Smoothing and Weighted Average Techniques

Shreyan Ganguly, Jeffrey Gory, Meng Li, Matt Pratola, and Angela Dean

October 28, 2014
Non-Stationarity

- Second order stationarity implies that the covariance structure between two distinct points in the spatial domain is only a function of the distance between them.
- Under the assumption of non-stationarity, this is not the case.
- One method of estimating the covariance structure of a non-stationary spatial process is with smoothing or weighted average techniques.
Notable Papers

- Fuentes (2001) represented a non-stationary spatial process as a weighted average of orthogonal local stationary processes.
- Fuentes and Smith (2001) extended this idea to a continuous convolution of local stationary processes.
- Kim, Mallick, and Holmes (2005) proposed a similar approach with piecewise Gaussian processes that allows for sharp transitions in the covariance structure.
- Nott and Dunsmuir (2002) used a weighted average of stationary processes to describe the conditional behaviour of a process given the values of the process at a set of monitoring sites.
Suppose $Z(\cdot)$ is a non-stationary process observed on a given spatial region $D$.

The region $D$ is divided into $k$ disjoint subregions $S_i$, $i = 1, 2, \ldots, k$ such that

$$D = \bigcup_{i=1}^{k} S_i$$

Here $k$ can be considered to be a variable valid for estimation.
Representation of $Z$

- $Z$ is represented as a weighted average of orthogonal local stationary processes $Z_i$

$$Z(s) = \sum_{i=1}^{k} Z_i(s)w_i(s)$$

- Each $w_i(\cdot)$ is a weight function (e.g., the inverse of the distance between $s$ and $S_i$)

- Each $Z_i$ is a stationary process with spatial covariance representing the spatial structure of $Z$ in the subregion $S_i$

- The non-stationary covariance of $Z$ is defined as

$$\text{cov}(Z(s), Z(t)) = \sum_{i=1}^{k} w_i(s)w_i(t)\text{cov}(Z_i(s), Z_i(t))$$

where $\text{cov}(Z_i(s), Z_i(t)) = K_{\theta_i}(s - t)$, $K_{\theta_i}(\cdot)$ is Matern
Algorithm to Find $k$

- Start with small number of subregions (i.e. a small value of $k$)
- Increase $k$ by dividing each subregion $S_i$ in half
- Repeat this until we have less than 36 observations in the subregions
  or until the AIC suggests no significant improvement in the estimation of $\theta_i$ for $i = 1, \ldots, k$ by increasing the number of subregions $k$
More Details on Finding $k$

- If data are in a regular grid, the Whittle (1954) approximation to the local stationary functions is used to calculate the global likelihood to obtain the AIC value.

- If the data are not in a regular grid, then the method proposed by Kitanidis (1993) and by Mardia and Marshall (1984) is used to efficiently maximize the global likelihood to obtain the AIC.

- The minimum number of observations in a subregion is restricted to 36 as it is not recommended to calculate the periodogram when the number of sample points available to compute the sample covariance is less than 36 (see Haas 1990, Journel and Huijbregts 1978, SAS/STAT Technical Report 1996).
Estimation of Covariance Parameters

- Recall that the non-stationary covariance of $Z$ is represented as

$$\text{cov}(Z(s), Z(t)) = \sum_{i=1}^{k} w_i(s)w_i(t)\text{cov}(Z_i(s), Z_i(t))$$

- Assuming $\text{cov}(Z_i(s), Z_i(t)) = K_{\theta_i}(s - t)$ is a Matern covariance structure we have

$$\text{cov}(Z(s), Z(t)) = \sum_{i=1}^{k} w_i(s)w_i(t)K_{\theta_i}(s - t)$$

- We need to estimate $\theta_i$ for each $S_i$
Using a spectral approach assume the spectral function

\[ f_i(\omega) = \phi_i(|\omega|^2)(-\nu_i-d/2) \]

for \( Z_i \) for each subgrid \( S_i \) so that \( \theta_i = (\phi_i, \nu_i)^T \)

Let \( G_i \) be the generalized covariance corresponding to \( f_i \)

\( \hat{\phi}_i \) and \( \hat{\nu}_i \) are obtained using a weighted non-linear least squares technique in the spectral domain

With these estimates we get the estimated generalized covariance

\[ \hat{G}(Z(s), Z(t)) = \sum_{i=1}^{k} w_i(s)w_i(t)\hat{G}_i(Z_i(s), Z_i(t)) \]
Continuous Case

- Spatially Weighted Average of Local Stationary Processes:
  \[ Z(s) = \sum_{i=1}^{k} Z_i(s)w_i(s) \]
  where \( w_i \) is the weight function and \( Z_i(s) \) are orthogonal stationary processes with \( Z_i(s) \perp Z_j(t) \), for any \( i \neq j \)

- Continuous Convolution of Local Stationary Processes:
  \[ Z(s) = \int_D K(s-u)Z_{\theta(u)}(s)du \]
  where \( K \) is kernel function and \( Z_{\theta(u)} \ u \in D \), is a family of independent stationary Gaussian processes indexed by \( \theta \), which is assumed to vary smoothly across space
Stationary Gaussian processes can be represented in the form

$$Z(s) = \int_D K(s - u)X(u)du$$

where $K(\cdot)$ is some kernel function and $X(\cdot)$ is a Gaussian white noise process.

