Introduction to Spatio-Temporal Statistics

Christopher K. Wikle, University of Missouri, Dept. of Statistics

Toryn Schafer, Texas A&M University, Dept. of Statistics

March 2024

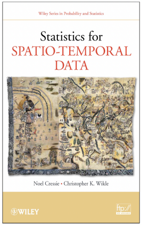
Talk Introduction

This Talk

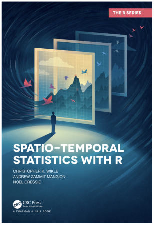
- Why Spatio-Temporal Statistics
- Spatio-Temporal Data
- Exploration and Visualization
- Gaussian Process-Based Descriptive Models
- Spatio-Temporal Dynamic Models
- R Examples

Primary References:

Cressie and Wikle (2011, 2015)

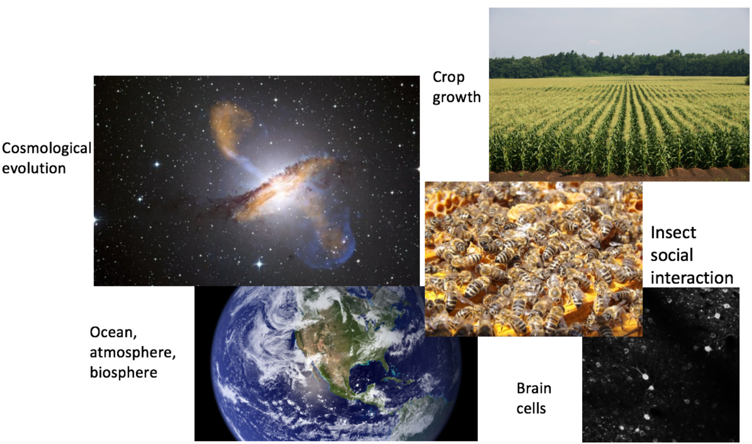


Wikle et al. (2019)
https://spacetimewithr.org
(free to download pdf)



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!

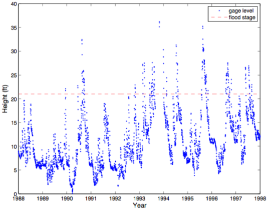


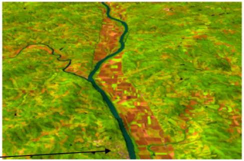
Space AND Time

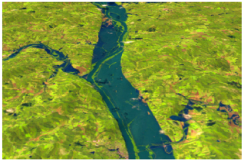
It is not sufficient to consider just snapshots of a spatial process at a given time, or time series at a spatial location – the behavior from one time point to the next is important!

Famous 1993 Missouri River Flood

Hydrologic gage time series; Missouri River







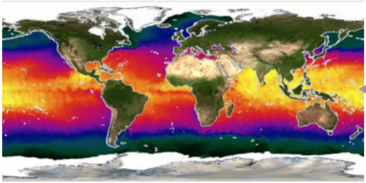
Images from NASA's Landsat Thematic Mapper. Each image shows a segment of the Missouri River near Hermann, MO (mile 96.5, at the bottom of the scene), and Gasconade, MO (mile 104.4, in the "V" in the middle of the scene). The river flows from west (top of the scene) to east (bottom of the scene). Left panel: September 1992, before a major flood event. Right panel: September 1993, after a record-breaking flood event in July 1993.

(Cressie and Wikle, 2011)

Interactions Across Space and Time

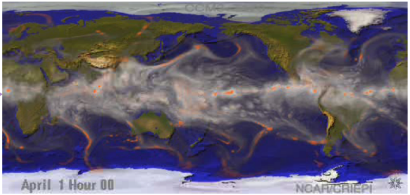
(Sea Surface Temperature, SST)

NASA AMSR: SST (scientific visualization studio)

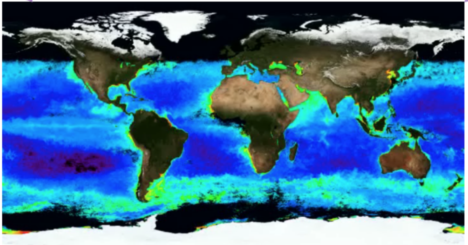


(Clouds/Precipitation)

NCAR CCM3 Cloud/Precipitation Simulation (NCAR VETS)



NASA SEAWIFS: Ocean Color (scientific visualization studio)



(Proxy for ocean phytoplankton)

Interaction across atmosphere, ocean, and ecosystem.

