# 5. Geostatistics

**JEAN-MICHEL FLOCH**
*INSEE*



**Abstract**

Geostatistics is a very important branch of spatial statistics. It jhas been developed on the basis of very practical concerns (mining research), and has undergone very significant methodological developments driven by Georges Matheron and his colleagues at the Fontainebleau Ecole des Mines. The simplest illustrations relate to problems such as the interpolation of temperatures or precipitation. But the most important work concerns geological and mining applications (*e.g.* Chiles et al. 2009). Applying geostatistics to demographic or social examples is more difficult, but it seems important to present the broad outline of the method — how to treat stationarity and introduce intrinsic stationarity; introducing the semivariogram to study spatial relations; data interpolation using the kriging method. Beyond mining applications, variogram analysis can be used in mixed models to analyse residuals.

R Prior reading of Chapter 1: "Descriptive spatial analysis", and Chapter 4: "Spatial distributions of points" is recommended.

Modelling spatial data is made difficult by the fact that there is only one realisation of the phenomenon. As in spatial point patterns, only one realisation is also observed, for which all the data are available. For continuous data (potentially observable at any point in the space), only partial data are available, from which values at unobserved points may be predicted. It is this lack of information that will lead to the use of probabilistic models.

Randomness is not a property of the phenomenon, but a characteristic of the model used to describe it. Geostatistics, which studies continuous phenomena, has enabled the development of specific methods to study spatial relationships between observations and to construct predictive tools.

Geostatistics owes its name to the discipline's origins in mining (Krige, Matheron). Many of the basic concepts of the discipline derive from the work of Georges Matheron (*regionalized variable, random function, intrinsic assumption, nugget effect* [1], Matheron et al. 1965). Figure 5.1 presents a summary illustrating the pathway from reality towards a more abstract model, a model that will itself enable us to act in expected best way (Chauvet 2008).

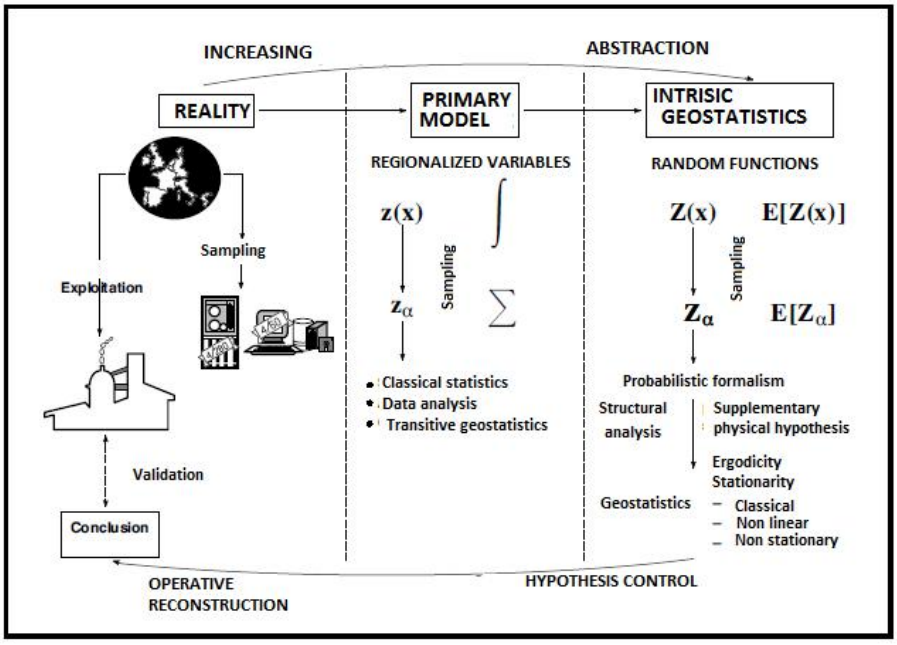


Figure 5.1 – Diagram of a geostatistical analysis
**Source:** *Chauvet 2008*

## 5.1 Random functions

### 5.1.1 Definitions

As in Figure 5.1, we designate $z(s)$ the *regionalized variable*, and $Z(s)$ the *random function*, the letter $s$ designating the position in space. We will retain this wording here as specific to geostatistics. A phenomenon occurring in space is qualified as regionalized. A regionalized variable is a function that describes this phenomenon satisfactorily. This is a first level of abstraction, where we remain

1. These terms are defined later in this chapter.

in the description, without resorting to a probabilistic model. If we make no additional assumption, we remain within the framework of *transitive geostatistics*.

The next step, qualified as *intrinsic geostatistics*, introduces the concept of random function. It results from a choice — considering the regionalized variable as the realisation of a random function. This choice makes it possible to use powerful probabilistic tools, the counterpart being moving further from reality. The probabilistic model is a calculation intermediary that is expected to be used in understanding the regionalized phenomenon.

The random function is fully characterised by the knowledge of its distribution function.

$$F(s_1, s_2, ...s_n; z_1, z_2, ...z_n) = P\{Z(s_1) \leq z_1, Z(s_2) \leq z_2, ..., Z(s_n) \leq z_n\}. \tag{5.1}$$

Since we have only one realisation of our regionalized phenomenon, we need to find another way to make the inference. Citing Matheron et al. 1965: "For inference to be possible, we need to introduce additional assumptions about random function $Z(s)$ so as to reduce the number of parameters on which its law depends. This is the purpose of the stationary assumption we are going to define: a stationary function is repeated itself in some form in space, and this repetition once again makes statistical inference possible from a single realisation." Each observation will therefore be treated as the realisation of a random variable.

### 5.1.2 Stationarity

Three meanings of stationarity are used in geostatistics:
— strict stationarity;
— second order stationarity;
— intrinsic stationarity.

**Definition 5.1.1 — Strict stationarity.** Strict stationarity directly refers to the probability law of the process. There is strict stationarity if by moving using translation, all the characteristics of the random function remain the same.
Formally, the joint distribution of $Z(s_i)$ is the same as that of $Z(s_i + h)$, $h$ indicating translation relative to the initial position. This form of stationarity is not operational and very restrictive.

**Definition 5.1.2 — Second order stationarity.** Second order stationarity or weak stationarity no longer imposes conditions on the probability law, but only on the mean and covariance. These indicators must be invariable by translation.
Given that $Z(s)$ breaks down into a deterministic component and a random component
$Z(s) = m(s) + R(s)$

second order stationarity requires the following conditions:
— $E[Z(s)] = m(s) \forall s$.
The invariance of the expected value by translation leads to constancy of the deterministic component.
$m(s+h) = m(s) = m \forall s$;
— The variance is constant: $E\left[(Z(s) - m)^2\right] = \sigma^2$;
— Covariance depends only on the spatial shift:
$Cov[Z(s+h), Z(s)] = E[(Z(s+h) - m)(Z(s) - m)] = C(h)$.

In practice, this stationarity assumption is often too strong. The most important limit is that the mean can change over the area of interest, and that the variance may not be bounded when this area of interest grows. It was George Matheron who drew the consequences of the limits of weak stationarity by suggesting the even weaker concept of intrinsic stationarity (Matheron et al. 1965).

**Definition 5.1.3 — Intrinsic stationarity.** The intrinsic stationarity assumption is as follows: $E\left[(Z(s+h)-Z(s))^2\right]=0$.
Increments can be stationary without the process itself being stationary.

A new function can then be defined, called *variogram*, based on differences between values and shifted values, and which depends only on the offset:

$$\gamma(h)=\frac{1}{2}E\left[Z(s+h)-Z(s)\right]^2 \tag{5.2}$$

Second order stationarity leads to intrinsic stationarity, but the reverse is not true. A random function can enable a variogram to be calculated, but this is not the case for covariance and the autocorrelation function.

## 5.2 Spatial variability

### 5.2.1 Covariance and correlogram

**Definition 5.2.1 — Covariance.** The covariance function will allow the relationships between all point pairs to be considered. If we consider two points $s_i$ and $s_j$, covariance can be defined by Equation 5.3.

$$Cov\left[Z(s_i),Z(s_j)\right]=E\left[\left(Z(s_i)-m\right)\left(Z(s_j)-m\right)\right] \tag{5.3}$$

When the process is second-order stationary, covariance will no longer depend only on the distance between the points $|s_i-s_j|$. If we designate $h$ this distance, we will define $C(h)$ calculated for all values of $h$ taking into account all point pairs located at a distance $h$ from each other. Covariance function $C(h)$ is defined by Equation 5.4.

$$C(h)=Cov\left[Z(s+h),Z(s)\right]=E\left[\left(Z(s+h)-m\right)\left(Z(s)-m\right)\right] \tag{5.4}$$

It reflects how the covariance of observations changes when their distance increases. When $h$ is equal to 0, covariance is equal to variance.

$$C(0)=E\left[(Z(s)-m)^2\right]=\sigma^2 \tag{5.5}$$

The covariance function has the following properties:

$$C(-h)=C(h) \tag{5.6}$$
$$|C(h)|\leq C(0). \tag{5.7}$$

For the covariance function to be called *allowable*, the variance of a linear combination of variables must be positive: $Var\left[\sum_{i=1}^{n}\lambda_i z(s_i)\right]=\sum_{i=1}^{n}\sum_{j=1}^{n}C(s_i-sj)$.

This results in $C$ being *positive semi-definite*.

**Definition 5.2.2 — Autocorrelation function.** The autocorrelation function $\rho(h)$ is defined as a function of $h$ by ratio $\frac{C(h)}{C(0)}$. Its value is between -1 and +1. The following relationships can be shown when second order stationarity is verified:

$$\begin{aligned}\gamma(h)&=C(0)-C(h)\\ \gamma(h)&=\sigma^2\left(1-\rho(h)\right).\end{aligned} \tag{5.8}$$

**Box 5.2.1 — Estimate of the covariance function.** The covariance function is estimated from $n(h)$ point pairs, as defined below, for $i$ variant from 1 to $n(h)$.

$$\widehat{C}(h) = \frac{1}{n(h)} \sum_{i=1}^{n(h)} (z(s_i) - m)(z(s_i + h) - m) \tag{5.9}$$

with $n(h) = Card\left\{(s_i, s_j) / \left|s_i - s_j\right| \approx h\right\}$

### 5.2.2 Variogram

The literature contains *variogram* or *semivariogram* expressions. Some authors (Matheron et al. 1965) believe that the term semivariogram should be used for $\gamma(h)$ as defined in Equation 5.10, the variogram corresponding to $2\gamma(h)$. This is the choice we make in this article.

