Statistics for spatio-temporal data

Reading group

Part I

Lucia Paci
lucia.paci@unicatt.it

Università Cattolica del Sacro Cuore

May 2022

Course outline[1]

1. Introduction to spatial and spatio-temporal data
   - what are spatial data? why spatial modeling? types of spatial data
2. Point-level modeling
   - spatial regression, Gaussian processes, Kriging, large datasets and spatio-temporal models
3. Areal data models
   - spatial association, modeling areal data and spatio-temoral models
4. Introduction to point patterns
   - modeling point pattern and spatio-temporal analysis

[1] Many of these slides were excerpted from Alan Gelfand's material.

Reference books

- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman and Hall/CRC, 2nd edition
- Cressie, N. and Wikle, C. (2011). *Statistics for Spatio-Temporal Data*. New York: Wiley.
- Finkenstädt B., Held L., Isham V. (2007) *Statistical methods for spatio-temporal systems*, Chapman and Hall/CRC.
- Gelfand, A.E., Diggle, P., Guttorp, P. and Fuentes, M. (2010) *Handbook of Spatial Statistics*, Boca Raton, FL: CRC Press.
- Wikle C. K., Zammit-Mangion A., Cressie N. (2019) *Spatio-Temporal Statistics with R*, Chapman and Hall/CRC.

Outline

What are spatial data?

What are spatial data?

- Any data equipped with spatial information

What are spatial data?

- Any data equipped with spatial information
- Usually, geographical information

What are spatial data?

- Any data equipped with spatial information
- Usually, geographical information
- Spatial data are:
  - usually response variable and many important predictors (highly multivariate)
  - are often presented as maps

What are spatial data?

- Any data equipped with spatial information
- Usually, geographical information
- Spatial data are:
  - usually response variable and many important predictors (highly multivariate)
  - are often presented as maps
- Common sources of spatial data: climatology, forestry, ecology, environmental health, epidemiology, economical and social sciences, real-estate marketing, etc...

Spatial data examples

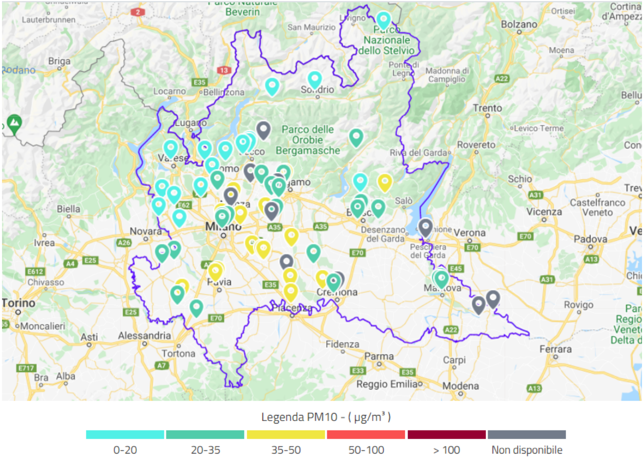


Figure: PM_{10} collected at the monitoring network over Lombardia on September 26, 2021. Source: ARPA Lombardia.

Spatial data examples

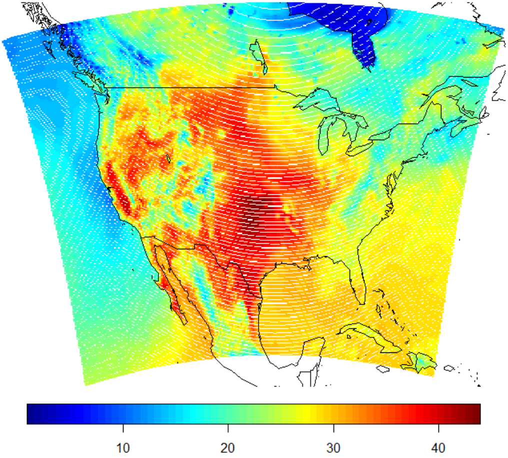


Figure: Temperature surface predicted by a computer model over US (Paci et al., 2015).

Spatial data examples

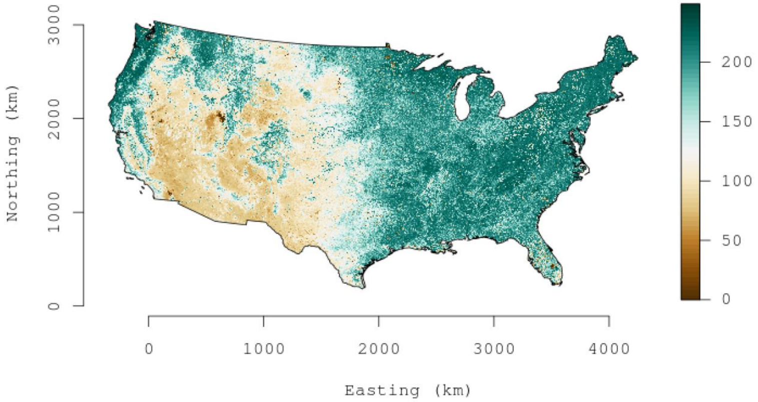


Figure: Normalized Difference Vegetation Index (NDVI) image from the MODerate-resolution Imaging Spectroradiometer (MODIS) sensor (Datta et al., 2016).

Spatial data examples

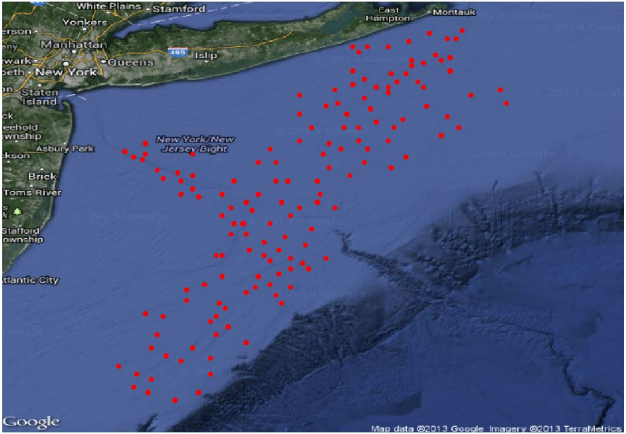


Figure: Locations of scallop counts off the New Jersey and Long Island coastline (Banerjee et al., 2014).

Spatial data examples

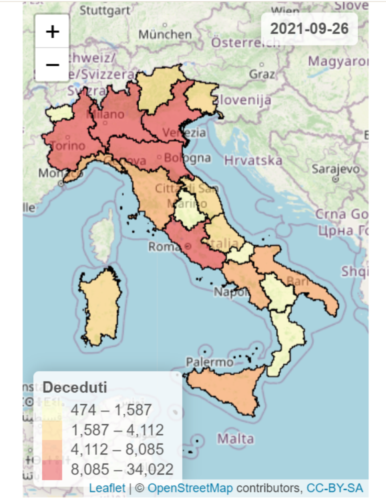


Figure: Covid-19 deaths from March 1, 2020 to September 26, 2021 in Italy. Source: StatGroup-19.

Spatial data examples

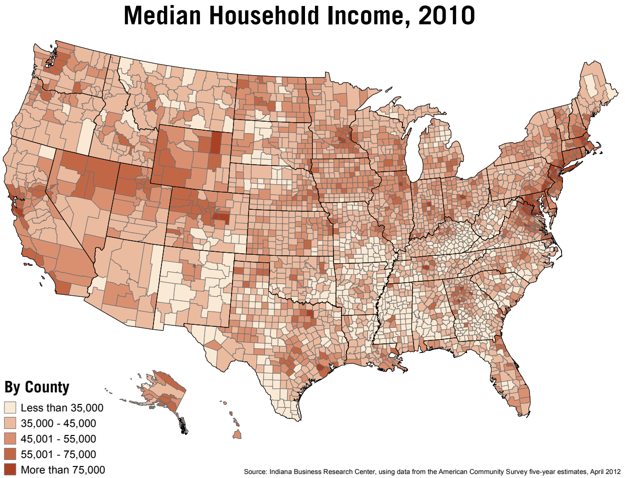


Figure: Median household income by US county in 2010. Source: Indiana's Public Data Utility.

Spatial data examples

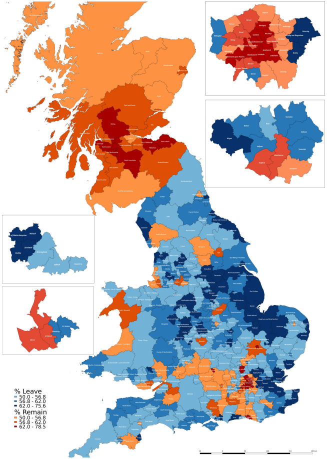


Figure: Brexit results by UK county. Source: Wikipedia.

Spatial data examples

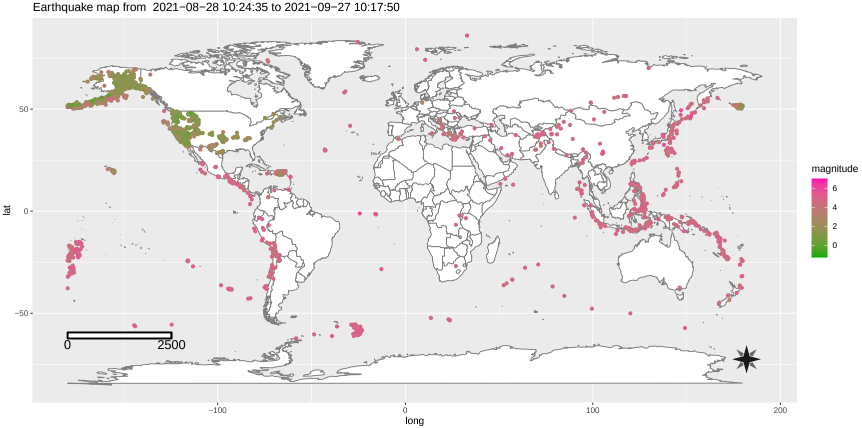


Figure: Earthquakes occurred from August 28 to September 27, 2021.

Spatial data examples

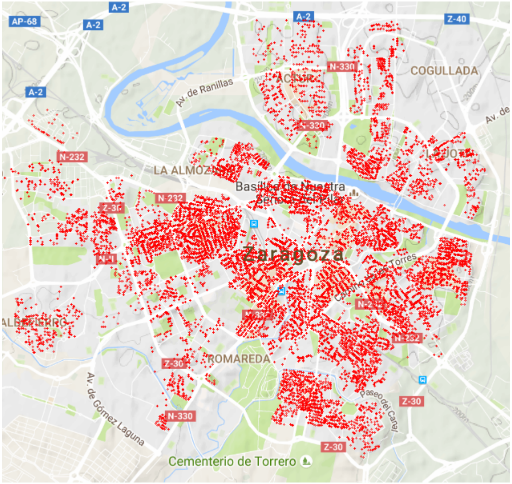


Figure: Property sales in Zaragoza during the period 2006-2014, consisting of $\sim 30,000$ transactions (Paci et al., 2017).

What are spatial data?

- Any data equipped with spatial information
- Usually, geographical information
- Spatial data are:
  - usually response variable and many important predictors (highly multivariate)
  - are often presented as maps
- Common sources of spatial data: climatology, forestry, ecology, environmental health, epidemiology, economics, social sciences, real-estate marketing, etc...

What are spatial data?

- Any data equipped with spatial information
- Usually, geographical information
- Spatial data are:
  - usually response variable and many important predictors (highly multivariate)
  - are often presented as maps
- Common sources of spatial data: climatology, forestry, ecology, environmental health, epidemiology, economics, social sciences, real-estate marketing, etc...
- Other examples where the space need not be the space on earth:

Further spatial data examples

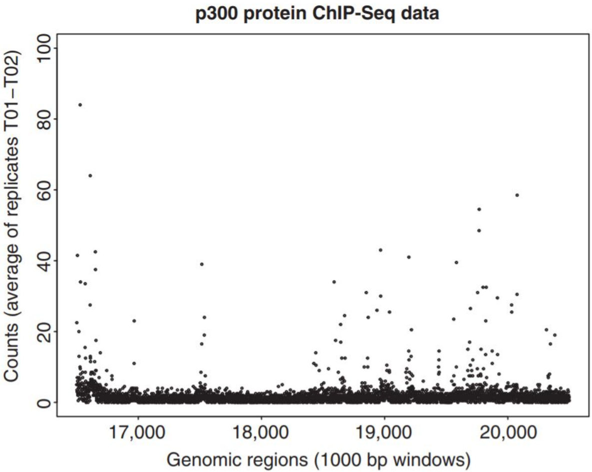


Figure: Chromatin Immunoprecipitation Sequencing (ChIP-Seq) data (Ranciati et al., 2017).

Further spatial data examples

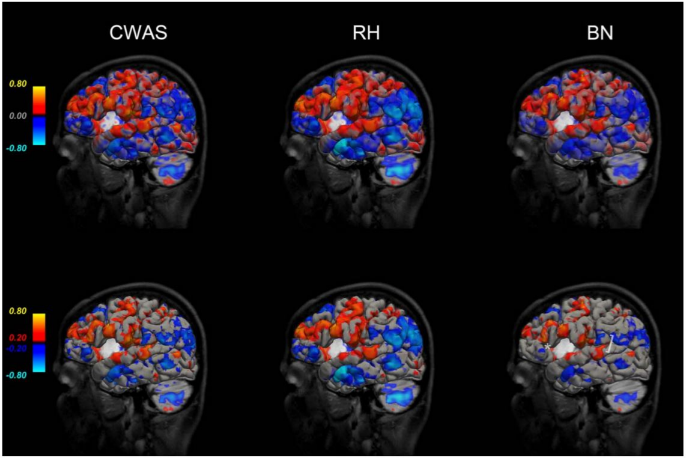


Figure: Pre-surgical functional magnetic resonance imaging (fMRI) data (Liu et al., 2016).

What are spatial data?

- Any data equipped with spatial (geographical) information
- Common sources of spatial data: climatology, forestry, ecology, environmental health, epidemiology, economics, social sciences, real-estate marketing, etc...
  - usually response variable and many important predictors (highly multivariate)
  - are often presented as maps
- Other examples where the space need not be the space on earth:
  - Genetics (position along a chromosome)
  - Neuroimaging (data for each voxel or region of the brain)

What are spatial data?

