HIERARCHICAL GAUSSIAN PROCESSES IN STAN

ROB TRANGUCCI

http://mc-stan.org

GOALS FOR TODAY

- Intro to GPs
- Hierarchical GPs in Stan
- Presidential forecasting model

GAUSSIAN PROCESSES

$$y_i \in \mathbb{R}, x_i \in \mathbb{R}^M \; i \in \{1, 2, \ldots\}$$

$$y_i \sim \mathrm{Normal}\left(f(x_i), \delta\right) \forall i$$

$$y_i \in \mathbb{R}, x_i \in \mathbb{R}^M \; i \in \{1, 2, \ldots\}$$

$$y_i \sim \mathrm{Normal}\left(f(x_i), \delta\right) \forall i$$

$$f(x) \sim \mathrm{GP}\left(\mu(x), K_\theta(x, x)\right)$$

$$y_i \in \mathbb{R}, x_i \in \mathbb{R}^M \; i \in \{1, 2, \ldots, N\}$$

$$y_i \sim \mathrm{Normal}\left(f(x_i), \delta\right) \forall i$$

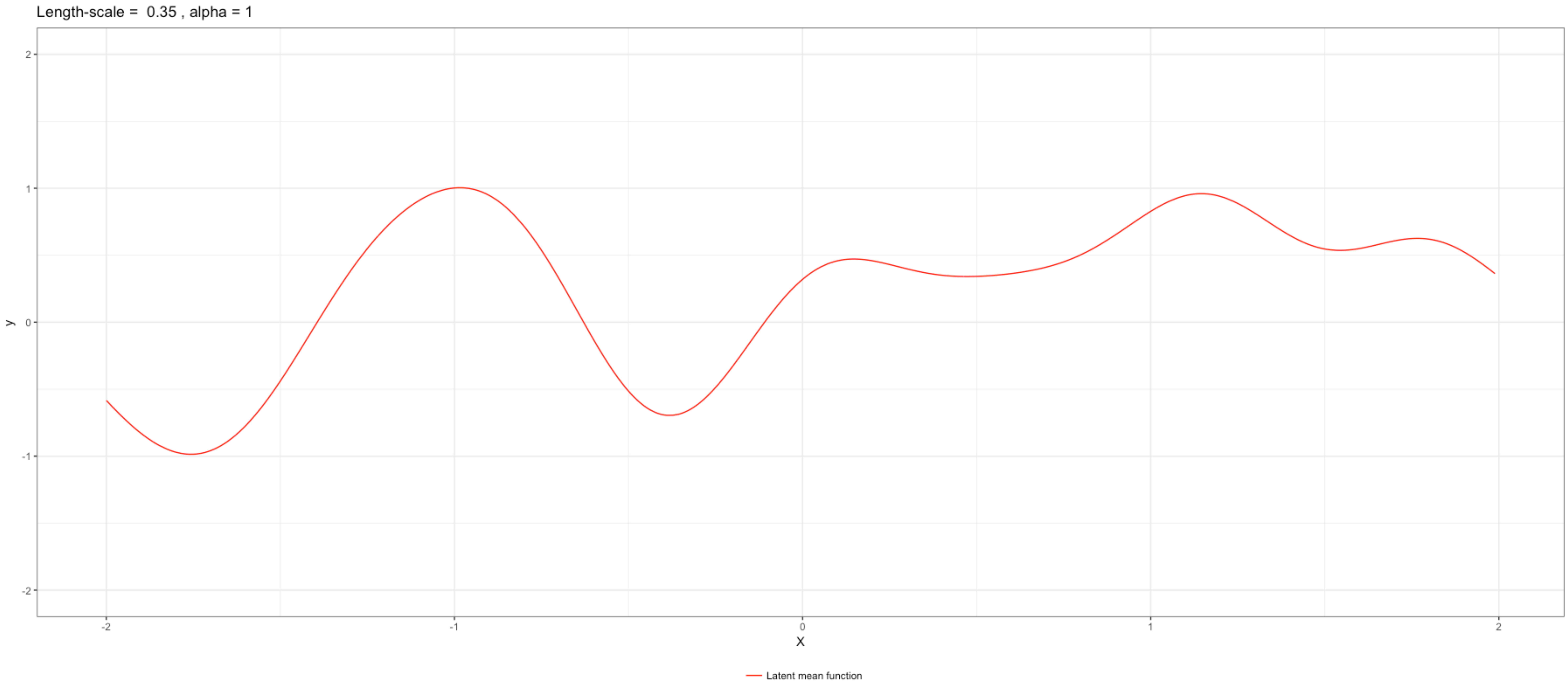
$$f(x) \sim \mathrm{MultiNormal}\left(\mu(x), K_\theta(x, x)\right)$$

