# Spatial Deformations

Casey Davis, Sophie Nguyen, Staci White
Advisors: Catherine Calder, William Notz , Elizabeth Stasny

Ohio State University

# General idea

- The assumption of nonstationary and isotropy is not realistic in various applications.
- Sampson and Guttorp (1992) first proposed deformation method in modeling nonstationary process.
- Essentially, deformation is about a nonlinear transformation that maps points in G-space into D-space such that the spatial structure is stationary and isotropic in D.
- Interpolation to ungauged sites is then done in D space.
- Difficulties include choosing the best function, accounting for uncertainty, ensuring bijectiveness (folding D space).
- Numerous additions/changes to the original model have been proposed. We will focus on the full Bayesian model by Schmidt and O'Hagan (2003).

# The setup

- G is the region of interest. $Y(\mathbf{x}, t)$ is a spatiotemporal process defined for all $\mathbf{x} \in G$ at any time $t$.
- Over G, we have n monitoring stations (gauged locations).
- At each location, repeated measurements were taken at time $t = 1, \ldots, T$.
- Observed data $Y(\mathbf{x_i}, t)$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T$.
- The main goal is to predict $Y(\mathbf{x}, t)$ at any ungauged location and time.

# The model assumptions

- All temporal effects have been removed and any correlation induced by the removal is ignored.
- The data are normally distributed (after suitable transformation).
- Let $\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{nt})^T$ denote the set of data collected from all gauged sites at time $t$.
- $\mathbf{Y}_1, \ldots, \mathbf{Y}_T$ are iid $N_n(\mu, \Sigma)$.

# Likelihood of $\Sigma$

- An $n \times n$ estimated covariance matrix **S** is calculated from the data.
- Assigning uniform prior and integrating out the mean $\mu$, we obtain the likelihood for $\Sigma$

$$f\left(S|\,\Sigma\right) \propto |\Sigma|^{-(T-1)/2}\, exp\left\{-\frac{T}{2} tr\left(S\Sigma^{-1}\right)\right\} \tag{1}$$

## $\Sigma$ element model

- For $i = 1, \ldots, n$ and $j = 1, \ldots, n$,

$$\begin{aligned} \Sigma_{ij} &= cov\left\{Y(\mathbf{x_i}, t), Y(\mathbf{x_j}, t)\right\} \\ &= \sqrt{v(\mathbf{x_i})\, v(\mathbf{x_j})} c_d(\mathbf{x_i}, \mathbf{x_j}) \end{aligned} \tag{2}$$

where for all $t$,

$$\begin{aligned} v(\mathbf{x}) &= Var\left\{Y(\mathbf{x}, t)\right\}, \\ c_d(\mathbf{x_i}, \mathbf{x_j}) &= corr\left\{Y(\mathbf{x_i}, t), Y(\mathbf{x_j}, t)\right\} \end{aligned}$$

- Prior for $v(\mathbf{x})$,

$$\begin{aligned} v(\mathbf{x}) | \tau^2, f &\sim IG\left\{\tau^2(f-2), f\right\}, \mathbf{x} \in G, \\ \pi(\tau^2) &\propto \tau^{-2} \end{aligned}$$

# $\Sigma$ element model (cont.): the spatial correlation

- $\mathbf{d}(\cdot)$ denotes the mapping from G-space to D-space, i.e.

$$\mathbf{d}(\mathbf{x}) : G \to D, \quad G \subset \mathbb{R}^2 \text{ and } D \subset \mathbb{R}^2.$$

- $\mathbf{d}(\cdot)$ is embedded in the correlation $c_d(\mathbf{x_i}, \mathbf{x_j})$ through

$$c_d(\mathbf{x_i}, \mathbf{x_j}) = g\left(\|\mathbf{d}(\mathbf{x_i}) - \mathbf{d}(\mathbf{x_j})\|\right)$$

  where $\|\cdot\|$ denotes Euclidean distance

- g is a mixture of K Gaussian correlation functions

$$g(h) = \sum_{k=1}^{K} a_k \exp\left(-b_k h^2\right)$$

  subjected to $\sum_{k=1}^{K} a_k = 1$, $b_1 > \ldots > b_K$ and, $\forall k$, $a_k, b_k > 0$.

# Latent Process $\mathbf{d}(\cdot)$

Recall: $\mathbf{d}(\cdot)$ maps the gauged sites to the stationary and isotropic $D$-space.

