



EN²M - 2021

01

Utilizando prioris de complexidade penalizada para estimação de tamanho de uma população a partir de filogenias

CRISTIANA APARECIDA NOGUEIRA COUTO
ORIENTADOR: LUIZ MAX CARVALHO

Tópicos desta Apresentação

Introdução

Justificativa

Métodos

Discussões

Considerações finais

Referências bibliográficas

Neste trabalho propõe-se um estudo de modelagem estatística bayesiana analisando a escolha de prioris para estimação de um fenômeno biológico.

1. Introdução

Sequência de apresentação



ENTENDENDO O
O FENÔMENO
BIOLÓGICO A
SER
INVESTIGADO

E porque deseja-se estudá-lo

MODELOS
MATEMÁTICOS
EXISTENTES

Quais as suas características?
Quais os parâmetros? Entre
outros ...

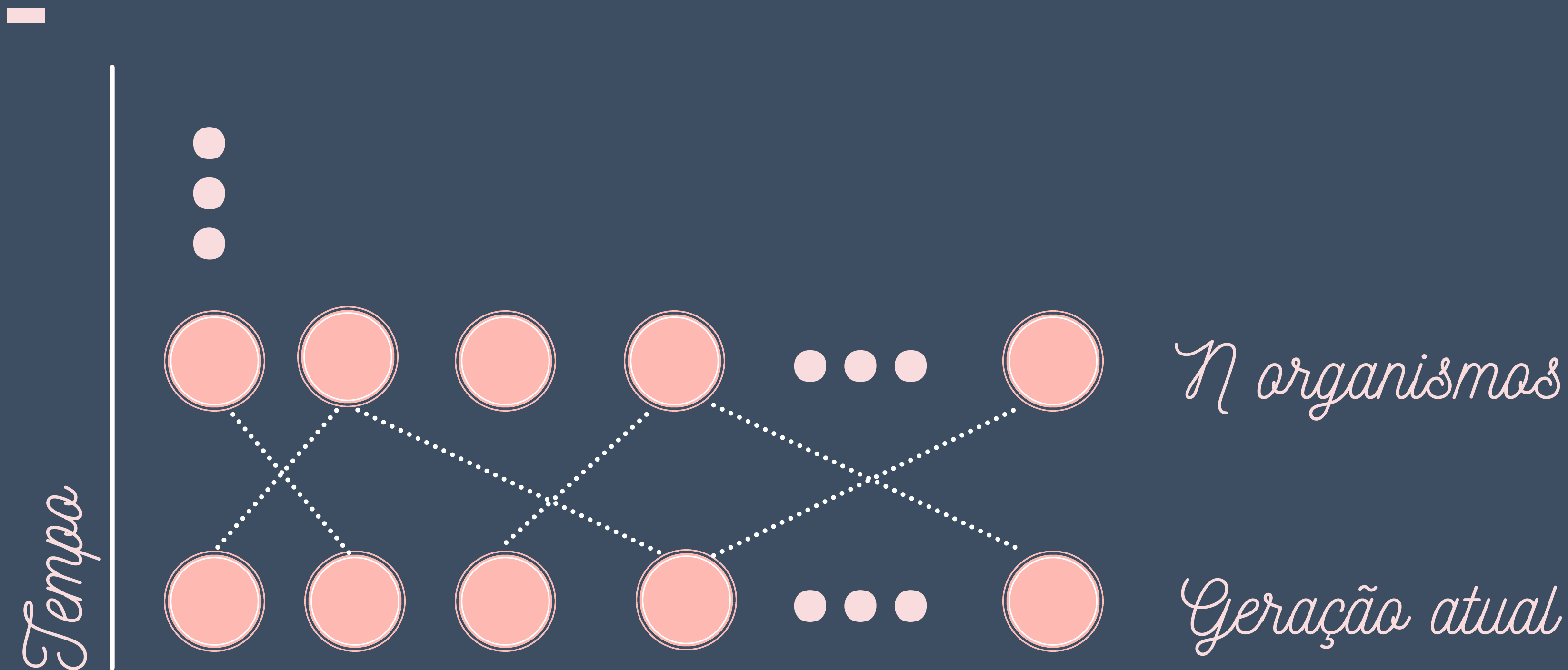
ANÁLISES
PROPOSTAS

Quais as modificações
propostas para o modelo e
qual a razão de fazê-las



O processo genealógico

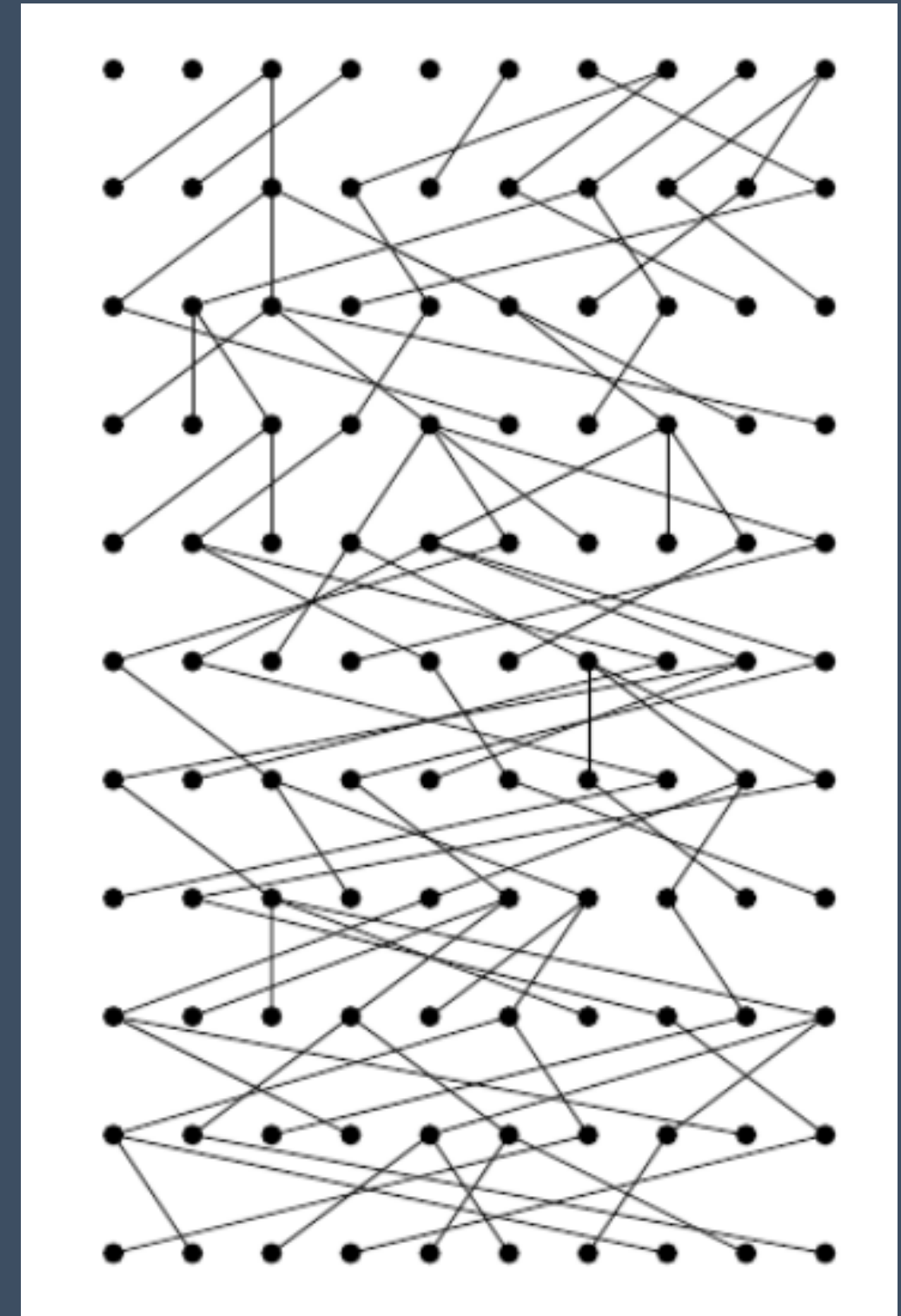
05



O processo genealógico

—

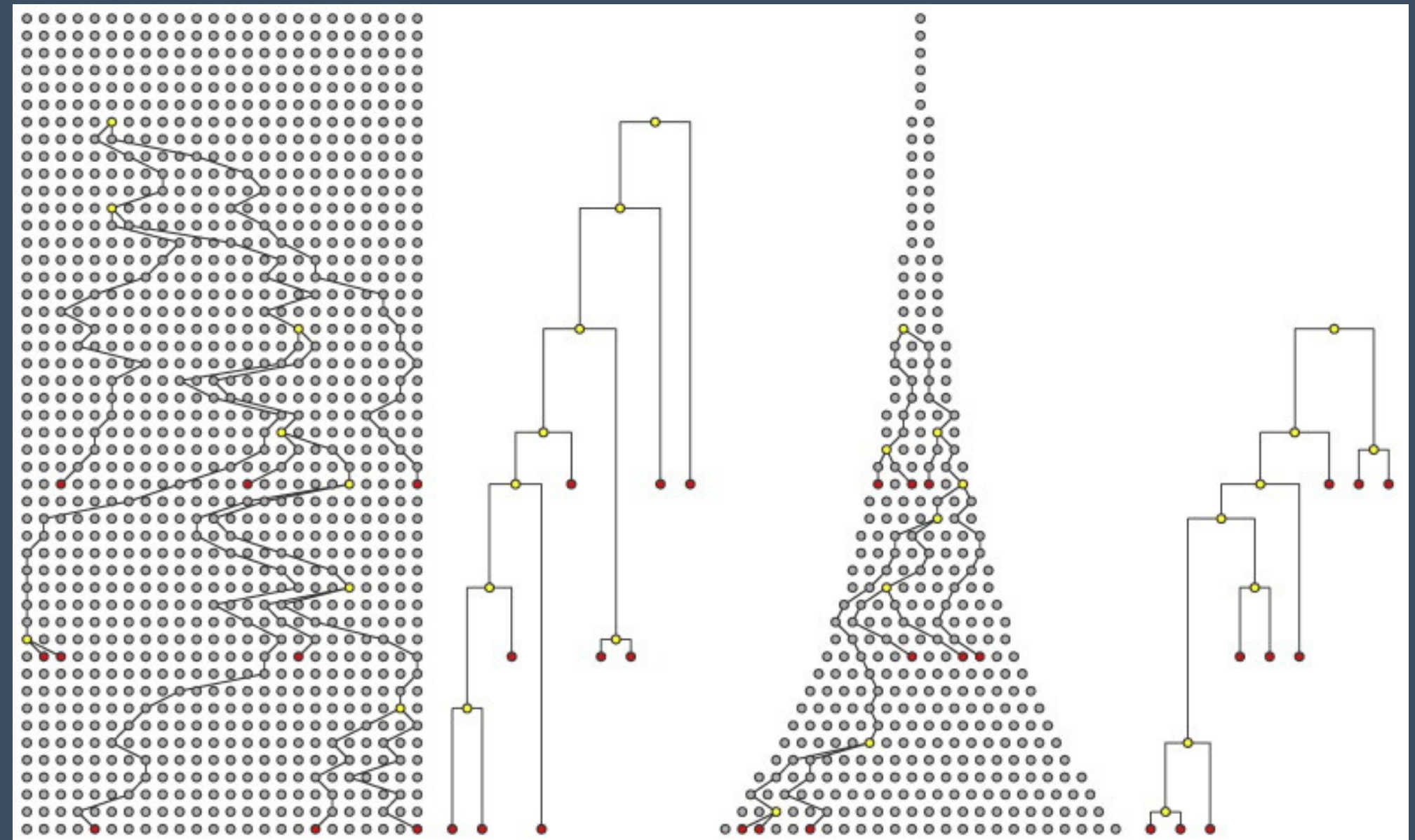
- Modelo Wright-Fisher de evolução;



FONTE: (NORDBORG, 2004)

Amostra da população e árvore filogenética

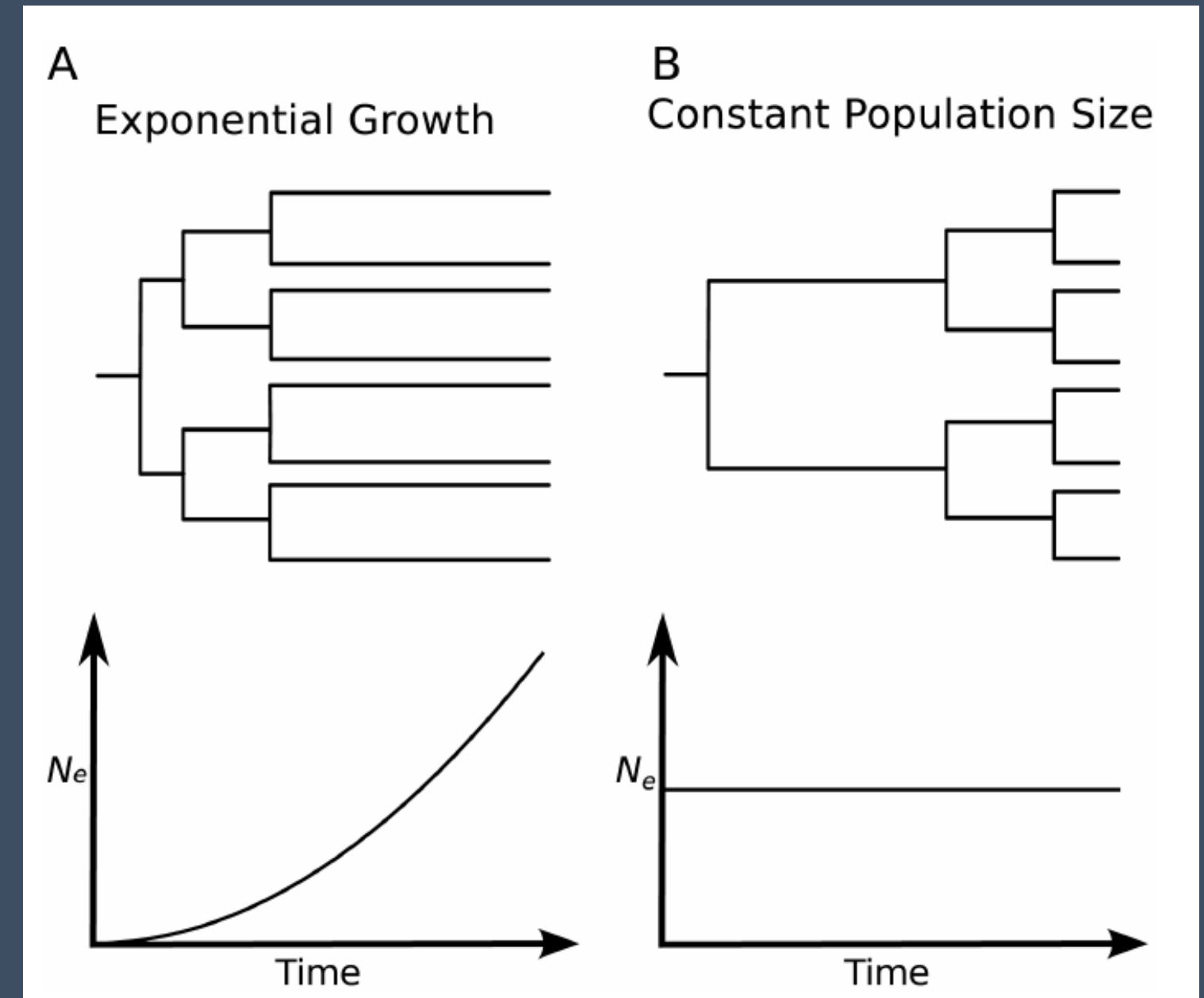
- Fazer inferências sobre a população total a partir de amostras da população;