- Any data equipped with spatial (geographical) information
- Common sources of spatial data: climatology, forestry, ecology, environmental health, epidemiology, economics, social sciences, real-estate marketing, etc...
  - usually response variable and many important predictors (highly multivariate)
  - are often presented as maps
- Other examples where the space need not be the space on earth:
  - Genetics (position along a chromosome)
  - Neuroimaging (data for each voxel or region of the brain)

$\Rightarrow$ Motivate data analysis and statistical modeling for complex spatial and spatio-temporal data sets.

What are spatio-temporal data?

Example

- Data collected across space and over time

What are spatio-temporal data?
Example

- Data collected across space and over time

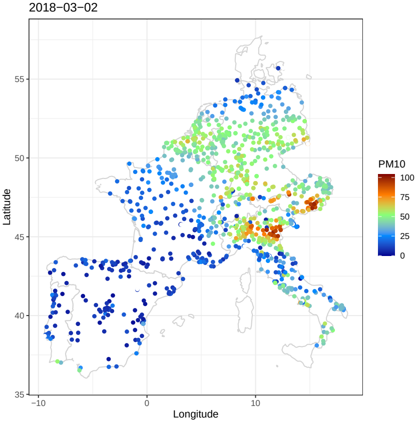


Figure: PM_{10} observed across Europe in two consecutive days.

What are spatio-temporal data?
Example

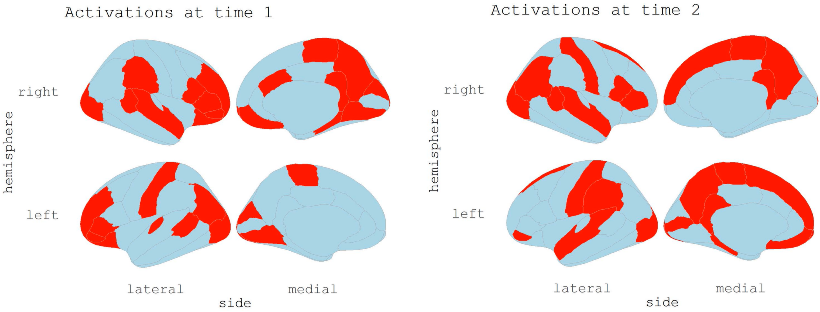


Figure: Spontaneous activations of brain regions over time, detected from resting state fMRI data (Gasperoni et al., 2021).

What are spatio-temporal data?
Example

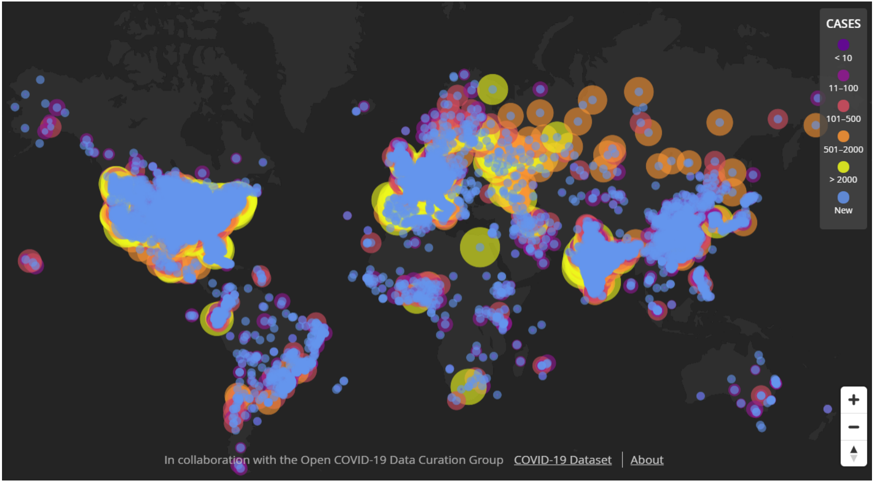


Figure: COVID-19 confirmed cases worldwide. Source: `www.healthmap.org`.

Why spatial models are *Statistical* models?

- Why do we need to consider randomness and uncertainty?

Why spatial models are *Statistical* models?

- Why do we need to consider randomness and uncertainty?
- The primary reason comes from the **uncertainty** resulting from incomplete knowledge of the science and of the mechanisms driving a spatio-temporal phenomenon.

Why spatial models are *Statistical* models?

- Why do we need to consider randomness and uncertainty?
- The primary reason comes from the **uncertainty** resulting from incomplete knowledge of the science and of the mechanisms driving a spatio-temporal phenomenon.
- Observations may have large gaps between them (in space and in time), they are observed with error, our understanding of the physical mechanisms is incomplete, we have limited knowledge about model parameters, and so on...

Why spatial models are *Statistical* models?

- Why do we need to consider randomness and uncertainty?
- The primary reason comes from the **uncertainty** resulting from incomplete knowledge of the science and of the mechanisms driving a spatio-temporal phenomenon.
- Observations may have large gaps between them (in space and in time), they are observed with error, our understanding of the physical mechanisms is incomplete, we have limited knowledge about model parameters, and so on...
- Answers to the problems come as **inference**, i.e., estimates or predictions along with a quantification of their uncertainties.

The goals of spatial statistics

1. Carrying out statistical **inference** tasks, such as:

The goals of spatial statistics

1. Carrying out statistical **inference** tasks, such as:
   - modeling of trends and correlation structures

The goals of spatial statistics

1. Carrying out statistical **inference** tasks, such as:
   - modeling of trends and correlation structures
   - estimation of underlying model parameters

The goals of spatial statistics

1. Carrying out statistical **inference** tasks, such as:
   - modeling of trends and correlation structures
   - estimation of underlying model parameters
   - hypothesis testing (or comparison of competing models)

The goals of spatial statistics

1. Carrying out statistical **inference** tasks, such as:
   - modeling of trends and correlation structures
   - estimation of underlying model parameters
   - hypothesis testing (or comparison of competing models)
   - regression specifications to explain spatial response

The goals of spatial statistics

1. Carrying out statistical **inference** tasks, such as:
    - modeling of trends and correlation structures
    - estimation of underlying model parameters
    - hypothesis testing (or comparison of competing models)
    - regression specifications to explain spatial response

The goals of spatial statistics

1. Carrying out statistical **inference** tasks, such as:
   - modeling of trends and correlation structures
   - estimation of underlying model parameters
   - hypothesis testing (or comparison of competing models)
   - regression specifications to explain spatial response
2. **Predict** at unobserved locations (smoothing, interpolation) and/or times (forecasting)

The goals of spatial statistics

1. Carrying out statistical **inference** tasks, such as:
   - modeling of trends and correlation structures
   - estimation of underlying model parameters
   - hypothesis testing (or comparison of competing models)
   - regression specifications to explain spatial response
2. **Predict** at unobserved locations (smoothing, interpolation) and/or times (forecasting)
3. More specific goals might include data assimilation, computer-model emulation, and design of spatio-temporal monitoring networks

Spatial data analysis

Statistical inference for spatial processes is often challenging, but is necessary when we try to draw conclusions about questions of interest, such as:

- Does the spatial patterning of disease incidences give rise to the conclusion that they are clustered, and if so, are the clusters found related to factors such as age, relative poverty, or pollution sources?
- Given a number of observed soil samples, which part of a study area is polluted?
- Given scattered air quality measurements, how many people are exposed to high levels of black smoke or particulate matter (e.g. PM_{10}) and where do they live?
- Do governments tend to compare their policies with those of their neighbors, or do they behave independently?

A bit of history

Spatial statistics has an unusual history as a field within the discipline of statistics.

A bit of history

Spatial statistics has an unusual history as a field within the discipline of statistics.

- The stochastic process theory that underlies much of the field was developed within the mathematical sciences by probabilists,

A bit of history

Spatial statistics has an unusual history as a field within the discipline of statistics.

- The stochastic process theory that underlies much of the field was developed within the mathematical sciences by probabilists,
- Early on, much of the statistical methodology was developed quite independently and grew primarily from the different areas of application, including:
  - mining engineering, which led to the development of *geostatistics* by Matheron and colleagues;
  - agriculture with spatial considerations informed by the thinking of Fisher on randomization and blocking;
  - forestry, which was the setting for the seminal PhD thesis of Matèrn.

A bit of history

Spatial statistics has an unusual history as a field within the discipline of statistics.

- The stochastic process theory that underlies much of the field was developed within the mathematical sciences by probabilists,
- Early on, much of the statistical methodology was developed quite independently and grew primarily from the different areas of application, including:
  - mining engineering, which led to the development of *geostatistics* by Matheron and colleagues;
  - agriculture with spatial considerations informed by the thinking of Fisher on randomization and blocking;
  - forestry, which was the setting for the seminal PhD thesis of Matèrn.
- For many years spatial statistics labored on the fringe of mainstream statistics. However, the past 30 years has seen an explosion of interest in space and space-time problems.

A bit of history

Spatial statistics has an unusual history as a field within the discipline of statistics.

- The stochastic process theory that underlies much of the field was developed within the mathematical sciences by probabilists,
- Early on, much of the statistical methodology was developed quite independently and grew primarily from the different areas of application, including:
  - mining engineering, which led to the development of *geostatistics* by Matheron and colleagues;
  - agriculture with spatial considerations informed by the thinking of Fisher on randomization and blocking;
  - forestry, which was the setting for the seminal PhD thesis of Matèrn.
- For many years spatial statistics labored on the fringe of mainstream statistics. However, the past 30 years has seen an explosion of interest in space and space-time problems.
- This has been largely fueled by the increased availability of inexpensive, high-speed computing that has:
  - enabled the collection of large spatial and spatio-temporal datasets across many fields,
  - facilitated the widespread usage of sophisticated geographic information systems (GIS) software to create attractive displays,
  - has endowed the ability to investigate (fit and infer under) challenging, evermore appropriate and realistic models.

Key ingredients

- Observations are equipped with spatial information
- Observations and/or locations are random variables

Key ingredients

- Observations are equipped with spatial information
- Observations and/or locations are random variables
- Variables of interest:
  - univariate, multivariate
  - continuous, categorical

Key ingredients

- Observations are equipped with spatial information
- Observations and/or locations are random variables
- Variables of interest:
  - univariate, multivariate
  - continuous, categorical
- Locations:
  - points
  - regions (regular or irregular)

Key ingredients

- Observations are equipped with spatial information
- Observations and/or locations are random variables
- Variables of interest:
  - univariate, multivariate
  - continuous, categorical
- Locations:
  - points
  - regions (regular or irregular)
- Location generation process
  - known or unknown
  - fixed or random

Notation

- $\mathcal{S}$: the space of all possible locations (we will consider $\mathbb{R}^2$)
- $D \subseteq \mathcal{S}$: fixed subset of $\mathcal{S}$ (study region)
- $\mathbf{s} \in D$: given point in D
- $\mathbf{s}_1, \ldots, \mathbf{s}_n$: locations of n observations
- $y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)$: observations of the variable of interest

Types of spatial data

There are three broad categories:

1. Point-referenced data (or geostatistical data)
2. Areal data
3. Point pattern data

Types of spatial data
Point-referenced data

- Each observation is associated with a fixed known location (point)
- Data represents a sample from a ***continuous spatial domain***
- The variable of interest does exist in all points of the domain but it is observed only at a finite set of points
- Also referred to as geocoded or geostatistical data

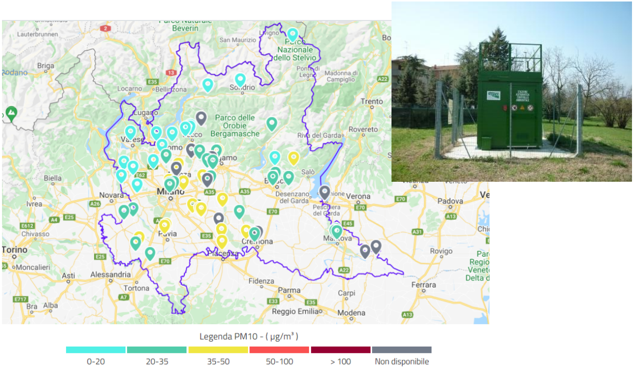


Figure: PM_{10} collected at the monitoring network over Lombardia on September 26, 2021. Source: ARPA Lombardia.

Types of spatial data
Areal data

- Each observation is associated with a areal unit with well-defined boundaries of ***regular*** or ***irregular*** shape, like pixels, regions, county etc...
- The variable of interest is observed only at a finite set of regions
- Usually a result of aggregating point level data
- Also referred to as lattice data

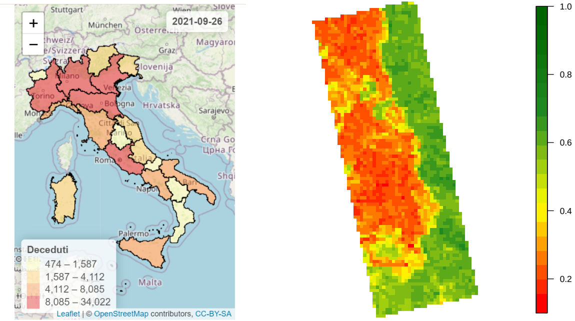


Figure: Left panel: Covid-19 deaths from March 1, 2020 to September 26, 2021 in Italy; source: StatGroup-19. Right panel: NDVI values over the Zuera region, Spain in 1986; source Landsat.

Types of spatial data

Point pattern data

- The ***locations*** are viewed as ***random***
- Need not have variables at locations, just the pattern of points
- Interest in the pattern of occurrences of an event like disease incidence, species distribution, crimes etc.
- Sometimes, also a response variable (mark) is associated with the random locations

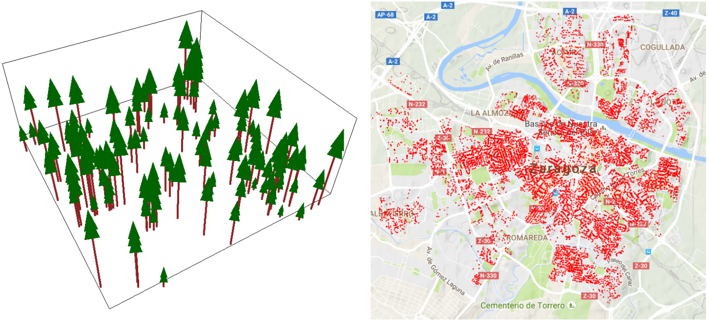


Figure: Left panel: Pines in a Finnish forest. Right panel: property sales in Zaragoza during the period 2006-2014 (Paci et al., 2017).

