STAT 8025

Lecture 8: Nonstationary Models

Dr. Emily Lei Kang

Division of Statistics & Data Science
Department of Mathematical Sciences
University of Cincinnati

Recall the model introduced in previous slides:

$$Y(\mathrm{s}) = \mu(\mathrm{s}) + w(\mathrm{s}) + \epsilon(\mathrm{s})$$

- $\mu(\cdot)$ is a deterministic mean function; often $\mu(\mathrm{s}) = \mathrm{X}(\mathrm{s})'\boldsymbol{\beta}$
- $w(\cdot) \sim \mathcal{GP}(0, C(\cdot,\cdot))$, a mean-zero Gaussian process
- $\epsilon(\cdot)$ is a spatially independent error process (nugget), independent of $w(\cdot)$

We observe $Y(\cdot)$ (or a noisy version of $Y(\cdot)$) at a finite set of locations $\mathrm{s}_1, \ldots, \mathrm{s}_n \in \mathcal{D}$, and would like to predict $Y(\mathrm{s}_0)$

- So far we have assumed a specific parametric covariance function, such as Matérn, exponential, which are stationary

$$C(\mathrm{s}, \mathrm{u}) = C(\mathrm{s} - \mathrm{u})$$

- In many (if not all) applications, this assumption is false.
- However, this assumption of stationarity can be a good simplification in some scenarios and results are not too sensitive to this assumption.
- In some cases, accounting for **nonstationary** spatial dependence is important and can give improved prediction.

- Incorporating nonstationarity into $\mu(\cdot)$ is usually easy, including $X(\mathrm{s})$
- Incorporating nonstationarity into variance and correlation is not that easy, as the covariance function must maintain valid.
- We will introduce (briefly) ways to construct nonstationary covariance functions proposed in literature.

$$C(\mathrm{s},\mathrm{u}) = cov(Y(\mathrm{s}), Y(\mathrm{u})) \neq C(\mathrm{s}-\mathrm{u})$$

Spatial Deformation

Sampson, P. D. and Guttorp, P. (1992) Nonparametric estimation of nonstationary spatial covariance structure, JASA, 87, 108-119.

- They propose to generate a nonstationary process by transforming a stationary process to a new coordinate system
  - For $Y(\cdot)$, we assume that there exists a function $f : \mathcal{R}^2 \rightarrow \mathcal{R}^d$:

$$C(\mathrm{s}, \mathrm{u}) = C(\|f(\mathrm{s}) - f(\mathrm{u})\|; \boldsymbol{\theta})$$

  - Here, $f(\cdot)$ is a (nonlinear) transformation from the original geographic domain to the deformed space. Usually, d is 2 or 3.
  - $f(\cdot)$ is assume to be one-to-one. In practice, people have modeled/estimated it in terms of observation locations using splines and penalty for non-smooth transformation such as the bending energy:

$$J(f) = \int\int \left[\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right] dxdy$$

  - In various papers, people have used variograms, MLE and Bayesian methods to estimate $f(\cdot)$ and $\boldsymbol{\theta}$.

- Extensions are made for space-time modeling
- It is computationally intensive, involving approximately 2 parameters per spatial monitoring site. Large problems are not practical.
  - constrained or regularized optimization with many parameters
  - MCMC in Bayesian inference involving many highly correlated parameters, making convergence problematic
- Open problem: Dimension reduction or other methods to facilitate deformation for large-scale problems?

Dimension Expansion

Bornn, L., Shaddick, G., and Zidek, J. V. (2012), Modeling non-stationary processes through dimension expansion, JASA, 107, 281-289.

- A related idea: Assuming data are stationary in 3D but we only observe the spatial coordinates in 2D
- $Y(lon, lat, Z)$ is stationary, but we don't have Z
- To construct the nonstationary structure means that we need to impute Z, requiring a joint model for $Y|Z$ and a spatial model for Z
- Bornn et al. (2012) used variograms for estimation. MLE and Bayesian inferences are also possible.
- Shand and Li (2017) extended this work to space-time modeling.
- Scale up for large data?