Out of the following three indicators – covariance function, autocorrelation function and variogram – the latter is most used to the extent that it refers to the weakest form of stationarity and therefore to the least restrictive conditions on the local behaviour of the mean.

$$\gamma(h) = \sigma^2 (1 - \rho(h)) \tag{5.10}$$

The variogram has the following properties:

$$\begin{aligned} \gamma(h) &= \gamma(-h) \\ \gamma(0) &= 0 \\ \frac{\gamma(h)}{\|h\|^2} &\to 0 \quad quand \quad \|h\| \to \infty \end{aligned} \tag{5.11}$$

For any set of real $\{a_1, a_2, ...a_m\}$ verifying $\sum_{i=1}^{m} a_i = 0$, we have the following property:

$$\sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \gamma(s_i - s_j) \leq 0 \tag{5.12}$$

When the process is isotropic:

$$\gamma(h) = \gamma(\|h\|) \tag{5.13}$$

**Box 5.2.2 — Estimate of the experimental variogram.** An experimental variogram can be estimated from point pairs defined as previously.

$$\widehat{\gamma}(h) = \frac{1}{n(h)} \sum_{i=1}^{n(h)} (z(s_i + h) - z(s_i))^2 \tag{5.14}$$

where $n(h) = Card\left\{(s_i, s_j) / \left|s_i - s_j\right| \approx h\right\}$
The variogram can be estimated in different directions to highlight any anisotropy of the phenomenon studied.

### 5.2.3 Empirical application

#### Geostatistics with R

The packages dedicated to geostatistics in this chapter are **gstat** and **geoR**, which are most commonly used. Many other packages are available on the CRAN website. Below is a list with comments produced by Roger Bivand.

Package **gstat** provides a wide range of functions for univariate and multivariate geostatistics, including larger data sets, while **geoR** and **geoRglm** contain functions for model-based geostatistics. Variogram diagnostics can be performed with **vardiag**. Automated interpolation using gstat is available in **automap**. This family of packages is supplemented by **intamap** with automatic interpolation procedures and **psgp**, which implements Gaussian process regression kriging. A broad range of similar functions is found in the **fields** package. The spatial package is delivered with the R database and contains several main functions. The **spBayes** package is compatible with Gaussian univariate and multivariate models with MCMC. The **rampes** package is another Bayesian geostatistical modelling set. The **geospt** package contains basic geostatistical and radial functions, including prediction and cross-validation. In addition, it includes functions to design optimal spatial sampling networks based on geostatistical modelling. **spsann** is another package that offers functions to optimise sample configurations, using spatial simulated annealing. The **geostatsp** package offers geostatistical modelling functions using Raster and SpatialPoints objects. Non-Gaussian models are adapted using INLA, and Gaussian geostatistical models use the estimate of maximum likelihood. The **FRK** package is a spatial/spatial-temporal modelling and prediction tool with large data sets. The approach, discussed in Cressie and Johannesson (2008), breaks down the field, and therefore the covariance function, using a fixed set of n basic functions, where n is generally much smaller than the number of data points (or polygons).

#### RGeostats package

RGeostats is an R-language package. It was developed by the geostatistics team at the Geosciences Centre of Mines ParisTech. It implements all geostatistical functions available from the (commercial) Geoslib library (written in C/C++). It therefore benefits from the experience accumulated in the mining field, and is especially dedicated to such applications. It makes it possible to carry out all the implementations described in this chapter, and to deal with the support effects.

RGeostats allows R users, by loading and installing it, to access the normal Geostatistics functionalities. It is also a platform that enables the Geostatistics team at the Geosciences Centre to develop prototypes for the application of new models (*e.g.* Boolean simulations, bi-plurigaussian simulations) or new techniques (flow of fluids, simulations of the first arrival time in geophysics). The functions are described, but the code is not accessible, and there are few examples.

RGeostats can be downloaded from the following address: `cg.ensmp.fr/rgeostats`.

#### Exploratory analyses

The proposed application uses *Swiss rainfall* data. This database is widely used in spatial studies, particularly in Diggle et al. 2003. It is supplied in the R *geoR* package. The observations are rainfall readings from 467 weather stations in Switzerland, and were made on 8 May, 1986. This is indeed a continuous data point, as rainfall can potentially be recorded at any point in the country. It can therefore fall within geostatistical modelling, but it is easier to understand than mining data. In addition to rainfall, measured in millimetres, data about the altitude of weather stations is provided.

In the *geoR* package, there are three Spatial Interpolation Comparison (SIC) databases:

— *Sic.100*: sample of 100 observations that may be used to make interpolations;
— *Sic.367*: observations not included in the sample that will enable estimates and observations to be compared;
— *Sic.all*: set.

The features of *geoR* will be used here, but *gstat* and *RGeostats* provide analysis tools.

Figure 5.2 shows the data sampled (in green) and the control data, the circles being proportional to the rainfall recorded.

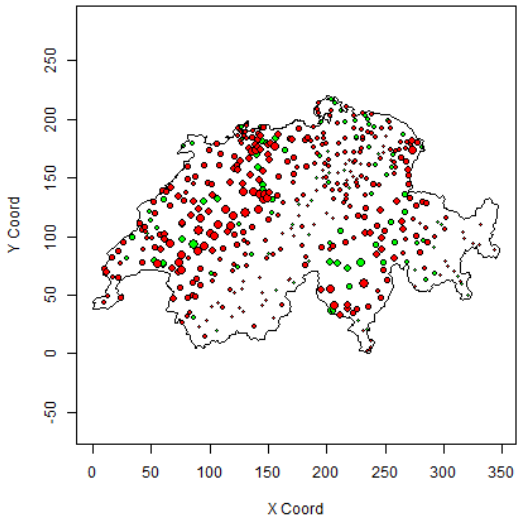


Figure 5.2 – Rainfall in Switzerland
**Source:** *Swiss rainfall from the geoR package*

```
library(geoR)
points(sic.100, borders=sic.borders,col="green")
points(sic.367, borders=sic.borders,col="red",add=TRUE)
```

The *geoR* package provides some descriptive representations using the `plotgeodata` function. Figure 5.3 shows, from left to right and top to bottom:

— representation of the level of rainfall, based on the quantiles of the variable;
— rainfall based on latitude;
— rainfall based on longitude;
— the histogram of rainfall data.

```
library(geoR)
plot.geodata(sic.100,bor=sic.borders)
```

The histogram in Figure 5.3 suggests that the variable's distribution is not Gaussian, and that a data transformation could be considered since the most common methods only have interesting properties in the Gaussian context.

### Variogram cloud and experimental variogram

In intrinsic geostatistics, the **variogram cloud** is a cloud of data points expressing their variability based on their interspacing. The variogram cloud provides the graphical representation of the values used to calculate the variogram. For a dataset of variable $Z$ at points $(s_1,.,s_i,.,s_n)$, it represents the abscissa points $\left\|s_i - s_j\right\|$ and ordinate points $\frac{1}{2}\left[z(s_i) - z(s_j)\right]^2$. It can be represented as a point cloud and as a boxplot (see Figure 5.4).

```
library(geoR)
library(fields)
vario.b<-  variog(sic.100,option =c ("bin", "cloud", "smooth"),
bin.cloud=TRUE)
```

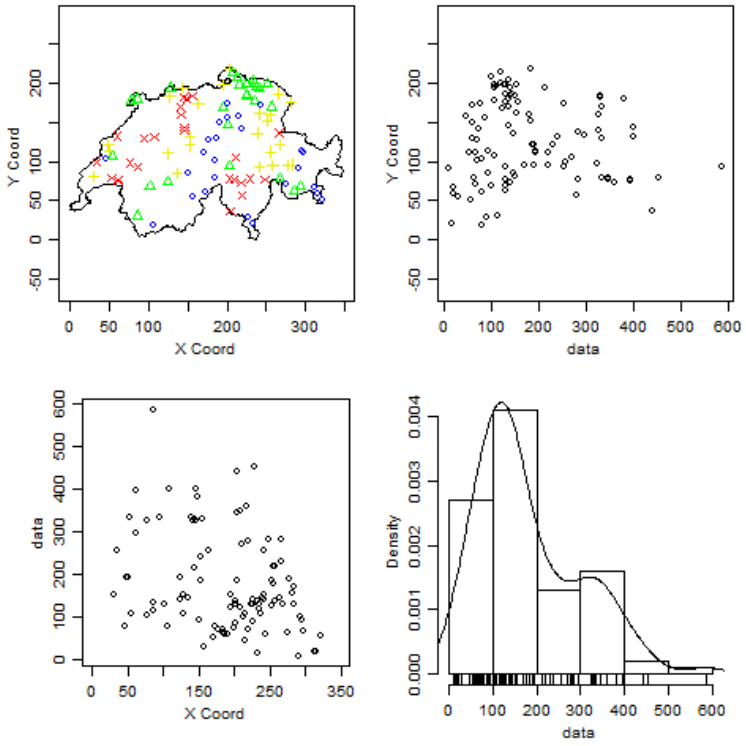


Figure 5.3 – Some descriptive statistics
**Source:** *Swiss rainfall from the geoR package*

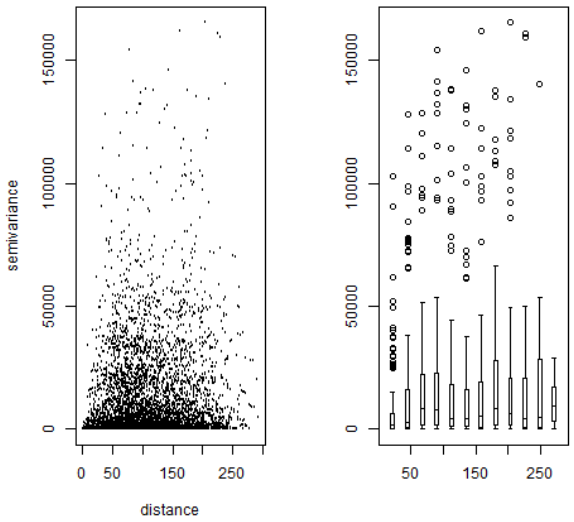


Figure 5.4 – Variogram cloud
**Source:** *Swiss rainfall from the geoR package*

```
vario.c <- variog(sic.100, op="cloud")
bplot.xy(vario.c$u,vario.c$v, breaks=vario.b$u,col="grey80",
lwd=2,cex=0.1,outline=FALSE)
```

Since these representations are hard to read, the most useful representation is the experimental variogram (defined in box 5.2.2), shown in Figure 5.6 based on a construction diagram shown in Figure 5.5.

