Nonstationary Covariance Functions for Spatial Modelling

Chris Paciorek

Department of Biostatistics
Harvard School of Public Health
paciorek@hsph.harvard.edu

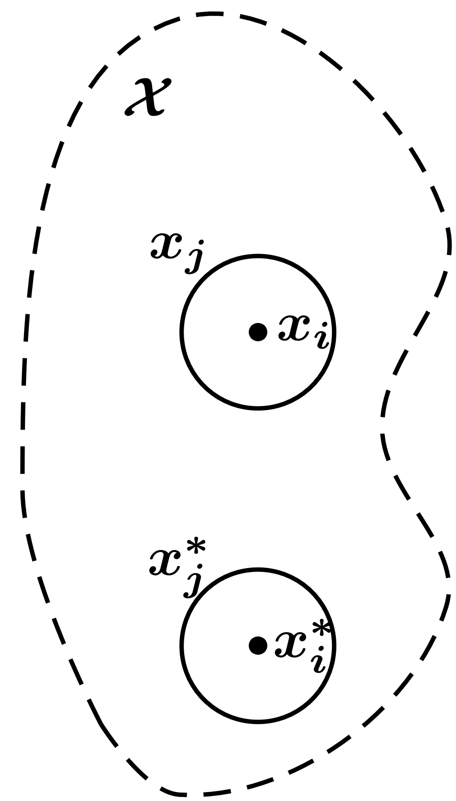
November 21, 2003

OUTLINE

- Gaussian processes and nonstationary covariance
- Generalized nonstationary covariance via convolution
- Application to nonstationary kriging
- A Bayesian model for nonstationary spatial processes
- Comparison with stationary modelling and free-knot splines
- Representations of stationary GPs for fast computation
  - ❖ Matérn-based basis functions
  - ❖ Fourier basis functions
- Efficient MCMC for generalized spatial models

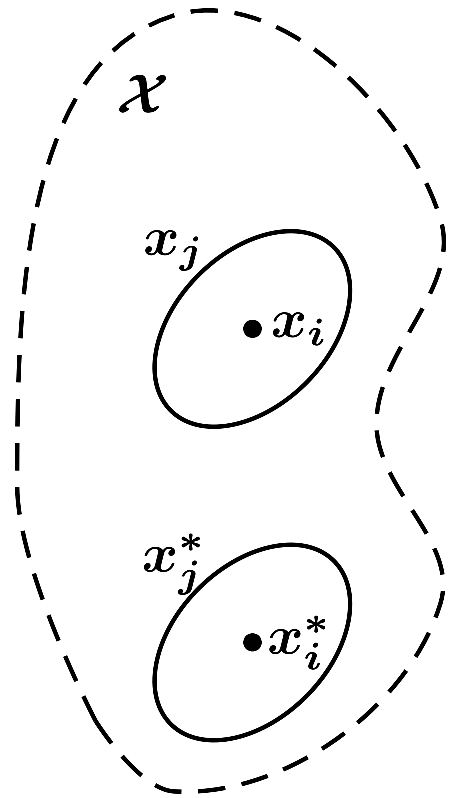
GAUSSIAN PROCESS DISTRIBUTION

- Infinite-dimensional joint distribution for $f(x),\ x \in \mathcal{X}$:
  - ❖ Example: $f(\cdot)$ a spatial process, $\mathcal{X} = \Re^2$
  - ❖ $f(\cdot) \sim \mathbf{GP}(\mu(\cdot), C(\cdot,\cdot))$
- Finite-dimensional marginals are normal
- Types of covariance functions, $C(x_i, x_j)$:
  - ❖ stationary, isotropic
  - ❖ stationary, anisotropic
  - ❖ nonstationary



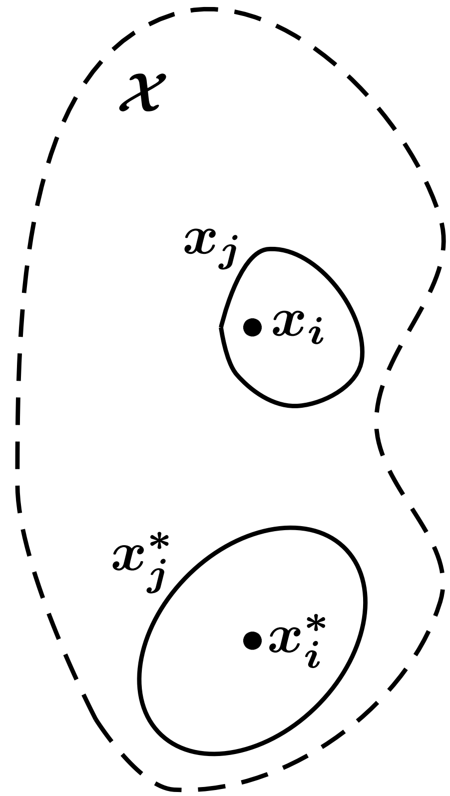
GAUSSIAN PROCESS DISTRIBUTION

- Infinite-dimensional joint distribution for $\boldsymbol{f(x)},\ \boldsymbol{x} \in \boldsymbol{\mathcal{X}}$:
  - ❖ Example: $\boldsymbol{f}(\cdot)$ a spatial process, $\boldsymbol{\mathcal{X}} = \boldsymbol{\Re^2}$
  - ❖ $\boldsymbol{f}(\cdot) \sim \mathbf{GP}(\mu(\cdot), C(\cdot,\cdot))$
- Finite-dimensional marginals are normal
- Types of covariance functions, $\boldsymbol{C(x_i, x_j)}$:
  - ❖ stationary, isotropic
  - ❖ stationary, anisotropic
  - ❖ nonstationary



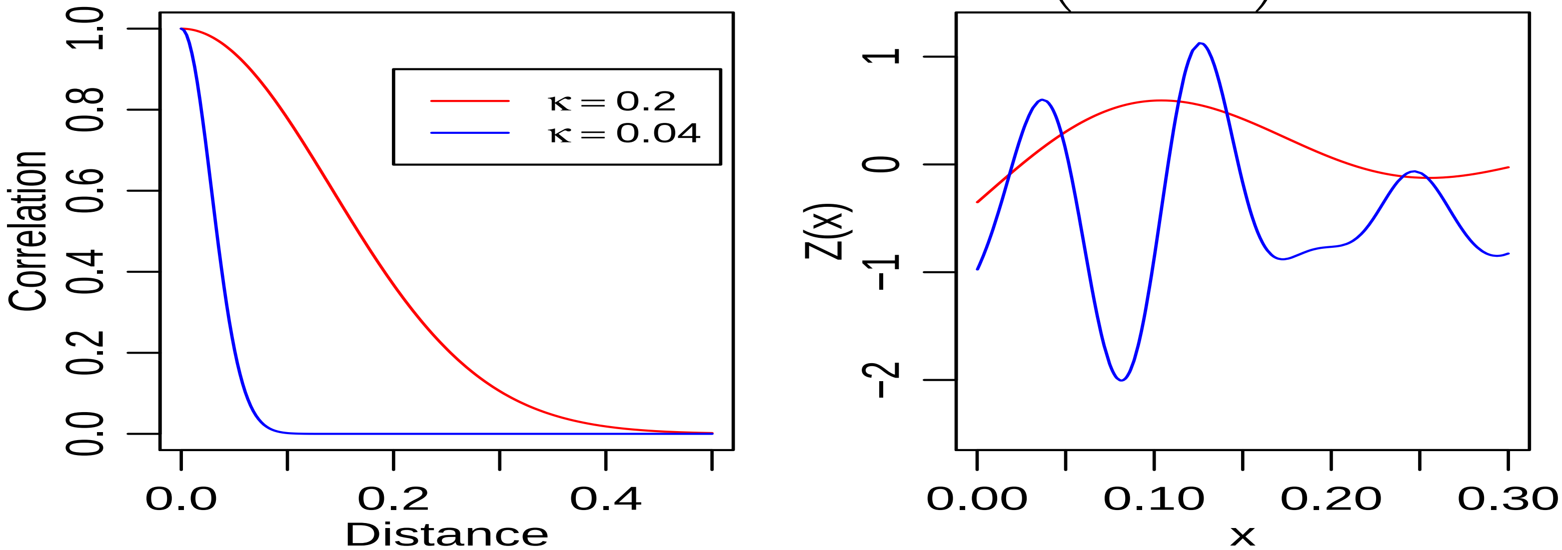
GAUSSIAN PROCESS DISTRIBUTION

- Infinite-dimensional joint distribution for $\boldsymbol{f(x)},\ \boldsymbol{x} \in \boldsymbol{\mathcal{X}}$:
  - ❖ Example: $\boldsymbol{f}(\cdot)$ a spatial process, $\boldsymbol{\mathcal{X}} = \Re^{\mathbf{2}}$
  - ❖ $\boldsymbol{f}(\cdot) \sim \mathbf{GP}(\boldsymbol{\mu}(\cdot), \boldsymbol{C}(\cdot, \cdot))$
- Finite-dimensional marginals are normal
- Types of covariance functions, $\boldsymbol{C}(\boldsymbol{x_i}, \boldsymbol{x_j})$:
  - ❖ stationary, isotropic
  - ❖ stationary, anisotropic
  - ❖ nonstationary

