STAT 510: Spatiotemporal Stats

Introduction and Exploring SPT Data

Prof. Taylor-Rodriguez

Spatio-Temporal Data Exploration

Flavors of temporal and spatial data

As I mentionend last time, this field is BIG, and methodology has branched off according to the specific flavor of the data.

In time

- Time intervals type: regular or irregular?
- Is time discrete or continuous?
- Is the random event the time at which an event occurs?

Flavors of temporal and spatial data

In space:

Spatial data corresponds to either a fixed time-point or a temporal aggregation over multiple timepoints and can be:

- Geostatistical: point-level/coordinate observations, space is treated as a continuous surface over a given spatial domain
- Areal: space is discretized into a grid, polygons or small areas
- Point-process: the locations where events occur are themselves random, and each location can be complemented with attributes (called marks)

Spatio-temporal data

Obviously spatio-temporal data will be defined by the combination of features in time and space the data have.

For this course, we'll focus on data that is discrete in time and *geostatistical* or *areal* in space

Exploratory Analysis of ST Data

Empirical Spatial and Temporal Statistics

Denote by $\{Z(s_i, t_j)\}_{i=1,j=1}$ the set of observations at

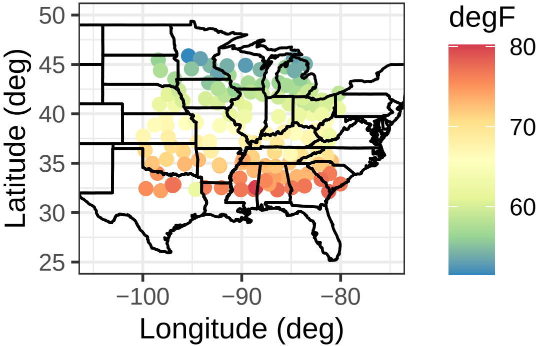
- locations $\{s_i : i = 1, \ldots, m\}$, and
- times $\{t_j : j = 1, \ldots, T\}$

The empirical spatial means vector is

$$\hat{\boldsymbol{\mu}}_{z,s} = \begin{bmatrix} \hat{\mu}_{z,s}(s_1) \\ \vdots \\ \hat{\mu}_{z,s}(s_m) \end{bmatrix} = \begin{bmatrix} \frac{1}{T}\sum_{j=1}^{T} Z(s_1, t_j) \\ \vdots \\ \frac{1}{T}\sum_{j=1}^{T} Z(s_m, t_j) \end{bmatrix} = \frac{1}{T}\sum_{j=1}^{T} \mathbf{Z}_{t_j}$$

Empirical Spatial and Temporal Statistics

Empirical Spatial Means (NOAA Data)

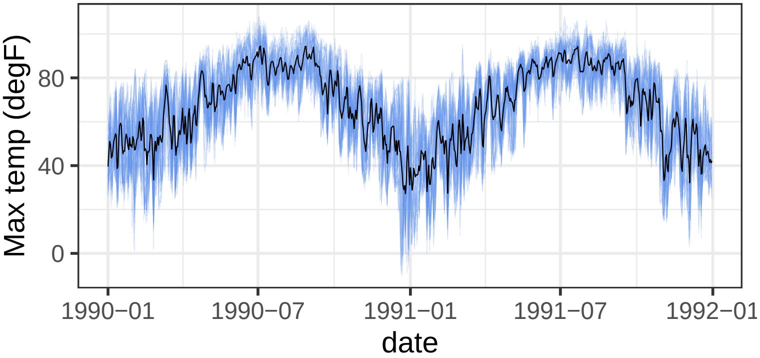


{In these data, stations have different number of measurements, so we use instead $\hat{\mu}_{z,s}(s_i) = \frac{1}{T_i}\sum_{j=1}^{T_i} Z(s_i, t_j)$}

Empirical Spatial and Temporal Statistics

Conversely, one could average over space to get the *empirical temporal mean*, given by

$$\hat{\mu}_{z,t}(t_j) = \frac{1}{m}\sum_{i=1}^{m} Z(s_i, t_j)$$



Empirical Spatial and Temporal Statistics

For 2 locations, the lag-τ empirical spatial covariance is

$$\hat{C}_z^{(\tau)}(s_i, s_k) = \frac{1}{T-\tau} \sum_{j=\tau+1}^{T} (Z(s_i, t_j) - \hat{\mu}_{z,s}(s_i))(Z(s_k, t_j - \tau) - \hat{\mu}_{z,s}(s_k))$$

for $\tau = 0, 1, \ldots, T-1$.

