Cressie and Wikle: Chapter 6

Serge Aleshin-Guendel

February 12, 2018

Notation

- We're going to be looking at spatio-temporal random processes of the form

$$\{Y(s,t) | s \in D_s, t \in D_t\}$$

  where D_s, D_t are index sets for space and time respectively.
- Want to model current value $Y(s,t)$, conditional on past values and current spatially close values, i.e.

$$Y(s,t) | \{Y(x,r) | x \in D_s, r < t\} \cup \{Y(x,t) | x \in D_s\}$$

Approaches to Spatio-Temporal Modelling

- Two approaches to space-time models that Cressie and Wikle emphasize: descriptive and dynamical
- When there's knowledge of process dynamics, you should use that to guide your modeling (this is the approach in chapter 7 that Johnny covered)
- When there's no knowledge of process dynamics, spatio-temporal covariance functions are compact and informative summaries of space-time processes, and can be used as building blocks for space-time models (e.g. kriging/Gaussian process regression)
- Gaussian processes and other related methods *describe* spatio-temporal variability, while methods motivated by dynamical systems *explain* spatio-temporal variability

Some Motivation

- Write out your space-time process as

$$Y(s,t) = \mu(s,t) + \beta(s) + \gamma(t) + \kappa(s,t) + \delta(s,t)$$

  where μ is the deterministic mean, and $\beta, \gamma, \kappa, \delta$ are all zero mean random effects
- Difference between κ and δ is that they're supposed to capture large and small scale spatial temporal variability respectively. δ assumed to be white noise
- All of the random effects have their own covariance functions
- Book describes classic versions of this model, where assumptions are made so that the spatial and temporal aspects are handled separately, in order to get covariance functions for $Y(s,t)$ that are manageable

Definition of Covariance Functions

Let $f : (D_s \times D_t) \times (D_s \times D_t) \to \mathbb{R}$

- f corresponds to a spatio-temporal covariance function if it's positive semi-definite, that is if the gram matrix generated by f and a finite number of points in $D_s \times D_t$ is positive semi-definite
- Most covariance functions that you'll work with are positive definite

Stationarity of Covariance Functions

- A spatio-temporal covariance function is stationary if it can be written for any $s, x \in D_s$ and $t, r \in D_t$ as

$$f((s,t),(x,r)) = C(s-x, t-r)$$

- So it only depends on the spatial and temporal lags h and τ
- For fixed τ, $C(h,\tau)$ is a stationary spatial covariance function
- for fixed h, $C(h,\tau)$ is a stationary temporal covariance function
- Possible to define a covariance function that's stationary in only space or only time
- Can define the stationary spatio-temporal correlation function as

$$\rho(h,\tau) = C(h,\tau)/C(0,0)$$

Separability

- A covariance function is separable, if it can be written for any $s, x \in D_s$ and $t, r \in D_t$ as

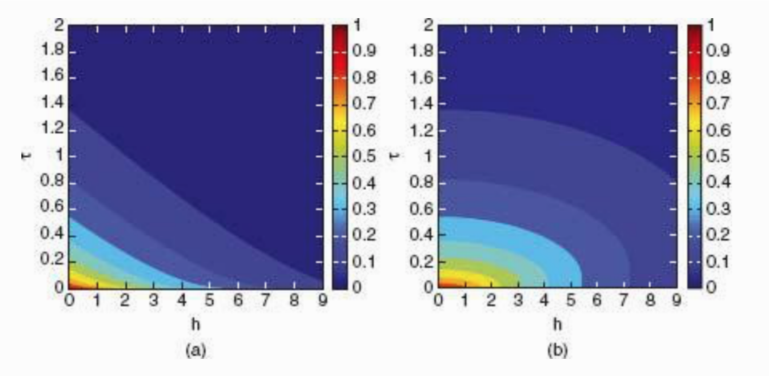
$$f((s,t),(x,r)) = C^s(s,x)C^t(t,r)$$

- When C^s, C^t are both stationary, this reduces to

$$f((s,t),(x,r)) = C^s(s-x)C^t(t-r)$$

- We can then visually test separability by comparing the contours of $C(h,\tau)$ and $C(h,0) \times C(0,\tau)$, or alternatively the contours of $\rho(h,\tau)$ and $\rho(h,0) \times \rho(0,\tau)$

Separability



(a)
(b)

(b) shows a correlation function $\rho(h,\tau)$ and (a) shows $\rho(h,0) \times \rho(0,\tau)$. They're not the same, so $C(h,\tau)$ is non-separable

