A Class of Convolution-based Nonstationary Covariance Functions

Chris Paciorek

Department of Biostatistics
Harvard School of Public Health

SAMSI/NCAR Workshop on Spatio-Temporal Statistics

June 2, 2003

ABSTRACT

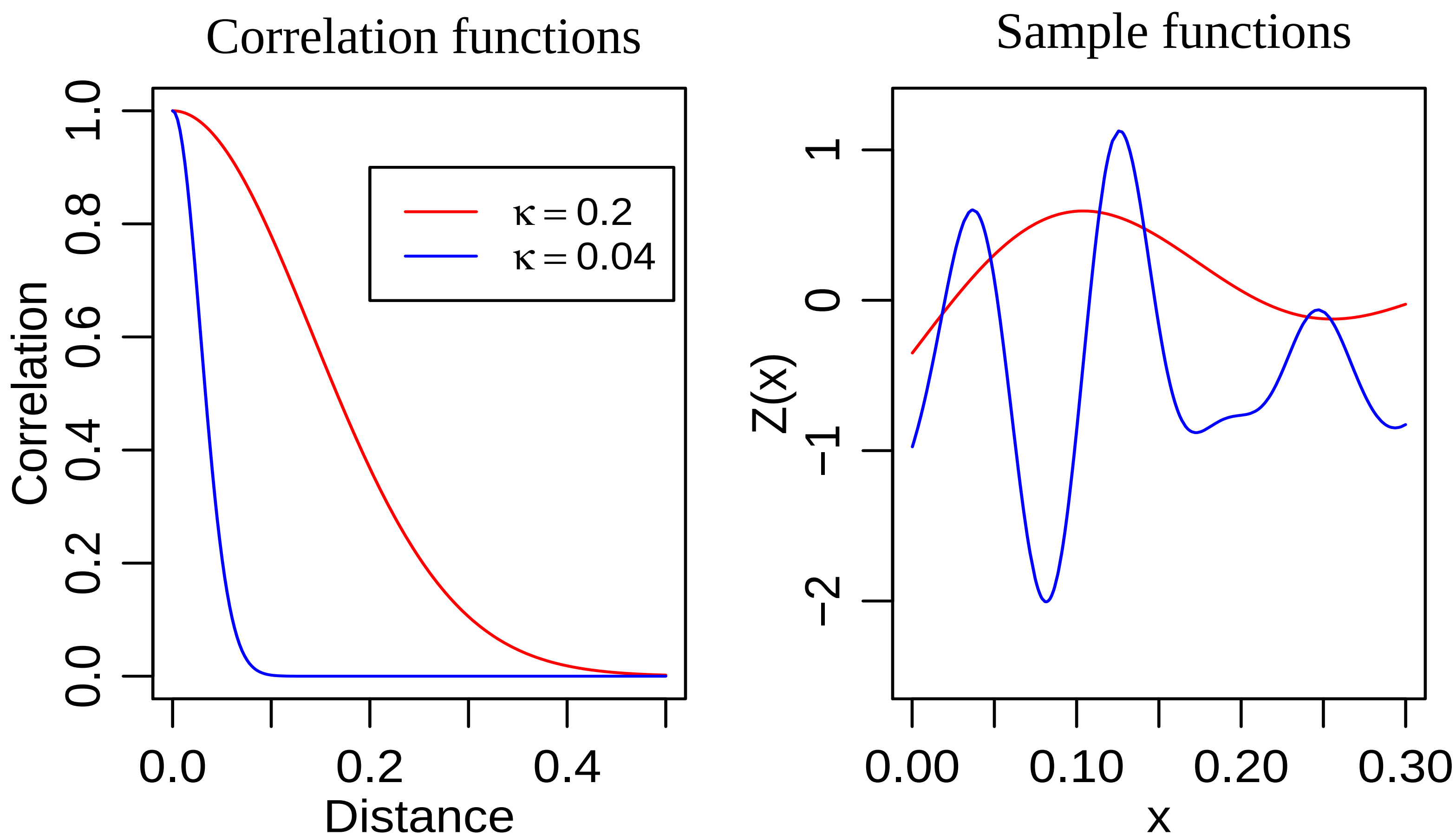
Nonstationary covariance models have received much recent attention in spatial statistics. Higdon et al. (1999) presented a nonstationary covariance function produced by convolving kernel functions. The evolution of the kernel functions in space produces nonstationary covariance. Using Gaussian kernels produces a closed form nonstationary covariance similar in form to the squared exponential (Gaussian) stationary covariance function. These covariance functions, when used in Gaussian processes, produce infinitely differentiable sample paths. I extend the function of Higdon et al. (1999) to produce a class of closed form nonstationary covariance functions that share the sample path differentiability properties of the stationary covariance functions upon which they are based. The class includes a Matérn nonstationary covariance function, which shares the sample path differentiability properties of the stationary Matérn covariance.

