Brief Introduction to Spatio-Temporal Statistics

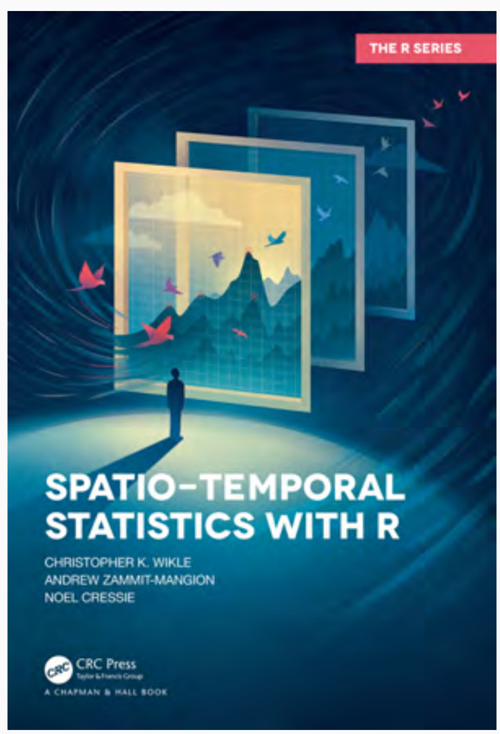
Christopher K. Wikle
Curators' Distinguished Professor and Chair of Statistics

University of Missouri

September 13, 2019

This Portion of the Course

- The portion of the course is based on the recently published book "Spatio-temporal Statistics with R" by Christopher K. Wikle, Andrew Zammit-Mangion, and Noel Cressie.
- Book downloadable for free from https://spacetimewithr.org
- These slides are extracted from a copyrighted short course developed by the authors.

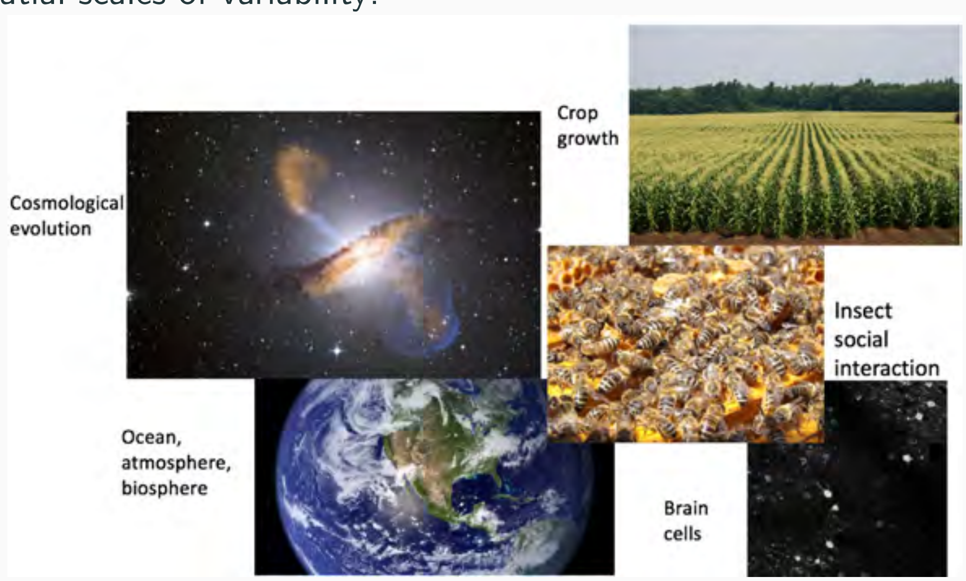


Spatio-Temporal Data are Not New

Lienzo de Quauhquechollan (digitally restored): from the indigenous people of Guatemala; documents the space and time history of the Spanish conquest from 1527 to 1530.

Complex System

Our universe is a complex system of interacting physical, biological, and social processes across a huge range of time and spatial scales of variability!

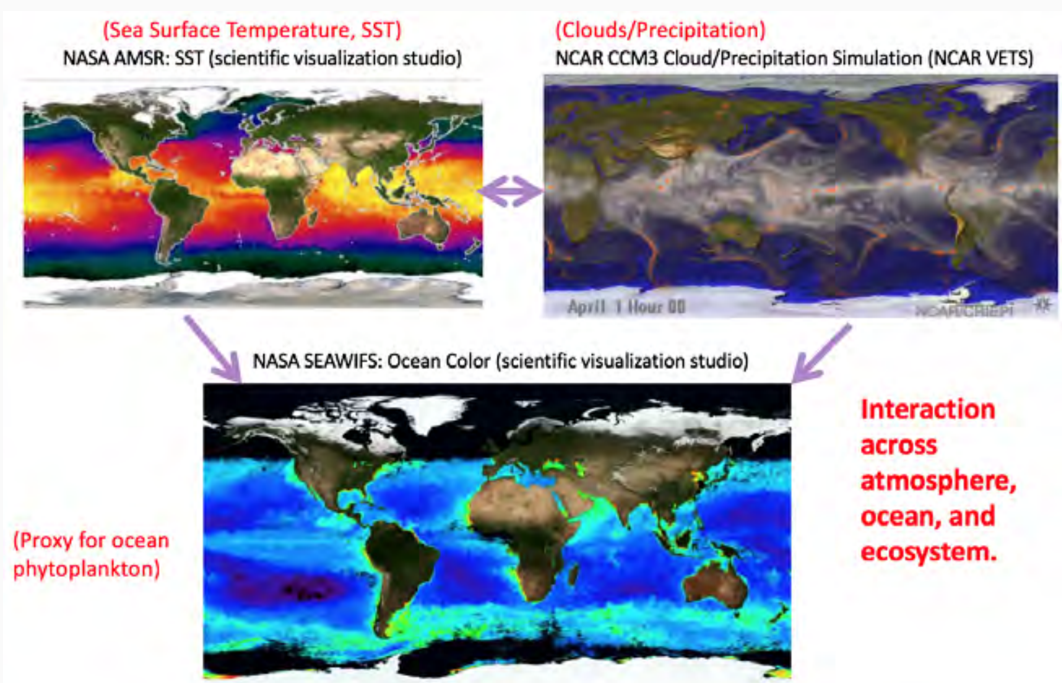


Space AND Time

It is not sufficient to consider just a snapshot of a spatial process at a given time, or time series at a spatial location or an average over spatial locations. Consider atmospheric CO_2:

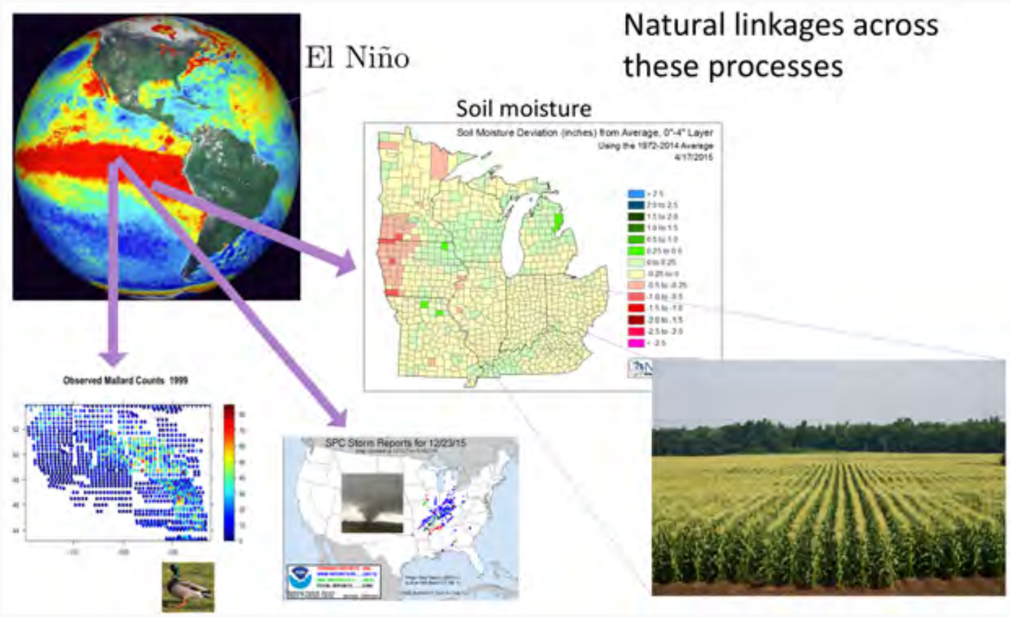
https://www.youtube.com/watch?time_continue=26&v=aogFbPOOFQI

Interactions Across Space and Time



Interaction and Impacts

Interaction and impacts across processes and spatio-temporal scales.



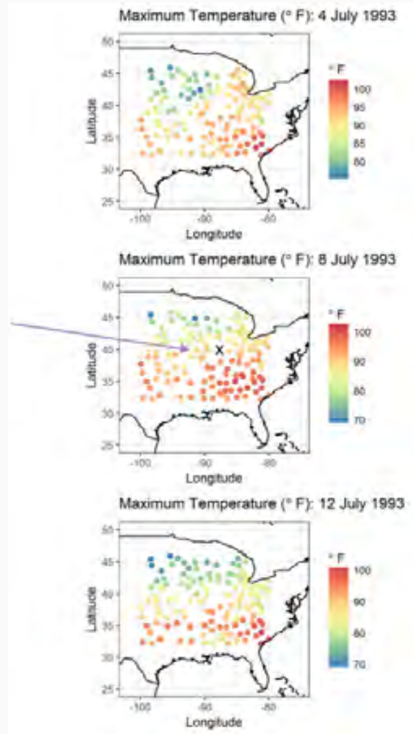
Goals of Spatio-Temporal Analysis

Characterize spatio-temporal processes in the presence of uncertain and (often) incomplete observations and system knowledge, for the purposes of:

- Prediction in space
- Forecasting in time
- Assimilation of observations and mechanistic models
- Accounting for dependence in parameter inference

Example: Prediction at a Spatial Location

What is the optimal prediction (interpolation) to get temperature at the location given by "$\times$"?

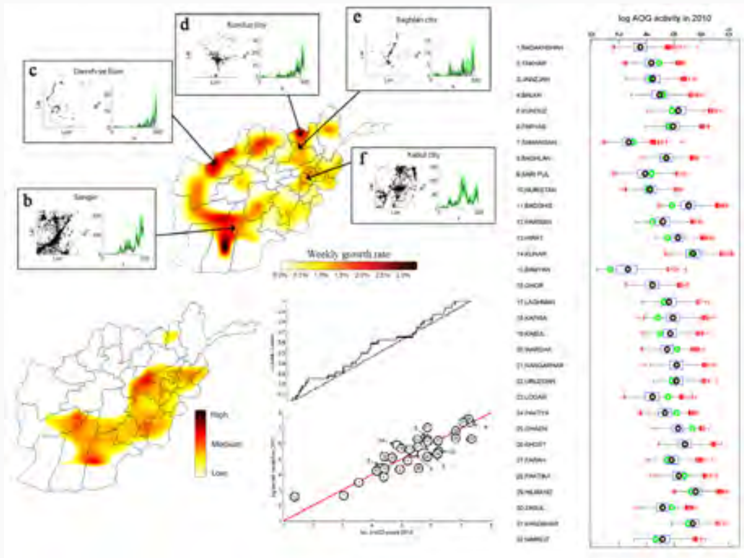


Example: Forecast in Time

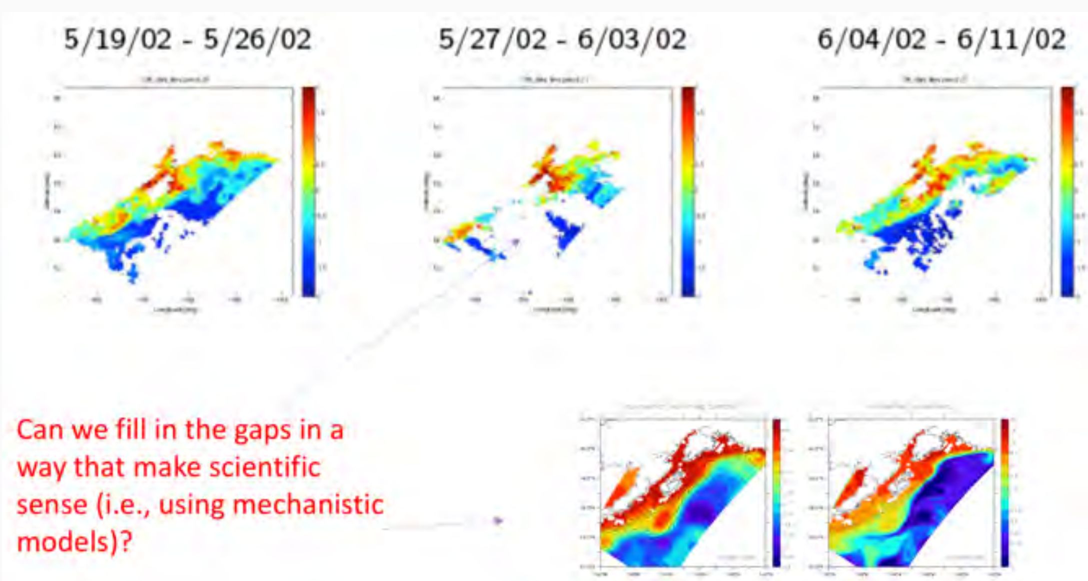
Forecasting the spread of an invasive species

Example: Parameter Estimation

How is armed conflict escalating temporally as a function of space?



Example: Assimilation of Observations and Mechanistic Models



Spatio-Temporal Modeling

Traditionally, there are two approaches for spatio-temporal modeling to address these goals: the "two D's"

Descriptive (Main focus in this course)

Dynamical

Spatio-Temporal Modeling

Descriptive (Marginal) Approach:

Characterize the first- and second-moment behavior of the process

- Convenient "optimal" prediction theory
- Precedence in spatial statistics
- Need only specify mean structure and covariability
- Most useful when knowledge of the process is limited and/or primary interest is with inference on fixed-effects parameters
- Can be difficult computationally in high dimensions
- Difficult to specify realistic spatio-temporal covariances for complex processes

Spatio-Temporal Modeling: Descriptive Approach

Tobler's (1970) "First Law of Geography" (spatio-temporal extension):

> *"everything is related to everything else, but near things [in space and time] are more related than distant things"*

(Except when they aren't! Exceptions "prove"' the rule.)

Spatio-Temporal Modeling

Dynamical (Conditional) Approach:

Current values of the process at a location evolve from past values of the process at various locations

- Need only specify conditional distributions
- Closer to the etiology of the phenomenon under study
- More likely to establish answers to the "why" question (causality) – better for forecasting and prediction in big gaps
- Are best when there is some *a priori* knowledge available concerning process behavior (i.e., mechanistic behavior)

Spatio-Temporal Modeling: Dynamics

- Study of how things change over time
- Pattern of change or growth of a system over time

Consider some examples ...

Dynamics: Examples

Invasive species

Example: Forecast in Time

Armed conflict

Dynamics: Examples

Collective animal movement

Spatio-Temporal Dependence in Statistics

The potential importance of spatial (and spatio-temporal) dependence has long been known in statistics:

> *"After choosing the area we usually have no guidance beyond the verifiable fact that patches in close proximity are commonly more alike, as judged by the yield of the crops, than those which are far apart."* (R.A. Fisher, 1935)

Spatio-Temporal Association and Inference

Consider averages over plots (assume variance = σ^2 for all)

Plots A and B are "closer" in space and time than plots A and C, or B and C. Consider the difference in plot means.

