Lecture 1: Introduction

Spatial statistics and image analysis

David Bolin
University of Gothenburg

Gothenburg
March 25, 2019

Course litterature

The course is mainly based on:

- Lecture notes by Mats Rudemo.

More details are found in:

- Handbook of Spatial Statistics by Gelfand et. al.
- Elements of statistical learning by Hastie et. al.
- Computer Age Statistical Inference by Efron and Hastie.

The books are available as eBooks, see homepage.
In the schedule, the relevant chapters are indicated for each lecture.

Practical information

Teachers:

David Bolin:	Lecturer and examiner
Room:	H3028
E-mail:	david.bolin@chalmers.se

Homepage:
`www.math.chalmers.se/Stat/Grundutb/CTH/tms016/1819/`

Schedule:

Lectures:	Mondays and Wednesdays (10-12)
Compute exercices:	Mondays and Wednesdays (13-15)

The lectures will cover the theory, which you will use in practice in the computer exercise directly after each lecture.

Examination

There will be two components in the examination:

- Written exam at the end of the course
- Project assignment.

these are weighted equally for the final grade.

Successful completion of the course will be rewarded by 7.5 hp.

The project:

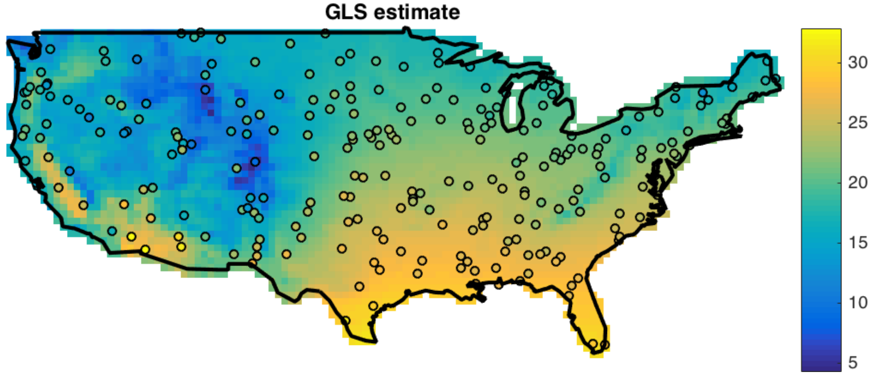
- can be in groups of 1-3 students.
- will consist of three parts: two problems introduced in the computer exercises and one problem you can choose on your own (with approval from me).
- Is presented at a seminar and as a written report at the end of the course.

Contents

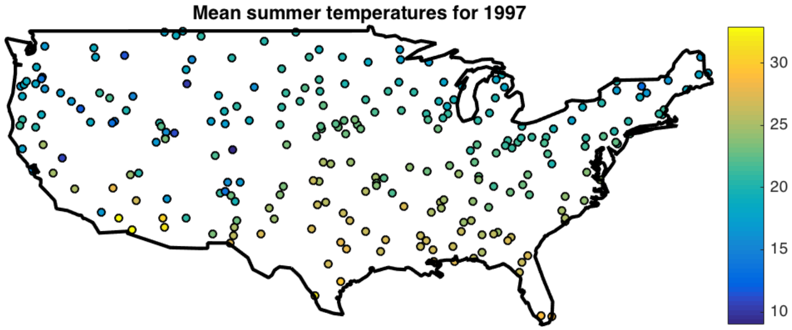
- Traditional method from spatial statistics.
- Statistical and machine learning methods for image analysis.
- Application areas:
  - Image analysis
  - climate science
  - environmental statistics
  - remote sensing
  - microscopy
  - medical imaging and fMRI
  - Disease mapping
  - +++

Kriging estimation

Using a statistical model, where we assume that there observations are noisy observations of the true temperatures, we obtain