$$y_i \in \mathbb{R}, x_i \in \mathbb{R}^M \; i \in 1, \ldots, n$$

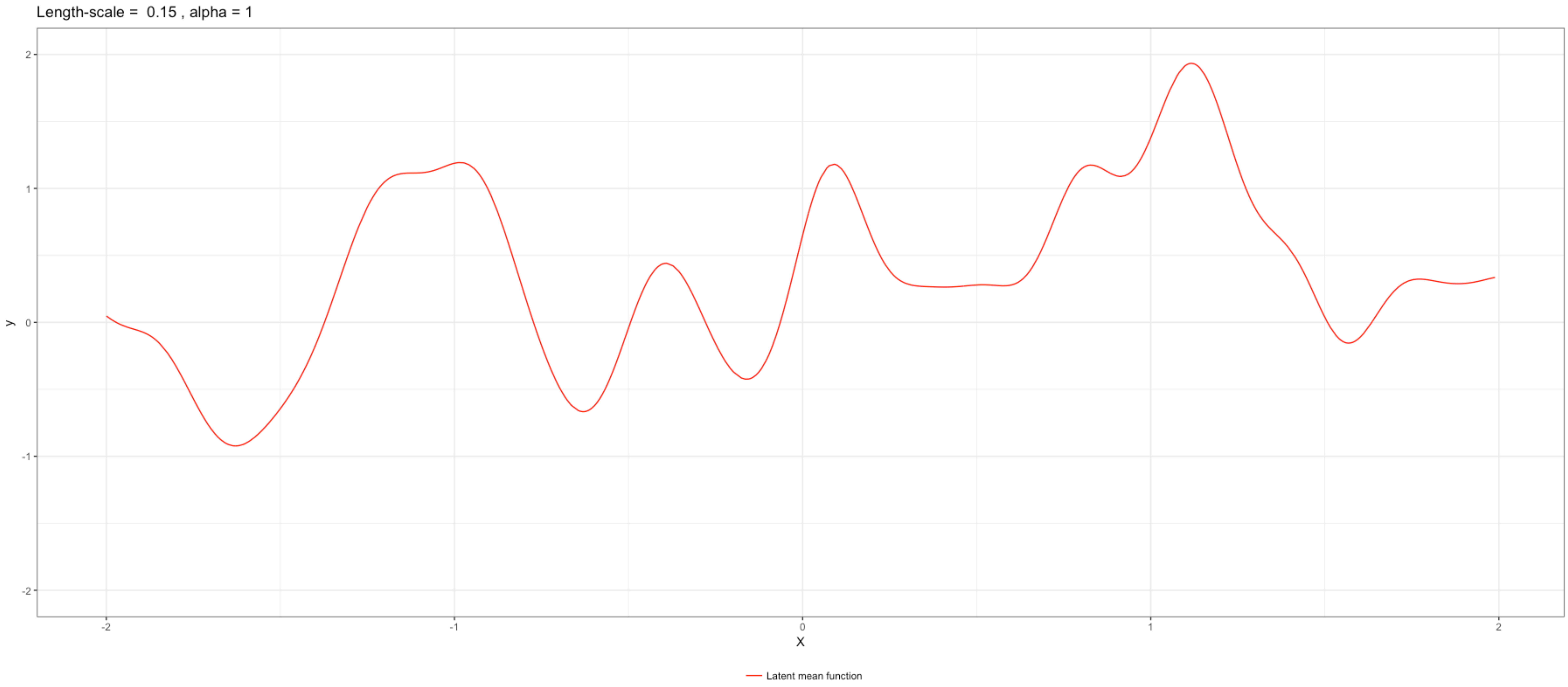
$$\begin{aligned}\operatorname{cov}(f(x_i), f(x_j)) &= k(x_i, x_j | \theta) \\ &= \alpha^2 \exp\left(-\frac{1}{2\ell^2}(x_i - x_j)^2\right)\end{aligned}$$

```
cov_exp_quad(real x,
             real alpha,
             real length_scale)
```