Spatio-Temporal Association and Inference

Consider averages over plots (assume variance = σ^2 for all)

Plots A and B are "closer" in space and time than plots A and C, or B and C. Consider the difference in plot means.

$$\text{var}(A-B) = \text{var}(A) + \text{var}(B) - 2\,\text{cov}(A,B) = 2\,\sigma^2 - 2\,\text{cov}(A,B)$$
$$\text{var}(A-C) = \text{var}(A) + \text{var}(C) - 2\,\text{cov}(A,C) = 2\,\sigma^2 - 2\,\text{cov}(A,C)$$

Spatio-Temporal Association and Inference

Consider averages over plots (assume variance = σ^2 for all)

Plots A and B are "closer" in space and time than plots A and C, or B and C. Consider the difference in plot means.

$$\text{var}(A-B) = \text{var}(A) + \text{var}(B) - 2\,\text{cov}(A,B) = 2\,\sigma^2 - 2\,\text{cov}(A,B)$$
$$\text{var}(A-C) = \text{var}(A) + \text{var}(C) - 2\,\text{cov}(A,C) = 2\,\sigma^2 - 2\,\text{cov}(A,C)$$

Now, if $\text{cov}(A,B) > \text{cov}(A,C)$ (because of 1st law of geography), then $\text{var}(A-B) < \text{var}(A-C)$

Spatio-Temporal Association and Inference

Consider averages over plots (assume variance = σ^2 for all)

Plots A and B are "closer" in space and time than plots A and C, or B and C. Consider the difference in plot means.

$$\text{var}(A-B) = \text{var}(A) + \text{var}(B) - 2\,\text{cov}(A,B) = 2\,\sigma^2 - 2\,\text{cov}(A,B)$$
$$\text{var}(A-C) = \text{var}(A) + \text{var}(C) - 2\,\text{cov}(A,C) = 2\,\sigma^2 - 2\,\text{cov}(A,C)$$

Now, if $\text{cov}(A,B) > \text{cov}(A,C)$ (because of 1st law of geography), then $\text{var}(A-B) < \text{var}(A-C)$

This would affect inference on differences!

Visualizing Spatio-Temporal Data (pp. 24–32 of STSwR)

Spatio-Temporal Data

Types of Spatio-Temporal Data we need to deal with:

- Univariate or Multivariate
- Time
  - regular or irregular intervals
  - continuous or discrete time
  - random events (i.e., point process)
- Space
  - "geostatistical" (continuous space)
  - lattice (finite or countable subset in space)
  - random events (i.e, spatial point process or point pattern)
  - objects (e.g., trajectories)

Data from spatio-temporal processes can be considered as some combination of these temporal and spatial process perspectives.

Visualization of Spatio-Temporal Data

A picture (or a video) is worth a thousand tables of data! Use of maps, color, and animation is a very powerful way to explore data.

There are challenges in visualizing spatio-temporal data due to the fact that multiple dimensions often have to be considered simultaneously (two or three spatial dimensions and time).

We consider a few of the more useful visualizations here.

Visualization of Spatio-Temporal Data: Spatial Plots

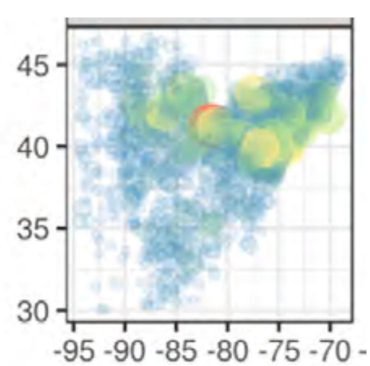
Snapshots of spatial processes for a given time period can be plotted in numerous ways.

- Irregular Data in Space: plot a symbol at data location and vary size and/or color to reflect the observation value
- Lattice Spatial Data
  - Choropleth maps for irregular lattice data
  - Image plots, contour plots, surface plots for regular lattice data

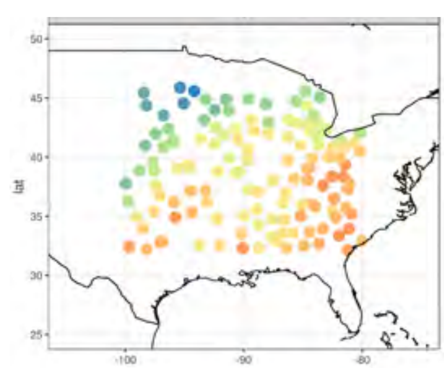
Visualization of Spatio-Temporal Data: Spatial Plots

Irregular Data in Space:

Color and/or circle size correspond to value



Breeding Bird Survey (BBS) House Finch counts in the NE US for 1994.



NOAA maximum daily temperature for 1 May 1993.

Visualization of Spatio-Temporal Data: Spatial Plots

Irregular Lattice Data: Choropleth Map: (from before)

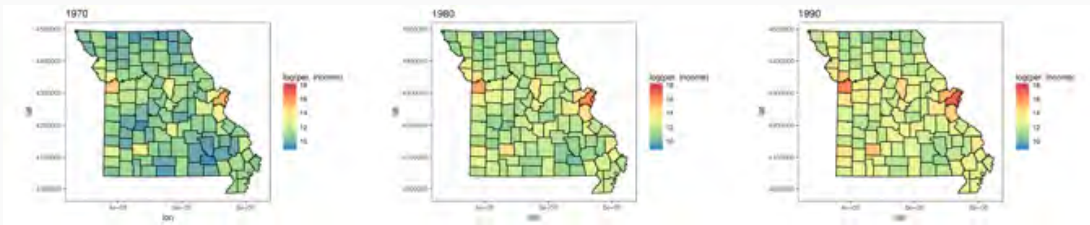
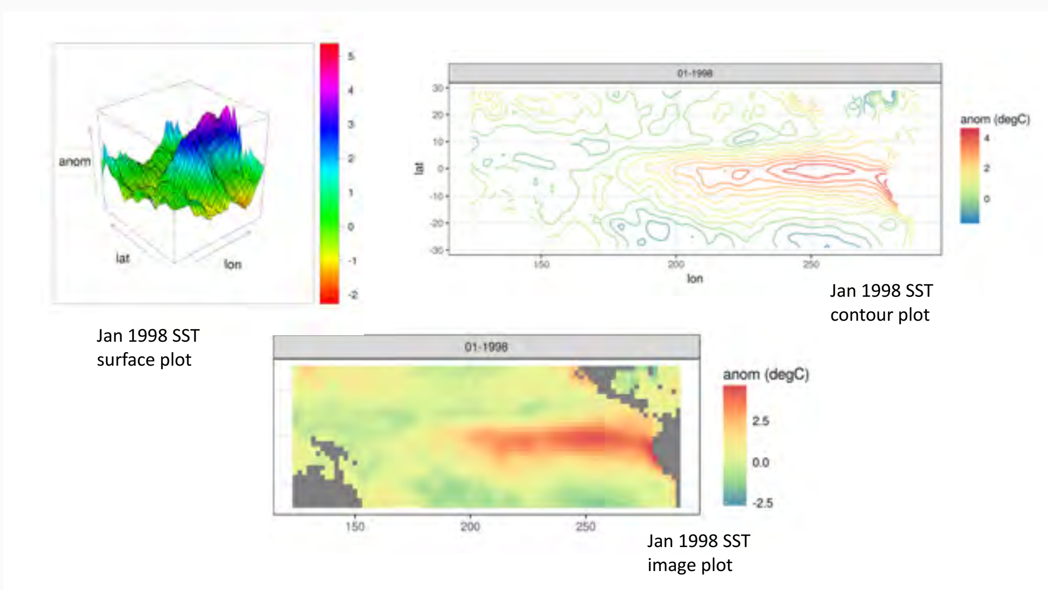


Figure: Personal income for residents in Missouri counties from the US Bureau of Economic Analysis (BEA) for the years 1970, 1980, and 1990.

Visualization of Spatio-Temporal Data: Spatial Plots

Regular Lattice Data: Image, Contour, Surface Maps:



Visualization of Spatio-Temporal Data: Spatial Uncertainty

It can be challenging to visualize both a value and its uncertainty simultaneously on a map. One way is to use a bivariate choropleth map where a color palette for the value of interest is merged with the color palette for the uncertainty.

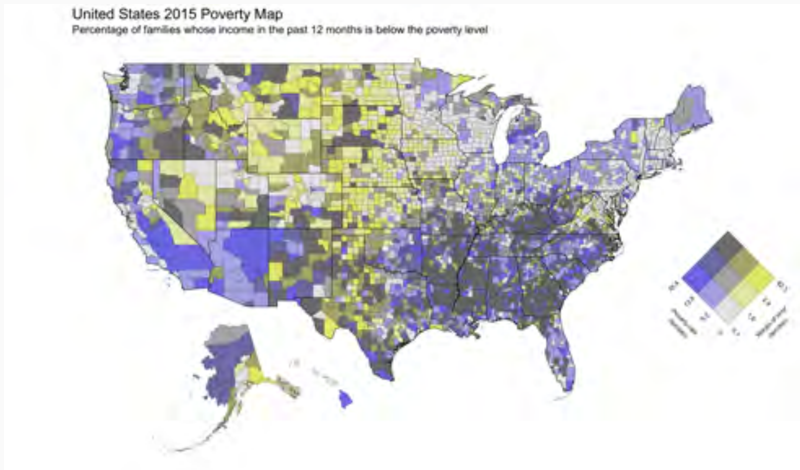


Figure: Bivariate choropleth map of ACS 5-year poverty rate estimates and associated margin of errors for US in 2015 (terciles) (**Vizumap** R package - Github).

Visualization of Spatio-Temporal Data: Spatial Uncertainty

Another way to visualize uncertainty is to "jitter" the color in a choropleth map so that more jitter corresponds to more uncertainty.

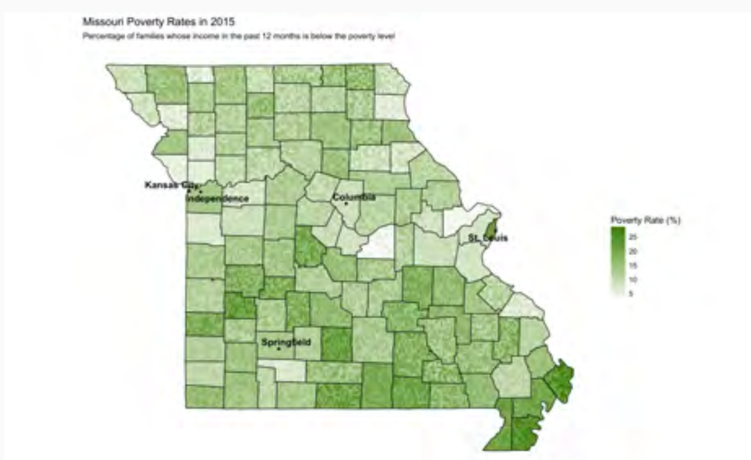


Figure: American Community Survey (ACS) 5-year poverty estimates for Missouri. (**Vizumap** R package - Github).

Visualization of Spatio-Temporal Data: Spatial Uncertainty

It can be particularly effective to animate the uncertainty jitter.

Figure: American Community Survey (ACS) 5-year poverty estimates for Missouri (**Vizumap** R package - Github).

Visualization of Spatio-Temporal Data: Time Series Plots

It can be instructive to plot time series corresponding to an observation location, an aggregation of locations, or multiple locations simultaneously.

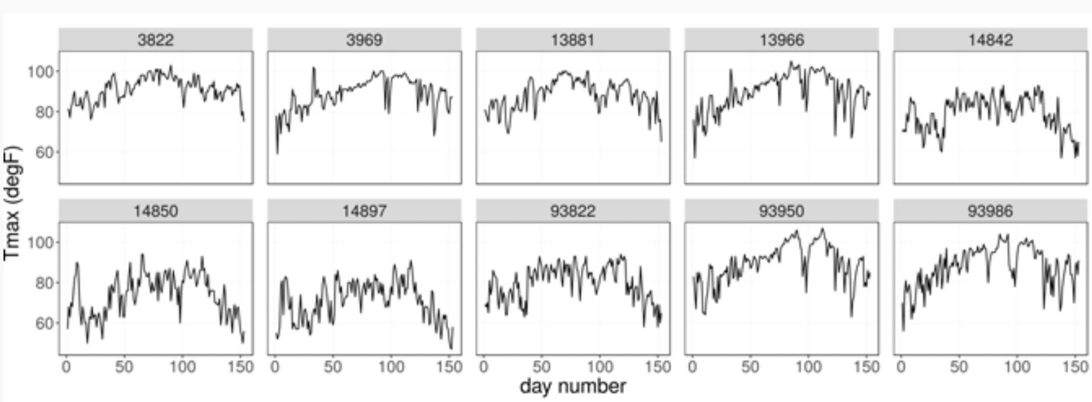


Figure: Maximum temperature for ten stations chosen from the NOAA dataset at random (first day corresponds to 01 May 1993 and last day 30 September 1993).

Visualization of Spatio-Temporal Data: Space and Time

It is often the *interaction* of space and time, or the evolution of spatial field through time that is of primary interest. In this case, it is helpful to try to visualize space and time together. There are several ways we can do this:

- Sequence of spatial maps
- Hovmöller plots
- Animations
- Interactive plots

Visualization of Spatio-Temporal Data: Map Sequences

Sequence of spatial (image) maps:

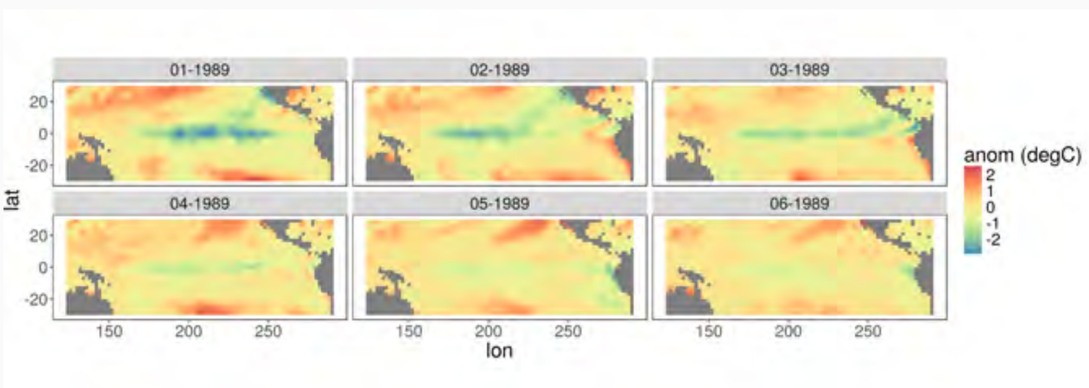


Figure: Sea-surface temperature (SST) anomalies for the months of January – June in 1989.

Visualization of Spatio-Temporal Data: Map Sequences

Sequence of spatial (point) maps:

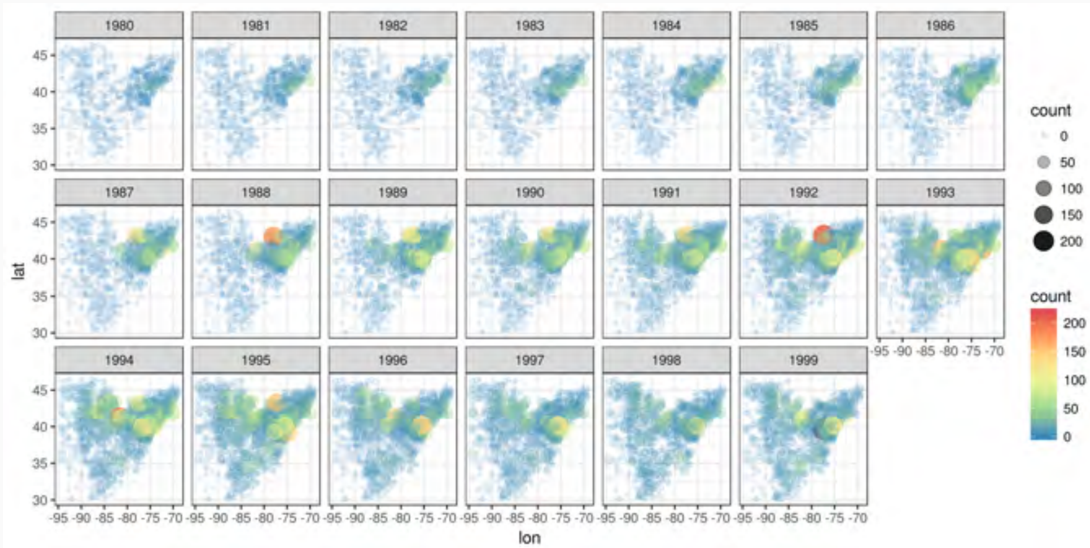


Figure: BBS House Finch counts between 1980-1999.

