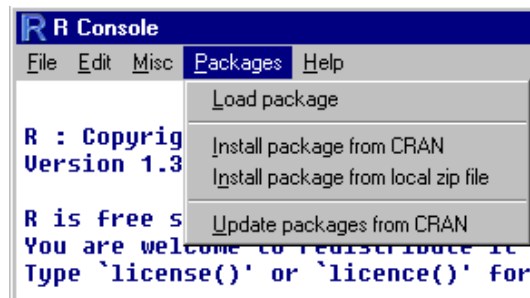


simple way to install them. Another function of `rwinst.exe` was to provide a point-and-click way to install pre-compiled packages. Now `rwinst.exe` is no longer supported<sup>3</sup>, we needed another way to ease the installation of packages. This is provided by the Packages menu introduced in R 1.3.0. This takes advantage of the new facilities to download files from URLs to allow installation of packages either from a local zip file or directly from a CRAN node.



Other items on that menu provide shortcuts to `library()` and `update.packages()`.

Brian D. Ripley  
University of Oxford, UK  
[ripley@stats.ox.ac.uk](mailto:ripley@stats.ox.ac.uk)

## Spatial Statistics in R

by Brian D. Ripley

The following two articles discuss two recent spatial statistics packages for R, and the Editors suggest that I write an overview of the area.

There are several packages available, almost all of which originated as S code and from quite different communities. As a result there is considerable overlap in their functionality. Many are quite old (pre-dating classes in S, for example, and from an era of much lower levels of computing resources). In roughly historical order there are

**akima** An R interface by Albrecht Gebhardt to spatial spline interpolation Fortran code by H. Akima. Closely modelled on the `interp` function in S-PLUS.

**tripack** Delaunay triangulation of spatial data. An R interface by Albrecht Gebhardt to Fortran code by R. J. Renka.

**spatial** Now part of the **VR** bundle. Contains trend-surface analysis, kriging and point-process code originally written by B. D. Ripley in 1990–1 for teaching support.

**VR** version 6.2-6 includes enhanced trend-surface code contributed by Roger Bivand.

**sgeostat** Geostatistical modelling code written for S by James J. Majure and ported to R by Albrecht Gebhardt.

**splancs** Originally commercial code for spatial and space-time point patterns by Barry Rowlingson. Roger Bivand has made a GPL-ed version available for R, with contributions from Giovanni Petris.

**spatstat** Code for point pattern analysis originally written for S-PLUS 5.1 by Adrian Baddeley and Rolf Turner. The version on the website (<http://www.maths.uwa.edu.au/~adrian/spatstat.html>) is said to work with R 1.0.1.

**geoR** Geostatistical code by Paulo J. Ribeiro. See the next article. Paulo does not mention that he also has a **geoS**.

**RandomFields** Specialized code to simulate from continuous random fields by Martin Schlather. See the next but one article.

Which should you use? Spatial statistics is not a single subject, and it depends which part you want. It is natural for me to use the classification of Ripley (1981).

**Smoothing and interpolation** is still a range of methods. Package **akima** provides one reasonable interpolation method, and package **spatial** provides the commonest of smoothing methods, trend surfaces. Kriging can be either interpolation or smoothing, and is covered at various depths in **spatial**, **sgeostat** and **geoR**.

For **spatial autocorrelation** there is nothing available yet. For spatial lattice process (e.g. CAR processes) there is no specialized code, but `gls` from **nlme** could be used.

For **spatial point patterns**, **spatial** provides the basic tools, and methods to fit Strauss processes. **splancs** handles polygonal boundaries, and it and **spatstat** have a wider range of methods (by no means all of which I would recommend).

<sup>3</sup>it can still be compiled from the sources.

It is worth mentioning the commercial module **S+SpatialStats** for S-PLUS, which covers all the areas mentioned here, and is **not** available for R. The prospect of such a module, and later of further development of it, has dampened enthusiasm for user-contributed spatial statistics code over much of the last decade. It is worth bearing in mind that a great deal of the current wealth of packages for R emanates from the work of users filling gaps they saw in S and its commercial offspring.

## Bibliography

- [1] B. D. Ripley. *Spatial Statistics*. Wiley, 1981.

Brian D. Ripley  
University of Oxford, UK  
[ripley@stats.ox.ac.uk](mailto:ripley@stats.ox.ac.uk)

# geoR: A Package for Geostatistical Analysis

by Paulo J. Ribeiro Jr and Peter J. Diggle

**geoR** is a package to perform geostatistical data analysis and spatial prediction, expanding the set of currently available methods and tools for analysis of spatial data in R. It has been developed at the Department of Mathematics and Statistics, Lancaster University, UK. A web site with further information can be found at: <http://www.maths.lancs.ac.uk/~ribeiro/geoR.html>.

Preliminary versions have been available on the web for the last two years. Based on users' feedback and on our own experiences, we judge that the package has been used mainly to support teaching material and to carry out data analysis and simulation studies for scientific publications.

Package **geoR** differs from the other R tools for geostatistical data analysis in following the model-based inference methods described in (3).

## Spatial statistics and geostatistics

*Spatial statistics* is the collection of statistical methods in which spatial locations play an explicit role in the analysis of data. The main aim of *geostatistics* is to model continuous spatial variation assuming a basic structure of the type  $Y(x) : x \in \mathbb{R}^d, d = 1, 2$  or  $3$  for a random variable  $Y$  of interest over a region. Characteristic features of geostatistical problems are:

- data consist of *responses*  $Y_i$  associated with *locations*  $x_i$  which may be non-stochastic, specified by the sampling design (e.g. a lattice covering the observation region  $A$ ), or stochastic but selected independently of the process  $Y(x)$ .
- in principle,  $Y$  could be determined from any location  $x$  within a continuous spatial region  $A$ .
- $\{Y(x) : x \in A\}$  is related to an unobserved stochastic process  $\{S(x) : x \in A\}$ , which we call the *signal*.

- scientific objectives include prediction of one or more functionals of the stochastic process  $\{S(x) : x \in A\}$ .

Geostatistics has its origins in problems connected with estimation of reserves in mineral exploration/mining (5). Its subsequent development, initially by Matheron and colleagues at École des Mines, Fontainebleau (8) was largely independent of "mainstream" spatial statistics. The term "kriging" was introduced to describe the resulting methods for spatial prediction. Earlier developments include work by Matérn (6, 7) and by Whittle (10). Ripley (9) re-casts kriging in the terminology of stochastic process prediction, and this was followed by significant cross-fertilisation during 1980's and 1990's (eg the *variogram* is now a standard statistical tool for analysing correlated data in space and/or time). However, there is still vigorous debate on practical issues such as how to perform inference and prediction, and the role of explicit probability models.

## The Gaussian model

The currently available functions on **geoR** assume a basic model specified by:

1. a signal  $S(\cdot)$  which is a stationary Gaussian process with
  - (a)  $E[S(x)] = 0$ ,
  - (b)  $\text{Var}\{S(x)\} = \sigma^2$ ,
  - (c)  $\text{Corr}\{S(x), S(x-u)\} = \rho(u)$ ;
2. the conditional distribution of  $Y_i$  given  $S(\cdot)$  is Gaussian with mean  $\mu + S(x_i)$  and variance  $\tau^2$ ;
3.  $Y_i : i = 1, \dots, n$  are mutually independent, conditional on  $S(\cdot)$ .