GAUSSIAN PROCESS DISTRIBUTION

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

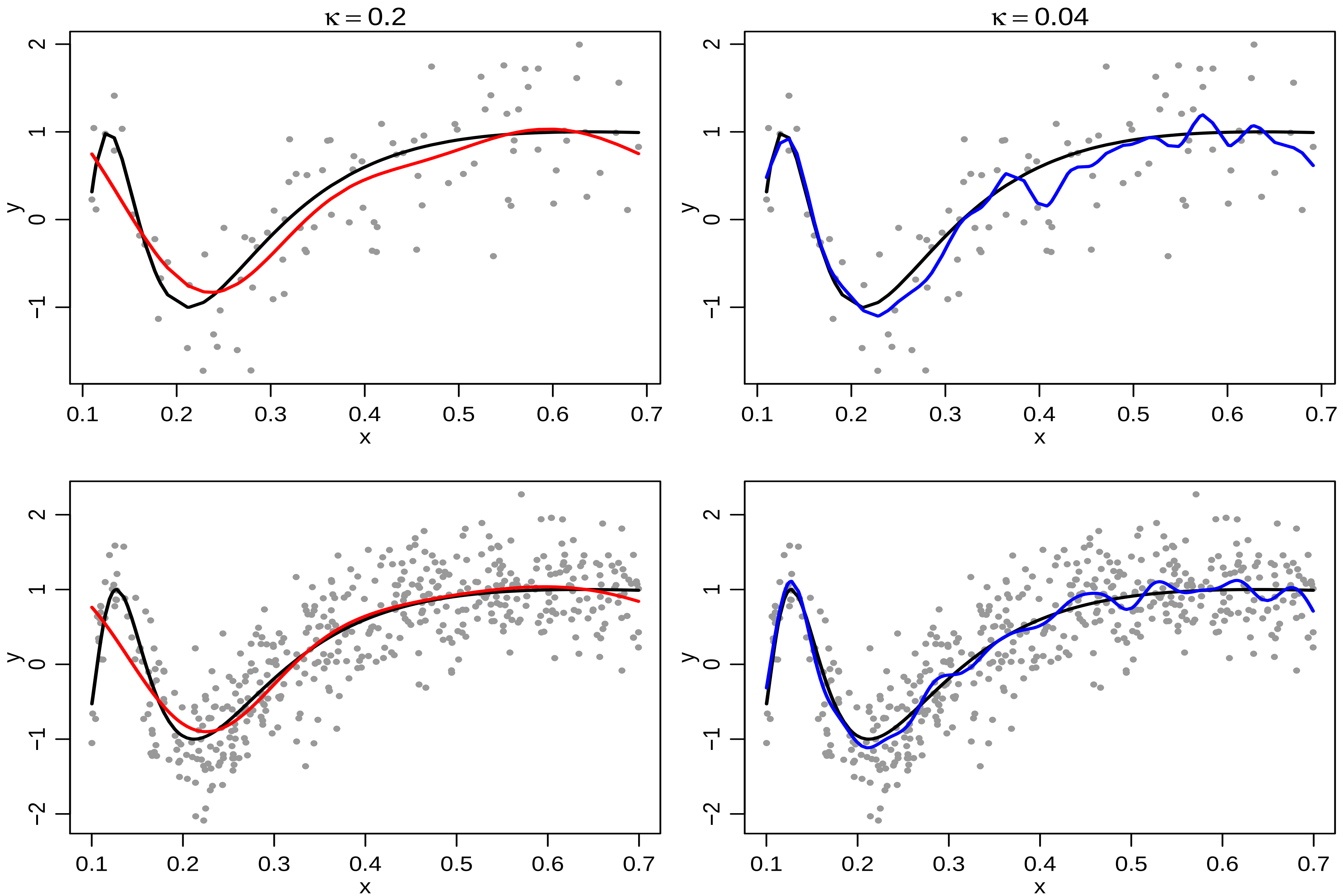
Stationary Correlation Functions

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

