
Nonstationary Covariance Functions for Spatial Modelling

Chris Paciorek

paciorek@hsph.harvard.edu

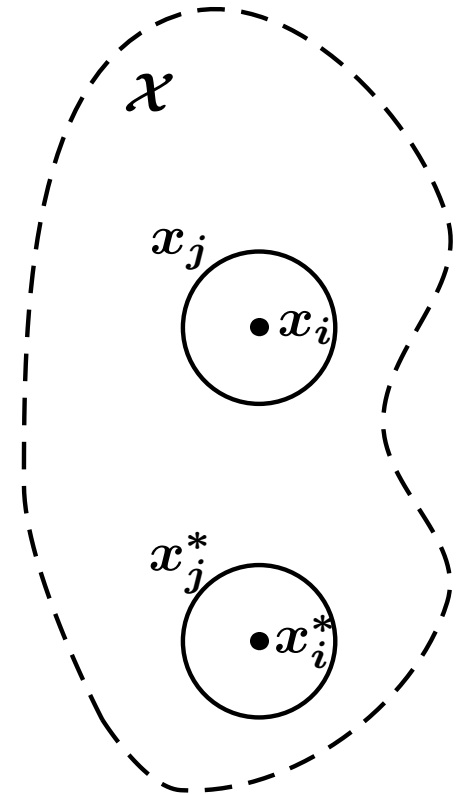
January 16, 2004

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

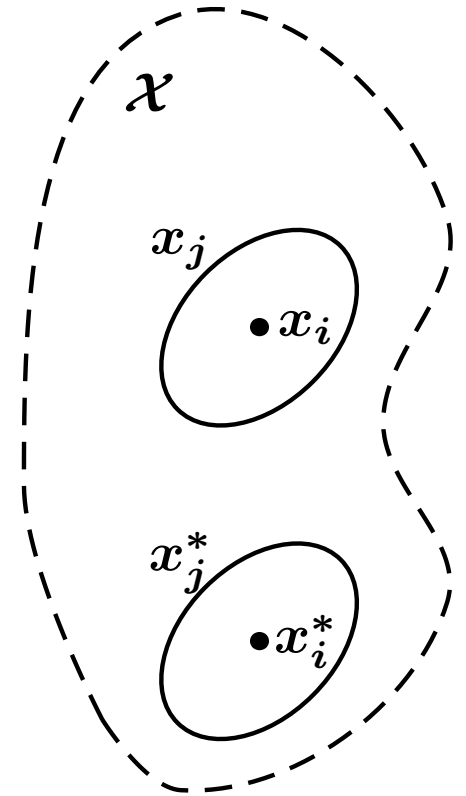
GAUSSIAN PROCESS DISTRIBUTION

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



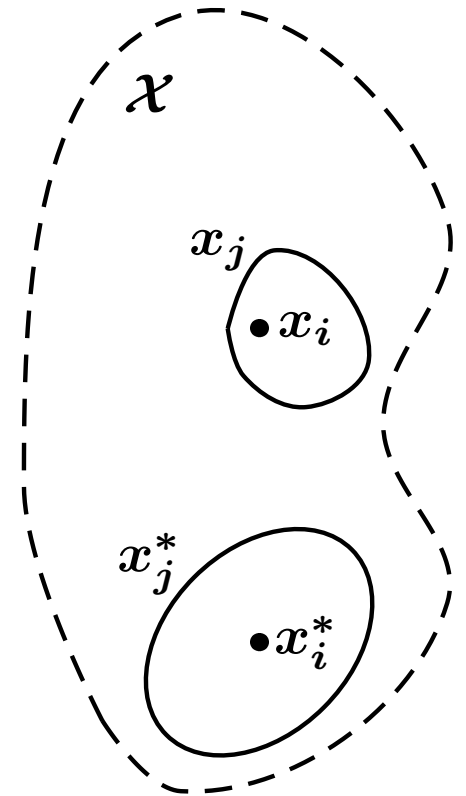
GAUSSIAN PROCESS DISTRIBUTION

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



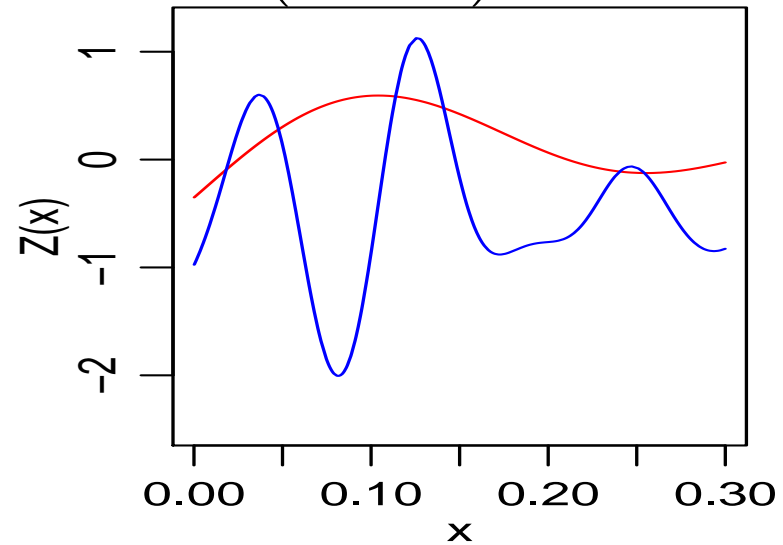
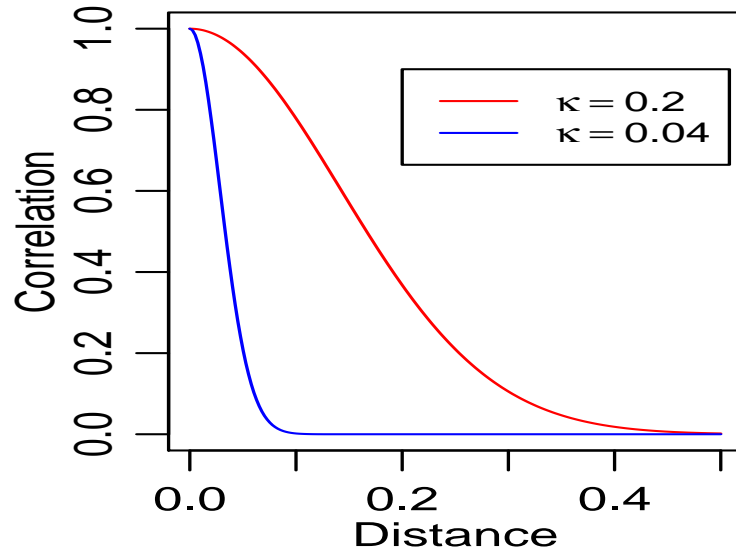
GAUSSIAN PROCESS DISTRIBUTION

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

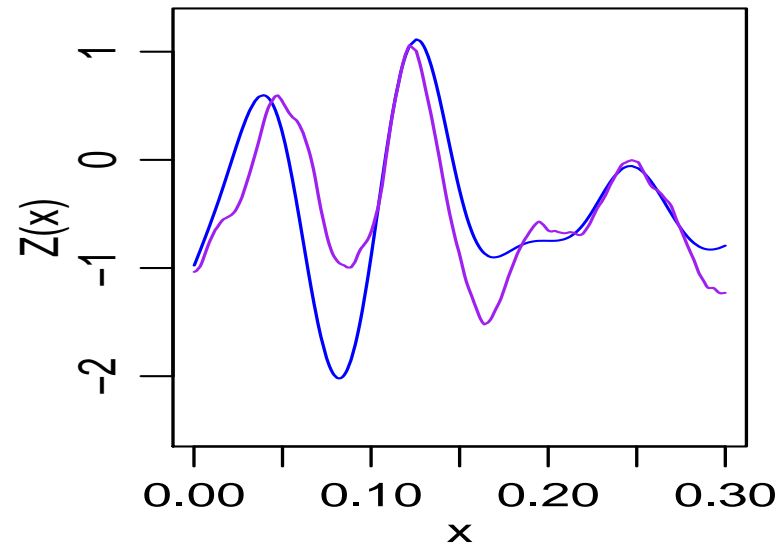
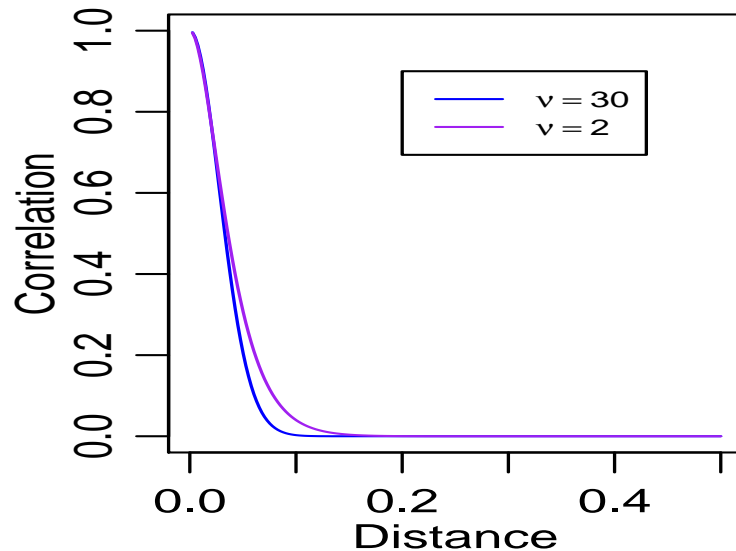