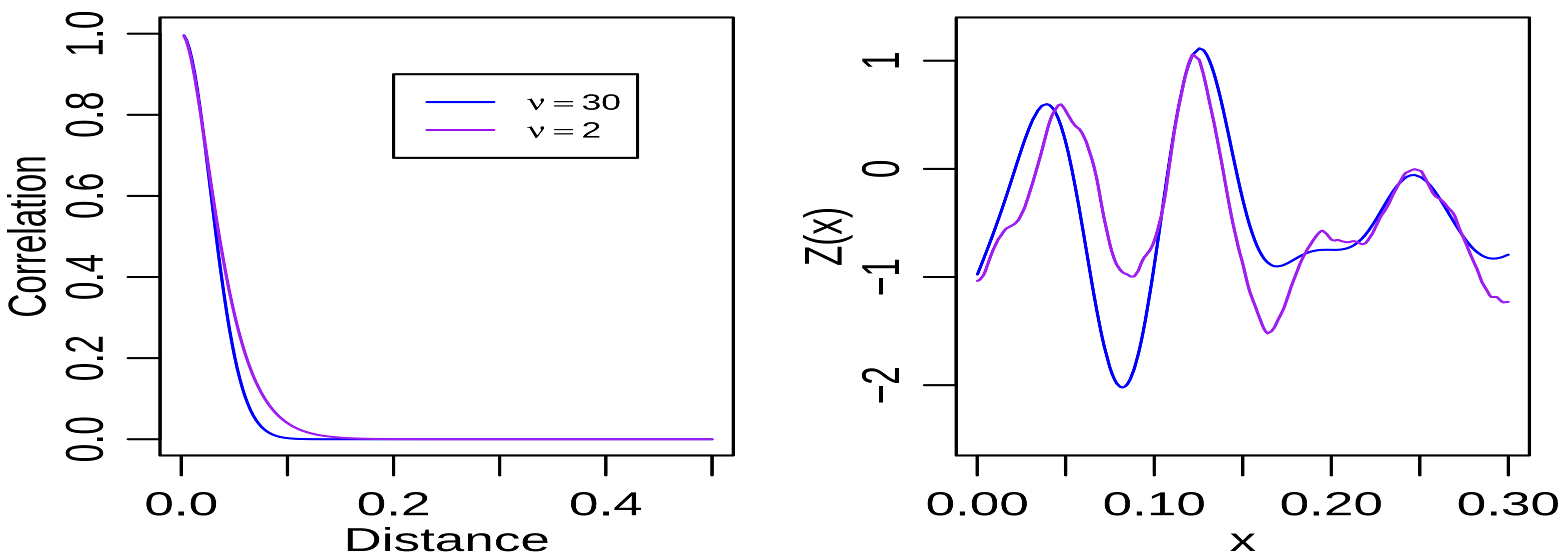


STATIONARY CORRELATION FUNCTIONS

Squared exponential: $R(\tau) = \exp\left(-\left(\frac{\tau}{\kappa}\right)^2\right)$

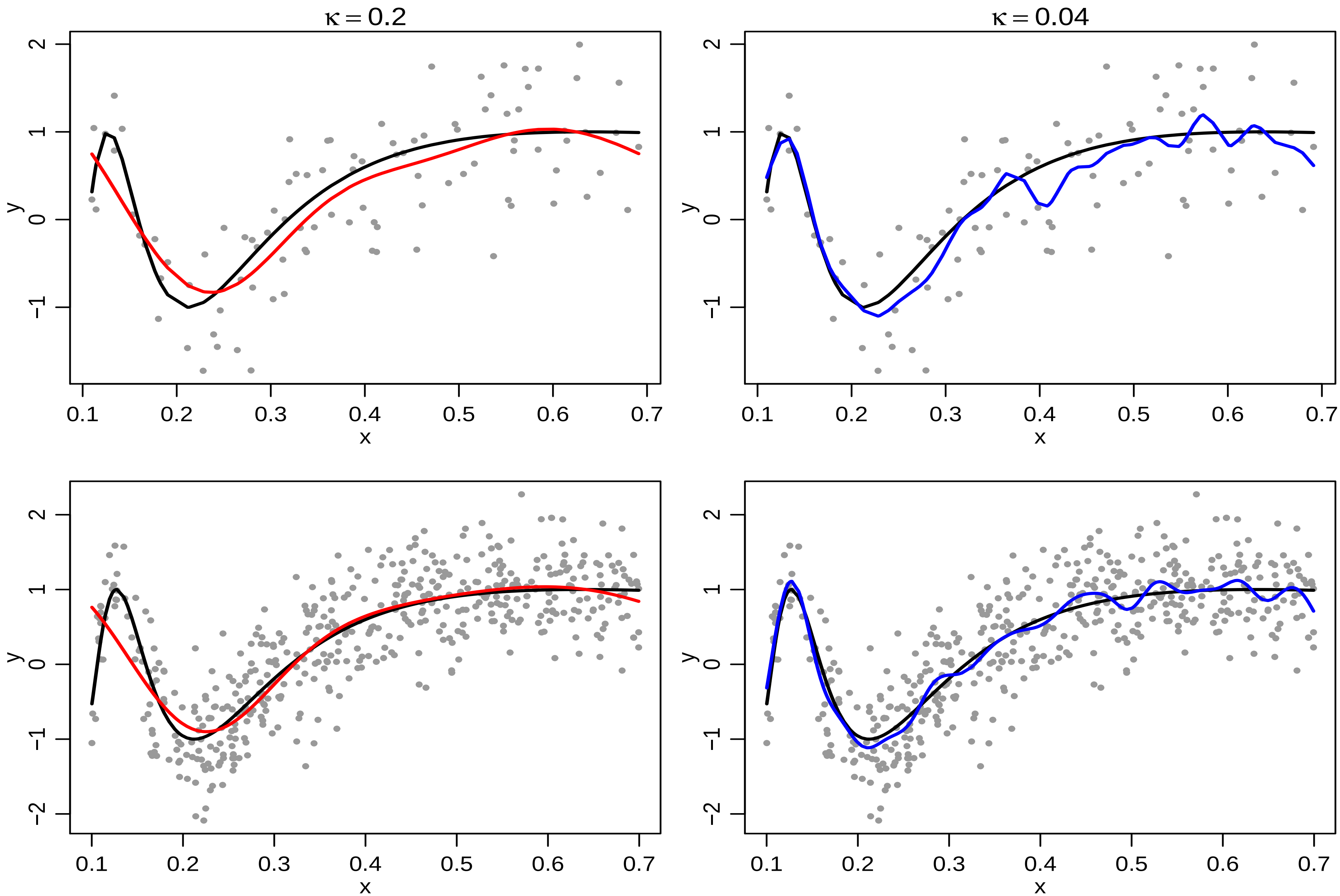


Matérn form: $R(\tau) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{2\sqrt{\nu\tau^2}}{\kappa}\right)^{\nu} K_\nu\left(\frac{2\sqrt{\nu\tau^2}}{\kappa}\right)$



- Differentiability controlled by ν, asymptotic advantages (Stein)

Degree of Smoothing



A Nonstationary Covariance

- Higdon, Swall, and Kern (1999) (HSK)