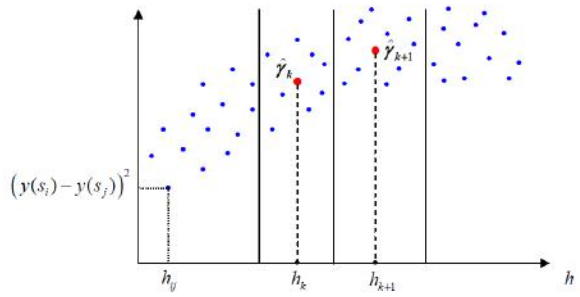


Figure 5.5 – Experimental variogram: construction diagram
**Source:** *Swiss rainfall from the geoR package*

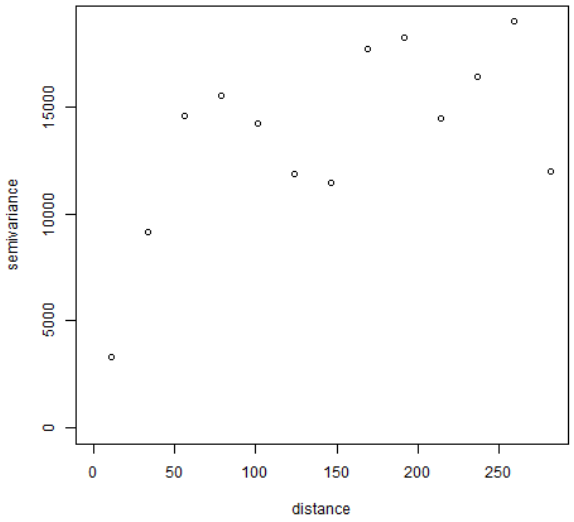


Figure 5.6 – Experimental variogram
**Source:** *Swiss rainfall from the geoR package*

```
library(geoR)
vario.ex<-  variog(sic.100, bin.cloud=TRUE)
plot(vario.ex)
```

In Figure 5.6, all observed points are used to calculate the variogram. But the phenomena studied are not necessarily isotropic, and it may be useful to calculate variograms based on several directions of space (see Figure 5.7).

```
library(geoR)
```

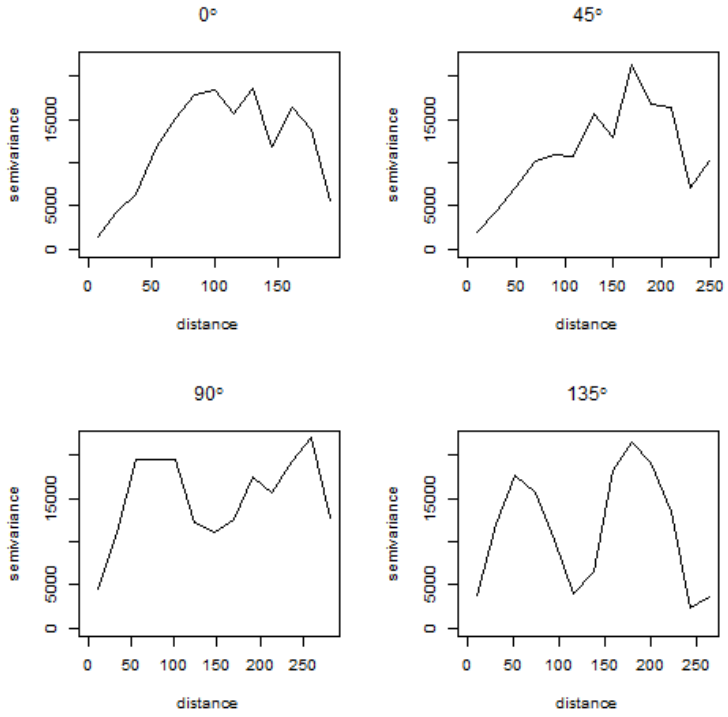


Figure 5.7 – Directional variograms
**Source:** *Swiss rainfall from the geoR package*

```
vario4<-variog4(sic.100)
plot(vario4,same=FALSE)
```

## 5.3 Fitting variogram

In section 5.4, which is dedicated to kriging, we see that the value of the estimators depends on observations and the spatial autocorrelation structure, understood by the variogram. Variogram analysis is therefore not just a passage point. It forms the central point of the geostatistical approach. The empirical variograms shown in section 5.2 cannot be used directly because they don't meet the properties listed in section 5.2.2. To be used in geostatistical models, they must first be adjusted to theoretical models with well-defined analytical forms, which implies a vision of what a semi-variogram should be.

### 5.3.1 General shape of the variogram

We begin by presenting the most classical model, from which theoretical variograms will be constructed, making it possible to define, among other things, the kriging equations (see section 5.4).

This variogram has a form that is initially increasing, up to a certain level. The value of $h$ corresponding to this plateau is called the *range*. We understand it by referring to the relationship between covariance, when it is defined, and the semi-variogram, *i.e.* $\gamma(h) = C(0) - C(h)$. Covariance is very often a decreasing function of distance, which implies the increase of the semi-variogram, but this is not always the case (*e.g.* cardinal sine model).

At a certain distance, which is called the range, covariance will be cancelled out. This range is the range of spatial dependence. There is no longer any relationship between the values observed at

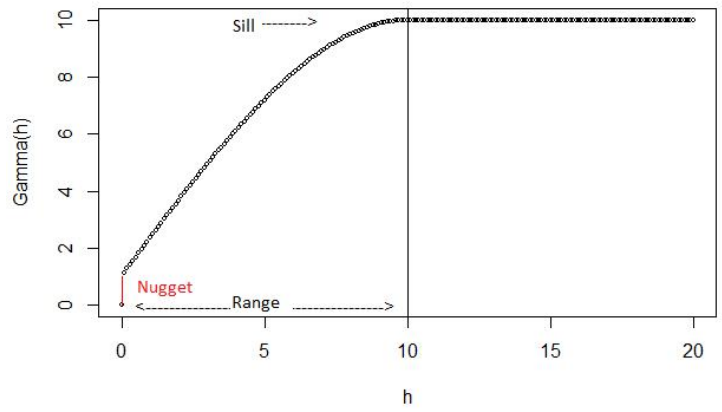


Figure 5.8 – Theoretical variogram

a distance beyond this range. For the semi-variogram this means that, beyond this range, its value is constant and we have: $C = C(0) = \sigma^2$.

For $h = 0$, the value of the variogram is zero, by definition. But in practice we see that for values very close to 0, the variogram takes values greater than 0 and there is therefore a discontinuity at the origin. We call the limit of the variogram at zero a *nugget*. As Matheron explains (Matheron et al. 1965): "The concept of scale plays a key role here. At a scale of ten metres, a transition phenomenon where the range is measured in centimetres is only seen on $\gamma(h)$ as a discontinuity at the origin, *i.e.* a nugget effect." It represents the variation between two measurements made at infinitely close locations, and from two effects there may arise:

— variability of the measuring instrument: the nugget therefore partly measures the statistical error of the measuring instrument.

— a real nugget effect: a sudden change in the measured parameter; the historical case is the passage without transition from a gold nugget to soil containing virtually no gold.

Other forms of variograms may be encountered. Figure 5.9 shows two classical cases. The first is the **linear variogram**, the second the **pure nugget effect**. When the variogram is unbounded, the mean and variance are not defined. This most frequently indicates a large-scale trend, which must be modelled. The pure nugget effect reflects the lack of spatial dependence.

### 5.3.2 Usual variograms

Geostatistical literature offers many functions that satisfy the properties of the semi-variogram as shown in Figure 5.8. These configured functions must be used to describe the different components (range, plateau, nugget). They must also handle the behaviour of the function at the origin (linear trend, horizontal or vertical tangent).

We will only discuss four examples of variogram models here (Figure 5.10), the others being described in the reference works (Armstrong 1998, Chiles et al. 2009, Waller et al. 2004).

**Definition 5.3.1 — Spherical model.**

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ c_0 + c_s \left[ \frac{3}{2}\left(\frac{h}{a}\right) - \frac{1}{2}\left(\frac{h}{a}\right)^3 \right] & 0 < h \leq a \\ c_0 + c_s & h > 0 \end{cases} \tag{5.15}$$

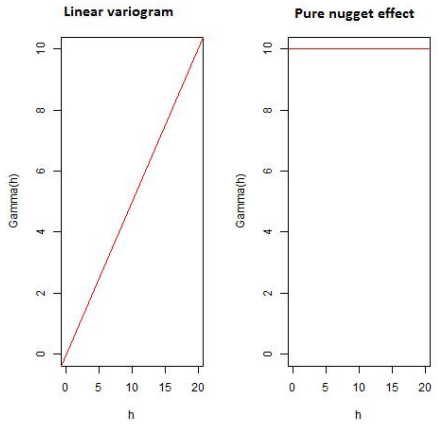


Figure 5.9 – Two atypical semi-variograms

**Definition 5.3.2 — Exponential model.**

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ c_0 + c_s \left[1 - exp\left(-\frac{h}{a}\right)\right] & h > 0 \end{cases} \tag{5.16}$$

**Definition 5.3.3 — Gaussian model.**

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ c_0 + c_s \left[1 - exp\left(-(\frac{h}{a})^2\right)\right] & h > 0 \end{cases} \tag{5.17}$$

**Definition 5.3.4 — Power model.**

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ c_0 + bh^p & h > 0 \end{cases} \tag{5.18}$$

**Definition 5.3.5 — Matern model.**

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ c_s \left[1 - \frac{\frac{h}{a}}{2^{\alpha-1}\Gamma(\alpha)} K_\alpha\left(\frac{h}{a}\right)\right] & h > 0 \end{cases} \tag{5.19}$$

where $\Gamma$ refers to the gamma function and $K_\alpha$, the modified Bessel of the second kind of parameter $\alpha$.

**Definition 5.3.6 — Cardinal sine model.**

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ c_s \left[1 - \frac{a}{h} sin\left(\frac{h}{a}\right)\right] & h > 0 \end{cases} \tag{5.20}$$

As explained by Matheron and the geostatisticians of the Ecole des Mines in Fontainebleau, modelling is a matter of choice. Choosing a theoretical model is a decisive moment in the geostatistician's approach, but we cannot associate *a priori* a theoretical variogram to any given type of process. We have to take account both empirical knowledge of the phenomenon and the shape of the experimental variogram obtained. As this is the nugget effect, one might think that it is not appropriate for data such as rainfall.

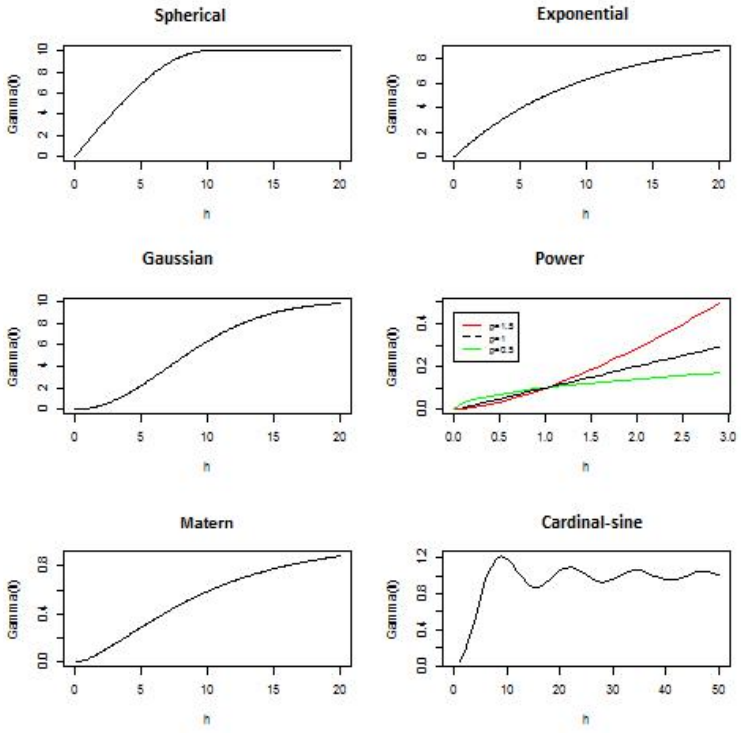


Figure 5.10 – Four examples of theoretical variograms

### 5.3.3 Variogram fitting

A functional form appropriate to the experimental variogram must then be found. An important first step is to obtain a smoothed representation of the variogram, which can be obtained using the `variog` function of *geoR*. As always, this smoothing depends on the choice of window. The smoothed representation is sometimes considered adequate to visually estimate the variogram. Many geostatisticians criticise this approach as too empirical, but this step can give some indications, particularly on behaviour at the origin.