Visualization of Spatio-Temporal Data: Hovmöller Plots

A 2-D plot where the x-dimension represents 1-D space and the y-dimension represents time (increasing from top to bottom).

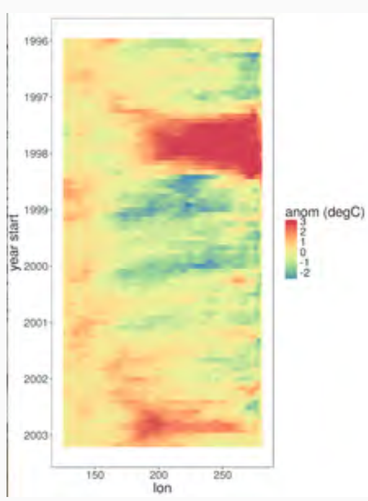


Figure: Hovmöller plots for SST anomalies for time vs. longitude averaged between 1N-1S.

Visualization of Spatio-Temporal Data: Animations

One of the most useful visualization tools for spatio-temporal data with complex interactions in two or three dimensions is through an animation. These spatio-temporal data are column-averaged carbon dioxide data from NASA's OCO-2 satellite.

https://www.youtube.com/watch?v=aogFbPOOFQI

Visualization of Spatio-Temporal Data: Animations

Figure: NASA AMSR Scientific Visualization Studio animation of global sea-surface temperature.

Visualization of Spatio-Temporal Data: Interactive

Programming tools for interactive visualization are becoming increasingly accessible. Such a "data-immersive" experience allows one to explore the data without "scripting" (e.g., "hovering" a cursor over a figure with a "linked brush"). The **plotly** and **ggvis** R packages are quite useful for this (see the following screen shot).

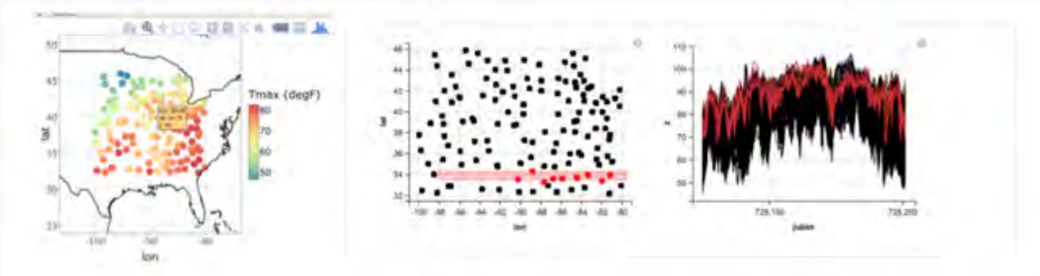


Figure: Left panel: Interactive exploration of NOAA maximum temperatures on 1 May 1993 using **ggplotly** from the package **plotly**. Right Panels: Linked brush exploration of time series corresponding to user-chosen set of spatial locations with the package **ggvis**.

Exploratory Analysis of Spatio-Temporal Data: Quantitative Summaries (pp. 32–47 of STSwR)

Exploration of Spatio-Temporal Data: Empirical Spatial Means

The empirical spatial mean for location $\mathbf{s}_i$ is given by:

$$\widehat{\mu}_z(\mathbf{s}_i) \equiv \frac{1}{T}\sum_{j=1}^{T} z(\mathbf{s}_i; t_j).$$

If we consider the means for all spatial data locations, we can write it more compactly as an m-dimensional vector:

$$\widehat{\boldsymbol{\mu}}_z \equiv \begin{bmatrix} \widehat{\mu}_z(\mathbf{s}_1) \\ \vdots \\ \widehat{\mu}_z(\mathbf{s}_m) \end{bmatrix} \equiv \frac{1}{T}\sum_{j=1}^{T} \mathbf{z}_{t_j},$$

where $\mathbf{z}_{t_j} \equiv (z(\mathbf{s}_1; t_j), \ldots, z(\mathbf{s}_m; t_j))'$. This mean vector is indexed in space and can be visualized on a map or as functions of latitude or longitude. [Note, it's OK if we don't have the same number of temporal observations at each location.]

Exploration of Spatio-Temporal Data: Empirical Spatial Means

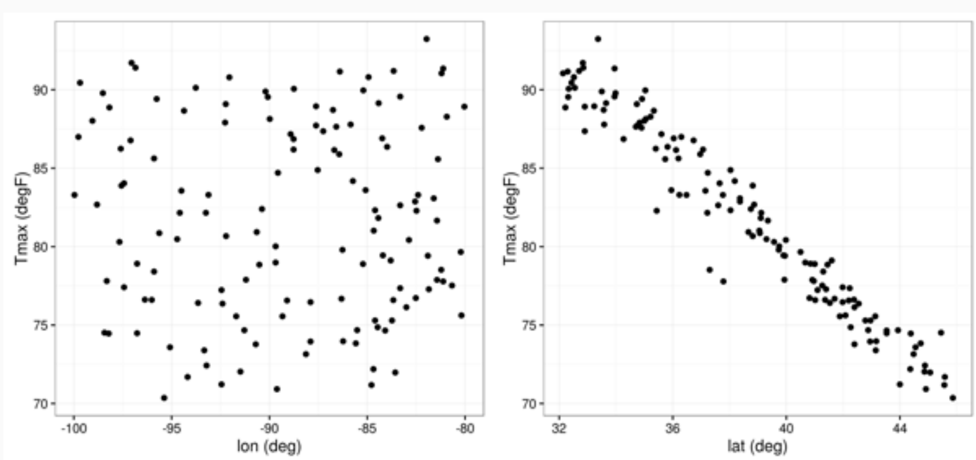


Figure: Spatial mean of the NOAA maximum temperature as a function of station longitude (left panel) and station latitude (right panel).

Exploration of Spatio-Temporal Data: Empirical Temporal Means

We can also average across space and plot the associated time series. The empirical temporal mean is given by:

$$\widehat{\mu}_z(t_j) \equiv \frac{1}{m} \sum_{i=1}^{m} z(\mathbf{s}_i; t_j).$$

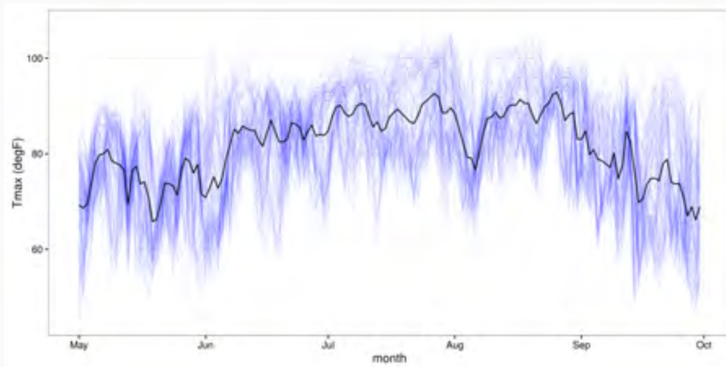


Figure: Raw data from the NOAA maximum temperature dataset (blue lines) and the empirical means $\widehat{\mu}_z(t)$ (black line).

Exploration of Spatio-Temporal Data: Empirical Spatial Covariance

We are often interested in the covariability between different locations (averaged over time). The empirical lag-τ covariance between spatial locations $\mathbf{s}_i$ and $\mathbf{s}_k$ is given by:

$$\widehat{c}_z^{(\tau)}(\mathbf{s}_i, \mathbf{s}_k) \equiv \frac{1}{T-\tau} \sum_{j=\tau+1}^{T} (z(\mathbf{s}_i; t_j) - \widehat{\mu}_z(\mathbf{s}_i))(z(\mathbf{s}_k; t_j - \tau) - \widehat{\mu}_z(\mathbf{s}_k)),$$

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

We can collect these empirical covariances and construct an $m \times m$ lag-τ covariance *matrix*, $\widehat{\mathbf{C}}_z^{(\tau)}$ or we can calculate it directly by

$$\widehat{\mathbf{C}}_z^{(\tau)} \equiv \frac{1}{T-\tau} \sum_{j=\tau+1}^{T} (\mathbf{z}_{t_j} - \widehat{\boldsymbol{\mu}}_z)(\mathbf{z}_{t_{j-\tau}} - \widehat{\boldsymbol{\mu}}_z)'; \quad \tau = 0, 1, \ldots, T-1.$$

Exploration of Spatio-Temporal Data: Empirical Spatial Covariance

In general, it can be difficult to get direct intuition from these covariance estimates since locations in a two-dimensional space don't have any specific ordering.

Sometimes it is helpful to plot the covariances for specific subsets of the data. For example, we might consider a specific longitude and consider the lag-τ spatial covariances for the locations as a function of latitude (for that longitude).

Exploration of Spatio-Temporal Data: Empirical Spatial Covariance

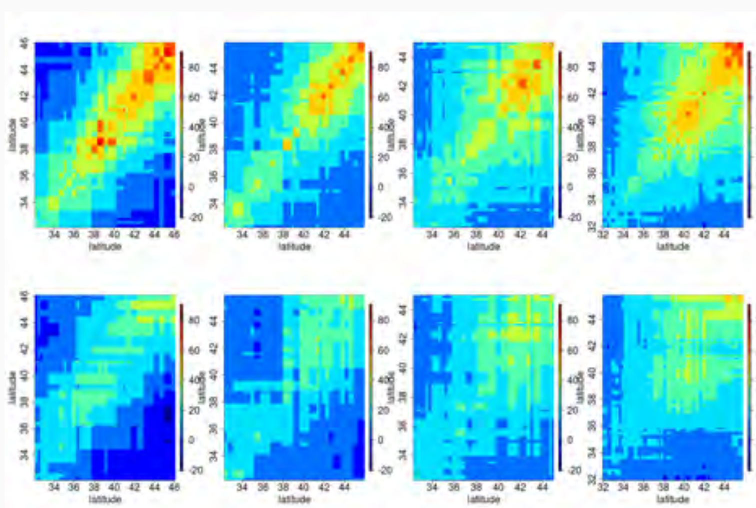


Figure: NOAA dataset lag-0 (top panels) and lag-1 (bottom panels) empirical spatial covariance image plots for four regular longitudinal strips in which the domain of interest is subdivided.

Exploration of Spatio-Temporal Data: Empirical Spatial Cross-Covariance

We can also calculate the empirical lag-τ cross-covariance matrix between two spatio-temporal datasets, $\{\mathbf{z}_{t_j}\}$ and $\{\mathbf{x}_{t_j}\}$, where $\{\mathbf{x}_{t_j}\}$ corresponds to data vectors at n spatial locations (potentially different from $\{\mathbf{z}_{t_j}\}$, but it is assumed that they are observed at the same times). The associated $m \times n$ lag-τ cross-covariance matrix is calculated by

$$\widehat{\mathbf{C}}_{z,x}^{(\tau)} \equiv \frac{1}{T-\tau} \sum_{j=\tau+1}^{T} (\mathbf{z}_{t_j} - \widehat{\boldsymbol{\mu}}_z)(\mathbf{x}_{t_j-\tau} - \widehat{\boldsymbol{\mu}}_x)',$$

for $\tau = 0, 1, \ldots, T-1$, where $\widehat{\boldsymbol{\mu}}_x$ is the empirical spatial mean vector for the $\{\mathbf{x}_{t_j}\}$ data.

Exploration of Spatio-Temporal Data: Empirical Spatial Cross-Covariance

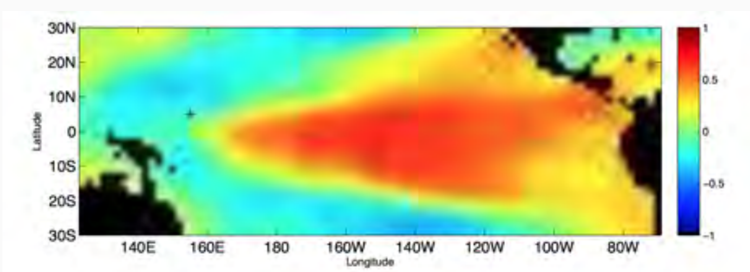


Figure: Empirical lag-6 cross-covariance between SST anomalies (the z-variable) and the east-west component of the wind at $(155°E, 5°N)$ lagged 6 months (the x-variable).

Exploration of Spatio-Temporal Data: Spatio-Temporal Covariograms

In the next part of the course we will see that it is necessary to characterize the joint spatio-temporal dependence structure of a spatio-temporal process in order to perform optimal prediction. We consider empirical spatio-temporal *covariograms*.

The biggest difference between what we are doing here and the covariance estimates we just considered, is that we are interested in characterizing the covariability in the spatio-temporal data as a function of specific lags in time *and* in space.

Note that the lag in time is a scalar, but the lag in space is a vector (corresponding to the displacement between locations in d-dimensional space).

Exploration of Spatio-Temporal Data: Empirical Spatio-Temporal Covariogram

Assuming that the first moment (mean) depends on space but not time, and that the second moment (covariance) depends only on the lag differences in space and time. The empirical *spatio-temporal covariogram* for spatial lag $\mathbf{h}$ and time lag τ is then given by

$$\widehat{C}_z(\mathbf{h};\tau) = \frac{1}{|N_{\mathbf{s}}(\mathbf{h})|}\frac{1}{|N_t(\tau)|}\sum_{\mathbf{s}_i,\mathbf{s}_k \in N_{\mathbf{s}}(\mathbf{h})}\ \sum_{t_j,t_\ell \in N_t(\tau)} (z(\mathbf{s}_i;t_j)-\widehat{\mu}_z(\mathbf{s}_i))(z(\mathbf{s}_k;t_\ell)-\widehat{\mu}_z(\mathbf{s}_k)),$$

where $N_{\mathbf{s}}(\mathbf{h})$ refers to the pairs of spatial locations with spatial lag within some tolerance of $\mathbf{h}$, $N_t(\tau)$ refers to the pairs of time points with time lag within some tolerance of τ, and $|N(\cdot)|$ refers to the number of elements in $N(\cdot)$.

Exploration of Spatio-Temporal Data: Empirical Spatio-Temporal Covariogram

We can further refine the neighborhoods in the empirical spatio-temporal covariogram so that the lags, $\mathbf{h}$, are directionally dependent (suggesting so-called *anisotropic* spatio-temporal dependence).

Alternatively, under *isotropy*, we might consider the lag only as a function of distance, $h = \|\mathbf{h}\|$, where $\|\cdot\|$ represents the Euclidean norm.

After calculating the empirical covariogram for various time and spatial lags, we can construct lag-τ empirical covariance matrices, $\widehat{\mathbf{C}}^{(\tau)}$ for $\tau = 0, 1, 2, \ldots$ similar to above.

