The mission of the Institute for Environment and Sustainability is to provide scientific-technical support to the European Union's Policies for the protection and sustainable development of the European and global environment.

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information:
Address: JRC Ispra, Via E. Fermi 1, I-21020 Ispra (VA), Italy
Tel.: +39- 0332-785349
Fax: +39- 0332-786394

http://ies.jrc.ec.europa.eu
http://www.jrc.ec.europa.eu

JRC 38153

EUR 22904 EN
ISBN 978-92-79-06904-8
ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

case of statistical models, we need to follow several statistical data analysis steps before we can generate maps. This makes the whole mapping process more complicated but it eventually helps us: (a) produce more reliable/objective maps, (b) understand the sources of errors in the data and (c) depict problematic areas/points that need to be revisited.

1.3.1 Kriging

Kriging has for many decades been used as a synonym for geostatistical interpolation. It originated in the mining industry in the early 1950's as a means of improving ore reserve estimation. The original idea came from the mining engineers D. G. Krige and the statistician H. S. Sichel. The technique was first published in Krige (1951), but it took almost a decade until a French mathematician G. Matheron derived the formulas and basically established the whole field of linear geostatistics[10] (Cressie, 1990; Webster and Oliver, 2001; Zhou et al., 2007).

A standard version of kriging is called **ordinary kriging** (**OK**). Here the predictions are based on the model:

$$Z(\mathbf{s}) = \mu + \varepsilon'(\mathbf{s}) \tag{1.3.1}$$

where μ is the constant *stationary* function (global mean) and $\varepsilon'(\mathbf{s})$ is the spatially correlated stochastic part of variation. The predictions are made as in Eq.(1.2.1):

$$\hat{z}_{\texttt{OK}}(\mathbf{s}_0) = \sum_{i=1}^{n} w_i(\mathbf{s}_0) \cdot z(\mathbf{s}_i) = \lambda_{\mathbf{0}}^{\mathbf{T}} \cdot \mathbf{z} \tag{1.3.2}$$

where $\lambda_{\mathbf{0}}$ is the vector of kriging weights (w_i), $\mathbf{z}$ is the vector of n observations at primary locations. In a way, kriging can be seen as a sophistication of the inverse distance interpolation. Recall from §1.2.1 that the key problem of inverse distance interpolation is to determine how much importance should be given to each neighbour. Intuitively thinking, there should be a way to estimate the weights in an objective way, so the weights reflect the true spatial autocorrelation structure. The novelty that Matheron (1962) and Gandin (1963) introduced to the analysis of point data is the derivation and plotting of the so-called **semivariances** — differences between the neighbouring values:

$$\gamma(\mathbf{h}) = \frac{1}{2} E\left[\left(z(\mathbf{s}_i) - z(\mathbf{s}_i + \mathbf{h})\right)^2\right] \tag{1.3.3}$$

where $z(\mathbf{s}_i)$ is the value of target variable at some sampled location and $z(\mathbf{s}_i + \mathbf{h})$ is the value of the neighbour at distance $\mathbf{s}_i + \mathbf{h}$. Suppose that there are n point observations, this yields $n \cdot (n-1)/2$ pairs for which a semivariance can be calculated. We can then plot all semivariances versus their distances, which will produce a variogram cloud as shown in Fig. 1.7b. Such clouds are not easy to describe visually, so the values are commonly averaged for standard distance called the **lag**. If we display such averaged data, then we get a standard **experimental variogram** as shown in Fig. 1.7c. What we usually expect to see is that semivariances are smaller at shorter distance and then they stabilize at some distance. This can be interpreted as follows: the values of a target variable are more similar at shorter distance, up to a certain distance where the

[10] Matheron (1962) named his theoretical framework the *Theory of Regionalized Variables*. It was basically a theory for modelling stochastic surfaces using spatially sampled variables.

differences between the pairs are more less equal the global variance[11]. This is known as the *spatial auto-correlation effect*.

