Process convolutions/spatially-varying parameters

Andrew Bean, Po-Hsu Chen, Marie Ozanne,
with Oksana Chkrebtii, Radu Herbei

The Ohio State University

November 25, 2014

Table of Contents

Introduction

- Motivation: estimation of the mean temperature field in the North Atlantic as a function of spatial location and time (Higdon 1998)
- Advantages over existing methodology:
  - Data dictates the amount of spatial smoothing
  - Accommodates temporal component
  - Adequately deals with large amounts of data
  - Allows dependence structure to evolve over space
  - Extensions provide multivariate approaches

(Discrete) Process Convolution (PC) Models

Higdon 1998

- Denote the temperature field (a Gaussian Process (GP)) over space s and time t

$$y(s,t) = z(s,t) + \epsilon(s,t). \tag{1}$$

where $\epsilon(s,t)$ is an independent error term and $z(s,t)$ is a smooth GP as shown in Equation 2:

$$z(s,t) = \sum_{j=1}^{M} \kappa(s - \omega_j, t - \tau_j) \cdot x_j. \tag{2}$$

- The kernel κ is separable so that both spatial and temporal components can be estimated.

$$\kappa(\Delta s, \Delta t) = \kappa_s(\Delta s) \cdot \kappa_t(\Delta t) \tag{3}$$

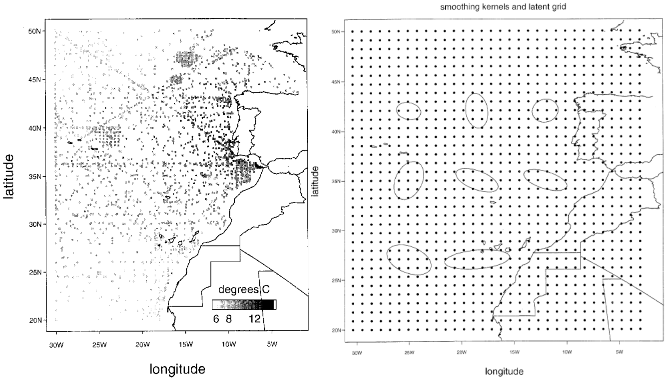

Figure : (Left) Temperature measurements taken between 1908 and 1988. The histogram shows the amount of data collected over time. (Right) Spatial locations of the underlying grid process x are given by the dots. One of the standard deviation ellipses corresponding to the covariance matrix of the Gaussian smoothing kernels are also shown.

(Discrete) PC Models - Ozone Concentrations

Higdon 2002

Conditional on the latent process values $x_t = (x_{1,t}, ..., x_{27,t})^T, t = 1, ..., 30$, a model for the data $y_t = (y_{1t}, ..., y_{n_t t})^T$ which are recorded at sites $s_{1t}, ..., s_{n_t t}$ can be written as:

$$y_t = K^t x_t + \epsilon_t \quad (4)$$

$$x_t = x_{t-1} + \nu_t \quad (5)$$

where

$$K_{ij}^t = \kappa(s_{it} - \omega_j) x_{jt}, t = 1, ..., 30 \quad (6)$$

$$\epsilon_t \overset{iid}{\sim} N(0, \sigma_\epsilon^2), t = 1, ..., 30 \quad (7)$$

$$\nu_t \overset{iid}{\sim} N(0, \sigma_\nu^2), t = 1, ..., 30 \text{ and} \quad (8)$$

$$x_1 \sim N(0, \sigma_x^2 I_{27}) \quad (9)$$

Table of Contents

Dynamic Process Convolution (DPC) Models

Calder 2007

- A DCP model constructs a space-time process that is continuous in space, but discrete in time by convolving the latent process, x, with a two-dimensional smoothing kernel at each time point.
- It can be represented as a multivariate state-space model that captures many types of temporal dependence structures, for example.