Crescimento da população

- Figura idealizada de filogenias de vírus que mostram os efeitos das mudanças no tamanho da população viral.

2



Revisão de literatura

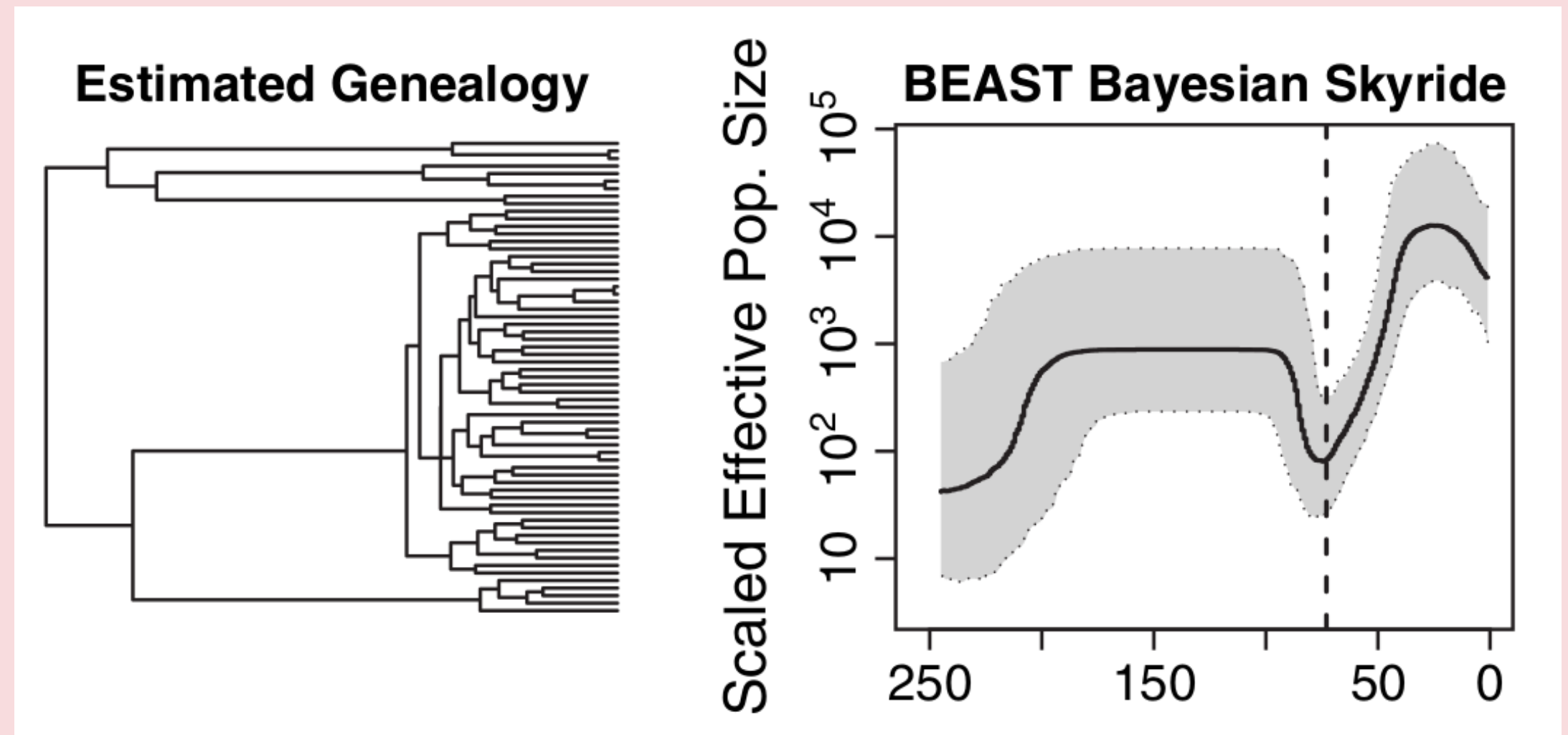


Conjunto de dados:

63 sequências de HCV, vírus causador da hepatite C, amostrados em 1993 no Egito;

Descrição da objetiva:

Investigar a dinâmica populacional do HCV no Egito (MININ; BLOOMQUIST; SUCHARD, 2008)