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
- Accounting for dependence in parameter inference
- Assimilation of observations and mechanistic models
- Model emulation
- Classification

Visualizing and Exploring Spatio-Temporal Data

(Wikle et al. 2019, Ch. 2; Cressie & Wikle, 2011, Ch. 5)

Spatio-Temporal Data

Types of Spatio-Temporal Data:

- Univariate or Multivariate
- Time-Series
    - regular or irregular intervals
    - continuous or discrete time
    - random events (i.e., point process)
- Spatial Random Process
    - "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: Spatial Plots

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 options here.

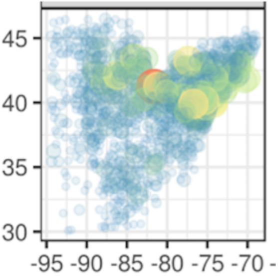
Snapshots of spatial processes for a given time period can be plotted in numerous ways.

- 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

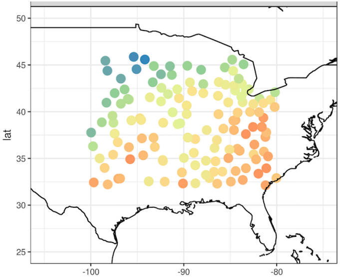
Visualization of Spatio-Temporal Data: Spatial Plots

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:

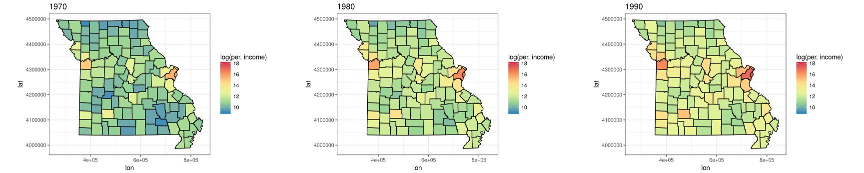
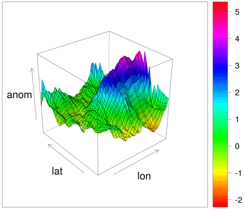
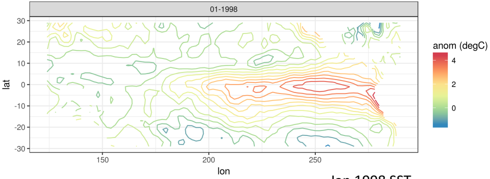


Figure: Personal income for residents in Missouri counties from the US Bureau of Economic Analysis (BEA) for the years 1970, 1980, and 1990.

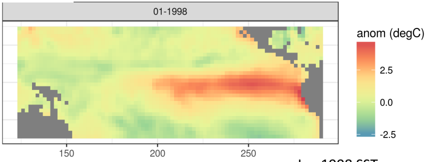
Regular Lattice Data: Image, Contour, Surface Maps:



Jan 1998 SST
surface plot



Jan 1998 SST
contour plot



Jan 1998 SST
image plot

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

Visualization of Spatio-Temporal Data: Map Sequences

Sequence of spatial (point) maps:

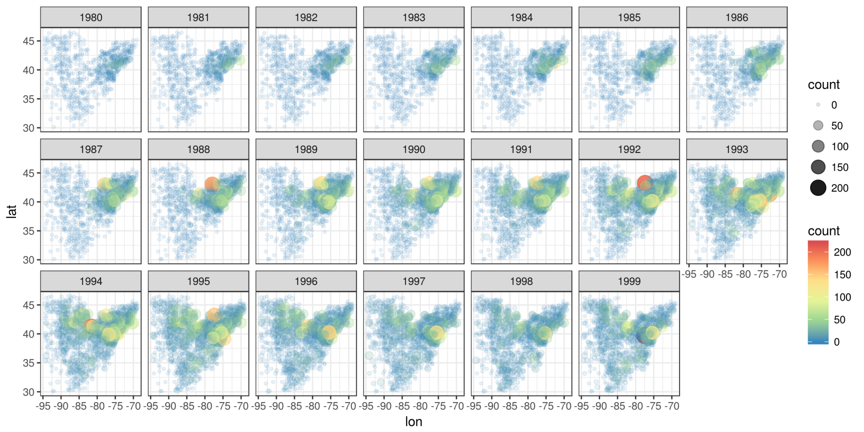


Figure: BBS House Finch counts between 1980-1999.

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).

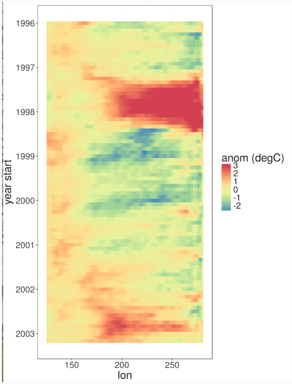
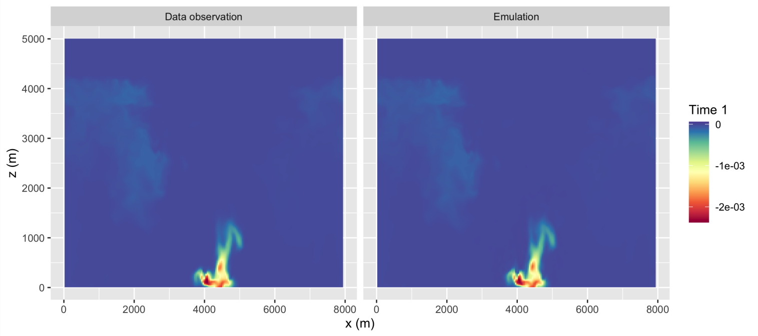


Figure: Hovmöller plots for SST anomalies for time vs. longitude averaged between 1N-1S.

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.



Exploratory Analysis via Quantitative Summaries

(Wikle et al. 2019, Ch. 2; Cressie and Wikle, 2011, Ch. 5)

Exploration of Spatio-Temporal Data:

In addition to visualization, we often wish to explore spatio-temporal data in terms of summaries of its first-order and second-order characteristics. In particular, useful summaries we consider here are

- empirical means
- empirical covariances
- empirical covariograms
- empirical orthogonal functions

Exploration of Spatio-Temporal Data: Empirical Spatial Means

Empirical spatial means: average over time replicates for each location in space.

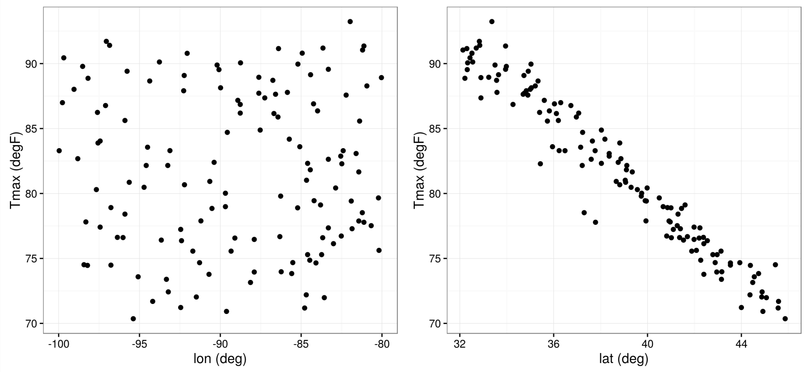


Figure: Spatial mean of the NOAA maximum temperature as a function of station longitude (left panel) and station latitude (right panel).