Outline

Point-referenced data

- Spatial data collected at locations referenced by coordinates (e.g., lat-long, Easting-Northing)

Point-referenced data

- Spatial data collected at locations referenced by coordinates (e.g., lat-long, Easting-Northing)
- Data from a spatial process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$, D is a fixed subset of $\mathbb{R}^r$

Point-referenced data

- Spatial data collected at locations referenced by coordinates (e.g., lat-long, Easting-Northing)
- Data from a spatial process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$, D is a fixed subset of $\mathbb{R}^r$
- Example: $Y(\mathbf{s})$ is a pollutant level at site $\mathbf{s}$

Point-referenced data

- Spatial data collected at locations referenced by coordinates (e.g., lat-long, Easting-Northing)
- Data from a spatial process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$, D is a fixed subset of $\mathbb{R}^r$
- Example: $Y(\mathbf{s})$ is a pollutant level at site $\mathbf{s}$
- Note: time series follows this approach with $r = 1$; here we will usually have $r = 2$ or $r = 3$

Point-referenced data

- Spatial data collected at locations referenced by coordinates (e.g., lat-long, Easting-Northing)
- Data from a spatial process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$, D is a fixed subset of $\mathbb{R}^r$
- Example: $Y(\mathbf{s})$ is a pollutant level at site $\mathbf{s}$
- Note: time series follows this approach with $r = 1$; here we will usually have $r = 2$ or $r = 3$
- In principle: Pollutant level exists at all possible sites in the domain

Point-referenced data

- Spatial data collected at locations referenced by coordinates (e.g., lat-long, Easting-Northing)
- Data from a spatial process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$, D is a fixed subset of $\mathbb{R}^r$
- Example: $Y(\mathbf{s})$ is a pollutant level at site $\mathbf{s}$
- Note: time series follows this approach with $r = 1$; here we will usually have $r = 2$ or $r = 3$
- In principle: Pollutant level exists at all possible sites in the domain
- In practice: Data will be a partial realization of the spatial process, i.e., $Y(\mathbf{s})$ is observed ONLY at fixed locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$

Point-referenced data

- Spatial data collected at locations referenced by coordinates (e.g., lat-long, Easting-Northing)
- Data from a spatial process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$, D is a fixed subset of $\mathbb{R}^r$
- Example: $Y(\mathbf{s})$ is a pollutant level at site $\mathbf{s}$
- Note: time series follows this approach with $r = 1$; here we will usually have $r = 2$ or $r = 3$
- In principle: Pollutant level exists at all possible sites in the domain
- In practice: Data will be a partial realization of the spatial process, i.e., $Y(\mathbf{s})$ is observed ONLY at fixed locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$
- Goal: **inference** about the process $Y(\mathbf{s})$ and **predictions** at new locations

Point-referenced data

- Spatial data collected at locations referenced by coordinates (e.g., lat-long, Easting-Northing)
- Data from a spatial process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$, D is a fixed subset of $\mathbb{R}^r$
- Example: $Y(\mathbf{s})$ is a pollutant level at site $\mathbf{s}$
- Note: time series follows this approach with $r = 1$; here we will usually have $r = 2$ or $r = 3$
- In principle: Pollutant level exists at all possible sites in the domain
- In practice: Data will be a partial realization of the spatial process, i.e., $Y(\mathbf{s})$ is observed ONLY at fixed locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$
- Goal: **inference** about the process $Y(\mathbf{s})$ and **predictions** at new locations
- We can learn about the entire surface $Y(\mathbf{s})$ for any $\mathbf{s}$ based on this data

Point-referenced data

- Spatial data collected at locations referenced by coordinates (e.g., lat-long, Easting-Northing)
- Data from a spatial process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$, D is a fixed subset of $\mathbb{R}^r$
- Example: $Y(\mathbf{s})$ is a pollutant level at site $\mathbf{s}$
- Note: time series follows this approach with $r = 1$; here we will usually have $r = 2$ or $r = 3$
- In principle: Pollutant level exists at all possible sites in the domain
- In practice: Data will be a partial realization of the spatial process, i.e., $Y(\mathbf{s})$ is observed ONLY at fixed locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$
- Goal: **inference** about the process $Y(\mathbf{s})$ and **predictions** at new locations
- We can learn about the entire surface $Y(\mathbf{s})$ for any $\mathbf{s}$ based on this data
- To achieve this: exploiting the **structured dependence**

Plotting the data

$$Y(\mathbf{s}_i), i = 1, \ldots, n = 213$$

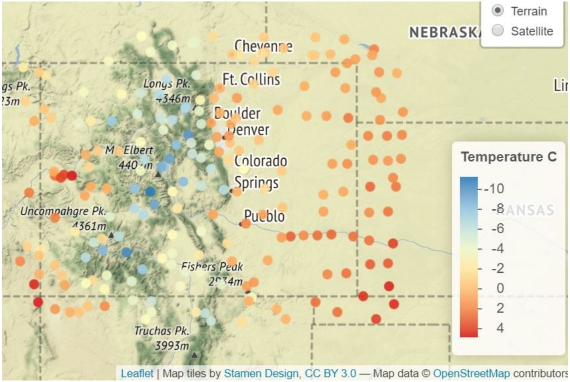


Figure: Temperature data collected at 213 monitoring sites in Colorado (US).

Plotting the data

$$Y(\mathbf{s}), \mathbf{s} \in D$$

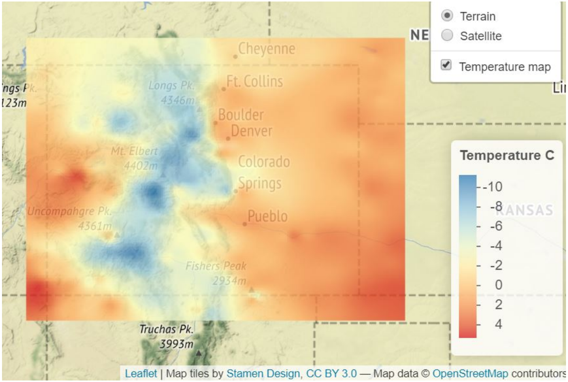


Figure: Interpolated map; **surface plots** of the data often helps to understand spatial patterns.

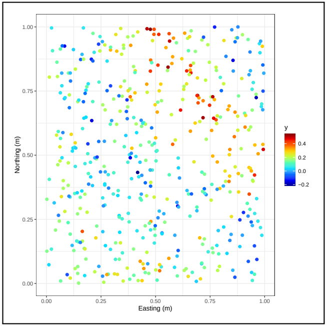
General set up

- Data observed at n fixed spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$

General set up

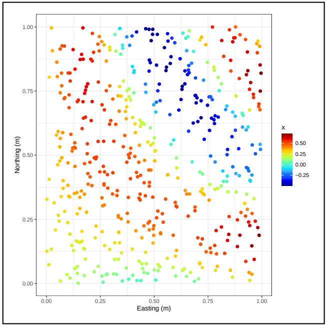
- Data observed at n fixed spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$
- At each location $\mathbf{s}_i$ we observe the response $y(\mathbf{s}_i)$ and a $p \times 1$ vector of covariates $\mathbf{x}(\mathbf{s}_i)$

Dataset 1



$y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)$

Covariate

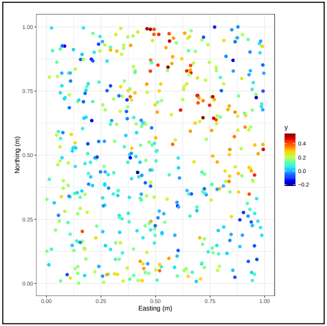


$x(\mathbf{s}_1), \ldots, x(\mathbf{s}_n)$

General set up

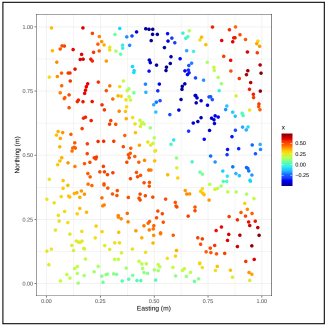
- Data observed at n fixed spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$
- At each location $\mathbf{s}_i$ we observe the response $y(\mathbf{s}_i)$ and a $p \times 1$ vector of covariates $\mathbf{x}(\mathbf{s}_i)$

Dataset 1



$y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)$

Covariate



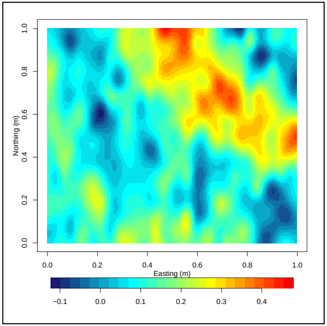
$x(\mathbf{s}_1), \ldots, x(\mathbf{s}_n)$

- **Goals**: Identify association between Y and X and predict $Y(\mathbf{s})$ at any arbitrary site $\mathbf{s}$

General set up

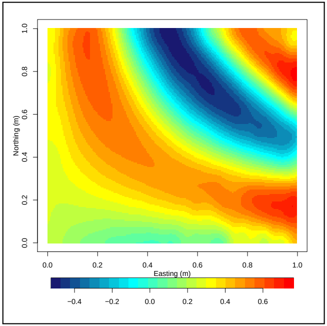
- Data observed at n fixed spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$
- At each location $\mathbf{s}_i$ we observe the response $y(\mathbf{s}_i)$ and a $p \times 1$ vector of covariates $\mathbf{x}(\mathbf{s}_i)$

Dataset 1



$y(\mathbf{s}), \mathbf{s} \in D$

Covariate



$x(\mathbf{s}), \mathbf{s} \in D$

- **Goals**: Identify association between Y and X and predict $Y(\mathbf{s})$ at any arbitrary site $\mathbf{s}$

What's so special about spatial?

- Data observed at n fixed spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$
- At each site $\mathbf{s}_i$ we observe the response $y(\mathbf{s}_i)$ and a $p \times 1$ vector of covariates $\mathbf{x}(\mathbf{s}_i)$
- **Goals**: Identify association between Y and $\mathbf{X}$ and predict $Y(\mathbf{s})$ at any arbitrary site $\mathbf{s}$

What's so special about spatial?

- Data observed at n fixed spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$
- At each site $\mathbf{s}_i$ we observe the response $y(\mathbf{s}_i)$ and a $p \times 1$ vector of covariates $\mathbf{x}(\mathbf{s}_i)$
- **Goals**: Identify association between Y and $\mathbf{X}$ and predict $Y(\mathbf{s})$ at any arbitrary site $\mathbf{s}$
- Linear regression model:

$$y(\mathbf{s}_i) = \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} + \varepsilon(\mathbf{s}_i), \quad \varepsilon(\mathbf{s}_i) \sim \text{iid N}(0, \sigma^2)$$

- Although the data is spatial, this is an ordinary linear regression model

What's so special about spatial?

- Data observed at n fixed spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$
- At each site $\mathbf{s}_i$ we observe the response $y(\mathbf{s}_i)$ and a $p \times 1$ vector of covariates $\mathbf{x}(\mathbf{s}_i)$
- **Goals**: Identify association between Y and $\mathbf{X}$ and predict $Y(\mathbf{s})$ at any arbitrary site $\mathbf{s}$
- Linear regression model:

$$y(\mathbf{s}_i) = \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} + \varepsilon(\mathbf{s}_i), \quad \varepsilon(\mathbf{s}_i) \sim \text{iid N}(0, \sigma^2)$$

- Although the data is spatial, this is an ordinary linear regression model
- $\mathbf{y} = (y(\mathbf{s}_1), y(\mathbf{s}_2), \ldots, y(\mathbf{s}_n))'$; $\mathbf{X} = (\mathbf{x}(s_1)', \mathbf{x}(s_2)', \ldots, \mathbf{x}(\mathbf{s}_n)')'$
- **Inference** - **OLS**: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
- **Prediction** at new location $\mathbf{s}_0$: $\hat{y}(\mathbf{s}_0) = \mathbf{x}(s_0)'\hat{\boldsymbol{\beta}}$

Residual plot
Dataset 1

- Surface plots of the residuals $(y(\mathbf{s}) - \hat{y}(\mathbf{s}))$ help to identify any spatial patterns left unexplained by the covariates

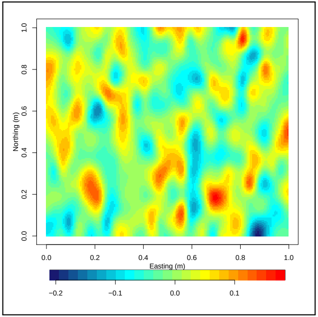


Figure: Residual plot of Dataset 1 after linear regression on $X(\mathbf{s})$.

Residual plot
Dataset 1

- Surface plots of the residuals $(y(\mathbf{s}) - \hat{y}(\mathbf{s}))$ help to identify any spatial patterns left unexplained by the covariates

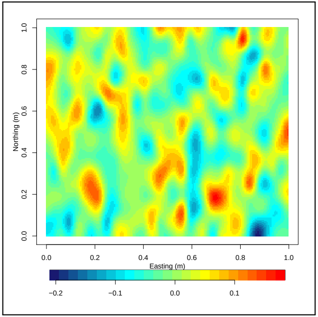
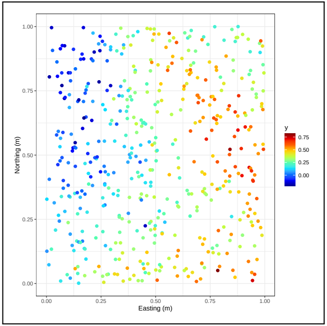


Figure: Residual plot of Dataset 1 after linear regression on $X(\mathbf{s})$.