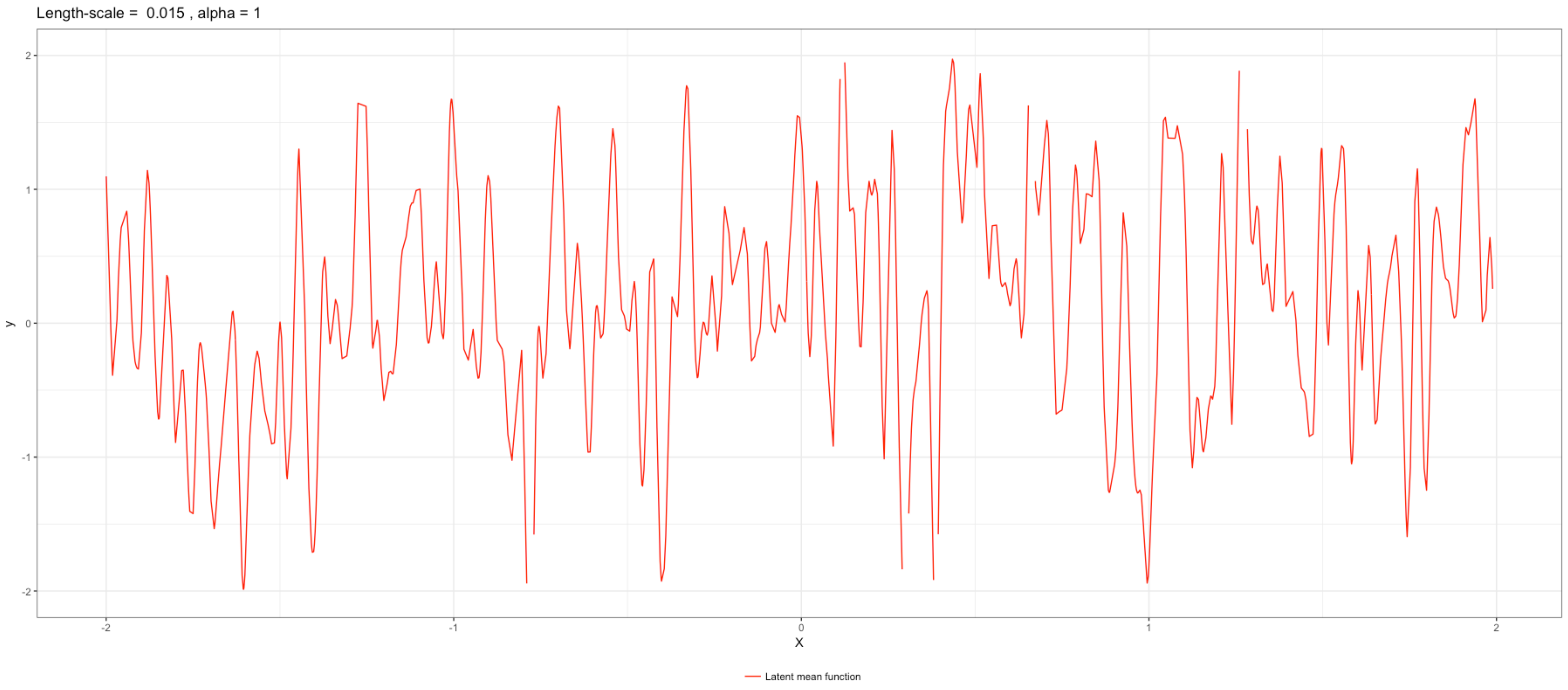
GENERATIVE MODEL – FINITE SAMPLE DRAW FROM GP



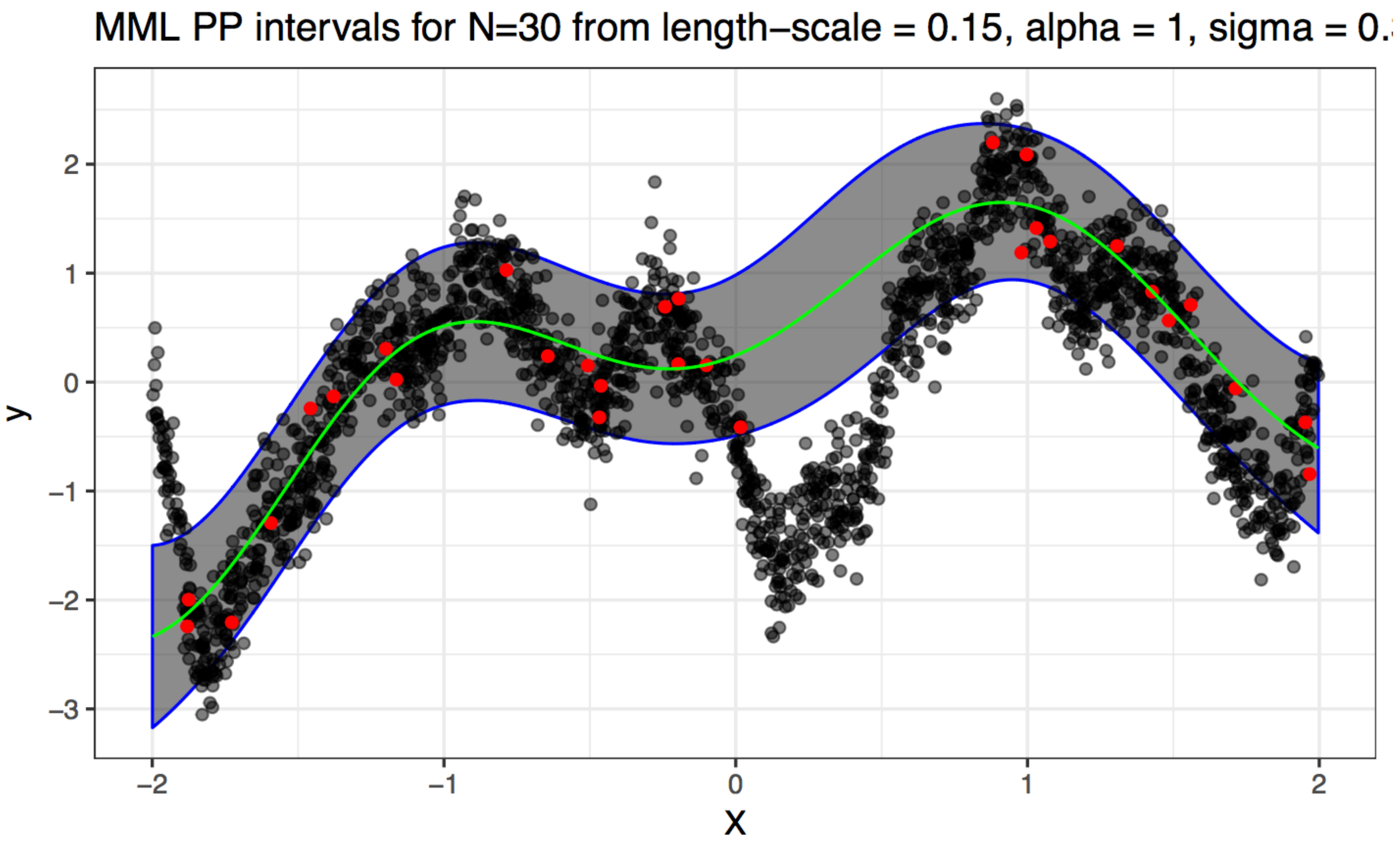
GENERATIVE MODEL – FINITE SAMPLE DRAW FROM GP



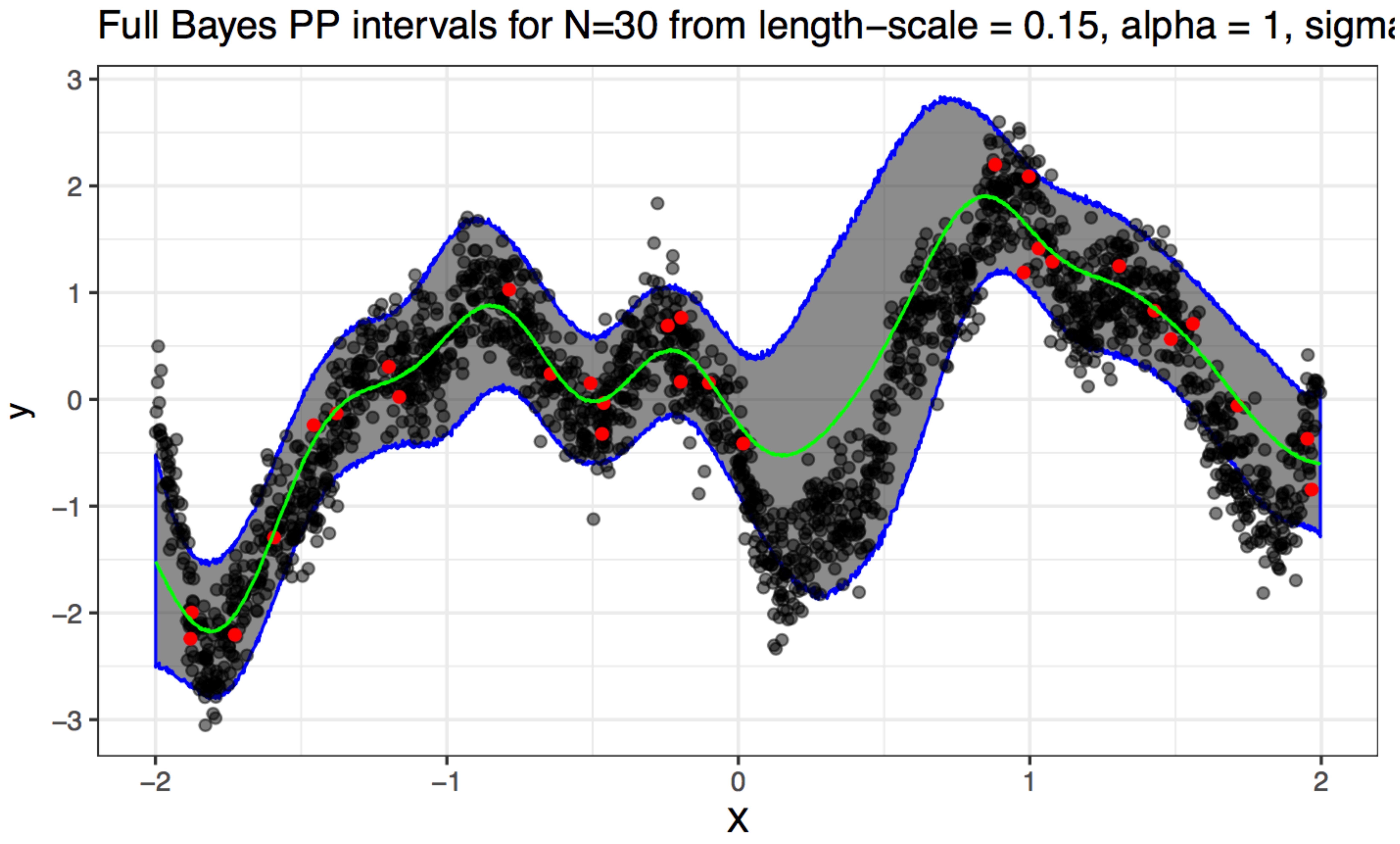
GENERATIVE MODEL – FINITE SAMPLE DRAW FROM GP



GENERATIVE MODEL – PUT PRIORS ON YOUR HYPER PARAMETERS!



GENERATIVE MODEL – PUT PRIORS ON YOUR HYPER PARAMETERS!