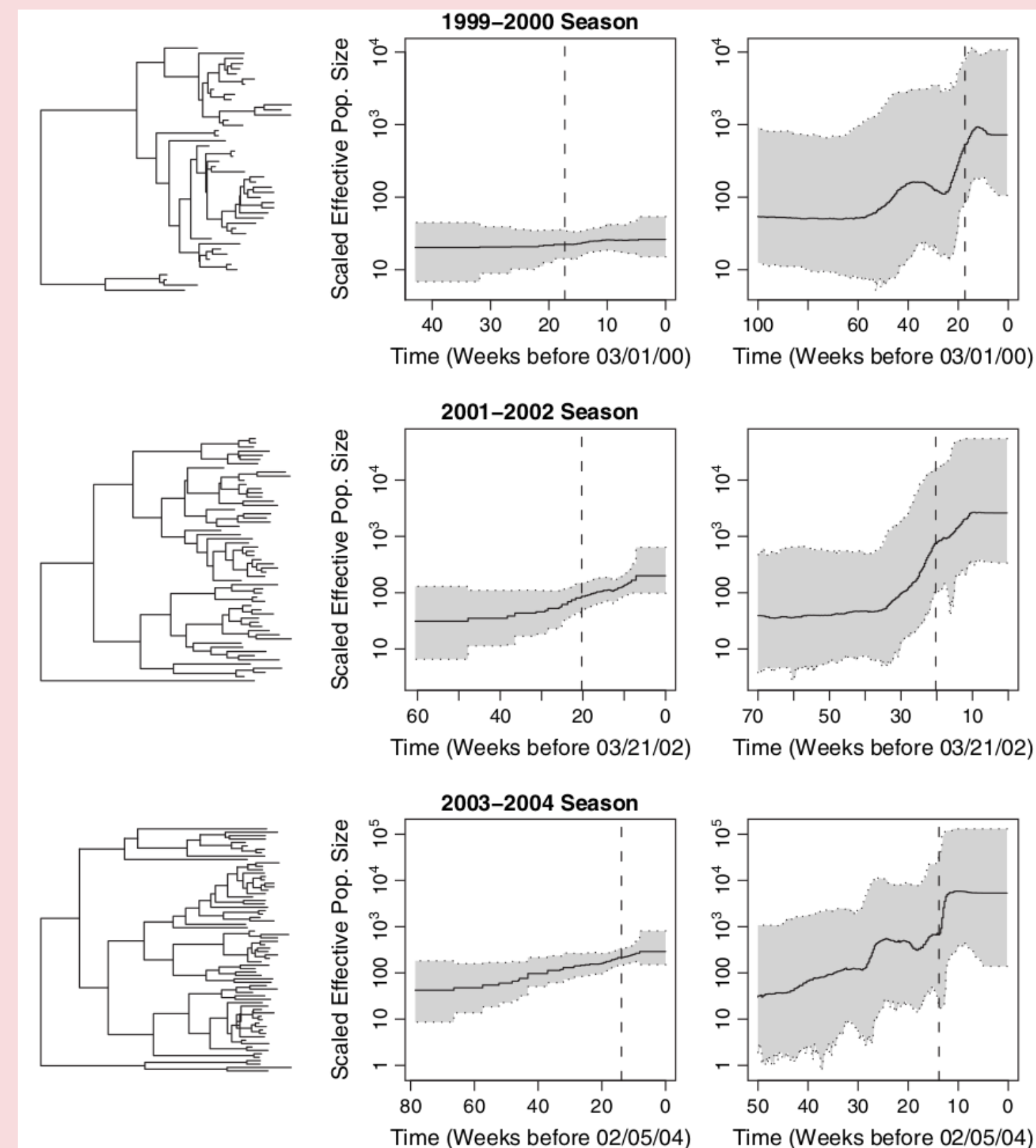
Revisão de literatura

Conjunto de dados:

Três conjuntos de dados correspondendo a três épocas gripais, período recorrente anual caracterizado pelos surtos de influenza;

Descrição da objetiva:

Investigar a dinâmica populacional intra-sazonal da influenza (MININ; BLOOMQUIST; SUCHARD, 2008)



Revisão de literatura

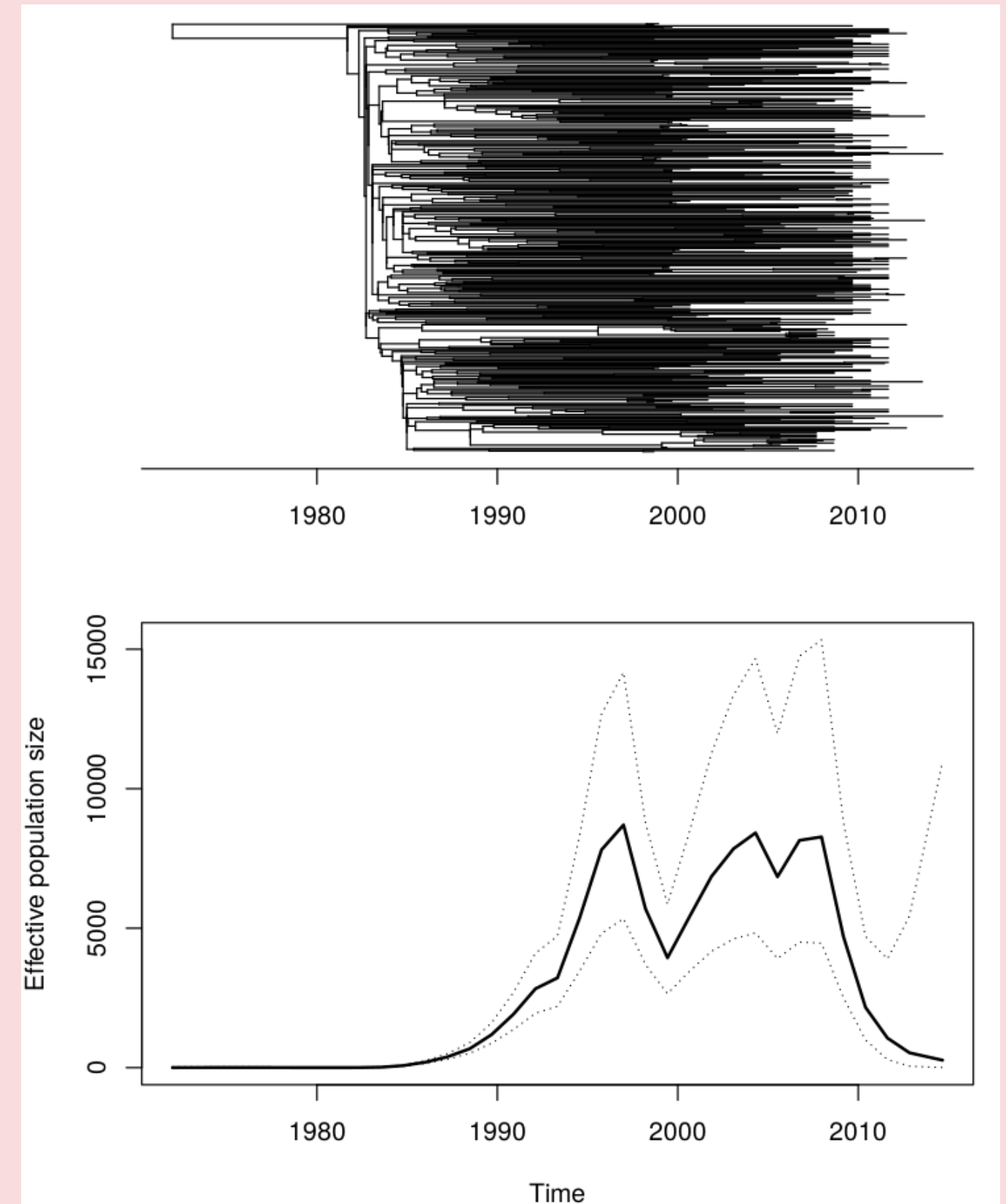


Conjunto de dados:

399 sequências de HIV-1, amostradas no Senegal entre 1990 e 2014. Todas as 179 sequências são do subtipo CRF02_AG;

Descrição da objetivo:

Inferir a função demográfica (DIDELOT; VOLZ, 2021)




Motivação



Num contexto de aplicação epidemiológica, por exemplo, quando aplicada a dados genéticos de patógenos, a inferência do tamanho da população efetiva pode fornecer informações importantes sobre a dinâmica epidemiológica de uma infecção.