- No evident spatial pattern in plot of the residuals
- The covariate $X(\mathbf{s})$ seems to explain all the spatial variation in $Y(\mathbf{s})$

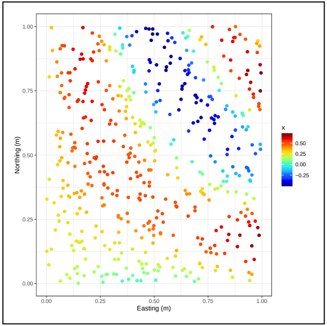
One more dataset

Dataset 2



$y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n)$

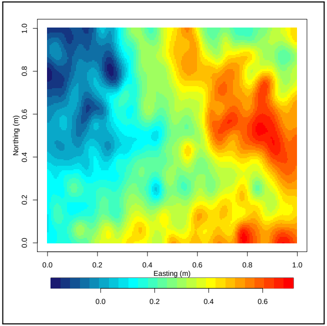
Same covariate



$x(\mathbf{s}_1), \ldots, x(\mathbf{s}_n)$

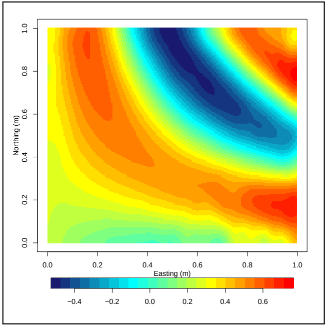
One more dataset

Dataset 2



$y(\mathbf{s}), \mathbf{s} \in D$

Same covariate



$x(\mathbf{s}), \mathbf{s} \in D$

Residual plot

Dataset 2

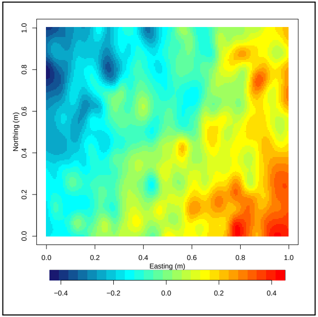


Figure: Residual plot of Dataset 2 after linear regression on $X(\mathbf{s})$.

Residual plot
Dataset 2

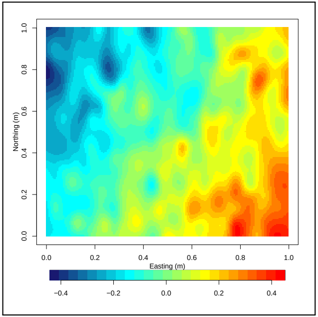


Figure: Residual plot of Dataset 2 after linear regression on $X(\mathbf{s})$.

- Strong residual spatial pattern in Dataset 2
- The covariate $X(\mathbf{s})$ does not explain all the spatial variation in $Y(\mathbf{s})$
- Simple regression seems to be not enough

EDA

- How to do more formal Exploratory Data Analysis (EDA) to identify spatial pattern?

EDA

- How to do more formal Exploratory Data Analysis (EDA) to identify spatial pattern?

Tobler's first law of geography

"Everything is related to everything else, but **near things** are **more related** than distant things." - Waldo Tobler

EDA

- How to do more formal Exploratory Data Analysis (EDA) to identify spatial pattern?

Tobler's first law of geography

"Everything is related to everything else, but **near things** are **more related** than distant things." - Waldo Tobler

- We expect $Y(\mathbf{s})$ and $Y(\mathbf{s}+\mathbf{h})$ to be very similar at short distance, i.e., $(Y(\mathbf{s}) - Y(\mathbf{s}+\mathbf{h}))^2 \to 0$ as $||\mathbf{h}|| \to 0$
- $(Y(\mathbf{s}) - Y(\mathbf{s}+\mathbf{h}))^2$ increasing with $||\mathbf{h}||$ increases $\Rightarrow$ spatial correlation

EDA

- How to do more formal Exploratory Data Analysis (EDA) to identify spatial pattern?

Tobler's first law of geography

"Everything is related to everything else, but **near things** are **more related** than distant things." - Waldo Tobler

- We expect $Y(\mathbf{s})$ and $Y(\mathbf{s}+\mathbf{h})$ to be very similar at short distance, i.e., $(Y(\mathbf{s}) - Y(\mathbf{s}+\mathbf{h}))^2 \to 0$ as $||\mathbf{h}|| \to 0$
- $(Y(\mathbf{s}) - Y(\mathbf{s}+\mathbf{h}))^2$ increasing with $||\mathbf{h}||$ increases $\Rightarrow$ spatial correlation
- Can this be formalized to identify spatial pattern?

Empirical variogram

- $N(t)$ is the set of pairs of sites such that $||\mathbf{s}_i - \mathbf{s}_j|| = t$
- $|N(t)|$ is the number of pairs in this set

Empirical variogram

- $N(t)$ is the set of pairs of sites such that $||\mathbf{s}_i - \mathbf{s}_j|| = t$
- $|N(t)|$ is the number of pairs in this set
- Empirical semivariogram:

$$\hat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(t)} \left(Y(\mathbf{s}_i) - Y(\mathbf{s}_j)\right)^2$$

Empirical variogram

- $N(t)$ is the set of pairs of sites such that $||\mathbf{s}_i - \mathbf{s}_j|| = t$
- $|N(t)|$ is the number of pairs in this set
- Empirical semivariogram:

$$\hat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(t)} \left(Y(\mathbf{s}_i) - Y(\mathbf{s}_j)\right)^2$$

- For spatial data, $\gamma(t)$ is expected to increase with t
- A flat semivariogram would suggest little spatial variation

Empirical variogram
Dataset 1

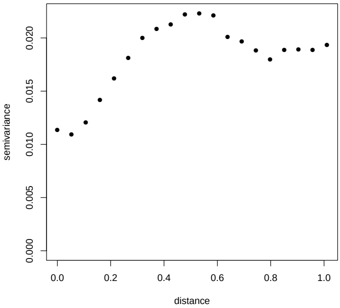


Figure: Empirical variogram on $y_1(\mathbf{s})$.

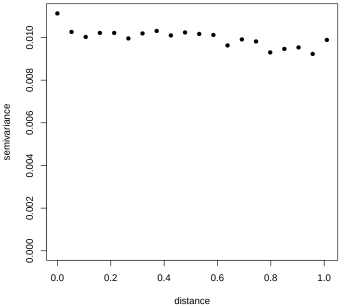


Figure: Empirical variogram on the residuals.

Empirical variogram
Dataset 1

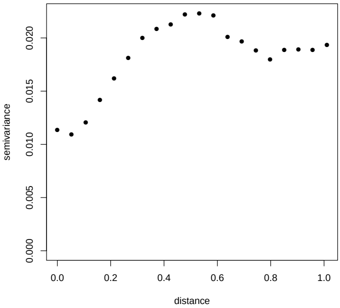


Figure: Empirical variogram on $y_1(\mathbf{s})$.

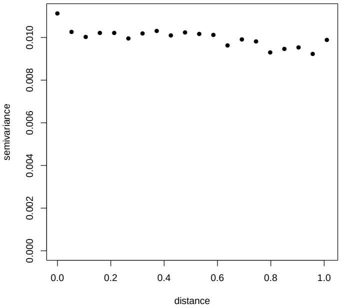


Figure: Empirical variogram on the residuals.

- Residuals display little spatial variation

Empirical variogram
Dataset 2

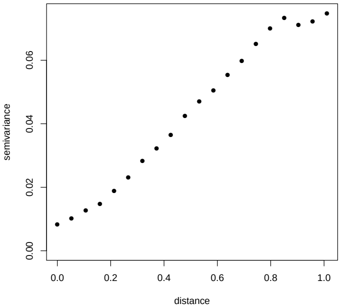


Figure: Empirical variogram of $y_2(\mathbf{s})$.

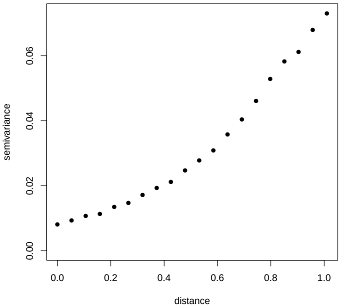


Figure: Empirical variogram on the residuals.

Empirical variogram
Dataset 2

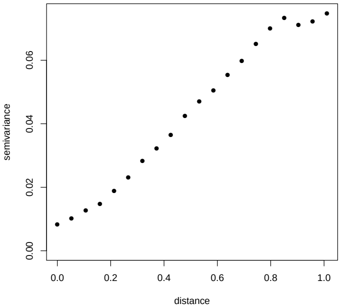


Figure: Empirical variogram of $y_2(\mathbf{s})$.

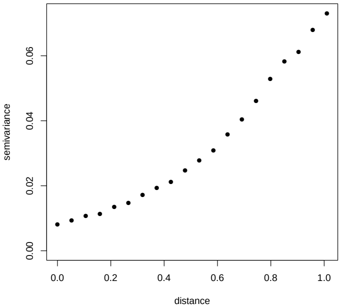


Figure: Empirical variogram on the residuals.

- Variogram of the residuals suggests **unexplained spatial variation**

Spatial regression

- When purely covariate-based models is not enough, one needs to leverage the information from locations

Spatial regression

- When purely covariate-based models is not enough, one needs to leverage the information from locations
- How to incorporate spatial information in linear regression?

Spatial regression

- When purely covariate-based models is not enough, one needs to leverage the information from locations
- How to incorporate spatial information in linear regression?

Spatial linear model

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\beta + w(\mathbf{s}) + \varepsilon(\mathbf{s})$$

Spatial regression

- When purely covariate-based models is not enough, one needs to leverage the information from locations
- How to incorporate spatial information in linear regression?

Spatial linear model

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\beta + w(\mathbf{s}) + \varepsilon(\mathbf{s})$$

- Since we want to predict at any location over the entire domain, we require a surface $w(\mathbf{s})$

Spatial regression

- When purely covariate-based models is not enough, one needs to leverage the information from locations
- How to incorporate spatial information in linear regression?

Spatial linear model

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\beta + w(\mathbf{s}) + \varepsilon(\mathbf{s})$$

- Since we want to predict at any location over the entire domain, we require a surface $w(\mathbf{s})$
- How to do this?

Spatial regression

- When purely covariate-based models is not enough, one needs to leverage the information from locations
- How to incorporate spatial information in linear regression?

Spatial linear model

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\beta + w(\mathbf{s}) + \varepsilon(\mathbf{s})$$

- Since we want to predict at any location over the entire domain, we require a surface $w(\mathbf{s})$
- How to do this?
- Model $w(\mathbf{s})$ as a stochastic process

Gaussian Processes (GPs)

- One popular approach to model $w(\mathbf{s})$ is via **Gaussian Processes**

Gaussian Processes (GPs)

- One popular approach to model $w(\mathbf{s})$ is via **Gaussian Processes**
- The collection of random variables $\{w(\mathbf{s}) \mid \mathbf{s} \in D\}$ is a GP if
  - it is a valid stochastic process
  - all finite dimensional densities $\{w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n)\}$ follow a multivariate Gaussian distribution

Gaussian Processes (GPs)

- One popular approach to model $w(\mathbf{s})$ is via **Gaussian Processes**
- The collection of random variables $\{w(\mathbf{s}) \mid \mathbf{s} \in D\}$ is a GP if
  - it is a valid stochastic process
  - all finite dimensional densities $\{w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n)\}$ follow a multivariate Gaussian distribution
- A GP is completely characterized by a mean function $m(\mathbf{s})$ and a covariance function $C(\cdot, \cdot)$

Gaussian Processes (GPs)

- One popular approach to model $w(\mathbf{s})$ is via **Gaussian Processes**
- The collection of random variables $\{w(\mathbf{s}) \mid \mathbf{s} \in D\}$ is a GP if
  - it is a valid stochastic process
  - all finite dimensional densities $\{w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n)\}$ follow a multivariate Gaussian distribution
- A GP is completely characterized by a mean function $m(\mathbf{s})$ and a covariance function $C(\cdot, \cdot)$
- Advantage: Likelihood based inference

$$\mathbf{w} = (w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n))' \sim \mathrm{N}(m, C)$$

where $m = (m(\mathbf{s}_1), \ldots, m(\mathbf{s}_n))'$ and $C = C(\mathbf{s}_i, \mathbf{s}_j)$.

Covariance function

- $C(\cdot, \cdot)$ needs to be valid, i.e., for all n and all $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$, the resulting covariance matrix $C = C(\mathbf{s}_i, \mathbf{s}_j)$ for $(w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n))$ must be positive definite
- So, $C(\cdot, \cdot)$ needs to be a positive definite function

Covariance function

- $C(\cdot, \cdot)$ needs to be valid, i.e., for all n and all $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$, the resulting covariance matrix $C = C(\mathbf{s}_i, \mathbf{s}_j)$ for $(w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n))$ must be positive definite
- So, $C(\cdot, \cdot)$ needs to be a positive definite function
- Construct valid covariance functions by using properties of characteristic functions:
  - Multiply valid covariance functions (corresponds to summing independent random variables)
  - Mixing covariance functions (corresponds to mixing distributions)
  - Convolving covariance functions, i.e., if c_1 and c_2 are valid then $c_{12}(\mathbf{s}) = \int c_1(\mathbf{s} - \mathbf{u}) c_2(\mathbf{u}) d\mathbf{u}$ is valid.

Covariance function

Simplifying assumptions

Covariance function
Simplifying assumptions

- Weak stationarity: $C(\mathbf{s}, \mathbf{s}+\mathbf{h}) = C(\mathbf{h}), \forall \mathbf{h} \in \mathbb{R}^r$ such that $\mathbf{s}$ and $\mathbf{s}+\mathbf{h}$ both lie within D.
  In words, the covariance between the values of the process at any two locations $\mathbf{s}$ and $\mathbf{s}+\mathbf{h}$ can be summarized by a covariance function $C(\mathbf{h})$ (sometimes called a covariogram), and this function depends only on the separation vector $\mathbf{h}$.

Covariance function

Simplifying assumptions

- Weak stationarity: $C(\mathbf{s}, \mathbf{s}+\mathbf{h}) = C(\mathbf{h}), \forall \mathbf{h} \in \mathbb{R}^r$ such that $\mathbf{s}$ and $\mathbf{s}+\mathbf{h}$ both lie within D.
  In words, the covariance between the values of the process at any two locations $\mathbf{s}$ and $\mathbf{s}+\mathbf{h}$ can be summarized by a covariance function $C(\mathbf{h})$ (sometimes called a covariogram), and this function depends only on the separation vector $\mathbf{h}$.