BUILDING GP MODELS IN STAN

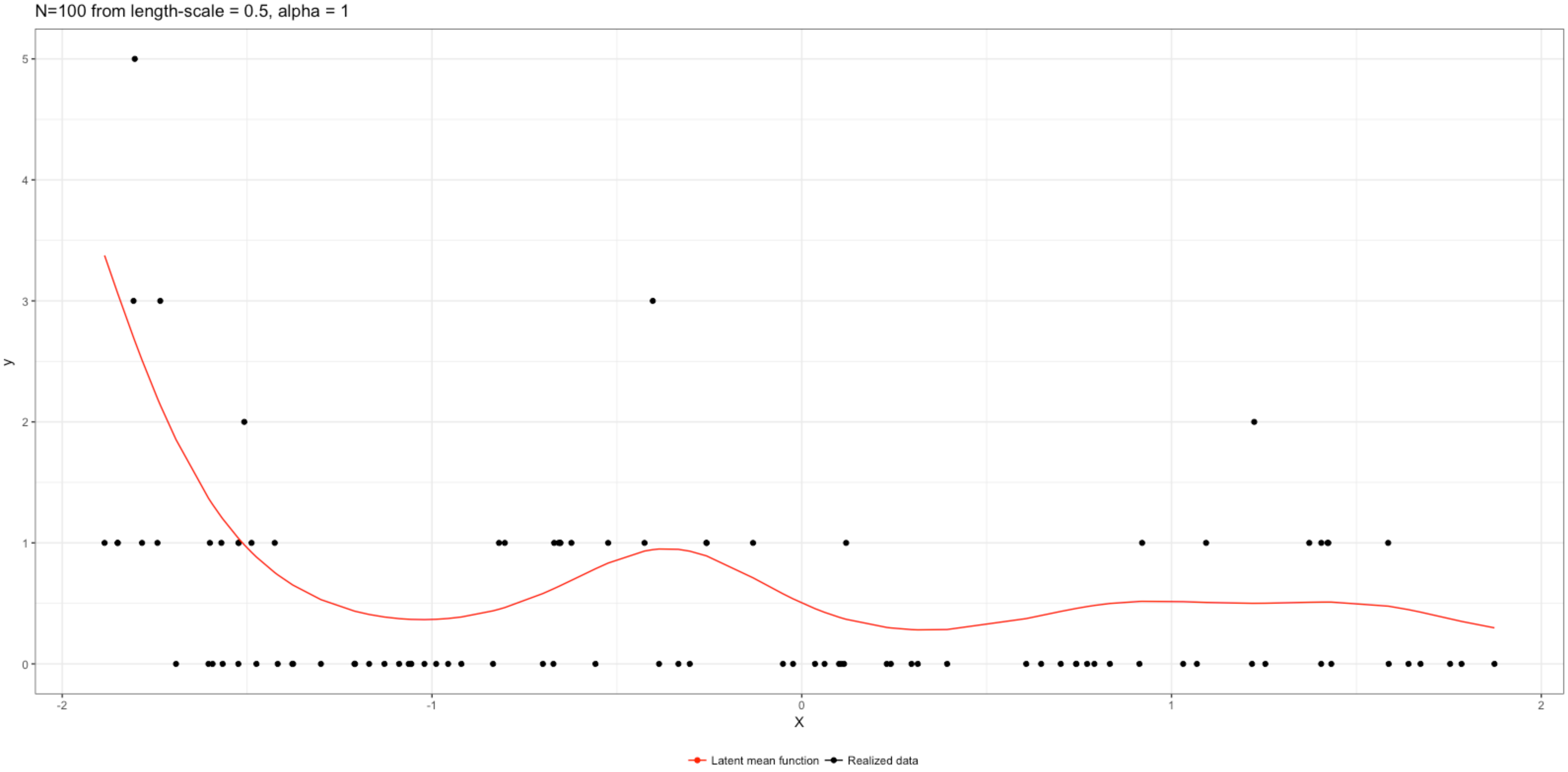
DEFINE PROBABILITY MODEL - LATENT VARIABLE

$$
\begin{aligned}
\mathbf{f} &\sim \text{MultiNormal}(0, K_{\ell,\alpha}(\mathbf{x}, \mathbf{x})) \\
\sigma &\sim \text{Half-Normal}(0, 1) \\
y_i &\sim \text{Normal}(f_i, \sigma) \\
\forall i &\in \{1, \dots, N\}, \mathbf{f}, \mathbf{x} \in R^N
\end{aligned}
$$