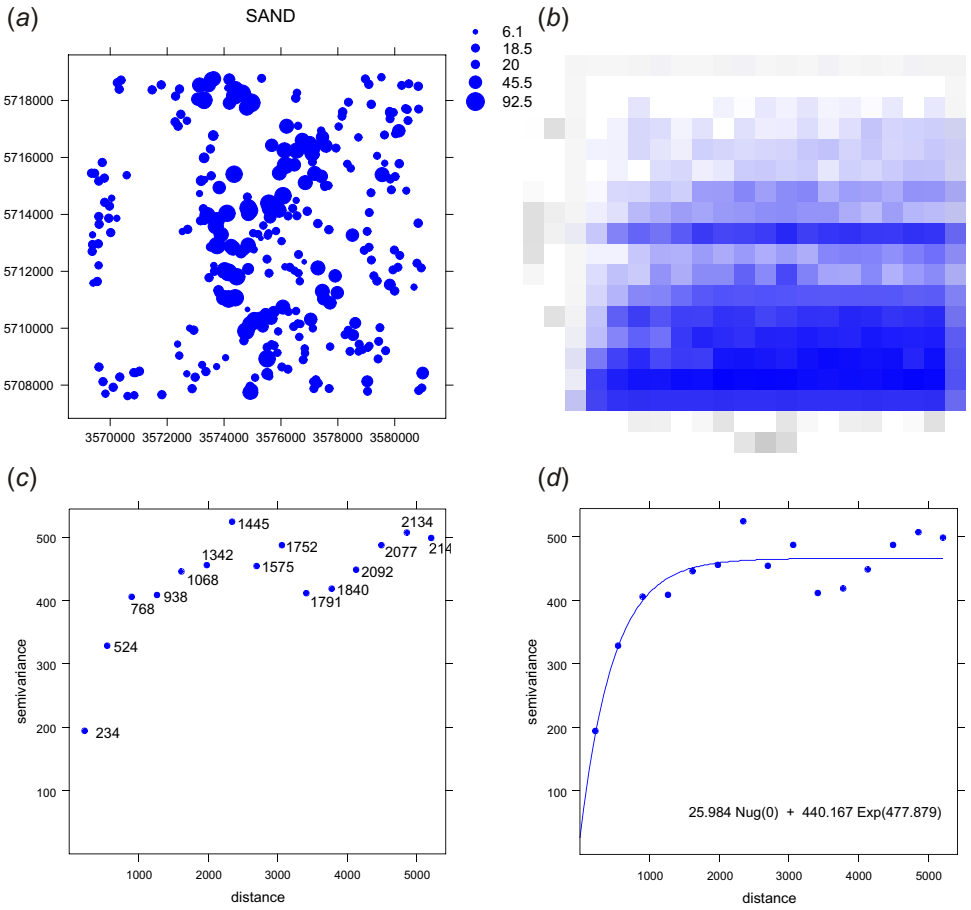

Fig. 1.7: Steps of variogram modelling: (a) location of points (300), (b) variogram cloud showing semivariances for 44850 pairs, (c) semivariances aggregated to lags of about 300 m, and (d) the final variogram model fitted using the default settings in gstat.

From a meta-physical perspective, spatial auto-correlation in the data can be considered as a result of **diffusion** — a random motion causing a system to decay towards uniform conditions. One can argue that, if there is a physical process behind a feature, one should model it using a deterministic function rather than to treat it as a stochastic component. However, diffusion is a random motion so that there is a meta-statistical argument to treat it is a stochastic component.

Once we calculated an experimental variogram, we can fit it using some of the **authorized variogram models**, such as *linear*, *spherical*, *exponential*, *circular*, *Gaussian*, *Bessel*, *power* and similar (Isaaks and Srivastava, 1989; Goovaerts, 1997). The variograms are commonly fitted by iterative reweighted least squares estimation, where the weights are determined based on the number of point pairs or based on the distance. Most commonly, the weights are determined using $N_j/\mathbf{h}_j^2$, where N_j is the number of pairs at certain lag, and $\mathbf{h}_j$ is the distance (Fig. 1.7d). This means that the algorithm will give much more importance to semivariances with large number of point pairs and to the shorter distances. Fig. 1.7d shows the result of automated variogram fitting given

[11]For this reason, many geostatistical packages (e.g. lsatis) automatically plot the global variance (horizontal line) directly in a variogram plot.