For all locations, the lag-τ empirical spatial covariance matrix is

$$\hat{\mathbf{C}}_z^{(\tau)} = \frac{1}{T-\tau} \sum_{j=\tau+1}^{T} (\mathbf{Z}_{t_j} - \hat{\boldsymbol{\mu}}_{z,s})(\mathbf{Z}_{t_j-\tau} - \hat{\boldsymbol{\mu}}_{z,s})'$$

where $\mathbf{Z}_{t_j} = (Z(s_1, t_j), Z(s_2, t_j), \ldots, Z(s_m, t_j))'$, and $\hat{\boldsymbol{\mu}}_{z,s}$ is defined analogously.

Empirical Spatial and Temporal Statistics

Between two different datasets, the lag-τ empirical cross-covariance matrix is

$$\hat{\mathbf{C}}_{z,x}^{(\tau)} = \frac{1}{T-\tau} \sum_{j=\tau+1}^{T} (\mathbf{Z}_{t_j} - \hat{\boldsymbol{\mu}}_{z,s})(\mathbf{X}_{t_j-\tau} - \hat{\boldsymbol{\mu}}_{x,s})'$$

Can be used to characterize ST dependence in two variables

Measures of Spatio-Temporal dependence

Empirical ST Covariogram

We want to describe covariability in the ST data as a function of temporal lags and space. Assume that

- the first moment depends on space but not on time
- the second moment depends on lag differences in space and time

Then the ST covariogram for spatial and temporal lags $\mathbf{h}$ and τ, respectively, is

{

$$\hat{\mathbf{C}}_z(\mathbf{h},\tau) = \frac{1}{|N_s(\mathbf{h})|}\frac{1}{|N_t(\tau)|}\sum_{s_i,s_k\in N_s(\mathbf{h})}\sum_{t_j,t_\ell\in N_t(\tau)}(Z(s_i,t_j)-\hat{\mu}_{z,s}(s_i))(Z(s_k,t_\ell)-\hat{\mu}_{z,s}(s_k))$$

}

Measures of Spatio-Temporal dependence

Empirical ST Semivariogram

$$\gamma_z(s_i, s_k; t_j, t_\ell) = \frac{1}{2}\mathrm{var}(Z(s_i; t_j) - Z(s_k; t_\ell)).$$

If the covariance is only in terms of displacement in space and differences in time, then

$$\begin{aligned}\gamma_z(s_i, s_k; t_j, t_\ell) &= \mathrm{var}(Z(s + \mathbf{h}; t + \tau) - Z(s; t)) \\ &= C_z(\mathbf{0}, 0) - C_z(\mathbf{h}, \tau).\end{aligned}$$

with $\mathbf{h} = s_i - s_k$ and $\tau = t_j - t_\ell$.

{Note: This last equation doesn't hold even if $\exists \gamma_z(\mathbf{h}; \tau)$ if there is no $C_z(\mathbf{h}; \tau)$ that is stationary.}

In class-problem

Show, under the conditions previously assumed, that

$$\mathrm{var}(Z(s+\mathbf{h};t+\tau)-Z(s;t)) \quad = \quad C_z(\mathbf{0},0)-C_z(\mathbf{h},\tau).$$

Measures of Spatio-Temporal dependence

ST Semivariogram

If $C_z(\mathbf{h};\tau)$ is well-defined, the the semivariogram depends on

Nugget effect

$$\lim_{\substack{\mathbf{h}\to\mathbf{0}\\ \tau\to 0}} \gamma_z(\mathbf{h};\tau)$$

Sill:

$$\lim_{\substack{\mathbf{h}\to\infty\\ \tau\to\infty}} \gamma_z(\mathbf{h};\tau)$$