Exploration of Spatio-Temporal Data: Empirical Temporal Means

We can also average across space and plot the associated time series; this is the empirical temporal mean:

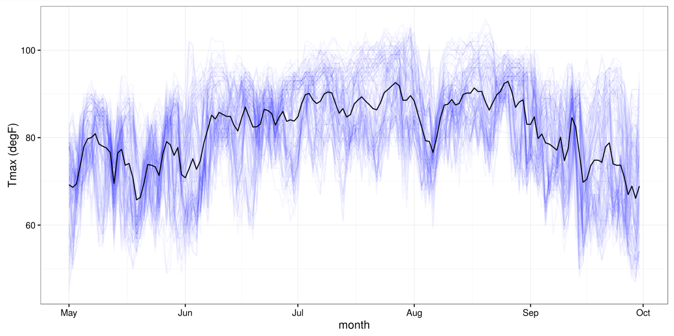


Figure: Raw data from the NOAA maximum temperature dataset (blue lines) and the empirical means $\widehat{\mu}_z(t)$ (black line).

Exploration of Spatio-Temporal Data: Empirical Spatial Covariability

There are several ways to look at second-order structure for spatio-temporal data. For example,

- Empirical lag-τ covariance or correlation between two spatial locations
- Empirical spatio-temporal covariograms: characterizes the covariability in the data as a function of lags in time and space
    - Often assumes isotropy: spatial lag only a funciton of distance (not direction)
    - Allows us to consider separability: spatial covariability independent of temporal covariability

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

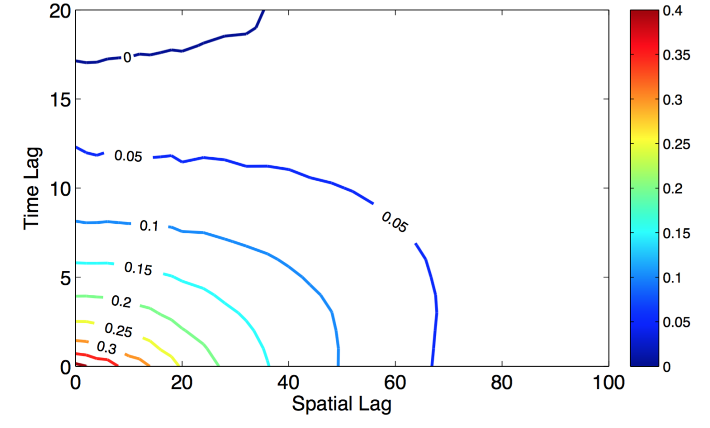


Figure: Empirical spatio-temporal covariogram for SST data; spatial lag (deg), time lag (months).

Exploration of Spatio-Temporal Data: Empirical Orthogonal Functions

Empirical Orthogonal Functions (EOFs) can reveal spatial structures that account for much of the spatio-temporal variability in the data; they can also be used as basis functions for dimensionality reduction in modeling applications as we describe below.

EOFs originated in the meteorology/climatology literature, and in the context of discrete space and time, EOF analysis is the spatio-temporal manifestation of principal component analysis (PCA) in statistics (see Cressie and Wikle (2011, Chap 5) or Wikle et al. (2019, Chap 2) for an extensive overview).

Exploration of Spatio-Temporal Data: EOFs

Recall that in PCA, one estimates a covariance or correlation matrix of the features (traits), averaging over replicates. Then, the symmetric decomposition of that covariance/correlation matrix gives orthogonal eigenvectors that correspond to loadings (weights) for new variables (principal components that are linear combinations of the original features and mutually independent.