Figure 5.11 shows three examples of fitting the experimental variogram to rainfall data in Switzerland, using a spherical variogram, an exponential variogram without nugget and an exponential variogram with nugget. These adjustments are made using the `lines.variomodel` function of *geoR* (Ribeiro Jr et al. 2006). This is the first visual approach. We can see that while the exponential variogram with nugget apparently fits better to the data than the variogram without nugget, the introduction of this very short distance effect has no physical justification for rainfall.

```
library(geoR)
vario.ex<-  variog(sic.100,option="bin")
vario.sphe<-(variofit(vario.ex,cov.model=  "spher",
ini.cov.pars=c(15000,200)))
par(mfrow=c(2,2), mar=c(3,3,1,1), mgp =c (2,1,0))
plot(vario.ex,main="Spherical")
lines.variomodel(cov.model="sphe",cov.pars=c(15000,100),
nug=0,max.dist=350)
plot(vario.ex,main="Exponential")
lines.variomodel(cov.model="exp",cov.pars=c(15000,100),
nug=0,max.dist=350)
```

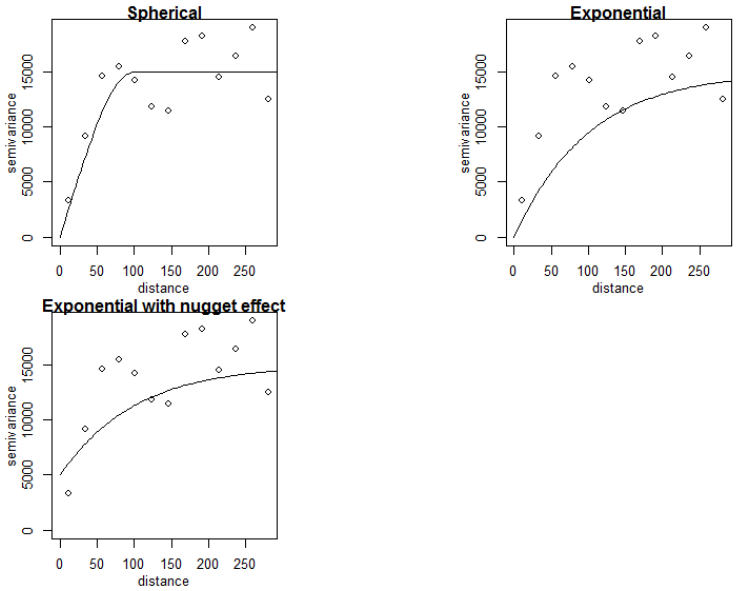


Figure 5.11 – Three examples of experimental variogram adjustment
**Source:** *Swiss rainfall from the geoR package*

```
plot(vario.ex,main="Exponential with nugget effect")
lines.variomodel(cov.model="exp",cov.pars=c(10000,100),
nug=5000,max.dist=350)
plot(vario.ex,main="Exponential with nugget effect")
lines.variomodel(cov.model="matern",cov.pars=c(10000,100),
nug=0,max.dist=350,kappa=0.5)
```

The choice is therefore a compromise as noted by Waller et al. 2004: "Even if a particular model is deemed better for a particular dataset using a statistical adjustment method, it may not necessarily be the best choice. For example, the Gaussian model is often selected using an automatic adjustment criterion, but this provides smoothing that often appears unrealistic. Ultimately, the final choice of model should reflect both the result of the procedure for adjusting the statistical model and a consistent interpretation with scientific understanding of the process being studied."

Many methods are suggested to fit the variogram — methods based on ordinary or weighted least squares, methods based on likelihood, as well as Bayesian methods. In *geoR*, the functions used are `variofit` (least squares) and `likfit` (maximum likelihood). The methods are quite technical and would require significant developments. Refer to Ribeiro Jr et al. 2006 for an illustration of these methods using simulated data. The R code is supplied in the article.

### Fitting by Ordinary Least Squares (OLS)

We look for the vector of the function's parameters that minimises a simple objective function, the sum of squares of the distances between the value of the experimental semi-variogram and the value of the theoretical variogram.

$$\hat{\theta}_{MCO} = \underset{\theta \in \Theta}{argmin} \sum_{i=1}^{k} (\hat{\gamma}(h_i) - \gamma(h_i;\theta))^2 \tag{5.21}$$

### Fitting by Weighted Least Squares (WLS)

Ordinary least squares do not take into account the number of point pairs involved in the calculation of each point of the experimental variogram, unlike the weighted least squares.

$$\hat{\theta}_{MCP} = \underset{\theta \in \Theta}{argmin} \sum_{i=1}^{k} \frac{\#N(h_i)}{\gamma(h_i;\theta)^2} \left(\hat{\gamma}(h_i) - \gamma(h_i;\theta)\right)^2 \tag{5.22}$$

#### Fitting by Generalised Least Squares (GLS)

Other authors have proposed generalised least squares to take heteroscedasticity into account.

$$\hat{\theta}_{MCG} = \underset{\theta \in \Theta}{argmin} \left(\hat{\gamma}_n - \gamma(\theta)\right)^T Cov(\gamma_n)^{-1} \left(\hat{\gamma}_n - \gamma(\theta)\right) \tag{5.23}$$

where $\gamma$ is the vector $(\gamma_1, \gamma_2, ..., \gamma_K)$.

#### Fitting by maximum likelihood (ML)

The parameters of the model are estimated by calculating the likelihood. In the non-Gaussian case, the estimates are not robust. The calculations are tedious, and this method must only be used for small samples. In addition, this method requires second order stationarity and cannot be applied to unbounded variograms. In the latter case, the weighted least squares must be used.

## 5.4 Ordinary kriging

### 5.4.1 Principle

The term kriging was coined by to Georges Matheron, and refers to the pioneering work by South African engineer Danie Krige. Kriging is a very powerful interpolation method. The examples provided here are very basic. Applications to mining or geological research provide many examples where we want to estimate volumes and not just simple interpolations. We will not discuss *simple* kriging here, which assumes the mean value is known, but *ordinary* kriging, which forms the highlight of geostatistics. In ordinary kriging, the mean value is unknown. Simple uses of it can be found in interpolating temperatures (Joly et al. 2009) or in air quality studies (Lloyd et al. 2004). Assume that $Z(.)$ is intrinsically stationary, that its variogram $\gamma(h)$ is known but its mean $m$ is unknown. We have a data set $Z = [Z(s_1), .., Z(s_i), .., Z(s_N)]^t$. We want to predict the value of $Z(.)$ as an unobserved point and calculate $Z(s_0)$. The ordinary kriging estimator will be defined as a linear combination of observations.

$$Z_{OK}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i) \tag{5.24}$$

The weighting values $\lambda_i$ are not calculated using a distance function but by using the semi-variogram and two statistical criteria. This way, there is no bias and minimisation of the mean squared prediction error. The absence of bias brings us to Equation 5.25:

$$E\left[\hat{Z}_{OK}(s_0)\right] = E\left[Z(s_0)\right] = m \tag{5.25}$$

$$E\left[\hat{Z}_{OK}(s_0)\right] = \sum_{i=1}^{N} \lambda_i E[Z(s_i)] = \sum_{i=1}^{N} \lambda_i m \quad \rightarrow \quad \sum_{i=1}^{N} \lambda_i = 1$$

Therefore, using the Lagrange multipliers method, we will minimise $E\left[\hat{Z}_{OK}(s_0) - Z(s_0)\right]^2$ under the constraint $\sum_{i=1}^{N} \lambda_i = 1$.

Box 5.4.1 shows how to introduce the variogram and end up with the kriging equations.

$$\begin{aligned}\sum_{j=1}^{N}\lambda_j\gamma(s_i-s_j)+m&=\gamma(s_0-s_i)\\ \sum_{i=1}^{N}\lambda_i&=1\end{aligned}\tag{5.26}$$

The value of $\hat{Z}_{OK}(s_0)$ is determined by points that depend on the correlation between the estimation point and the observation points, but also correlations between the observation points. These kriging equations are generally best written in matrix form:

$$\begin{bmatrix}\lambda_1\\ \lambda_2\\ .\\ .\\ \lambda_N\\ m\end{bmatrix}=\begin{bmatrix}\gamma(s_1-s_1) & .. & \gamma(s_1-s_N) & 1\\ \gamma(s_2-s_1) & .. & \gamma(s_2-s_N) & 1\\ . & .. & . & .\\ . & .. & . & .\\ \gamma(s_N-s_1) & .. & \gamma(s_N-s_N) & 1\\ 1 & .. & 1 & 0\end{bmatrix}^{-1}\begin{bmatrix}\gamma(s_0-s_1)\\ \gamma(s_0-s_2)\\ .\\ .\\ \gamma(s_0-s_N)\\ 1\end{bmatrix}\tag{5.27}$$

or in a more condensed form:

$$\lambda_0=\Gamma_{ij}^{-1}\gamma_0.\tag{5.28}$$

Matrix $\Gamma$ does not depend on the estimation point, and does not therefore have to be recalculated every time. The values of all $\gamma(s_i-s_j)$ and $\gamma(s_0-s_i)$ are calculated using values from the estimated variogram. The mean squared prediction error r, known as *kriging variance*, is equal to $\lambda_0^t\gamma_0$. In short, kriging provides an unbiased estimator of minimum variance that is also an exact interpolator since, for every known point, it returns an estimated value equal to the observed value.

**Box 5.4.1 — Estimate of kriging equations.** Using the Lagrange multipliers method, we minimise:
$E\left[\hat{Z}_{OK}(s_0)-Z(s_0)\right]^2$ under the constraint $\sum_{i=1}^{N}\lambda_i=1$.