Partial sill: Sill minus Nugget

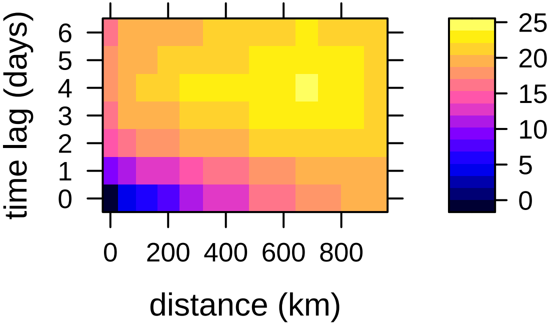
Note: Before fitting a semivariogram, non-stationary behavior in $C_z(\mathbf{h};\tau)$ may be avoided by fitting (and removing) linear and quadratic trends in ST coordinates.

Measures of Spatio-Temporal dependence

Empirical ST Semivariogram

Assuming that $\mu_{z,s}$ is constant across all locations, then

$$\begin{aligned}\hat{\gamma}_z(\mathbf{h};\tau) &= \hat{C}_z(\mathbf{0};0) - \hat{C}_z(\mathbf{h};\tau) \\ &= \frac{1}{|N_s(\mathbf{h})|}\frac{1}{|N_t(\tau)|}\sum_{s_i,s_k\in N_s(\mathbf{h})}\sum_{t_j,t_\ell\in N_t(\tau)}(Z(s+\mathbf{h},t+\tau)-Z(s,t))^2\end{aligned}$$



Empirical Orthogonal Functions

- Help identify spatial patterns in ST data
- Can be used for dimension reduction
- Are ST manifestation of Principal Components

Interlude: A Review of PCA

- Assume iid random $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ with $p \times p$ covariance matrix $\text{var}(\mathbf{x}_i) = \mathbf{C}_x$.
- Suppose we want to build variables

$$a_{ik} = w_{k1}x_{i1} + w_{k2}x_{i2} + \cdots + w_{kp}x_{ip}$$

  for $k = 1, \ldots, p$, that are *orthogonal* to each other.
- How should the w_{kj}'s be chosen?

Interlude: A Review of PCA

Using spectral decomposition:

$$\mathbf{C}_x = W \Lambda_x W',$$

with $WW' = W'W = \mathbf{I}_p$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_p)$, with

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0.$$

Dropping the i

$$\mathbf{a} = W'\mathbf{x} = \begin{bmatrix} \mathbf{w}_1'\mathbf{x} \\ \vdots \\ \mathbf{w}_p'\mathbf{x} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix},$$

Note that $\text{var}(\mathbf{a}) = W'\text{var}(\mathbf{x})W = W'(W\Lambda_x W')W = \Lambda_x$.

Interlude: A Review of PCA

- Since Λ_x is diagonal then the a_k's are independent (orthogonal to each other)
- Furthermore, note that

$$\begin{aligned}
\mathrm{var}(x_1) + \cdots + \mathrm{var}(x_p) &= \mathrm{trace}(C_x) \\
&= \lambda_1 + \cdots + \lambda_p \\
&= \mathrm{var}(a_1) + \cdots + \mathrm{var}(a_p)
\end{aligned}$$

- Lastly, since $\lambda_1 \geq \ldots \geq \lambda_p$.
  a_1 captures the largest amount of variability in $\mathbf{x}$, a_2 the second largest, etc.

Note:In practice $\mathbf{C}_x$ is unknown, so we use the estimator $\hat{\mathbf{C}}_x$ instead.

Empirical Orthogonal Functions

- Replace $\mathbf{Z}_{t_j}$ for $\mathbf{x}_i$, and use the lag-0 empirical spatial covariance matrix $\hat{\mathbf{C}}_z^{(0)}$
- Spectrally decompose $\hat{\mathbf{C}}_z^{(0)}$ into $\Psi\Lambda\Psi$, where

$$\begin{aligned} \Psi_{m\times m} &= (\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_m)' \text{ with} \\ \boldsymbol{\psi}_k &= (\psi_k(s_1), \ldots, \psi_k(s_m))'. \end{aligned}$$

- The kth principal component times series is

$$\mathbf{a}_k = \begin{bmatrix} a_k(t_1) \\ a_k(t_2) \\ \vdots \\ a_k(t_T) \end{bmatrix} = \begin{bmatrix} \psi_k'\mathbf{Z}_{t_1} \\ \psi_k'\mathbf{Z}_{t_2} \\ \vdots \\ \psi_k'\mathbf{Z}_{t_T} \end{bmatrix}.$$

Empirical Orthogonal Functions

The ψ_k's ($k = 1, \dots, m$) are known as the Empirical Orthogonal Functions (EOFs). Main uses are:

1. Identify salient spatial patterns in ST data
2. Reduce the dimensionality of spatial or ST random effects (more on this later)

Empirical Orthogonal Functions

To do dimension reduction, we need to decide the number of EOFs to use. How?

