

RGeostats Manual

D. Renard, F. Ors, N. Desassis

August 20th, 2017

Abstract

This document constitutes the users manual for the package RGeostats. It gives an overall presentation of the package, developed using R language. For a more detailed description of each function, the reader will refer to its on-line documentation. Finally some tutorials are also available in the standard RGeostats distribution which enable the interested user to run some examples on provided data sets. You should refer to the Getting Started manual for installation of RGeostats package.

1 History of the RGeostats package

The Geostatistics Team of Centre de Géosciences of Mines ParisTech spent several years developing different commercial libraries or software in the past. Let us mention:

- *GEOSLIB*: the first geostatistical library in FORTRAN
- *BLUEPACK*: a geostatistical package that lasted over 10 years and was famous in most mining and oil companies over the world
- *SIMPACK*: a package dedicated to geostatistical stochastic simulations
- *HERESIM*: a package, developed jointly with IFPEN, based on the Pluri-gaussian simulation technique
- *ISATIS*: the geostatistical toolbox, developed jointly and commercialized by Géovariances

It is therefore a tradition for the Centre de Géostatistique to imagine, produce and commercialize the algorithms developed by scientists so that practitioner can apply these fancy techniques to the different fields in their own domain of activity, without having to bother writing lines of code.

However, these packages do not allow the user to modify the code in order to test new ideas or algorithms. This is the reason why, starting in late 1990's,

some researchers started introducing some algorithms in the framework of the R package. This was initiated with the *GEFA* package developed for the fisheries community, which uses geostatistical techniques to forecast the fish density by species and age. The user could benefit from all the advantages due to the large number of contributors of R developers, combined to the procedures established by the researchers of the Centre de Géostatistique.

As usual with writing packages using the R interpreted language, the GEFA package needed to be strongly improved for improving the calculation speed. This usually involves writing pieces of the package in using a compiled language (such as C or C++).

For that reason, the package *RGeoS* was created in the year 2001 containing a set of R objects to manipulate data, parameters and results. The package *RGeoS* is based on a library of geostatistical code written in C (and C++) called *Geoslib*.

Recently, the package *RGeoS* has been renamed ***RGeostats*** during the year 2014 to better explain its contents and avoid conflict with packages with similar names.

The main characteristics of the RGeostats package is to perform geostatistical operations simultaneously on several (p) variables in a space of any dimension (n): we will designate this package as R_n-R_p package. However, some techniques are not defined for any space dimension, nor any number of variables treated simultaneously: a special test restricts their usage.

2 What is RGeostats ?

RGeostats is provided as a binary R package for Windows, Linux and Mac platform . It provides:

- all the R procedures with the corresponding on-line help (use the command `?func_name` to access the on-line help of the function `func_name`)
- the object library of the geostatistical code Geoslib
- some demonstration case studies: the user can run them using the command `demo()` with the available data sets.

Note that the *source code corresponding to the Geoslib library is not available*.

3 Who can use RGeostats ?

RGeostats can be downloaded by anyone.

RGeostats can be used free of charge in a non-commercial use.

4 The reference to RGeostats

When you use this software for publication, please use the following reference:

Renard D., Bez N., Desassis N., Beucher H., Ors F., Freulon X.
RGeostats: The Geostatistical package [version number].
MINES ParisTech.
Free download from: <http://cg.ensmp.fr/rgeostats>

5 Where can I find RGeostats ?

The package RGeostats must be downloaded from the web site of the Centre de Géosciences of Mines ParisTech:

<http://cg.ensmp.fr/rgeostats>

This site contains several directories, such as:

- the user community Board where you can:
 - Download RGeostats package (according to the Operating System where you want to use the package): this operation requires that you register to the Board first
 - Ask any question about any issue you may encounter
 - Learn on how to use specific parts of the package by reviewing the corresponding Tutorial
- the Documentation directory where you can find several case studies provided as vignettes
- the Demo directory where you can find several demonstration scripts
- the Function directory where you can find the on-line help for all functions

The package is provided for few platforms. For each platform, RGeostats is provided as a single file in an archive format. The extension of the archive file depends upon the platform:

- *Windows* 32 or 64 bits: file with extension *zip*
- *LINUX 32-bits*: file with extension *linux32.tar.gz*
- *LINUX 64-bits*: file with extension *linux64.tar.gz*
- *Mac*: file with extension *tgz*

6 Installing the RGeostats package

6.1 Installing R

The package R must be installed first. R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. This package can be downloaded from the site:

`http://cran.r-project.org`

If available for your Operating System, it is easier to install directly the dedicated binary version. Otherwise, one can always download the source code, configure it and compile it. Then please follow official information provided on the site.