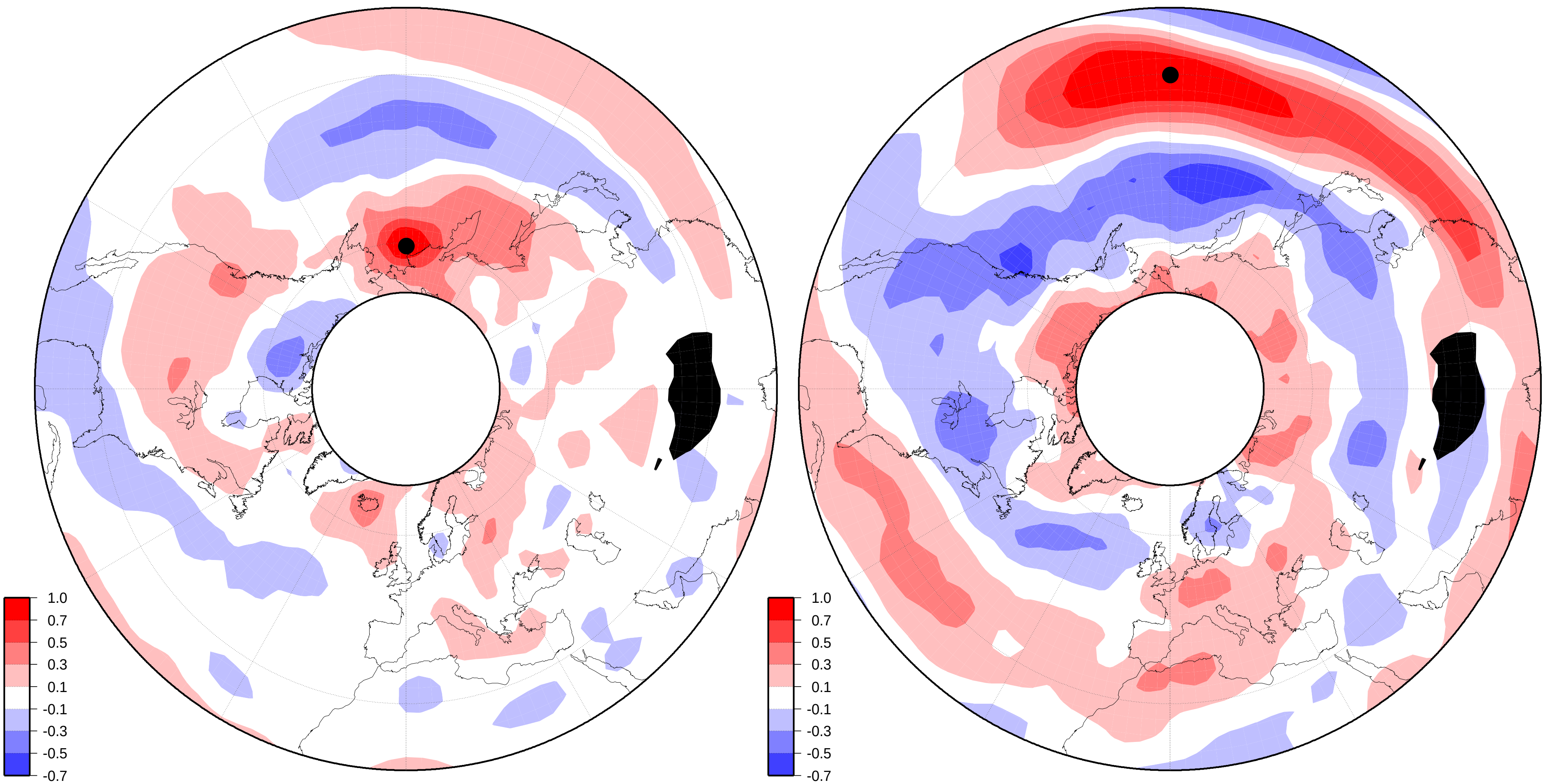


- Sample paths of GPs with this correlation are infinitely differentiable
- Replacing the power of 2 by 1 in the equation gives the exponential correlation, with no sample path derivatives