Full Symmetry

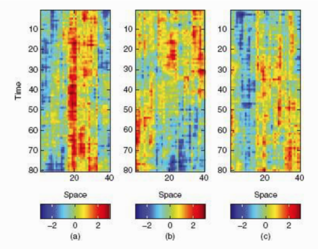
- A generalization of separability is full symmetry, which occurs if a covariance function can be written for any $s, x \in D_s$ and $t, r \in D_t$ as

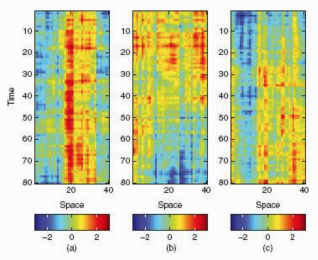
$$f((s,t),(x,r)) = f((s,r),(x,t))$$

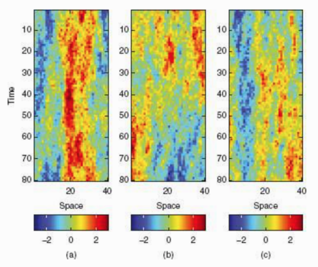
- Full symmetry implies separability, but separability doesn't imply full symmetry
- Covariance functions derived from dynamical systems aren't usually separable or fully symmetric. Both conditions are strong, and usually imply regularity conditions that aren't actually met in practice

Separability and Full Symmetry

Three realizations from Gaussian processes with non-separable, fully symmetric, and separable covariance matricies. Which is which?







Separability and Full Symmetry

- Easy to diagnose separability from covariance functions
- Hard to diagnose separability from realizations of processes with certain covariance functions

Spectral Representations

- By Bochner's Theorem, we can represent a stationary covariance function in terms of a non-negative function $f(\omega, \xi)$, called the spectral density:

$$C(h, \tau) = \int \int e^{-ih^T\omega - i\tau\xi} f(\omega, \xi) d\omega d\xi$$

- Equivalently, a non-negative function $f(\omega, \xi)$ can be represented in terms of a positive semi-definite function $C(h, \tau)$:

$$f(\omega, \xi) = (2\pi)^{-(d+1)} \int \int e^{-ih^T\omega - i\tau\xi} C(h, \tau) dh d\tau$$

- Can recognize these two representations as an inverse Fourier transform and a Fourier transform respectively

Spectral Representations

- These representations give us two new ways to create covariance functions
- Verifying positive semi-definiteness of a function is hard. If we have a potential covariance function in mind, you can find it's spectral density and verify that it's non-negative
- Alternatively, you can start wish a non-negative function $f(\omega, \xi)$ and use the induced positive semi-definite function as a covariance function
- Covariance functions induced by non-negative functions are usually going to be non-separable!

Spatio-Temporal Kriging (Gaussian Process Regression)

- Suppose we have observed data $Z(s_i, t_{ij})$, where $j \in \{1, \cdots, T_i\}$ and $i \in \{1, \cdots, m\}$
- Assume the model $Z(s_i, t_{ij}) = Y(s_i, t_{ij}) + \varepsilon(s_i, t_{ij})$, where ε is independent of Y, representing i.i.d. mean zero and variance σ_ε^2 measurement error, and Y is the latent spatio-temporal process
- Goal is to predict $Y(s_0, t_0)$ for unobserved s_0, t_0

Spatio-Temporal Kriging (Gaussian Process Regression)

- Assume that $Y \sim \text{GP}(\mu, f)$ where μ is a known mean function and f is a covariance function with no unknown parameters
- Assume that $\epsilon \sim \text{Norm}(0, \sigma_\varepsilon^2)$
- Can write

$$Z^i = [Z(s_i, t_{ij}) | j \in \{1, \cdots, T_i\}]^T,$$

$$Z = [(Z^1)^T \cdots (Z^m)^T]^T,$$

$$\mu_Z = [\mu(s_i, t_{ij}) | j \in \{1, \cdots, T_i\}, i \in \{1, \cdots, m\}]^T$$

Spatio-Temporal Kriging (Gaussian Process Regression)

- Let $Cov(Z) = f(Z, Z) + \sigma_\varepsilon^2 I$,
  $Var(Y(s_0, t_0)) = f(Y(s_0, t_0), Y(s_0, t_0))$, and
  $C_0 = Cov(Y(s_0, t_0), Z) = f(Y(s_0, t_0), Z)$
- The joint of the observed and unobserved data is

$$\begin{bmatrix} Y(s_0, t_0) \\ Z \end{bmatrix} \sim \text{Norm}\left(\begin{bmatrix} \mu(s_0, t_0) \\ \mu_Z \end{bmatrix}, \begin{bmatrix} Var(Y(s_0, t_0)) & C_0^T \\ C_0 & Cov(Z) \end{bmatrix} \right)$$