We'll look for $\lambda_1,...,\lambda_N$ and multiplier $m$ that enable the constraint to be introduced. The objective function is therefore written:

$$E\left[\left(\sum_{i=1}^{N}\lambda_iZ(s_i)-Z(s_0)\right)^2\right]-2m\left(\sum_{i=1}^{N}\lambda_i-1\right).\tag{5.29}$$

Due to the constraint, we can write:

$$\left[\sum_{i=1}^{N}\lambda_iZ(s_i)-Z(s_0)\right]^2=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[(Z(s_i)-Z(s_j))^2\right]+\sum_{i=1}^{N}\left[(Z(s_i)-Z(s_0))^2\right].\tag{5.30}$$

By taking the expected value of the expressions, we have:

$$E\left[\sum_{i=1}^{N}\lambda_iZ(s_i)-Z(s_0)\right]^2=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_jE\left[(Z(s_i)-Z(s_j))^2\right]+\sum_{i=1}^{N}\lambda_iE\left[(Z(s_i)-Z(s_0))^2\right].\tag{5.31}$$

This expression brings out the variogram and we can rewrite the constraint as:

$$-\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j\gamma(s_i-s_j)+2\sum_{i=1}^{N}\gamma(s_0-s_i)-2m\left(\sum_{i=1}^{N}\lambda_i-1\right). \tag{5.32}$$

We will minimise this expression by deriving in relation to $\lambda_1,...,\lambda_N$ and $m$, which leads to the kriging equations:

$$\begin{aligned}&\sum_{j=1}^{N}\lambda_j\gamma(s_i-s_j)+m=\gamma(s_0-s_i)\\&\sum_{i=1}^{N}\lambda_i=1.\end{aligned} \tag{5.33}$$

### 5.4.2 Application to rainfall data

#### Raw data

An initial kriging was carried out from raw data using a spherical model for the variogram. Figure 5.12 shows the spherical variogram used, after estimating the parameters using maximum likelihood compared to the experimental variogram.

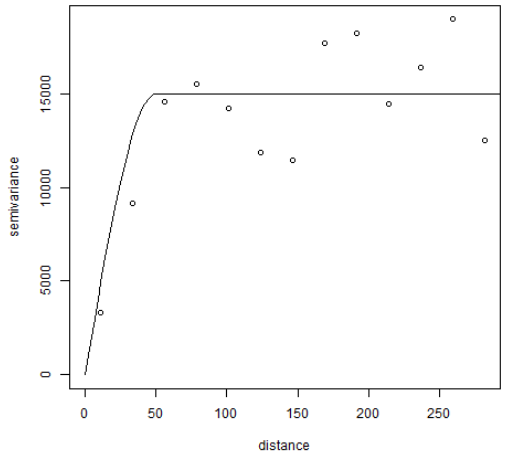


Figure 5.12 – Spherical variogram for raw data
**Source:** *Swiss rainfall from the geoR package*

```
library(geoR)
vario.ex<-  variog(sic.100, bin.cloud=TRUE)
plot(vario.ex,main="")
lines.variomodel(cov.model="spher",cov.pars=c(15000,50),
nug=0,max.dist=300)
```

This way, we can calculate kriged values on a grid, as well as the kriging variances represented in Figure 5.13. The values are represented according to graphical semiology conventions, the warm colours corresponding to the high values.

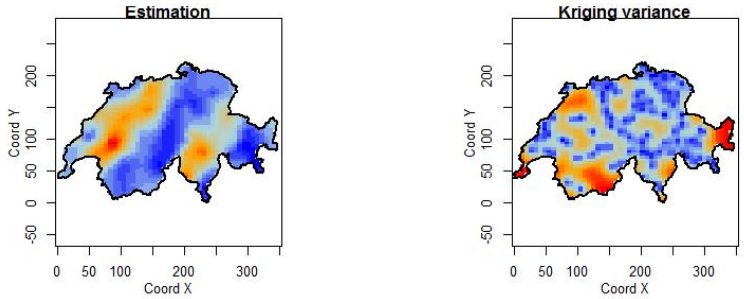


Figure 5.13 – Kriging estimates and variances for raw data
**Source:** *Swiss rainfall from the geoR package*

```
library(geoR)
pred.grid <-  expand.grid(seq(0,350, l=51),seq (0,220, l=51))
rgb.palette <- colorRampPalette(c("blue", "lightblue",
"orange", "red"),space = "rgb")
kc<-  krige.conv(sic.100, loc = pred.grid,
 krige=krige.control(cov.model="spherical",cov.pars=c(15000,50)))
image(kc, loc = pred.grid,col =rgb.palette(20) ,xlab="Coord X",
ylab="Coord Y",borders=sic.borders,main="Estimation")
image(kc, krige.var,loc = pred.grid,col=rgb.palette(20),
xlab="Coord X",ylab="Coord Y",borders=sic.borders,
main="Kriging variance")
```

The estimate was made using the 100 points of the sample. We can:
— confirm that kriging is unbiased on the points observed;
— measure differences between estimated values and observed values.

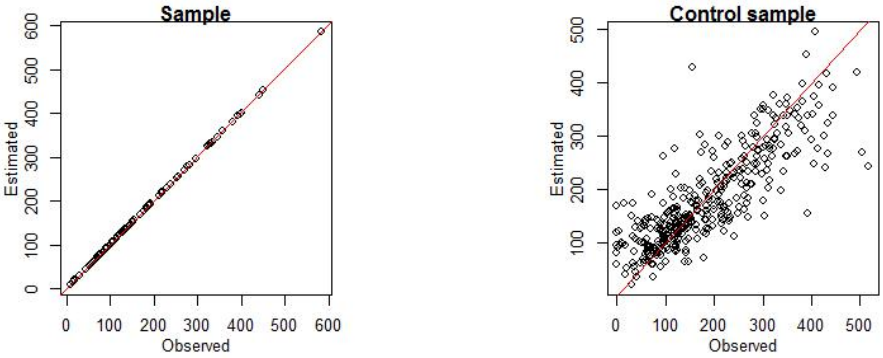


Figure 5.14 – Estimated and observed values
**Source:** *Swiss rainfall from the geoR package*

```
library(geoR)
kc1<-  krige.conv(sic.100, loc = sic.100$coords,
krige=krige.control(cov.model="spherical",cov.pars=c(16000,47)))
kc2<-  krige.conv(sic.100, loc = sic.367$coords,
krige=krige.control(cov.model="spherical",cov.pars=c(16000,47)))
plot(sic.100$data,kc1$predict,xlab="Observed",ylab="Estimated",
```

```
main="Control sample")
abline(a=0,b=1,col="red")
plot(sic.367$data,kc2$predict,,xlab="Observed",ylab="Estimated",
main="Control")
abline(a=0,b=1,col="red")
```

### Transformed data

The histogram in Figure 5.3 indicates that the distribution of rainfall data deviated from a Gaussian distribution. A first possibility, classical in statistics, is to modify variables. In fact, it may be better to work with data that obey a normal law. This is not absolutely necessary in the kriging model, but the linearity assumption in kriging is only really effective when the data are Gaussian. Classical transformations are logarithmic transformations, or more generally Box-Cox transformations. The `boxcoxfit` function of *geoR* suggests a coefficient close to 0.5. Exploratory data are shown in Figure 5.15.

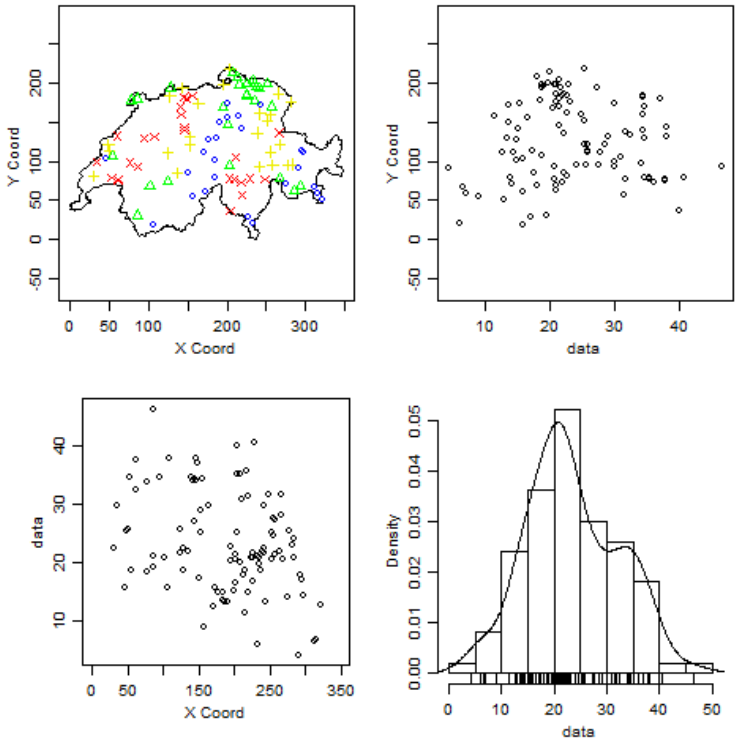


Figure 5.15 – Exploratory data, after Box-Cox transformation
**Source:** *Swiss rainfall from the geoR package*

```
library(geoR)
plot.geodata(sic.100,bor=sic.borders,lambda=0.5)
```

Rainfall data have frequently been studied. Ribeiro Jr et al. 2004 recommends the same transformation of variables. For the variogram, they suggest using a Matern model for which $K = 1$. The parameters of the model are determined using maximum likelihood. The experimental and theoretical variograms for transformed data are shown in Figure 5.16.

```
library(geoR)
```

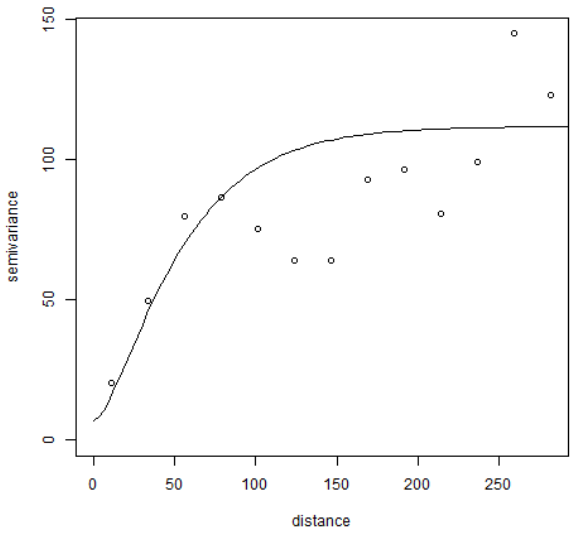


Figure 5.16 – Experimental variogram and theoretical variogram after transformation
**Source:** *Swiss rainfall from the geoR package*

```
vario.ext<-  variog(sic.100,option="bin",lambda=0.5)
plot(vario.ext)
lines.variomodel(cov.m = "mat",cov .p =c (105, 36), nug = 6.9,
                 max.dist = 300,kappa  = 1, lty = 1)
```

As for raw data, we can provide mapping of kriging estimates and values (Figure 5.17).

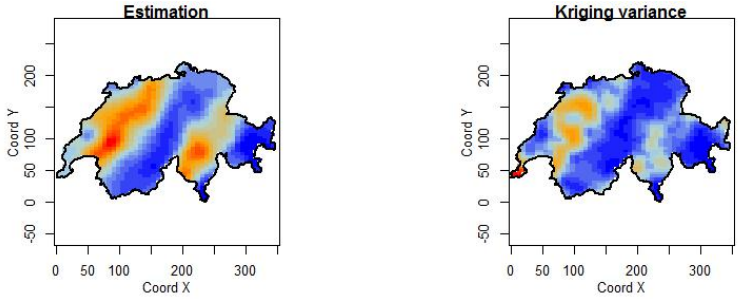


Figure 5.17 – Kriging estimates and variance of rainfall in Switzerland after transformation
**Source:** *Swiss rainfall from the geoR package*

```
library(geoR)
kct<-  krige.conv(sic.100, loc = pred.grid,
krige=krige.control(cov.model="matern",cov.pars=c(105, 36),
kappa=1,nugget=6.9,lambda=0.5))
pred.grid <-  expand.grid(seq(0,350, l=51),seq (0,220, l=51))
rgb.palette <- colorRampPalette(c("blue", "lightblue",
"orange", "red"),space = "rgb")
image(kct, loc = pred.grid,col =rgb.palette(20) , xlab="Coord X",
ylab="Coord Y",borders=sic.borders,main="Estimation")
```

```
image(kct, krige.var,loc = pred.grid,col =rgb.palette(20) ,
xlab="Coord X",ylab="Coord Y",borders=sic.borders,
main="Kriging variance")
```

The estimated values and observed values can be compared (5.18).