Degree of Smoothing



$\hat{R}((60°, 180°), x)$ $\hat{R}((30°, 180°), x)$



BLACK= NO DATA

Each map shows the empirical correlation between the focal point and all other points on the map, whose domain is the Northern Hemisphere,

A NONSTATIONARY COVARIANCE FUNCTION

- Higdon, Swall, and Kern (1999)

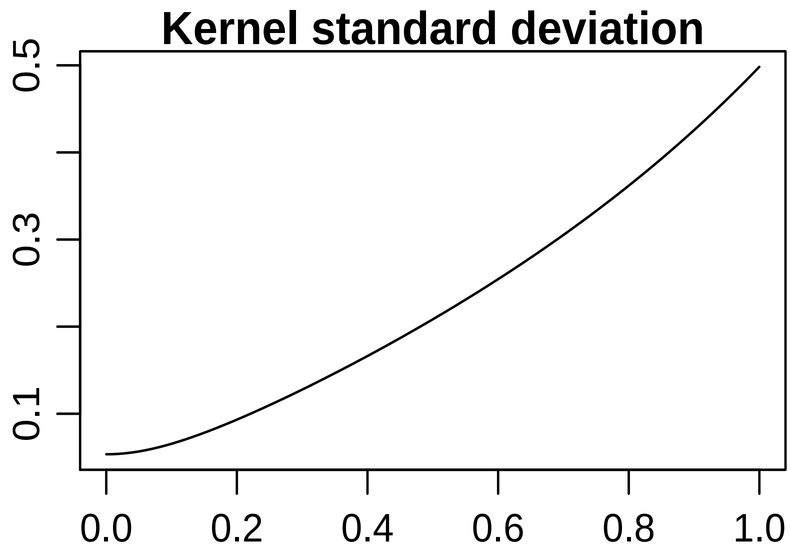
  $$R^{NS}(x_i, x_j) = c\int_{\Re^P} k_{x_i}(u) k_{x_j}(u)\,du$$

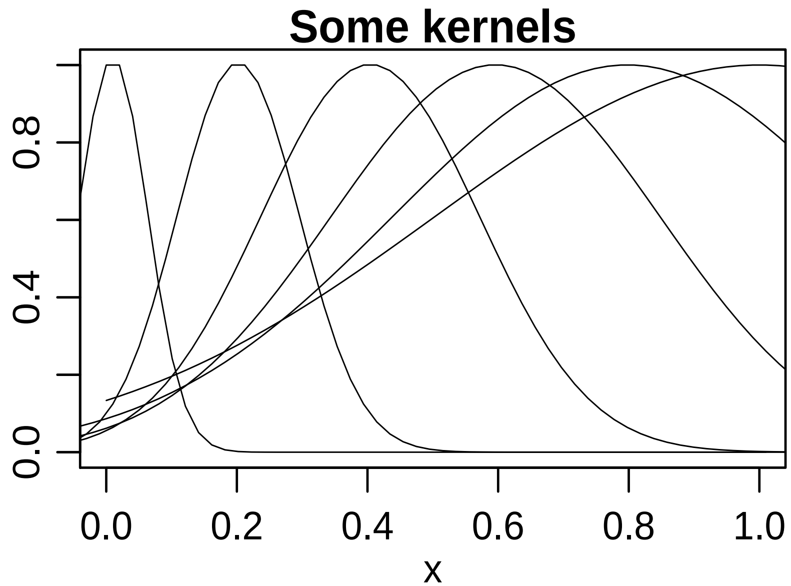
- k_x are kernel functions centered at x
- Guaranteed positive definite
- Normal kernels:

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

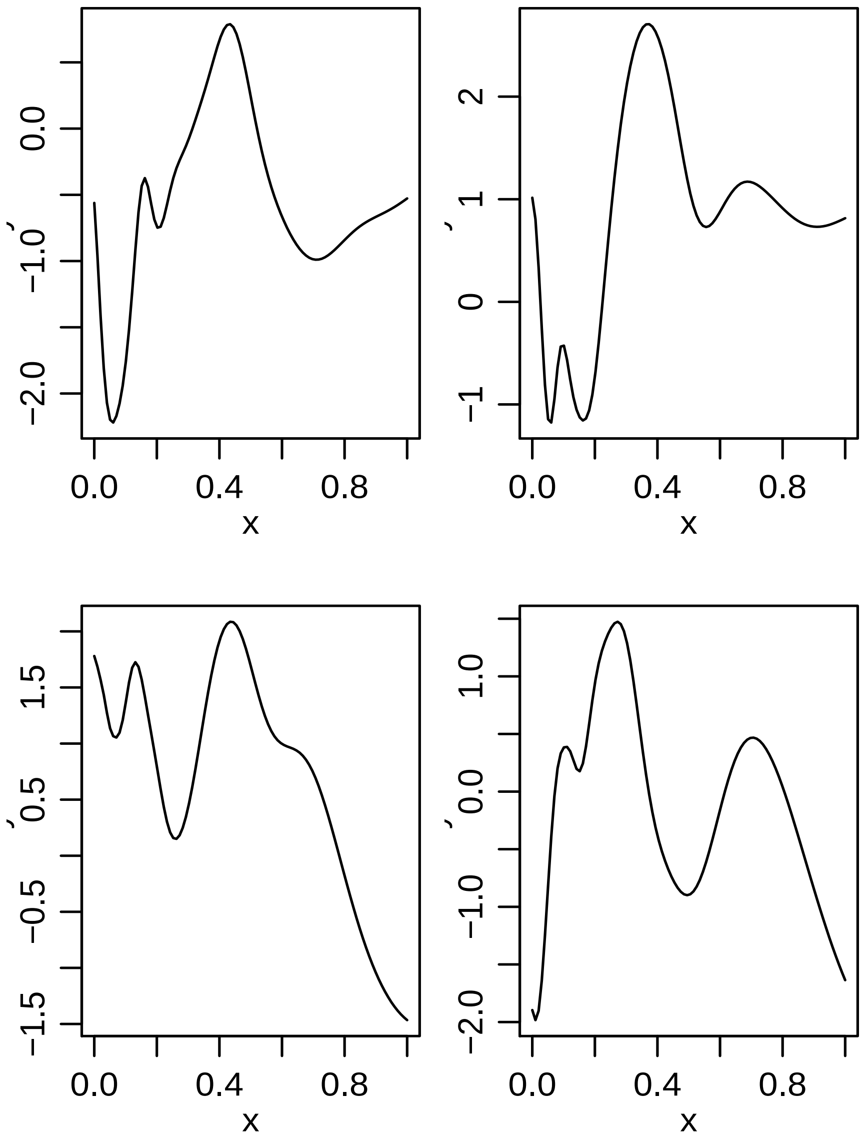
- $Z(\cdot) \sim \mathrm{GP}(\mu, \sigma^2 R^{NS}(\cdot,\cdot))$ is a nonstationary Gaussian process