STATIONARY CORRELATION FUNCTIONS

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

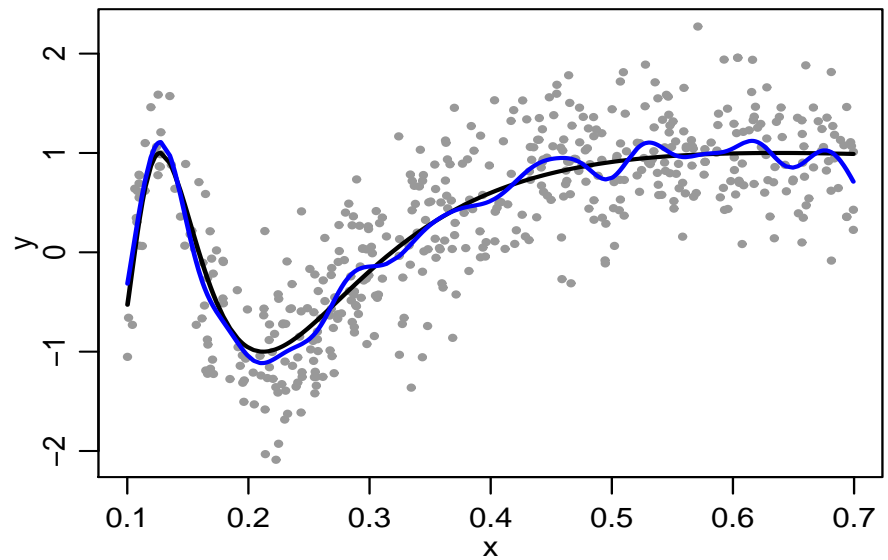
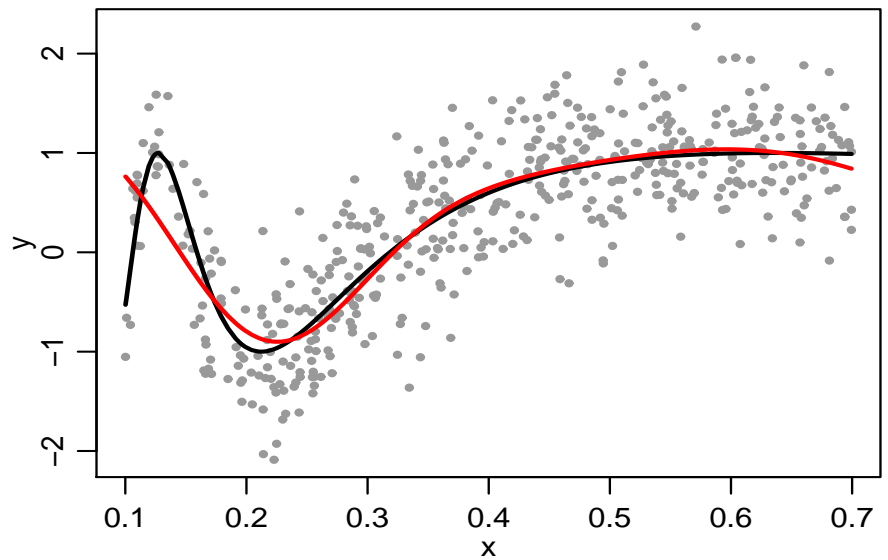
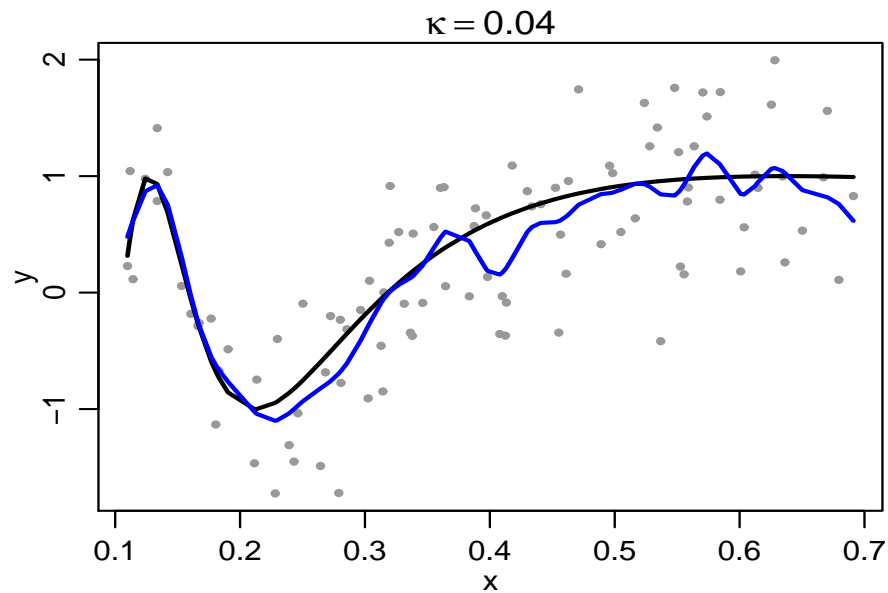
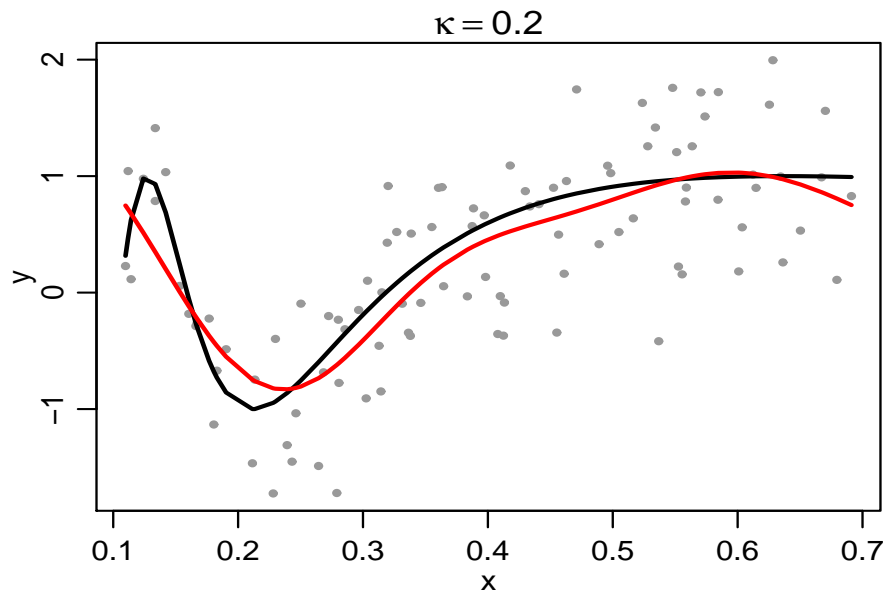


Matérn form: $R(\tau) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{2\sqrt{\nu}\tau}{\kappa}\right)^\nu K_\nu\left(\frac{2\sqrt{\nu}\tau}{\kappa}\right)$



- Differentiability controlled by ν , asymptotic advantages (Stein)

DEGREE OF SMOOTHING



A NONSTATIONARY COVARIANCE

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

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

- Guaranteed positive definite
- Gaussian kernels:

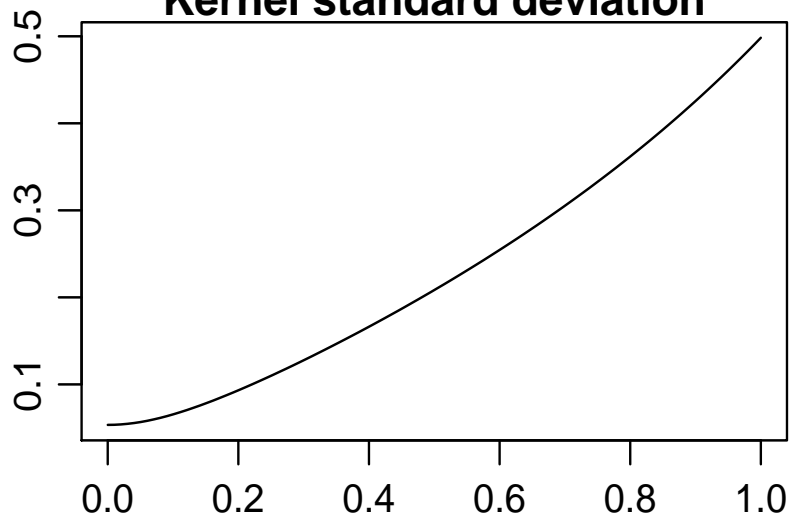
$$k_{x_i}(u) \propto \exp\left(- (u - x_i)^T \Sigma_i^{-1} (u - x_i)\right)$$

$$R^{NS}(x_i, x_j) = c_{ij} \exp\left(- (x_i - x_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1} (x_i - x_j)\right)$$

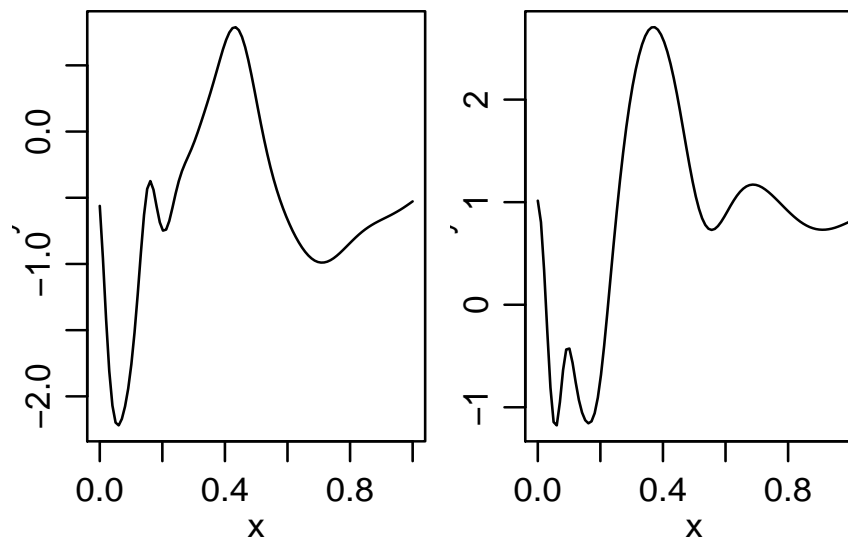
- $f(\cdot) \sim \text{GP}(\mu, \sigma^2 R^{NS}(\cdot, \cdot))$