Exploration of Spatio-Temporal Data: Empirical Spatio-Temporal Covariogram

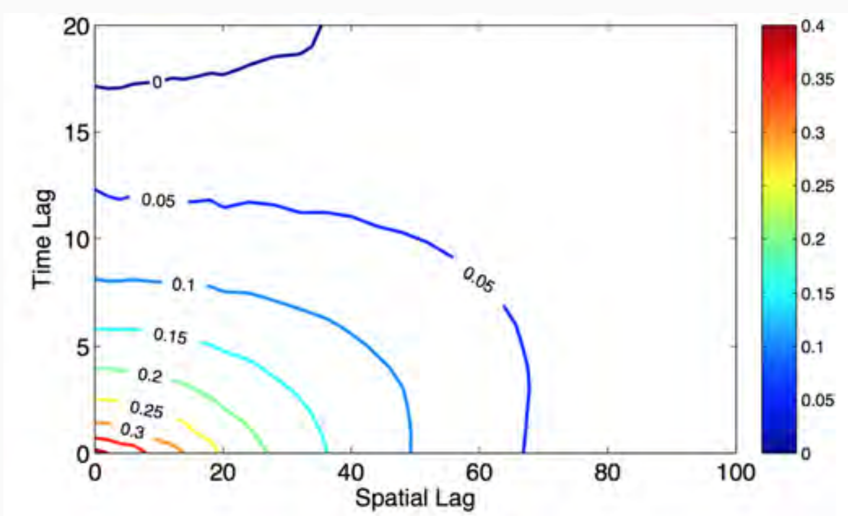


Figure: Empirical spatio-temporal covariogram for SST data; spatial lag (deg), time lag (months).

Exploration of Spatio-Temporal Data: Empirical Spatio-Temporal Semivariogram

In some spatio-temporal prediction kriging applications, you might see a spatio-temporal semivariogram given by:

$$\gamma(\mathbf{h};\tau) \equiv \frac{1}{2}\mathrm{var}(Z(\mathbf{s}+\mathbf{h};t+\tau) - Z(\mathbf{s};t)).$$

In the case where we assume that the mean depends on space but not time, and that the second moment (covariance) depends only on the lag differences in space and time, we can write the semivariogram as: $\gamma(\|\mathbf{h}\|;\tau) = C(\mathbf{0};0) - C(\|\mathbf{h}\|;\tau)$.

Thus, it is simple to go back and forth between the empirical semivariogram and the empirical covariogram in this case:

$$\widehat{\gamma}(\|\mathbf{h}\|;\tau) = \widehat{C}(\mathbf{0};0) - \widehat{C}(\|\mathbf{h}\|;\tau).$$

Exploration of Spatio-Temporal Data: Empirical Spatio-Temporal Semivariogram

Example (spatial):

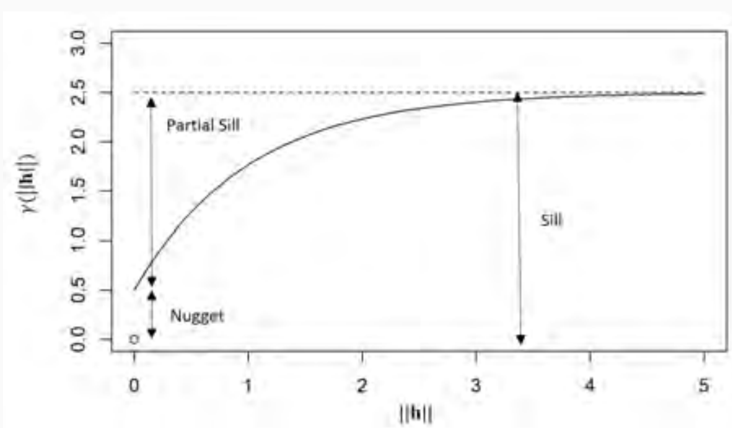


Figure: Basic spatial semivariogram as a function of spatial distance $\|\mathbf{h}\|$.

Exploration of Spatio-Temporal Data: Empirical Spatio-Temporal Semivariogram

More generally, note:

$$\gamma_z(\mathbf{h};\tau) = \frac{1}{2} E\left(Z(\mathbf{s}+\mathbf{h}; t+\tau) - Z(\mathbf{s}; t)\right)^2,$$

and hence an alternative empirical semivariogram estimate is

$$\widehat{\gamma}_z(\mathbf{h};\tau) = \frac{1}{|N_{\mathbf{s}}(\mathbf{h})|}\frac{1}{|N_t(\tau)|}\sum_{\mathbf{s}_i,\mathbf{s}_k \in N_{\mathbf{s}}(\mathbf{h})}\ \sum_{t_j,t_\ell \in N_t(\tau)} (Z(\mathbf{s}_i;t_j) - Z(\mathbf{s}_k;t_\ell))^2,$$

Note that this calculation does not need any information about the spatial means.

Exploration of Spatio-Temporal Data: Empirical Spatio-Temporal Semivariogram

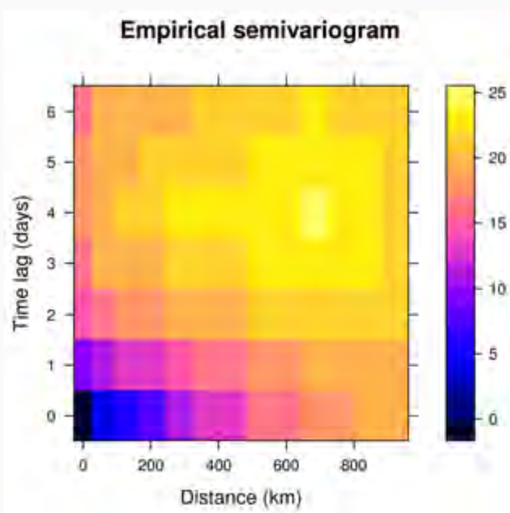


Figure: Empirical spatio-temporal semivariogram of daily max temperature from the NOAA data set during July 2003, computed using the function `variogram` in **gstat**.

Exploration of Spatio-Temporal Data: Some Other Approaches

Some other approaches for exploring spatio-temporal data (not covered here; see Cressie and Wikle, 2011; STSwR, 2019):

- Spectral and Cross-Spectral Analysis
- Empirical Orthogonal Function (EOF) Analysis (spatio-temporal PCA)
- Spatio-temporal CCA
- Principal Oscillation Pattern (POP) Analysis
- Singular Spectrum Analysis (SSA)
- etc.

Visualization and Exploration of Spatio-Temporal Data: R Examples

Chapter Labs in STSwR:

- Lab 2.1: Data Wrangling
- Lab 2.2: Visualization
- Lab 2.3: Exploratory Data Analysis

Labs in the free pdf of STSwR from:
`https://spacetimewithr.org`

Lab R Code available at: `https://spacetimewithr.org/code`

Descriptive Spatio-Temporal Modeling: Part I (Covariance-Based Approach) (pp. 138–151 of STSwR)

Spatio-Temporal Prediction

As an example, consider prediction ("interpolation") of maximum temperature at location "x" in the central US given data from 138 measurement locations.

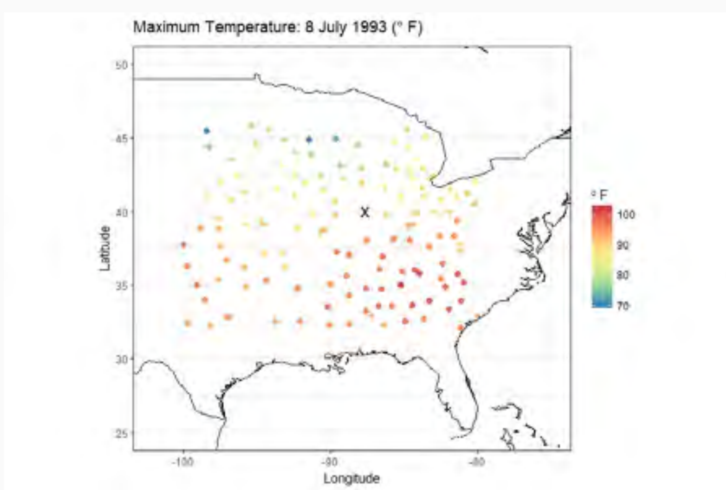


Figure: NOAA maximum temperature observations for 8 July 1993.

Spatio-Temporal Prediction

But, we have data from other times as well; Tobler's "law" suggests that we should consider "nearby" observations in both space AND time to help with the prediction (interpolation).

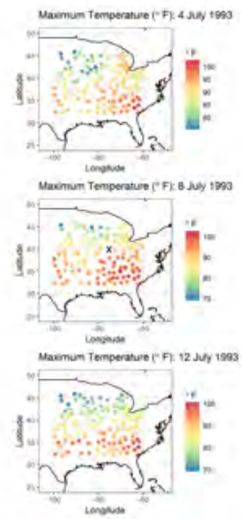


Figure: NOAA maximum temperature observations for 4, 8, and 12 July 1993.

General Statistical Model Framework

We consider models of the form:

$$\begin{aligned} \textit{observations} &= \textit{latent process} + \textit{observation error} \\ \textit{latent process} &= \textit{trend term} + \textit{dependent random process} \end{aligned}$$

Here we focus on the more traditional "descriptive" approach that considers the dependent random process in terms of the first-order and second-order moments (means, variances, and covariances) of its marginal distribution.

Spatio-Temporal Prediction: Notation

$\{Y(\mathbf{s}; t) : \mathbf{s} \in D_s, t \in D_t\}$ denotes a latent (unobserved) spatio-temporal random process with spatial location $\mathbf{s}$ in spatial domain D_s (a subset of d-dimensional real space), and time index t in temporal domain D_t (along the one-dimensional real line).

Data: $\{z(\mathbf{s}_{ij}; t_j)\}$ for spatial locations $\{\mathbf{s}_{ij} : i = 1, \ldots, m_j\}$ and times $\{t_j : j = 1, \ldots, T\}$.

It is convenient to consider all observations at a given time (say, t_j) using the vector:

$$\mathbf{z}_{t_j} = (z(\mathbf{s}_{1j}; t_j), \ldots, z(\mathbf{s}_{m_j j}; t_j))'.$$

Spatio-Temporal Prediction: Data Model

Now, we represent the data in terms of the latent spatio-temporal process of interest plus an error term:

$$z(\mathbf{s}_{ij}; t_j) = Y(\mathbf{s}_{ij}; t_j) + \epsilon(\mathbf{s}_{ij}; t_j); \quad i = 1, \ldots, m_j, \ j = 1, \ldots, T,$$

Assumptions:

- $\{\epsilon(\mathbf{s}_{ij}; t_j)\}$ - independent and identically distributed mean-zero observation errors that are independent of Y and have variance σ_ϵ^2
- the errors are assumed to have a normal (Gaussian) distribution here: $\epsilon(\mathbf{s}_{ij}; t_j) \sim iid\ N(0, \sigma_\epsilon^2)$ (we might also write this as $\epsilon(\mathbf{s}_{ij}; t_j) \sim iid\ Gau(0, \sigma_\epsilon^2)$)

Spatio-Temporal Prediction: Goal

We seek to predict the "true" (latent) process at some location $(\mathbf{s}_0; t_0)$ (for which we might not have data), $Y(\mathbf{s}_0; t_0)$, given all of the observations $\{z(\mathbf{s}_{ij}; t_j) : i = 1, \ldots, m_j; j = 1, \ldots, T\}$.

Consider the linear predictor of the form:

$$\widehat{Y}(\mathbf{s}_0; t_0) = \sum_{j=1}^{T} \sum_{i=1}^{m_j} w_{ij}\, z(\mathbf{s}_{ij}; t_j).$$

We need to choose these weights. How? Instead of *ad hoc* approaches (e.g., inverse distance weighting), we want to find the "optimal" weights. This requires that we consider spatio-temporal dependence more directly.

Spatio-Temporal Prediction: Cartoon

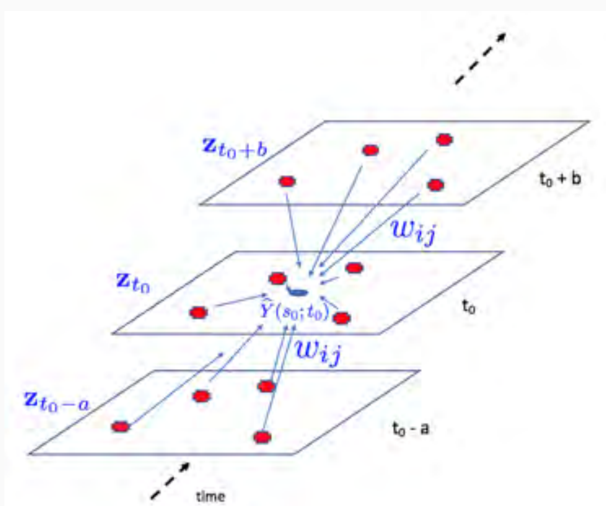


Figure: Cartoon showing spatio-temporal prediction of latent process $Y(\mathbf{s}_0; t_0)$ given observations at time t_0 and before and after t_0. The solid arrows depict the prediction weights w_{ij} for the various observations.

Spatio-Temporal Prediction: Latent Process Model

We assume that the latent process follows the model

$$Y(\mathbf{s}; t) = \mu(\mathbf{s}; t) + \eta(\mathbf{s}; t),$$

for all $(\mathbf{s}; t)$ in our space-time domain of interest (e.g., $D_s \times D_t$).

- $\mu(\mathbf{s}; t)$ represents the process mean: We may choose to let $\mu(\mathbf{s}; t)$ be: (i) known, (ii) constant but unknown, or (iii) modeled in terms of p covariates, $\mu(\mathbf{s}; t) = \mathbf{x}(\mathbf{s}; t)'\boldsymbol{\beta}$, where the p-dimensional vector $\boldsymbol{\beta}$ is unknown, which corresponds to (i) simple, (ii) ordinary, and (iii) universal *spatio-temporal kriging*, respectively
- $\eta(\mathbf{s}; t)$ represents a zero-mean Gaussian process (we will define this next) with spatial and temporal dependence

Spatio-Temporal Prediction: Gaussian Processes

What is the difference between a Gaussian “distribution” and a Gaussian “process”?

- **Gaussian Distribution**: a distribution over *vectors*
  - Fully specified by a mean vector and covariance matrix: e.g., $\mathbf{Y} \sim Gau(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (i.e., a multivariate normal distribution)
  - In spatio-temporal applications, the elements of $\mathbf{Y}$ are indexed by space-time location, for example, $Y_i = Y(\mathbf{s}_i; t_i)$
- **Gaussian Process**: a distribution over *functions*
  - Fully specified by a mean *function* (m) and covariance *function* (c): e.g., $Y \sim GP(m, c)$; thus, we will need to specify a mean function and covariance function to define a GP
  - In spatio-temporal applications, the *arguments* of the functions are the spatio-temporal locations: $(\mathbf{s}; t)$

Spatio-Temporal Prediction: Gaussian Processes

Why GPs?

- A GP is an infinite-dimensional object
  - This will allow us to make a prediction *anywhere* in our space-time domain (if we know the mean and covariance functions)
- But, in practice we will only ever need to consider finite-dimensional distributions; why?
  - We only have a finite set of data and are eventually only interested in a finite set of prediction locations
  - Any finite collection of GP random variables has a joint Gaussian *distribution*
  - This allows us to use the traditional machinery of multivariate normal distributions and linear mixed models to do calculations

Spatio-Temporal Prediction: Gaussian Processes

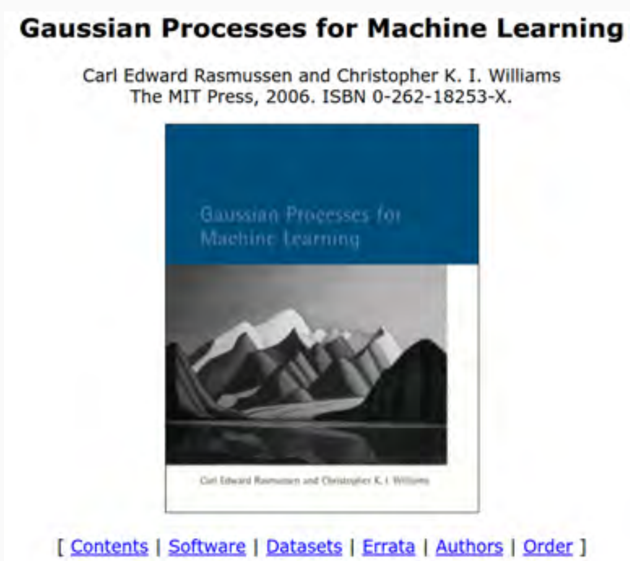


Figure: One of the most accessible books on GPs. Freely available from http://www.gaussianprocess.org/gpml/.