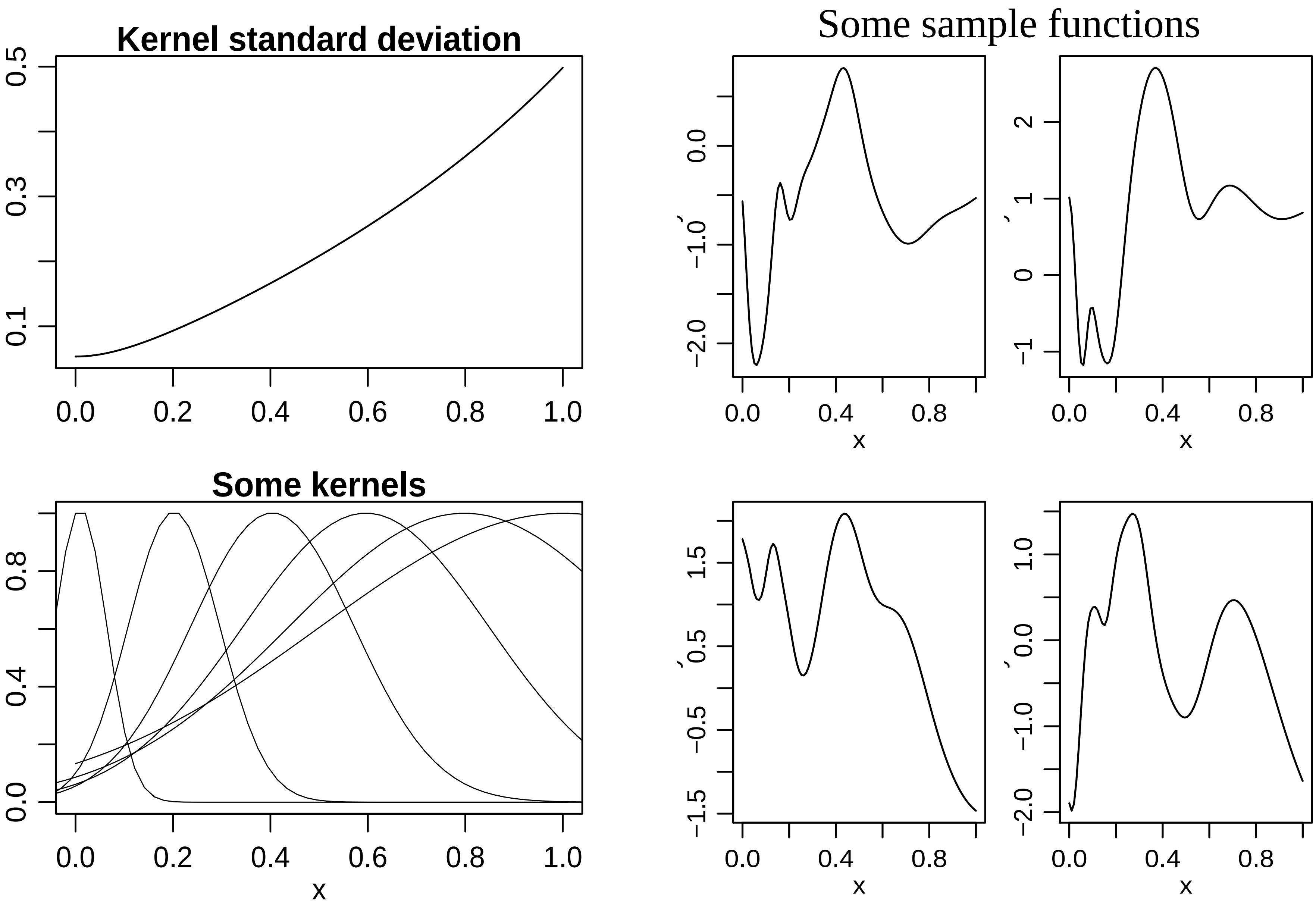
  $$R^{NS}(x_i, x_j) = c_{ij} \int_{\Re^P} k_{x_i}(u) k_{x_j}(u) du$$

- Guaranteed positive definite
- Gaussian kernels:

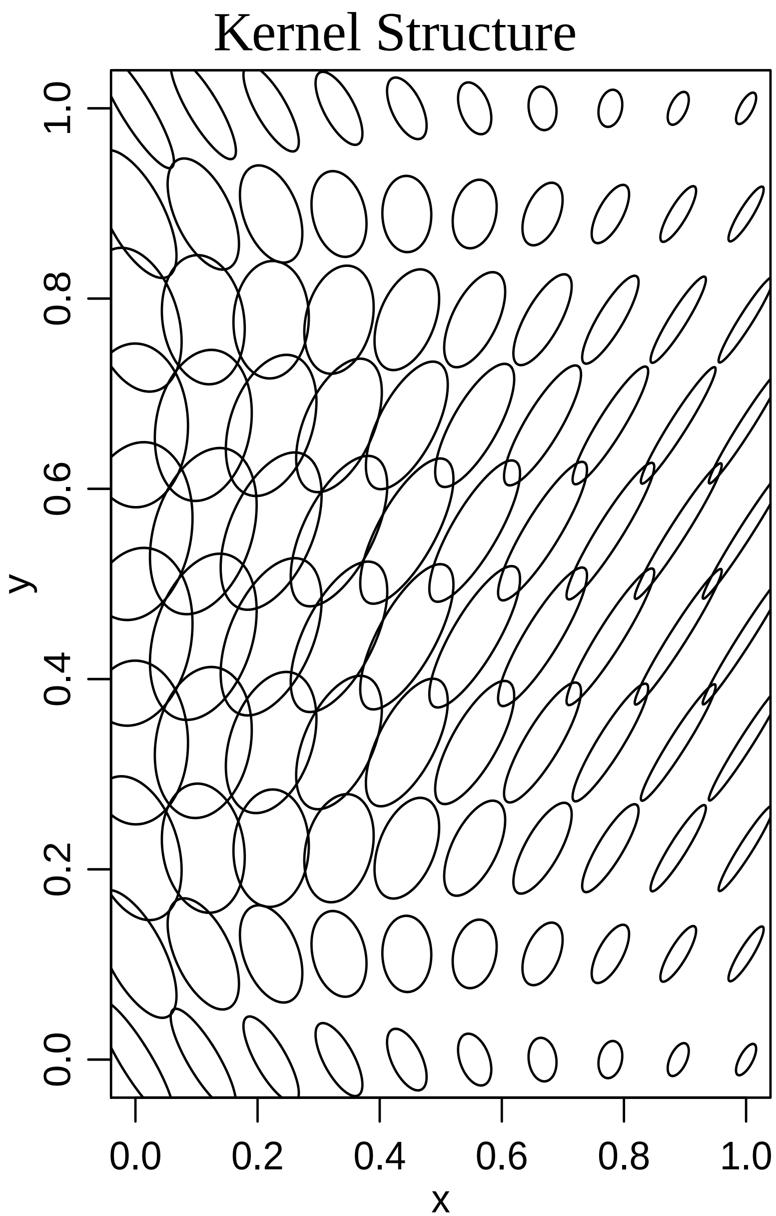
$$\begin{aligned} k_{x_i}(u) &\propto \exp\left(-(u - x_i)^T \Sigma_i^{-1} (u - x_i)\right) \\ R^{NS}(x_i, x_j) &= c_{ij} \exp\left(-(x_i - x_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1} (x_i - x_j)\right) \end{aligned}$$