NONSTATIONARY GPs IN 1D

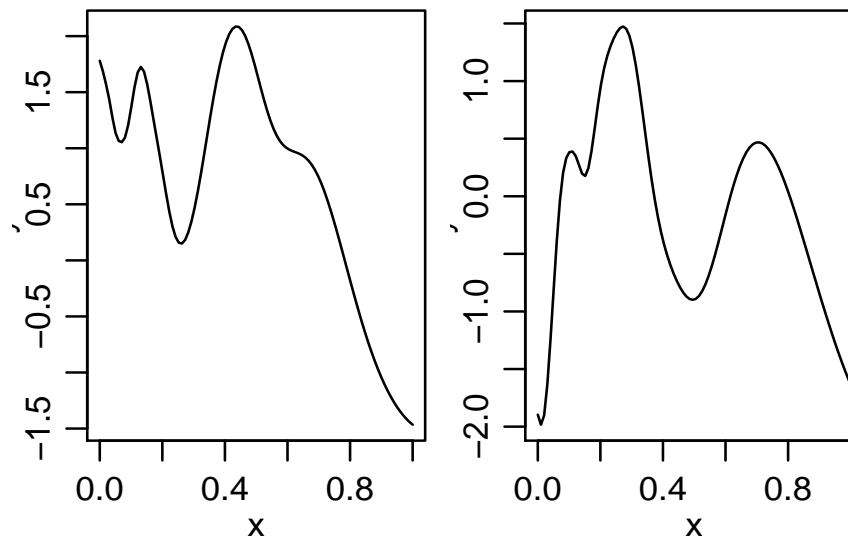
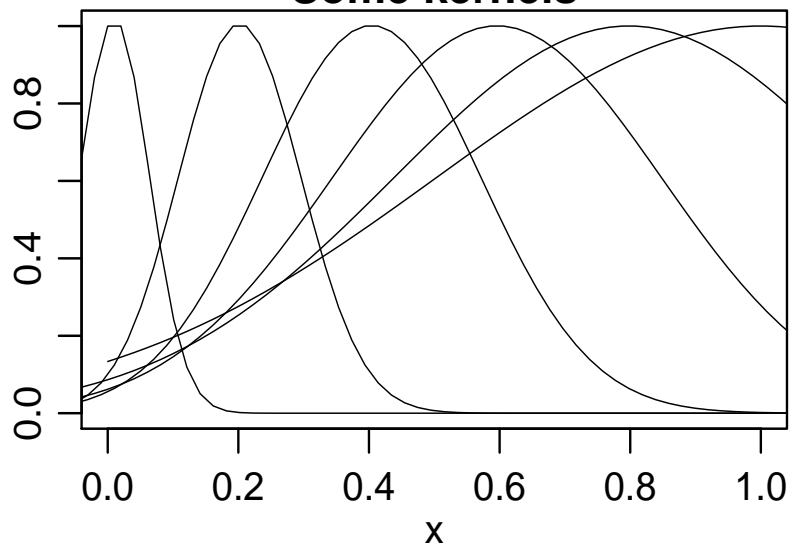
Kernel standard deviation



Some sample functions

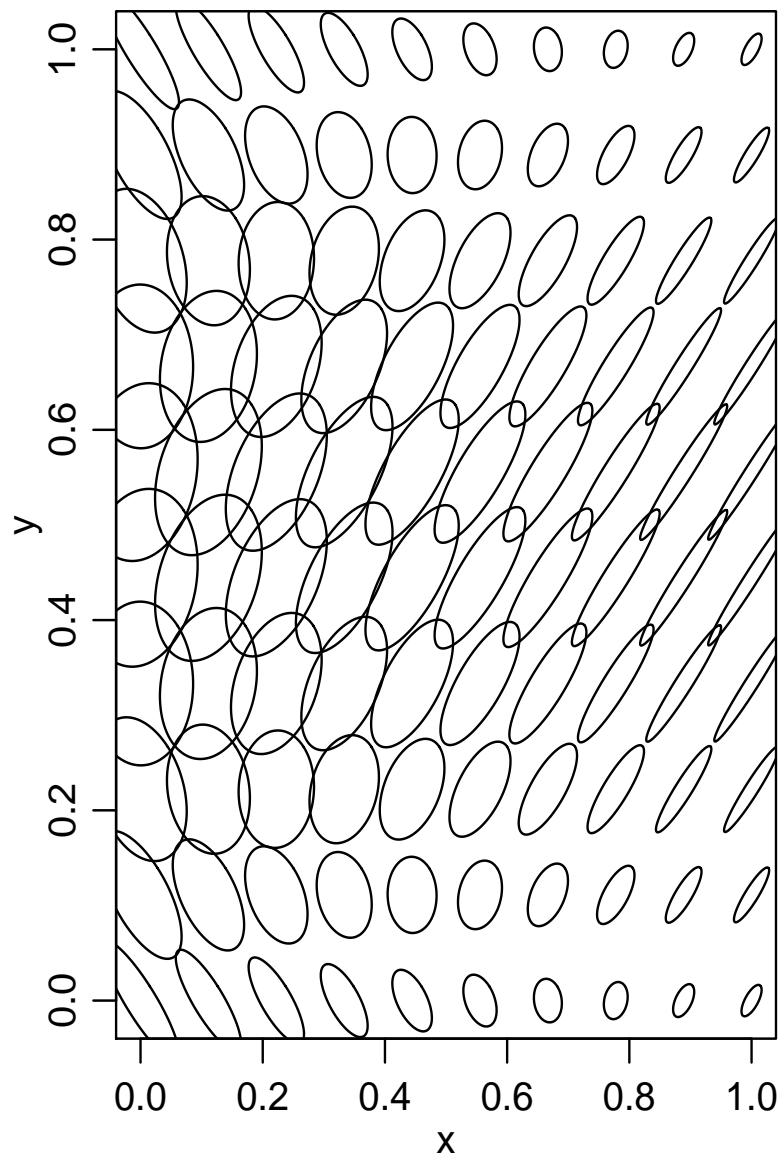


Some kernels

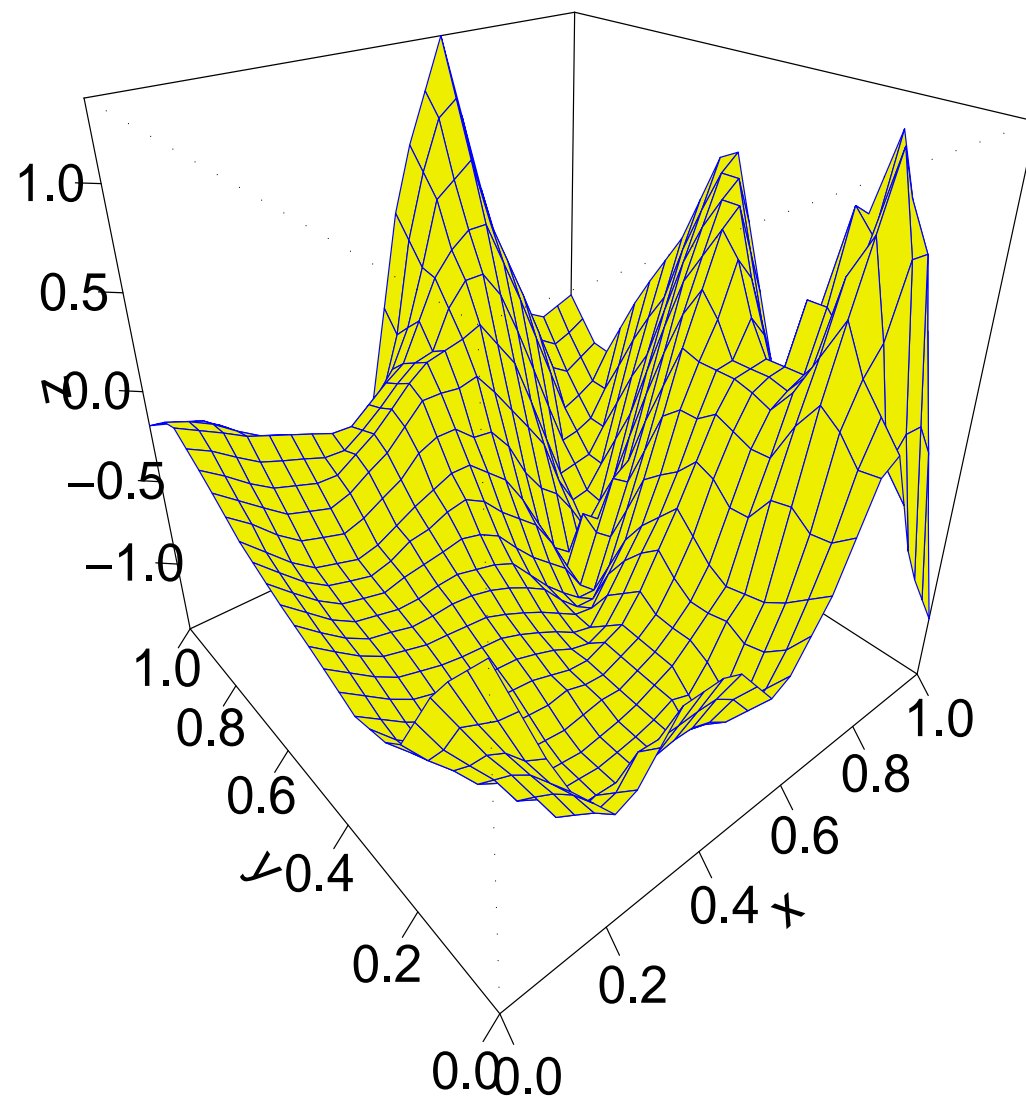


NONSTATIONARY GPs IN 2D

Kernel Structure



Sample Function



GENERALIZING THE HSK KERNEL METHOD

- Squared exponential form:

$$\exp\left(-\left(\frac{\tau}{\kappa}\right)^2\right) \Rightarrow c_{ij} \exp\left(-(\mathbf{x}_i - \mathbf{x}_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1} (\mathbf{x}_i - \mathbf{x}_j)\right)$$