Assume the space-time process is observed at N spatial locations at each time point. The observation at location s_n at time t is modeled as

$$y(s_n, t) = \mathbf{K}(n)\mathbf{x}_t + \epsilon_{n,t}, \tag{10}$$

where the latent process evolves as

$$\mathbf{x}_t = G(D_{t-1}) + \nu_t. \tag{11}$$

In Equation 10, $\mathbf{K}(n) = [\kappa(\omega_1 - s_n), ..., \kappa(\omega_M - s_n)]$. M is the number of locations where the process is defined. $G(\cdot)$ controls the evolution of the latent process and is taken to be a function of $D_t = \{x(\omega_{1,\tau}), ..., x(\omega_{M,\tau}) : \tau \leq t\}$.

Bivariate DPC model for $PM_{2.5}$ and PM_{10} concentrations:

$$\begin{pmatrix} \mathbf{y}_t^{2.5} \\ \mathbf{y}_t^{10} \end{pmatrix} = \begin{pmatrix} \mu_t^{2.5} \\ \mu_t^{10} \end{pmatrix} + \begin{pmatrix} \mathbf{K}_t^{fine.2.5} & \mathbf{0} \\ \mathbf{K}_t^{fine.10} & \mathbf{K}_t^{coarse.10} \end{pmatrix} \begin{pmatrix} \mathbf{x}_t^{fine} \\ \mathbf{x}_t^{coarse} \end{pmatrix} + \epsilon_t \qquad (12)$$

where $\mathbf{y}_t^{2.5}$ and $\mathbf{y}_t^{10}$ are $(N_t^{2.5} \times 1)$ and $(N_t^{10} \times 1)$ vectors and

$$\begin{pmatrix} \mathbf{x}_t^{fine} \\ \mathbf{x}_t^{coarse} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{t-1}^{fine} \\ \mathbf{x}_{t-1}^{coarse} \end{pmatrix} + \nu_t; \qquad (13)$$

and $\epsilon_t \sim N\left(\mathbf{0}, \begin{pmatrix} \lambda_{\epsilon_t}^{2.5} & \mathbf{0} \\ \mathbf{0} & \lambda_{\epsilon_t}^{10} \end{pmatrix}\right)$, $\nu_t \sim N\left(\mathbf{0}, \begin{pmatrix} \lambda_{\nu_t}^{fine} & \mathbf{0} \\ \mathbf{0} & \lambda_{\nu_t}^{coarse} \end{pmatrix}\right)$

The $\mathbf{K}_t$ matrices perform the convolution of the underlying $\mathbf{x}_t$ processes from the $M^{fine} + M^{coarse}$ lattice locations to the $N_t^{2.5} + N_t^{10}$ monitor locations at time t, so they are determined by the choice of convolution kernels κ_t^{fine} and κ_t^{coarse}. The $(i, j)^{th}$ element of $\mathbf{K}_t^{fine.2.5}$ is

$$\mathbf{K}_t^{fine.2.5} = \kappa_t^{fine}(u(y_t^{2.5}(i)) - u(x_t^{fine}(j)), \nu(y_t^{2.5}(i)) - \nu(x_t^{fine}(j))) \quad (14)$$

and $u(\cdot)$ and $\nu(\cdot)$ return the longitude and latitude, respectively.

DPC Models

Calder 2008

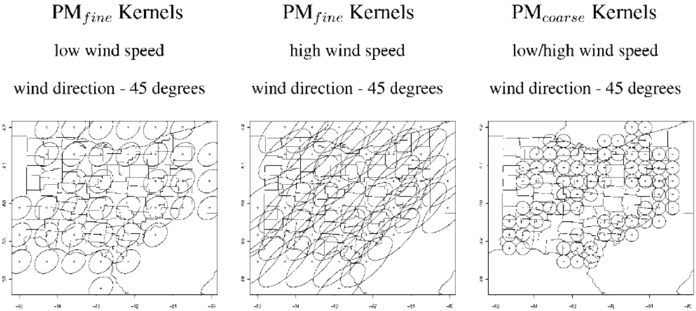

Figure : Underlying lattice locations for the PM_{fine} and PM_{coarse} processes along with examples of one standard deviation ellipses of the PM_{fine} smoothing kernel (κ_t^{fine}) and PM_{coarse} smoothing kernel (κ_t^{coarse}) (shrunk by a factor of two)