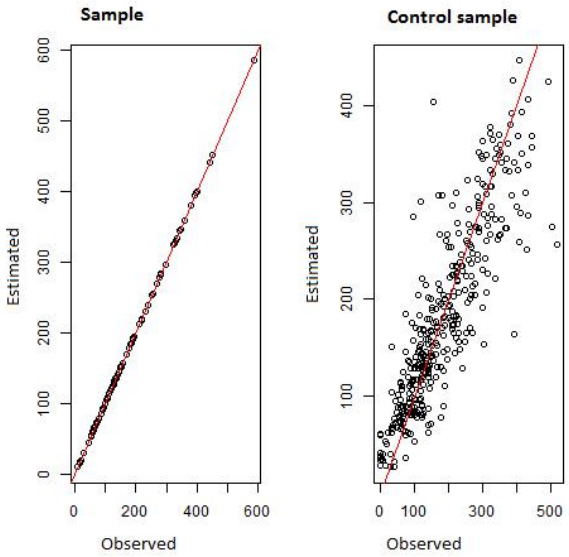


Figure 5.18 – Estimated and observed values
**Source:** *Swiss rainfall from the geoR package*

```
library(geoR)
kct1<-  krige.conv(sic.100, loc = sic.100$coords,
krige=krige.control(cov.model="spherical",cov.pars=c(16000,47),
kappa=1,nugget=6.9,lambda=0.5))
kct2<-  krige.conv(sic.100, loc = sic.367$coords,
krige=krige.control(cov.model="spherical",cov.pars=c(16000,47),
kappa=1,nugget=6.9,lambda=0.5))
plot(sic.100$data,kct1$predict,xlab="Observed",ylab="Estimated",
main="Sample")
abline(a=0,b=1,col="red")
plot(sic.367$data,kct2$predict,,xlab="Observed",ylab="Estimated",
main="Control sample")
abline(a=0,b=1,col="red")
```

If we take the square root of the mean quadratic deviation as a criterion,

$$RMSE(y) = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n}}$$

we see an improvement in prediction if we use the transformed data modelled by the Matern model. In the case of raw data, the value of RMSE is 62.3, while for transformed data it is 55.2.

## 5.5 Support and change of support

The question of analysis scales is of primary importance in spatial analysis. Geographers, especially Openshaw, use the term *Modifiable areal unit problem* (MAUP), discussed in Chapter 1: "Descriptive spatial analysis". The MAUP can be summarised as overlaying a zoning effect and an aggregation effect. This can be illustrated simply using surface data, by showing the variability of results obtained depending on the territorial breakdown used.

In geostatistics, we talk about *change of support problem* (COSP). This problem is actually more general than MAUP, since it refers to size, shape and directions. The COSP in geostatistics originates from very practical concerns linked to mining research, the origins of which can be found in Krige's work, which was then expanded on by Matheron. Krige's articles are contemporary with those by Yule and Kendall in classical statistics, which anticipate those by Openshaw in geography. From a practical viewpoint, geostatisticians quickly realised that it was more important to predict a value over a large block than on one point, even though the first prediction derived from the second.

The support can be a point, a larger or smaller block, or a meeting of points in a given geometric configuration. There is a relationship between the values taken on different volumes. In the context of additive variables, for a volume $V$ partitioned in units $v_i$ of same support $v$ ($V$ being a multiple of $v$), we have:

$$z(V) = \frac{1}{n}\sum_{i=1}^{n} z(v_i) \tag{5.34}$$

or if $v$ is unique:

$$z(V) = \frac{1}{V}\int_V z(x)dx. \tag{5.35}$$

$Z(V)$ is qualified as a *regularised* variable, because it increases statistical regularity. In fact, it is a specific form of regularised variable, the more general form being a *Z convolute* (Chiles et al. 2009).

### 5.5.1 Empirical dispersion variance and Krige additivity relationships

**Definition 5.5.1 — Empirical dispersion variation.**

$$s^2(v|V) = \frac{1}{n}\sum_{i=1}^{n} [z(v_i) - z(V)]^2 \tag{5.36}$$

If V is included in a broader domain called $D$, we can demonstrate relationship 5.37, called the **Krige additivity relationship**:

$$s^2(v|D) = s^2(v|V) + s^2(V|D). \tag{5.37}$$

Armstrong 1998 gives an educational example from the yields in millet on 16 blocks of 2m side, divided into 64 parcels of 1m side. The average is 201 in the case of blocks as in the case of parcels. The variance of the dispersion of the blocks in the field is 16.64, that of the parcels 27.59, which means that the dispersion of the parcels in the blocks is 10.85. (see 5.1).

### 5.5.2 Variogram of the regularised variable

Variance by block can be defined based on information about single data points (covariance function).

$$Var\,[Z(V)] = \overline{C}\,(V,V) = \frac{1}{|V|^2}\int_V\int_V C(x-y)\,dx\,dy \tag{5.38}$$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 735 | 325 | 45 | 140 | 125 | 175 | 167 | 485 |
| 540 | 420 | 260 | 128 | 20 | 30 | 105 | 70 |
| 450 | 200 | 337 | 190 | 95 | 260 | 245 | 278 |
| 180 | 250 | 380 | 405 | 250 | 80 | 515 | 605 |
| 124 | 120 | 430 | 175 | 230 | 120 | 460 | 260 |
| 40 | 135 | 240 | 35 | 190 | 135 | 160 | 170 |
| 75 | 95 | 20 | 35 | 32 | 95 | 20 | 450 |
| 200 | 35 | 100 | 59 | 2 | 45 | 58 | 90 |

| | | | |
|---|---|---|---|
| 505 | 143 | 88 | 207 |
| 270 | 328 | 171 | 411 |
| 102 | 220 | 154 | 263 |
| 101 | 54 | 44 | 155 |

Table 5.1 – Observed values on parcels (top) and blocks (bottom)
**Source :** *Armstrong 1998.*

where $C$ designates the covariance function for single data points and $\overline{C}$ designates covariance of the data by block. Covariance is written:

$$Cov\left[Z(V),Z(V')\right] = \overline{C}(V,V') = \frac{1}{|V||V'|}\int_V\int_{V'} C(x-y)\,dx\,dy. \tag{5.39}$$

Covariance function $C_V$ is defined by introducing $V_h$ which is the translation of support $V$ by vector $h$:

$$C_V(h) = Cov\left[Z(V),Z(V_h)\right] = \overline{C}(V,V_h). \tag{5.40}$$

From the unique variogram we can also deduce the variogram of the regularised variable:

$$\gamma_V(h) = \overline{\gamma}(V,Vh) - \overline{\gamma}(V,V) \tag{5.41}$$

with $\gamma(V,V) = \frac{1}{|V|^2}\int_V\int_V \gamma(x-y)\,dx\,dy$ and $\overline{\gamma}(V,V_h) = \frac{1}{|V|^2}\int_V\int_{V_h} \gamma(x-y)\,dx\,dy$ where $\gamma$ is the variogram calculated from unique observations.

We can then show that $\gamma_V(h) \sim \gamma(h) - \overline{\gamma}(V,V)$, leading to the graph in Figure 5.19.

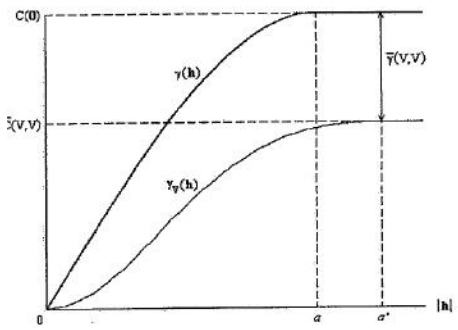


Figure 5.19 – Switch from unique variogram to regularised variogram

To switch from unique to regularised variogram, the same type of theoretical model is retained by correcting the sill and the range.

### 5.5.3 Block kriging

We can determine kriging equations, in the logic of ordinary kriging. From $E[Z(x)] = m$, we deduce that $E[Z_V] = m$.

As for ordinary kriging, we are looking for an estimator that is a linear combination of the observations collected.

$$Z_V = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{5.42}$$

The approach is the same as that shown in box 5.4.1 (minimisation under constraint) and we get the following kriging equations:

$$\begin{aligned} &\sum_{j=1}^{n} \lambda_j \gamma(x_i, x_j) + \mu = \gamma(x_i, V) \quad pour \quad i = 1, 2, ..., n \\ &\sum_{i=1}^{n} \lambda_i = 1 \end{aligned} \tag{5.43}$$

where $\overline{\gamma}(x_i, V) = \frac{1}{|V|} \int_V \gamma(x_i - x)\, dx$.

## 5.6 Extensions

Ordinary kriging is the basic method of geostatistics. But in the end, it is only part of it, and there have been many developments, particularly at the Ecole des Mines in Fontainebleau.

Intrinsic stationarity assumptions remain fairly restrictive, in particular the constancy of the mean. Many methods have been developed to introduce less constraining assumptions, or to use supplementary information. In this paragraph, a number of modifications and variants will be presented, mainly from Swiss rainfall data, and another frequently-used dataset on the contents of various minerals in a meander of the River Meuse.

### 5.6.1 Cokriging

Geostatistics developed multivariate methods long ago. One of them is Cokriging, which will consider several variables. It was defined by Waller et al. 2004 as follows:

**Definition 5.6.1 — cokriging.** "Cokriging is an extension of kriging to the case of two or more spatial variables. It was originally developed as a technique for improving the prediction of a variable for which only a few samples could be taken, by using its spatial correlation with other more easily measured variables. Cokriging differs from kriging with external drift in that the explanatory variables are no longer assumed to be fixed variables that indicate the nature of a trend in the primary variable, but are themselves spatial random variables with expected values and variograms.".