Here are some possibilities

1. Use the λ_k's to determine the proportion of overall variance to retain

2. Use scree plot and find after including how many EOFs the variance contribution flattens out

3. Identify the *significant* EOFs by obtaining EOFs for several spatially permuted versions of the data. See where the tue scree plot intersects with those of the permuted datasets

Empirical Orthogonal Functions

A more efficient method to obtain EOF's

Note that letting

$$\tilde{\mathbf{Z}} = \frac{1}{\sqrt{T-1}}(\mathbf{Z} - \mathbf{1}_T \hat{\mu}'_{z,s})$$

Implies that

$$\Longrightarrow \hat{\mathbf{C}}_z^{(0)} = \tilde{\mathbf{Z}}'\tilde{\mathbf{Z}} = \mathbf{\Psi}\mathbf{\Lambda}\mathbf{\Psi}'$$

Also, considering the SVD $\tilde{\mathbf{Z}} = \mathbf{UDV}'$, we have

$$\hat{\mathbf{C}}_z^{(0)} = (\mathbf{UDV}')'(\mathbf{UDV}') = \underbrace{\mathbf{V}}_{\Psi}\underbrace{\mathbf{D}'\mathbf{D}}_{\Lambda}\underbrace{\mathbf{V}'}_{\Psi'}$$

This equality implies that SVD can be used (and is recommended) to obtain the EOF's.

ST Canonical Correlation Analysis (CCA)

- **Goal:** create 2 new sets of variables, which arise as linear combinations of two multivariate (mvt) datasets and that are maximally correlated
- The first few pairs of canonical variables are very similar (since maximaly correlated)
- The spatial patterns in the weights show the areas in space that are most responsible for the high correlations

In the ST case, each *spatial location* corresponds to a *variable*, and each *time-point* corresponds to a *sample*

ST Canonical Correlation Analysis (CCA)

This, from two datasets denoted by

$$
\begin{gathered}
\{\mathbf{Z}_{t_t} = (Z(s_1;t_j), \ldots, Z(s_m;t_j))' : j = 1, \ldots, T\} \\
\{\mathbf{X}_{t_t} = (X(r_1;t_j), \ldots, X(r_n;t_j))' : j = 1, \ldots, T\}
\end{gathered}
$$

ST-CCA, for $k = 1, 2, \ldots, \min\{m, n\}$, seeks to generate variables

$$
\begin{aligned}
a_k(t_j) &= \sum_{i=1}^{m} \xi_{ik} Z(s_i;t_j) = \boldsymbol{\xi}_k' \mathbf{Z}_{t_j} \\
b_k(t_j) &= \sum_{\ell=1}^{n} \psi_{\ell k} X(r_\ell;t_j) = \boldsymbol{\psi}_k' \mathbf{X}_{t_j}
\end{aligned}
$$

with weights $\boldsymbol{\xi}_k$ and $\boldsymbol{\psi}_k$ a.k.a. *canonical variables*

In-class Problem

The kth canonical correlation r_k is given by

$$r_k = \mathrm{corr}(a_k, b_k) = \frac{\boldsymbol{\xi}_k' \mathbf{C}_{z,x}^{(0)} \boldsymbol{\psi}_k}{(\boldsymbol{\xi}_k' \mathbf{C}_z^{(0)} \boldsymbol{\xi}_k)^{1/2} (\boldsymbol{\psi}_k' \mathbf{C}_x^{(0)} \boldsymbol{\psi}_k)^{1/2}}$$

show this result.