This can be extended to non-stationary processes.
Kernel Smoothing Models

- **Model 1: Process Convolution Model**

\[
Z(s) = \int_D K_s(u)X(u)du
\]  \hspace{1cm} (1)

where \( K_s \) depends on the location \( s \)

- **Model 2: Continuous Convolution of Local Stationary Processes**

\[
Z(s) = \int_D K(s - u)Z_{\theta(u)}(s)du
\]  \hspace{1cm} (2)

where \( Z_{\theta(u)}(s) \), \( s \in D \) is a family of independent stationary Gaussian processes indexed by \( \theta(u) \)
Generally, Model (1) and (2) are not equivalent.

Since $Z_{\theta(u)}$ is a family of independent stationary Gaussian process, it can be written in the form,

$$Z_{\theta(u)}(s) = \int_{\mathbb{R}^2} K_u(x-s)X_u(x)dx$$

where $K_u$ is a kernel for each location $u$, and $X_u$ is an independent white noise process for each location $u$.

Only in the case that the process $X_u$ is common across all $u$ (instead of being an independent white noise process for each location $u$) can Model (1) and (2) be expressed interchangeably.

However, $Z_{\theta(u)}(s), s \in D$ are correlated through kernel smoothing.
Covariance Structure of Non-Stationary Process $Z$

- Let $Z_{\theta(u)}$ be a stationary Gaussian process with covariance matrix $C_{\theta(u)}$, that is,

$$\text{cov}(Z_{\theta(u)}(x_1), Z_{\theta(u)}(x_2)) = C_{\theta(u)}(x_1 - x_2)$$

- The covariance between locations $x_1$ and $x_2$ of the non-stationary process $Z$ is a convolution of the local covariances $C_{\theta(u)}(x_1 - x_2)$,

$$C(x_1, x_2; \theta) = \int_D K(x_1 - u)K(x_2 - u)C_{\theta(u)}(x_1 - x_2)du$$

- Assume that $\theta(u)$ is a continuous function of $u$
- In implementation, the integration can be approximated by Riemann sum or Monte Carlo integration.
Hierarchical Bayesian Model

\[ Z(s) = \int_D K(s - u)Z_{\theta(u)}(s)du \]

- **Stage 1:**
  Conditional on model parameters \( \{\theta(u), u \in D\} \), \( Z \) is Gaussian

- **Stage 2:**
  \( \theta(u) = a + r_i + c_j + \epsilon_{\theta}(u) \), where the process \( \epsilon_{\theta}(u) \) is a stationary spatial process with mean zero and its covariance structure can be modeled by some well-known covariance function with unknown parameters.

  Another parameter that needs to be estimated is the bandwidth \( h \) that defines the kernel function \( K \).
Piecewise Gaussian Processes

- Much like Fuentes (2001), partition region into several subregions and fit a stationary process within each subregion.
- Adapted to deal with sharp transitions in covariance structure, so processes assumed independent across subregions.
  - The resulting non-stationary model is not smooth.
  - One possible application is modeling soil permeability.
- Voronoi tessellation (Green and Sibson 1978) used to determine number of subregions and their centers.
- Fit model within a Bayesian framework.
Voronoi Tessellation
General Idea

- Measurements of a spatial process are available at a set of monitoring sites
- An empirical covariance matrix can be estimated from these measurements
- Want to extend this covariance matrix to a covariance function over the entire region of interest
- Do this with a weighted average of local stationary processes
Illustration

Image from Haas 1990
Notation

- $Z = \{Z(s), s \in \mathbb{R}^d\}$ is the spatial process of interest
- $s_1, \ldots, s_n$ are the sites at which $Z$ is observed
- $\Gamma$ is the covariance matrix for $(Z(s_1), \ldots, Z(s_n))^T$
- $W_i(s), i = 1, \ldots, I$ are a collection of independent, zero-mean, stationary processes which describe the behavior of $Z$ locally about a collection $z_1, \ldots, z_I$ of $I$ (unmonitored) locations
- $R_i(h), h \in \mathbb{R}^d$ is the covariance function corresponding to $W_i(s)$
- $C_i = [R_i(s_j - s_k)]_{j,k=1}^n$ is the covariance matrix of $W_i = (W_i(s_1), \ldots, W_i(s_n))^T$
- $c_i(s)$ is the vector $(R_i(s - s_1), \ldots, R_i(s - s_n))^T$ of cross-covariances between $W_i(s)$ and $W_i$
We can represent $W_i(s)$ as

$$W_i(s) = c_i(s)^T C_i^{-1} W_i + \delta_i(s)$$

where $\delta_i(s)$ is a zero-mean, non-stationary random field with covariance function

$$R_{\delta_i}(s, t) = R_i(t - s) - c_i(s)^T C_i^{-1} c_i(t)$$

Note that $c_i(s)^T C_i^{-1} w$ is the simple kriging predictor of $W_i(s)$ given $W_i = w$