Spatio-Temporal Prediction: Gaussian Distributions

Technical Review:

Recall the famous result relating joint and conditional Gaussian random variables:

$$\begin{pmatrix} \mathbf{v} \\ \mathbf{v}_* \end{pmatrix} \sim Gau\left(\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}_* \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}_* \\ \boldsymbol{\Sigma}_*' & \boldsymbol{\Sigma}_{**} \end{pmatrix}\right)$$

where $E(\mathbf{v}) = \boldsymbol{\mu}$, $E(\mathbf{v}_*) = \boldsymbol{\mu}_*$, $\boldsymbol{\Sigma} = \mathrm{cov}(\mathbf{v}, \mathbf{v})$, $\boldsymbol{\Sigma}_* = \mathrm{cov}(\mathbf{v}, \mathbf{v}_*)$, and $\boldsymbol{\Sigma}_{**} = \mathrm{cov}(\mathbf{v}_*, \mathbf{v}_*)$.

Then,

$$\mathbf{v}_* | \mathbf{v} \sim Gau(\boldsymbol{\mu}_* + \boldsymbol{\Sigma}_*' \boldsymbol{\Sigma}^{-1}(\mathbf{v} - \boldsymbol{\mu}), \boldsymbol{\Sigma}_{**} - \boldsymbol{\Sigma}_*' \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_*)$$

Spatio-Temporal Prediction: Optimal Linear Prediction

Let's get back to the problem at hand:

Recall we are interested in predicting $Y(\mathbf{s}_0; t_0)$ (i.e., at some arbitrary location in space-time) given the data $\mathbf{z} \equiv (z(\mathbf{s}_{11}; t_1), \ldots, z(\mathbf{s}_{m_T T}; t_T))'$. Assume for simplicity that the mean function $\mu(\mathbf{s}; t)$ is known. We are looking for linear predictors of the form:

$$\widehat{Y}(\mathbf{s}_0; t_0) = k(\mathbf{s}_0; t_0) + \sum_{j=1}^{T} \sum_{i=1}^{m_j} w_{ij} \, z(\mathbf{s}_{ij}, t_j) = k(\mathbf{s}_0; t_0) + \mathbf{w}_0' \mathbf{z}.$$

The optimal linear predictor in this Gaussian case is also the conditional expectation $\mathrm{E}[Y(\mathbf{s}_0; t_0) \mid \mathbf{z}\,]$. Thus, we seek the conditional distribution $[Y(\mathbf{s}_0; t_0) \mid \mathbf{z}\,]$. Why not $[Z(\mathbf{s}_0; t_0) \mid \mathbf{z}\,]$?

Spatio-Temporal Prediction: Optimal Linear Prediction

It helps to rewrite the data model in vector form:

$$\mathbf{z} = \mathbf{Y} + \varepsilon, \quad \varepsilon \sim Gau(\mathbf{0}, \mathbf{C}_\epsilon),$$

where $\mathbf{Y}$ and ε are ordered in the same way as the data vector $\mathbf{z}$.

Now, $\mathrm{cov}(\mathbf{Y},\mathbf{Y}) \equiv \mathbf{C}_y$, $\mathrm{cov}(\varepsilon,\varepsilon) \equiv \mathbf{C}_\epsilon$, $\mathrm{cov}(\mathbf{z},\mathbf{z}) = \mathbf{C}_y + \mathbf{C}_\epsilon$.

In addition, $\mathbf{c}_0' \equiv \mathrm{cov}(Y(\mathbf{s}_0;t_0),\mathbf{z})$, $c_{0,0} \equiv var(Y(\mathbf{s}_0;t_0))$, and $\mathbf{X}$ is the $(\sum_{j=1}^T m_j) \times p$ matrix of known covariates. Then, consider the joint Gaussian distribution:

$$\begin{bmatrix} Y(\mathbf{s}_0;t_0) \\ \mathbf{z} \end{bmatrix} \sim Gau\left(\begin{bmatrix} \mathbf{x}(\mathbf{s}_0;t_0)' \\ \mathbf{X} \end{bmatrix} \boldsymbol{\beta}, \begin{bmatrix} c_{0,0} & \mathbf{c}_0' \\ \mathbf{c}_0 & \mathbf{C}_z \end{bmatrix}\right).$$

Spatio-Temporal Prediction: Optimal Linear Prediction

Using the joint/conditional Gaussian distribution formula from before (and, assuming for the moment that $\boldsymbol{\beta}$ is known, i.e., *simple spatio-temporal kriging*), we get

$$Y(\mathbf{s}_0;t_0)\,|\mathbf{z} \sim Gau\left(\mathbf{x}(\mathbf{s}_0;t_0)'\boldsymbol{\beta} + \mathbf{c}_0'\mathbf{C}_z^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})\,,\; c_{0,0} - \mathbf{c}_0'\mathbf{C}_z^{-1}\mathbf{c}_0\right).$$

Spatio-Temporal Prediction: Optimal Linear Prediction

The prediction formula is given by the mean of this distribution:

$$\widehat{Y}(\mathbf{s}_0; t_0) = \mathbf{x}(\mathbf{s}_0; t_0)'\boldsymbol{\beta} + \mathbf{c}_0'\mathbf{C}_z^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta}),$$

and the prediction variance is then $\sigma^2_{Y_0} = c_{0,0} - \mathbf{c}_0'\mathbf{C}_z^{-1}\mathbf{c}_0$.

This predictor is the *optimal predictor* in that it minimizes the mean squared prediction error:

$$MSPE = E[(Y(\mathbf{s}_0; t_0) - \widehat{Y}(\mathbf{s}_0; t_0))^2].$$

Thus, in this case, the weights are given by $\mathbf{w}_0 = \mathbf{c}_0'\mathbf{C}_z^{-1}$ and $k(\mathbf{s}_0; t_0) = \mathbf{x}(\mathbf{s}_0; t_0)'\boldsymbol{\beta} - \mathbf{c}_0'\mathbf{C}_z^{-1}\mathbf{X}\boldsymbol{\beta}$.

Spatio-Temporal Prediction: Optimal Linear Prediction

Let's take a closer look at the predictive distribution:

$$Y(\mathbf{s}_0; t_0)\,|\mathbf{z} \sim Gau\left(\mathbf{x}_0'\boldsymbol{\beta} + \mathbf{c}_0'\mathbf{C}_z^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})\ ,\ c_{0,0} - \mathbf{c}_0'\mathbf{C}_z^{-1}\mathbf{c}_0\right)$$

"prior" mean without observing **z**

Weights: only a function of GP covariances and observation error (co)variances

"residuals" between observations and their (known) means

"prior" variance is reduced given the information from the data; how much it is reduced is related to the GP covariances and measurement process (co)variances

(Note, it is simple to predict at many new locations, say $\mathbf{Y}_0$. The formulas are slightly modified in terms of matrices for $\mathbf{X}_0$, $\mathbf{C}_0$, and $\mathbf{c}_{0,0}$.)

Spatio-Temporal Prediction: Optimal Prediction Examples

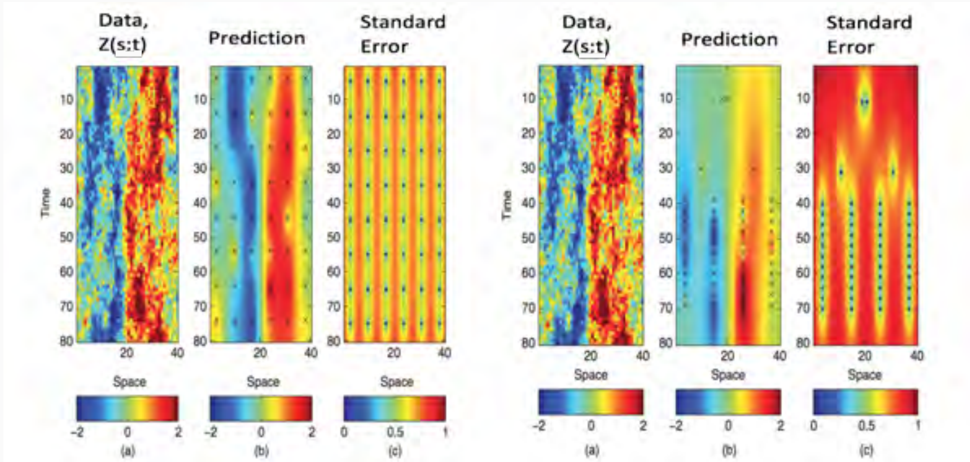


Figure: Simulated data in 1-D space and time, with predictions and standard errors. Observation locations are given by "x". Note, prediction errors are much less when data are "nearby" in space and time.

Spatio-Temporal Prediction: Optimal Prediction

In most real-world problems, one does not know $\boldsymbol{\beta}$.

In this case, the optimal *universal spatio-temporal kriging predictor* of $Y(\mathbf{s}_0; t_0)$ is given by:

$$\hat{Y}(\mathbf{s}_0; t_0) = \mathbf{x}(\mathbf{s}_0; t_0)'\hat{\boldsymbol{\beta}}_{gls} + \mathbf{c}_0'\mathbf{C}_z^{-1}(\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}_{gls}),$$

where the generalized-least-squares (gls) estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_{gls} \equiv (\mathbf{X}'\mathbf{C}_z^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}_z^{-1}\mathbf{z}.$$

Spatio-Temporal Prediction: Optimal Prediction

The associated universal spatio-temporal kriging variance is given by

$$\sigma^2_{Y,uk}(\mathbf{s}_0; t_0) = c_{0,0} - \mathbf{c}_0' \mathbf{C}_z^{-1} \mathbf{c}_0 + \kappa,$$

where

$$\kappa \equiv (\mathbf{x}(\mathbf{s}_0; t_0) - \mathbf{X}' \mathbf{C}_z^{-1} \mathbf{c}_0)' (\mathbf{X}' \mathbf{C}_z^{-1} \mathbf{X})^{-1} (\mathbf{x}(\mathbf{s}_0; t_0) - \mathbf{X}' \mathbf{C}_z^{-1} \mathbf{c}_0),$$

represents the additional uncertainty brought to the prediction (relative to simple kriging) due to the estimation of $\boldsymbol{\beta}$.

Spatio-Temporal Prediction: Optimal Prediction Example

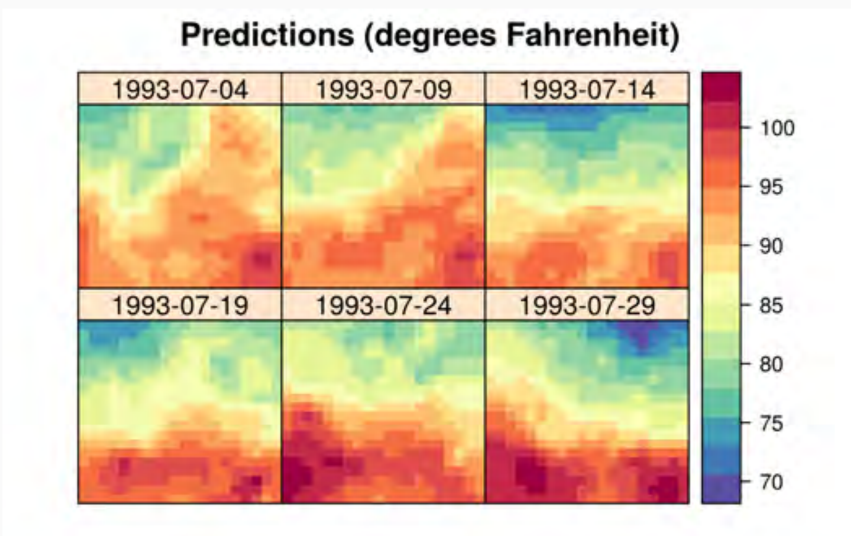


Figure: Universal kriging predictions of NOAA maximum temperature (in degrees Fahrenheit) within a square box enclosing the domain of interest for six days in July 1993, using the R package **gstat**. Data for 14 July 1993 were omitted from the original dataset.

Spatio-Temporal Prediction: Optimal Prediction Example

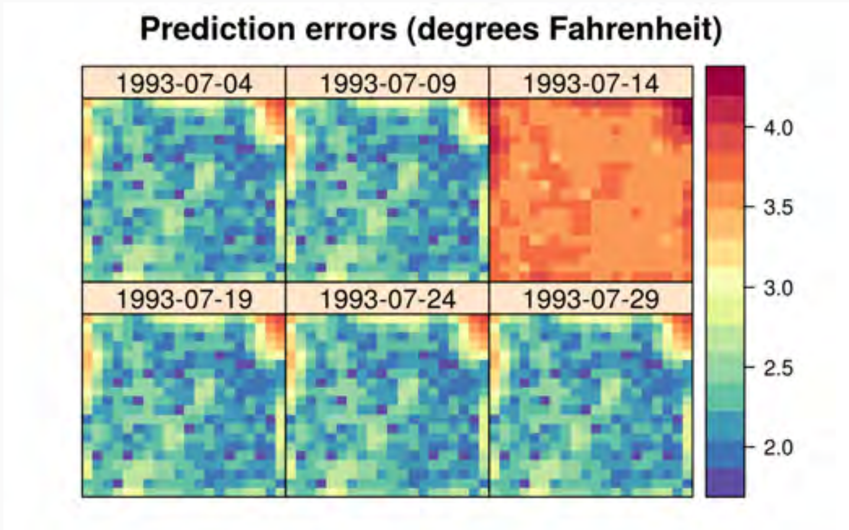


Figure: Universal kriging standard errors for predictions of NOAA maximum temperature (in degrees Fahrenheit) within a square box enclosing the domain of interest for six days in July 1993, using the R package **gstat**. Data for 14 July 1993 were omitted from the original dataset.

Covariance Functions (pp. 143–151 of STSwR)

Spatio-Temporal Prediction: Covariance Functions

Recall that we considered the spatio-temporal process as a *Gaussian process* here and thus we need the covariance between any two space-time locations. This requires that we specify a mean function (easy) and a covariance function (not so easy).

So far, we assumed that we knew the variances and covariances that make up $\mathbf{C}_y$, $\mathbf{C}_\epsilon$, $\mathbf{c}_0$, and $c_{0,0}$. In the real-world we would rarely (if ever) know these.

Spatio-Temporal Prediction: Covariance Functions

Seemingly simple solution: specify covariance functions in terms of parameters, $\boldsymbol{\theta}$, that can give a covariance between any pair of space-time coordinates, and then estimate the parameters (see below).

In spatio-temporal data analysis, as with spatial statistics, the specification of these covariance functions becomes one of the most challenging components of the problem. Why?

- Covariance functions **must be** positive-semidefinite (this ensures that prediction variances are non-negative!)
- Covariance functions should be realistic (that is, they should account for realistic dependencies)

Spatio-Temporal Prediction: Covariance Functions

In practice, classical kriging implementations typically make a stationarity assumption.

For example, if the random process is assumed to be *second-order (weakly) stationary* then it has constant expectation and a covariance function that can be expressed in terms of the spatial and temporal lags

$$c(\mathbf{h}\,;\tau) \equiv c(\mathbf{s}_{ij} - \mathbf{s}_{k\ell}\,;\; t_j - t_\ell), \tag{1}$$

where $\mathbf{h} \equiv \mathbf{s}_{ij} - \mathbf{s}_{k\ell}$ and $\tau \equiv t_j - t_\ell$ are the spatial and temporal lags, respectively.

Recall from our exploratory analyses that if the spatial lag does not depend on direction (i.e., $h = ||\mathbf{h}||$), we say there is spatial *isotropy*.