- Isotropic: $C(\mathbf{h}) = C(||\mathbf{h}||)$
  Isotropy implies circular contours in terms of decay in spatial dependence, i.e., association does not depend upon direction.

Covariance function

Simplifying assumptions

- Weak stationarity: $C(\mathbf{s}, \mathbf{s}+\mathbf{h}) = C(\mathbf{h}), \forall \mathbf{h} \in \mathbb{R}^r$ such that $\mathbf{s}$ and $\mathbf{s}+\mathbf{h}$ both lie within D.
  In words, the covariance between the values of the process at any two locations $\mathbf{s}$ and $\mathbf{s}+\mathbf{h}$ can be summarized by a covariance function $C(\mathbf{h})$ (sometimes called a covariogram), and this function depends only on the separation vector $\mathbf{h}$.
- Isotropic: $C(\mathbf{h}) = C(||\mathbf{h}||)$
  Isotropy implies circular contours in terms of decay in spatial dependence, i.e., association does not depend upon direction.
- Anisotropic: Stationary but not isotropic

Covariance function
Simplifying assumptions

- Weak stationarity: $C(\mathbf{s}, \mathbf{s}+\mathbf{h}) = C(\mathbf{h}), \forall \mathbf{h} \in \mathbb{R}^r$ such that $\mathbf{s}$ and $\mathbf{s}+\mathbf{h}$ both lie within D.
  In words, the covariance between the values of the process at any two locations $\mathbf{s}$ and $\mathbf{s}+\mathbf{h}$ can be summarized by a covariance function $C(\mathbf{h})$ (sometimes called a covariogram), and this function depends only on the separation vector $\mathbf{h}$.
- Isotropic: $C(\mathbf{h}) = C(||\mathbf{h}||)$
  Isotropy implies circular contours in terms of decay in spatial dependence, i.e., association does not depend upon direction.
- Anisotropic: Stationary but not isotropic

Isotropic models are popular because of their simplicity, interpretability, and because a number of relatively simple parametric forms are available as candidates for C.

Common isotropic covariance functions

Model	Covariance function, $C(t)$
Spherical	$C(t) = \begin{cases} 0 & \text{if } t \geq 1/\phi \\ \sigma^2 \left[1 - \frac{3}{2}\phi t + \frac{1}{2}(\phi t)^3\right] & \text{if } 0 < t \leq 1/\phi \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Exponential	$C(t) = \begin{cases} \sigma^2 \exp(-\phi t) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Powered exponential	$C(t) = \begin{cases} \sigma^2 \exp(-\lvert\phi t\rvert^p) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Gaussian	$C(t) = \begin{cases} \sigma^2 \exp(-\phi^2 t^2) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Rational quadratic	$C(t) = \begin{cases} \sigma^2 \left(1 - \frac{t^2}{(1+\phi t^2)}\right) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Wave	$C(t) = \begin{cases} \sigma^2 \frac{\sin(\phi t)}{\phi t} & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Matérn	$C(t) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(2\sqrt{\nu} t\phi\right)^\nu K_\nu(2\sqrt{\nu} t\phi) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$

Exponential covariance function

$$C(t) = \begin{cases} \tau^2 + \sigma^2 & \text{if } t = 0 \\ \sigma^2 \exp(-\phi t) & \text{if } t > 0 \end{cases}$$

- The effective range, t_0, as the distance at which the correlation has dropped to only 0.05. Setting $\exp(-\phi t_0)$ equal to this value we obtain $t_0 \approx 3/\phi$, since $\log(0.05) \approx -3$.
- The nugget τ^2 is often viewed as a "nonspatial effect variance".
- The partial sill σ^2 is viewed as a "spatial effect variance".
- The sill $\tau^2 + \sigma^2$ gives the maximum total variance
- Note discontinuity at 0 due to the nugget. This is intentional! To account for measurement error or micro-scale variability.

Covariance functions and variograms

- Recall the empirical semivariogram:

$$\hat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(t)} (Y(\mathbf{s}_i) - Y(\mathbf{s}_j))^2$$

Covariance functions and variograms

- Recall the empirical semivariogram:

$$\hat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(t)} \left(Y(\mathbf{s}_i) - Y(\mathbf{s}_j)\right)^2$$

- The theoretical semivariogram is defined as

$$\gamma(\mathbf{h}) = \frac{1}{2} E\left[\left(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})\right)^2\right]$$

  If the variogram depends only on **h** and not the particular choice of **s**, we say the process is intrinsically stationary.

Covariance functions and variograms

- Recall the empirical semivariogram:

$$\hat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{(\mathbf{s}_i,\mathbf{s}_j)\in N(t)} (Y(\mathbf{s}_i) - Y(\mathbf{s}_j))^2$$

- The theoretical semivariogram is defined as

$$\gamma(\mathbf{h}) = \frac{1}{2} E\left[(Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s}))^2\right]$$

  If the variogram depends only on $\mathbf{h}$ and not the particular choice of $\mathbf{s}$, we say the process is intrinsically stationary.

- For any stationary GP:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$$

Covariance functions and variograms

- Recall the empirical semivariogram:

$$\hat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(t)} \left(Y(\mathbf{s}_i) - Y(\mathbf{s}_j)\right)^2$$

- The theoretical semivariogram is defined as

$$\gamma(\mathbf{h}) = \frac{1}{2} E\left[\left(Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})\right)^2\right]$$

  If the variogram depends only on $\mathbf{h}$ and not the particular choice of $\mathbf{s}$, we say the process is intrinsically stationary.

- For any stationary GP:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$$

- So given $C(\mathbf{h})$, we are able to determine $\gamma(\mathbf{h})$.

Covariance functions and variograms

- Recall the empirical semivariogram:

$$\hat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(t)} \left(Y(\mathbf{s}_i) - Y(\mathbf{s}_j)\right)^2$$

- The theoretical semivariogram is defined as

$$\gamma(\mathbf{h}) = \frac{1}{2} E\left[\left(Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})\right)^2\right]$$

  If the variogram depends only on $\mathbf{h}$ and not the particular choice of $\mathbf{s}$, we say the process is intrinsically stationary.

- For any stationary GP:

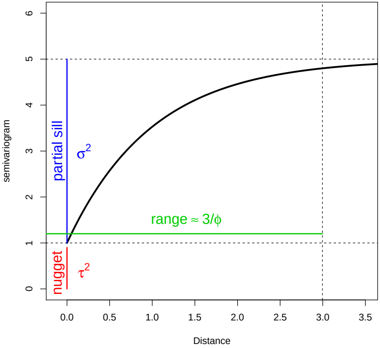
$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$$

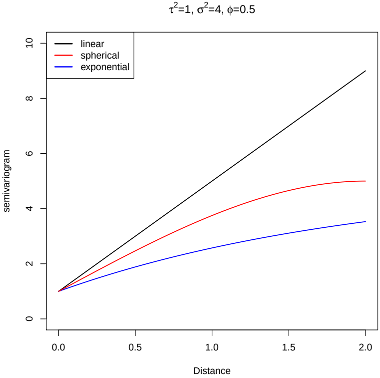
- So given $C(\mathbf{h})$, we are able to determine $\gamma(\mathbf{h})$.
- Example: For exponential (isotropic) GP

$$\gamma(t) = \begin{cases} 0 & \text{if } t = 0 \\ \tau^2 + \sigma^2(1 - \exp(-\phi t)) & \text{if } t > 0 \end{cases}$$

Variograms

Examples





The Matèrn Correlation Function

- The Matèrn is a very versatile family:

$$C(t) = \begin{cases} \tau^2 + \sigma^2 & \text{if } t = 0 \\ \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\phi t)^\nu K_\nu(\phi t) & \text{if } t > 0 \end{cases}$$

  where K_ν is the modified Bessel function of order ν (computationally tractable)

- ν is a smoothness parameter controlling process smoothness:
  - $\nu = 1/2$ gives the exponential covariance function
  - $\nu = 3/2$ gives a convenient closed form
  - $\nu = \infty$ gives the Gaussian covariance function

Modeling point-referenced data

Spatial linear model

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\beta + w(\mathbf{s}) + \varepsilon(\mathbf{s})$$

Modeling point-referenced data

Spatial linear model

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\beta + w(\mathbf{s}) + \varepsilon(\mathbf{s})$$

- $w(\mathbf{s})$ is the spatial error (spatial random effects) and it is modeled as a stationary GP $(0, C(\cdot, \cdot \mid \theta))$
- Under isotropy: $C(\mathbf{s} + \mathbf{h}, \mathbf{s} \mid \theta) = \sigma^2 \mathbf{R}\,(||h||; \phi)$
- $\mathbf{w} = (w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n))' \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{R}(\phi))$ where $\mathbf{R}(\phi)_{ij} = \rho(||\mathbf{s}_i - \mathbf{s}_j||; \phi))$, introducing the partial sill (σ^2) and range (ϕ) parameters.

Modeling point-referenced data

Spatial linear model

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\beta + w(\mathbf{s}) + \varepsilon(\mathbf{s})$$

- $w(\mathbf{s})$ is the spatial error (spatial random effects) and it is modeled as a stationary $\mathrm{GP}\left(0, C(\cdot, \cdot \mid \theta)\right)$
- Under isotropy: $C(\mathbf{s}+\mathbf{h}, \mathbf{s} \mid \theta) = \sigma^2 \mathbf{R}\left(||h||; \phi\right)$
- $\mathbf{w} = (w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n))' \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{R}(\phi))$ where $\mathbf{R}(\phi)_{ij} = \rho(||\mathbf{s}_i - \mathbf{s}_j||; \phi))$, introducing the partial sill (σ^2) and range (ϕ) parameters.

- $\varepsilon(\mathbf{s}) \sim$ iid $\mathrm{N}(0, \tau^2)$ is the non-spatial error and contributes to the nugget τ^2
- Interpretations attached to $\varepsilon(\mathbf{s})$:
  - pure error term; model is not perfectly spatial;
  - measurement error or replication variability causing discontinuity in spatial surface $Y(\mathbf{s})$;
  - microscale uncertainty; distances smaller than the smallest inter-location distance.

Parameter estimation

- Gaussian spatial models are special cases of of the general linear model, with a particular specification of the covariance matrix: $\Sigma_\theta = \sigma^2 \mathbf{R}(\phi) + \tau^2 I_n$
- The sampling model is given by

$$\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\beta} \sim \mathrm{N}_n \left(\mathbf{X}\boldsymbol{\beta}, \Sigma_\theta \right)$$

  where $\boldsymbol{\theta} = (\sigma^2, \tau^2, \phi)$.

- So the likelihood is

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) \propto |\Sigma_\theta|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \Sigma_\theta^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}$$

Parameter estimation

- Gaussian spatial models are special cases of of the general linear model, with a particular specification of the covariance matrix: $\Sigma_\theta = \sigma^2 \mathbf{R}(\phi) + \tau^2 I_n$
- The sampling model is given by

$$\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\beta} \sim \mathrm{N}_n \left(\mathbf{X}\boldsymbol{\beta}, \Sigma_\theta \right)$$

where $\boldsymbol{\theta} = (\sigma^2, \tau^2, \phi)$.

- So the likelihood is

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) \propto |\Sigma_\theta|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \Sigma_\theta^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}$$

- The Maximum Likelihood Estimates (MLE) of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ that maximizes the likelihood above
- $\hat{\boldsymbol{\beta}}_{MLE} = \hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}' \Sigma_\theta^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma_\theta^{-1} \mathbf{y}$

Parameter estimation

- In practice, the likelihood is often very flat with respect to the spatial covariance parameters and choice of initial values is important
- Initial values can be eyeballed from empirical semivariogram of the residuals from ordinary linear regression
- Estimated parameter values can be used for prediction (*kriging*)

Parameter estimation

Example: Ozone data

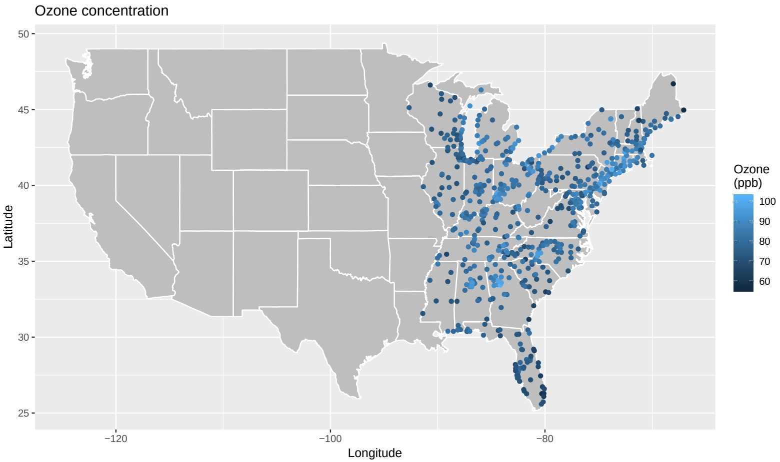


Figure: Ozone concentrations at $n = 631$ monitoring stations in the Eastern US.