$\delta_i(s)$ describes the residual variation, so $\delta_i(s) = 0$ when $s \in \{s_1, \ldots, s_n\}$
Representations of Non-Stationary Process

Define a non-stationary process $W^*(s)$ by

$$W^*(s) = \sum_i \nu_i(s)\mu_i(s) + \sum_i \nu_i(s)^{1/2}\delta_i(s)$$

- Each $\nu_i(s)$ is a weight function
- Each $\delta_i(s)$ is a residual process
- Each $\mu_i(s)$ is a mean process given by

$$\mu_i(s) = c_i(s)^T C_i^{-1} W^*$$

where $W^*$ is a zero-mean $n\times1$ random vector with covariance matrix $\Gamma$ uncorrelated with each $\delta_i(s)$.
With this representation, $W^*(s)$ has covariance function

$$R^*(s, t) = \sum_{i,j} \nu_i(s)\nu_j(t)c_i(s)^T C_i^{-1} \Gamma C_j^{-1} c_j(t)$$

$$+ \sum_i \nu_i(s)^{1/2}\nu_i(t)^{1/2}R_{\delta i}(s, t)$$

We use a covariance function of this form to estimate the covariance of $Z(s)$.
Changing notation we can write

\[
\text{Cov}(W^*(s), W^*(t)) = \Sigma_0(s, t) + \sum_{i=1}^{I} w_i(s)w_i(t)C_{\theta_i}(s - t)
\]

\(\Sigma_0(s, t)\) is a function of the empirical covariance matrix \(\Gamma\) and the local stationary models \(W_i(s)\)

- The \(W_i(s)\) are computed so that \(\text{Cov}(W^*(s_k), W^*(s_l)) = \Gamma_{kl}\)
- The 2nd term is similar to the model proposed by Fuentes (2001)
- Weighted average of residual covariance functions
- Condition on observed values at the monitoring sites
Choice of Weight Functions

- The weight functions $\nu_i(s)$ should have the following properties:
  - $\nu_i(s) \geq 0 \ \forall i = 1, \ldots, I \text{ and } \forall s \in \mathbb{R}^d$
  - $\sum_i \nu_i(s) = 1$
  - $\text{argmax}(\nu_i(s)) = z_i \ \forall i = 1, \ldots, I$
  - $\nu_i(s)$ decays smoothly to zero as $\|s - z_i\| \rightarrow \infty$

- One choice of $\nu_i(s)$ is
  $$\nu_i(s) = \frac{f_\eta(s-z_i)}{\sum_j f_\eta(s-z_j)}$$
  where $f_\eta(\cdot)$ is a kernel function given by
  $$f_\eta(t) = \exp\left(-\frac{\|t\|^2}{\eta}\right)$$
  and $\eta$ is a smoothing parameter.
Estimation of $\Gamma$

- Can estimate covariance matrix $\Gamma$ by averaging over time

$$\hat{\Gamma}_{n \times n} = \frac{1}{M} \sum_{i=1}^{M} (z_i - \bar{z})(z_i - \bar{z})^T$$

- Could alternatively use the Loader & Switzer (1992) empirical Bayes shrinkage estimator

$$\hat{\Gamma}_{LS} = \lambda \hat{\Gamma} + (1 - \lambda) \hat{C}$$
Estimation of the $R_i(h)$

- For each $i = 1, \ldots, I$ fit a variogram within a window centered on the point $z_i$
  - $z_1, \ldots, z_I$ could be on a grid covering the region of interest
  - Might instead be based on knowledge of non-stationarity
  - Recommend at least 10 sites in each window
  - Related to moving window kriging (Haas 1990, 1995)
- One can then estimate $\hat{c}_i(s)$, $\hat{c}_j(t)$, $\hat{C}_i$, $\hat{C}_j$, and $\hat{R}_{\delta_i}(s, t)$ and ultimately compute $\hat{R}^*(s, t)$
Advantages of this Method

- Produces a valid non-negative definite non-stationary covariance function
- "Computationally attractive," especially in comparison to deformation techniques
- Can be extended to multivariate processes
References