Table of Contents

A derived class of nonstationary covariance functions

Paciorek and Schervish (2004 & 2006)

- The nonstationary Gaussian process

$$z(\mathbf{x}) = \int k_{\mathbf{x}}(\mathbf{s}) x(\mathbf{s}) \, d\mathbf{s}$$

discussed by Higdon, Swall and Kern (1999) has the covariance function

$$C(\mathbf{x}_i, \mathbf{x}_j) = \int k_{\mathbf{x}_i}(\mathbf{s}) k_{\mathbf{x}_j}(\mathbf{s}) \, d\mathbf{s}.$$

- We can evaluate $C(\mathbf{x}_i, \mathbf{x}_j)$ when the kernels $k_{\mathbf{x}}(\mathbf{s})$ are Gaussian. In that case,

$$C(\mathbf{x}_i, \mathbf{x}_j) = \sigma^2 \left|\boldsymbol{\Sigma}_i\right|^{\frac{1}{4}} \left|\boldsymbol{\Sigma}_j\right|^{\frac{1}{4}} \left|\frac{\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j}{2}\right|^{-\frac{1}{2}} \exp\{-Q_{ij}\},$$

where $Q_{ij} = (\mathbf{x}_i - \mathbf{x}_j)^T \left(\frac{\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j}{2}\right)^{-1} (\mathbf{x}_i - \mathbf{x}_j)$.

A derived class of nonstationary covariance functions

Paciorek and Schervish (2004 & 2006)

- Having arrived at the nonstationary covariance function

$$C(\mathbf{x}_i, \mathbf{x}_j) = \sigma^2 \left|\boldsymbol{\Sigma}_i\right|^{\frac{1}{4}} \left|\boldsymbol{\Sigma}_j\right|^{\frac{1}{4}} \left|\frac{\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j}{2}\right|^{-\frac{1}{2}} \exp\{-Q_{ij}\},$$

  Paciorek and Schervish suggest $\exp\{-Q_{ij}\}$ could be replaced with any stationary, isotropic correlation function.

- For example, replacing $\exp\{-Q_{ij}\}$ with

$$\frac{1}{\Gamma(\nu)2^{\nu-1}} \left(2\sqrt{\nu Q_{ij}}\right)^{\nu} \mathcal{K}_{\nu}\left(2\sqrt{\nu Q_{ij}}\right)$$

  gives a nonstationary version of the Matèrn covariance.

Covariance estimation and surface fitting

Paciorek and Schervish (2004 & 2006)

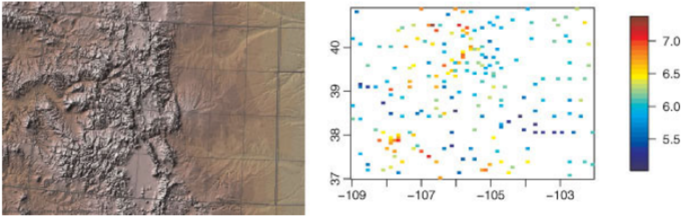

Figure 1. (a) Topography of Colorado, with thicker line indicating state boundary, and (b) image plot of log-transformed annual precipitation observations in 1981, with the color of the box indicating the magnitude of the observation

Using these Colorado precipitation measurements, Paciorek and Schervish illustrate two approaches for estimation and spatial prediction.