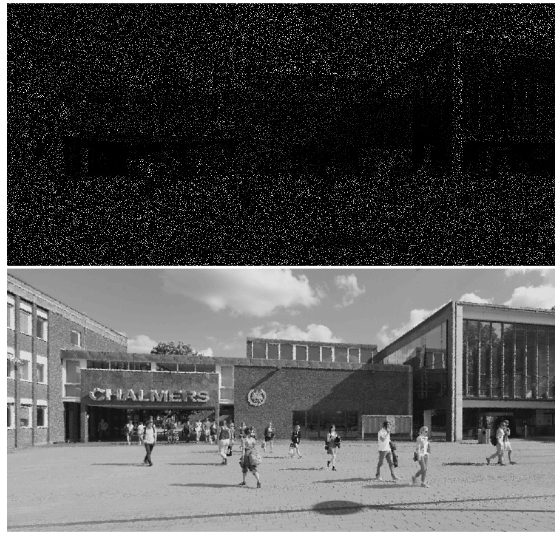


A common problem in geostatistics



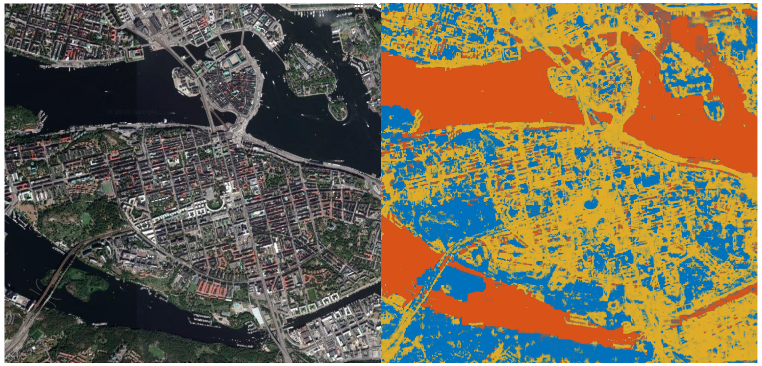
- Mean summer temperatures (June-August) in the continental US 1997 recorded at 250 weather stations.
- We want to estimate all US temperatures based on the data.

Image reconstruction

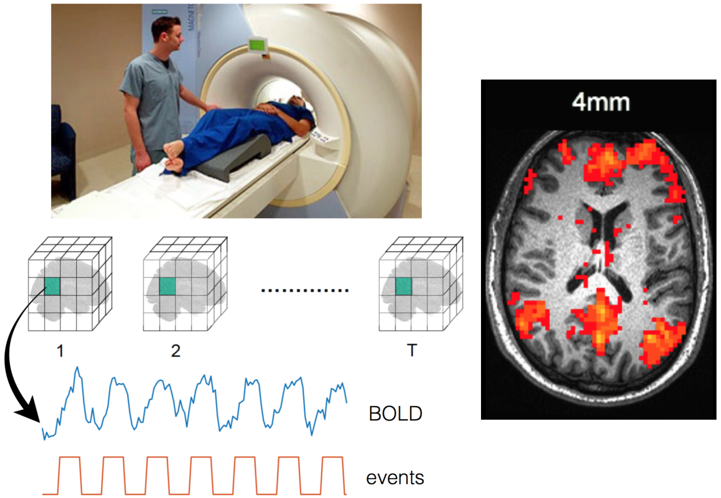


Noise reduction

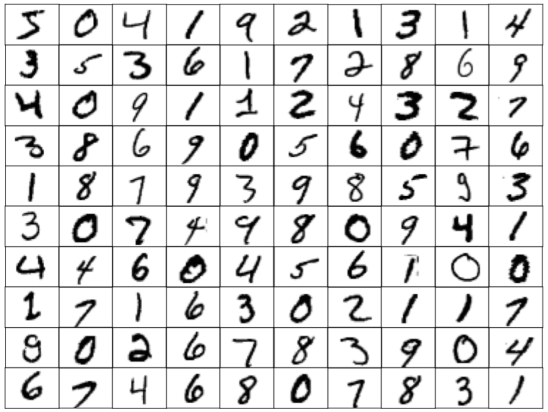
Segmentation



Brain imaging



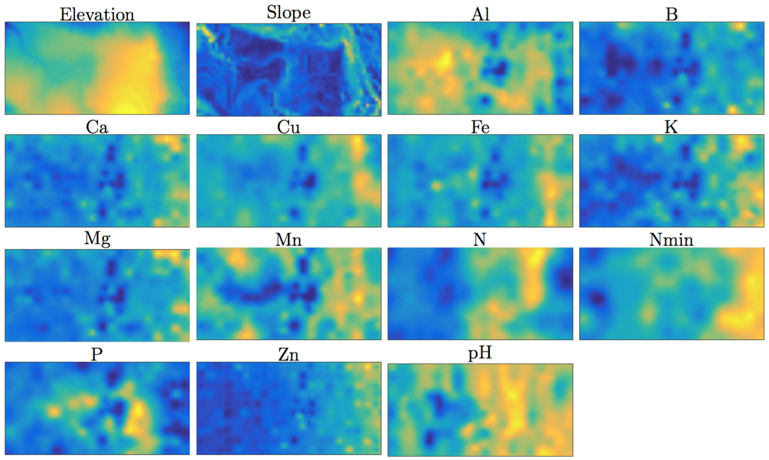
Classification



Puppy or bagel?