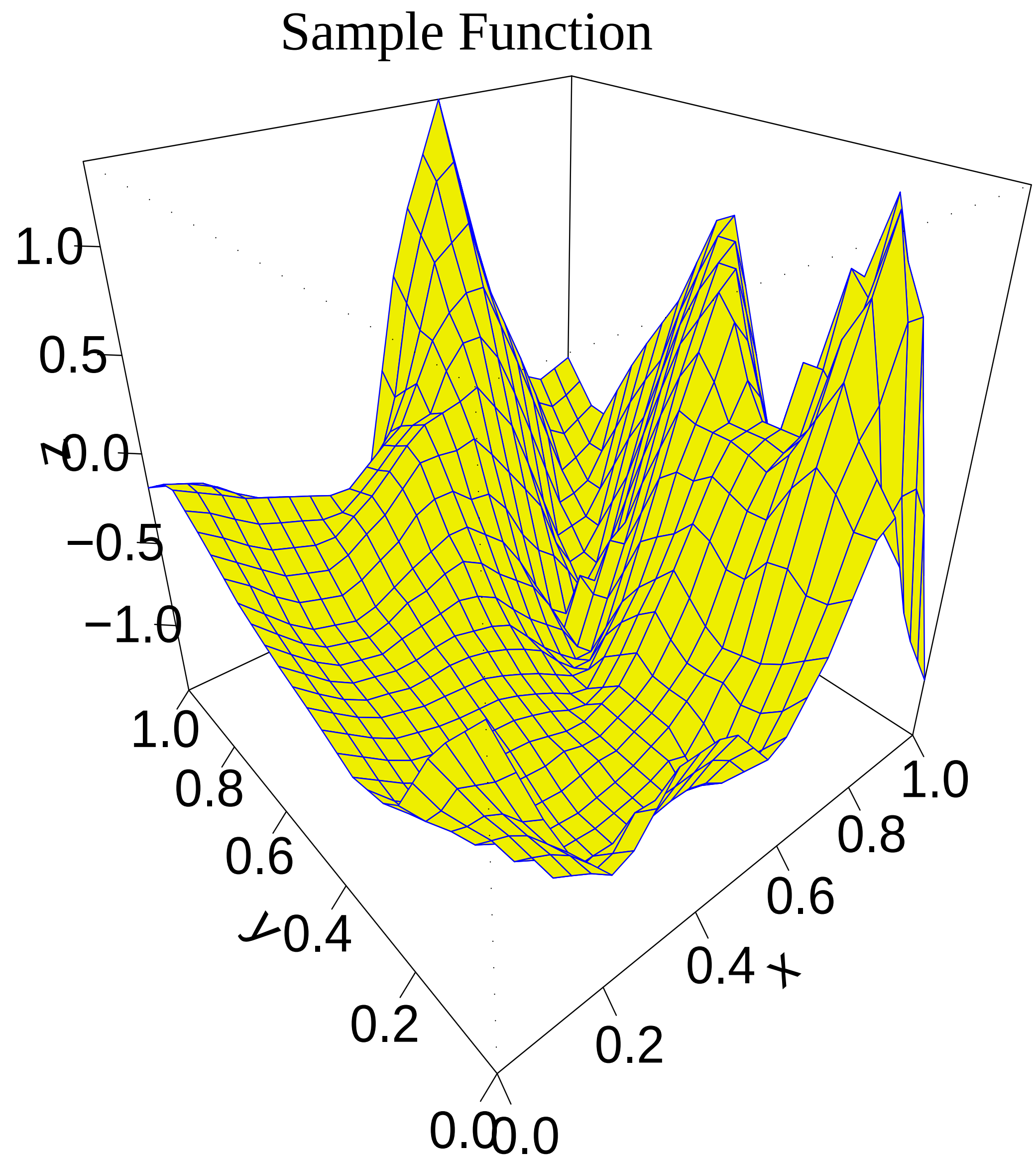
- $f(\cdot) \sim \mathrm{GP}(\mu, \sigma^2 R^{NS}(\cdot, \cdot))$

NONSTATIONARY GPS IN 1D



NONSTATIONARY GPS IN 2D





Generalizing the HSK Kernel Method

- Squared exponential form:

$$\exp\left(-\left(\frac{\tau}{\kappa}\right)^2\right) \Rightarrow c_{ij}\exp\left(-(x_i - x_j)^T\left(\frac{\Sigma_i+\Sigma_j}{2}\right)^{-1}(x_i - x_j)\right)$$

Infinitely-differentiable sample paths

- 'Distance measures'

isotropy	τ_{ij}^2	$= (x_i - x_j)^T(x_i - x_j)$
anisotropy	τ^{*2}_{ij}	$= (x_i - x_j)^T\Sigma^{-1}(x_i - x_j)$
nonstationarity	Q_{ij}	$= (x_i - x_j)^T\left(\dfrac{\Sigma_i + \Sigma_j}{2}\right)^{-1}(x_i - x_j)$

- Can we replace τ_{ij}^2 with Q_{ij} in other stationary correlation functions?

Generalized Nonstationary Covariance

- Theorem 1: if $\boldsymbol{R(\tau)}$ is positive definite for $\Re^{\boldsymbol{P}}, \boldsymbol{P = 1, 2, \ldots}$, then

$$\boldsymbol{R^{NS}(x_i, x_j) = \frac{|\Sigma_i|^{\frac{1}{4}} |\Sigma_j|^{\frac{1}{4}}}{\left|\frac{\Sigma_i + \Sigma_j}{2}\right|^{\frac{1}{2}}} R\left(\sqrt{Q_{ij}}\right)}$$

  is positive definite for $\Re^{\boldsymbol{P}}, \boldsymbol{P = 1, 2, \ldots}$

- Theorem 2: Smoothness properties of original stationary correlation retained

Proof (sketch)

- $$\boldsymbol{R(\tau) = \int_0^\infty \exp(-\tau^2 w) h(w) dw} \quad \text{(Schoenberg, 1938)}$$

- $$\begin{aligned} R^{NS}(x_i, x_j) &= \frac{2^{\frac{P}{2}} \left|\Sigma_i\right|^{\frac{1}{4}} \left|\Sigma_j\right|^{\frac{1}{4}}}{\left|\Sigma_i + \Sigma_j\right|^{\frac{1}{2}}} \int_0^\infty \exp(-Q_{i,j} w) h(w) dw \\ &= \frac{2^{\frac{P}{2}} \left|\Sigma_i\right|^{\frac{1}{4}} \left|\Sigma_j\right|^{\frac{1}{4}}}{\left|\Sigma_i + \Sigma_j\right|^{\frac{1}{2}}} \cdot \\ &\quad \int_0^\infty \exp\left(-\tfrac{1}{2}(x_i - x_j)^T \left(\tfrac{\Sigma_i + \Sigma_j}{2w}\right)^{-1} (x_i - x_j)\right) h(w) dw \\ &= \int_0^\infty \int_{\Re^P} k_{i,w}(u) k_{j,w}(u) du h(w) dw \end{aligned}$$

$$
\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(x_i, x_j) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \int_0^\infty \int_{\Re^P} k_{i,w}(u) k_{j,w}(u) du h(w) dw \\
&= \int_0^\infty \int_{\Re^P} \sum_{i=1}^{n} a_i k_{i,w}(u) \sum_{j=1}^{n} a_j k_{j,w}(u) du h(w) dw \\
&= \int_0^\infty \int_{\Re^P} \left(\sum_{i=1}^{n} a_i k_{i,w}(u) \right)^2 du h(w) dw \geq 0.
\end{aligned}
$$

- Covariance must depend only on location-specific kernels

NONSTATIONARY MATÉRN COVARIANCE

$$\frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{2\sqrt{\nu}\tau}{\kappa}\right)^{\nu} K_{\nu}\left(\frac{2\sqrt{\nu}\tau}{\kappa}\right) \Rightarrow \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(2\sqrt{\nu Q_{ij}}\right)^{\nu} K_{\nu}\left(2\sqrt{\nu Q_{ij}}\right)$$

Advantages: more flexible form, differentiability not constrained, possible asymptotic advantages

NONSTATIONARY KRIGING

- basic kriging model: $\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{\mu}, (\boldsymbol{\sigma}^2 \boldsymbol{R_f}(\boldsymbol{\kappa}, \boldsymbol{\nu}) + \boldsymbol{\eta}^2 \boldsymbol{I}))$
- $\sqrt{\boldsymbol{Q_{ij}}}$ is an anisotropic distance if $\boldsymbol{\Sigma_i} = \boldsymbol{\Sigma_j}$
- To krige, estimate parameters of $\boldsymbol{\Sigma}_{(\cdot)}$ locally and proceed as usual
- Possibilities
  - ❖ knit together region-specific covariance structures
  - ❖ estimate local covariance parameters based on a moving window
  - ❖ any approach that creates location-specific kernels produces a legitimate covariance structure