1. “Ad-hoc nonstationary kriging”
2. A Bayesian hierarchical model for spatial smoothing

Approach 1: “Ad-hoc nonstationary kriging”

Paciorek and Schervish (2004 & 2006)

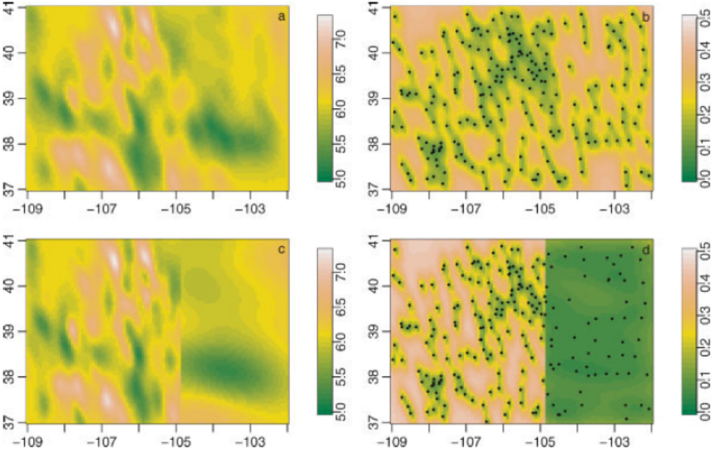

Figure 2. Surface estimates from stationary (a) and nonstationary (c) kriging with corresponding standard deviations (data locations overlaid as points), (b) and (d)

Approach 2: A Bayesian hierarchical model

Paciorek and Schervish (2004 & 2006)

The goal is to construct a Bayesian hierarchical model that makes use of the new nonstationary covariance functions.

In doing so, a kernel matrix process $\boldsymbol{\Sigma}(\cdot)$ must be defined so that the kernel matrices vary smoothly across space. This raises (at least) two questions.

1. How to parametrize $\boldsymbol{\Sigma}(\cdot)$?
2. What is an appropriate prior on those parameters?

Approach 2: A Bayesian hierarchical model

Paciorek and Schervish (2004 & 2006)

The goal is to construct a Bayesian hierarchical model that makes use of the new nonstationary covariance functions.

In doing so, a kernel matrix process $\mathbf{\Sigma}(\cdot)$ must be defined so that the kernel matrices vary smoothly across space. This raises (at least) two questions.

1. How to parametrize $\mathbf{\Sigma}(\cdot)$?
   - Use the spectral decomposition $\mathbf{\Sigma}(\mathbf{x}_i) \equiv \mathbf{\Sigma}_i = \mathbf{\Gamma}_i \mathbf{\Lambda}_i \mathbf{\Gamma}_i^T$, and set

$$\mathbf{\Gamma}_i = \frac{1}{\sqrt{\gamma_1^2(\mathbf{x}_i) + \gamma_2^2(\mathbf{x}_i)}} \begin{pmatrix} \gamma_1(\mathbf{x}_i) & -\gamma_2(\mathbf{x}_i) \\ \gamma_2(\mathbf{x}_i) & \gamma_1(\mathbf{x}_i) \end{pmatrix}.$$

2. What is an appropriate prior on those parameters?

Approach 2: A Bayesian hierarchical model

Paciorek and Schervish (2004 & 2006)

The goal is to construct a Bayesian hierarchical model that makes use of the new nonstationary covariance functions.

In doing so, a kernel matrix process $\mathbf{\Sigma}(\cdot)$ must be defined so that the kernel matrices vary smoothly across space. This raises (at least) two questions.

1. How to parametrize $\mathbf{\Sigma}(\cdot)$?
   - Use the spectral decomposition $\mathbf{\Sigma}(\mathbf{x}_i) \equiv \mathbf{\Sigma}_i = \mathbf{\Gamma}_i \mathbf{\Lambda}_i \mathbf{\Gamma}_i^T$, and set

$$\mathbf{\Gamma}_i = \frac{1}{\sqrt{\gamma_1^2(\mathbf{x}_i) + \gamma_2^2(\mathbf{x}_i)}} \begin{pmatrix} \gamma_1(\mathbf{x}_i) & -\gamma_2(\mathbf{x}_i) \\ \gamma_2(\mathbf{x}_i) & \gamma_1(\mathbf{x}_i) \end{pmatrix}.$$

2. What is an appropriate prior on those parameters?

$$\left\{ \underbrace{\log(\lambda_2(\cdot))}_{\text{eigenvalue process}} , \underbrace{\gamma_1(\cdot), \gamma_2(\cdot)}_{\text{eigenvector processes}} \right\} \sim \text{independent G.P. priors}$$