Nonstationary GPs in 1D

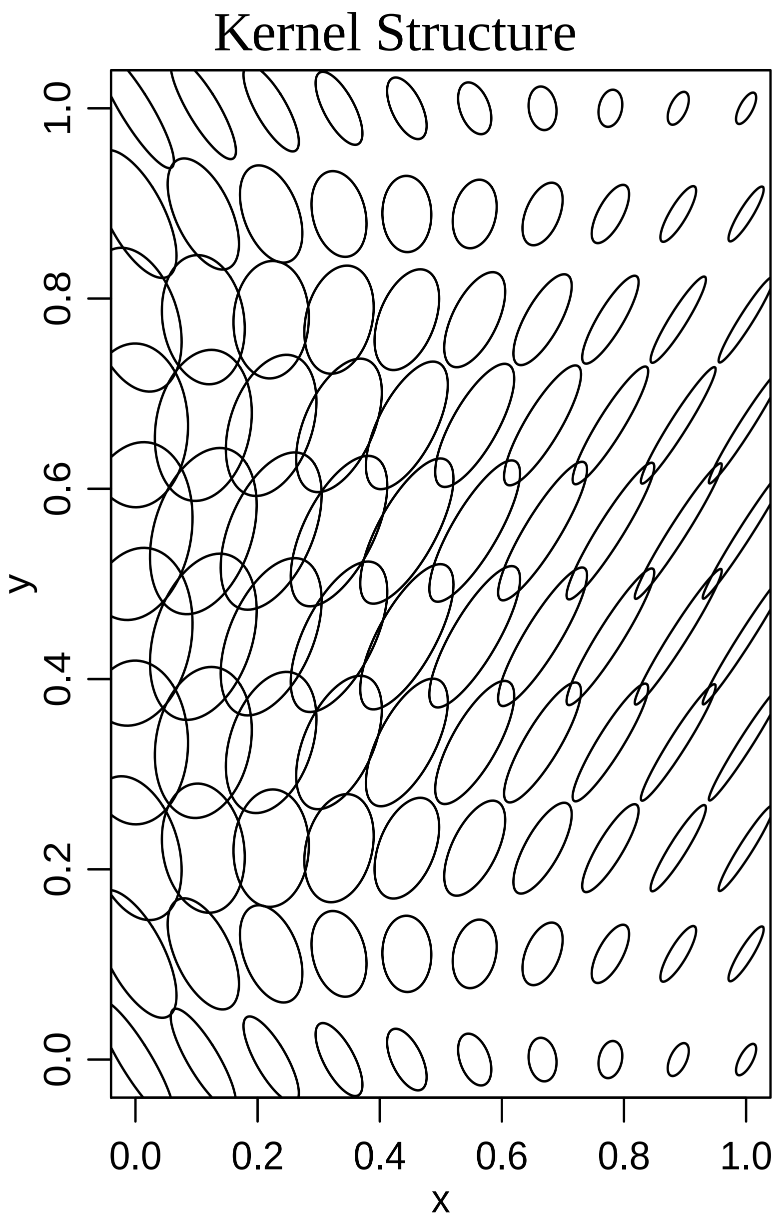


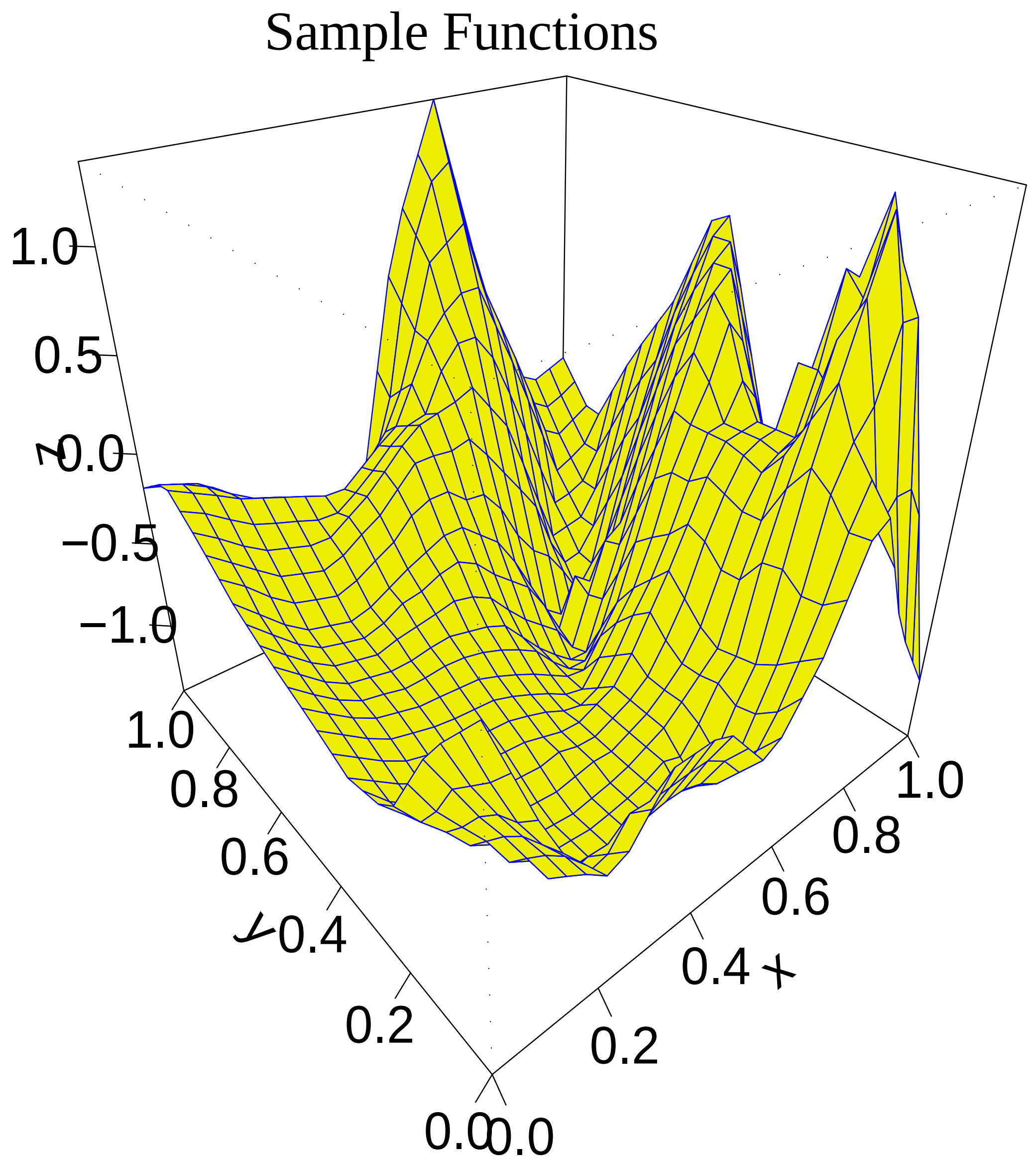


Some sample functions



NONSTATIONARY GPS IN 2D





GENERALIZING THE NONSTATIONARY COVARIANCE