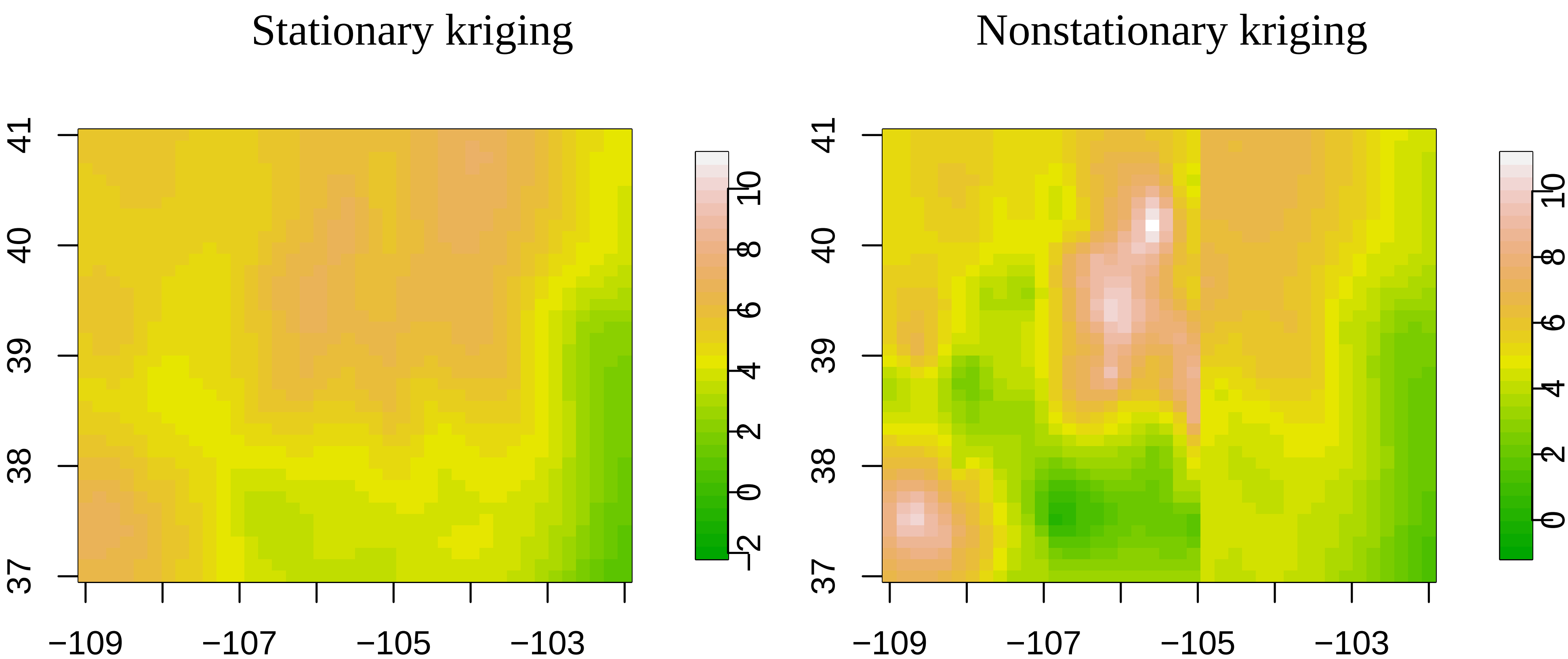
NONSTATIONARY KRIGING EXAMPLE

- precipitation anomalies in Colorado, August 1963
- fit covariance structure by maximizing marginal likelihood (EB):

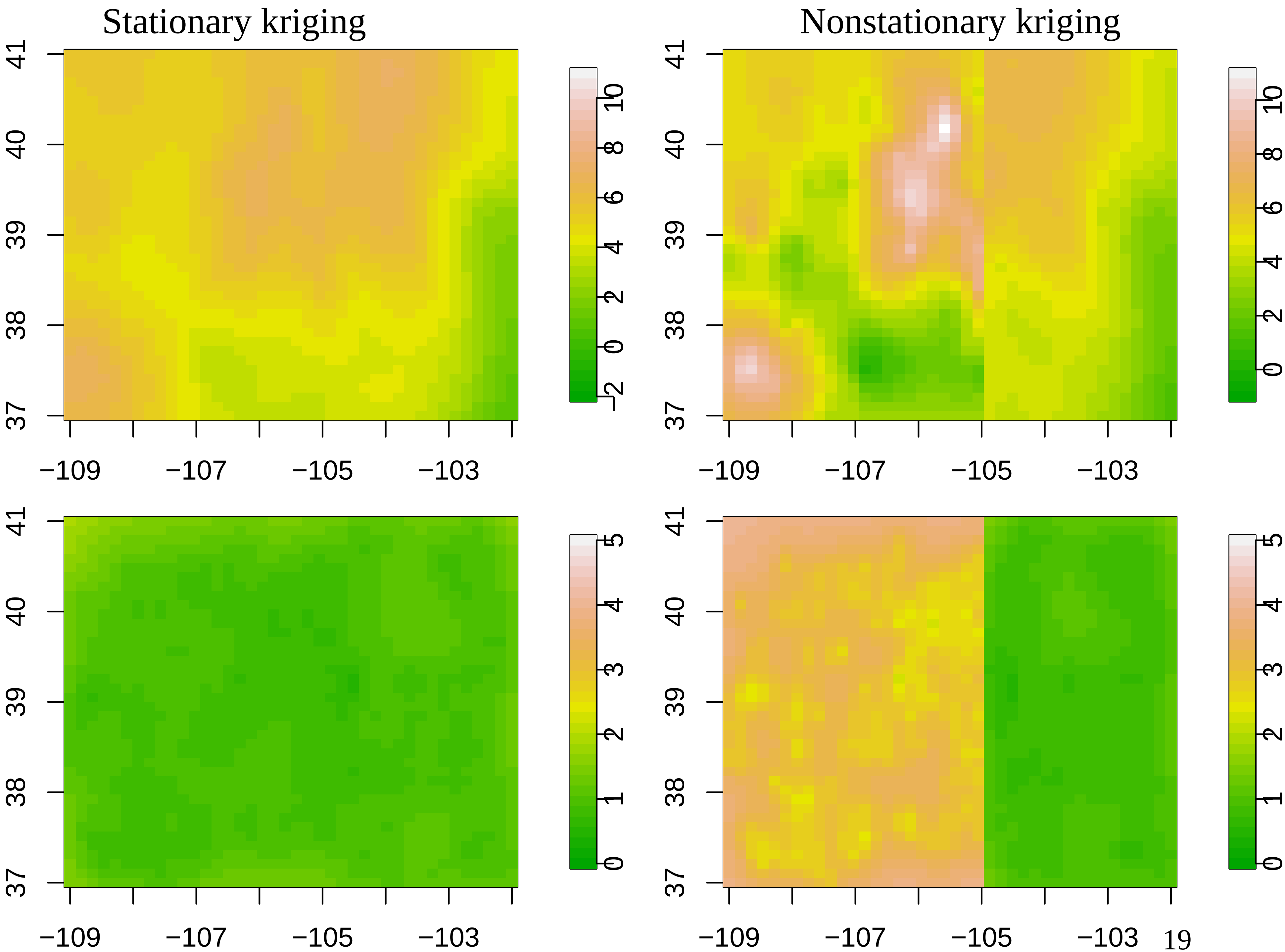
	η	σ	κ
whole state	3.75	7.10	5.49
eastern CO	3.63	9.50	7.92
western CO	3.46	3.23	0.69