2. Modelos

Modelo Demográfico por Partes:

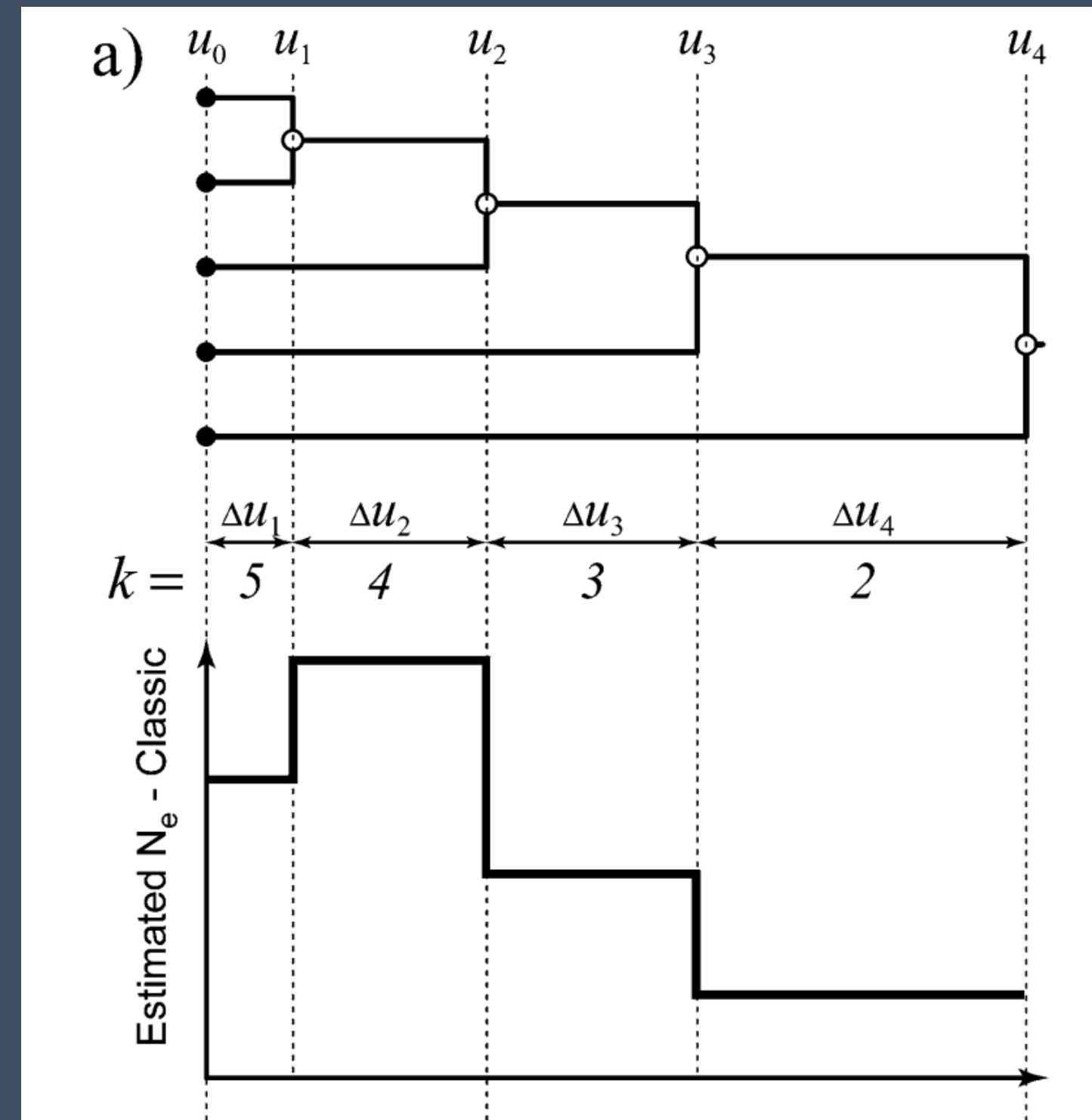


Seja uma árvore filogenética T com n folhas. Seja $s = (s_2, s_3, \dots, s_n)$ os intervalos de inter-coalescência de T , ou seja, intervalos entre os eventos de coalescência.

Seja θ o tamanho da população, onde $\theta = (\theta_2, \dots, \theta_n)$.

Modelo skyline clássico

- A mudança no tamanho efetivo da população coincide com os tempos de coalescência;
- Função escada;



- Supõe que o tamanho da população muda suavemente ao longo do tempo;
- Define um processo GMRF como priori no logaritmo do tamanho efetivo da população com um parâmetro de precisão τ .

Nesse modelo, precisamos inferir o parâmetro de precisão τ que é desconhecido.

Modelo Skyrider

$$Pr(s|\theta) = \prod_{i=2}^n Pr(s_k|\theta_k) \quad (1)$$

Transformação $\gamma_k = \log(\theta_k)$, $k = 2, \dots, n$.

$$Pr(\gamma|\tau) \propto \tau^{\frac{n-2}{2}} \exp\left(\frac{-\tau}{2} \sum_{k=2}^{n-1} \frac{(\gamma_{k+1} - \gamma_k)^2}{2\theta}\right) \quad (2)$$