A cross co-variogram can be defined:

$$\gamma_{ZY}(h) = \frac{1}{2p(h)} \sum_{i=1}^{p(h)} (z(s_i) - z(s_i + h))\,(y(s_i) - y(s_i + h)) \tag{5.44}$$

with $p(h) = Card\left\{(s_i, s_j) \middle| \left|s_i - s_j\right| \approx h\right\}$

Must like kriging, there will be several versions for cokriging. We will only discuss ordinary cokriging. We will restrict ourselves to the case where only one auxiliary variable is introduced, which will be called $Y$. The estimator we calculate takes the form:

$$Z(s_0) = \sum_{i=1}^{n_Z} \lambda_i Z(s_i) + \sum_{i=1}^{n_Y} \alpha_i Y(s_i) \tag{5.45}$$

with bias-free constraints:

$$\begin{aligned} &\sum_{i=1}^{n_Z} \lambda_i = 1 \\ &\sum_{i=1}^{n_Y} \alpha_i = 0. \end{aligned} \tag{5.46}$$

In matrix form, the cokriging equations are written:

$$\begin{bmatrix} \Gamma_{ZZ} & \Gamma_{ZY} & 1 & 0 \\ \Gamma_{YZ} & \Gamma_{YY} & 0 & 1 \\ 1' & 0' & 0 & 0 \\ 0 & 1' & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \alpha \\ \mu_Z \\ \mu_Y \end{bmatrix} = \begin{bmatrix} \gamma_{ZZ} \\ \gamma_{YZ} \\ 1 \\ 0 \end{bmatrix} \tag{5.47}$$

This method will be illustrated using River *Meuse* data supplied in the *sp* package and particularly studied by Pebesma (Pebesma 2001) and Rossiter (Rossiter 2017). The latter author provides R programs on his website.

■ **Example 5.1** Analysis of River *Meuse* data by Cokriging

The River *Meuse* data provide localised measurements of lead, zinc and cadmium contents, but also other variables such as altitude and organic matter content of the soil. The *sp* package contains a table called *Meuse*, which can be loaded with the `data(meuse)` function. It also provides a 40x40 m grid – *meuse.grid* – and boundaries of the "département" – *meuse.riv*.

The data comprise 155 observations on 15x15 m supports, over the upper 20 cm of alluvial soils on the right bank of the River Meuse. Associated data provide the geographical coordinates of the observations, their altitude, and concentrations of cadmium, copper, lead, zinc and organic matter. There is also the distance to the River Meuse and the flooding frequency. In the example provided by Rossiter 2017, the lead content is studied (after logarithmic transformation). The organic matter content that will be used as covariable.

The variogram analysis for Cokriging is based on studying the two simple variograms and the cross-variogram.

The same exercise is performed using the logarithm of the zinc content as covariable. The figure below shows the kriging results for the lead content, by ordinary kriging and by cokriging, using the two variables mentioned above (zinc content and organic matter). Figure 5.22 shows the kriged values (left column) and the residual values (right column). It successively shows ordinary kriging, cokriging with organic matter as co-variable, and cokriging with the zinc content.

The code required for this processing is quite long. The link below provides access to the R programme made available by Rossiter (Rossiter 2007)

`http://www.css.cornell.edu/faculty/dgr2/teach/R/ck_plotfns.R`.

If we compare the three models using RMSE, we find the following results:

— 0.166 for ordinary kriging;

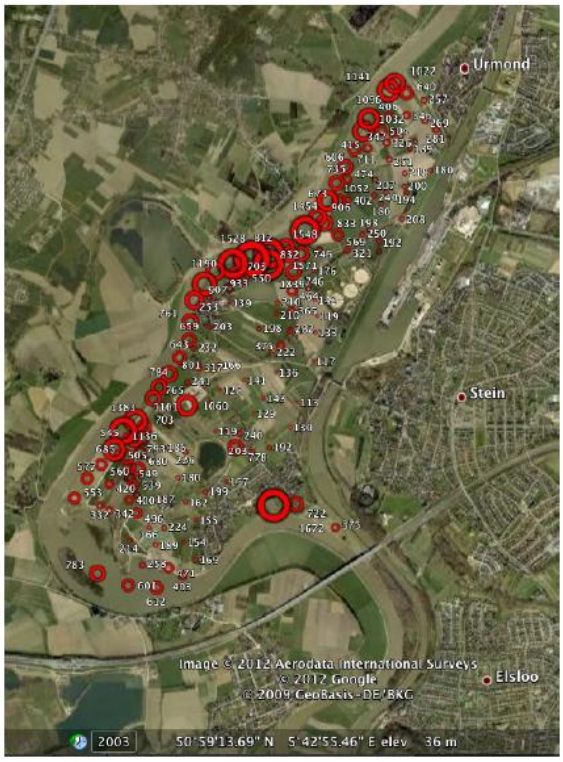


Figure 5.20 – Geography and sample locations
**Source:** *Meuse data from the sp package*

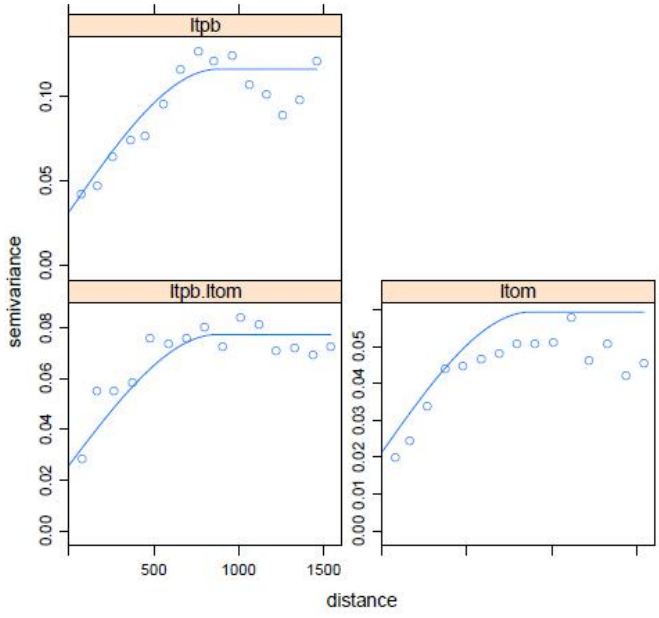


Figure 5.21 – Simple and cross variograms
**Source:** *Meuse data from the sp package*

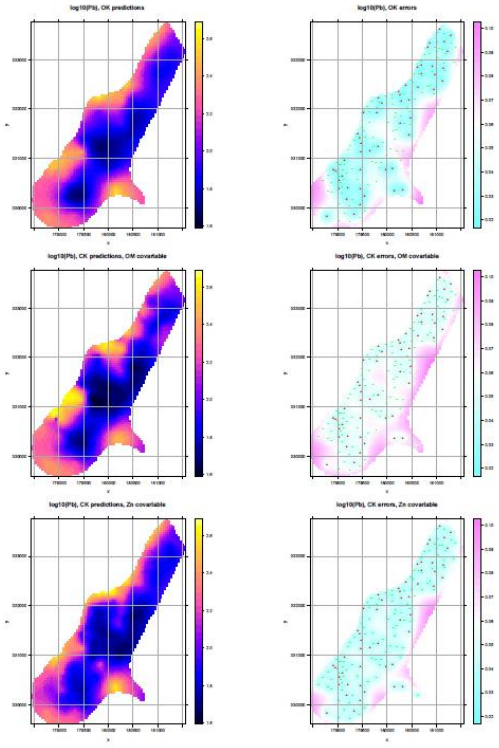

Figure 5.22 – Kriging and Cokriging estimate and variance
**Source:** *Meuse data from the sp package*

— 0.226 for cokriging with organic matter content;
— 0.078 for cokriging with zinc content.

Cokriging involving zinc content therefore improves the performance of ordinary kriging. This is not true for cokriging with organic matter content. ■

### 5.6.2 Universal kriging

In many cases, the average value is not constant, and ordinary kriging cannot be used. This applies when deterministic relationships are observed between the value of the variable and its position in space. The regionalized variable can then be written:

$$Z(s) = m(s) + Y(s) \tag{5.48}$$

where $m(s)$ represents the deterministic component.

■ **Example 5.2 — Analysis of River Meuse data by cokriging.** The River Meuse data provide two variables likely to build a deterministic component — distance to the river and flooding frequency (Figure 5.23). They are different from the co-variables used previously in cokriging.

On his website, Rossiter provides examples used to compare the predictions from normal kriging and from two universal kriging models.

If we compare the three models using RMSE, we find the following results:
— 0.173 for ordinary kriging;
— 0.141 for the model with flooding frequency;
— 0.145 for the model with flooding frequency and distance to the river.

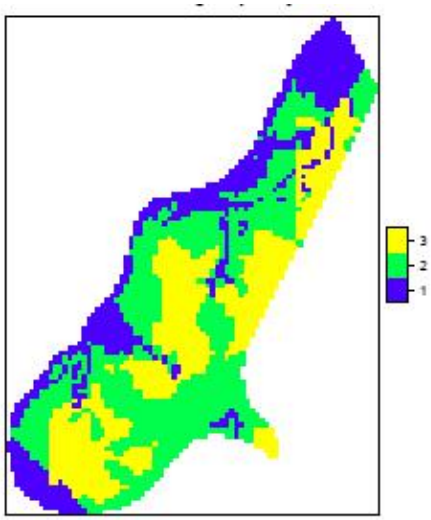


Figure 5.23 – Flooding frequency
**Source:** *Meuse data from the sp package*

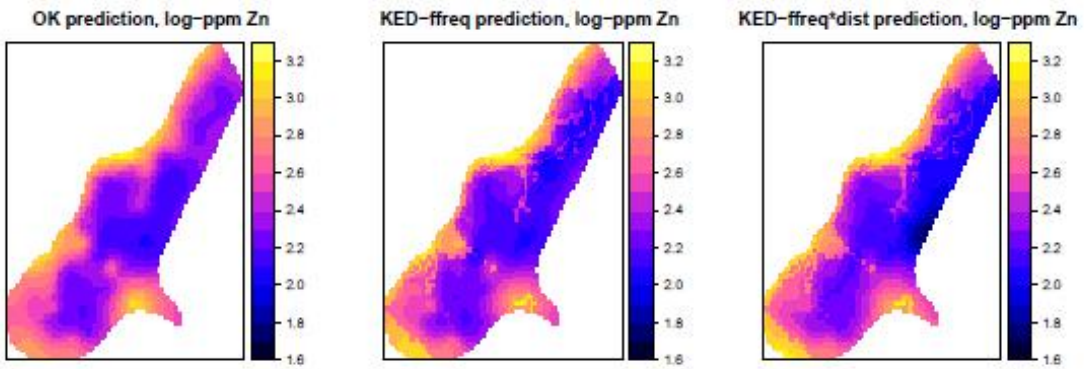


Figure 5.24 – Three estimates
**Note:** ordinary kriging (left), flooding frequency (middle), flooding frequency and distance to the river (right)
**Source:** *Meuse data from the sp package*

Introducing flooding frequency improves predictions, but introducing the distance also degrades it (probably due to correlation of the two variables). The R code used is available at the address below. It is a good example to illustrate and extend what has been discussed in this chapter.

`http://www.css.cornell.edu/faculty/dgr2/teach/R/gs_short_ex.pdf` ■

## 5.7 Combined models with variogram

Other uses can be found for tools developed in geostatistics. When performing regressions on spatial data, the residuals are frequently spatially autocorrelated. This correlation can be highlighted by the Moran spatial autocorrelation indicator (see Chapter 3: "Spatial autocorrelation indices"). This autocorrelation can be taken into account using spatial econometric models (see Chapter 6: "Space econometrics: current models") or geographically-weighted regression (see Chapter 9: "Geographically-weighted regression").