Spatio-Temporal Prediction: Covariance Functions

Two biggest benefits of the stationarity assumption (both of which facilitate estimation):

- allows for more parsimonious parameterizations of the covariance function
- provides pseudo-replication of dependencies at given lags in space and time

The next question is how do we specify (or construct) such valid covariance functions? Traditionally,

- separable covariance functions
- sums-and-products formulations
- spectral construction
- stochastic partial differential equation (SPDE) solutions

Spatio-Temporal Prediction: Covariance Functions

Separable Covariance Functions:

$$c(\mathbf{h};\tau) \equiv c^{(s)}(\mathbf{h}) \cdot c^{(t)}(\tau),$$

which is valid if both the spatial covariance, $c^{(s)}(\mathbf{h})$, and temporal covariance $c^{(t)}(\tau)$ are valid.

Spatio-Temporal Prediction: Covariance Functions

There are a large number of classes of valid spatial and temporal covariance functions in the literature (e.g., the Matérn, powered exponential, and Gaussian classes, to name but a few). For example, the exponential covariance function is given by

$$c^{(t)}(\tau) = \sigma^2 \exp\left\{-\frac{|\tau|}{a}\right\},$$

where σ^2 is the variance parameter and a is the dependence parameter (sometimes called the e-folding time).

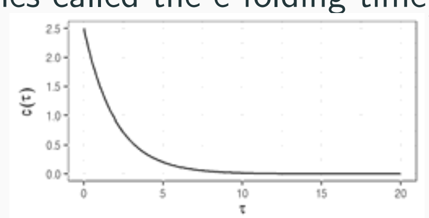


Figure: Exponential covariance function for time lag, τ; $\sigma^2 = 2.5$ and $a = 2$.

Spatio-Temporal Prediction: Covariance Functions

Benefit of Separable Covariance Functions:

Separable models help with computation in problems with many spatial and/or temporal locations (e.g., matrix inverses).

Technical Note:

For example, assume that $Z(\mathbf{s}_{ij}; t_j)$ is observed at each of $m_j = m$ locations and each time point, $j = 1, \ldots, T$.

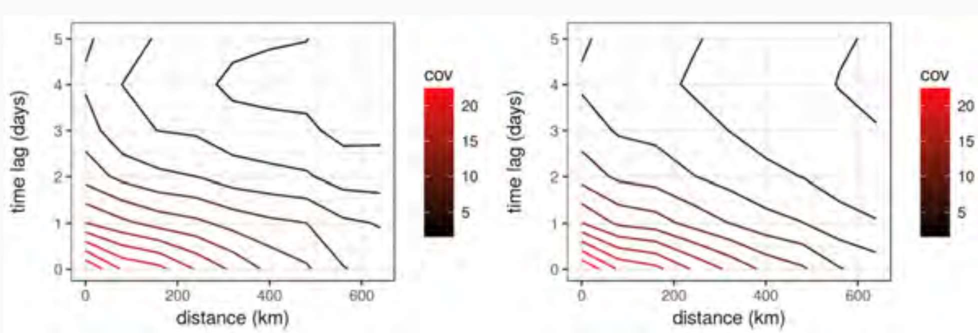
Separability allows us to write: $\mathbf{C}_z = \mathbf{C}_z^{(t)} \otimes \mathbf{C}_z^{(s)}$, where $\otimes$ is the Kronecker product, $\mathbf{C}_z^{(s)}$ is the $m \times m$ spatial covariance matrix, and $\mathbf{C}_z^{(t)}$ is the $T \times T$ temporal covariance matrix.

In this case, $\mathbf{C}_z^{-1} = (\mathbf{C}_z^{(t)})^{-1} \otimes (\mathbf{C}_z^{(s)})^{-1}$, which shows that one only has to take the inverse of m and T dimensional matrices to get the inverse of the much larger $(mT) \times (mT)$ matrix.

Spatio-Temporal Prediction: Covariance Functions

Separable Covariance Functions:

Consider the NOAA maximum temperature dataset. The left panel below shows a contour plot of the empirical spatio-temporal covariance function and the right panel shows the product of the empirical spatial $(\hat{c}^{(s)}(\|\mathbf{h}\|; 0))$ and temporal $(\hat{c}^{(t)}(\mathbf{0}; |\tau|))$ marginal covariance functions. These plots suggest that separability may be a reasonable assumption for these data.



Spatio-Temporal Prediction: Covariance Functions

Separability is unusual in spatio-temporal processes; it says that temporal evolution of the process at a given spatial location does not depend directly on the process' temporal evolution at other locations. That is, separability comes from a lack of spatio-temporal interaction in $Y(\cdot;\cdot)$.

E.g., Storvik (2002) showed that a process with covariance function

$$c(\mathbf{h};\tau) \equiv c^{(s)}(\mathbf{h}) \cdot a^{|\tau|},$$

corresponds to a spatio-temporal dynamic model of the form

$$Y(\mathbf{s};t) = aY(\mathbf{s};t-1) + \epsilon(\mathbf{s};t)$$

where ϵ has spatial correlation function proportional to $c^{(s)}(\mathbf{h})$.

Spatio-Temporal Prediction: Covariance Functions

What other options do we have to specify S-T covariance functions?

- Sums and products of covariance functions
- Construction (e.g., spectrally, via Bochner's Theorem, which relates the spectral representation to the covariance representation)
- Solving a stochastic partial differential equation (SPDE)

Descriptive Spatio-Temporal Modeling: Part II (Estimation) (pp. 151–154 of STSwR)

Spatio-Temporal Prediction: Estimation

The spatio-temporal covariance models presented above depend on unknown parameters; these must be estimated from the data.

There is a history in spatial statistics of fitting such models directly to the empirical estimates (e.g., by using a least-squares or weighted-least-squares approach; see Cressie 1993, Section 2.6 for an overview).

In the spatio-temporal context, one typically considers covariance models and estimates parameters through likelihood-based methods or through fully Bayesian methods. This follows closely the approaches in linear mixed model parameter estimation (for an overview see McCulloch and Searle, 2001).

Spatio-Temporal Prediction: Estimation

Likelihood Estimation:

Recall $\mathbf{C}_z = \mathbf{C}_y + \mathbf{C}_\epsilon$ and this depends on parameters $\boldsymbol{\theta} \equiv \{\boldsymbol{\theta}_y, \boldsymbol{\theta}_\epsilon\}$ for the latent and error processes, respectively. The likelihood can then be written as,

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) \propto |\mathbf{C}_z(\boldsymbol{\theta})|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{z} - \mathbf{X}\beta)'(\mathbf{C}_z(\boldsymbol{\theta}))^{-1}(\mathbf{z} - \mathbf{X}\beta)\right\},$$

and we maximize this likelihood with respect to $\{\boldsymbol{\beta}, \boldsymbol{\theta}\}$, thus obtaining the *maximum likelihood estimates* (MLEs).

The fact that the covariance parameters appear in the matrix inverse and determinant prohibit analytical maximization, but numerical methods can be used (e.g., Newton-Raphson) if the dimension of $\boldsymbol{\theta}$ isn't too large and $\mathbf{C}_z(\boldsymbol{\theta})^{-1}$ is tractable.

Spatio-Temporal Prediction: Estimation

Restricted Maximum Likelihood Estimation (REML):

REML considers the likelihood of a linear transformation of the data vector such that the errors are orthogonal to the **X**s that make up the mean function.

Numerical maximization of the associated likelihood, which is only a function of the parameters $\boldsymbol{\theta}$, is typically more computationally efficient than for MLEs (and, conventional wisdom is that the bias properties of the REML estimates are better than those for MLEs for linear mixed models.)

Spatio-Temporal Prediction: Estimation

Fully (Hierarchical) Bayesian Estimation:

Recall, we can decompose an arbitrary joint distribution in terms of a hierarchical sequence of conditional distributions and a marginal distribution, for example, $[A, B, C] = [A|B, C][B|C][C]$.

In the context of the general spatio-temporal model,

$$\begin{aligned}[\mathbf{z}, \mathbf{Y}, \boldsymbol{\beta}, \boldsymbol{\theta}] &= [\mathbf{z}|\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\beta}][\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\theta}][\boldsymbol{\beta}|\boldsymbol{\theta}][\boldsymbol{\theta}] \\ &= [\mathbf{z}|\mathbf{Y}, \boldsymbol{\theta}_\epsilon][\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\theta}_y][\boldsymbol{\theta}][\boldsymbol{\beta}],\end{aligned}$$

where $\boldsymbol{\theta}$ contains all of the variance and covariance parameters from the data model and the process model.

Note that the first equality is based on the probability decomposition, and the second equality is based on writing $\boldsymbol{\theta} = \{\boldsymbol{\theta}_\epsilon, \boldsymbol{\theta}_y\}$ and assuming the priors on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are independent.

Spatio-Temporal Prediction: Estimation

Fully (Hierarchical) Bayesian Estimation:
Now, Bayes' rule gives the *posterior distribution*

$$[\mathbf{Y}, \boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{z}] \propto [\mathbf{z}|\mathbf{Y}, \boldsymbol{\theta}_\epsilon][\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\theta}_y][\boldsymbol{\theta}][\boldsymbol{\beta}],$$

where $[\mathbf{z}|\mathbf{Y}, \boldsymbol{\theta}_\epsilon]$ is the data model and $[\mathbf{Y}|\boldsymbol{\beta}, \boldsymbol{\theta}_y]$ is the latent process model. The prior distributions for the parameters, $[\boldsymbol{\theta}]$ and $[\boldsymbol{\beta}]$, are then specified according to the specific choices one makes for the error and process covariance functions.

In general, the normalizing constant required to fully specify the posterior distribution is not available analytically and numerical sampling methods (e.g., Markov chain Monte Carlo, MCMC) must be used.

Advantage of the Bayesian hierarchical model (BHM) approach:
parameter uncertainty is accounted for directly.

Descriptive Spatio-Temporal Modeling: Part III (Basis Functions and Random Effects Parameterizations) (pp. 157–164 of STSwR)

Spatio-Temporal Basis-Function and Random-Effects Models

Two problems with spatio-temporal Kriging:

- difficulty in working with large spatio-temporal covariance matrices (e.g., $\mathbf{C}_z^{-1}$) in situations with large numbers of prediction or observation locations
- realistic covariance structure

Possible solution:

Expand the spatio-temporal process in terms of basis functions and associated random effects and take advantage of conditional specifications that a hierarchical modeling framework allows.

This allows us to build the dependence rather than specify it directly!

Basis Functions

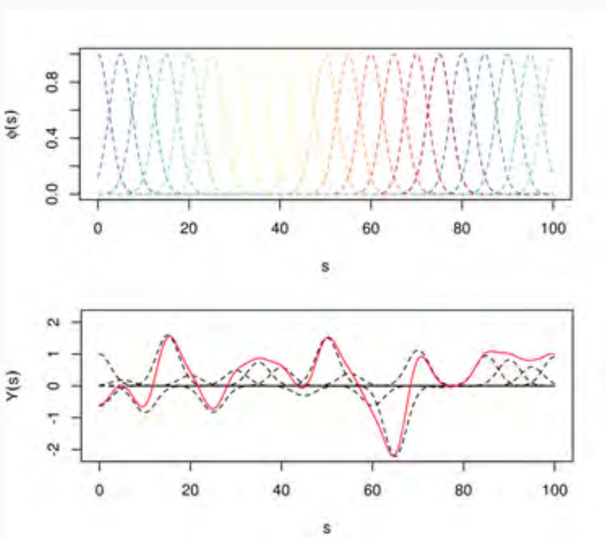
- Imagine that we have a complex curve or surface. Informally, we can decompose this curve or surface as a linear combination of some "elemental" functions. For example,

$$Y(x) = \alpha_1\phi_1(x) + \alpha_2\phi_2(x) + \ldots + \alpha_p\phi_p(x),$$

  where α_i are expansion coefficients and $\phi_i(x)$ are known *basis functions*.
- The $\{\alpha_i\}$ coefficients are weights that describe how important each basis function is in representing the function $Y(x)$.

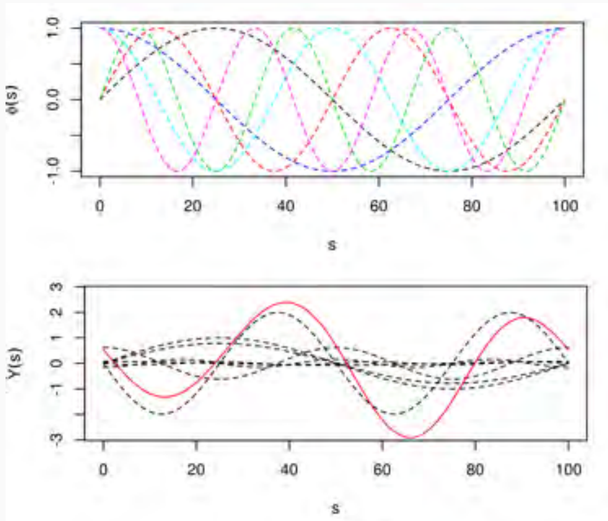
Basis Functions

The basis functions can be *local* with compact support. For example, consider below a set of local basis functions in a one-dimensional domain, as well as the reconstructed $Y(x)$.



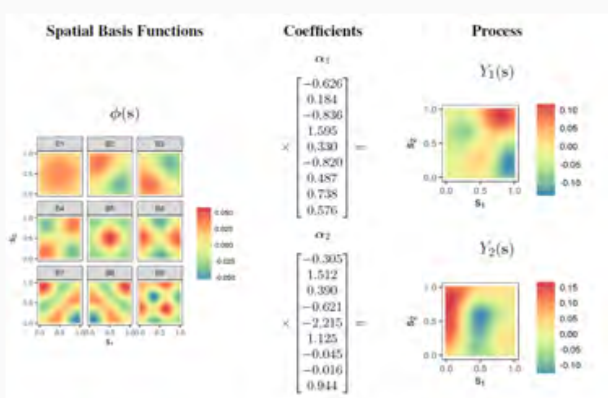
Basis Functions

The basis functions can also be *global*, taking values across the whole domain. For example, consider below a set of local basis functions in a one-dimensional domain, as well as the reconstructed $Y(x)$.



Basis Functions

- When $Y(x)$ is a random process and the basis functions are known, then the coefficients (weights) are random.
- Consider below nine two-dimensional spatial basis functions and associated random coefficients (α_1, α_2) that lead to two different spatial-process realizations.



Basis Functions

- Examples of basis functions include polynomials, splines, wavelets, sines and cosines, empirical orthogonal functions, among many others.
- Often, spatio-temporal basis functions are constructed from a spatial basis function and a temporal basis function via a *tensor product.*

Spatio-Temporal Basis Function and Random Effects Models

As in traditional linear mixed models, consider:

$$\mathbf{z} = \mathbf{Y} + \varepsilon,$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Phi}\boldsymbol{\alpha},$$

where $\varepsilon \sim Gau(\mathbf{0}, \mathbf{C}_\epsilon)$, and $\boldsymbol{\alpha} \sim Gau(\mathbf{0}, \mathbf{C}_\alpha)$.

Alternatively, we can write this conditionally:

$$\mathbf{z}|\mathbf{Y} \sim Gau(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Phi}\boldsymbol{\alpha}, \mathbf{C}_\epsilon),$$

$$\boldsymbol{\alpha} \sim Gau(\mathbf{0}, \mathbf{C}_\alpha).$$

Integrating out the random effects ($\boldsymbol{\alpha}$) we get the marginal representation:

$$\mathbf{z} \sim Gau(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Phi}\mathbf{C}_\alpha\boldsymbol{\Phi}' + \mathbf{C}_\epsilon).$$

Thus, averaging across the random effects induces marginal dependence (e.g., $\boldsymbol{\Phi}\mathbf{C}_\alpha\boldsymbol{\Phi}' + \mathbf{C}_\epsilon$ versus just $\mathbf{C}_\epsilon$).

