Using R to statistically solve an engineering problem

Reinhard Furrer, UZH
LV-ID.UZH, 13-11-06
Big picture

Spatial statistics

An engineering problem

A “statistical” solution
Big picture: statistics

Mathematical Statistics ← Statistics → Applied Statistics

2013 – International Year of Statistics

Probability Theory
WIKIPEDIA: The branch of mathematics concerned with probability, the analysis of random phenomena.

Statistics
WIKIPEDIA: The study of the collection, organization, analysis, interpretation and presentation of data.

Test Theory
WIKIPEDIA: A method of making decisions using data from a scientific study.

Estimation Theory
WIKIPEDIA: Estimating the values of parameters based on measured data that has a random component.

Source: Carsten Magnus
Big picture: statistics

What is statistics:

“Statistics is the science of extracting information from data to solve real world problems”

or:

“Applied statistics is the science and art of extracting information from data to solve real world problems”
Big picture: missing link

Source: www.cisl.ucar.edu

Source: wikipedia.org

Source: www.cisl.ucar.edu

Source: allvoices.com, artvehicle.com, centralasiaonline.com

Source: wikipedia.org

Source: wikipedia.org

Source: wikipedia.org
Big picture: missing link

Source: wikipedia.org
Big picture: spatial statistics

First law of geography (Waldo Tobler):

“Everything is related to everything else, but near things are more related than distant things.”

Examples:

- temperature,
- photosynthetic activity,
- mRNA hybridization,
- enemy attack intensity: in space, in time, in space-time

Exploit correlation for prediction (interpolation, extrapolation)
Spatial statistics: prediction

Observations: \( y(s_1), \ldots, y(s_n) \)

Model:
\[
Y(s) = \text{signal} + \text{noise}
\]
\[
Y(s) = \text{trend} + \text{stochastic part} + \text{noise}
\]
\[
Y(s) = X(s) + \epsilon(s)
\]
Spatial statistics: prediction

Predict the quantity of interest at an arbitrary location.

Why?

▶ Fill-in missing data
▶ Force data onto a regular grid
▶ Smooth out the measurement error

How?

▶ By eye
▶ Linear interpolation
▶ The correct way . . .
Spatial statistics: prediction

Describing the covariance structure

Covariance matrix $\Sigma$ contains elements $C(\text{dist}(s_i, s_j))$. 
Spatial statistics: prediction

Predict $X(s_0)$ given $y(s_1), \ldots, y(s_n)$.

Minimize mean squared prediction error (over all linear unbiased predictors)

$\rightarrow$ Best Linear Unbiased Predictor:

$$BLUP = \text{Cov}[X(s_{\text{predict}}), Y(s_{\text{obs}})] \cdot \text{Var}[Y(s_{\text{obs}})]^{-1}_{\text{obs}}$$

$$\hat{X}(s_0) = c^T \Sigma^{-1} y$$

(one spatial process, no trend; otherwise almost the same)
Applied statistician’s workflow

Solving real world problems (art of statistics):

Scientific problem

exploratory data analysis (EDA) ("where to go")
theoretical aspects ("allowed to do")
simulation ("able to do")
application to dataset ("useful to do")
Engineering problem

Compaction for road construction:

- one vibrating drum (smooth drum or padfoot)
- rolling at 1m/s
- 20cm material per layer
- typical bed is 12–15m wide and 30–150m long
- sufficient compaction is ‘manually’ tested after several layers of material (USA)

Is an automatic quality assurance and intelligent compaction possible?
Engineering problem

Current “intelligent” compaction:

- precise GPS positioning
- on-board visualization
- off-board processing
Engineering problem

- Relation between measurement and actual soil modulus is unknown.

- (Linear) relationship is determined with a second measurement device (lightweight deflection, nuclear density ...
Applied statistician’s workflow

- EDA: get to know your data inside and out
- Model: spatial, additive mixed effects model
- Probe the model: do we get what we need?
- Application: solve the problem
Sparse covariance matrices

Calculate $\Sigma$:  

Distances:

Covariance( Distances ):
Sparse covariance matrices

Sparseness is guaranteed when:

- the covariance function has a compact support
- a compact support is (artificially) imposed $\Rightarrow$ tapering

![Graph showing covariance functions](image)

Covariance for

- exponential
- spherical
- exponential * spherical

![Color map](image)
Backfitting

Minnesota testbed:

Subgrade, subbase, base (top to bottom).
Backfitting

Minnesota testbed:

Model:

\[
Y_1(s) = \mathbf{X}_1\beta_1 + \gamma_1(s) + \varepsilon_1(s)
\]

\[
Y_2(s) = \mathbf{X}_2\beta_2 + c\gamma_1(s) + \gamma_2(s) + \varepsilon_2(s)
\]

\[
Y_3(s) = \mathbf{X}_3\beta_3 + c^2\gamma_1(s) + c\gamma_2(s) + \gamma_3(s) + \varepsilon_3(s)
\]

Extending the ‘classical’ backfitting approach to dependent data:

repeat until convergence

repeat until convergence

estimate fixed effects

for all ‘stochastic’ effects

estimate parameters

predict smooth field
Backfitting

Minnesota testbed:

Model:

\[
Y_1(s) = X_1 \beta_1 + \gamma_1(s) + \varepsilon_1(s)
\]
\[
Y_2(s) = X_2 \beta_2 + c \gamma_1(s) + \gamma_2(s) + \varepsilon_2(s)
\]
\[
Y_3(s) = X_3 \beta_3 + c^2 \gamma_1(s) + c \gamma_2(s) + \gamma_3(s) + \varepsilon_3(s)
\]

Cell 27:
Fitted smooths:
Multiresolution analysis

- Simultaneous inference for IC and QA
  - “which features are ‘really there’?”

- Using Holmström et al. (2010), extensions from original SiZER
  - Ready-to-use software

- Decomposition into different scales
  - What scales to use?
Multiresolution analysis

Minnesota testbed:

Subgrade, subbase, base (top to bottom).
Multiresolution analysis

Cell 27:
Fitted smooths:

Multiresolution analysis:

subgrade
subbase
base
Afterthoughts/outlook

- Flexible setting . . . methodological toolbox

- Pick and extend code . . . software toolbox

- Using toolboxes to solve other real world problems!
Collaboration with:

— Daniel Heersink, now Research Scientist at CSIRO, Canberra
— Mike Mooney, CSM
— Roland Anderegg, FHNW . . .

URPP Systems Biology / Functional Genomics
References


