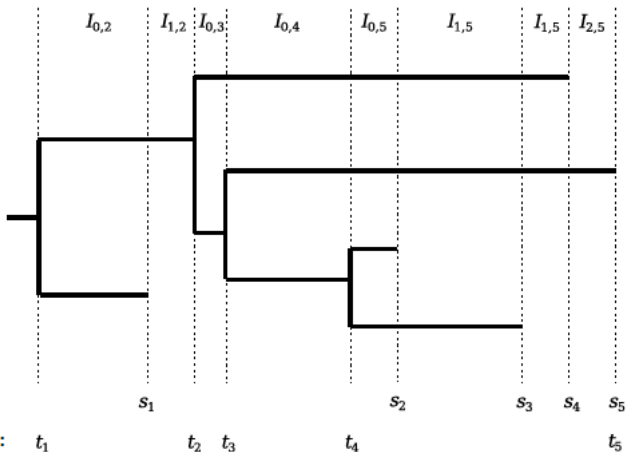


Figure 1: Illustration of an example heterochronous genealogy with $n = 5$ lineages. Sampling times s_1, \dots, s_5 and coalescent times t_1, \dots, t_5 are marked below the genealogy.

- Likelihood

$$\Pr[\mathbf{t}|N_e(\mathbf{t}), \mathbf{s}] \propto \prod_{k=2}^n \frac{C_{0,k}}{N_e(t_{k-1})} \exp \left[- \sum_{i=0}^{m_k} \int_{l_{i,k}} \frac{C_k}{N_e(t)} dt \right]$$

Current nonparametric methods

- Piecewise constant – "Skyline" (Pybus et al 2000)
- Gaussian Markov random field – "Skyride" (Minin et al 2008)
- Gaussian process (Palacios and Minin 2013)
- GMRF with preferential sampling (Karcher et al 2016)

GMRF smoothing prior

Assume effective population size follows an unknown function $N_e(t)$, where t is a continuous index of time.

Let $N_e(j)$ be effective population size at discrete time $j \in \{1, \dots, m\}$, and let $\theta_j = \ln[N_e(j)]$

GMRF smoothing prior

Assume effective population size follows an unknown function $N_e(t)$, where t is a continuous index of time.

Let $N_e(j)$ be effective population size at discrete time $j \in \{1, \dots, m\}$, and let $\theta_j = \ln[N_e(j)]$

Then a simple k th-order GMRF prior for θ is induced by letting:

$$\Delta^k \theta_i \sim N(0, \gamma^2), \quad i = 1, \dots, m - k$$

where $\Delta^k \theta_i$ is a k th-order forward difference operator.

Adaptive smoothing prior

We can allow locally-adaptive behavior and increase smoothing properties by putting a **shrinkage prior** on $\Delta^k \theta_i$:

$$\Delta^k \theta_i \sim \text{Horseshoe}(0, \gamma)$$

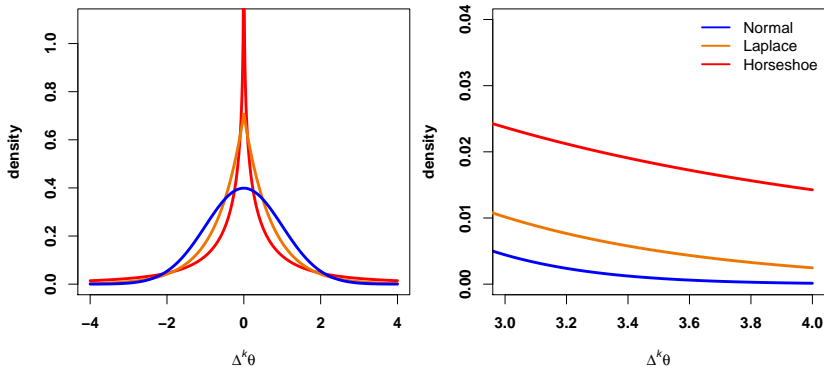
$$\gamma \sim C^+(0, \zeta)$$

where γ is the global smoothing parameter.

The result is non-Gaussian (horseshoe) Markov random field prior for θ .