Parameter estimation
Example: Ozone data

- Model: $y(\mathbf{s}) = \beta_0 + w(\mathbf{s}) + \varepsilon(\mathbf{s})$
- $w(\mathbf{s}) \sim \mathrm{GP}\,(0, C)$, $C(\mathbf{s}_i, \mathbf{s}_j) = \sigma^2 \exp(-\phi||\mathbf{s}_i - \mathbf{s}_j||)$, $\varepsilon(\mathbf{s}) \sim \mathrm{N}(0, \tau^2)$

Parameter estimation

Example: Ozone data

- Model: $y(\mathbf{s}) = \beta_0 + w(\mathbf{s}) + \varepsilon(\mathbf{s})$
- $w(\mathbf{s}) \sim \mathrm{GP}\,(0, C)$, $C(\mathbf{s}_i, \mathbf{s}_j) = \sigma^2 \exp(-\phi||\mathbf{s}_i - \mathbf{s}_j||)$, $\varepsilon(\mathbf{s}) \sim \mathrm{N}(0, \tau^2)$
- MLE obtained using `likfit` function of `geoR` `R` package

```
fit_mle <- likfit(Y, coords, fix.nugget = FALSE,
                  cov.model="exponential", ini = c(var(Y), 1))
summary(fit_mle)

Parameters of the mean component (trend): beta0  76.1311

Parameters of the spatial component:
   correlation function: exponential
      (estimated) variance parameter sigmasq (partial sill)  =  47.82
      (estimated) cor. fct. parameter phi (range parameter)  =  197

   anisotropy parameters:
      (fixed) anisotropy angle = 0  ( 0 degrees )
      (fixed) anisotropy ratio = 1

Parameter of the error component:
      (estimated) nugget =  15.7

Practical Range with cor=0.05 for asymptotic range: 590.0498
```

Hierarchical modeling

- Alternatively, the **Bayesian** posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ can be derived (and sampled from). A prior distribution of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ is required.
- The spatial model can be recast as a hierarchical model as follows:

Hierarchical modeling

- Alternatively, the **Bayesian** posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ can be derived (and sampled from). A prior distribution of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ is required.
- The spatial model can be recast as a hierarchical model as follows:

First stage (Data Model): [data | process, parameters]

$$\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{w}, \tau^2 \sim \mathrm{N}_n\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{w}, \tau^2 I_n\right)$$

The $Y(\mathbf{s}_i)$ are conditionally independent given the $w(\mathbf{s}_i)$'s.

Hierarchical modeling

- Alternatively, the **Bayesian** posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ can be derived (and sampled from). A prior distribution of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ is required.
- The spatial model can be recast as a hierarchical model as follows:

First stage (Data Model): [data | process, parameters]

$$\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{w}, \tau^2 \sim \mathrm{N}_n\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{w}, \tau^2 I_n\right)$$

The $Y(\mathbf{s}_i)$ are conditionally independent given the $w(\mathbf{s}_i)$'s.

Second stage (Process Model): [process | parameters]

$$\mathbf{w} \mid \sigma^2, \phi \sim \mathrm{N}_n\left(\mathbf{0}, \sigma^2 \mathbf{R}(\phi)\right)$$

Hierarchical modeling

- Alternatively, the **Bayesian** posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ can be derived (and sampled from). A prior distribution of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ is required.
- The spatial model can be recast as a hierarchical model as follows:

First stage (Data Model): [data | process, parameters]

$$\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{w}, \tau^2 \sim \mathrm{N}_n\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{w}, \tau^2 I_n\right)$$

The $Y(\mathbf{s}_i)$ are conditionally independent given the $w(\mathbf{s}_i)$'s.

Second stage (Process Model): [process | parameters]

$$\mathbf{w} \mid \sigma^2, \phi \sim \mathrm{N}_n\left(\mathbf{0}, \sigma^2 \mathbf{R}(\phi)\right)$$

Third stage (Parameter Model): [(hyper)parameters]

priors on $(\boldsymbol{\beta}, \boldsymbol{\theta})$

Bayesian posterior inference

- Markov chain Monte Carlo methods can be used to sample from the posterior distribution $p(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \mathbf{y})$
- The model can be fitted either as:
  - $\star$ $f(\mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\theta}) p(\boldsymbol{\beta}, \boldsymbol{\theta})$
  - $\star$ $f(\mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{w}) p(\mathbf{w} \mid \boldsymbol{\theta}) p(\boldsymbol{\beta}, \boldsymbol{\theta})$

$\Rightarrow$ Exercise: derive the forms of the full conditional distributions for $\boldsymbol{\beta}$, σ^2, τ^2, ϕ and $\mathbf{w}$ in the exponential covariance model

Parameter estimation

Example: Ozone data

Posterior inference obtained from `spBayes` `R` package (Finley et al., 2015)

```
starting <- list(phi = 3/(0.5 * d.max), sigma.sq = 5, tau.sq = 5)
tuning   <- list(phi = 0.5, sigma.sq = 0.05, tau.sq = 0.05)
priors   <- list("beta.Flat", phi.Unif = c(3/d.max, 3/(0.01 * d.max)),
               sigma.sq.IG = c(2, 5), tau.sq.IG = c(2, 5))

fit_bayes <- spLM(Y ~ 1, coords = coords.utm,  cov.model = "exponential",
                  starting = starting, tuning = tuning, priors = priors,
                  n.samples = 5000, n.report = 2000)

round(apply(fit_bayes$p.theta.samples,2,quantile))

      sigma.sq   tau.sq    phi
  0%     4.776    6.356  0.001
 25%    40.310   14.439  0.004
 50%    45.558   15.588  0.005
 75%    53.145   16.857  0.006
100%   191.853   36.265  0.010
```

Kriging: spatial prediction at new locations

- Kriging: Named in honor of D.G. Krige, a South African mining engineer whose seminal work on empirical methods for geostatistical data inspired the general approach.
- **Goal**: given the observations $\mathbf{y}$, predict Y at new site $\mathbf{s}_0$ where $\mathbf{s}_0$ has not been observed

Conditional normal distribution

- $\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$

Conditional normal distribution

- $\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$
- Then, $\mathbf{X}_1 \mid \mathbf{X}_2 \sim \mathrm{N}\left(\boldsymbol{\mu}_{1|2}, \Sigma_{1|2}\right)$, where
- $\mathrm{E}[\mathbf{X}_1 \mid \mathbf{X}_2] = \mu_{1|2} = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}(\mathbf{X}_2 - \boldsymbol{\mu}_2)$ is the conditional mean
- $\mathrm{Var}[\mathbf{X}_1 \mid \mathbf{X}_2] = \Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is the conditional variance

Conditional normal distribution

- $\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$
- Then, $\mathbf{X}_1 \mid \mathbf{X}_2 \sim \mathrm{N}\left(\boldsymbol{\mu}_{1|2}, \Sigma_{1|2}\right)$, where
- $\mathrm{E}[\mathbf{X}_1 \mid \mathbf{X}_2] = \mu_{1|2} = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}(\mathbf{X}_2 - \boldsymbol{\mu}_2)$ is the conditional mean
- $\mathrm{Var}[\mathbf{X}_1 \mid \mathbf{X}_2] = \Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is the conditional variance
- The conditional mean is the “best” predictor of $\mathbf{X}_1$ based on $\mathbf{X}_2$, i.e., $\boldsymbol{\mu}_{1|2}$ minimize the mean-squared prediction error

Kriging: spatial prediction at new locations

- Kriging: Named in honor of D.G. Krige, a South African mining engineer whose seminal work on empirical methods for geostatistical data inspired the general approach.
- **Goal**: given the observations $\mathbf{y}$, predict Y at new site $\mathbf{s}_0$ where $\mathbf{s}_0$ has not been observed
- Recalling the linear spatial model with GP, we have

$$\begin{pmatrix} y(\mathbf{s}_0) \\ \mathbf{y} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{x}_0'\boldsymbol{\beta} \\ \mathbf{X}\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \tau^2 + \sigma^2 & \boldsymbol{\gamma}_\phi' \\ \boldsymbol{\gamma}_\phi & \Sigma_\theta \end{pmatrix} \right)$$

where $\boldsymbol{\gamma}_\phi' = \left(\sigma^2\rho(||\mathbf{s}_1 - \mathbf{s}_0||;\phi), \ldots, \sigma^2\rho(||\mathbf{s}_n - \mathbf{s}_0||;\phi)\right)$ and $\Sigma_\theta = \sigma^2\mathbf{R}(\phi) + \tau^2 I_n$.

Kriging: spatial prediction at new locations

- Kriging: Named in honor of D.G. Krige, a South African mining engineer whose seminal work on empirical methods for geostatistical data inspired the general approach.
- **Goal**: given the observations $\mathbf{y}$, predict Y at new site $\mathbf{s}_0$ where $\mathbf{s}_0$ has not been observed
- Recalling the linear spatial model with GP, we have

$$\begin{pmatrix} y(\mathbf{s}_0) \\ \mathbf{y} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{x}_0'\boldsymbol{\beta} \\ \mathbf{X}\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \tau^2 + \sigma^2 & \boldsymbol{\gamma}_\phi' \\ \boldsymbol{\gamma}_\phi & \Sigma_\theta \end{pmatrix} \right)$$

  where $\boldsymbol{\gamma}_\phi' = \left(\sigma^2\rho(||\mathbf{s}_1 - \mathbf{s}_0||;\phi), \ldots, \sigma^2\rho(||\mathbf{s}_n - \mathbf{s}_0||;\phi)\right)$ and $\Sigma_\theta = \sigma^2\mathbf{R}(\phi) + \tau^2 I_n$.

- Substituting these values into the conditional mean and variance, we obtain

$$\mathrm{E}[Y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}] = \mathbf{x}_0'\boldsymbol{\beta} + \boldsymbol{\gamma}_\phi'\Sigma_\theta^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad \textbf{kriging estimator}$$

$$\mathrm{Var}[Y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}] = \tau^2 + \sigma^2 - \boldsymbol{\gamma}_\phi'\Sigma_\theta^{-1}\boldsymbol{\gamma}_\phi$$

Kriging: spatial prediction at new locations

- Plugging-in ML estimates:

$$\hat{Y}(\mathbf{s}_0) = \mathbf{x}_0'\hat{\boldsymbol{\beta}}_{MLE} + \hat{\boldsymbol{\gamma}}_\phi'\hat{\Sigma}_\theta^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MLE}\right)$$

$$\mathrm{Var}[\hat{Y}(\mathbf{s}_0)] = \hat{\tau}^2 + \hat{\sigma}^2 - \hat{\boldsymbol{\gamma}}_\phi'\hat{\Sigma}_\theta^{-1}\hat{\boldsymbol{\gamma}}_\phi$$

Kriging: spatial prediction at new locations

- Plugging-in ML estimates:

$$\hat{Y}(\mathbf{s}_0) = \mathbf{x}_0'\hat{\boldsymbol{\beta}}_{MLE} + \hat{\boldsymbol{\gamma}}_\phi'\hat{\Sigma}_\theta^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{MLE}\right)$$
$$\mathrm{Var}[\hat{Y}(\mathbf{s}_0)] = \hat{\tau}^2 + \hat{\sigma}^2 - \hat{\boldsymbol{\gamma}}_\phi'\hat{\Sigma}_\theta^{-1}\hat{\boldsymbol{\gamma}}_\phi$$

Thus we can write $\hat{Y}(\mathbf{s}_0) = \boldsymbol{\lambda}'\mathbf{y}$, where

$$\boldsymbol{\lambda} = \hat{\Sigma}_\theta^{-1}\hat{\boldsymbol{\gamma}}_\phi + \hat{\Sigma}_\theta^{-1}\mathbf{X}\left(\mathbf{X}'\hat{\Sigma}_\theta^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{x}_0 - \mathbf{X}'\hat{\Sigma}_\theta^{-1}\hat{\boldsymbol{\gamma}}_\phi\right)$$

Kriging: spatial prediction at new locations

- Plugging-in ML estimates:

$$\hat{Y}(\mathbf{s}_0) = \mathbf{x}_0' \hat{\boldsymbol{\beta}}_{MLE} + \hat{\boldsymbol{\gamma}}_\phi' \hat{\Sigma}_\theta^{-1} \left(\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{MLE} \right)$$
$$\text{Var}[\hat{Y}(\mathbf{s}_0)] = \hat{\tau}^2 + \hat{\sigma}^2 - \hat{\boldsymbol{\gamma}}_\phi' \hat{\Sigma}_\theta^{-1} \hat{\boldsymbol{\gamma}}_\phi$$

  Thus we can write $\hat{Y}(\mathbf{s}_0) = \boldsymbol{\lambda}' \mathbf{y}$, where

$$\boldsymbol{\lambda} = \hat{\Sigma}_\theta^{-1} \hat{\boldsymbol{\gamma}}_\phi + \hat{\Sigma}_\theta^{-1} \mathbf{X} \left(\mathbf{X}' \hat{\Sigma}_\theta^{-1} \mathbf{X} \right)^{-1} \left(\mathbf{x}_0 - \mathbf{X}' \hat{\Sigma}_\theta^{-1} \hat{\boldsymbol{\gamma}}_\phi \right)$$

- Bayesian Kriging computes (simulates) posterior predictive density:

$$p(y(\mathbf{s}_0) \mid \mathbf{y}) = \int p\left(\mathbf{s}_0 \mid \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}\right) p\left(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \mathbf{y}\right) d\boldsymbol{\beta} d\boldsymbol{\theta}$$

Kriging: spatial prediction at new locations

Example: Ozone data

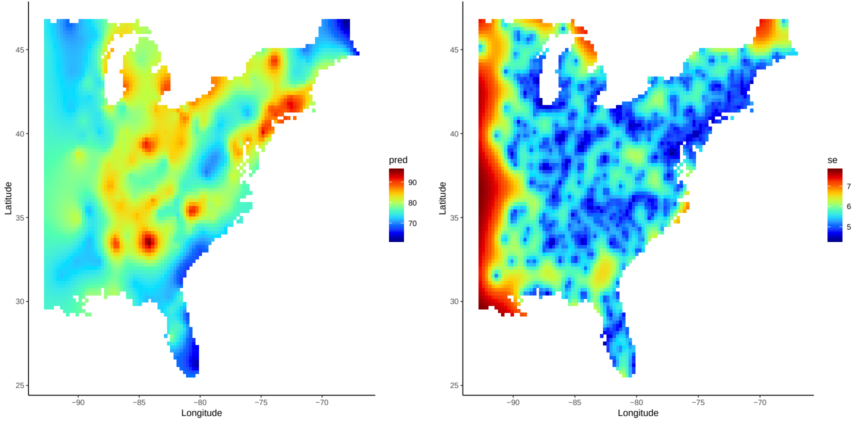


Figure: Kriged surface (left panel) and predicted standard error (right panel).

Kriging: spatial prediction at new locations

- What is the kriging estimator of $w(\mathbf{s}_0)$ when $\mathbf{s}_0 = \mathbf{s}_i$ for some i ?