- If the underlying spatial process is isotropic, then $\mathbf{d}(\cdot)$ is the identity function.
- In the case of elliptical anisotropy, $\mathbf{d}(\cdot)$ is a linear transformation $\mathbf{d}(\mathbf{x}) = \mathbf{A}\mathbf{x}$.

# Elliptical Anisotropy Illustration

$$\mathbf{d}(\mathbf{x}) = \mathbf{A}\mathbf{x}, \text{ where } \mathbf{A} = \begin{pmatrix} 0.1905 & -0.0476 \\ -0.1429 & 0.2857 \end{pmatrix}$$

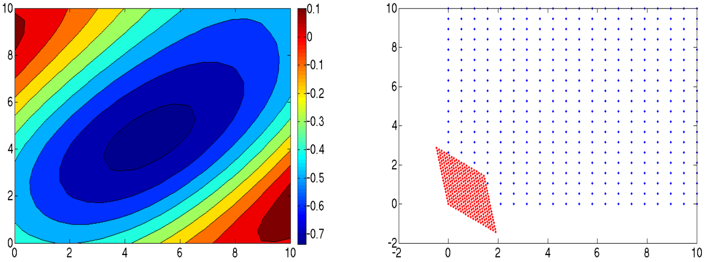


Figure : (Left) Spatial process on the *G*-space. (Right) *G*-space locations in blue, *D*-space locations in red.

# Elliptical Anisotropy Illustration

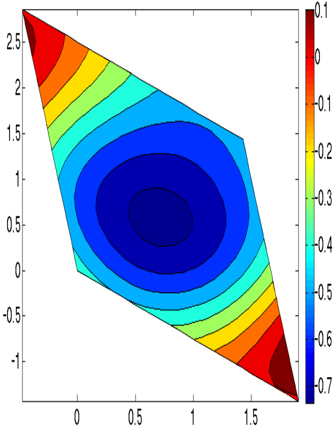


Figure : Spatial process after transforming to the $D$-space.

# Nonstationary Illustration

$$\mathbf{d}(\mathbf{x}) = \begin{pmatrix} 0.1x_1^2 + 0.75x_2 \\ 0.75x_2 \end{pmatrix}$$

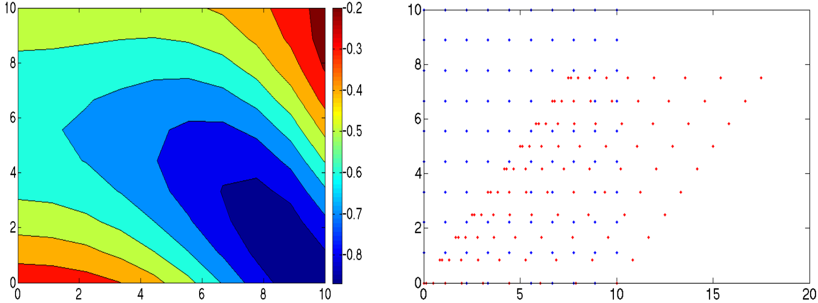


Figure : (Left) Spatial process on the $G$-space. (Right) $G$-space locations in blue, $D$-space locations in red.

# Nonstationary Illustration

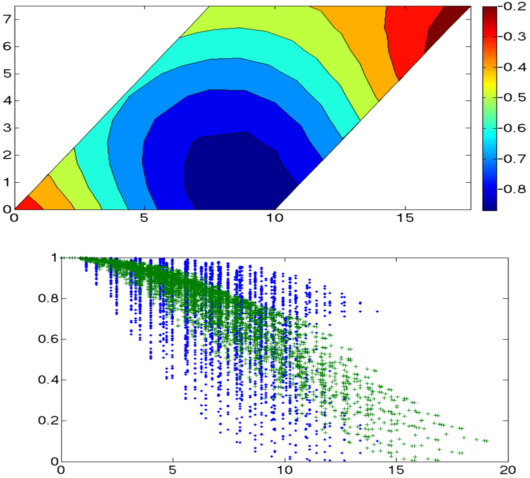


Figure : (Top) Spatial process on the *D*-Space. (Bottom) Observed correlations between locations versus their distance on the *G*-space (blue) and *D*-space (green).

# The $\mathbf{d}(\cdot)$ Process

What if you don't know the functional form of the deformation?