Prior comparisons

Good shrinkage prior has high density at zero and fat tails



Simulations

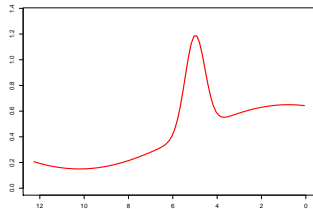
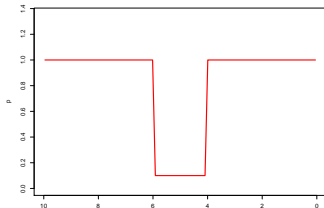
Simulated coalescent data

Scenario 1:

- 500 sample times
- Bottleneck

Scenario 2:

- 2000 sample times
- Spike increase



Simulation results

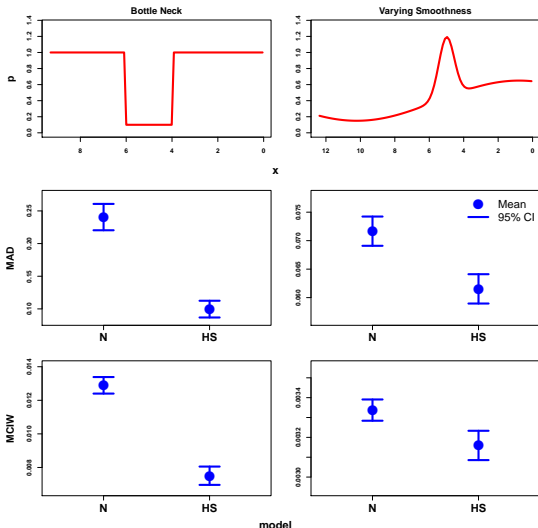
100 Simulations
Hamiltonian Monte Carlo

Models:

- Normal (N)
- Horseshoe (HS)

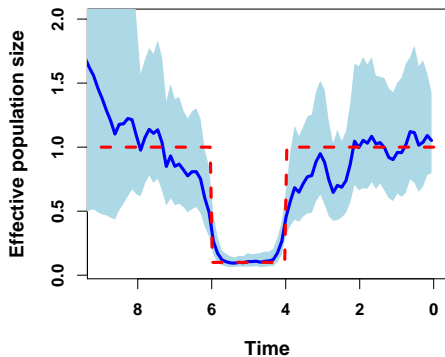
Metrics:

- Mean absolute deviation (MAD)
- Mean credible interval width (MCIW)