Approach 2: A Bayesian hierarchical model

Paciorek and Schervish (2004 & 2006)

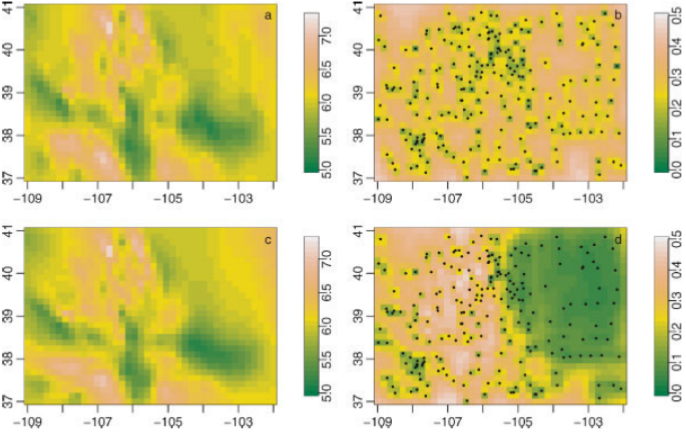

Figure 3. Posterior mean surface estimates from stationary (a) and nonstationary (c) GP models with corresponding posterior standard deviations (data locations overlaid as points), (b) and (d)

Table of Contents

Multivariate Spatial Models

- Multivariate spatial models are increasingly required in the geophysical sciences.

Multivariate Spatial Models

- Multivariate spatial models are increasingly required in the geophysical sciences.
- Consider a multivariate process $\boldsymbol{Z}(\boldsymbol{x}) = (Z_1(\boldsymbol{x}), \ldots, Z_p(\boldsymbol{x}))'$ with matrix valued covariance function

$$C(\boldsymbol{x}, \boldsymbol{y}) = \begin{pmatrix} C_{11}(\boldsymbol{x}, \boldsymbol{y}) & \cdots & C_{1p}(\boldsymbol{x}, \boldsymbol{y}) \\ \vdots & \ddots & \vdots \\ C_{p1}(\boldsymbol{x}, \boldsymbol{y}) & \cdots & C_{pp}(\boldsymbol{x}, \boldsymbol{y}) \end{pmatrix},$$

where $C_{ij}(\boldsymbol{x}, \boldsymbol{y}) = Cov(Z_i(\boldsymbol{x}), Z_j(\boldsymbol{y}))$.

Multivariate Spatial Models

- Multivariate spatial models are increasingly required in the geophysical sciences.
- Consider a multivariate process $\boldsymbol{Z}(\boldsymbol{x}) = (Z_1(\boldsymbol{x}), \ldots, Z_p(\boldsymbol{x}))'$ with matrix valued covariance function

$$C(\boldsymbol{x}, \boldsymbol{y}) = \begin{pmatrix} C_{11}(\boldsymbol{x}, \boldsymbol{y}) & \cdots & C_{1p}(\boldsymbol{x}, \boldsymbol{y}) \\ \vdots & \ddots & \vdots \\ C_{p1}(\boldsymbol{x}, \boldsymbol{y}) & \cdots & C_{pp}(\boldsymbol{x}, \boldsymbol{y}) \end{pmatrix},$$

  where $C_{ij}(\boldsymbol{x}, \boldsymbol{y}) = Cov(Z_i(\boldsymbol{x}), Z_j(\boldsymbol{y}))$.
- Cross-Covariance function $C(\cdot, \cdot)$ needs to be nonnegative definiteness.

Matérn Cross-Covariance Functions for Multivariate

- Gneiting et al. (2010) developed a multivariate Matérn model where each constituent component $Z_i(\boldsymbol{x})$ has a stationary Matérn covariance function.