When the data are suitable, the variogram can also be used to study the spatial structure of residuals in linear models. Examples are more frequently described in books dealing with ecology or epidemiology. Illustrations can be found in general spatial statistical handbooks such as Waller et al. 2004 or Schabenberger et al. 2017, as well as books dealing with ecology, such as Plant 2012 or Zuur et al. 2009, the last two providing examples of implementation in R.

■ **Example 5.3 — Analysis of the spatial structure of residuals with a variogram.** An example of use on ecological data is provided by Zuur et al. 2009 [2]. The example comes from data collected over forests in the Raifa section of the Voljsko-Kamsky state natural biosphere.

The interest variable is a 'boreality' index (*Bor*) defined as the share of specifically boreal species compared to the total number of species on a site. There are also explanatory variables provided by satellite images:

1. standardised vegetation difference index;
2. temperature;
3. moisture index;
4. greening index.

Due to the strong co-linearity between these variables, only moisture was used to explain the boreality index. The variance analysis performed using ordinary least squares provides the following results:

| Variable | Estimated value | standard deviation |
|---|---|---|
| Constant | 27.63 | 0.981 |
| Wet | 429.609 | 27.45 |

Table 5.2 – Least squares estimate

The addresses of the sites are used to provide an exploratory vision of the spatialisation of residues of the OLS model.

```
library(sp)
library(nlme)
Boreality<-read.table("C:/jmf/Boreality.txt",header=TRUE)
B.lm <- lm(boreal ~ Wet, data = Boreality)
```

2. https://github.com/James-Thorson/2016_class_CMR/tree/master/Other%20material/Zuur%20et%20al.%202007/ZuurDataMixedModelling

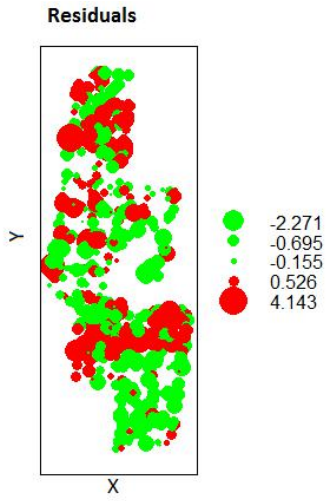


Figure 5.25 – Residuals
**Source:** *Meuse data from the sp package*

```
E <- rstandard(B.lm)
graphic <- data.frame(E, Boreality$x, Boreality$y)
library(sp)
coordinates(graphic) <- c("Boreality.x","Boreality.y")
bubble(graphic, "E",col = c ("green","red"),main = "Residuals",
xlab = "X", ylab = "Y")
```

The OLS model does not allow the introduction of spatial structure on residues. It can only be introduced in a generalised linear model. It will be estimated using the `gls` function of the R *nlme* package. This package contains a function to estimate the variogram.
The experimental variogram is shown below (Figure 5.26)

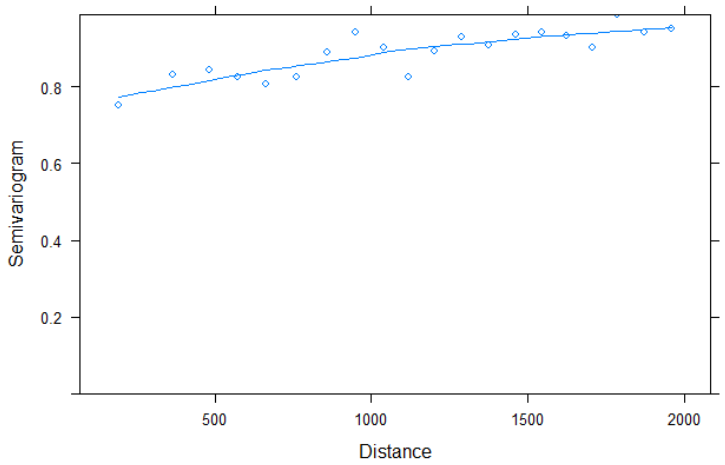


Figure 5.26 – Variogram of residuals
**Source:** *Meuse data from the sp package*

```
mod<-gls(boreal~Wet,data=bor)
summary(mod)
plot(Variogram(mod,form=~x+y,maxdist=10000),xlim=c(0,10000))
```

The initial estimate from the model, without the introduction of the spatial structure, gives the same results as OLS. The `update` command is used to introduce a spatial structure using a variogram analysis and to re-estimate the model.

Table 5.3 shows the result of the comparison between OLS and models using spherical, Gaussian and exponential variograms.

| Variogram | AIC | Likelihood | L | Significance |
|---|---|---|---|---|
| OLS | 3855 | -1924 | | |
| Spherical | 3859 | -1924 | 1 | 1 |
| Gaussian | 3750 | -1870 | 109 | <0.001 |
| Exponential | 3740 | -1865 | 119 | <0.001 |

Table 5.3 – AIC and L criteria based on variogram type

The AIC and L criteria show that the model can be improved by using a spherical variogram to model the residues. The regression results with the exponential variogram are shown in Table 5.4.

| Variable | Estimated value | standard deviation |
|---|---|---|
| constant | 18.099 | 2.333 |
| Wet | 180.247 | 34.932 |

Table 5.4 – Estimates

```
f1 <- formula(boreal ~ Wet)
B1.gls <- gls(f1, data = Boreality)
Vario.gls <- Variogram(B1.gls, form =~ x + y,robust = TRUE,
maxDist = 2000,resType = "pearson")
B1A <- gls(f1, correlation = corSpher(form =~ x + y,nugget = TRUE),
data = Boreality)
B1B <- gls(f1, correlation = corLin(form =~ x + y,nugget = TRUE),
data = Boreality)
B1C <- gls(f1, correlation = corRatio(form =~ x + y,nugget = TRUE),
data = Boreality)
B1D <- gls(f1, correlation = corGaus(form =~ x + y,nugget = TRUE),
data = Boreality)
B1E <- gls(f1, correlation = corExp(form =~ x + y,nugget = TRUE),
data = Boreality)
AIC(B1.gls, B1A, B1B, B1C, B1D, B1E)
B1 <- lm(f1, data = Boreality)
anova(B1.gls,B1A)
anova(B1.gls,B1D)
anova(B1.gls,B1E)
summary(B1E)
```

The parameters of the model are significant. The influence of moisture is less pronounced when spatial autocorrelation is introduced into the model. ■

## Conclusion

The first chapter of this handbook presents the three main areas of spatial statistics appropriate to the analysis of continuous, surface or single data points. Geostatistics, used for continuous data, is less directly linked to the work of public statistics. Nonetheless, it seemed useful to provide a quick description in the handbook. From an educational viewpoint, geostatistical methods illustrate particularly well how considering spatial autocorrelation (through the variogram) makes it possible to improve the estimators. From a more operational viewpoint, without going into the complexity of mining research work, geostatistics *using* kriging methods is useful for modelling simpler continuous data (*e.g.* climate data). The Fontainebleau Ecole des Mines, which has played a crucial part in developing of these methods, has used quite unusual language for statisticians, but many discussions have been held since Cressie's work to link the different approaches. The classic book by Chilès and Delfiner is a good example of this (Chiles et al. 2009). In the healthcare field, significant work has involved geostatistical methods to model epidemiological data, particularly that by Diggle (Diggle et al. 2003), who is better known for his work on *ad hoc* methods. Finally, we can only recommend that statisticians who may use models in the future should read the exploratory article by the founder of geostatistics (Matheron 1978).

# Appendices

## Mathematical reminders

The expressions of theoretical variograms, in particular Matern's variogram, use quite unusual mathematical expressions, in particular the expression below (Chiles et al. 2009).

### The Gamma function

$\Gamma(x) = \int_0^\infty e^{-u} u^{x-1} du$

For whole number values:

$\Gamma(n+1) = n!$

### The Bessel functions

The Bessel function of the first kind is as follows:

$J_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^\infty \frac{(-1)^k}{k!\Gamma(\nu+k+1)} \left(\frac{x}{2}\right)^{2k}$

The modified Bessel function of the first kind is as follows:

$I_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^\infty \frac{1}{k!\Gamma(\nu+k+1)} \left(\frac{x}{2}\right)^{2k}$

The modified Bessel function of the second kind is defined from the previous type

$K_\nu(x) = \frac{\pi}{2} \frac{I_{-\nu}(x)+I_\nu(x)}{sin\pi\nu}$

## References - Chapter 5


Armstrong, Margaret (1998). *Basic linear geostatistics*. Springer Science & Business Media.

Chauvet, Pierre (2008). *Aide-mémoire de géostatistique linéaire*. Presses des MINES.

Chiles, Jean-Paul and Pierre Delfiner (2009). *Geostatistics: modeling spatial uncertainty*. Vol. 497. John Wiley & Sons.

Diggle, Peter J, Paulo J Ribeiro Jr, and Ole F Christensen (2003). « An introduction to model-based geostatistics ». *Spatial statistics and computational methods*. Springer, pp. 43–86.

Joly, Daniel et al. (2009). « Interpolation par régressions locales: application aux précipitations en France ». *L'Espace géographique* 38.2, pp. 157–170.

Lloyd, Christopher D and Peter M Atkinson (2004). « Increased accuracy of geostatistical prediction of nitrogen dioxide in the United Kingdom with secondary data ». *International Journal of Applied Earth Observation and Geoinformation* 5.4, pp. 293–305.

Matheron, Georges et al. (1965). *Les variables régionalisées et leur estimation*. Masson et Cie.

Matheron, Georges (1978). *Estimer et choisir: essai sur la pratique des probabilités*. Ecole nationale supérieure des mines de Paris.

Pebesma, Edzer J (2001). « Gstat user's manual ». *Dept. of Physical Geography, Utrecht University, Utrecht, The Netherlands*.

Plant, Richard E (2012). *Spatial data analysis in ecology and agriculture using R*. cRc Press.

Ribeiro Jr, Paulo J and Peter J Diggle (2006). « geoR: Package for Geostatistical Data Analysis An illustrative session ». *Artificial Intelligence* 1, pp. 1–24.

Ribeiro Jr, Paulo Justiniano and Peter J Diggle (2004). « Model Based Geostatistics ». *Springer Series in Statistics*.

Rossiter, David G (2017). « An introduction to geostatistics with R/gstat Version 3.7, 12-May-2017. »

Rossiter, DG (2007). « Co-kriging with the gstat package of the R environment for statistical computing ». *Web: http://www. itc. nl/rossiter/teach/R/R ck. pdf*.

Schabenberger, Oliver and Carol A Gotway (2017). *Statistical methods for spatial data analysis*. CRC press.

Waller, Lance A and Carol A Gotway (2004). *Applied spatial statistics for public health data*. Vol. 368. John Wiley & Sons.

Zuur, AF et al. (2009). « Mixed effects models and extensions in ecology with R. Gail M, Krickeberg K, Samet JM, Tsiatis A, Wong W, editors ». *New York, NY: Spring Science and Business Media*.