Infinitely-differentiable sample paths

- ‘Distance measures’

isotropy $\tau_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)$

anisotropy $\tau_{ij}^{*2} = (\mathbf{x}_i - \mathbf{x}_j)^T \Sigma^{-1} (\mathbf{x}_i - \mathbf{x}_j)$

nonstationarity $Q_{ij} = (\mathbf{x}_i - \mathbf{x}_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1} (\mathbf{x}_i - \mathbf{x}_j)$

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

GENERALIZED NONSTATIONARY COVARIANCE

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

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

is positive definite for \mathfrak{R}^P , $P = 1, 2, \dots$

- Theorem 2: Smoothness properties of original stationary correlation retained

PROOF (SKETCH)

•

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

•

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

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

- Covariance must depend only on location-specific kernels

NONSTATIONARY MATÉRN COVARIANCE

stationary form

nonstationary form

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

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

NONSTATIONARY KRIGING

- basic kriging model: $Y \sim N(\mu, (\sigma^2 R_f(\kappa, \nu) + \eta^2 I))$
- $\sqrt{Q_{ij}}$ is an anisotropic distance if $\Sigma_i = \Sigma_j$
- To kriging, estimate parameters of $\Sigma_{(\cdot)}$ 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

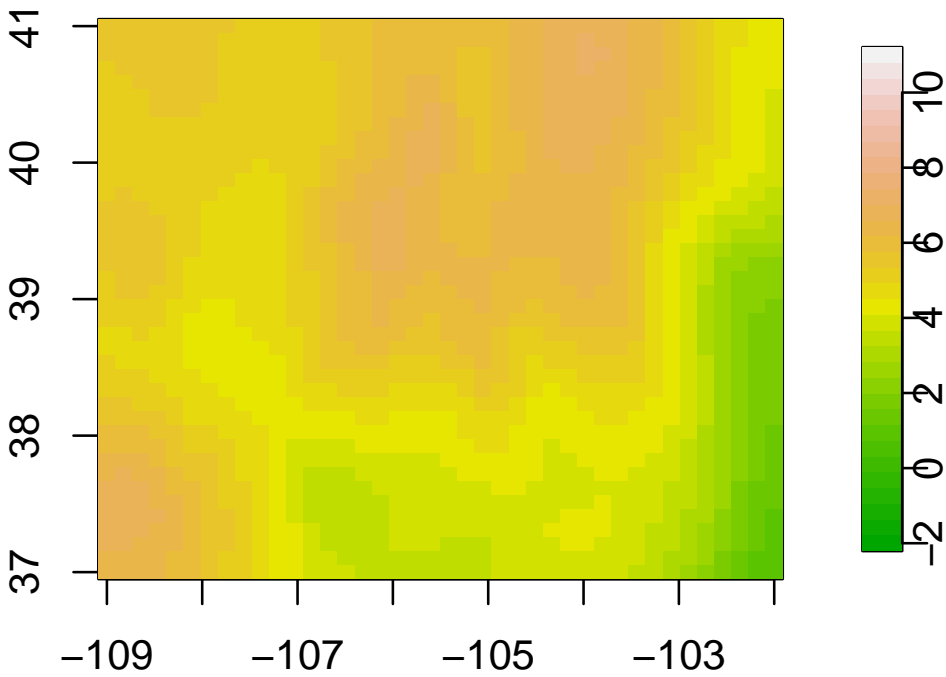
- precipitation anomalies in Colorado, August 1963
- fit covariance structure by maximizing marginal likelihood (EB):

	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

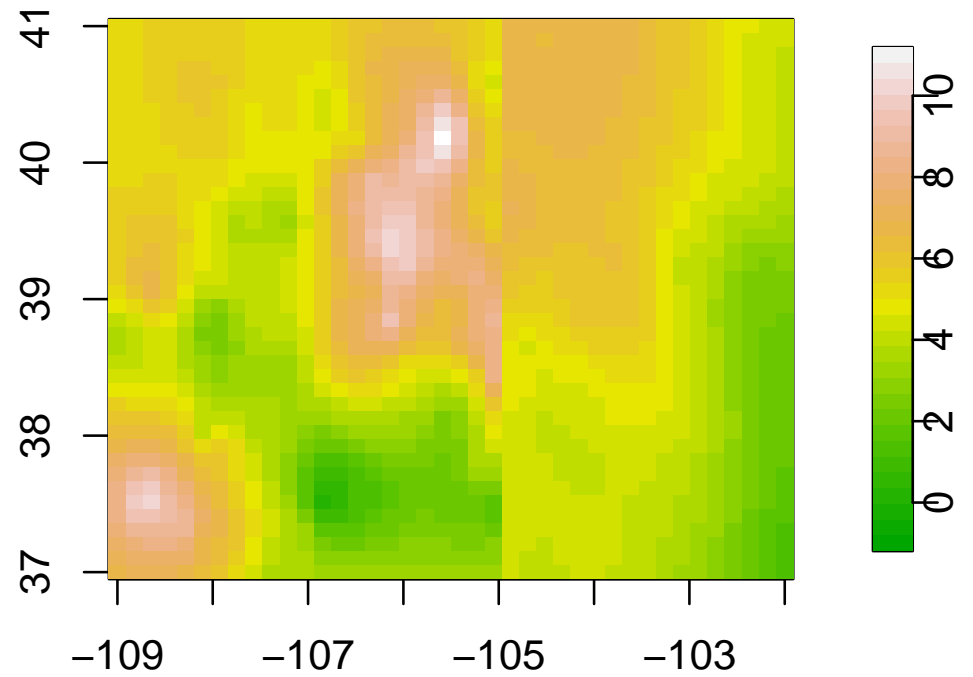
- could also use standard variogram fitting