Schmidt & O'Hagan (2003) suggest using a Gaussian process.

$$\mathbf{d}(\cdot)|\mathbf{m}(\cdot), \boldsymbol{\sigma}_d^2, R_d(\cdot,\cdot) \sim GP(\mathbf{m}(\cdot), \boldsymbol{\sigma}_d^2 R_d(\cdot,\cdot)),$$

where

- $\mathbf{m}(\cdot)$ is the prior mean function
- $\boldsymbol{\sigma}_d^2 = \text{Var}(\mathbf{d}(\mathbf{x}))$ is a $2 \times 2$ diagonal matrix
- $R_d(\cdot,\cdot)$ measures prior correlation of the gauged sites in $D$-space.

Let $\mathbf{D} = (\mathbf{d}_1, ..., \mathbf{d}_n)$ be the $2 \times n$ matrix whose elements are the coordinates of the gauged sites in $D$-space:

$$vec(\mathbf{D})|vec(\mathbf{m}), \boldsymbol{\sigma}_d^2, \mathbf{R}_d \sim N_{2n}(vec(\mathbf{m}), \boldsymbol{\sigma}_d^2 \otimes \mathbf{R}_d).$$

# Gaussian Process Specification

- If there is no prior information on how $D$ and $G$ differ, $\mathbf{m}(\mathbf{x}) = \mathbf{x}$.
- The elements of $R_d(\cdot, \cdot)$ are modeled as

$$R_d(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2a}||\mathbf{x} - \mathbf{x}'||^2\right),$$

  where $a$ is the square of a typical distance among locations in $G$.
- $\boldsymbol{\sigma}_d^2$ controls the amount of distortion in mapping $G$ to $D$,

$$\boldsymbol{\sigma}_{d_{ii}}^2 \sim IG(\beta_i, \alpha_i), \; i = 1, 2.$$

# The Posterior Distribution

- Let $\boldsymbol{\theta} = \left[\mathbf{v}, \mathbf{D}, \mathbf{a}, \mathbf{b}, \tau^2, \boldsymbol{\sigma}_d^2\right]^T$
- Then the posterior distribution is:

$$\pi(\boldsymbol{\theta}|S) \propto f\left(S|\,\Sigma\right) \times \pi\left(\mathbf{D}|\mathbf{m}, \boldsymbol{\sigma}_d^2, \mathbf{R}_d\right) \times \pi\left(\mathbf{v}|\tau^2\right)$$
$$\times\, \pi\left(\tau^2\right) \times \pi\left(\sigma_{d_{11}}^2\right) \times \pi\left(\sigma_{d_{22}}^2\right) \times \pi\left(\mathbf{b}\right)$$

where

$$\pi\left(\mathbf{b}\right) \propto \prod_{i=1}^{K} \frac{1}{b_k} exp\left\{\frac{-\left(log(b_k) - \mu_b\right)}{2\sigma_b^2}\right\}$$

- This is an analytically intractable distribution so MCMC methods are used to obtain samples from this distribution.

# Prediction

- One goal is to predict $Y^*$ at a new, ungauged location, $\mathbf{x}^*$ at some time $t \in \{1, \ldots, T\}$, conditioned on the observations $\mathbf{Y}$ of the $n$ gauged sites.
- Augment the parameter vector $\boldsymbol{\theta}$ such that $\boldsymbol{\theta}^* = [\boldsymbol{\theta}, v^*, \mathbf{d}^*]^T$, where $v^*$ is the variance of the process at the ungauged location, and $\mathbf{d}^*$ is its position in $D$-space.
- The predictive distribution of $Y^*|\mathbf{Y}$ is

$$p\left(Y^*|\mathbf{Y}\right) = \int_{\boldsymbol{\theta}^*} p\left(Y^*|\mathbf{Y}, \boldsymbol{\theta}^*\right) p\left(\boldsymbol{\theta}^*|\mathbf{Y}\right) d\boldsymbol{\theta}^*$$

# Prediction

From usual multivariate normal theory we have

$$(Y^*|\mathbf{Y}, \boldsymbol{\theta}^*) \sim N\left(\mu^* + \boldsymbol{\psi}^T \Sigma^{-1}(\mathbf{Y} - \boldsymbol{\mu}), v^* - \boldsymbol{\psi}^T \Sigma^{-1} \boldsymbol{\psi}\right)$$

where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_n)^T$ with

$$\psi_i = \sqrt{v^* v_i} \sum_{k=1}^{K} a_k \exp\left(-b_k \left(\|\mathbf{d}^* - \mathbf{d}(\mathbf{x_i})\|\right)^2\right)$$