Spatio-Temporal Basis Function and Random Effects Models

In spatial and spatio-temporal cases, we sometimes include an additional random effect, e.g.,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Phi}\boldsymbol{\alpha} + \boldsymbol{\nu},$$

where $\boldsymbol{\nu} \sim Gau(\mathbf{0}, \mathbf{C}_{\nu})$ describes fine-scale variation.

Spatio-Temporal Basis Function and Random Effects Models

Consider three basis function approaches for modeling spatio-temporal processes:

- spatio-temporal basis functions: the basis functions are indexed in space and time and the random effects have no specific location index (e.g., package **FRK**)
- temporal basis functions: the basis functions are indexed by time and the random effects are indexed by space (e.g., package **SpatioTemporal**)
- spatial basis functions: the basis functions are indexed by space and the random effects are indexed by time (e.g., package **INLA**)

Spatio-Temporal Basis Function and Random Effects Models

Spatio-Temporal Basis Functions:

Rewrite the process model in terms of fixed and random effects ($\boldsymbol{\beta}$ and $\{\alpha_i : i = 1, \ldots, n_\alpha\}$, respectively):

$$Y(\mathbf{s}; t) = \mathbf{x}(\mathbf{s}; t)'\boldsymbol{\beta} + \sum_{i=1}^{n_\alpha} \phi_i(\mathbf{s}; t)\alpha_i + \nu(\mathbf{s}; t),$$

where $\{\phi_i(\mathbf{s}; t) : i = 1, \ldots, n_\alpha\}$ are (typically) specified basis functions corresponding to location $(\mathbf{s}; t)$, $\{\alpha_i\}$ are random effects, and $\nu(\mathbf{s}; t)$ is sometimes needed to represent left-over small-scale spatio-temporal random effects.

Let $\boldsymbol{\alpha} \sim Gau(\mathbf{0}, \mathbf{C}_\alpha)$, where $\boldsymbol{\alpha} \equiv (\alpha_1, \ldots, \alpha_{n_\alpha})'$.

Spatio-Temporal Basis Function and Random Effects Models

Note, we are interested in the Y-process at n_y spatio-temporal locations, which we denote by the vector $\mathbf{Y}$.

The process model then becomes

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Phi}\boldsymbol{\alpha} + \boldsymbol{\nu},$$

where the ith column of the $n_y \times n_\alpha$ matrix $\boldsymbol{\Phi}$ corresponds to the ith basis function, $\phi_i(\cdot;\cdot)$, at all of the spatio-temporal locations ordered as given in $\mathbf{Y}$; and the vector $\boldsymbol{\nu}$ also corresponds to the spatio-temporal ordering given in $\mathbf{Y}$ such that $\boldsymbol{\nu} \sim Gau(\mathbf{0}, \mathbf{C}_\nu)$.

Spatio-Temporal Basis Function and Random Effects Models

The marginal distribution of $\mathbf{Y}$ is then

$$\mathbf{Y} \sim Gau(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Phi}\mathbf{C}_\alpha\boldsymbol{\Phi}' + \mathbf{C}_\nu),$$

so that $\mathbf{C}_Y = \boldsymbol{\Phi}\mathbf{C}_\alpha\boldsymbol{\Phi}' + \mathbf{C}_\nu$.

In this case, the spatio-temporal dependence is primarily accounted for by the spatio-temporal basis functions, $\boldsymbol{\Phi}$, and the random effects covariance matrix $\mathbf{C}_\alpha$; in general, this could accommodate non-separable dependence.

The benefits of this approach are that the modeling effort then focuses on the random effects, $\boldsymbol{\alpha}$ (which are typically of relatively low dimension and do not have spatial or temporal dependence).

Spatio-Temporal Basis Function and Random Effects Models

Low-Rank Representation:

In situations where n_α is much smaller than the dimension of $\mathbf{Y}$ (i.e., a *low-rank representation*), then obvious benefits come from being able to perform matrix inverses in terms of n_α-dimensional matrices (through well-known matrix algebra relationships).

Technical Note:

Specifically, let $\mathbf{V} \equiv \mathbf{C}_\nu + \mathbf{C}_\epsilon$, then $\mathbf{C}_z$ in the spatio-temporal kriging formulas above can be written $\mathbf{C}_z = \boldsymbol{\Phi}\boldsymbol{\Sigma}_\alpha\boldsymbol{\Phi}' + \mathbf{V}$. Thus, using the well-known Sherman-Morrison-Woodbury matrix identities

$$\mathbf{C}_z^{-1} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\mathbf{V}^{-1}\boldsymbol{\Phi} + \mathbf{C}_\alpha^{-1})^{-1}\boldsymbol{\Phi}'\mathbf{V}^{-1}.$$

If $\mathbf{V}^{-1}$ has a simple structure (i.e., is diagonal or very sparse) and $n_\alpha \ll n_Y$, then this inverse is easy to calculate. This is the essence of "fixed rank kriging."

Spatio-Temporal Basis Function and Random Effects Models

It is important to note that even in the *full-rank* (n_α equal to the dimension of $\mathbf{Y}$) and *over-complete* (n_α larger than the dimension of $\mathbf{Y}$) cases, there can still be dramatic computational benefits through induced sparsity in $\mathbf{C}_\alpha$ and efficient matrix multiplication routines that utilize multi-resolution algorithms.

In addition, some implementations assume that $\boldsymbol{\nu} = \mathbf{0}$. One may also choose basis functions that are *orthonormal*, so that $\boldsymbol{\Phi}\, \boldsymbol{\Phi}' = \mathbf{I}$, to further simplify computations.

Finally, we note that specific basis functions and methodologies take advantage of other properties of the various matrices, for example, *sparse structure* on the random-effects covariance matrix, $\mathbf{C}_\alpha$, or the random-effects precision matrix, $\mathbf{C}_\alpha^{-1}$ (e.g., **LatticeKrig**).

Spatio-Temporal Basis Function and Random Effects Models

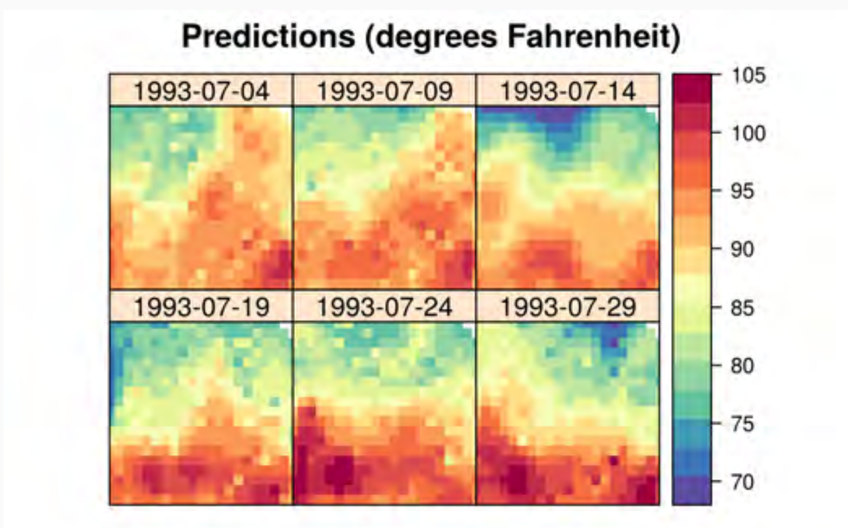


Figure: Predictions of NOAA maximum temperature for six days spanning the temporal window of the data, 01 July 1993 – 30 July 1993 using spatio-temporal bisquare basis functions implemented in the **FRK** (fixed rank kriging) R package.

Spatio-Temporal Basis Function and Random Effects Models

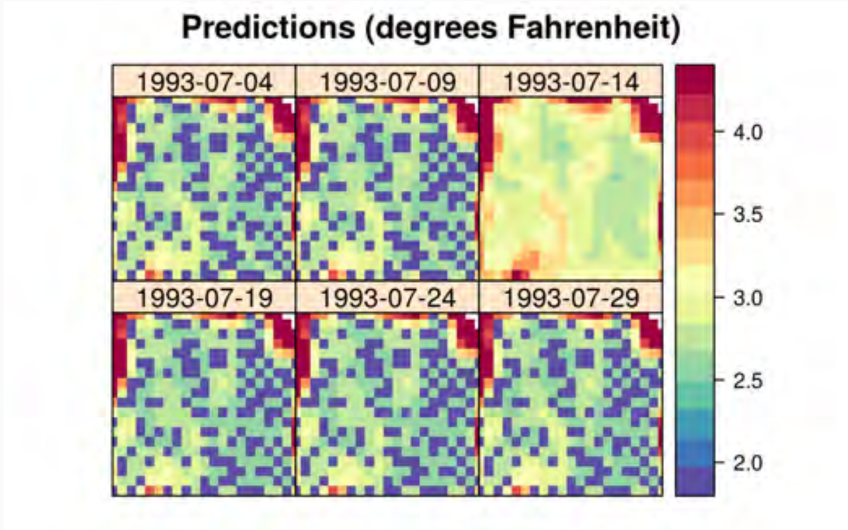


Figure: Prediction standard errors of NOAA maximum temperature for six days spanning the temporal window of the data, 01 July 1993 – 30 July 1993 using spatio-temporal bisquare basis functions implemented in the **FRK** (fixed rank kriging) R package.

Spatio-Temporal Basis Function and Random Effects Models

Random Effects with Temporal Basis Functions:

We can also express the spatio-temporal random process in terms of *temporal basis functions* and *spatially-indexed random effects*

$$Y(\mathbf{s};t) = \mathbf{x}(\mathbf{s};t)'\beta + \sum_{i=1}^{n_\alpha} \phi_i(t)\alpha_i(\mathbf{s}) + \nu(\mathbf{s};t),$$

where $\{\phi_i(t) : i = 1, \ldots, n_\alpha; t \in D_t\}$ are temporal basis functions and $\alpha_i(\mathbf{s})$ are spatially-indexed random effects.

In this case, one could model $\{\alpha_i(\mathbf{s})\}$ using the methods from spatial statistics. That is, either n_α separate spatial processes, or an n_α-dimensional multivariate spatial process (see Cressie and Wikle, 2011, Ch. 4).

Spatio-Temporal Basis Function and Random Effects Models

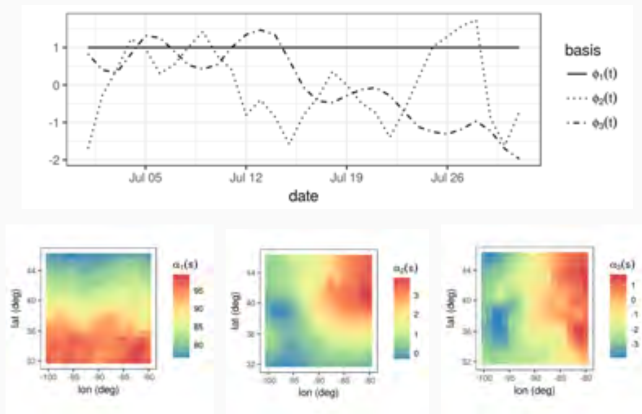


Figure: NOAA maximum temperature data temporal basis functions example. Top panel: Basis functions $\phi_1(t), \phi_2(t)$, and $\phi_3(t)$, where the first basis function is a constant and latter two were obtained from the temporal EOFs from the data matrix. Bottom panels: $E(\alpha_0(\mathbf{s}) \mid \mathbf{z}), E(\alpha_1(\mathbf{s}) \mid \mathbf{z})$, and $E(\alpha_2(\mathbf{s}) \mid \mathbf{z})$. Note: modeling and prediction using temporal basis functions can be carried out with the **SpatioTemporal** R package.

Spatio-Temporal Basis Function and Random Effects Models

Random Effects with Spatial Basis Functions

We can also express the spatio-temporal random process in terms of *spatial basis functions* and *temporally-indexed random effects*

$$Y(\mathbf{s}; t) = \mathbf{x}(\mathbf{s}; t)'\boldsymbol{\beta} + \sum_{i=1}^{n_\alpha} \phi_i(\mathbf{s})\alpha_t(i) + \nu(\mathbf{s}; t),$$

where $\{\phi_i(\mathbf{s}) : i = 1, \ldots, n_\alpha; \mathbf{s} \in D_s\}$ are spatial basis functions and $\alpha_t(i)$ are temporal random effects.

The modeling focus is then on modeling the random temporal effects vectors, $\boldsymbol{\alpha}_t \equiv (\alpha_t(1), \ldots, \alpha_t(n_\alpha))'$.

Spatio-Temporal Basis Function and Random Effects Models

If the vectors $\boldsymbol{\alpha}_t$ are independent in time, then the marginal dependence structure of $\{Y(\mathbf{s};t)\}$ is typically unrealistic.

Interesting spatio-temporal dependence arises when these effects are dependent; such models are simplified by assuming conditional temporal dependence (dynamics!), as we will see in the next section when we talk about dynamical representations of spatio-temporal processes.

Descriptive Spatio-Temporal Modeling: Part IV (Non-Gaussian Data Models with Latent Gaussian Processes) (pp. 165–170 of STSwR); Not covered here

Deterministic, Regression, and Descriptive Modeling of Spatio-Temporal Data: R Examples

Chapter Labs in STSwR:

- Lab 3.1: Deterministic Prediction Methods
- Lab 3.2: Trend Prediction
- Lab 3.3: Regression Models for Forecasting
- Lab 3.4: Generalized Linear Spatio-Temporal Regression
- Lab 4.1: Spatio-Temporal Kriging with **gstat**
- Lab 4.2: Spatio-Temporal Basis Functions with **FRK**
- Lab 4.3: Temporal Basis Functions with **SpatioTemporal**
- Lab 4.4: Non-Gaussian Spatio-Temporal GAMs with **mgcv**
- Lab 4.5: Non-Gaussian Spatio-Temporal Models with **INLA**

Lab R Code available at: https://spacetimewithr.org/code

Statistical Dynamical Spatio-Temporal Models (DSTMs) (pp. 205–209 of STSwR)

Motivation

One of the biggest challenges with the *descriptive* approach to spatio-temporal modeling that we talked about before is that most real-world spatio-temporal processes are more complex than can be realistically specified by the relatively simple classes of spatio-temporal covariance functions that are available.

We can improve on this by using basis-function expansions and random effects, but it is often the case that the random effects have temporal dependence.

We can accommodate this temporal dependence by considering a *conditional* or *dynamic* perspective, in which we think of spatial processes (or basis coefficients) evolving through time in some probabilistic manner.

Dynamics

- Spatio-temporal dynamics are due to the interaction of the process components across space and time and/or across scales of variability
  - Some types of interaction make sense for some processes, and some don't (e.g., process knowledge should not be ignored if available)
  - Statisticians have often ignored such knowledge!
- Dimensionality can prevent the (efficient) estimation of model parameters
  - Requires sensible science-based parameterizations and/or dimension reduction; sparse structures; regularization
  - Deep (hierarchical) representations can help here as well

September 25, 2010: Radar Nowcasting Application!

It may look like something out of a movie, but it's not — this was the scene from early Saturday afternoon in Kansas. (AP Photo/Charlie Riedel)

DSTM Radar Data

Weather Radar Reflectivity: How can we model this behavior?