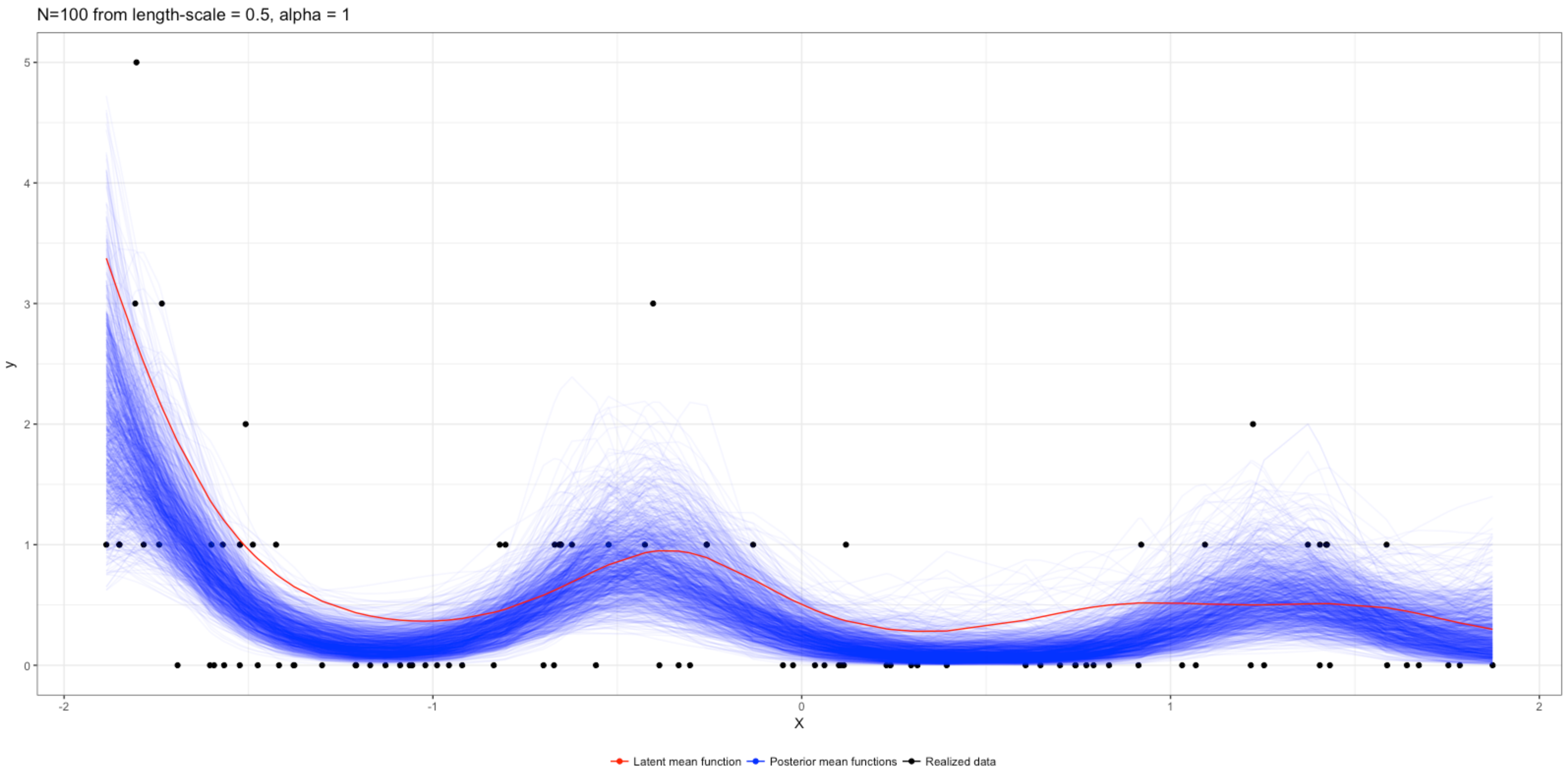
DEFINE PROBABILITY MODEL - LATENT VARIABLE

$$\mathbf{f} \sim \mathrm{MultiNormal}(0, K_{\ell,\alpha}(\mathbf{x}, \mathbf{x}))$$
$$y_i \sim \mathrm{Poisson}(f_i)$$
$$\forall i \in \{1, \ldots, N\}, \mathbf{f}, \mathbf{x} \in R^N$$

COUNT DATA


N=100 from length-scale = 0.5, alpha = 1
y
5
4
3
2
1
0
-2
-1
0
1
2
X
Latent mean function
Realized data

RESULTS


N=100 from length-scale = 0.5, alpha = 1
y
X
0
1
2
3
4
5
-2
-1
0
1
2
Latent mean function
Posterior mean functions
Realized data