COLORADO PRECIPITATION ANOMALIES

Stationary kriging

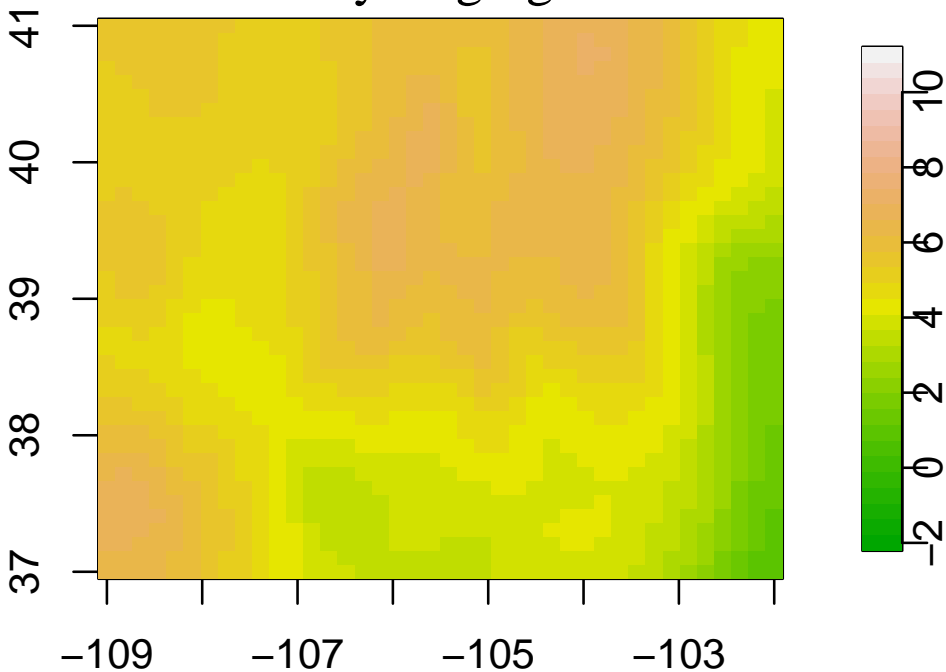


Nonstationary kriging

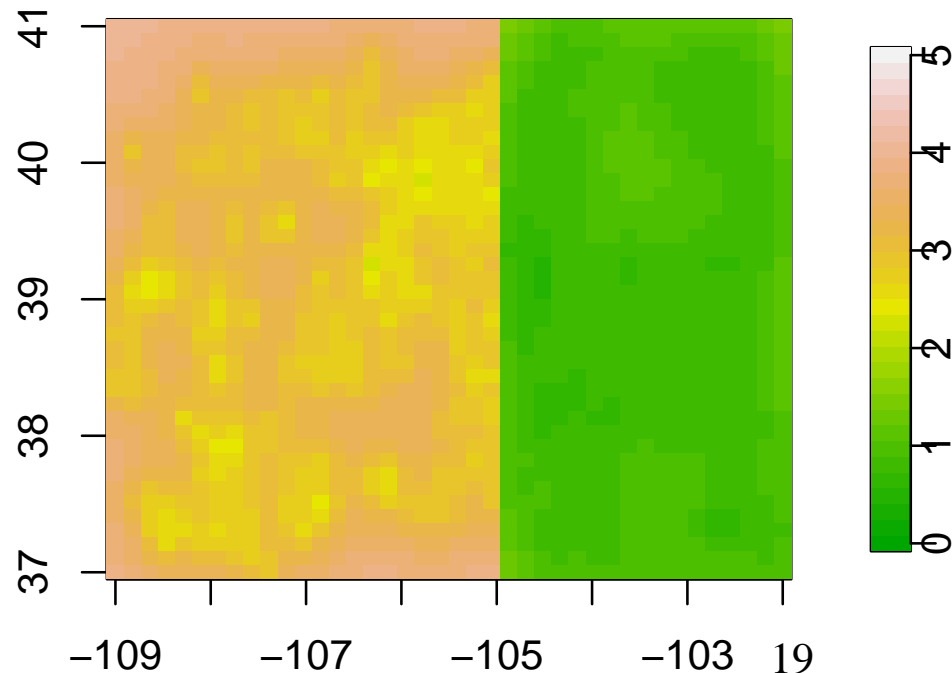
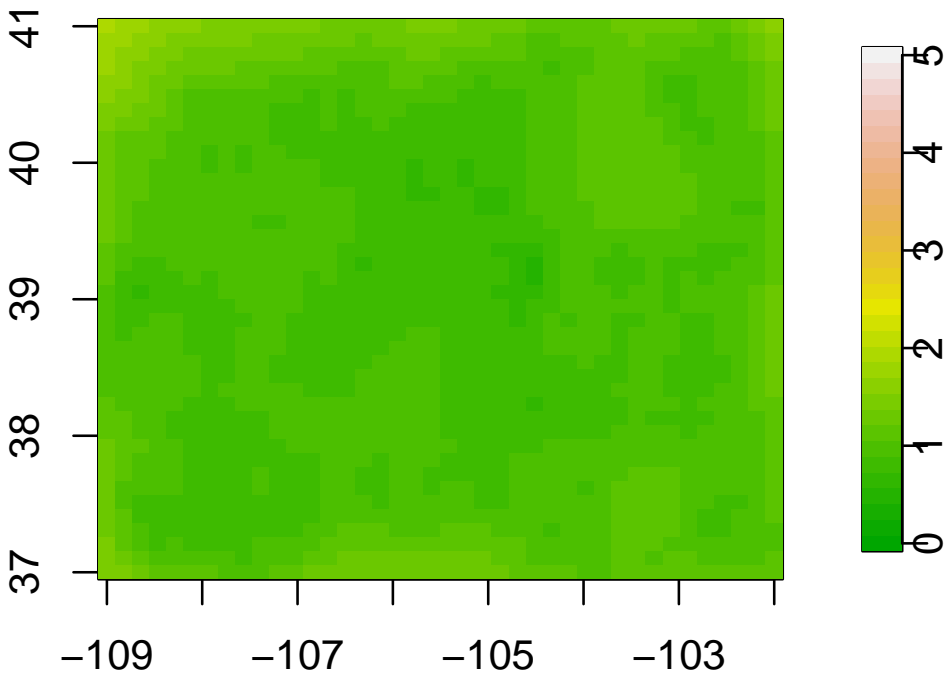
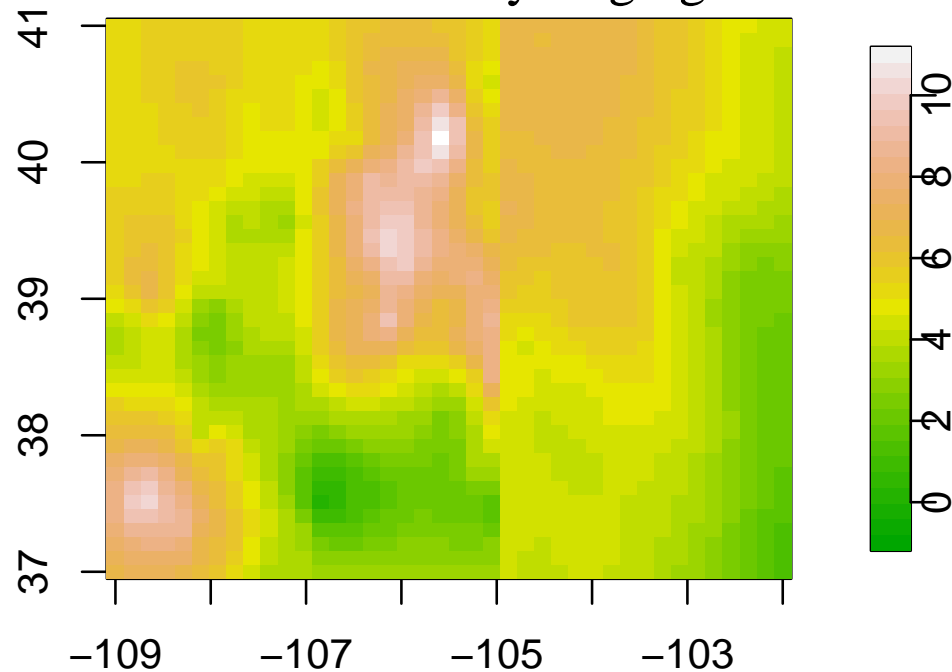


COLORADO PRECIPITATION ANOMALIES

Stationary kriging

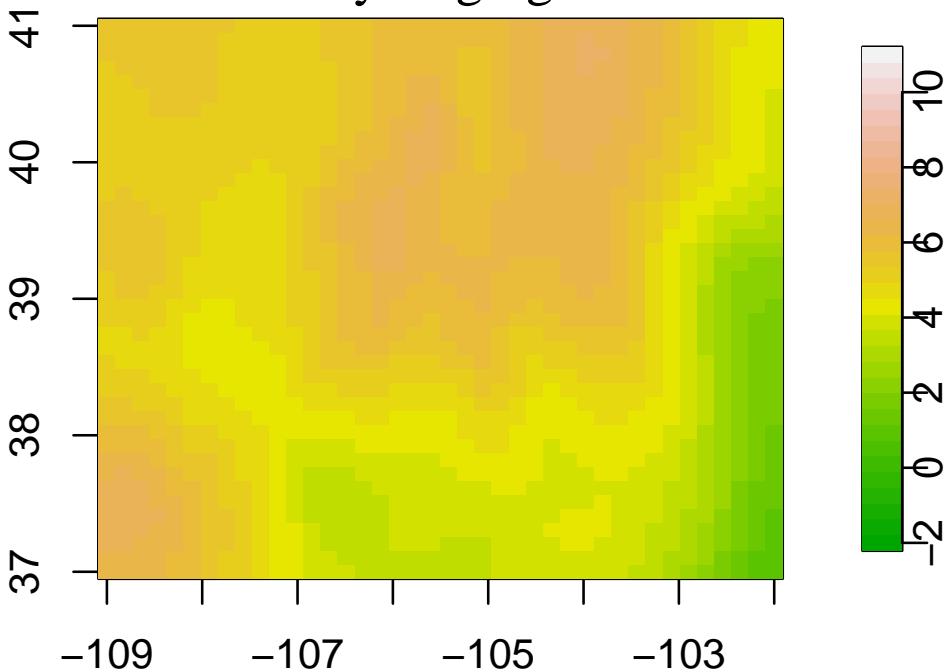


Nonstationary kriging

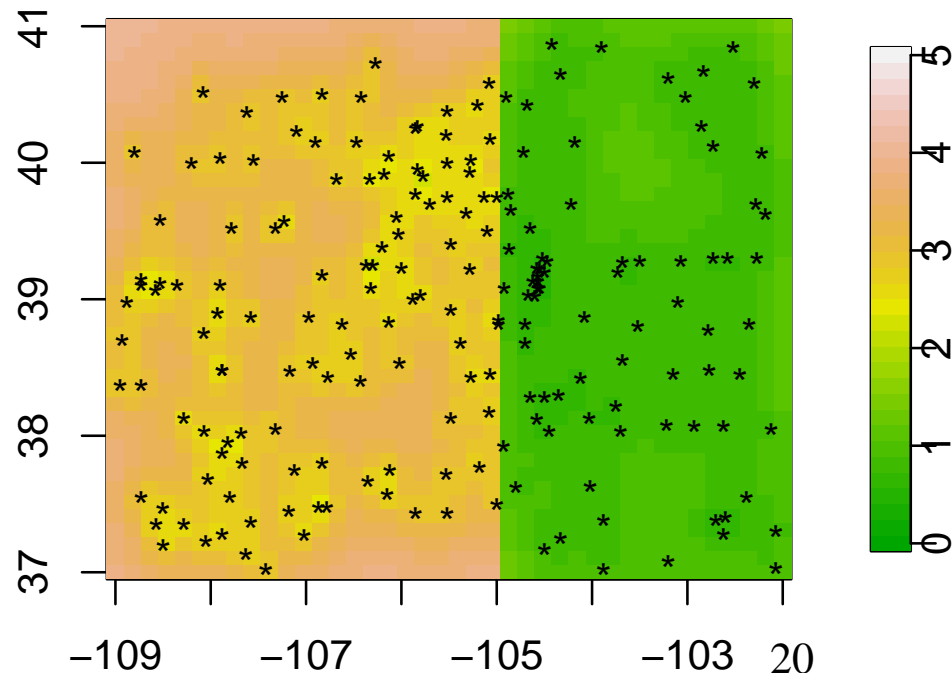
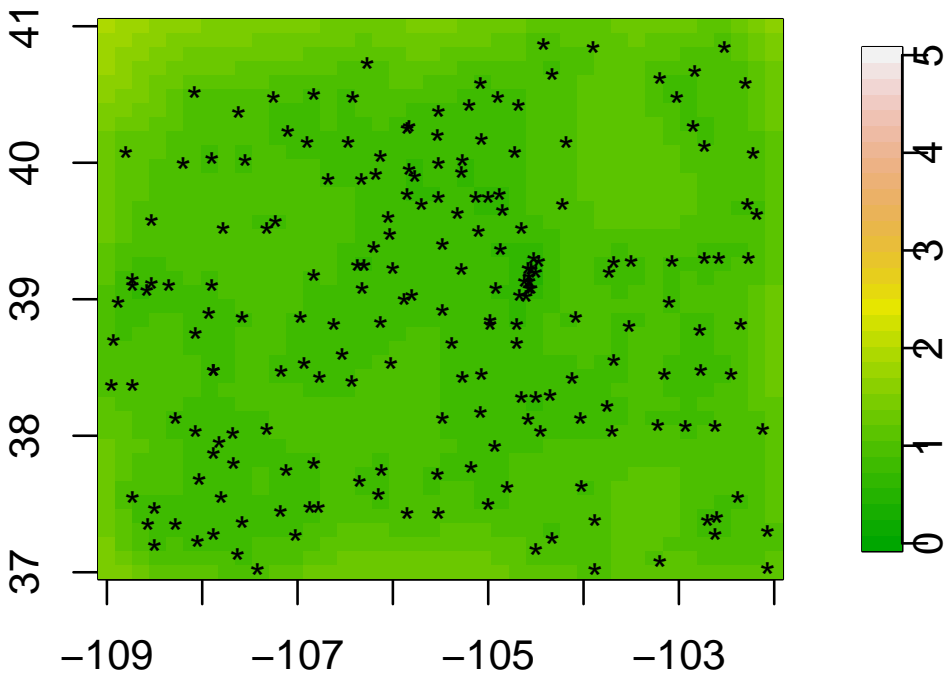
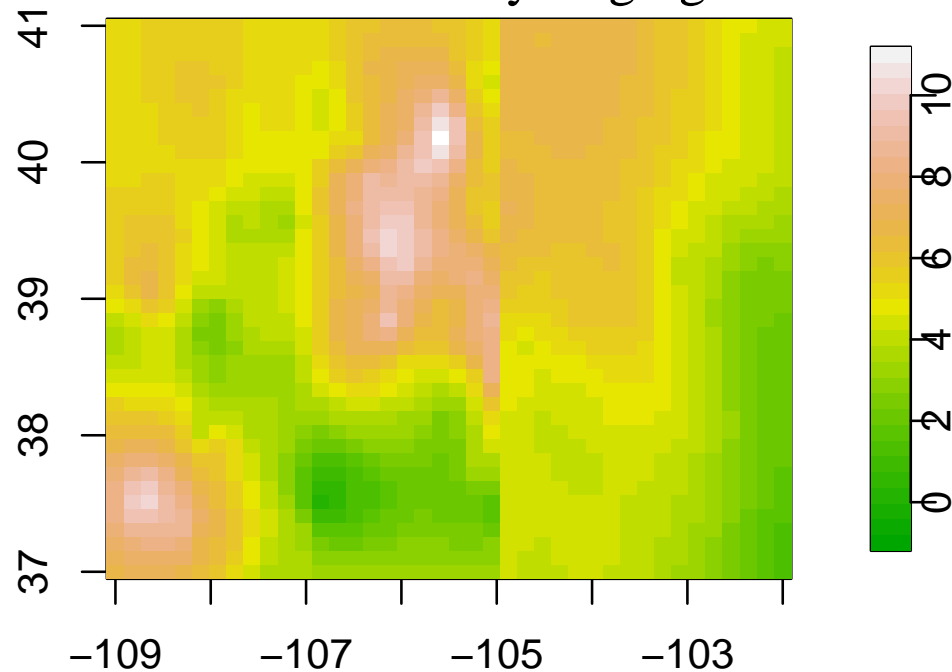