ST Canonical Correlation Analysis (CCA)

Calculating ST-CCA in practice

$$
\begin{aligned}
r_k = \mathrm{corr}(a_k, b_k) &= \frac{\xi_k' \mathbf{C}_{z,x}^{(0)} \psi_k}{(\xi_k' \mathbf{C}_z^{(0)} \xi_k)^{1/2} (\psi_k' \mathbf{C}_x^{(0)} \psi_k)^{1/2}} \\
&= \frac{\overbrace{\xi_k' (\hat{\mathbf{C}}_z^{(0)})^{1/2}}^{\equiv \tilde{\xi}_k'} (\hat{\mathbf{C}}_z^{(0)})^{-1/2} \mathbf{C}_{z,x}^{(0)} (\hat{\mathbf{C}}_x^{(0)})^{-1/2} \overbrace{(\hat{\mathbf{C}}_x^{(0)})^{1/2} \psi_k}^{\equiv \tilde{\psi}_k}}{(\xi_k' (\hat{\mathbf{C}}_z^{(0)})^{1/2} (\hat{\mathbf{C}}_z^{(0)})^{1/2} \xi_k)^{1/2} (\psi_k' (\hat{\mathbf{C}}_x^{(0)})^{1/2} (\hat{\mathbf{C}}_x^{(0)})^{1/2} \psi_k)^{1/2}} \\
&= \frac{\tilde{\xi}_k' (\hat{\mathbf{C}}_z^{(0)})^{-1/2} \mathbf{C}_{z,x}^{(0)} (\hat{\mathbf{C}}_x^{(0)})^{-1/2} \tilde{\psi}_k}{(\tilde{\xi}_k' \tilde{\xi}_k)^{1/2} (\tilde{\psi}_k' \tilde{\psi}_k)^{1/2}}
\end{aligned}
$$

It is well-known that from the SVD of

$$
(\hat{\mathbf{C}}_z^{(0)})^{-1/2} \mathbf{C}_{z,x}^{(0)} (\hat{\mathbf{C}}_x^{(0)})^{-1/2},
$$

r_1^2 corresponds to the largest singular value, and $\tilde{\xi}_1$ and $\tilde{\psi}_1$ are the first left and right singular vectors, respectively.

ST Canonical Correlation Analysis (CCA)

Calculating ST-CCA in practice

```
data("SSTlandmask", package = "STRbook")
data("SSTlonlat", package = "STRbook")
data("SSTdata", package = "STRbook")

delete_rows <- which(SSTlandmask == 1)
SSTdata <- SSTdata[-delete_rows, 1:396]

## Put data into space-wide form
Z <- t(SSTdata)
```

ST Canonical Correlation Analysis (CCA)

Calculating ST-CCA in practice

```
## First find the matrix we need to subtract:
spat_mean <- apply(SSTdata, 1, mean)
nT <- ncol(SSTdata)

## Then subtract and standardize:
Zspat_detrend <- Z - outer(rep(1, nT), spat_mean)
Zt <- 1/sqrt(nT - 1)*Zspat_detrend

## get the SVD
E <- svd(Zt)
```

ST Canonical Correlation Analysis (CCA)

Calculating ST-CCA in practice

```
nEOF <- 10
EOFset1 <- E$u[1:(nT-7), 1:nEOF] * sqrt(nT - 1)
EOFset2 <- E$u[8:nT, 1:nEOF] * sqrt(nT - 1)

cc <- cancor(EOFset1, EOFset2) # compute CCA

options(digits = 3)
CCA_df <- data.frame(t = 1:(nT - 7),
                     CCAvar1 = (EOFset1 %*% cc$xcoef[,1])[,1],
                     CCAvar2 = (EOFset2 %*% cc$ycoef[,1])[,1])
t_breaks <- seq(1, nT, by = 60) # breaks for x-labels
year_breaks <- seq(1970,2002,by=5) # labels for x-axis
ggplot(CCA_df) +
  geom_line(aes(t, CCAvar1), col = "dark blue") +
  geom_line(aes(t, CCAvar2), col = "dark red") +
  scale_x_continuous(breaks = t_breaks, labels = year_breaks) +
  ylab("CCA variables") + xlab("Year") + theme_bw()
```

ST Canonical Correlation Analysis (CCA)

Calculating ST-CCA in practice