Matérn Cross-Covariance Functions for Multivariate

- Gneiting et al. (2010) developed a multivariate Matérn model where each constituent component $Z_i(\boldsymbol{x})$ has a stationary Matérn covariance function.
- Each marginal covariance function,

$$C_{ii}(\boldsymbol{h}) = \sigma_i^2 \boldsymbol{M}(\boldsymbol{h}|\nu_i, a_i) \quad \text{for } i = 1, \ldots, p, \tag{15}$$

  is of the Matérn type with variance parameter $\sigma_i^2 > 0$, smoothness parameter $\nu_i > 0$ and scale parameter $a_i > 0$, where $\boldsymbol{h} = ||\boldsymbol{x} - \boldsymbol{y}||$.
- Each cross-covariance function,

$$C_{ij}(\boldsymbol{h}) = C_{ji}(\boldsymbol{h}) = \rho_{ij}\sigma_i\sigma_j \boldsymbol{M}(\boldsymbol{h}|\nu_{ij}, a_{ij}) \text{ for } 1 \leq i \neq j \leq p, \tag{16}$$

  is also a Matérn function, with collocated correlation coefficient ρ_{ij}, smoothness parameter ν_{ij}, and scale parameter a_{ij}.

Nonstationary Multivariate Spatial Processes

- Kleiber and Nychka 2012 developed a parametric nonstationary multivariate spatial model with locally varying parameter functions.

Nonstationary Multivariate Spatial Processes

- Kleiber and Nychka 2012 developed a parametric nonstationary multivariate spatial model with locally varying parameter functions.
- Let $\Sigma_{ij}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}\left(\Sigma_i(\boldsymbol{x}) + \Sigma_j(\boldsymbol{y})\right)$, $\nu_{ij}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}\left(\nu_i(\boldsymbol{x}) + \nu_j(\boldsymbol{y})\right)$, $Q_{ij}(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})'\Sigma_{ij}(\boldsymbol{x}, \boldsymbol{y})(\boldsymbol{x} - \boldsymbol{y})$, where $\beta_{ii} = 1$ and $\beta_{ij} \in [-1, 1]$ for $i \neq j$. Then the matrix with diagonal entries

$$C_{ii}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sigma_i(\boldsymbol{x})\sigma_i(\boldsymbol{y})}{|\Sigma_{ii}(\boldsymbol{x}, \boldsymbol{y})|^{1/2}} \boldsymbol{M}(Q_{ii}(\boldsymbol{x}, \boldsymbol{y})^{1/2} | \nu_{ii}(\boldsymbol{x}, \boldsymbol{y})) \quad \text{for } i = 1, \ldots, p.$$

Nonstationary Multivariate Spatial Processes

- Kleiber and Nychka 2012 developed a parametric nonstationary multivariate spatial model with locally varying parameter functions.
- Let $\Sigma_{ij}(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{2}\left(\Sigma_i(\boldsymbol{x}) + \Sigma_j(\boldsymbol{y})\right)$, $\nu_{ij}(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{2}\left(\nu_i(\boldsymbol{x}) + \nu_j(\boldsymbol{y})\right)$, $Q_{ij}(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{x}-\boldsymbol{y})'\Sigma_{ij}(\boldsymbol{x},\boldsymbol{y})(\boldsymbol{x}-\boldsymbol{y})$, where $\beta_{ii} = 1$ and $\beta_{ij} \in [-1,1]$ for $i \neq j$. Then the matrix with diagonal entries

$$C_{ii}(\boldsymbol{x},\boldsymbol{y}) = \frac{\sigma_i(\boldsymbol{x})\sigma_i(\boldsymbol{y})}{|\Sigma_{ii}(\boldsymbol{x},\boldsymbol{y})|^{1/2}} \boldsymbol{M}(Q_{ii}(\boldsymbol{x},\boldsymbol{y})^{1/2}|\nu_{ii}(\boldsymbol{x},\boldsymbol{y})) \quad \text{for } i = 1,\ldots,p.$$

- and off-diagonal entries

$$C_{ij}(\boldsymbol{x},\boldsymbol{y}) = \beta_{ij}\frac{\sigma_i(\boldsymbol{x})\sigma_j(\boldsymbol{y})}{|\Sigma_{ij}(\boldsymbol{x},\boldsymbol{y})|^{1/2}} \boldsymbol{M}(Q_{ij}(\boldsymbol{x},\boldsymbol{y})^{1/2}|\nu_{ij}(\boldsymbol{x},\boldsymbol{y})) \text{ for } 1 \leq i \neq j \leq p,$$

is a multivariate covariance function.