COLORADO PRECIPITATION ANOMALIES

Stationary kriging



Nonstationary kriging



A BASIC BAYESIAN SPATIAL MODEL

- Bayesian model

$$Y_i \sim N(f(x_i), \eta^2), x_i \in \mathbb{R}^2$$
$$f(\cdot) \sim \text{GP}(\mu, \sigma^2 R^{NS}(\cdot, \cdot; \nu, \Sigma_{(\cdot)}))$$

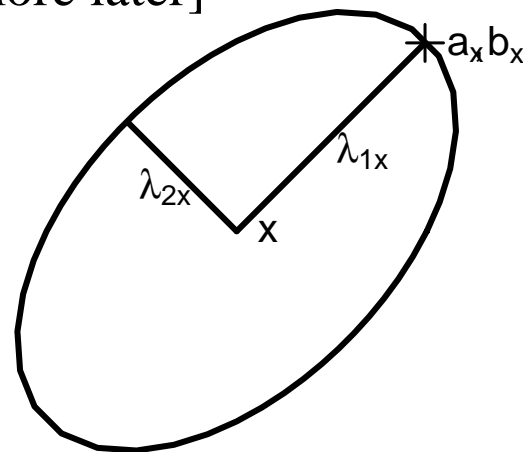
- ❖ Let R^{NS} be the nonstationary Matérn correlation
- ❖ Kernels (Σ_x) constructed based on stationary GP priors

SMOOTHLY-VARYING KERNEL MATRICES

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

SMOOTHLY-VARYING KERNEL MATRICES (2)

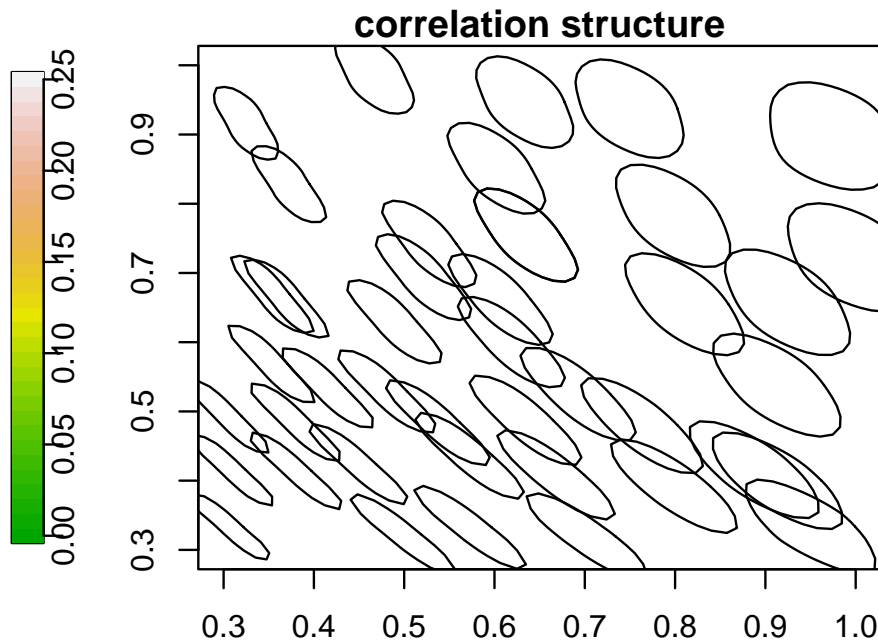
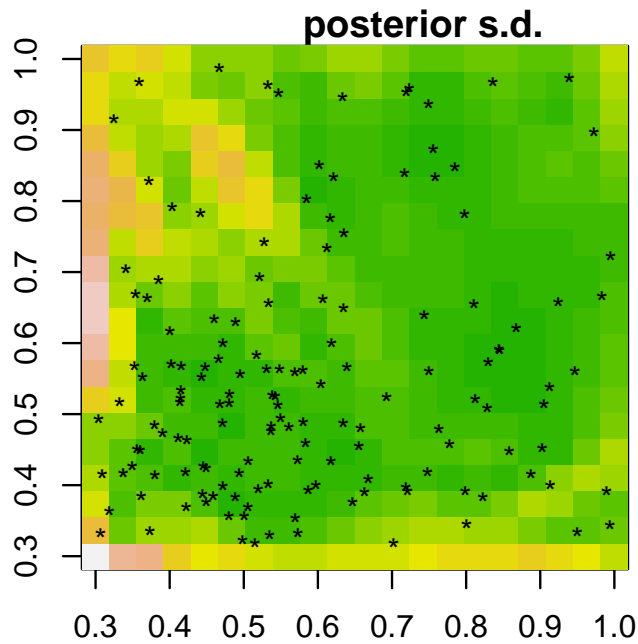
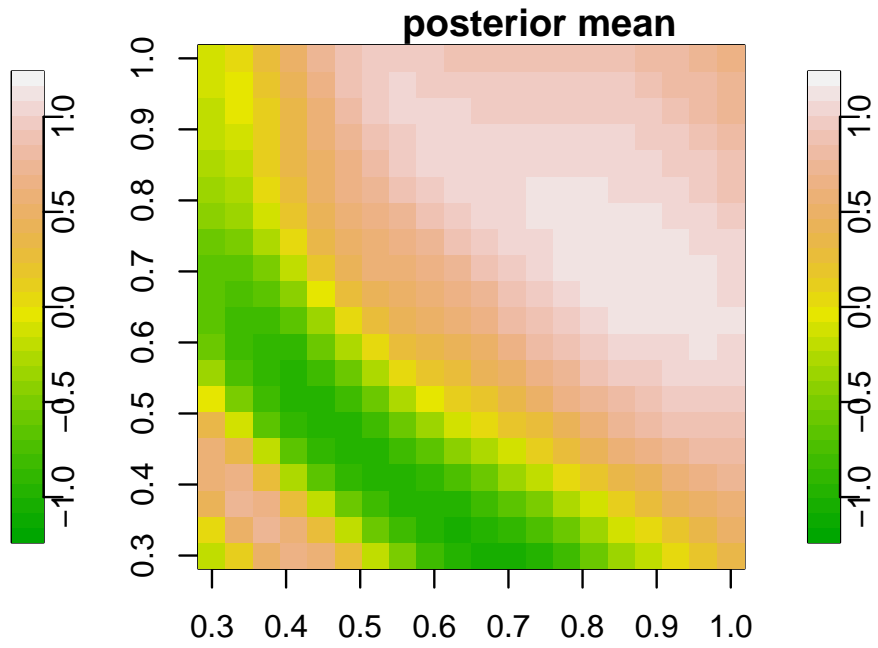
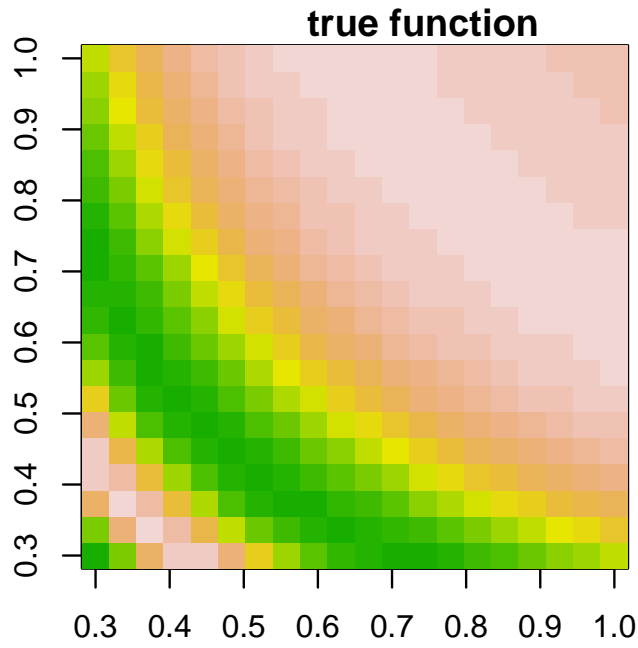
- Spectral decomposition (\mathfrak{R}^P): $\Sigma_x = \Gamma_x^T \Lambda_x \Gamma_x$
 - ❖ in \mathfrak{R}^P , Γ_x parameterized as first eigenvector plus successive orthogonal vectors in reduced-dimension subspaces
 - ❖ in \mathfrak{R}^2 , stationary GP priors on unnormalized eigenvector coordinates (a_x, b_x) and on logarithm of eigenvalues $(\lambda_{x,1}, \lambda_{x,2})$
 - ❖ efficient parameterizations of stationary GPs for $\Phi(\cdot) \in \{a(\cdot), b(\cdot), \lambda_1(\cdot), \lambda_2(\cdot)\}$ [more later]