The installation requires the Administrator rights.

6.2 Required package

The package Rcpp is required and can be downloaded from the CRAN web site.

6.3 Additional contributions

Moreover, some additional contributions can be downloaded (from the same site): such as *maps* and *mapproj* which are only necessary in some parts of the package *RGeostats* and will be only needed upon request.

Each extension comes as an archive file.

6.4 Installing an additional contribution

When installing a package, one may choose between:

- installing it as a permanent extension of R: this operation requires the Administrator rights as the RGeostats add-on package is written on the directory where R distribution is installed. The installation is performed by typing:

```
R CMD INSTALL mypack
```

- installing it as a personal extension: this is the case when an extension often varies. This installation does not require the Administrator rights. The package is installed on a user's dedicated directory (say *my_dir*) by typing:

```
R CMD INSTALL mypack --library=my_dir
```

7 Getting started with RGeostats

7.1 Loading the package

You must first start R in a working directory. You may have one working directory by project. You launch R by typing the corresponding command (or clicking the corresponding icon on Windows for example)

Within the R session, you must load the RGeostats. If RGeostats has been installed in the R distribution directory, simply type:

```
library(RGeostats)
```

Otherwise type:

```
library(RGeostats, lib.loc="/my_dir")
```

This information can be stored in a specific (hidden) file, called *.First*, which is automatically started each time R is loaded in the working directory.

In order to create it, the best solution is to enter the R session and to define it interactively by typing:

```
fix(.First)
```

The previous command launches a text editor. The name of the text editor can also be parametrized in the *.First* file for future use.

The contents of the *.First* file could be something as :

```
.First
function ()
{
  library(RGeostats, lib.loc="/my_dir")
}
```

7.2 Additional information on RGeostats

When RGeostats is loaded successfully, the user can check the version of the RGeostats package. This information may become useful for further discussion concerning the ability of the package to perform a given task or to describe a mys-functioning:

```
acknowledge("RGeostats")
```

The following message is displayed (which may evolve with time):

Package RGeostats (Version:XX.X.X – Date:mm,dd,yy)

Geoslib Library (Version:XX.X.X – Date:mm,dd,yy)

Authors:

Didier RENARD (didier.renard@mines-paristech.fr)
Nicolas BEZ (nicolas.bez@ird.fr)
Nicolas DESASSIS (nicolas.desassis@mines-paristech.fr)
Helene BEUCHER (helene.beucher@mines-paristech.fr)
Fabien ORS (fabien.ors@mines-paristech.fr)
Xavier FREULON (xavier.freulon@mines-paristech.fr)

Another interesting function (R standard) gives the position where the package has been loaded by typing:

```
search()
```

The following information is obtained in the R session (the contents depends upon the R version, the user's environment and the list of packages already loaded):

```
[1] ".GlobalEnv"           "package:RGeostats" "package:Rcpp"  
[4] "package:stats"        "package:graphics" "package:grDevices"  
[7] "package:utils"        "package:datasets" "package:methods"  
[10] "Autoloads"            "package:base"
```

The order of the loaded packages may vary depending on the user's preferences. It is easy to see that here RGeostats is loaded in position 2.

The user can then type the following command in order to get the list of all the procedures included in RGeostats:

```
ls(pos=2)
```

Another way to learn about each command (say my_command) is to ask for its calling arguments by typing:

```
args(my_command)
```

But obviously the best solution is to get the information on the command by typing:

```
?my_command
```

The information can even be displayed in a more sophisticated manner if the user has launched a HTML browser beforehand by typing the following command at least once in the R session:

```
help.start()
```

Have fun !!!.

8 List of Models

The package RGeostats offers a list of basic structures that can be used in order to construct a Model. Each basic structure is now described with its exact formula.