*controla a suavidade
da trajetória*

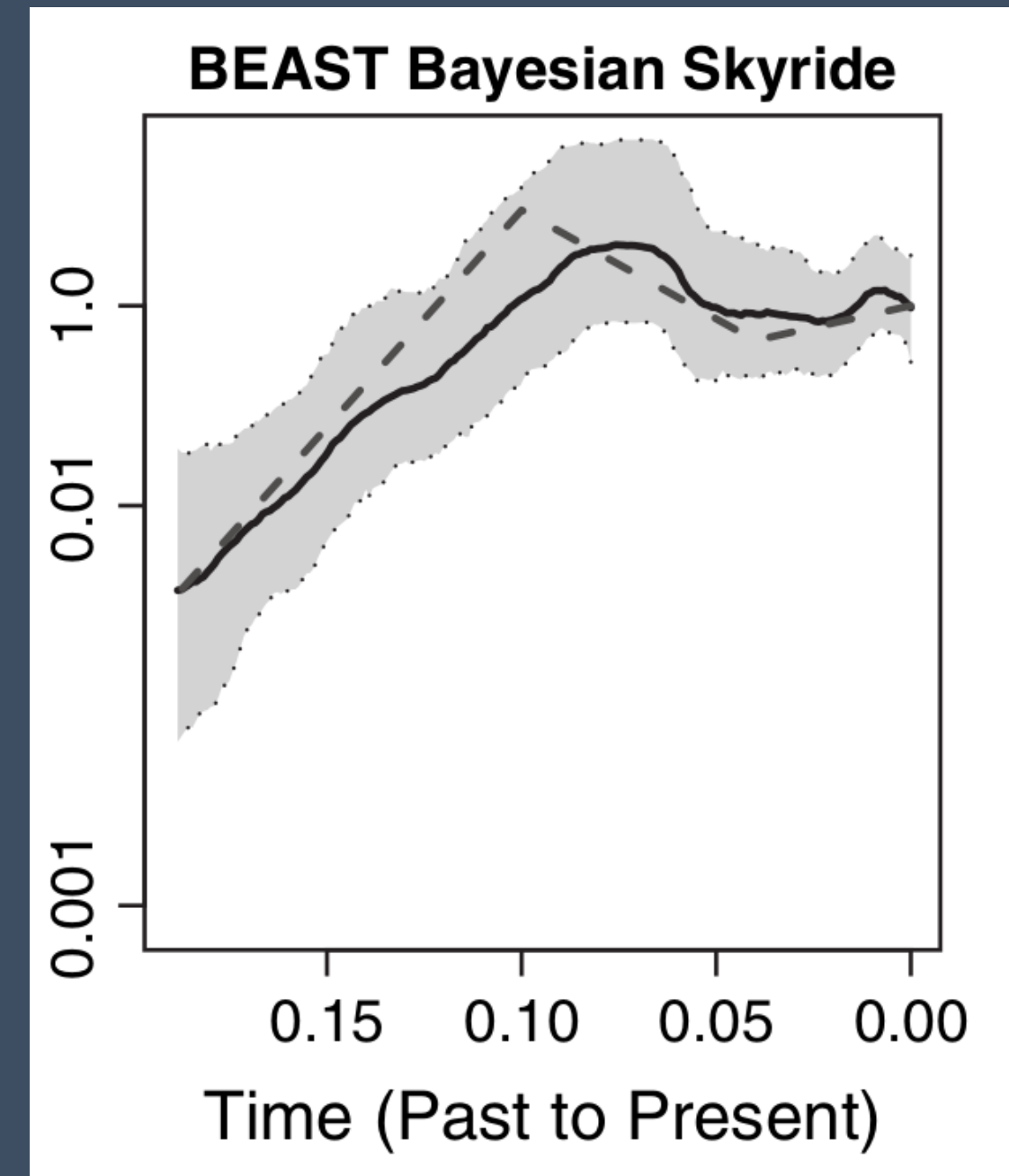


Modelo Skyrider

Estimação do tamanho efetivo de uma população com efeito gargalo

—

Ao lado, há um exemplo de estimação usando o Modelo Skyride;



Objetivo



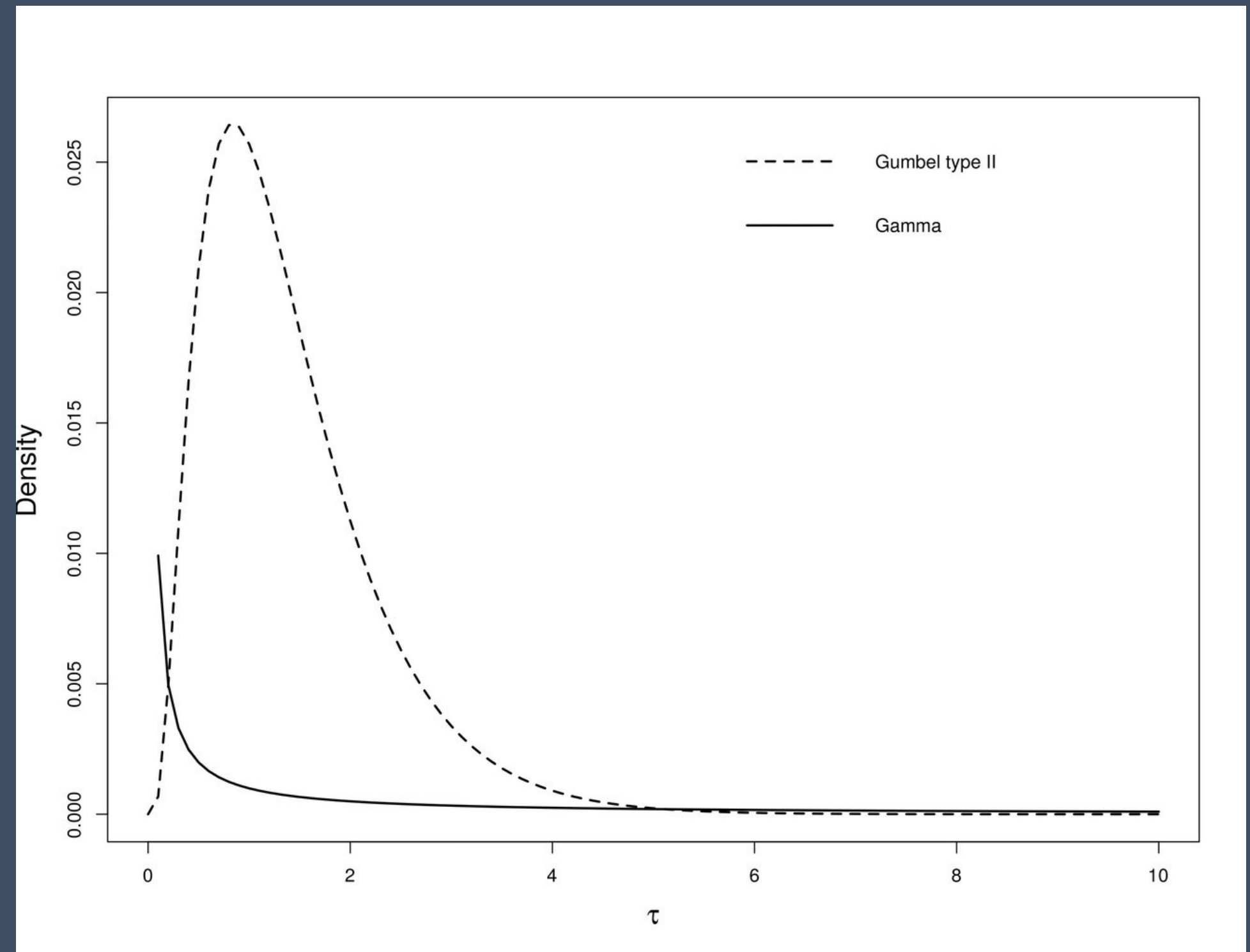
Busca-se entender a influência da escolha da distribuição a priori nas inferências sobre a dinâmica populacional obtidas a partir de árvores filogenéticas.

Em particular, das chamadas prioris de complexidade penalizada (Penalised Complexity Prior).