Weighted-Average Methods

Idea: Construct a nonstationary spatial process by **smoothing several locally stationary processes**

- Assume J independent GPs $Z_j(\mathrm{s})$ each with mean zero and covariance $C(\mathrm{s}-\mathrm{u};\boldsymbol{\theta}_j)$
- Divide the spatial region $\mathcal{D}$ into J disjoint subregions S_j and attribute each process to a subregion.
- Construct a global nonstationary process as a weighted average of the locally stationary processes:

$$Y(\mathrm{s}) = \sum_{j=1}^{J} w_j(\mathrm{s}) Z_j(\mathrm{s})$$

  - $w_j(\mathrm{s}) = I(\mathrm{s} \in S_j)$ or $w_j(\mathrm{s}) \propto exp(-\|\mathrm{s}-\mathrm{u}_j\|^2/\psi)$; u_j is the 'center' of subregion S_j
  - Choose J using BIC
  - Parameters $\boldsymbol{\theta}_j$, $j = 1, \ldots, J$ can be estimated from the responses or local MLEs from subregions.

- Fuentes, M. (2001). A high frequency kriging approach for non-stationary environmental processes. Environmetrics, 12(5):469-483.
- Kim, H.M., Mallick, B.K., and Holmes, C.C. (2005) Analyzing nonstationary spatial data using piecewise Gaussian processes. Journal of the American Statistical Association, 100, 653-658.
  - Automatically partitions the spatial domain into disjoint regions and then fits a piecewise Gaussian process model
- Nott, D.J. and Dunsmuir, W.T.M. (2002). Estimation of nonstationary spatial covariance structure. Biometrika, 89, 819-829.
  - Propsing

$$C(Y(\mathrm{s}), Y(\mathrm{u})) = \Sigma_0 + \sum_{j=1}^{J} w_j(\mathrm{s}) w_j(\mathrm{u}) C_{\theta_j}(\mathrm{s} - \mathrm{u})$$

where the second term is fitted based on local residual covariance structure.

Basis Function Models

Idea: Decompose spatial covariance functions in terms of basis functions

- The Karhunen-Loéve (KL) expansion of a covariance function (no need to assume stationarity) is

$$C(\mathrm{s}, \mathrm{u}) = \sum_{k=1}^{\infty} \lambda_k \phi_k(\mathrm{s}) \phi_k(\mathrm{u})$$

  - $\phi_k(\cdot)$ are eigenfunctions with eigenvalues λ_k
- Then after truncation

$$Y(\mathrm{s}) = \sum_{k=1}^{K} a_k \phi_k(\mathrm{s})$$

- In practice, $\phi_k(\cdot)$ can be obtained when we have repeated observations (e.g., over time)

$$\hat{\Sigma}_Y = \Phi \Lambda \Phi'$$

  - Φ is the matrix of eigenvectors, called the empirical orthogonal functions or EOFs
  - Λ is the corresponding diagonal matrix with eigenvalues on the diagonal
- We then use $\Phi\boldsymbol{\alpha}$ in place of Y and treat $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)'$ as unknown parameters/random effects
  - Naturally nonstationary
  - EOFs not available without repeated observation; not adaptive for prediction; not incorporating measurement errors directly

Process Convolution

Idea: Using a constructive specification to induce nonstationarity

- Any spatial GP can be written as

$$Y(\mathrm{s}) = \int K(\mathrm{s}, \mathrm{u}) Z(\mathrm{u}) d\mathrm{u}$$

  where $K(\cdot, \cdot)$ is a kernel function and $Z(\cdot)$ is a white noise process

- Example: $K(\mathrm{s}, \mathrm{u}) = exp(-\phi \|\mathrm{s} - \mathrm{u}\|^2)$. Can you derive $Cov(Y(\mathrm{s}), Y(\mathrm{u}))$? (Assuming mean zero)

$$C(\mathrm{s}, \mathrm{u}) = \sigma^2 \int K(\mathrm{s}, \mathrm{v}) K(\mathrm{u}, \mathrm{v}) d\mathrm{v}$$

Higdon, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic ocean, Environmental and Ecological Statistics, 5, 173-190.