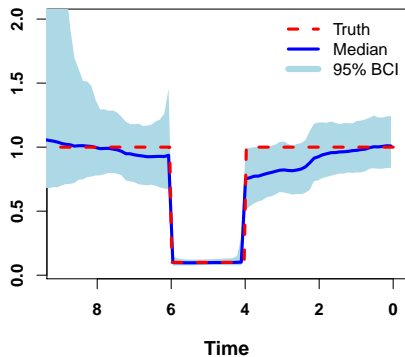


Example results - Bottleneck

Normal

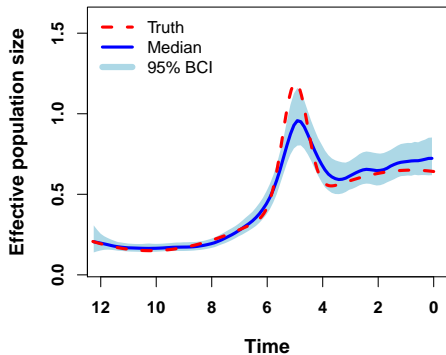


Horseshoe

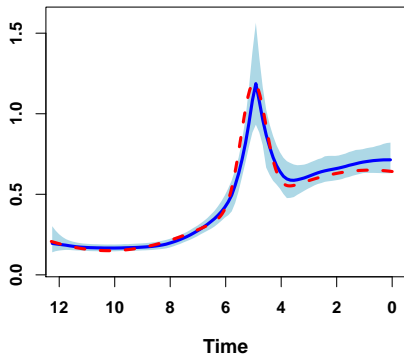


Example results - Spike increase

Normal



Horseshoe

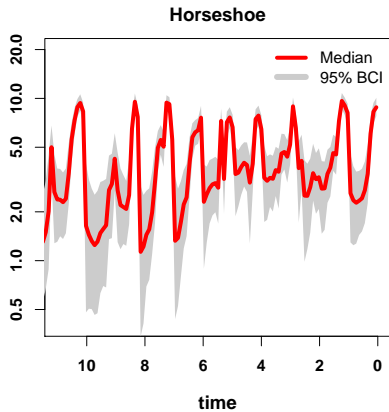
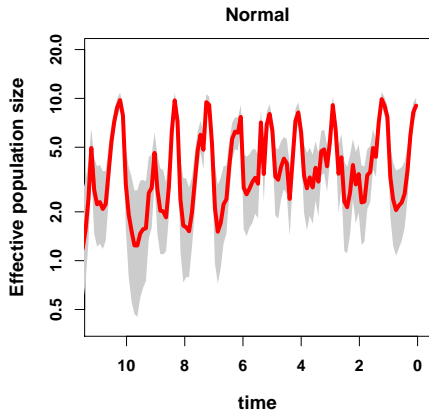


Data Example: Regional Influenza

- H3N2 data collected monthly across several regions of the world over 10-12 years (Zinder et al. 2014).
- Regional seasonal variation in climate can affect infection rates, with temperate regions expected to have greater fluctuations.
- We look at 3 regions: USA/Canada, Oceania, and South China.
- Models account for preferential sampling.

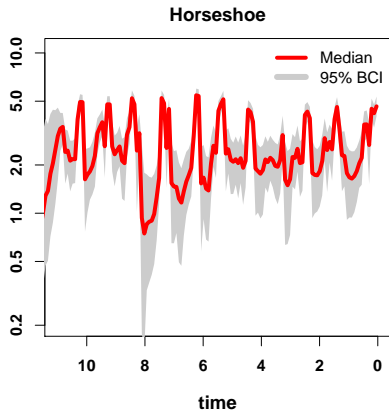
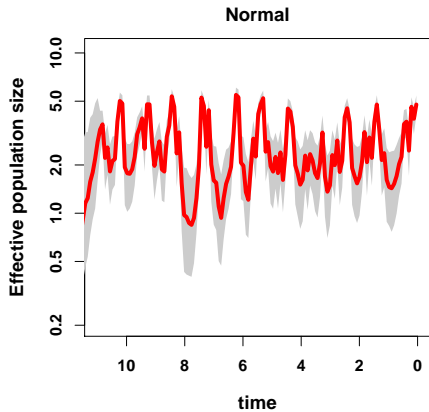
Influenza results

USA / Canada



Influenza results

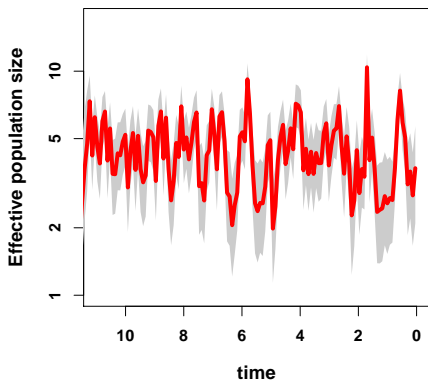
Oceania



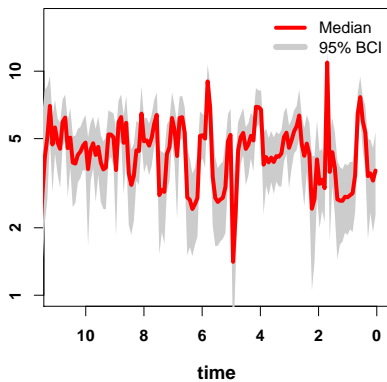
Influenza results

South China

Normal



Horseshoe



Conclusions

- Our adaptive smoothing method captures abrupt jumps or changes in trend better than non-adaptive approaches, as evidenced by improved bias and precision in simulations.
- Can be useful for smoothing over weaker signals while still maintaining strong signals.