Estimation

- Unless there is a simplifying parametric form for the nonstationary covariance function parameters $\Sigma(\cdot), \sigma(\cdot), \nu(\cdot)$, and $\beta(\cdot)$, estimation can be difficult.

Estimation

- Unless there is a simplifying parametric form for the nonstationary covariance function parameters $\Sigma(\cdot), \sigma(\cdot), \nu(\cdot)$, and $\beta(\cdot)$, estimation can be difficult.
- Paciorek and Schervish 2006 introduced an approach to local univariate covariance function estimation. The main concern with this approach is that there is no way to estimate the locally varying coefficient $\beta(\cdot)$.

Estimation

- Unless there is a simplifying parametric form for the nonstationary covariance function parameters $\Sigma(\cdot), \sigma(\cdot), \nu(\cdot)$, and $\beta(\cdot)$, estimation can be difficult.
- Paciorek and Schervish 2006 introduced an approach to local univariate covariance function estimation. The main concern with this approach is that there is no way to estimate the locally varying coefficient $\beta(\cdot)$.
- Kleiber and Nychka 2012 proposed an estimation approach that is feasible for large datasets, and only imposes the parametric assumptions of those contained within the covariance model. The approach has two steps, first estimating the cross-ocrrelation $\beta(\cdot)$, and then the local Matérn parameters.

Cross-correlation Coefficient: $\beta_{ij}(\cdot)$

- Kernel smoothed empirical covariance matrix

$$\hat{C}_{ij}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sum_{k=1}^{n} \mathcal{K}_\lambda(||\boldsymbol{x} - \boldsymbol{s}_k||)^{1/2} \mathcal{K}_\lambda(||\boldsymbol{y} - \boldsymbol{s}_k||)^{1/2} Z_i(\boldsymbol{s}_k) Z_j(\boldsymbol{s}_k)}{\left(\sum_{k=1}^{n} \mathcal{K}_\lambda(||\boldsymbol{x} - \boldsymbol{s}_k||)\right)^{1/2} \left(\sum_{k=1}^{n} \mathcal{K}_\lambda(||\boldsymbol{y} - \boldsymbol{s}_k||)\right)^{1/2}}$$

  Here, $\mathcal{K}_\lambda(\cdot)$ is a nonnegative kernel function with bandwidth λ, such as $\mathcal{K}_\lambda(h) = \exp(-h/\lambda)$.

- Then,

$$\hat{\gamma}_{ij}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\hat{C}_{ij}(\boldsymbol{x}, \boldsymbol{y})}{\sqrt{\hat{C}_{ii}(\boldsymbol{x}, \boldsymbol{x}) \hat{C}_{jj}(\boldsymbol{y}, \boldsymbol{y})}}$$

  for all $i \neq j$, where they set $\hat{\gamma}_{ii}(\boldsymbol{x}, \boldsymbol{y}) = 1$. To convert from $\gamma_{ij}(\boldsymbol{x}, \boldsymbol{y})$ to $\beta_{ij}(\boldsymbol{x}, \boldsymbol{y})$, they used

$$\hat{\beta}_{ij}(\boldsymbol{x}, \boldsymbol{y}) = \hat{\gamma}_{ij}(\boldsymbol{x}, \boldsymbol{y}) \frac{\sqrt{\Gamma(\nu_i(\boldsymbol{x}))\Gamma(\nu_j(\boldsymbol{y}))}}{\Gamma(\nu_{ij}(\boldsymbol{x}, \boldsymbol{y}))}.$$

Marginal Parameter Functions: $\sigma_i(\cdot)$, $\Sigma_i(\cdot)$, and $\nu_i(\cdot)$