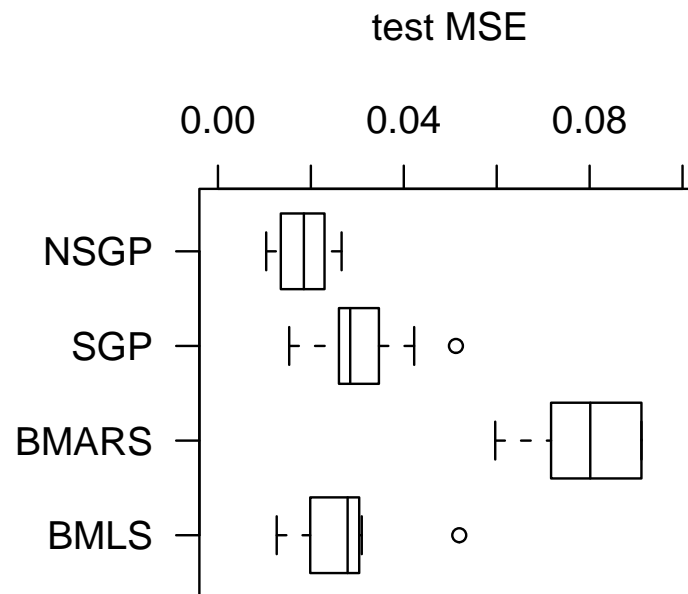
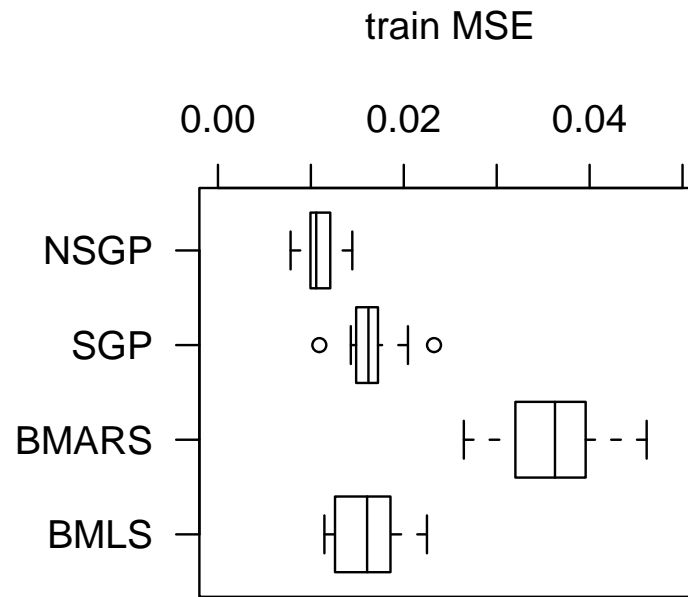
EMPIRICAL ASSESSMENT

- Compare performance to:
 - ❖ stationary GP 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

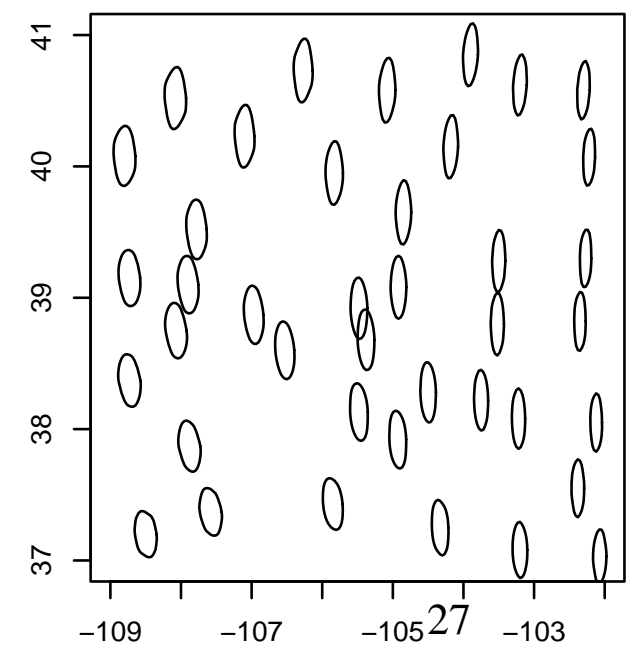
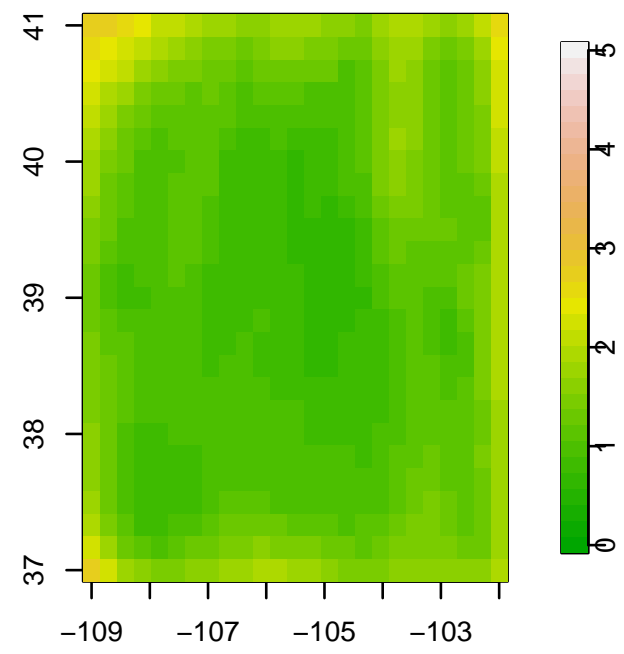
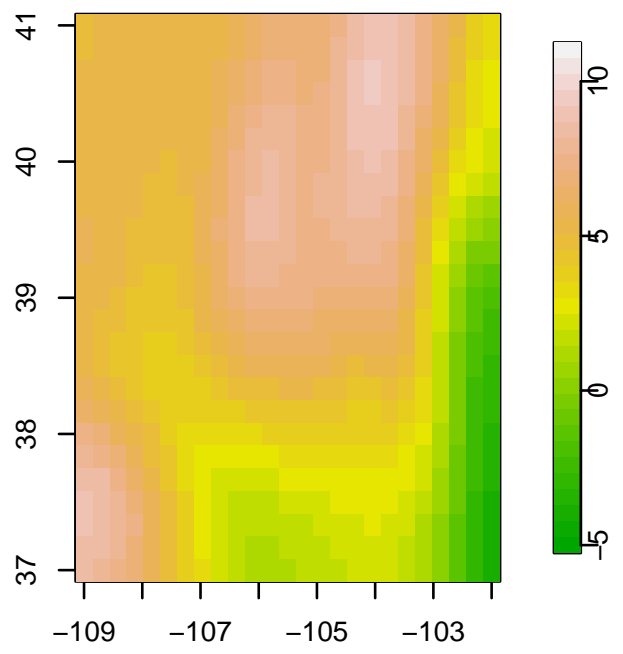
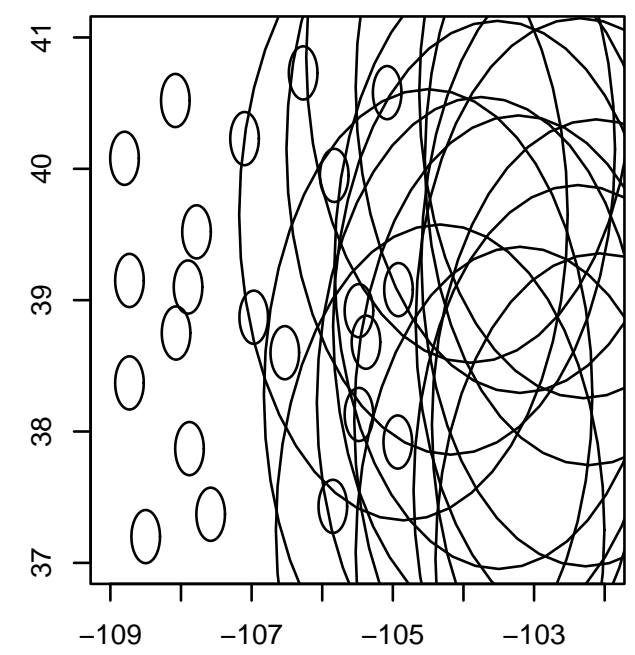
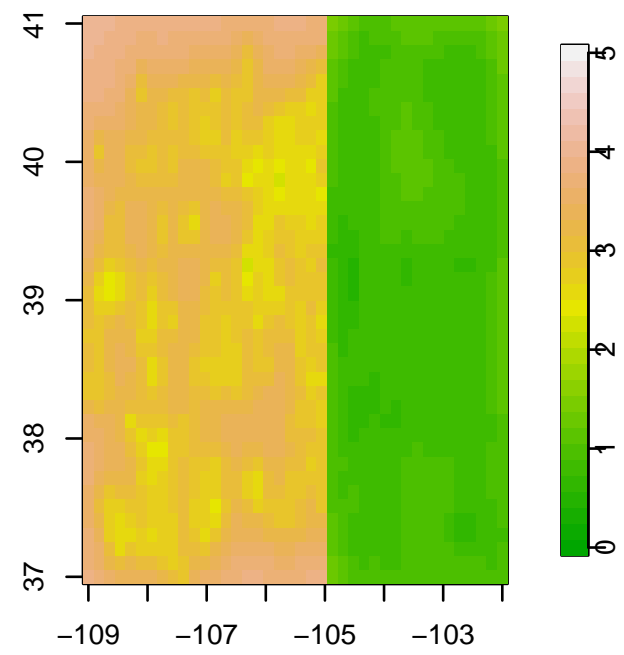
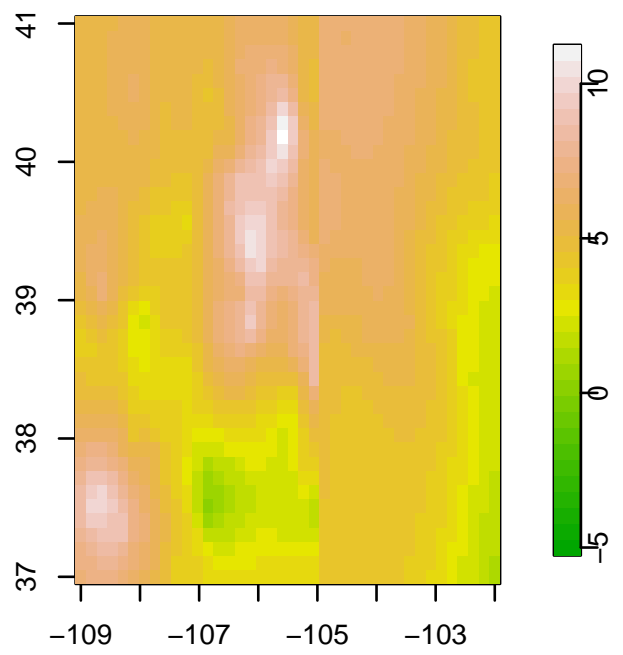
RESULTS - SIMULATED NONSTATIONARY FUNCTION