- spatial locations are features and time are replicates
- loadings are spatial maps (called EOFs) and the new principal component variables are time series corresponding to the projection of the data onto those loading maps
- Patterns in the EOFs show areas of highest variability and when the associated PC time series is large in magnitude, those areas are most present in the data

Exploration of Spatio-Temporal Data: EOFs

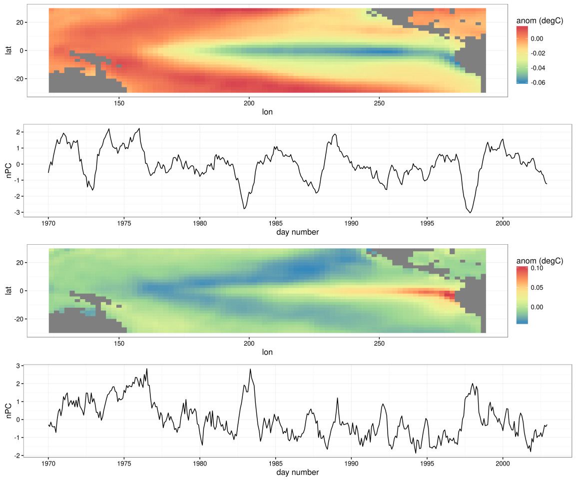


Figure: The first two empirical orthogonal functions and principal-component time series for the SST dataset.

Spatio-Temporal Modeling: Descriptive (Gaussian Processes)

(Wikle et al. 2019, Ch. 4; Cressie and Wikle, 2011, Ch. 6)

Traditionally, there are two approaches for spatio-temporal modeling:

Descriptive

or

Dynamic

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

Notation

- Data: $Z(s;t)$ or $Z(s_i;t)$ or $Z_t(s)$
- Data (vector): $\mathbf{Z}_t$ (vector elements are indexed in space)
- Latent (hidden) Process: $Y(s;t)$ or $Y(j;t)$ or $Y_t(\cdot)$ (may or may not be indexed in space)
- Latent Process (vector): $\mathbf{Y}_t$
- Input (covariate) vector: $\mathbf{x}_t$ (potentially spatially indexed)
- Parameters: $\boldsymbol{\theta}_t$ (may be spatial or temporal or neither)
- Distribution (conditional): $[A|B]$
- Transpose (prime): $\mathbf{Z}'$

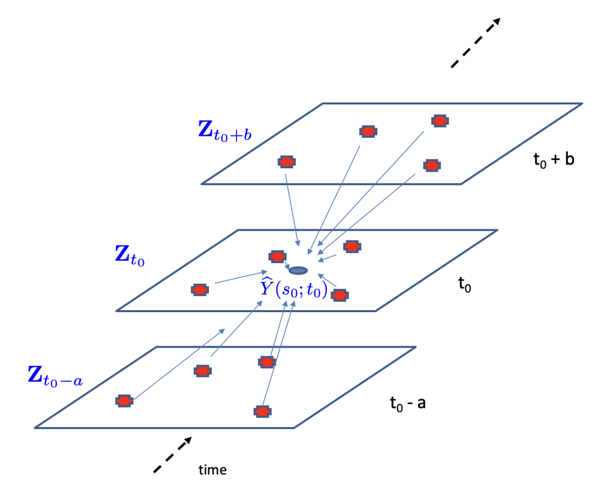
Tobler's (1970) "First Law of Geography" (spatio-temporal extension):

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

(Except when they aren't!)

Descriptive Spatio-Temporal Prediction

Prediction of $Y(s_0; t_0)$ as a weighted average of observations at nearby spatial locations and times.



Descriptive Spatio-Temporal Prediction

A linear spatio-temporal predictor requires that we find the weights, w_{ij} in the linear combination:

$$Y(s_0; t_0) = \sum_{i=1}^{n} \sum_{j=1}^{T} w_{ij} Z(s_i; t_j).$$

These weights typically reflect the first "law" of geography so the closer (s_i, t_j) is to $(s_0; t_0)$, the larger the weight.

This is similar to a kernel-weighted regression smoother, but the "optimal" (in terms of minimizing mean squared error) linear ("kriging") predictor considers spatio-temporal dependence more explicitly.