3. Métodos

Prioris para o parâmetro de precisão τ

- A escolha usual para τ é a distribuição gamma com parâmetros $\alpha = 0.001$ e $\beta = 0.001$,



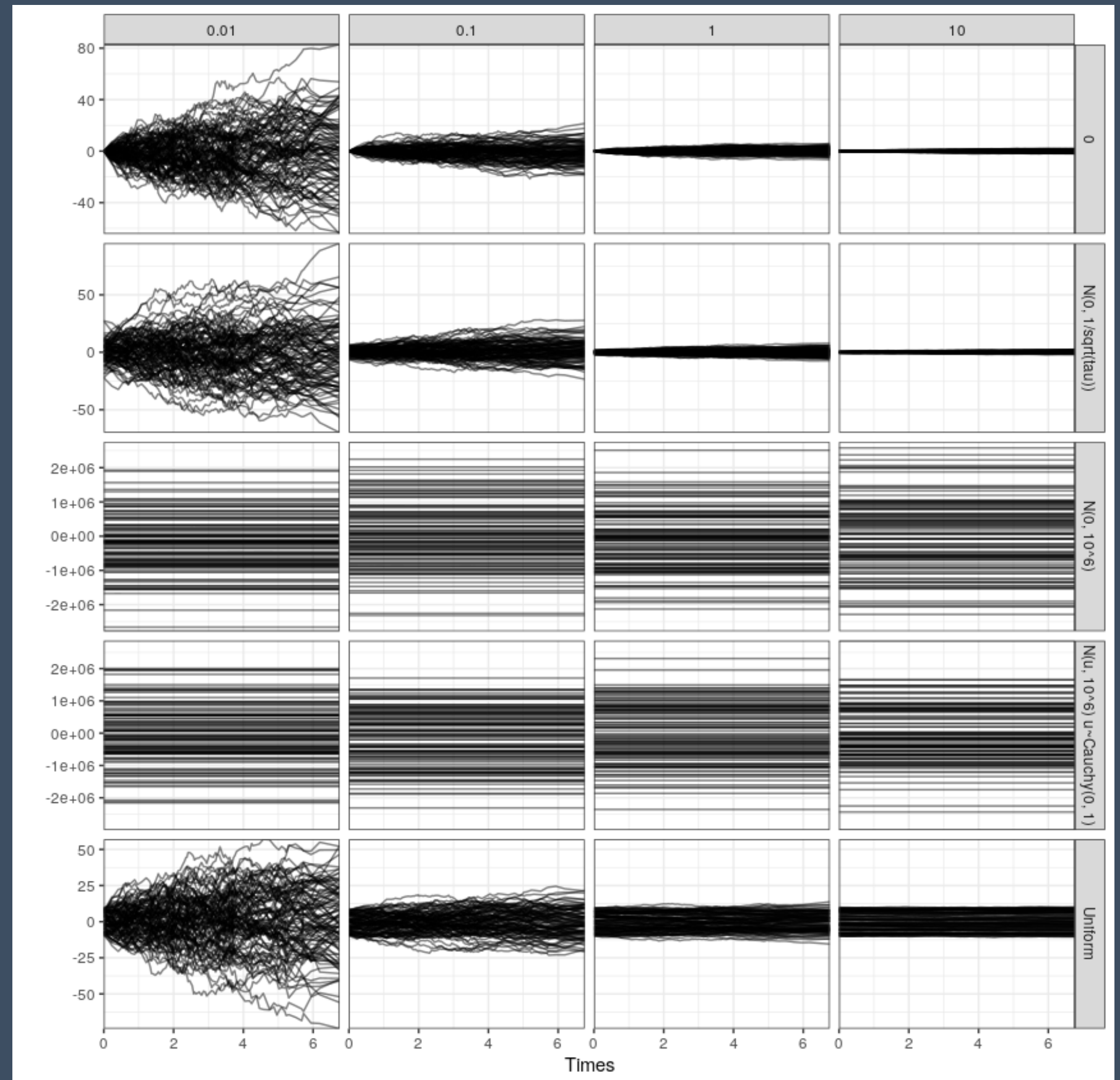
Prioris de Complexidade Penalizada



O trabalho de (SIMPSON et al., 2017) propõe um conceito para a construção de distribuições a priori que são robustas, invariantes à reparametrização e baseada em alguns princípios.

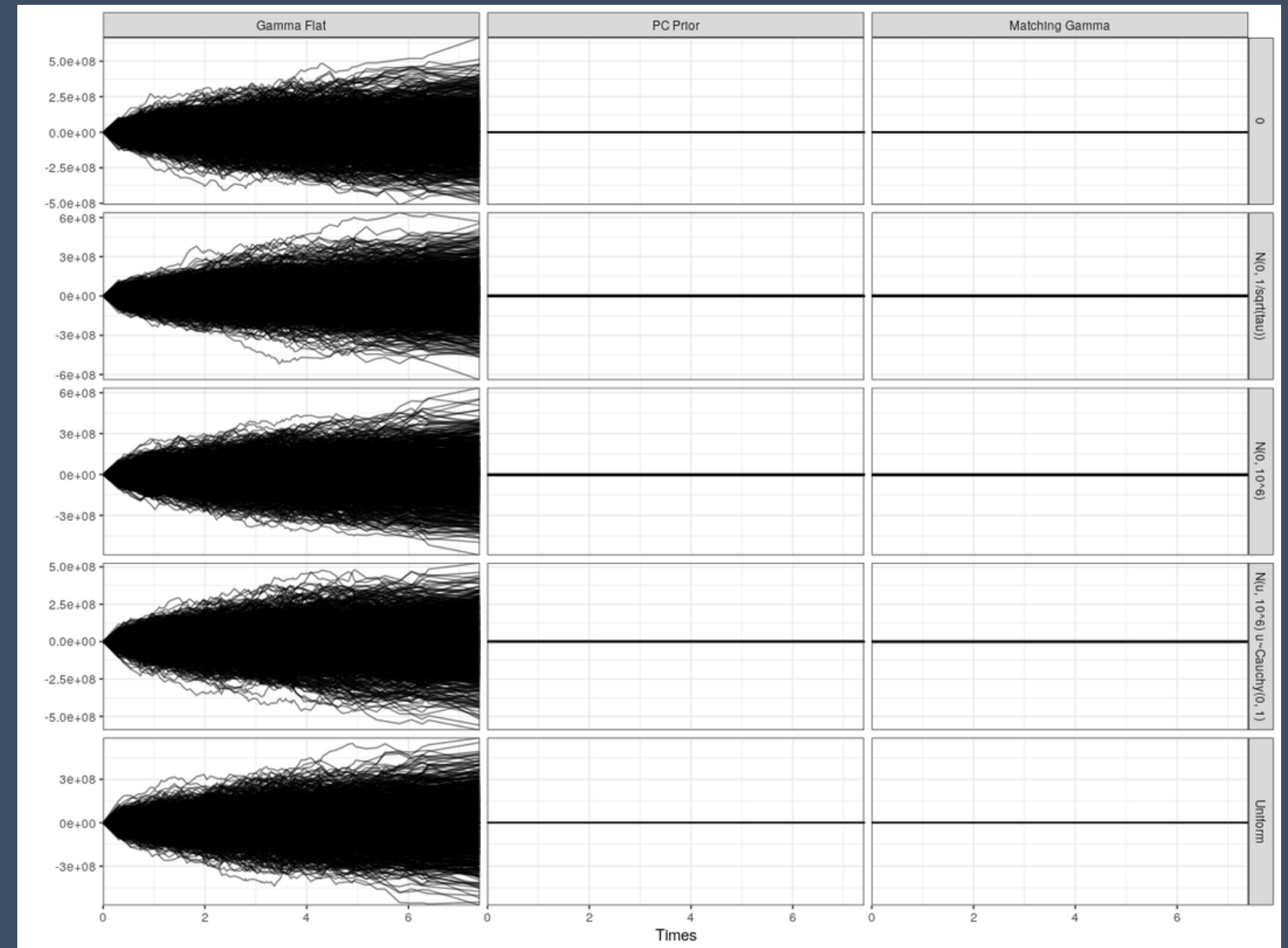
Trajeto rias do GMRF para τ fixo

- A simula  o ao lado foi repetida 100 vezes para 4 valores fixos de τ atribuindo diferentes priors ao valor da trajet ria no tempo $t = 0$;