- could also use standard variogram fitting

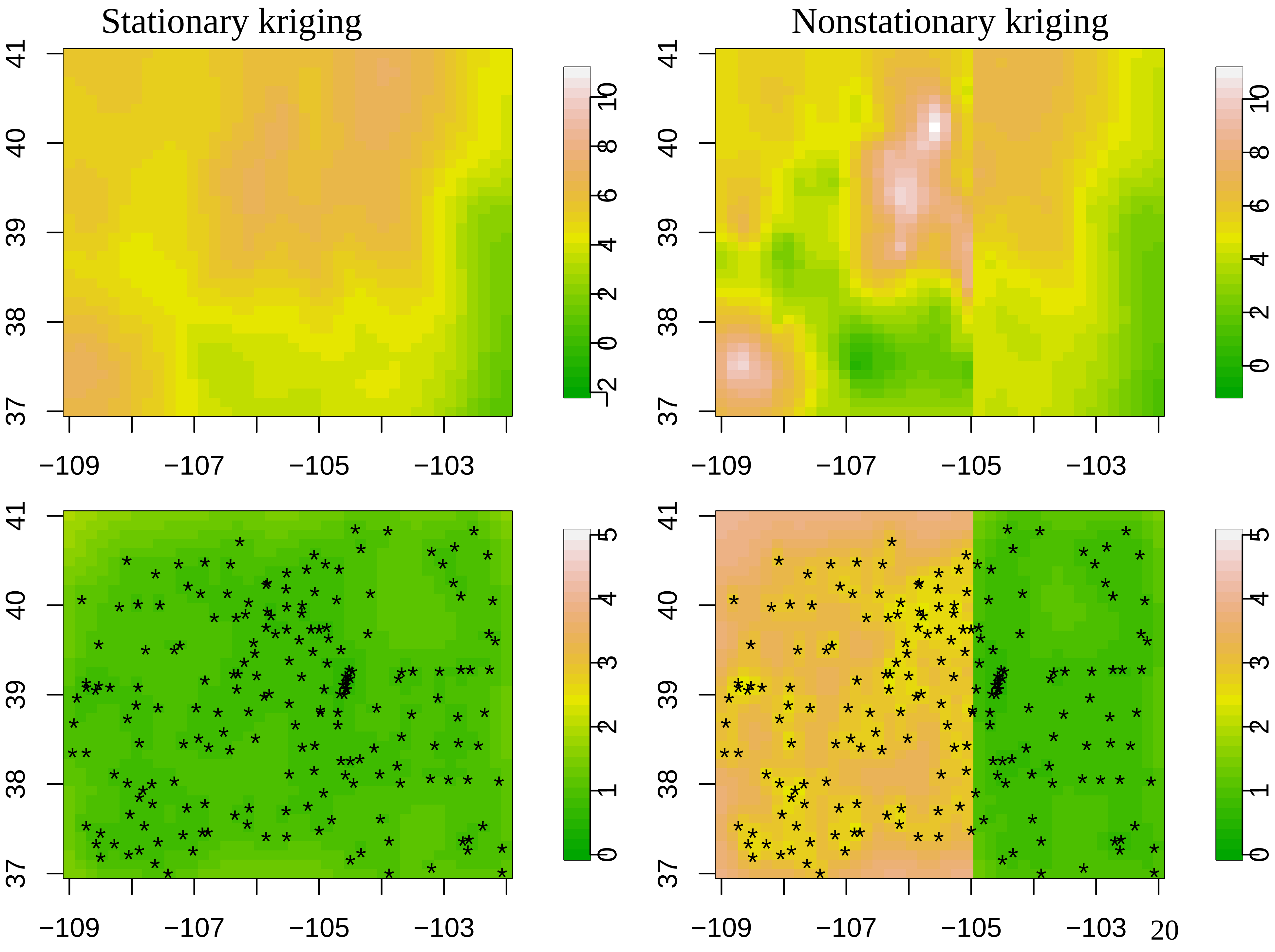
COLORADO PRECIPITATION ANOMALIES



Colorado precipitation anomalies



Colorado precipitation anomalies



A BASIC BAYESIAN SPATIAL MODEL

- Bayesian model

$$\begin{aligned} \boldsymbol{Y_i} &\sim \boldsymbol{N(f(x_i), \eta^2)},\ \boldsymbol{x_i} \in \boldsymbol{\Re^2} \\ \boldsymbol{f(\cdot)} &\sim \mathrm{GP}(\boldsymbol{\mu}, \boldsymbol{\sigma^2 R^{NS}}(\cdot, \cdot; \boldsymbol{\nu}, \boldsymbol{\Sigma}_{(\cdot)})) \end{aligned}$$

❖ Let $\boldsymbol{R^{NS}}$ be the nonstationary Matérn correlation

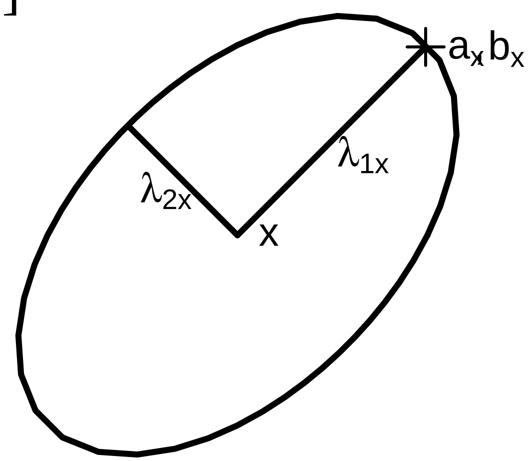
❖ Kernels ($\boldsymbol{\Sigma_x}$) constructed based on stationary GP priors

SMOOTHLY-VARYING KERNEL MATRICES

- Goals:
  - ❖ Define multiple kernel matrices, $\boldsymbol{\Sigma}_{\boldsymbol{x}},\ \boldsymbol{x} \in \boldsymbol{\mathcal{X}}$
  - ❖ Smoothly-varying (element-wise) in covariate space
  - ❖ Positive definite
- Approaches:
  - ❖ Parameterize ellipse foci and size ($\boldsymbol{\mathcal{X}} = \Re^{\boldsymbol{2}}$) (HSK)
    - ✦ mixing issues and non-generalizability to higher dimensions
  - ❖ Cholesky decomposition ($\boldsymbol{\mathcal{X}} = \Re^{\boldsymbol{P}}$): $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{L}_{\boldsymbol{x}}\boldsymbol{L}_{\boldsymbol{x}}^{T}$
    - ✦ hard to simultaneously control direction and size

SMOOTHLY-VARYING KERNEL MATRICES (2)

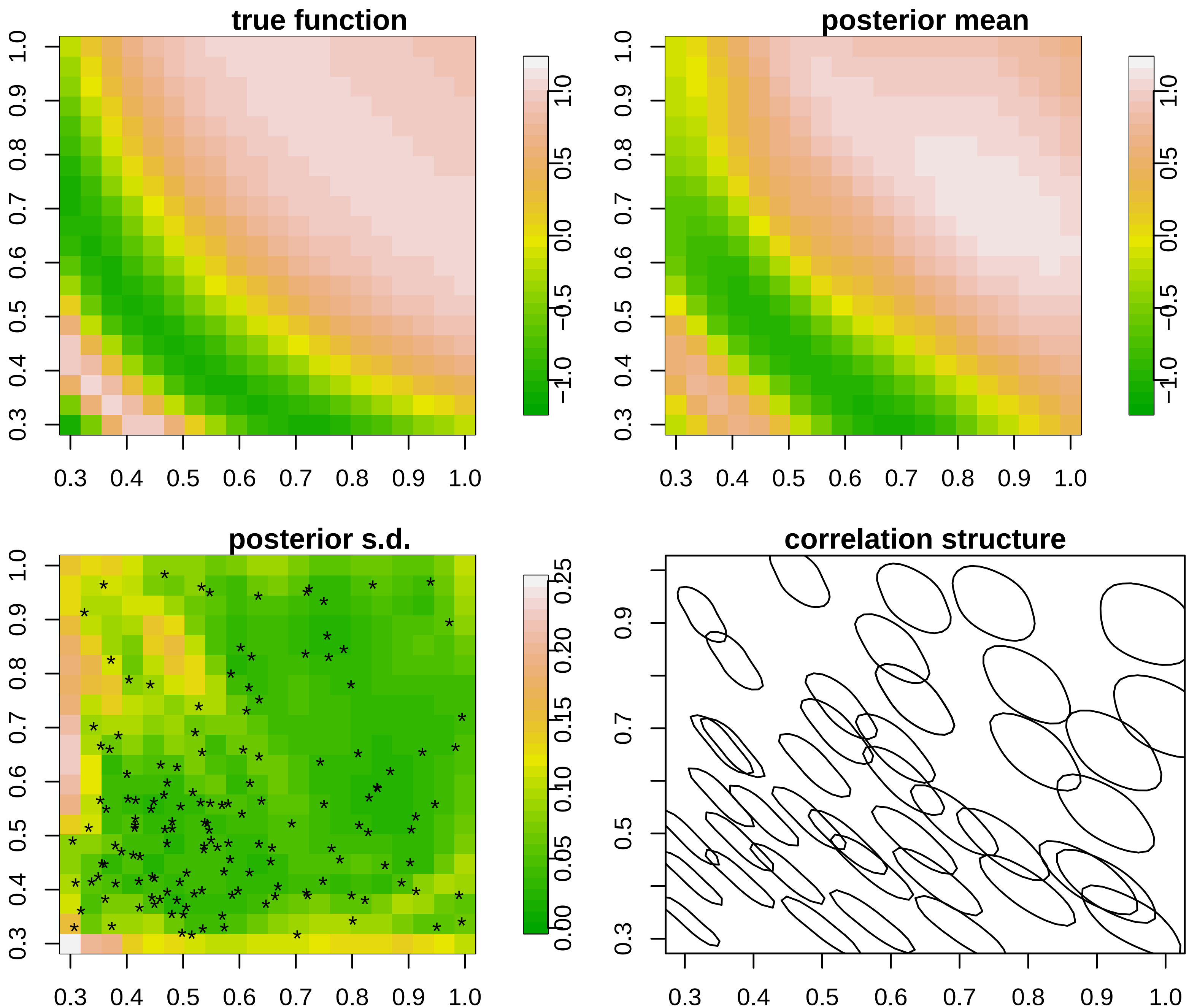
- Spectral decomposition ($\Re^P$): $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{\Gamma}_{\boldsymbol{x}}^{\boldsymbol{T}} \boldsymbol{\Lambda}_{\boldsymbol{x}} \boldsymbol{\Gamma}_{\boldsymbol{x}}$)
  - ❖ in $\Re^P$, $\boldsymbol{\Gamma}_{\boldsymbol{x}}$ parameterized as first eigenvector plus successive orthogonal vectors in reduced-dimension subspaces
  - ❖ in $\Re^{\mathbf{2}}$, stationary GP priors on unnormalized eigenvector coordinates $(\boldsymbol{a}_{\boldsymbol{x}}, \boldsymbol{b}_{\boldsymbol{x}})$ and on logarithm of eigenvalues $(\boldsymbol{\lambda}_{\boldsymbol{x},\mathbf{1}}, \boldsymbol{\lambda}_{\boldsymbol{x},\mathbf{2}})$
  - ❖ efficient parameterizations of stationary GPs for $\boldsymbol{\Phi} \in \{\boldsymbol{a}_{\boldsymbol{x}}, \boldsymbol{b}_{\boldsymbol{x}}, \boldsymbol{\lambda}_{\mathbf{1},\boldsymbol{x}}, \boldsymbol{\lambda}_{\mathbf{2},\boldsymbol{x}}\}$ [more later]