Basic Modeling Framework

There are two critical assumptions for DSTMs:

- Data conditioned on the latent process can be factored into the product of independent distributions:

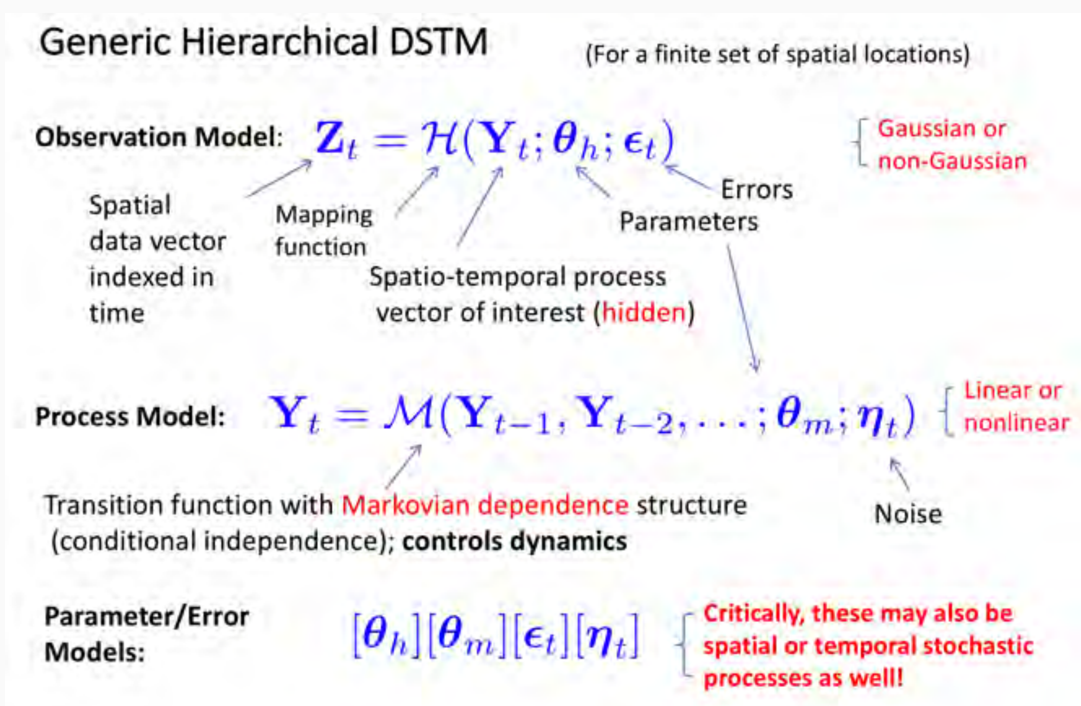
$$[\mathbf{z}_T, \ldots, \mathbf{z}_1 | \mathbf{Y}_T, \ldots, \mathbf{Y}_1, \boldsymbol{\theta}_d] = \prod_{t=1}^{T} [\mathbf{z}_t | \mathbf{Y}_t, \boldsymbol{\theta}_d]$$

- The joint distribution of the latent process can be factored into conditional (in time) models (e.g., first-order model):

$$[\mathbf{Y}_T, \ldots, \mathbf{Y}_1, \mathbf{Y}_0 | \boldsymbol{\theta}_p] = \prod_{t=1}^{T} [\mathbf{Y}_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}_p][\mathbf{Y}_0 | \boldsymbol{\theta}_p]$$

Challenge: specification of the models associated with these component distributions

General Hierarchical DSTM



DSTM Modeling Issues

- Linear DSTMs
  - Data Models
  - Process Models
  - Efficient Parameterization
  - Basis-Function Representations
  - Estimation and Prediction
- Nonlinear DSTMs
  - Threshold Models
  - General Quadratic Nonlinear Models
  - Individual-Based Models
  - Analog Models
  - Recurrent Neural Network Models

Linear DSTMs: Process Model

We need a model corresponding to the distribution:

$$[\mathbf{Y}_t | \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots, \boldsymbol{\theta}_m] \quad \text{(often assume } [\mathbf{Y}_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}_m])$$

Assumption: the process at the current time is related to the process at a previous time (or times). In general, we can consider spatio-temporal processes in continuous time and/or space. For brevity, we focus here on the case where time is discrete and equally spaced.

We refer to such a process as a space-time dynamical process.

Linear DSTM: Process Model

In the linear process case, the conditional distribution $[\mathbf{Y}_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}_m]$ implies a first-order Markov model of the form:

$$Y_t(\mathbf{s}_i) = \sum_{j=1}^{n} m_{ij}(\boldsymbol{\theta}_m) Y_{t-1}(\mathbf{s}_j) + \eta_t(\mathbf{s}_i),$$

where $m_{ij}(\boldsymbol{\theta}_m)$ describes how the spatial process at the previous time at location j gets redistributed to location i at the current time, according to the parameters, $\boldsymbol{\theta}$

These equations imply a matrix model:

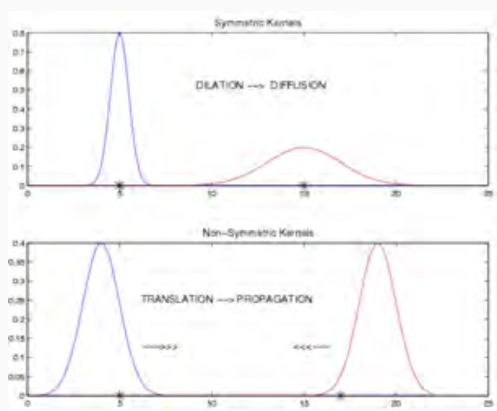
$$\mathbf{Y}_t = \mathbf{M}(\boldsymbol{\theta}_m)\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim Gau(\mathbf{0}, \mathbf{C}_\eta)$$

Challenge: parsimonious parameterization of $\mathbf{M}(\boldsymbol{\theta}_m)$!

Linear DSTM: Process Behavior

Dynamical behavior is implied by changes in the transition operator "shape": e.g., linear spatio-temporal processes often exhibit advective and diffusive behavior:

- "width" (decay rate) of the transition operator neighborhood controls the rate of spread (diffusion)
- degree of "asymmetry" in the transition operator controls the speed and direction of propagation (advection)
- "long range dependence" can be accommodated by "multimodal" operators and/or heavy tails

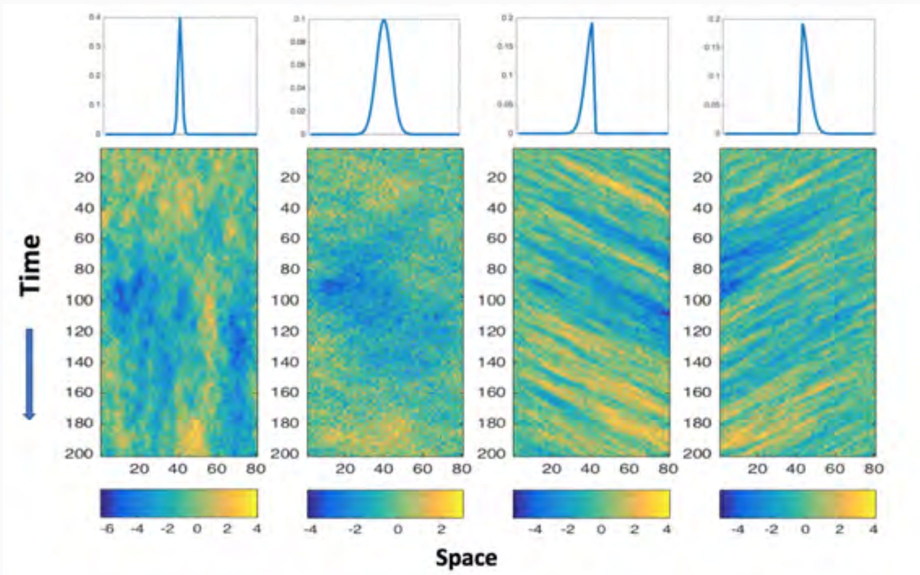


This suggests ways that we might parameterize the transition operator and/or induce sparse structure.

Linear DSTM: Process Behavior

Integro-difference equation (IDE) (kernel) representation:

$$Y_t(\mathbf{s}) = \int_{D_s} m(\mathbf{s}, \mathbf{x}; \boldsymbol{\theta}_m) Y_{t-1}(\mathbf{x}) d\mathbf{x} + \eta_t(\mathbf{s}), \quad \mathbf{s}, \mathbf{x} \in D_s$$



Linear DSTM: Process Behavior

Simulation of an advection-diffusion process from a linear DSTM.

Basic Hierarchical Linear DSTM

Data:

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{Y}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim Gau(\mathbf{0}, \mathbf{C}_{\epsilon,t}(\boldsymbol{\theta}_d))$$

Process:

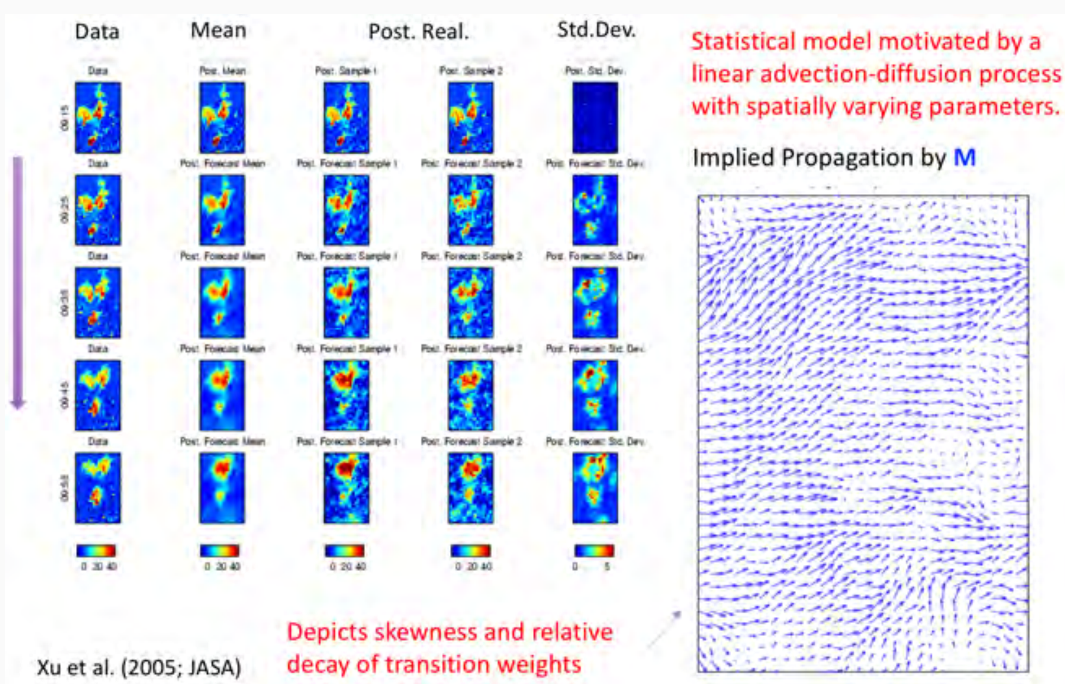
$$\mathbf{Y}_t = \mathbf{M}(\boldsymbol{\theta}_{m,1}) \mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim Gau(\mathbf{0}, \mathbf{C}_{\eta}(\boldsymbol{\theta}_{m,2}))$$

Parameters:

$\boldsymbol{\theta}_d, \ \boldsymbol{\theta}_{m,1}, \ \boldsymbol{\theta}_{m,2}$

These parameters may be estimated empirically, but we get more flexibility if they are given dependent prior distributions, such as Gaussian random process priors (that may depend on other variables), and they can easily be allowed to vary with time and/or space.

Example: Radar Nowcasting (Sydney, pre 2000 Olympics)



Nonlinear Spatio-Temporal Dynamic Models (pp. 224–228 of STSwR)

Nonlinear Spatio-Temporal Processes

Few environmental processes are linear (e.g., growth, nonlinear advection, density dependence, shock waves, repulsion, predator-prey, etc.)

- Nonlinear dynamical behavior arises from the complicated interactions across spatio-temporal scales of variability and interactions across multiple processes
- It is important to consider this if forecasting or filling in big gaps is important
- Examples in mechanistic models across many disciplines

Nonlinear DSTMs: Statistical Approaches

Over the past few years, several parametric and non-parametric statistical approaches have proven useful for hierarchical formulations of nonlinear DSTMs. Some of these have been motivated by non-statistical models from other disciplines. It is still early in the development of these methods.

- Time-Varying Parameters
- General Quadratic Nonlinearity (GQN)
- Individual (Agent)-Based Models
- "Mechanism-Free" Analog Embedding Models
- Neural Network Models

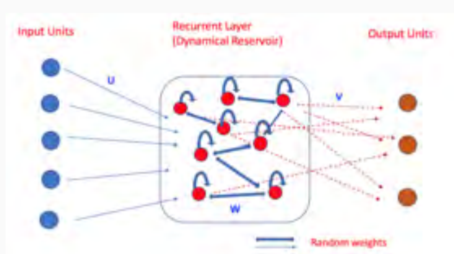
Very Basic Recurrent Neural Network: Echo State Network

$$\mathbf{Y}_t = g(\mathbf{V}\mathbf{h}_t)$$

$$\mathbf{h}_t = \tanh(\mathbf{W}\mathbf{h}_{t-1} + \mathbf{U}\mathbf{x}_t)$$

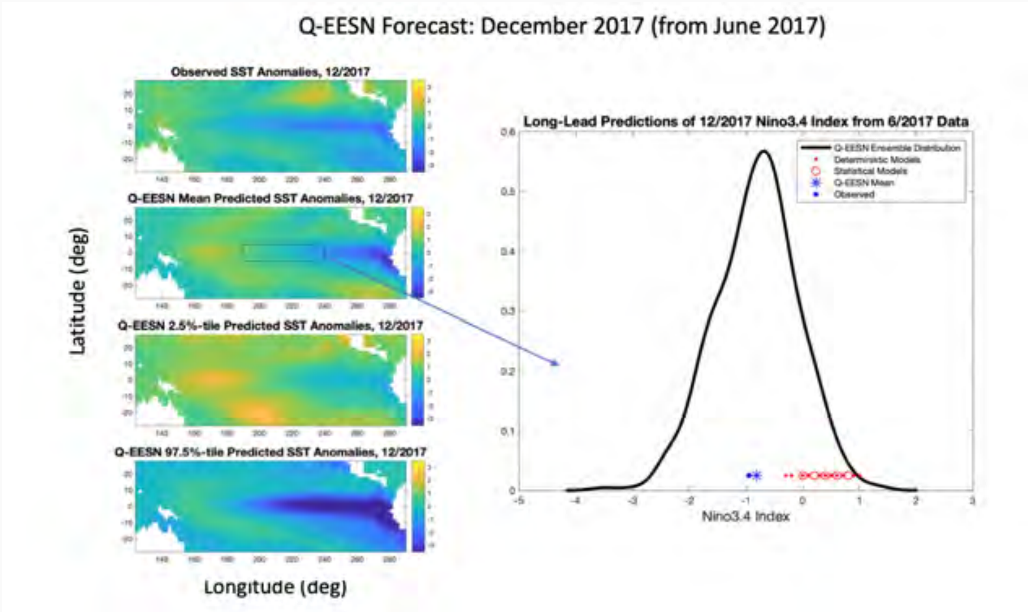
where $\mathbf{Y}_t$ is the output vector, $\mathbf{h}_t$ are hidden states, $\mathbf{x}_t$ are inputs, $\mathbf{V}$, $\mathbf{W}$, and $\mathbf{U}$ are parameter (weight) matrices.

In the so-called Echo State Network (ESN), the weight matrices $\mathbf{W}$ and $\mathbf{U}$ are sparse and chosen randomly. Only the output matrix, $\mathbf{V}$ is trained.



(See Appendix F of STSwR, and McDermott and Wikle, 2017, 2019)

ESN Example



Dynamic Modeling of Spatio-Temporal Data: R Examples

Chapter Labs in STSwR:

- Lab 5.1: Implementing an IDE Model in One-Dimensional Space
- Lab 5.2: Spatio-Temporal Inference using the IDE Model
- Lab 5.3: Spatio-Temporal Inference with Unknown Evolution Operator
- Appendix E: Case Study: Physical-Statistical Bayesian Hierarchical Model for Predicting Mediterranean Surface Winds
- Appendix F: Case Study: Quadratic Echo State Networks for Sea Surface Temperature Long Lead Prediction

Lab R Code available at: https://spacetimewithr.org/code

Evaluating Spatio-Temporal Statistical Models (Chapter 6 of STSwR) [Not Covered Here]