# Prediction

Note

$$
\begin{aligned}
p\left(\boldsymbol{\theta}^{*}|\mathbf{Y}\right) &= p\left(\boldsymbol{\theta}, v^{*}, \mathbf{d}^{*}|\mathbf{Y}\right) \\
&= p\left(v^{*}, \mathbf{d}^{*}|\boldsymbol{\theta}, \mathbf{Y}\right) p\left(\boldsymbol{\theta}|\mathbf{Y}\right) \\
&= p\left(v^{*}, \mathbf{d}^{*}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}|\mathbf{Y}\right) \\
&= p\left(v^{*}|\boldsymbol{\theta}\right) p\left(\mathbf{d}^{*}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}|\mathbf{Y}\right)
\end{aligned}
$$

where $v^{*}|\boldsymbol{\theta} \sim IG\left\{\tau^{2}\left(f-2\right), f\right\}$ and

$$
\mathbf{d}^{*}|\boldsymbol{\theta} \sim N\left(\mathbf{m}_{\mathbf{d}^{*}}, \Sigma_{\mathbf{d}^{*}}\right)
$$

where

$$
\begin{aligned}
\mathbf{m}_{\mathbf{d}^{*}} &= \mathbf{m}^{*} + \left(\mathbf{I}_{2} \otimes \mathbf{R}_{d}^{*T}\mathbf{R}^{-1}\right)\left(\mathbf{D}-\mathbf{m}\right) \\
\Sigma_{\mathbf{d}^{*}} &= \boldsymbol{\sigma}_{d}^{2}\left(1-\mathbf{R}_{d}^{*T}\mathbf{R}^{-1}\mathbf{R}_{d}^{*}\right)
\end{aligned}
$$

# Prediction

- Now MCMC algorithms can be used to obtain a sample from the posterior distribution of $\boldsymbol{\theta}$ and then for each $m = 1, \ldots, n_{mcmc}$
  1. Sample $v^*|\boldsymbol{\theta}^{[m]} \sim IG\left\{\tau^{2^{[m]}}(f-2)\,,\, f\right\}$
  2. Sample $\mathbf{d}^*|\boldsymbol{\theta}^{[m]} \sim N\left(\mathbf{m}_{\mathbf{d}^*}, \Sigma_{\mathbf{d}^*}\right)$
  3. Compute $\Sigma_{ij}^{[m]} = \sqrt{v^{[m]}(\mathbf{x}_i)\, v^{[m]}(\mathbf{x}_j)} \sum_{k=1}^{K} a_k^{[m]} \exp\left(-b_k^{[m]}\left(\|\mathbf{d}(\mathbf{x_i}) - \mathbf{d}(\mathbf{x_j})\|\right)^2\right)$
- You now have a sample, $\Sigma^{[1]}, \ldots, \Sigma^{[n_{mcmc}]}$, from the posterior distribution of this covariance matrix with the new location involved.
- To make a prediction, a Rao-Blackwellized point prediction would then be

$$
\begin{aligned}
\widehat{Y}^* &= E\left(Y^*|Y\right) \\
&= E_{\boldsymbol{\theta}^*|S}\left(E\left(Y^*|Y, \boldsymbol{\theta}^*\right)\right) \\
&\approx \frac{1}{n_{mcmc}} \sum_{m=1}^{n_{mcmc}} \mu^* + \boldsymbol{\psi}^{[m]^T} \Sigma^{[m]-1}(\mathbf{Y} - \boldsymbol{\mu})
\end{aligned}
$$

# Comments

- Similarly to Sampson and Guttorp, this model maps the space of interest to a latent space where stationarity and isotropy hold.
- This model is an improvement on Sampson and Guttorp in that it takes into account the uncertainty in the mapping into the latent space.
- This model is very computationally expensive and uses a complicated MCMC algorithm.

# References I


Schmidt ,A. M. , O'Hagan , A.
Bayesian Inference for Nonstationary Spatial Covariance Structure via Spatial Deformations.
*Journal Of The Royal Statistical Society, Series B*, 65, 745–758 2009.

Sampson, P. D. and Guttorp, P.
Nonparametric Estimation of Nonstationary Spatial Covariance Structure.
*Journal of the American Statistical Association*, Vol. 87, No. 417:108-119, 1992.