DEFINE PROBABILITY MODEL – LATENT VARIABLE

$$L \times L^T = K_{\ell,\alpha}$$

DEFINE PROBABILITY MODEL - LATENT VARIABLE

$$L = \text{cholesky\_decompose}(K_{\ell,\alpha})$$

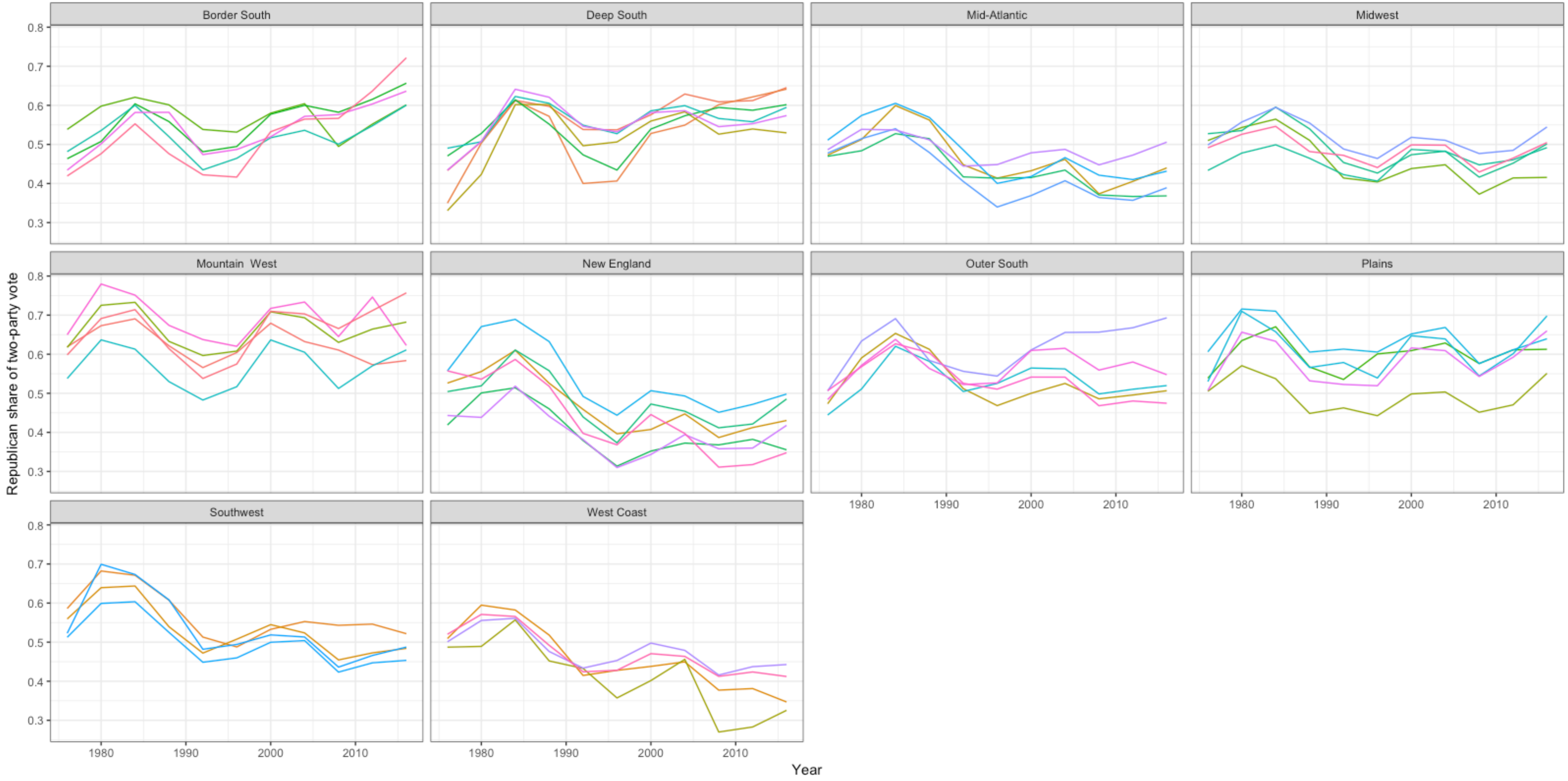
$$\boldsymbol{\eta} \sim \text{Normal}(0, 1)$$

$$\mathbf{f} = L \times \boldsymbol{\eta}$$

$$\mathbf{f} \sim \text{MultiNormal}(0, K_{\ell,\alpha}(x, x))$$

PRESIDENTIAL FORECASTING

FORECASTING US PRESIDENTIAL ELECTIONS 2020 – 2028



FORECASTING US PRESIDENTIAL ELECTIONS 2020 – 2028

$$y_{t,j} \sim \mathrm{Beta}(\mu_{t,j}\,\nu, (1-\mu_{t,j})\,\nu)$$

$$\operatorname{logit}\mu_{t,j} = \theta_t^{\text{year}} + \theta_j^{\text{state}} + \theta_{k[j]}^{\text{region}}$$

$$+ \gamma_{t,j} + \delta_{t,k[j]}$$

Short-term wiggle

Long-term trend

State $\rightarrow$ $$\boldsymbol{\gamma_j} \sim \mathrm{MultiNormal}(0, K_{\ell_1^\gamma,\alpha_1^\gamma} + K_{\ell_2^\gamma,\alpha_2^\gamma})$$

Region $\rightarrow$ $$\boldsymbol{\delta_k} \sim \mathrm{MultiNormal}(0, K_{\ell_1^\delta,\alpha_1^\delta} + K_{\ell_2^\delta,\alpha_2^\delta})$$