- Through standard properties of the multivariate Gaussian, the posterior predictive distributions is

$$Y(s_0, t_0) | Z \sim \text{Norm}(\mu_{post}, \Sigma_{post}),$$

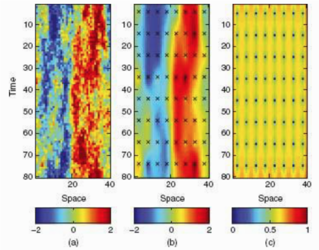
where $\mu_{post} = \mu(s_0, t_0) + C_0^T Cov(Z)^{-1}(Z - \mu_Z)$ and
$\Sigma_{post} = Var(Y(s_0, t_0)) - C_0^T Cov(Z)^{-1} C_0$.

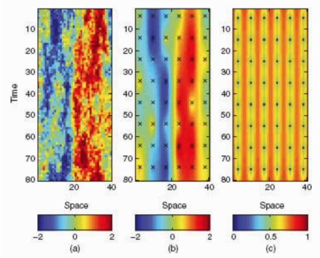
Spatio-Temporal Kriging (Gaussian Process Regression)

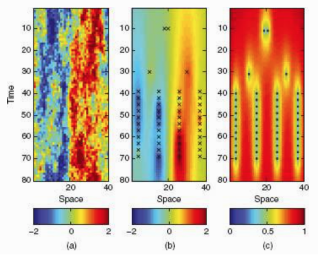
- Can generalize by placing prior on σ_ε^2, covariance function parameters, etc.
- There are no restrictions on f, can be separable, non-separable, stationary, non-stationary, etc.
- However, have to compute the inverse of the gram matrix, which is of the order $\mathcal{O}((\sum_{i=1}^{m} T_i)^3)$ in general. Separability brings this cost down since gram matrix will have Kronecker product structure

Spatio-Temporal Kriging (Gaussian Process Regression)

Example of kriging on realization of stationary Gaussian process, with varying sample sizes (64, 48, 48)







Spatio-Temporal Point Processes

(Going to be changing notation slightly for point processes)

- A spatio-temporal point process is a stochastic counting process on a bounded subset $D_{s,t}$ of $\mathbb{R}^d \times \mathbb{R}$
- Usually we let $D_{s,t} = D_t \times [0, T]$ where T is the largest observed time
- For $A \subset D_{s,t}$, we let $Z(A)$ be the number of events in A. So $\{Z(A) | A \subset D_{s,t}\}$ completely characterizes the process

Spatio-Temporal Point Processes

Some examples:

- Homogeneous Poisson Process: Given a constant intensity λ_0, $Z(A) \sim \text{Pois}(\lambda_0 |A|)$, where $|A|$ is the volume of A
- Inhomogeneous Poisson Process: Given intensity function $\lambda(s, t)$, $Z(A) \sim \text{Pois}(\int_A \lambda(s, t) ds dt)$ (so the homogeneous PP is a special case)

Spatio-Temporal Point Processes

If we have an inhomogeneous Poisson process, we can write down the likelihood using a conditonal intensity function

- Suppose we have the history of the counting process up to time t, denoted by $\mathcal{H}_t$. Then we can define the conditional intensity of a process as

$$\psi(s,t|\theta) = \lim_{|ds|,dt\to 0} \frac{\mathbb{E}(Z(ds,dt)|\mathcal{H}_t)}{|ds||dt|}$$

  where θ are any parameters governing the process
- Writing the events of Z be written as $\{(s_i,t_i)|i\in\{1,\cdots,N\}\}$, where $N = Z(D_{s,t})$, we can then write the likelihood as

$$\mathcal{L}(\theta|Z) \propto \prod_{i=1}^{N} \psi(s,t|\theta)\exp\left[\int_{D_{s,t}} \psi(s,t|\theta)dsdt\right]$$

- Evaluating the integral in the likelihood can be tough, and the condtional intensity can be hard to interpret in actual data settings

Spatio-Temporal Point Processes

- Can generalize the inhomogenous Poisson process in a hierarchical fashion, by letting the intensity function be a stochastic process
- This results in a Cox process in general
- If you let $Y(s,t) \sim \text{GP}$, and define the intensity function as $\lambda(s,t) = \exp[Y(s,t)]$, you arrive at a log Gaussian Cox process