Descriptive Spatio-Temporal Modeling

We typically consider the two-stage model:

observation = true process + observation error

$$Z(s_i;t) = Y(s_i;t) + \epsilon(s_i;t), \quad \epsilon(s_i;t) \sim iid\ N(0, \sigma_\epsilon^2)$$

true process = trend term + dependent random process

$$\begin{aligned} Y(s;t) &= \mu(s;t) + \eta(s;t) \\ &= \mathbf{x}(s;t)'\boldsymbol{\beta} + \eta(s;t) \end{aligned}$$

where $\eta(s;t)$ is a mean-zero spatio-temporal Gaussian process (GP; fully specified by a covariance function for **any** locations in space and time).

Predictive Distribution

The space-time kriging predictive distribution is given by:

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

where

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

is the generalized least squares estimator.

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

Predictive Distribution

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

- $\mathbf{x}(\mathbf{s}_0; t_0)'\hat{\boldsymbol{\beta}}$: GLS regression estimate

Predictive Distribution

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

- $(\mathbf{z} - \mathbf{X}\hat{\beta})$: residuals at data locations

Predictive Distribution

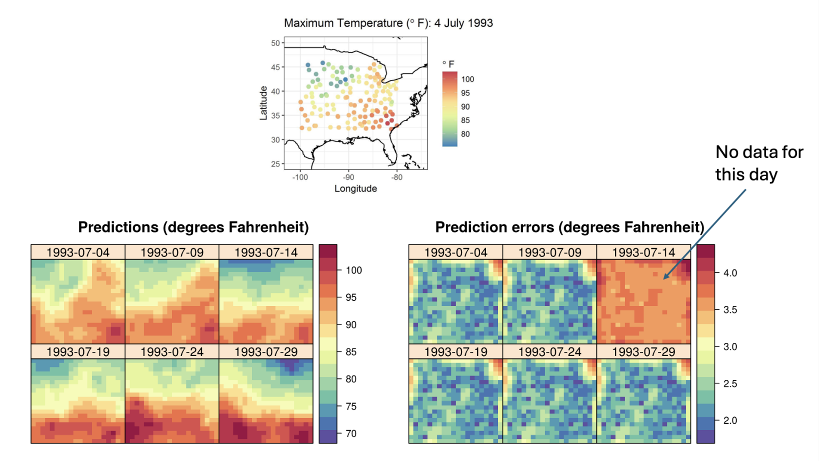
$$Y(\mathbf{s}_0; t_0) \,|\mathbf{z} \;\sim\; Gau\left(\mathbf{x}(\mathbf{s}_0; t_0)'\hat{\beta} + \mathbf{c}_0'\mathbf{C}_z^{-1}(\mathbf{z} - \mathbf{X}\hat{\beta})\,,\; c_{0,0} - \mathbf{c}_0'\mathbf{C}_z^{-1}\mathbf{c}_0\right)$$

- $\mathbf{w}_0' \equiv \mathbf{c}_0'\mathbf{C}_z^{-1}$: weights
- $\mathbf{c}_0' \equiv \mathrm{cov}(Y(\mathbf{s}_0; t_0), \mathbf{z})$, $c_{0,0} \equiv var(Y(\mathbf{s}_0; t_0))$; (note, locations that are more correlated with location $(s_0; t_0)$ get the most weight) [This is why we need GPs!]

So, the GLS regression estimate is corrected by borrowing strength from linear combinations of residuals of GLS fits, with those that are more correlated getting the most weight.

Example: Spatio-Temporal Kriging

Prediction of NOAA midwest US maximum temperature on 20 x 20 grid for 6 times in July 1993 (using the **gstat** R package).



