[9] B. D. Ripley. *Spatial Statistics*. Wiley, 1981. 15

[10] P. Whittle. On stationary processes in the plane. *Biometrika*, 41:434–449, 1954. 15

Paulo J. Ribeiro Jr
Universidade Federal do Paraná, Brasil
and *Lancaster University, UK*
Paulo.Ribeiro@est.ufpr.br

Peter J. Diggle
Lancaster University, UK
p.diggle@lancaster.ac.uk

Simulation and Analysis of Random Fields

by Martin Schlather

Random fields are the d-dimensional analogues of the one-dimensional stochastic processes; they are used to model spatial data as observed in environmental, atmospheric, and geological sciences. They are traditionally needed in mining and exploration to model ore deposits, oil reservoirs, and related structures.

The contributed package **RandomFields** allows for the simulation of Gaussian random fields defined on Euclidean spaces up to dimension 3. It includes some geostatistical tools and algorithms for the simulation of extreme-value random fields.

In the following two sections we give an example of an application, and a summary of the features of **RandomFields**.

A brief geostatistical analysis

To demonstrate key features we consider soil moisture data collected by the Soil Physics Group at the University of Bayreuth (see Figure 1), and perform a simple geostatistical analysis.

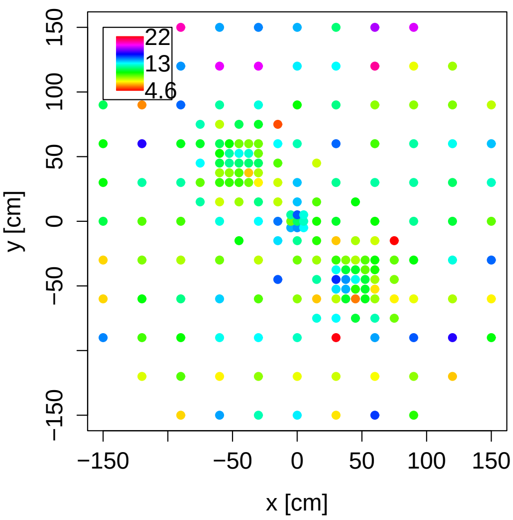

Figure 1: Measured soil moisture content (%)

The coordinates of the sample locations are assumed to be stored in `pts` and the moisture measurements in `d`. (See the example to the data set `soil` for the complete code.)

In geostatistics the variogram γ is frequently used to describe the spatial correlation structure. It can be expressed in terms of the (auto-)covariance function C if the random field is stationary: $\gamma(h) = C(0) - C(h)$ for $h \in \mathbb{R}^d$. We assume isotropy, i.e. γ depends only on the distance $|h|$. Then, we can find variogram values by

```
ev <- EmpiricalVariogram(pts, data=d,
        grid=FALSE,
        bin=c(-1, seq(0, 175, l=20)))
```

and select an appropriate model by

```
ShowModels(0:175, x=x, y=y, emp=ev)
```

see Figure 4 where `x` and `y` equal `seq(-150, 150, l=121)`. The parameters of the variogram model (here the Whittle-Matérn model) might be fitted by eye, but we prefer maximum likelihood,

```
p <- mleRF(pts, d, "whittle", param=rep(NA,5),
           lower.k=0.01, upper.k=30).
```

Now,

```
Kriging("O", x=x, y=y, grid=TRUE,
        model="whittle", par=p, given=pts,
        data=d)
```

yields the expected moisture content on the grid given by `x` and `y` (Figure 2).

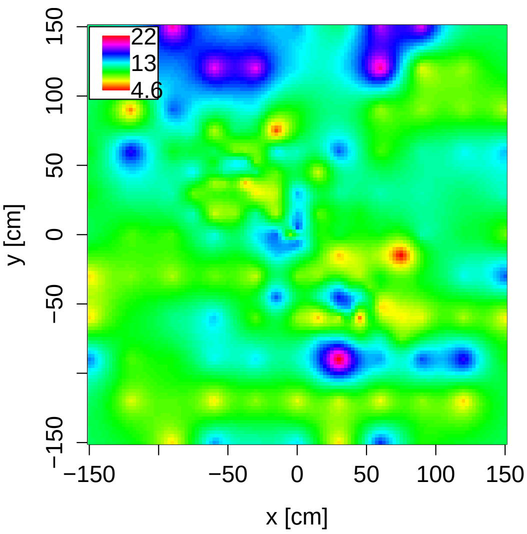

Figure 2: Kriged field

Conditional simulation, see Figure 3, allows for further inference.

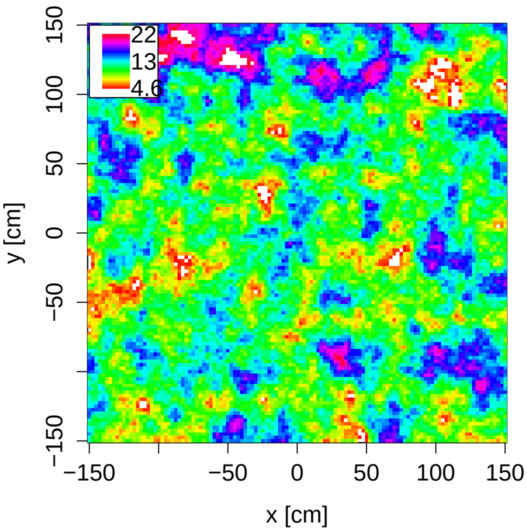

Figure 3: Conditional simulation

For instance, we may ask for the conditional probability, given the data at the respective locations, that the maximal moisture content is not greater than 24%. To this end, a random sample, here of size 100, is drawn from the conditional distribution

```
CondSimu("O", x=x, y=y, grid=TRUE, n=100,
         model="whittle", param=p, given=pts,
         data=d),
```

which yields an estimate of 40%.