FORECASTING US PRESIDENTIAL ELECTIONS 2020 - 2028

$$\boldsymbol{\eta} \in R^{T,J}, L \in R^{T,T}$$

$$L = \text{cholesky\_decompose}(K_{\ell,\alpha})$$

$$\boldsymbol{\eta} \sim \text{Normal}(0, 1)$$

$$\boldsymbol{\gamma} = L \times \boldsymbol{\eta}$$

$$\boldsymbol{\gamma}_{[,j]} \sim \text{MultiNormal}(0, K_{\ell,\alpha}(x, x))$$

FORECASTING US PRESIDENTIAL ELECTIONS 2020 - 2028

$$\ell_1^\gamma \sim \mathrm{Weibull}(30, 8)$$

$$\ell_2^\gamma \sim \mathrm{Weibull}(30, 3)$$

$$\ell_1^\delta \sim \mathrm{Weibull}(30, 8)$$

$$\ell_2^\delta \sim \mathrm{Weibull}(30, 3)$$

FORECASTING US PRESIDENTIAL ELECTIONS 2020 – 2028

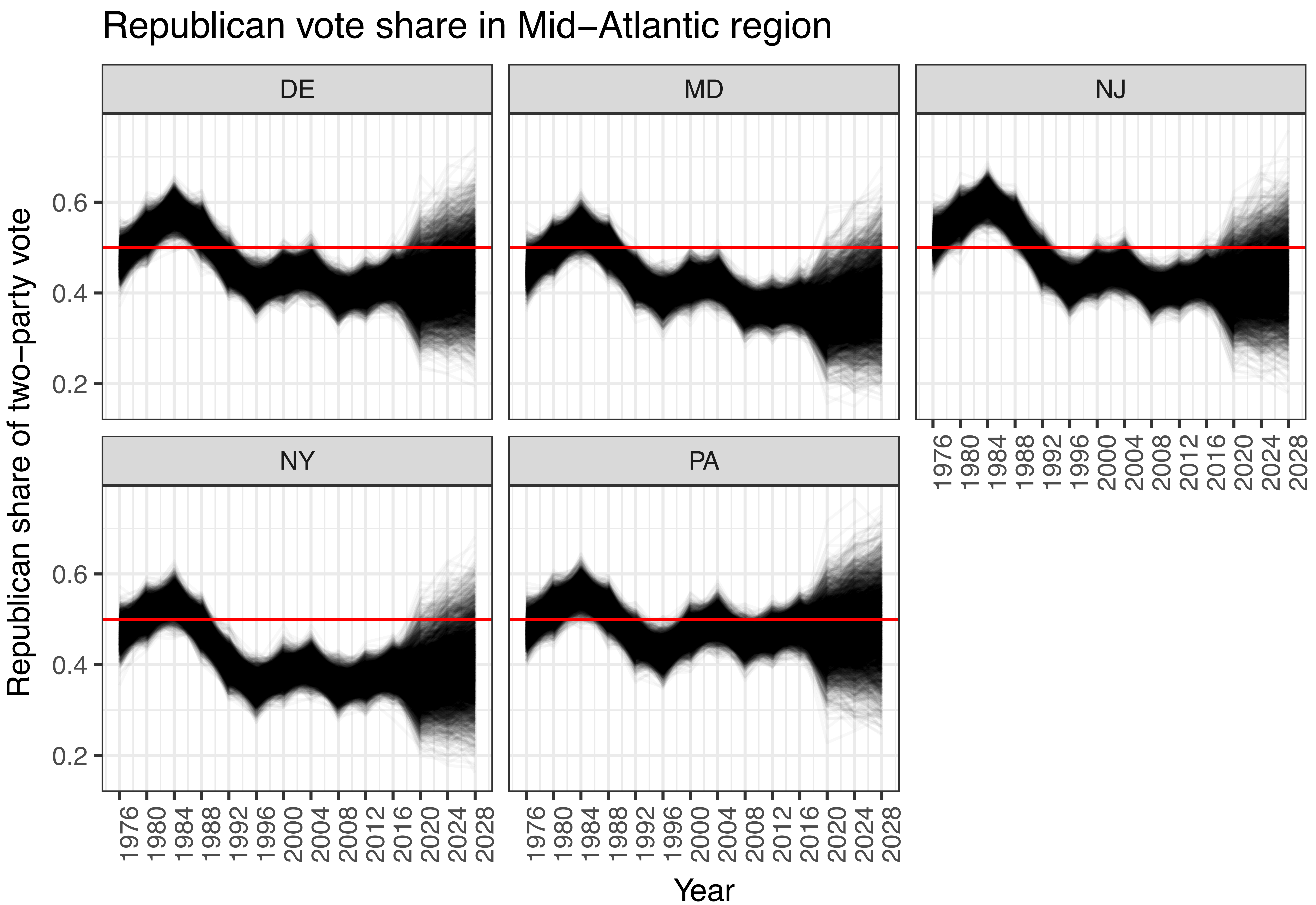
$$\sum_{v=1}^{V} \sigma_v^2 \propto \tau^2$$

$$(\sigma_1^2, \ldots, \sigma_V^2) = (\pi_1, \ldots, \pi_V) \times \tau^2$$

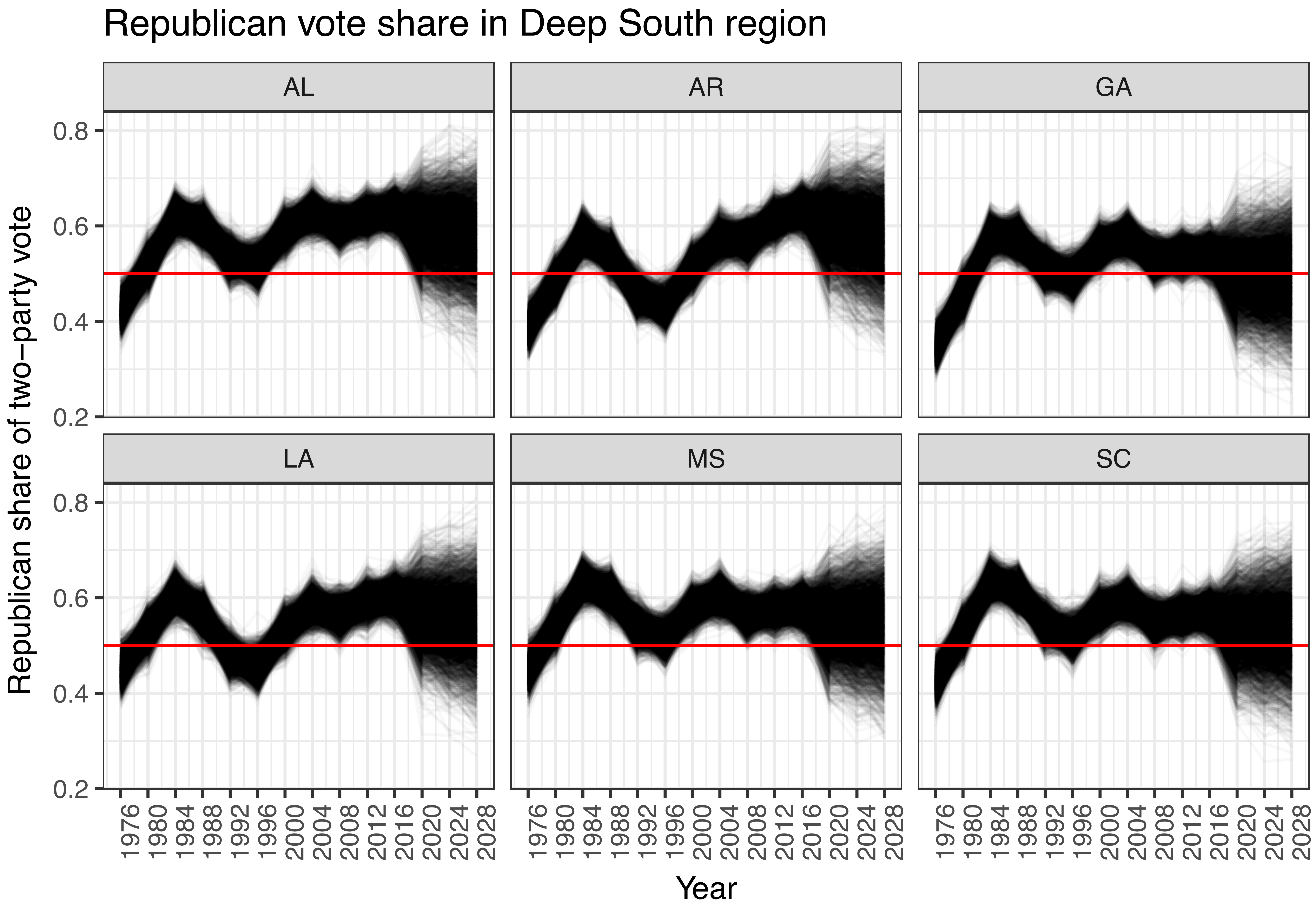
$$\boldsymbol{\pi} \sim \mathrm{Dirichlet}(A)$$