Challenges: Descriptive Spatio-Temporal Modeling

- Covariance functions:
  - Assumed we know (we don't). Must be a valid (non-negative definite) function; we must parameterize it in terms of a few parameters (e.g., Matérn model)
  - Realism of spatio-temporal covariance function: typically assume stationarity (same variance/covariance structure over the entire domain) and often separability (product of space and time covariance functions)
- Computational Complexity: matrix inverse (order n^3 computations)
- Solutions: (approximations)
  - neighborhood-based methods (nearest neighbor Gaussian processes, Vecchia approximations)
  - stochastic PDE methods (INLA)
  - basis function representations w/ random coefficients

Other Types of Models

Gaussian process (GP) spatio-temporal models can also serve as latent processes for other models such as spatio-temporal point processes (e.g., log-Gaussian Cox processes).

GP models can also be used as latent processes for areal (gridded, county, etc.) data as well. But, these types of data are often considered with Markov Random Field (MRF) (Conditional Autoregressive) or Simultaneous Autoregressive (SAR) models. For more details with software examples, see Sahu (2022) or Lawson (2018).

R Packages for Descriptive Spatio-Temporal Modeling

Some R packages to do spatio-temporal descriptive modeling (from the CRAN Task View on SpatioTemporal):

- **gstat**: provides kriging, methods of moments variogram estimation and model fitting for a limited range of spatio-temporal models
- **spTimer**: fit, spatially predict and temporally forecast large amounts of space-time data using Bayesian Gaussian Process (GP) Models, Bayesian Auto-Regressive (AR) Models, and Bayesian Gaussian Predictive Processes (GPP) based AR Models
- **FRK**: spatial/spatio-temporal modelling and prediction with large datasets

Non-Gaussian Extensions

As with linear mixed models, we can easily extend descriptive spatio-temporal models to non-Guassian data (in principle) by simply conditioning on a latent spatio-temporal Gaussian process. Of course, computationally this introduces additional challenges. Some approaches/packages:

- (Mixed) Model-based geostatistics (**GeoR**)
- Laplace approximations
    - Integrated Nested Laplace Approximations (**INLA**)
    - Template Model Builder: **sdmTMB**: spatial and spatiotemporal predictive-process generalized linear mixed effect models using 'TMB', 'INLA', and the SPDE approximation to Gaussian random fields
- Generalized Additive Mixed Models (**mgcv** R package)
- Hierarchical Bayesian Methods (**spBayes**)

Non-Gaussian Data and Latent Gaussian Processes: Example

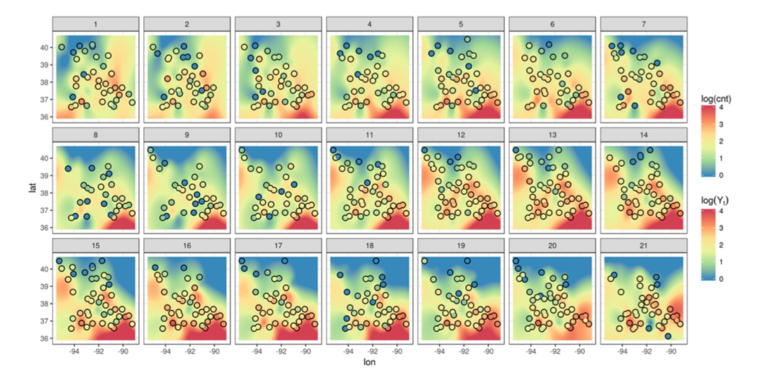


Figure: Log intensity of Carolina Wren BBS counts in Missouri on a grid for t = 1 (1994) to t = 21 (2014). The log of the observed count is shown in circles using the same color scale. This was fit using a Poisson data model and GAM basis functions in the **mgcv** R package.

Dynamic Spatio-Temporal Models (DSTMs)

(Wikle et al. 2019, Ch. 5; Cressie and Wikle, 2011, Ch. 7)

Dynamic Models: Motivation

One of the biggest challenges with the *descriptive* approach to spatio-temporal modeling 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.