- To estimate the parameter functions $\sigma_i(\cdot)$, $\Sigma_i(\cdot)$, and $\nu_i(\cdot)$, they used a smoothed full empirical covariance matrix $\hat{C}_e$ and consider a local minimum Frobenius distance, $||\cdot||_F$.
- The smoothed full empirical covariance matrix is

$$\hat{C}_{e,ij}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sum_{k=1}^{n} \sum_{l=1}^{n} \mathcal{K}_\lambda(||\boldsymbol{x} - \boldsymbol{s}_k||)\mathcal{K}_\lambda(||\boldsymbol{y} - \boldsymbol{s}_l||) Z_i(\boldsymbol{s}_k) Z_j(\boldsymbol{s}_l)}{\sum_{k=1}^{n} \sum_{l=1}^{n} \mathcal{K}_\lambda(||\boldsymbol{x} - \boldsymbol{s}_k||)\mathcal{K}_\lambda(||\boldsymbol{y} - \boldsymbol{s}_l||)}.$$

- At a fixed location $\boldsymbol{s}$, the local estimates of $\sigma_i(\boldsymbol{s}), \Sigma_i(\boldsymbol{s})$ and $\nu_i(\boldsymbol{s})$ are found via

$$\min_{\sigma_i(\boldsymbol{s}), \Sigma_i(\boldsymbol{s}), \nu_i(\boldsymbol{s}), \forall i} ||W_\lambda(\boldsymbol{s})(\boldsymbol{C}_M(\boldsymbol{s}) - \hat{C}_e)||_F.$$

Here, $\boldsymbol{C}_M(\boldsymbol{s})$ is the theoretical multivariate Matérn covariance matrix holding all parameter functions equal to the local function values, and the matrix $W_\lambda(\boldsymbol{s})$ is a weight matrix.

Example

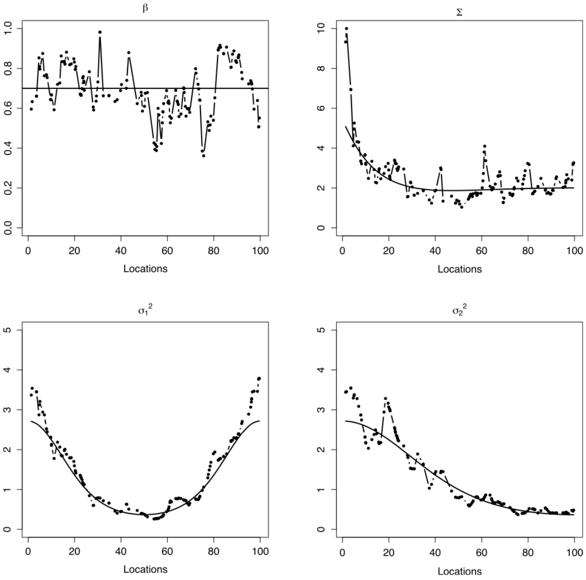

Fig. 1. Local estimates (dotted lines) and true parameter functions (solid lines) for $\beta(t)$, $\Sigma(t)$, $\sigma_1^2(t)$ and $\sigma_2^2(t)$.

Summary of Nonstationary Multivariate Process

- Kleiber and Nychka 2012 proposed a really flexible model for nonstationary multivariate spatial processes,

  and they introduced an estimation procedure which is computationally feasible for a large data set.

Summary of Nonstationary Multivariate Process

- Kleiber and Nychka 2012 proposed a really flexible model for nonstationary multivariate spatial processes,

  and they introduced an estimation procedure which is computationally feasible for a large data set.

- But two concerns here (Personally)

  1. At each location, it seems like that this model needs many (50) realizations (replications).

Summary of Nonstationary Multivariate Process

- Kleiber and Nychka 2012 proposed a really flexible model for nonstationary multivariate spatial processes,

  and they introduced an estimation procedure which is computationally feasible for a large data set.

- But two concerns here (Personally)

  1. At each location, it seems like that this model needs many (50) realizations (replications).
  2. How to do the prediction at a new location?

Thanks for your attention !