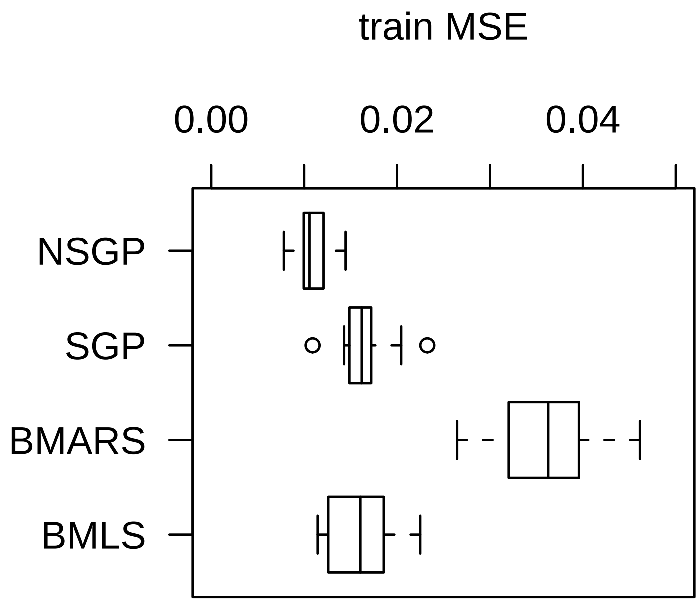
Empirical Assessment

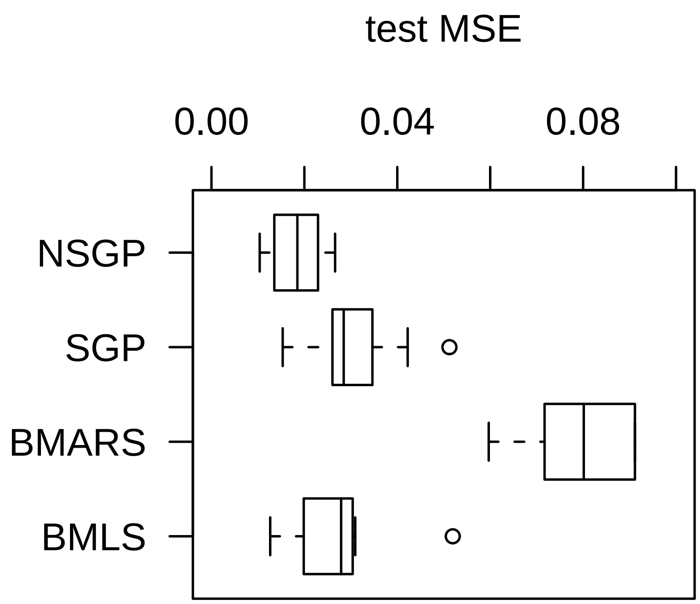
- Compare performance to:
  - ❖ stationary spatial model
  - ❖ Bayesian models in which $f(\cdot)$ is a spline
    - ✥ BMARS (Denison, Mallick & Smith 1998) - tensor products of univariate splines
    - ✥ BMLS (Holmes & Mallick 2001) - multivariate linear splines
  - ❖ spatial deformation approach (Sampson, Guttorp, et al.) (software?)

RESULTS - SIMULATED NONSTATIONARY FUNCTION

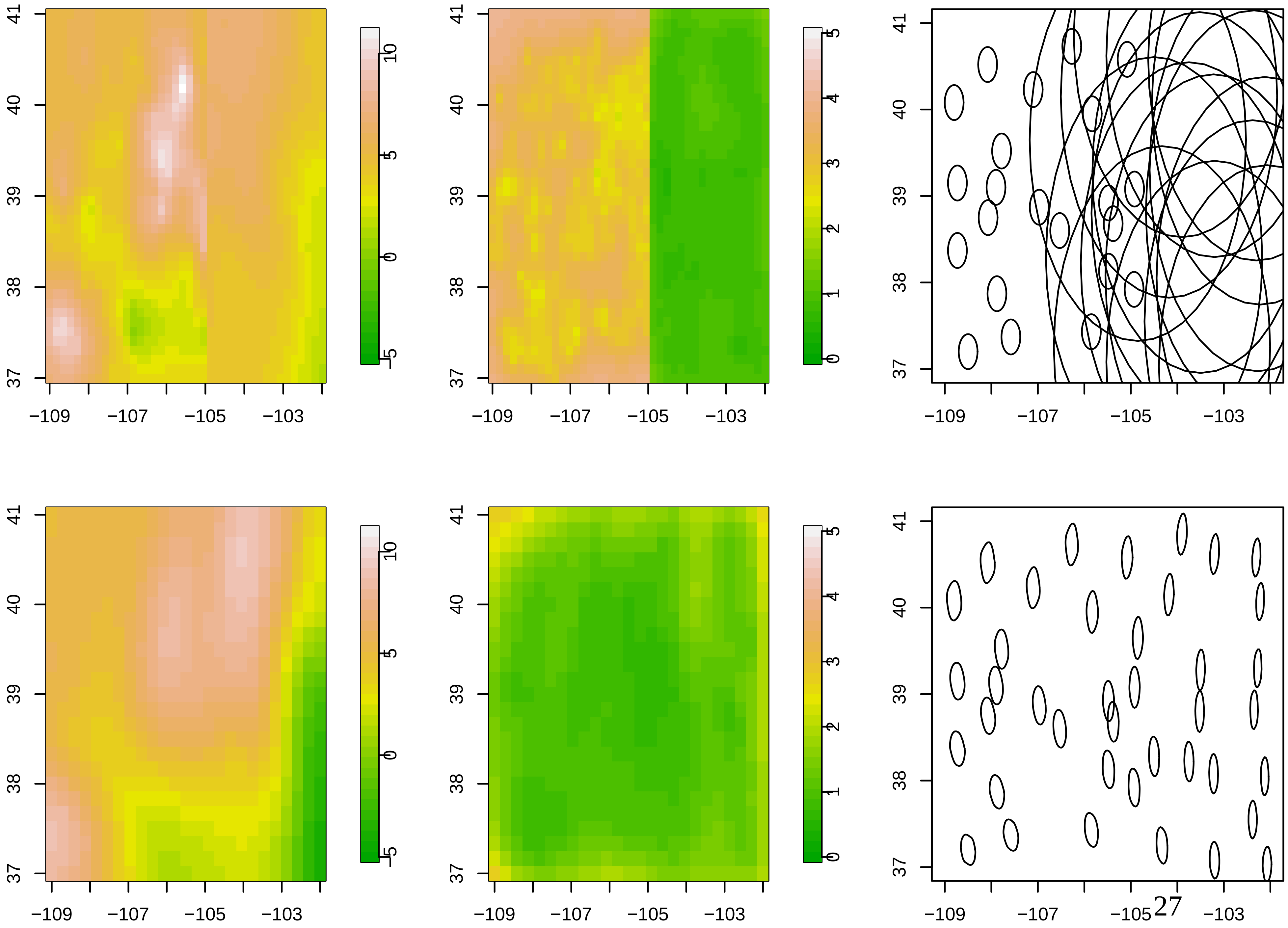


Results - simulated nonstationary function (2)