Kriging: spatial prediction at new locations

- What is the kriging estimator of $w(\mathbf{s}_0)$ when $\mathbf{s}_0 = \mathbf{s}_i$ for some i ?
- Kriging predictions of $w(\mathbf{s}_0)$ at data locations are the observed values themselves with prediction variance equaling zero; kriging interpolates!

Kriging: spatial prediction at new locations

- What is the kriging estimator of $w(\mathbf{s}_0)$ when $\mathbf{s}_0 = \mathbf{s}_i$ for some i ?
- Kriging predictions of $w(\mathbf{s}_0)$ at data locations are the observed values themselves with prediction variance equaling zero; kriging interpolates!
- What happens when $\mathbf{s}_0$ is far away from all the $\mathbf{s}_i$'s ?

Kriging: spatial prediction at new locations

- What is the kriging estimator of $w(\mathbf{s}_0)$ when $\mathbf{s}_0 = \mathbf{s}_i$ for some i ?
- Kriging predictions of $w(\mathbf{s}_0)$ at data locations are the observed values themselves with prediction variance equaling zero; kriging interpolates!
- What happens when $\mathbf{s}_0$ is far away from all the $\mathbf{s}_i$'s ?
- $(\rho(\mathbf{s}_1, \mathbf{s}_0; \phi), \ldots, \rho(\mathbf{s}_n, \mathbf{s}_0; \phi))' \approx (0, \ldots, 0)'$
- Kriging mean: $\approx \mu(\mathbf{s}_0) =$ unconditional mean
- Kriging variance: $\approx C(\mathbf{s}_0, \mathbf{s}_0) =$ unconditional variance
- $w(\mathbf{s}_0)$ is almost independent of the $w(\mathbf{s}_i)$'s i.e., we are not exploiting much information on the process at far away locations

Model comparison

- Likelihood-based criteria. For k total parameters and sample size n:
  - AIC: $2k - 2log(L(\mathbf{y} \mid \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}))$
  - BIC: $log(n)k - 2log(L(\mathbf{y} \mid \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}))$

Model comparison

- Likelihood-based criteria. For k total parameters and sample size n:
  - AIC: $2k - 2log(L(\mathbf{y} \mid \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}))$
  - BIC: $log(n)k - 2log(L(\mathbf{y} \mid \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}))$
- Prediction-based approaches using holdout data:
  - Rqoot Mean Square Predictive Error (RMSPE): $\sqrt{\frac{1}{n_{out}} \sum_{i=1}^{n_{out}} \left(y(\mathbf{s}_i) - \hat{y}(\mathbf{s}_i)\right)^2}$
  - Coverage probability (CP): $\frac{1}{n_{out}} \sum_{i=1}^{n_{out}} I\left(y(\mathbf{s}_i) \in (\hat{y}(\mathbf{s}_i)_{0.025}, \hat{y}(\mathbf{s}_i)_{0.975})\right)$
  - Width of 95% confidence interval (CIW): $\frac{1}{n_{out}} \sum_{i=1}^{n_{out}} \left(\hat{y}(\mathbf{s}_i)_{0.025} - \hat{y}(\mathbf{s}_i)_{0.975}\right)$
  - The last two approaches compares the distribution of $y(\mathbf{s}_i)$ instead of comparing just their point predictions

Spatial Generalized Linear Models

- Some data sets preclude Gaussian modeling; $Y(\mathbf{s})$ need not be continuous

Spatial Generalized Linear Models

- Some data sets preclude Gaussian modeling; $Y(\mathbf{s})$ need not be continuous
- Example: $Y(\mathbf{s})$ is a binary or count variable
  - Presence/absence of a species at a location;
  - Abundance of a species at a location;
  - Precipitation or deposition was measurable or not at a location
  - Number of insurance claims by residents of a single-family home at a given location
  - Land use classification at a location (not ordinal)

Spatial Generalized Linear Models

- Some data sets preclude Gaussian modeling; $Y(\mathbf{s})$ need not be continuous
- Example: $Y(\mathbf{s})$ is a binary or count variable
  - Presence/absence of a species at a location;
  - Abundance of a species at a location;
  - Precipitation or deposition was measurable or not at a location
  - Number of insurance claims by residents of a single-family home at a given location
  - Land use classification at a location (not ordinal)
- Replace Gaussian likelihood by an appropriate non-Gaussian likelihood $f\left(y(\mathbf{s}_i)|\boldsymbol{\beta}, w(\mathbf{s}_i), \delta\right)$ such that

$$g\left(\mathrm{E}(Y(\mathbf{s}_i))\right) = \eta(\mathbf{s}_i) = \mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i),$$

where η is a canonical link function (such as a logit) and δ is a dispersion parameter.

The "big n problem"

- $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$ are locations where data are observed
- $y(\mathbf{s}_i)$ is outcome at the i-th location and $\mathbf{y} = (y(\mathbf{s}_1), y(\mathbf{s}_2), \ldots, y(\mathbf{s}_n))'$
- Model: $\mathbf{y} \sim \mathrm{N}\left(\mathbf{X}\boldsymbol{\beta}, \Sigma_\theta\right)$, where $\Sigma_\theta = \Sigma(\sigma^2, \phi)$
- Estimating process parameters from the likelihood:

$$-\frac{1}{2}\log\det(\Sigma_\theta) - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\Sigma_\theta^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

The "big n problem"

- $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$ are locations where data are observed
- $y(\mathbf{s}_i)$ is outcome at the i-th location and $\mathbf{y} = (y(\mathbf{s}_1), y(\mathbf{s}_2), \ldots, y(\mathbf{s}_n))'$
- Model: $\mathbf{y} \sim \mathrm{N}\left(\mathbf{X}\boldsymbol{\beta}, \Sigma_\theta\right)$, where $\Sigma_\theta = \Sigma(\sigma^2, \phi)$
- Estimating process parameters from the likelihood:

$$-\frac{1}{2}\log\det(\Sigma_\theta) - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\Sigma_\theta^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

- Σ_θ is usually dense with no exploitable structure

The “big n problem”

- $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$ are locations where data are observed
- $y(\mathbf{s}_i)$ is outcome at the i-th location and $\mathbf{y} = (y(\mathbf{s}_1), y(\mathbf{s}_2), \ldots, y(\mathbf{s}_n))'$
- Model: $\mathbf{y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \Sigma_\theta)$, where $\Sigma_\theta = \Sigma(\sigma^2, \phi)$
- Estimating process parameters from the likelihood:

$$-\frac{1}{2}\log\det(\Sigma_\theta) - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\Sigma_\theta^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

- Σ_θ is usually dense with no exploitable structure
- Need to calculate $\det(\Sigma_\theta)$ and quadratic forms of Σ_θ^{-1}.

The "big n problem"

- Conditional predictive density

$$p\left(y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}\right) = \mathrm{N}\left(y(\mathbf{s}_0) \mid \mu(\mathbf{s}_0), \sigma^2(\mathbf{s}_0)\right)$$

- Classig Kriging, spatial prediction/interpolation

$$\mathrm{E}[Y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}] = \mathbf{x}_0'\boldsymbol{\beta} + \boldsymbol{\gamma}'\Sigma_\theta^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$
$$\mathrm{Var}[Y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}] = \tau^2 + \sigma^2 - \boldsymbol{\gamma}'\Sigma_\theta^{-1}\boldsymbol{\gamma}$$

- Bayesian Kriging computes (simulates) posterior predictive density:

$$p\left(y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}\right) = \int p\left(y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}\right) p\left(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \mathbf{y}\right) d\boldsymbol{\beta} d\boldsymbol{\theta}$$

The "big n problem"

- Conditional predictive density

$$p\left(y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}\right) = \mathrm{N}\left(y(\mathbf{s}_0) \mid \mu(\mathbf{s}_0), \sigma^2(\mathbf{s}_0)\right)$$

- Classig Kriging, spatial prediction/interpolation

$$\begin{aligned} \mathrm{E}[Y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}] &= \mathbf{x}_0'\boldsymbol{\beta} + \boldsymbol{\gamma}'\Sigma_\theta^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ \mathrm{Var}[Y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}] &= \tau^2 + \sigma^2 - \boldsymbol{\gamma}'\Sigma_\theta^{-1}\boldsymbol{\gamma} \end{aligned}$$

- Bayesian Kriging computes (simulates) posterior predictive density:

$$p\left(y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}\right) = \int p\left(y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}\right) p\left(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \mathbf{y}\right) d\boldsymbol{\beta} d\boldsymbol{\theta}$$

- Again need to evaluate quadratic forms of Σ_θ^{-1}.

The "big n problem"

- Conditional predictive density

$$p\left(y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}\right) = \mathrm{N}\left(y(\mathbf{s}_0) \mid \mu(\mathbf{s}_0), \sigma^2(\mathbf{s}_0)\right)$$

- Classig Kriging, spatial prediction/interpolation

$$\mathrm{E}[Y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}] = \mathbf{x}_0'\boldsymbol{\beta} + \boldsymbol{\gamma}'\Sigma_\theta^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$
$$\mathrm{Var}[Y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta}] = \tau^2 + \sigma^2 - \boldsymbol{\gamma}'\Sigma_\theta^{-1}\boldsymbol{\gamma}$$

- Bayesian Kriging computes (simulates) posterior predictive density:

$$p\left(y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}\right) = \int p\left(y(\mathbf{s}_0) \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\beta}\right) p\left(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \mathbf{y}\right) d\boldsymbol{\beta} d\boldsymbol{\theta}$$

- Again need to evaluate quadratic forms of Σ_θ^{-1}.
- **Challenge**: evaluating the density requires $O(n^3)$ operations and $O(n^2)$ memory which can quickly overwhelm computing systems when n is only moderately large.

Literature on spatial big data

1. Low rank methods: typically involve reducing the rank of the $n \times n$ matrix Σ_θ (Fixed Rank Kriging, Predictive process);

Literature on spatial big data

1. Low rank methods: typically involve reducing the rank of the $n \times n$ matrix Σ_θ (Fixed Rank Kriging, Predictive process);

2. Sparse covariance matrix methods: work by introducing “0's”' into Σ allowing for sparse matrix computations (Spatial Partitioning, Covariance Tapering);

Literature on spatial big data

1. Low rank methods: typically involve reducing the rank of the $n \times n$ matrix Σ_θ (Fixed Rank Kriging, Predictive process);
2. Sparse covariance matrix methods: work by introducing "0's"' into Σ allowing for sparse matrix computations (Spatial Partitioning, Covariance Tapering);
3. Sparse precision matrix methods: induce sparsity in the precision matrix Σ^{-1} to allow for efficient computation (Multiresolution Approximations, Stochastic PDEs, Nearest Neighbor Processes);

Literature on spatial big data

1. Low rank methods: typically involve reducing the rank of the $n \times n$ matrix Σ_θ (Fixed Rank Kriging, Predictive process);
2. Sparse covariance matrix methods: work by introducing “0's”' into Σ allowing for sparse matrix computations (Spatial Partitioning, Covariance Tapering);
3. Sparse precision matrix methods: induce sparsity in the precision matrix Σ^{-1} to allow for efficient computation (Multiresolution Approximations, Stochastic PDEs, Nearest Neighbor Processes);
4. Algorithmic approaches: differ from the previous approaches in that they focus more on fitting schemes than model building (Metakriging, Gapfill, Local Approximate Gaussian Processes).

A case study competition

- Heaton, M.J., Datta, A., Finley, A.O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R.B., Hammerling, D., Katzfuss, M. and Lindgren, F. (2018), A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, pp.1-28.

- A case study competition between various research groups across the globe who each implemented their own method to analyze the same spatial datasets.

- Data and code available at: `https://github.com/finnlindgren/heatoncomparison`

Outline

Spatio-temporal analysis

- Spatio-temporal data arranged into a matrix with n rows and T columns, where n is the number of sites and T is the number of time points

Spatio-temporal analysis

- Spatio-temporal data arranged into a matrix with n rows and T columns, where n is the number of sites and T is the number of time points
- Space is continuous (so far) and time?

Spatio-temporal analysis

- Spatio-temporal data arranged into a matrix with n rows and T columns, where n is the number of sites and T is the number of time points
- Space is continuous (so far) and time?

Continuous time vs discretized time

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

- Spatio-temporal domain: $D = D_s \times D_t$, where $D_s \subset \mathbb{R}^r$ and $D_t \subset \mathbb{R}^+$
- Each observation has a space-time coordinate

$$\{Y_t(\mathbf{s}), \mathbf{s} \in D_s, t = 1, \ldots, T\}$$

- Geo-referenced time series
- Restricted to data observed at *regular* time intervals (hourly, daily, etc)
- Interpolation at finer temporal resolution not possible

Continuous space-time models

- Extending spatial linear regression by connecting space and time

Continuous space-time models

- Extending spatial linear regression by connecting space and time

$$y(\mathbf{s},t) = \mu(\mathbf{s},t) + w(\mathbf{s},t) + \varepsilon(\mathbf{s},t)$$

- $\mu(\mathbf{s},t)$ is the mean structure, usually set to as $\mu(\mathbf{s},t) = \mathbf{x}'(\mathbf{s},t)\boldsymbol{\beta}(\mathbf{s},t)$;

- $w(\mathbf{s},t)$ is the spatio-temporal random effect;

- $\varepsilon(\mathbf{s},t)$ is the unstructured error

Continuous space-time models

- Extending spatial linear regression by connecting space and time

$$y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + w(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$$

- $\mu(\mathbf{s}, t)$ is the mean structure, usually set to as $\mu(\mathbf{s}, t) = \mathbf{x}'(\mathbf{s}, t)\boldsymbol{\beta}(\mathbf{s}, t)$;
  - $\boldsymbol{\beta}(\mathbf{s}, t) = \boldsymbol{\beta}$; $\boldsymbol{\beta}(\mathbf{s}, t) = \boldsymbol{\beta}(\mathbf{s})$; $\boldsymbol{\beta}(\mathbf{s}, t) = \boldsymbol{\beta}(t)$
- $w(\mathbf{s}, t)$ is the spatio-temporal random effect;
- $\varepsilon(\mathbf{s}, t)$ is the unstructured error

Continuous space-time models

- Extending spatial linear regression by connecting space and time

$$y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + w(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$$

- $\mu(\mathbf{s}, t)$ is the mean structure, usually set to as $\mu(\mathbf{s}, t) = \mathbf{x}'(\mathbf{s}, t)\boldsymbol{\beta}(\mathbf{s}, t)$;
  - $\boldsymbol{\beta}(\mathbf{s}, t) = \boldsymbol{\beta}$; $\boldsymbol{\beta}(\mathbf{s}, t) = \boldsymbol{\beta}(\mathbf{s})$; $\boldsymbol{\beta}(\mathbf{s}, t) = \boldsymbol{\beta}(t)$

- $w(\mathbf{s}, t)$ is the spatio-temporal random effect;
  - $w(\mathbf{s}, t)$ is often modeled as a spatio-temporal GP, i.e., $w(\mathbf{s}, t) \sim \mathrm{GP}\left(0, C(\cdot, \cdot \mid \theta)\right)$

- $\varepsilon(\mathbf{s}, t)$ is the unstructured error

Continuous space-time models

- Extending spatial linear regression by connecting space and time

$$y(\mathbf{s},t) = \mu(\mathbf{s},t) + w(\mathbf{s},t) + \varepsilon(\mathbf{s},t)$$

- $\mu(\mathbf{s},t)$ is the mean structure, usually set to as $\mu(\mathbf{s},t) = \mathbf{x}'(\mathbf{s},t)\boldsymbol{\beta}(\mathbf{s},t)$;
  - $\boldsymbol{\beta}(\mathbf{s},t) = \boldsymbol{\beta}$; $\boldsymbol{\beta}(\mathbf{s},t) = \boldsymbol{\beta}(\mathbf{s})$; $\boldsymbol{\beta}(\mathbf{s},t) = \boldsymbol{\beta}(t)$

- $w(\mathbf{s},t)$ is the spatio-temporal random effect;
  - $w(\mathbf{s},t)$ is often modeled as a spatio-temporal GP, i.e., $w(\mathbf{s},t) \sim \mathrm{GP}\left(0, C(\cdot,\cdot \mid \theta)\right)$

- $\varepsilon(\mathbf{s},t)$ is the unstructured error
  - $\varepsilon(\mathbf{s},t) \sim \text{iid } \mathrm{N}(0,\tau^2)$.