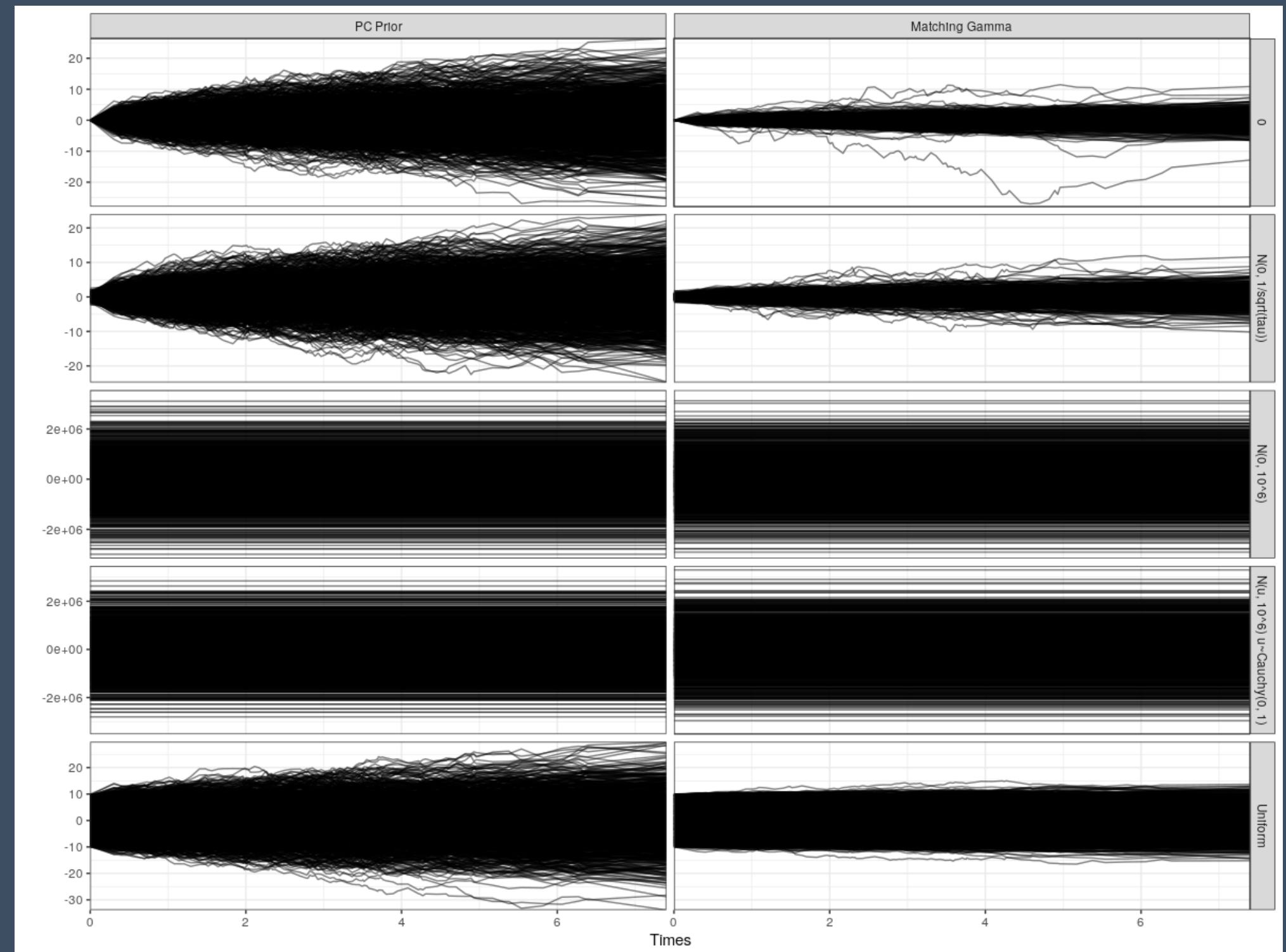
Trajетórias do GMRF para diferentes priors de τ

- Novamente a simulação foi repetida 1000 vezes dessa vez atribuindo à τ e ao valor da trajetória no tempo $t = 0$ diferentes priors;



Trajetórias do GMRF para diferentes prioris de τ

- Novamente a simulação foi repetida 1000 vezes dessa vez atribuindo à τ e ao valor da trajetória no tempo $t = 0$ diferentes prioris;



Experimentos com dados simulados:



- População constante;
- População exponencial;
- População que teve um efeito gargalo, ou seja, um evento que reduz drasticamente o tamanho de uma população.

Os conjuntos de dados estão sendo simulados com o pacote phylodyn do R.



Dados reais



Experimentos com dados reais:

- Dengue do tipo 4;
- Influenza H3N2;
- Outros conjuntos de dados comumente usadas na literatura, como as sequências de HIV amostradas no Egito,



Referências



CARVALHO, L. M. F. de. A better prior for the precision parameter in skyride/grid/track. 2021.



DIDELOT, X.; VOLZ, E. M. Maximum likelihood inference of pathogen population size history from a phylogeny. bioRxiv, Cold Spring Harbor Laboratory, 2021. Disponível em: <<https://www.biorxiv.org/content/early/2021/01/19/2021.01.18.427056>>.



DRUMMOND, A. J. et al. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. Molecular Biology and Evolution, v. 22, n. 5, p. 1185–1192, 02 2005. ISSN 0737-4038. Disponível em: <<https://doi.org/10.1093/molbev/msi103>> .



KÜHNERT, D.; WU, C.-H.; DRUMMOND, A. J. Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. Infection, Genetics and Evolution, Elsevier BV, v. 11, n. 8, p. 1825–1841, dez. 2011. Disponível em: <<https://doi.org/10.1016/j.meegid.2011.08.005>>.

MININ, V. N.; BLOOMQUIST, E. W.; SUCHARD, M. A. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Molecular Biology and Evolution, Oxford University Press (OUP), v. 25, n. 7, p. 1459–1471, abr. 2008. Disponível em: <<https://doi.org/10.1093/molbev/msn090>>.

NORDBORG, M. Coalescent theory. Handbook of statistical genetics, Wiley Online Library, 2004.



Referências



SIMPSON, D. et al. Penalising model component complexity: A principled, practical approach to constructing priors. Statistical Science, Institute of Mathematical Statistics, v. 32, n. 1, fev. 2017. Disponível em: <<https://doi.org/10.1214/16-sts576>>.



VOLZ, E. M.; KOELLE, K.; BEDFORD, T. Viral phylodynamics. PLoS Computational Biology, Public Library of Science (PLOS), v. 9, n. 3, p. e1002947, mar. 2013. Disponível em: <<https://doi.org/10.1371/journal.pcbi.1002947>>.



OBRIGADA PELA ATENÇÃO!



CONTATO:

email:

cristiana.couto@fgv.edu.br

luiz.fagundes@fgv.br



AGRADECIMENTOS:

- Centro para o Desenvolvimento da Matemática e Ciências – CDMC

See Ya!