- Higdon (1998) propose a discrete approximation:

$$Y(\mathrm{s}) = \sum_{i=1}^{J} K(\mathrm{s}, \mathrm{u}_i) x_i$$

where x_i's are iid $N(0, \lambda^2)$ associated with knot location u_i

- The kernels are further assumed to be weighted average of fixed basis kernels

$$K(\mathrm{s}, \mathrm{u}_i) = \sum_{j=1}^{M} w_j(\mathrm{s}) K_j^*(\mathrm{s}, \mathrm{u}_i)$$

$$w_j(\mathrm{s}) \propto exp\left(-\frac{1}{2}\|\mathrm{s} - \mathrm{s}_j^*\|^2\right)$$

$$K_j^*(\mathrm{s}, \mathrm{u}_i) = \frac{1}{\sqrt{2\pi}}|\Sigma_{\mathrm{s}_j^*}|^{-1} exp\left(-\frac{1}{2}(\mathrm{s} - \mathrm{u}_i)'\Sigma_{\mathrm{s}_j^*}^{-1}(\mathrm{s} - \mathrm{u}_i)\right)$$

Extension: Letting the Kernel parameters be spatial varying

$$K(\mathrm{s},\mathrm{u}) = exp(-\phi(\mathrm{s})\|\mathrm{s}-\mathrm{u}\|^2)$$

where $\phi(\mathrm{s})$ is a spatially-varying kernel bandwidth.

- Paciorek and Schervish (2006) and Stein (2005) use this idea to develop a general class of nonstationary covariance functions (including the Matérn model):

$$C(\mathrm{s},\mathrm{u}) = \sigma(\mathrm{s})\sigma(\mathrm{u})|\Sigma(\mathrm{s})|^{1/4}|\Sigma(\mathrm{u})|^{1/4}|\frac{\Sigma(\mathrm{s})+\Sigma(\mathrm{u})}{2}|^{-1/2}g(-\sqrt{Q(\mathrm{s},\mathrm{u})})$$

  where $g(\cdot)$ is a valid correlation function (e.g., the stationary Matérn correlation function).
  - Spatially-varying standard deviation $\sigma(\cdot)$
  - $\Sigma(\cdot)$ is $d \times d$ spatially-varying local anisotropy and

$$Q(\mathrm{s},\mathrm{u}) = (\mathrm{s}-\mathrm{u})'\left(\frac{\Sigma(\mathrm{s})+\Sigma(\mathrm{u})}{2}\right)^{-1}(\mathrm{s}-\mathrm{u})$$

    is a Mahalonabis distance.
  - It is possible to allow $k(\mathrm{s})$, spatially-varying smoothness parameter. Then in $g(\cdot)$ we use $[k(\mathrm{s})+k(\mathrm{u})]/2$.

- We no longer need to specify the kernel functions
- Kleiber, W. and Nychka, D. (2012). Nonstationary modeling for multivariate spatial processes. Journal of Multivariate Analysis, 112, 76-91.
  - further extend this model to the multivariate setting.
- Calder, C. A. (2008). A dynamic process convolution approach to modeling ambient particulate matter concentrations. Environmetrics, 19, 39-48. AND Calder, C.A. (2007). Dynamic factor process convolution models for multivariate space-time data with application to air quality assessment. Environmental and Ecological Statistics, 14, 229-247.
  - extend to space-time versions of the Hidgon model
- Heaton, M. J., Katzfuss, M., Berrett, C., and Nychka, D.W. (2014). Constructing valid spatial processes on the sphere using kernel convolutions, Environmetrics, 25, 2-14.
  - extends process convolution models to spherical spatial domains

Discussion

- Computation can be a challenge for many methods in literature
- Recent work related to computation:
  - Risser M. and Calder. C.A. (2015) Local Likelihood Estimation for Covariance Functions with Spatially-Varying Parameters: The `convoSPAT` Package for R. Journal of Statistical Software. **discretized basis kernel approach as in Higdon (1998)** and local likelihood estimation (rather than optimizing the full log-likelihood)
  - Li, Y. and Sun, Y. (2019) Efficient estimation of nonstationary spatial covariance functions with application to high-resolution climate model emulation, Statistica Sinica, 29, 1209-1231. **Local polynomial approximation**

Reading Assignments - without submission

- Risser M. and Calder. C.A. (2015) Local Likelihood Estimation for Covariance Functions with Spatially-Varying Parameters: The `convoSPAT` Package for R. Journal of Statistical Software.
- Li, Y. and Sun, Y. (2019) Efficient estimation of nonstationary spatial covariance functions with application to high-resolution climate model emulation, Statistica Sinica, 29, 1209-1231.
- Andrew Zammit-Mangion, Tin Lok James Ng, Quan Vu and Maurizio Filippone (2021) Deep Compositional Spatial Models, Journal of the American Statistical Association, DOI: 10.1080/01621459.2021.1887741.

Summary

- Nonstationary methods

Preview:

- Spatio-temporal modeling