Spatio-temporal covariance function

- Don't treat time as a third coordinate - scale issue, i.e., distance in space has nothing to do with distance on time scale

$$\mathrm{Cov}\left(w(\mathbf{s},t), w(\mathbf{s}',t')\right)$$

Spatio-temporal covariance function

- Don't treat time as a third coordinate - scale issue, i.e., distance in space has nothing to do with distance on time scale

$$\mathrm{Cov}\left(w(\mathbf{s},t), w(\mathbf{s}',t')\right)$$

- Weak stationarity:

$$\mathrm{Cov}\left(w(\mathbf{s},t), w(\mathbf{s}',t')\right) = \mathrm{Cov}(\mathbf{s}-\mathbf{s}', t-t')$$

Spatio-temporal covariance function

- Don't treat time as a third coordinate - scale issue, i.e., distance in space has nothing to do with distance on time scale

$$\mathrm{Cov}\left(w(\mathbf{s},t), w(\mathbf{s}',t')\right)$$

- Weak stationarity:

$$\mathrm{Cov}\left(w(\mathbf{s},t), w(\mathbf{s}',t')\right) = \mathrm{Cov}(\mathbf{s}-\mathbf{s}', t-t')$$

- Isotropy:

$$\mathrm{Cov}\left(w(\mathbf{s},t), w(\mathbf{s}',t')\right) = \mathrm{Cov}(||\mathbf{s}-\mathbf{s}'||, |t-t'|)$$

Spatio-temporal covariance function

- Don't treat time as a third coordinate - scale issue, i.e., distance in space has nothing to do with distance on time scale

$$\text{Cov}\left(w(\mathbf{s},t), w(\mathbf{s}',t')\right)$$

- Weak stationarity:

$$\text{Cov}\left(w(\mathbf{s},t), w(\mathbf{s}',t')\right) = \text{Cov}(\mathbf{s}-\mathbf{s}', t-t')$$

- Isotropy:

$$\text{Cov}\left(w(\mathbf{s},t), w(\mathbf{s}',t')\right) = \text{Cov}(||\mathbf{s}-\mathbf{s}'||, |t-t'|)$$

- Separability:

$$\text{Cov}\left(w(\mathbf{s},t), w(\mathbf{s}',t')\right) = \sigma^2 \rho_1(\mathbf{s}-\mathbf{s}';\phi_1)\rho_2(t-t';\phi_2)$$

  where ρ_1 is a valid two-dimensional correlation function and ρ_2 is a valid one-dimensional correlation function.

  - Separable models do **not** allow for space-time interaction

Spatio-temporal covariance function
Non-separable models

- Non-separable covariance function (e.g., De Iaco et al. 2002; Gneiting 2002; Stein 2005) allows space-time interaction

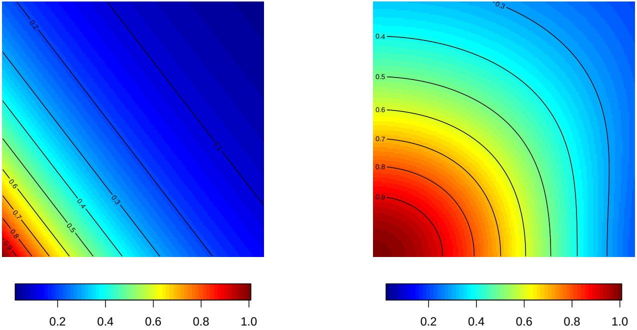


Figure: Separable vs non-separable correlation function.

Spatio-temporal models with GPs

- For locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and times $1, \ldots, T$

Spatio-temporal models with GPs

- For locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and times $1, \ldots, T$
- $\mathbf{w}(\mathbf{s}_i) = (w(\mathbf{s}_i, 1), \ldots, w(\mathbf{s}_i, T))'$, $\mathbf{w} = (\mathbf{w}'(\mathbf{s}_1), \ldots, \mathbf{w}'(\mathbf{s}_n))$

Spatio-temporal models with GPs

- For locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and times $1, \ldots, T$
- $\mathbf{w}(\mathbf{s}_i) = (w(\mathbf{s}_i, 1), \ldots, w(\mathbf{s}_i, T))'$, $\mathbf{w} = (\mathbf{w}'(\mathbf{s}_1), \ldots, \mathbf{w}'(\mathbf{s}_n))$
- $\mathbf{R}(\phi_1)_{ii'} = \rho_1(\mathbf{s}_i - \mathbf{s}_{i'} \mid \phi_1)$ and $\mathbf{R}(\phi_2)_{tt'} = \rho_2(t - t' \mid \phi_2)$

Spatio-temporal models with GPs

- For locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and times $1, \ldots, T$
- $\mathbf{w}(\mathbf{s}_i) = (w(\mathbf{s}_i, 1), \ldots, w(\mathbf{s}_i, T))'$, $\mathbf{w} = (\mathbf{w}'(\mathbf{s}_1), \ldots, \mathbf{w}'(\mathbf{s}_n))$
- $\mathbf{R}(\phi_1)_{ii'} = \rho_1(\mathbf{s}_i - \mathbf{s}_{i'} \mid \phi_1)$ and $\mathbf{R}(\phi_2)_{tt'} = \rho_2(t - t' \mid \phi_2)$
- The covariance matrix of $\mathbf{w}$ is

$$\Sigma_w(\boldsymbol{\theta}) = \sigma^2 \mathbf{R}(\phi_1) \otimes \mathbf{R}(\phi_2)$$

  where $\otimes$ is the Kronecker product (separable model).
- Fast likelihood evaluation by properties of Kronecker.

Spatio-temporal models with GPs

- For locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and times $1, \ldots, T$
- $\mathbf{w}(\mathbf{s}_i) = (w(\mathbf{s}_i, 1), \ldots, w(\mathbf{s}_i, T))'$, $\mathbf{w} = (\mathbf{w}'(\mathbf{s}_1), \ldots, \mathbf{w}'(\mathbf{s}_n))$
- $\mathbf{R}(\phi_1)_{ii'} = \rho_1(\mathbf{s}_i - \mathbf{s}_{i'} \mid \phi_1)$ and $\mathbf{R}(\phi_2)_{tt'} = \rho_2(t - t' \mid \phi_2)$
- The covariance matrix of $\mathbf{w}$ is

$$\Sigma_w(\boldsymbol{\theta}) = \sigma^2 \mathbf{R}(\phi_1) \otimes \mathbf{R}(\phi_2)$$

  where $\otimes$ is the Kronecker product (separable model).
- Fast likelihood evaluation by properties of Kronecker.
- Spatio-temporal predictions (kriging) through $w(\mathbf{s}_0, t_0) \mid \mathbf{w}$

Discrete-time spatial models

$$y_t(\mathbf{s}) = \mu_t(\mathbf{s}) + w_t(\mathbf{s}) + \varepsilon_t(\mathbf{s})$$

- $\mu_t(\mathbf{s})$ is the mean structure, usually set to as $\mu_t(\mathbf{s}) = \mathbf{x}_t'(\mathbf{s})\boldsymbol{\beta}_t(\mathbf{s})$;
- $w_t(\mathbf{s})$ is the spatio-temporal random effect;
- $\varepsilon_t(\mathbf{s})$ is the unstructured error

Discrete-time spatial models

$$y_t(\mathbf{s}) = \mu_t(\mathbf{s}) + w_t(\mathbf{s}) + \varepsilon_t(\mathbf{s})$$

- $\mu_t(\mathbf{s})$ is the mean structure, usually set to as $\mu_t(\mathbf{s}) = \mathbf{x}_t'(\mathbf{s})\boldsymbol{\beta}_t(\mathbf{s})$;
- $w_t(\mathbf{s})$ is the spatio-temporal random effect;
- $\varepsilon_t(\mathbf{s})$ is the unstructured error
  - $\varepsilon_t(\mathbf{s}) \sim \text{iid N}(0, \tau_t^2)$.

Discrete-time spatial models

$$y_t(\mathbf{s}) = \mu_t(\mathbf{s}) + w_t(\mathbf{s}) + \varepsilon_t(\mathbf{s})$$

- $\mu_t(\mathbf{s})$ is the mean structure, usually set to as $\mu_t(\mathbf{s}) = \mathbf{x}_t'(\mathbf{s})\boldsymbol{\beta}_t(\mathbf{s})$;
- $w_t(\mathbf{s})$ is the spatio-temporal random effect;
  - *additive form* (or multiplicative) $w_t(\mathbf{s}) = \alpha_t + v(\mathbf{s})$
  - *independent-in-time* spatial processes $w_t(\mathbf{s}), \forall t$
  - **dynamic spatio-temporal models**

    $$w_t(\mathbf{s}) = w_{t-1}(\mathbf{s}) + \eta_t(\mathbf{s}),$$

    where $\eta_t(\mathbf{s})$ are independent-in-time spatial processes
- $\varepsilon_t(\mathbf{s})$ is the unstructured error
  - $\varepsilon_t(\mathbf{s}) \sim \text{iid N}(0, \tau_t^2)$.

Discrete-time spatial models

Dynamic spatio-temporal models

- State-space formulation (West and Harrison, 1997)
- $\mathbf{y}_t = (y_t(\mathbf{s}_1), \ldots, y_t(\mathbf{s}_n))'$

Discrete-time spatial models

Dynamic spatio-temporal models

- State-space formulation (West and Harrison, 1997)
- $\mathbf{y}_t = (y_t(\mathbf{s}_1), \ldots, y_t(\mathbf{s}_n))'$

  1. Stage: Measurement equation

  $$\mathbf{y}_t = f(\mathbf{z}_t \mid \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_t,$$

  where $\mathbf{z}_t$ is the state vector, not observable but generated by a first-order Markovian process and $\boldsymbol{\varepsilon}_t$ is a white noise process

Discrete-time spatial models

Dynamic spatio-temporal models

- State-space formulation (West and Harrison, 1997)
- $\mathbf{y}_t = (y_t(\mathbf{s}_1), \ldots, y_t(\mathbf{s}_n))'$

1. Stage: Measurement equation

$$\mathbf{y}_t = f(\mathbf{z}_t \mid \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{z}_t$ is the state vector, not observable but generated by a first-order Markovian process and $\boldsymbol{\varepsilon}_t$ is a white noise process

2. Stage: Transition equation

$$\mathbf{z}_t = g(\mathbf{z}_{t-1} \mid \boldsymbol{\gamma}) + \boldsymbol{\eta}_t,$$

where $\boldsymbol{\eta}_t$ is a discrete-time spatial process.

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, Boca Raton, FL, second edition.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812.

De Iaco, S., Myers, D. E., and Posa, D. (2002). Nonseparable space-time covariance models: Some parametric families. *Mathematical Geology*, 34:23 – 42.

Finley, A. O., Banerjee, S., and Gelfand, A. E. (2015). spbayes for large univariate and multivariate point-referenced spatio-temporal data models. *Journal of Statistical Software*, 63(13):1–28.

Gasperoni, F., Luati, A., Paci, L., and D'Innocenzo, E. (2021). Score driven modeling of spatio-temporal data. *Journal of the American Statistical Association*, in press.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97(458):590–600.

Liu, Z., Berrocal, V. J., Bartsch, A. J., and Johnson, T. D. (2016). Pre-surgical fMRI data analysis using a spatially adaptive conditionally autoregressive model. *Bayesian Analysis*, 11(2):599–625.

Paci, L., Beamonte, M. A., Gelfand, A. E., Gargallo, P., and Salvador, M. (2017). Analysis of residential property sales using space–time point patterns. *Spatial Statistics*, 21:149 – 165.

Paci, L., Gelfand, A. E., and Cocchi, D. (2015). Quantifying uncertainty for temperature maps derived from computer models. *Spatial Statistics*, 12:96 – 108.

Ranciati, S., Viroli, C., and Wit, E. C. (2017). Mixture model with multiple allocations for clustering spatially correlated observations in the analysis of ChIP-Seq data. *Biometrical Journal*, 59(6):1301–1316.

Stein, M. L. (2005). Space–time covariance functions. *Journal of the American Statistical Association*, 100(469):310–321.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer, 2nd edition.