Dynamic (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)
- Study of how things change over time
- Pattern of change or growth of a system over time
- Dynamics are due to the interaction of the process components across space and time and/or across scales of variability

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 models):

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

Challenge: specification of the models associated with these component distributions

Generic Hierarchical DSTM

(For a finite set of spatial locations)

Observation Model: $\mathbf{Z}_t = \mathcal{H}(\mathbf{Y}_t; \boldsymbol{\theta}_h; \boldsymbol{\epsilon}_t)$

Gaussian or non-Gaussian

Spatial data vector indexed in time

Mapping function

Spatio-temporal process vector of interest (hidden)

Parameters

Errors

Process Model: $\mathbf{Y}_t = \mathcal{M}(\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots; \boldsymbol{\theta}_m; \boldsymbol{\eta}_t)$

Linear or nonlinear

Transition function with Markovian dependence structure (conditional independence); **controls dynamics**

Noise

Parameter/Error Models: $[\boldsymbol{\theta}_h][\boldsymbol{\theta}_m][\boldsymbol{\epsilon}_t][\boldsymbol{\eta}_t]$

Critically, these may also be spatial or temporal stochastic processes as well!

Example: Basic Gaussian Hierarchical Linear DSTM

Data:

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

Process:

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

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.

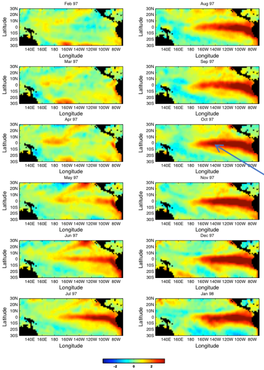
Parameterization of the dynamic operator $\mathbf{M}$ can be facilitated by mechanistic motivation (e.g., Wikle and Hooten, 2010), basis function representation, and/or regularization.

DSTM Basis Example: Long-Lead SST Forecasting

Long-lead (7 mo) Prediction of Tropical Pacific Sea Surface Temperature (SST)

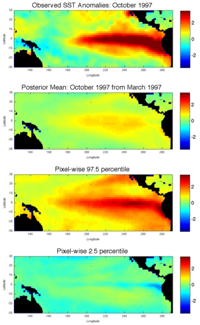
SST Feb 1997-Jan 1998

3/97



El Nino

Model Prediction for 10/97 starting from 3/97



mean prediction

Upper 97.5% Uncertainty

Lower 2.5% Uncertainty

R Packages for DSTMs

- **IDE**
- **spTimer**
- **spBayes**
- **spTDyn**
- **CarBayesST**

More Complex Models

Non-linear and highly complex spatio-temporal processes can be addressed through various statistical (e.g., generalize quadratic nonlinearity; Wikle and Hooten, 2010) and neural network approaches (e.g., echo state networks).

In addition, estimation of complex simulation-based models (agent-based models, cellular automata, point processes, spatial extremes) have been shown to be facilitated significantly by neural estimation.

There is also increased interest in learning the dynamics that are present in the data.

These are fast growing areas of spatial and spatio-temporal statistics (see the recent review by Wikle and Zammit-Mangion (2023) and North et al. (2023)).

References

Cressie, N. and C.K. Wikle, 2011: *Statistics for Spatio-Temporal Data*. John Wiley & Sons.

Lawson, A.B., 2018: *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC press.

North, J.S., Wikle, C.K. and E.M. Schliep, 2023: A Review of data-driven discovery for dynamic systems. *International Statistical Review*. 91, 464-492

Sahu, S., 2022: *Bayesian modeling of spatio-temporal data with R*. CRC Press.

Wikle, C.K. and M.B. Hooten, 2010: A general science-based framework for spatio-temporal dynamical models. *Test*, 19, 417-451.

Wikle, C.K., Zammit-Mangion, A., and Cressie, N., 2019: *Spatio-Temporal Statistics with R*. Chapman & Hall/CRC.

Wikle, C.K. and A. Zammit-Mangion, 2023: Statistical deep learning for spatial and spatiotemporal data. *Annual Review Statistics and its Application*. 10:247-270