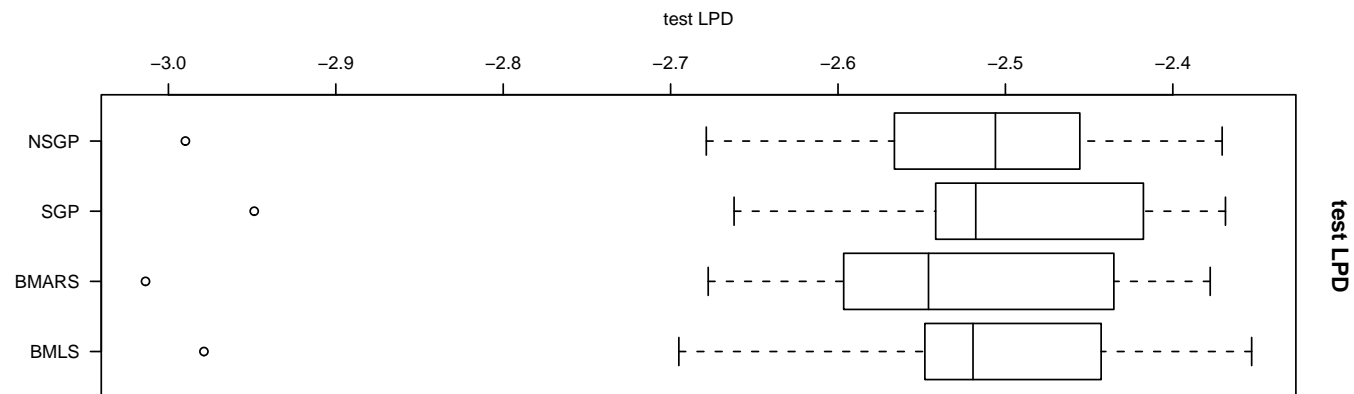
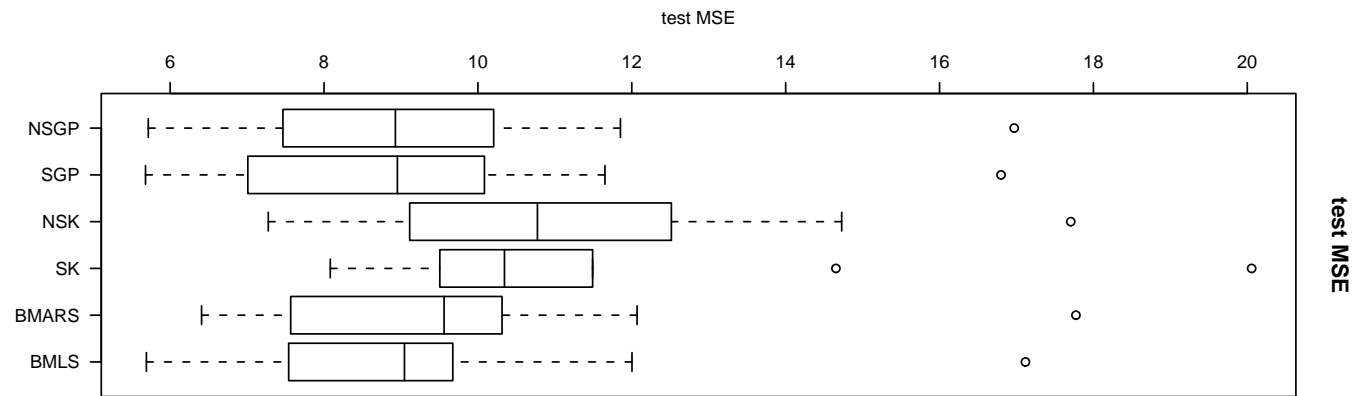
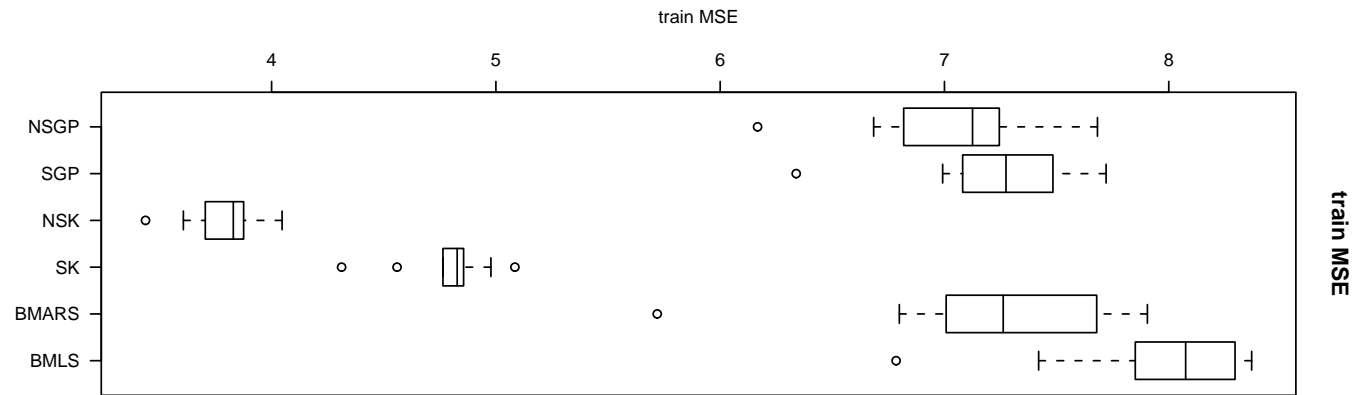
RESULTS - SIMULATED NONSTATIONARY FUNCTION (2)



RESULTS - COLORADO PRECIPITATION



RESULTS - COLORADO PRECIPITATION (2)



ONGOING ASSESSMENT

- Compare performance on several datasets to:
 - ❖ stationary spatial models
 - ❖ kriging
 - ❖ thin-plate spline model
 - ❖ mixed model representation of spatial surface
 - ❖ stationary GP model
 - ❖ Bayesian nonstationary models
 - ❖ BMARS (Denison, Mallick & Smith 1998) - tensor products of univariate splines
 - ❖ BMLS (Holmes & Mallick 2001) - multivariate linear splines
 - ❖ neural network model (R. Neal software)

REPRESENTATIONS OF STATIONARY PROCESSES

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

MATÉRN BASIS FUNCTIONS (ZHAO, KAMMANN, WAND)

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

FOURIER BASIS FUNCTIONS (WIKLE)

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

APPLICATION TO THE NONSTATIONARY MODEL

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

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
 - ❖ $Y_i \sim \text{Ber}(\text{logit}^{-1}((\mathbf{X}\boldsymbol{\beta})_i + \Phi(\mathbf{x}_i)))$
 - ❖ stationary models for $\Phi(\cdot)$
- FFT approach is particularly efficient
 - ❖ for fixed grid, likelihood scales as $O(n)$
- Work in progress to compare FFT and mixed model/Matérn basis 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