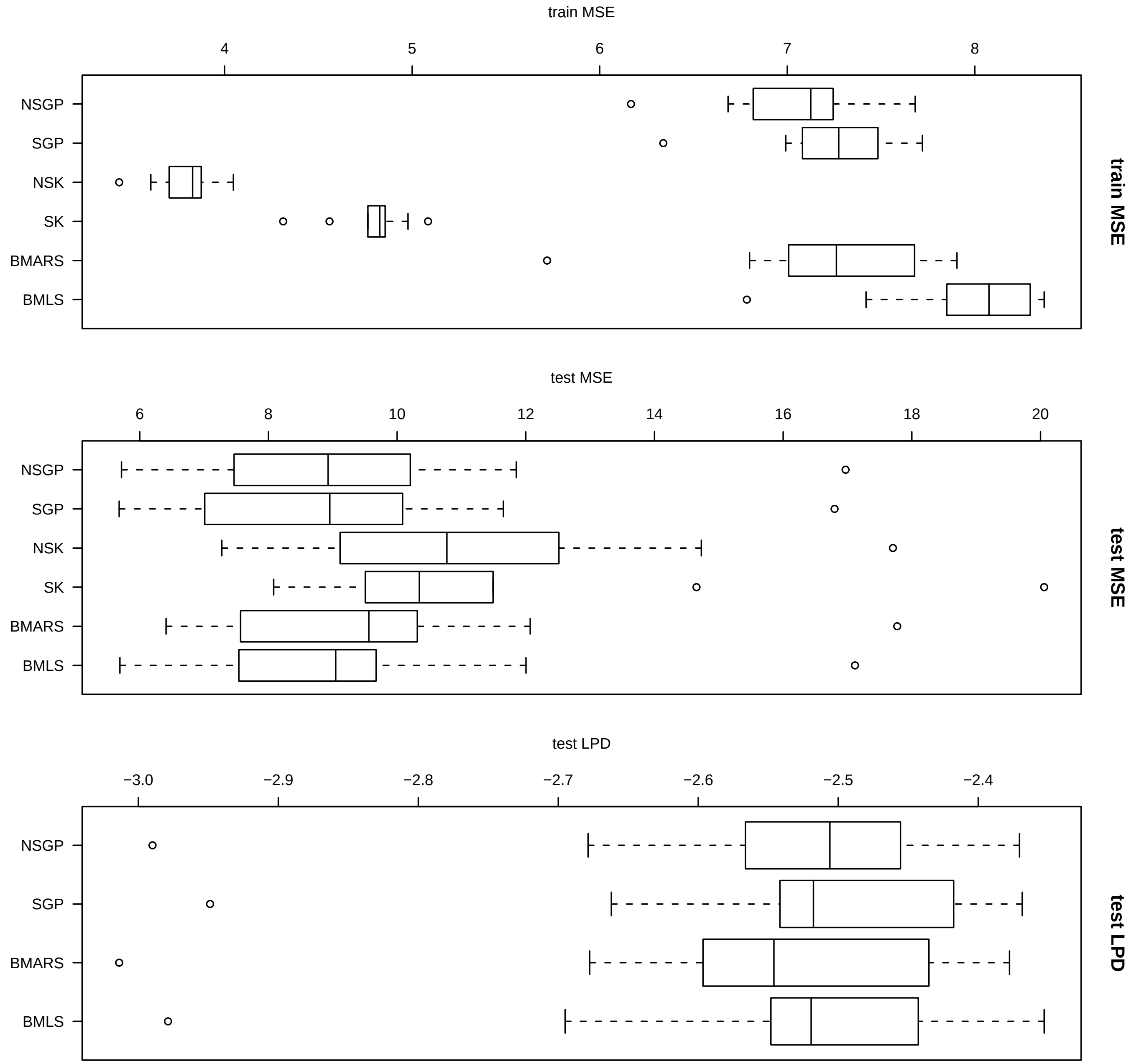




RESULTS - COLORADO PRECIPITATION



Results - Colorado precipitation (2)



REPRESENTATIONS OF STATIONARY PROCESSES

- Applications
  - ❖ Eigenprocesses in the nonstationary model
  - ❖ Function representation in generalized spatial modelling
    $\boldsymbol{Y} \sim \boldsymbol{D}(\boldsymbol{g}^{-1}(\boldsymbol{X\beta} + \boldsymbol{\Phi}(\boldsymbol{x})))$
- Goals
  - ❖ Efficient computation
  - ❖ Close approximation to the covariance structure of a stationary GP

MATÉRN BASIS FUNCTIONS (KAMMANN & WAND)

- $\Phi = \mu + \sigma Z \Omega^{-\frac{1}{2}} u$
  - ❖ $Z = \left(C(\| x_i - \kappa_k \|)\right)_{,1 \le i \le n, 1 \le k \le K}$
  - ❖ $\Omega = \left(C(\| \kappa_j - \kappa_k \|)\right)_{,1 \le j \le K, 1 \le k \le K}$
  - ❖ $C(\cdot)$ a stationary covariance function
- matrix operations based on K knots, so more efficient
- motivation: if $\{\kappa_k\} = \{x_i\}$, Cov$(\Phi(\cdot)) = C(\cdot)$

Fourier basis functions (Wikle)

- $\boldsymbol{\Phi}_{\text{dat}} = \boldsymbol{\mu} + \boldsymbol{\sigma} \boldsymbol{A} \boldsymbol{\Phi}_{\text{grid}}$
- $\boldsymbol{\Phi}_{\text{grid}} = \boldsymbol{\Psi} \boldsymbol{u}$ (discretized process)
- $\boldsymbol{u}$ elements are independent, complex-valued RVs
  - ❖ variance based on spectral density of stationary $\boldsymbol{C}(\cdot)$
- $\boldsymbol{\Psi} \boldsymbol{u}$ is the inverse FFT ($\boldsymbol{\Psi}$ are Fourier basis vectors)
- propose blocks of values of $\boldsymbol{u}$ with focus on low-frequency coefficients

APPLICATION TO THE NONSTATIONARY MODEL

- Represent each eigenprocess as a stationary GP
- Fix some hyperparameters
  - Fix $\boldsymbol{\kappa}$ and let $\boldsymbol{\sigma}$ do the smoothing
  - Fix $\boldsymbol{\sigma}$ and let $\boldsymbol{\kappa}$ do the smoothing
- When sample hyperparameters, sample eigenprocess as well:
  - $\boldsymbol{\Phi} = \boldsymbol{\mu} + \boldsymbol{\sigma} \boldsymbol{B}(\boldsymbol{\kappa}, \boldsymbol{\nu}) \boldsymbol{u}$

Application to Public Health Data

- Features of spatial disease modelling
  - ❖ case-control binary outcomes common
  - ❖ relatively large sample sizes (100s to 1000s)
  - ❖ models must include individual-level covariates
  - ❖ $\boldsymbol{Y} \sim \mathrm{Ber}(\mathrm{logit}^{-1}(\boldsymbol{X\beta} + \boldsymbol{\Phi}(\boldsymbol{x})))$
- FFT approach is particularly efficient
  - ❖ for fixed grid, likelihood scales as $\boldsymbol{O}(\boldsymbol{n})$
- Work in progress to compare to frequentist estimation:
  - ❖ mixed model approximation (penalized quasi-likelihood)
  - ❖ efficient thin-plate splines (Simon Wood, R mgcv library)

Conclusions & Comments

- Generalized HSK nonstationary covariance provides a family of nonstationary covariances
- Accounting for nonstationarity can improve fit
- Limitations: sharp changes in function, boundaries and other complicated structure not well captured
- Computational speed and mixing remain issues
- Nonstationary GPs can be used in general nonparametric regression setting