Package features

The currently implemented methods for the simulation of stationary and isotropic random fields include circulant embedding, turning bands, and methods based on matrix decomposition or Poisson point processes. Important features of the package are

- User friendly interface: depending on his/her preferences, the user can either specify the desired simulation method, or the package will automatically pick an appropriate technique, according to the variogram model, the dimension of the space, and the spatial configuration of the simulation domain (1).
- Increased speed if multiple simulations with identical specifications are performed. To this end, a specification is first compared to that of the preceding simulation. In case the parameters match, the stored results of the deterministic part of the algorithm are used instead of being recalculated.
- The function `ShowModels` is an instructive tool for teaching, which allows for the interactive choice of variogram models and parameters. For each specification the graph of the variogram model and one- or two-dimensional simulations are provided. `ShowModels` can also be used to fit an empirical variogram by eye.

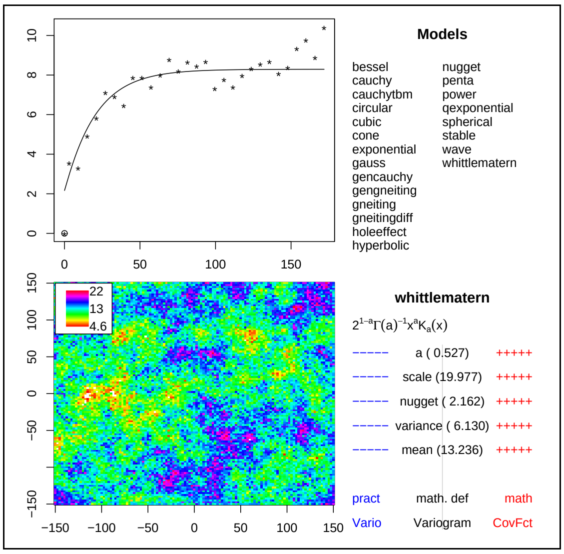

Figure 4: Snapshot of the interactive plot `ShowModels`

Further functionalities include:

- variogram analysis: calculation of the empirical variogram and MLE parameter estimation
- conditional simulation based on simple or ordinary kriging
- checks whether the parameter specifications are compatible with the variogram model
- simulation of max-stable random fields

Future extensions may provide further simulation algorithms for Gaussian and non-Gaussian random fields, and a basic toolbox for the analysis of geostatistical and spatial extreme value data.

Use `help(RandomFields)` to obtain the main man page. To start with, the examples in `help(GaussRF)` are recommended.

Acknowledgement. The work has been supported by the EU TMR network ERB-FMRX-CT96-0095 on "Computational and statistical methods for the analysis of spatial data" and the German Federal Ministry of Research and Technology (BMFT) grant PT BEO 51-0339476C. The author is thankful to Tilmann Gneiting, Martin Mächler, and Paulo Ribeiro for hints and discussions.

Bibliography

[1] M. Schlather. An introduction to positive definite functions and to unconditional simulation of random fields. Technical Report ST-99-10, Lancaster University, 1999. 19

Martin Schlather
University of Bayreuth, Germany
Martin.Schlather@uni-bayreuth.de

mgcv: GAMs and Generalized Ridge Regression for R

by Simon N. Wood

Generalized Additive Models (GAMs) have become quite popular as a result of the work of Wahba (1990) and co-workers and Hastie & Tibshirani (1990). Package **mgcv** provides tools for GAMs and other generalized ridge regression. This article describes how GAMs are implemented in **mgcv**: in particular the innovative features intended to improve the GAM approach. The package aims to provide the convenience of GAM modelling in S-PLUS, combined with much improved model selection methodology. Specifically, the degrees of freedom for each smooth term in the model are chosen simultaneously as part of model fitting by minimizing the Generalized Cross Validation (GCV) score of the whole model (not just component wise scores). At present **mgcv** only provides one dimensional smooths, but multi-dimensional smooths will be available from version 0.6, and future releases will include anisotropic smooths. GAMs as implemented in **mgcv** can be viewed as low rank approximations to (some of) the generalized spline models implemented in **gss** — the idea is to preserve most of the practical advantages with which elegant underlying theory endows the generalized smoothing spline approach, but without the formidable computational burden that accompanies full gss models of moderately large data sets.

GAMs in mgcv

GAMs are represented in **mgcv** as penalized generalized linear models (GLMs), where each smooth term of a GAM is represented using an appropriate set of basis functions and has an associated penalty measuring its wiggliness: the weight given to each penalty in the penalized likelihood is determined by its "smoothing parameter". Models are fitted by the usual iteratively re-weighted least squares scheme for GLMs, except that the least squares problem at each iterate is replaced by a penalized least squares problem, in which the set of smoothing parameters must be estimated alongside the other model parameters: the smoothing parameters are chosen by GCV. This section will sketch how this is done in a little more detail.

A GLM relating a univariate response variable y to a set of explanatory variables $x_1, x_2, \ldots$, has the general form:

$$g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots \qquad (9.1)$$

where $E(y_i) \equiv \mu_i$ and the y_i are independent observations on r.v.s all from the same member of the exponential family. g is a smooth monotonic "link-function" that allows a useful degree of non-linearity into the model structure. The β_i are model parameters: likelihood theory provides the means for estimation and inference about them. The r.h.s. of (9.1) is the "linear predictor" of the GLM, and much of the statistician's modelling effort goes into finding an appropriate form for this.

The wide applicability of GLMs in part relates to the generality of the form of the of the linear predictor: the modeller is not restricted to including explanatory variables in their original form, but can include transformations of explanatory variables and dummy variables in whatever combinations are appropriate. Hence the class of models is very rich, including, for example, polynomial regression models and models for designed experiments. However the