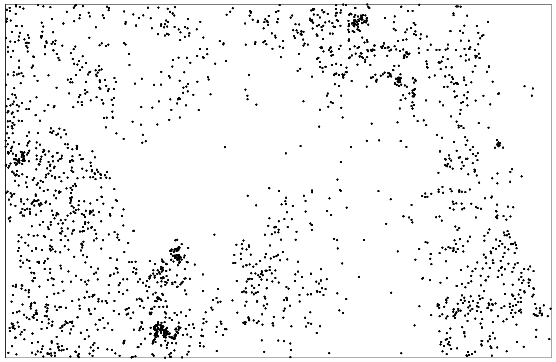
See `twistedsifter.com/2016/03/puppy-or-bagel-meme-gallery/` for more important classification problems.

Point processes



Possible covariates that can be used for drawing conclusions on the association of habitat preferences.

Point processes



The locations of the tree species Beilschmiedia Pendula in the tropical rainforest plot on Barro Colorado Island.

Outline of course

Current plan for lectures:

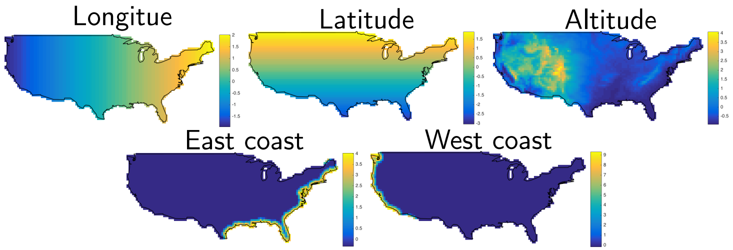
- 1 Introduction and background
- 2 Gaussian random fields
- 3-4 Kriging and parameter estimation
- 5 Gaussian Markov random fields
- 6-7 Image segmentation and mixture models
- 8-9 Discrete Markov random fields
- 10-11 Machine learning methods and neural nets
- 12 Point processes
- 13 Recap
- 14-15 Project seminars

Example: Interpolation of the temperature data

- A first idea is to use linear regression to interpolate the data:

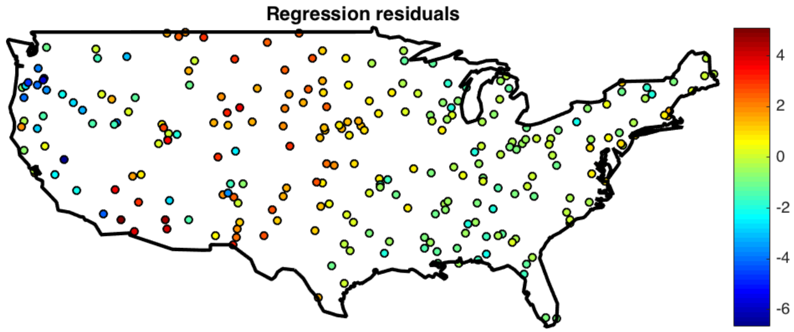
$$Y(\mathbf{s}) = \sum_{i=1}^{k} \beta_i B_i(\mathbf{s}) + \varepsilon_{\mathbf{s}}, \quad \text{where } \varepsilon_{\mathbf{s}} \text{ are iid } \mathsf{N}(0, \sigma^2)$$

- Possible covariates



Residudals

- How do we test whether the prediction is reasonable?
- If the model assumptions hold, the residuals $Y(\mathbf{s}) - \hat{X}(\mathbf{s})$ should be independent identically distributed.



OLS estimate

- Estimate the parameters using ordinary least squares:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{B}\boldsymbol{\beta}\| \quad \Rightarrow \quad \hat{\boldsymbol{\beta}} = (\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{Y},$$

  where $\mathbf{B}_{ij} = B_i(\mathbf{s}_j)$ and $\mathbf{Y}_i = Y(\mathbf{s}_i)$.
- Calculate the prediction $\hat{X}(\mathbf{s}) = \sum_{i=1}^{k} \hat{\beta}_i B_i(\mathbf{s})$.