- Squared exponential form:

Stationary $\Rightarrow$ Nonstationary

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

Infinitely-differentiable sample paths

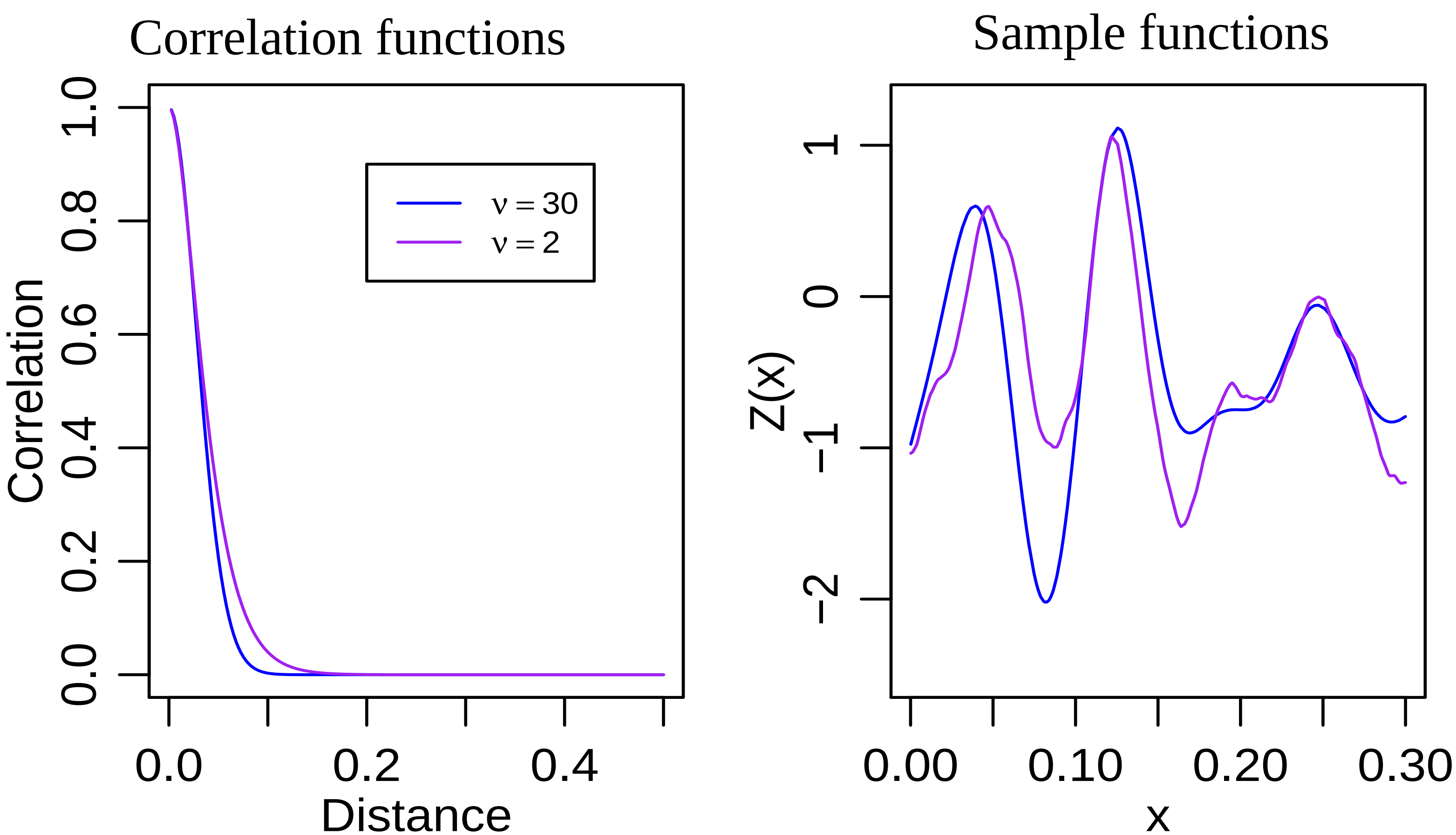
- 'Distance measures'