We recall that:

- the basic structure includes covariances, variograms or generalized covariances
- a covariance is a particular variogram (bounded), a variogram (and therefore a covariance) is a particular generalized covariance
- a basic structure is valid for certain space dimensions
- in all subsequent formulae, the value h defines the modulus of the (isotropic) distance: therefore, this distance is always positive
- some covariance use a practical range which corresponds to the distance beyond which the covariance reaches 95% of the sill value

8.1 List of the Basic structures

8.1.1 Nugget Effect

This is a covariance, defined for any space dimension:

$$C(h) = C_0\delta(h)$$

where:

- C_0 is the sill
- $\delta(h)$ is a function which returns 1 if $h = 0$, and 0 for strictly positive distance

8.1.2 Exponential

This is a covariance, defined for any space dimension:

$$C(h) = C \times \exp\left(-\frac{h}{a/s}\right)$$

where:

- C is the sill
- a is the (practical) range
- $s = 2.995732$

8.1.3 Spherical

This is a covariance, defined up to the third space dimension:

$$C(h) = C \times \left[1 - \frac{3h}{2a} + \frac{1}{2} \left(\frac{h}{a} \right)^3 \right]$$

where:

- C is the sill
- a is the range

8.1.4 Gaussian

This is a covariance, defined for any space dimension:

$$C(h) = C \times \exp \left(- \left(\frac{h}{a/s} \right)^2 \right)$$

where:

- C is the sill
- a is the (practical) range
- $s = 1.730818$

8.1.5 Cubic

This is a covariance, defined up to the third space dimension:

$$C(h) = C \times \left[1 - 7 \left(\frac{h}{a} \right)^2 + \frac{35}{4} \left(\frac{h}{a} \right)^3 - \frac{7}{2} \left(\frac{h}{a} \right)^5 + \frac{3}{4} \left(\frac{h}{a} \right)^7 \right]$$

where:

- C is the sill
- a is the range

8.1.6 Cardinal Sine

This is a covariance, defined for any space dimension:

$$C(h) = C \times \left[\frac{\sin\left(\frac{h}{a/s}\right)}{\frac{h}{a/s}} \right]$$

where:

- C is the sill
- a is the (practical) range
- $s = 20.371$

8.1.7 J-Bessel

This is a covariance defined for any space dimension.

$$C(h) = C \times \left[2^\alpha \Gamma(\alpha + 1) \frac{J_\alpha\left(\frac{h}{a}\right)}{\left(\frac{h}{a}\right)^\alpha} \right]$$

where

- C is the sill
- a is the range
- $\alpha > \frac{d}{2} - 1$ is the shape factor (third parameter)

8.1.8 K-Bessel

This is a covariance defined for any space dimension. It is also known as the Matérn variogram.

$$C(h) = C \times \left[\frac{\left(\frac{h}{a}\right)^\alpha}{2^{\alpha-1} \Gamma(\alpha)} K_{-\alpha}\left(\frac{h}{a}\right) \right]$$

where

- C is the sill
- a is the range
- $\alpha > 0$ is the shape factor (third parameter)

8.1.9 Gamma

This is a covariance, defined for any space dimension:

$$C(h) = C \times \left[\frac{1}{\left(1 + \frac{h}{a}\right)^\alpha} \right]$$

where:

- C is the sill
- a is the range
- α is the (positive) exponent defined as the third parameter

8.1.10 Cauchy

This is a covariance, defined for any space dimension:

$$C(h) = C \times \left[\frac{1}{\left(1 + \left(\frac{h}{a}\right)^2\right)^\alpha} \right]$$

where:

- C is the sill
- a is the range
- α is the (positive) exponent defined as the third parameter

8.1.11 Stable

This is a covariance, defined for any space dimension:

$$C(h) = C \times \exp\left(-\left(\frac{h}{a}\right)^\alpha\right)$$

where:

- C is the sill
- a is the range
- α is the exponent defined as the third parameter which lies within $[0; 2]$

8.1.12 Linear

This is a variogram, defined for any space dimension:

$$\gamma(h) = C \times \frac{h}{a}$$

where:

- C is the multiplicative coefficient (also called *sill* in the interface)
- a is the scale factor (also called *range* in the interface)

8.1.13 Power

This is a variogram, defined for any space dimension:

$$\gamma(h) = C \times \left(\frac{h}{a}\right)^\alpha$$

where:

- C is the multiplicative coefficient (also called *sill* in the interface)
- a is the scale factor (also called *range* in the interface)
- α is the exponent defined as the third argument which must lie within $[0; 2]$

8.1.14 Order-1 GC

This is a generalized covariance defined for an intrinsic random function, defined for any space dimension:

$$\gamma(h) = C \times \left(\frac{h}{a}\right)$$

where:

- C is the multiplicative coefficient (also called *sill* in the interface)
- a is the scale factor (also called *range* in the interface)

8.1.15 Order-3 GC

This is a generalized covariance defined for a first order random function (it needs a first order polynomial drift), defined for any space dimension:

$$K(h) = C \times \left(\frac{h}{a}\right)^3$$

where:

- C is the multiplicative coefficient (also called *sill* in the interface)
- a is the scale factor (also called *range* in the interface)

8.1.16 Spline GC

This is a generalized covariance defined for a first order random function (it needs a first order polynomial drift), defined for any space dimension:

$$K(h) = C \times \left(\frac{h}{a}\right)^2 \ln\left(\frac{h}{a}\right)$$

where:

- C is the multiplicative coefficient (also called *sill* in the interface)
- a is the scale factor (also called *range* in the interface)

8.1.17 Order-5 GC

This is a generalized covariance defined for a second order random function (it needs a second order polynomial drift), defined for any space dimension:

$$K(h) = C \times \left(\frac{h}{a}\right)^5$$

where:

- C is the multiplicative coefficient (also called *sill* in the interface)
- a is the scale factor (also called *range* in the interface)

8.1.18 Cosinus

This is a covariance, defined for a one dimension space:

$$C(h) = C \times \cos\left(\frac{2\pi h}{a}\right)$$

where:

- C is the sill
- a is the period

8.1.19 Triangle

This is a covariance, defined for a one dimension space:

$$C(h) = C \times \left(1 - \frac{h}{a}\right) \times \delta(h < a)$$

where:

- C is the sill
- a is the range
- $\delta(f)$ is a function which returns 1 if f is true and 0 otherwise

8.1.20 Cosexp

This is a covariance, defined for a one dimension space:

$$C(h) = C \times \cos\left(\frac{2\pi h}{\alpha}\right) \times \exp\left(-\frac{h}{a/s}\right)$$

where:

- C is the sill
- a is the (practical) range
- $s = 2.995732$
- α is the period

8.1.21 Exp2dfact

This is a covariance, defined for any space dimension:

$$C(h) = C \times \exp\left(-\frac{h_{2D}}{a_{2D}/s}\right) \times \prod \exp\left(\frac{h_i}{a_i/s}\right)$$

where:

- C is the sill
- a is the (practical) range
- $s = 2.995732$
- h_{2D} refers to the distance in the 2-D plane
- h_i refers to the distance in any subsequent space dimension

8.1.22 Expfact

This is a covariance, defined for any space dimension:

$$C(h) = C \times \prod \exp\left(-\frac{h_i}{a_i/s}\right)$$

where:

- C is the sill
- a is the (practical) range
- $s = 2.995732$
- h_i refers to the distance in any space dimension

8.1.23 Reg1D

This is a covariance, defined for 1 dimension only:

$$\begin{aligned} C(h) &= C \times \left[1 - 3\frac{h}{a} \left(1 - \frac{h}{2a} \left(1 + \frac{h}{6a}\right)\right)\right] & h < 0.5 \\ C(h) &= C \times \left[-2 + 3\frac{h}{a} \left(1 - \frac{h}{2a} \left(1 - \frac{h}{6a}\right)\right)\right] & 0.5 < h < 1 \\ C(h) &= 0 & h > 1 \end{aligned}$$

where:

- C is the sill
- a is the range
- h refers to the distance in 1 dimension

8.1.24 Pentamodel

This is a covariance, which corresponds to the spherical model calculated in R^7 , after fourth order "montée" (upscaling)

$$C(h) = C \times \begin{cases} 1 - \frac{22}{3} \left(\frac{h}{a}\right)^2 + 33 \left(\frac{h}{a}\right)^4 - \frac{77}{2} \left(\frac{h}{a}\right)^5 + \frac{33}{2} \left(\frac{h}{a}\right)^7 - \frac{11}{2} \left(\frac{h}{a}\right)^9 + \frac{5}{6} \left(\frac{h}{a}\right)^{11} & h < 1 \\ C(h) = 0 & 1 < h \end{cases}$$

where:

- C is the sill
- a is the range
- h refers to the isotropic distance

8.1.25 Spline-2

This is a generalized covariance defined for a first order random function (it needs a second order polynomial drift), defined for any space dimension:

$$K(h) = C \times \left(1 - 2 \left(\frac{h}{a}\right)^2 + \left(\frac{h}{a}\right)^3 \right) \quad h < a$$

where:

- C is the multiplicative coefficient (also called *sill* in the interface)
- a is the scale factor (also called *range* in the interface)

8.1.26 Storkey

This is a covariance:

$$C(h) = \frac{C}{3} \begin{cases} 2 \times \left(1 - \frac{h}{a}\right) \times \left(1 + \frac{\cos\left(\frac{2\pi h}{a}\right)}{2}\right) + \frac{3}{2\pi} \sin\left(\frac{2\pi h}{a}\right) & h \leq 1 \\ C(h) = 0 & h > 1 \end{cases}$$

where:

- C is the sill
- a is the range
- h refers to the isotropic distance

8.1.27 Wendland1

This is a covariance:

$$C(h) = C \begin{cases} \left[1 - \left(\frac{h}{a}\right)^2 \times \left(10 - \frac{h}{a} \times \left(20 - \frac{h}{a} \times \left(15 - \frac{4h}{a}\right)\right)\right) \right] & h \leq 1 \\ C(h) = 0 & h > 1 \end{cases}$$

where:

- C is the sill
- a is the range
- h refers to the isotropic distance

8.1.28 Wendland2

This is a covariance:

$$C(h) = C \begin{cases} \left[1 - \left(\frac{h}{a}\right)^2 \times \left(\frac{28}{3} - \left(\frac{h}{a}\right)^2 \times \left(70 - \frac{h}{a} \times \left(\frac{448}{3} - \frac{h}{a} \times \left(140 - \frac{h}{a} \times \left(64 - \frac{35h}{3a}\right)\right)\right)\right)\right) \right] & h \leq 1 \\ C(h) = 0 & h > 1 \end{cases}$$

where:

- C is the sill
- a is the range
- h refers to the isotropic distance

8.2 Anisotropy

This paragraph describes the way RGeostats handles the anisotropy.

We first recall that a model is a combinaison of several basic structures. Each basic structure corresponds to one of the structures listed and illustrated in the previous paragraph. Each basic structure (at least the ones which require a range definition) can be anisotropic.

Traditionally, two sorts of anisotropies can be distinguished:

- the geometric anisotropy: the ranges are different in different directions, so that a simple stretch of the space (in the relevant direction) would bring the structure back to isotropy.

- the zonal anisotropy: in one given direction of the space, the variability (corresponding to the sill) is smaller than in any other direction of the space.

First, let us note that both anisotropies require the definition of the rotation system in which the anisotropy will be expressed: we will then speak of the main anisotropy (orthogonal) directions.

8.2.1 Geometric anisotropy

When the main anisotropy directions have been defined, the geometric anisotropy consists in defining the ranges along these different directions.

For the covariance evaluation, the generic formula (defined in the previous paragraph) is used. For example, the isotropic spherical covariance is defined in the 2-D space as:

$$C(h) = C \times \left[1 - \frac{3h}{2a} + \frac{1}{2} \left(\frac{h}{a} \right)^3 \right]$$

where:

- C is the sill
- a is the (isotropic) range

We now consider the anisotropic spherical covariance considering that the main anisotropy directions coincide with the main axes of the system (otherwise, we must simply perform the rotation beforehand). The ranges are defined in the two main anisotropy directions (therefore along X and Y), denoted as a_x and a_y .

The value $\frac{h}{a}$ of the general formula must be replaced by the weighted distance:

$$\sqrt{\left(\frac{h_x}{a_x} \right)^2 + \left(\frac{h_y}{a_y} \right)^2}$$

8.3 Zonal anisotropy

When the main anisotropy directions have been defined, we must define the direction in which the variable shows no variability: say the Y direction for example. Then it can almost be considered as a geometric anisotropy where the range along the Y direction is set to a large arbitrary distance.

However, it is not trivial to enter such as a large distance. For that reason, in the interactive way to input a model (function *model.input*), it suffices to set conventionally the anisotropy range, in the Y direction, to NA. The range printed as N/A in the anisotropy direction confirms that this direction corresponds to the zonal anisotropy main direction.

When reading model from an ASCII file using the *model.read* procedure, the range in the Y direction should be set to NA again. But the NA string cannot be read: instead, it must be replaced by the conventional string “-999.0”.