$$\tau \sim \mathrm{Gamma}(3, 3)$$

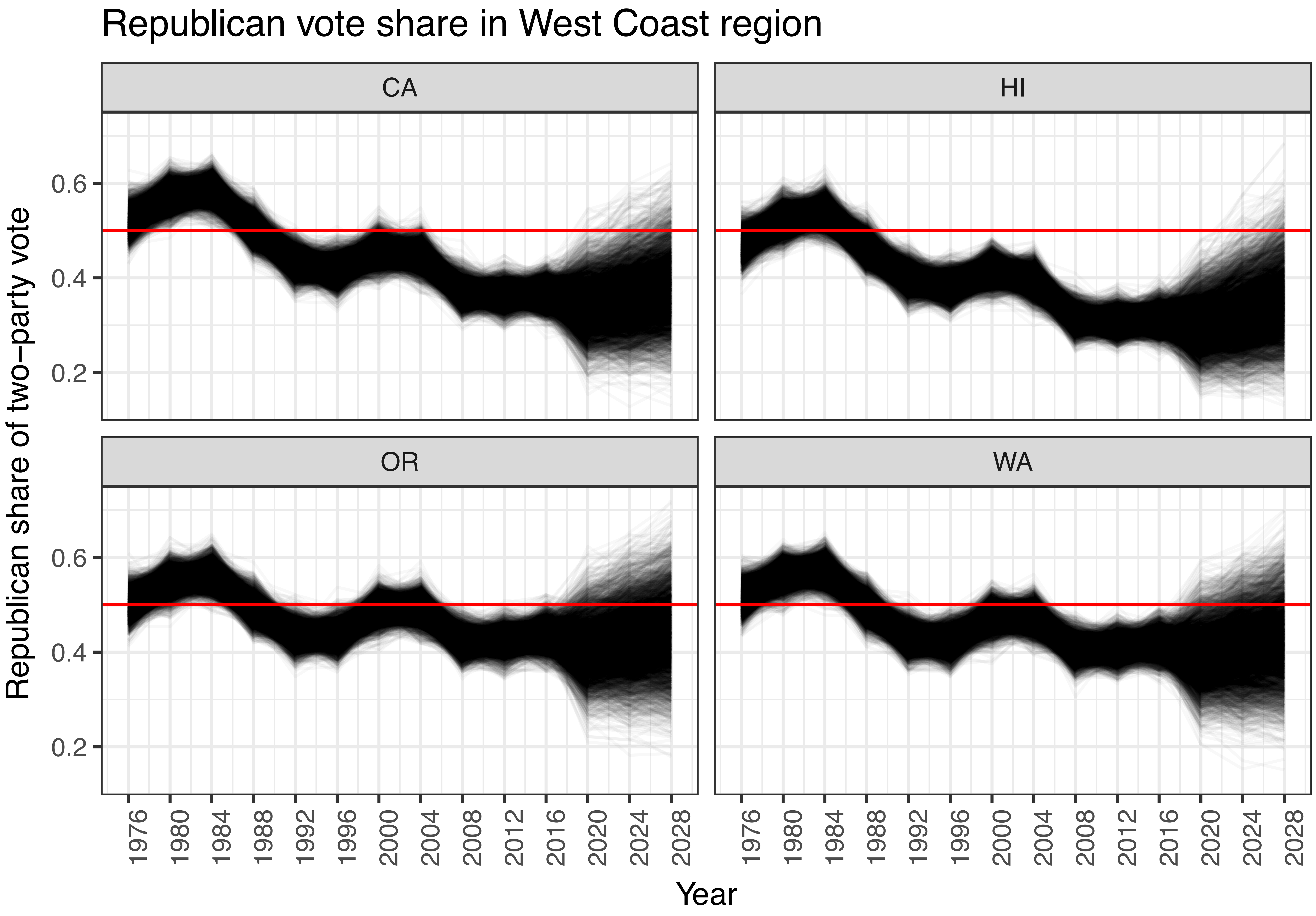
FORECASTING US PRESIDENTIAL ELECTIONS 2020 – 2028



FORECASTING US PRESIDENTIAL ELECTIONS 2020 – 2028



FORECASTING US PRESIDENTIAL ELECTIONS 2020 – 2028



FUTURE OF GAUSSIAN PROCESSES IN STAN

SPEED

- Blocked cholesky gradient implemented on a branch on stan-dev/math GitHub
  https://github.com/stan-dev/math/tree/feature/issue-384-blocked-cholesky
- On blocked-cholesky branch, possible to generate ~1000 effective samples for each hyperparameter in an N=1280 cov_exp_quad GP with unknown hyperparameters in about 45 minutes on a compute-intensive EC2 instance
- Will be implemented in 2.14++
- Kronecker inference: if data are on a D-dimensional grid: $K_1, \ldots, K_D$

$$K_{\text{all}} = K_1 \otimes K_2 \otimes \cdots \otimes K_D$$

- More kernels!

THANKS

- To Stan dev-team: http://mc-stan.org/team/
- Aki Vehtari, Michael Betancourt, and Andrew Gelman
- YouGov