$$\begin{aligned}
\tau^2_{x_i,x_j} &= (x_i - x_j)^T(x_i - x_j) \\
{\tau^*}^2_{x_i,x_j} &= (x_i - x_j)^T\Sigma^{-1}(x_i - x_j) \\
Q_{x_i,x_j} &= (x_i - x_j)^T\left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1}(x_i - x_j)
\end{aligned}$$

- Can we replace τ^2 with Q_{x_i,x_j} in other stationary correlation functions?

MATÉRN STATIONARY CORRELATION FUNCTION

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



- Sample path differentiability controlled by $\boldsymbol{\nu}$
- Matérn form has asymptotic advantages (Stein 1999)

GENERALIZED KERNEL METHOD

- Theorem (Paciorek 2003): if $R(\tau)$ is positive definite for $\Re^p, p = 1, 2, \ldots$, then

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

  is positive definite for $\Re^p, p = 1, 2, \ldots$

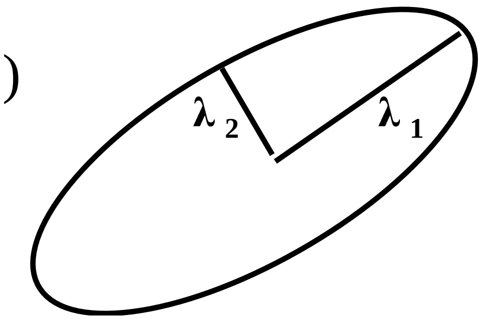
- Matérn form:

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

- Summary of theorems (Paciorek 2003) on smoothness properties of sample paths:
  - ❖ Smoothness is based on original stationary correlation function
  - ❖ Provided kernel matrices vary sufficiently smoothly in covariate space

SMOOTHLY-VARYING KERNEL MATRICES

- Goals:
  - ❖ Define multiple kernel matrices, $\boldsymbol{\Sigma_x}$ (the cov. matrices of the Gaussian kernels, $\boldsymbol{k_x}$)
  - ❖ Smoothly-varying in covariate space
  - ❖ Positive definite



- Use spectral decomposition ($\boldsymbol{\Sigma_x = \Gamma_x^T \Lambda_x \Gamma_x}$)
  - ❖ $\boldsymbol{\Gamma_x}$ parameterized as first eigenvector plus successive orthogonal vectors in reduced-dimension subspaces
  - ❖ stationary GP priors on unnormalized eigenvector coordinates $(\boldsymbol{a_x}, \boldsymbol{b_x})$ and on logarithm of eigenvalues ($\boldsymbol{\lambda_{x,1}}, \boldsymbol{\lambda_{x,2}}$)
  - ❖ gets unwieldy and highly-parameterized for large $\boldsymbol{P}$

Conclusions

I propose a new class of nonstationary covariance functions, generalizing the nonstationary covariance function of Higdon et al. (1999). The class includes a Matérn nonstationary covariance function, which shares the desirable sample path differentiability properties of the stationary Matérn covariance, in which the number of derivatives of sample paths from Gaussian processes with the specified Matérn covariance is controlled by a parameter.

These nonstationary covariance functions can be used in various models that rely on Gaussian processes. These include kriging, Bayesian spatial models, and nonparameteric regression modelling. Ongoing work includes approaches for simplifying the parameterization of the kernels that define the nonstationary covariance. The goal is to make the fitting (using either MCMC or other faster methods) simpler and less computationally intensive.