an experimental variogram (Fig. 1.7c) and using the $N_j/\mathbf{h}_j^2$-weights: in this case, we obtained an exponential model with the nugget parameter = 26, sill parameter = 440, and the range parameter = 478 m. Note that this is only a **sample variogram** — if we would go and collect several point samples, each would lead to somewhat different variogram plot. The target variable is said to be *stationary* if several sample variograms are very similar (constant), which is referred to as the **covariance stationarity**. Otherwise, if the variograms differ much locally and/or globally, then we speak about a non-stationary inherent properties. In principle, assumptions of kriging are that the target variable is stationary and that it has a normal distribution, which is probably the biggest limitation of kriging[12]. It is also important to note that there is a difference between the range factor and the range of spatial dependence, also known as the **practical range**. A practical range is the Lag **h** for which $\gamma(\mathbf{h})$=0.95 $\gamma(\infty)$, i.e. that distance at which the semivariance is close to 95% of the sill (Fig. 1.8b).

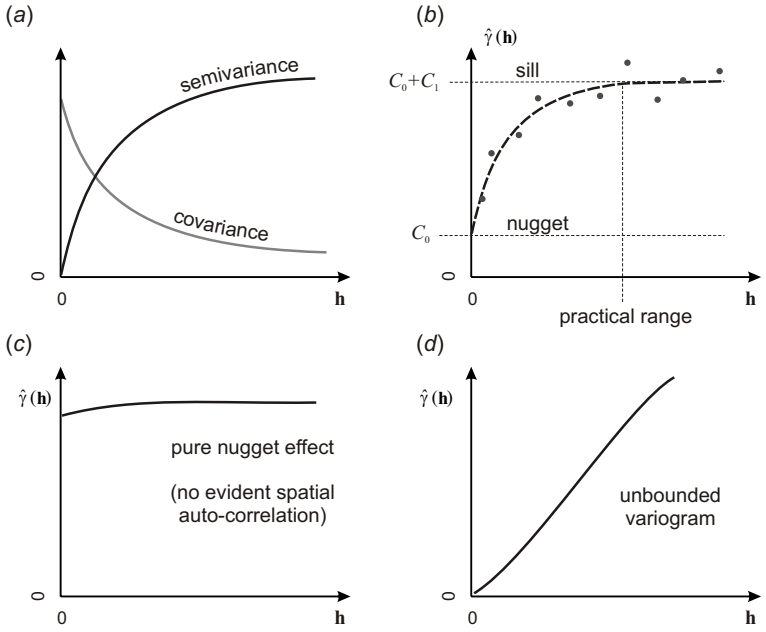

Fig. 1.8: Some basic concepts of variograms: (a) the difference between semivariance and covariance; (b) it often important in geostatistics to distinguish between the sill variation ($C_0 + C_1$) and the sill parameter (C_1) and between the range parameter (R) and the practical range; (c) a variogram that shows no spatial correlation can be defined by a single parameter (C_0); (d) an unbounded variogram typically leads to predictions similar to inverse distance interpolation.

Once we have estimated[13] the variogram model, we can use it to derive semivariances at all locations and solve the kriging weights. The kriging OK weights are solved by multiplying the covariances:

$$\lambda_\mathbf{0} = \mathbf{C}^{-1} \cdot \mathbf{c_0}; \qquad C(|\mathbf{h}| = 0) = C_0 + C_1 \tag{1.3.4}$$

where $\mathbf{C}$ is the covariance matrix derived for $n \times n$ observations and $\mathbf{c_0}$ is the vector of

[12]The constant variogram/histogram and normality are rarely tested in real case studies, which can lead to poor predictions (although the output maps might appear to be fine). In the case of regression-kriging (see further §2.1), the variable does not have to be stationary, so no need to test this property.

[13]We need to determine the parameters of the variogram model: e.g. the nugget (C_0), sill (C_1) and the range (R) parameter. By knowing these parameters, we can estimate the semivariance at any location in the area of interest.

covariances at new location. Note that the $\mathbf{C}$ is in fact $(n+1) \times (n+1)$ matrix if it is used to derive kriging weights. One extra row and column are used to ensure that the sum of weights is equal to one:

$$\begin{bmatrix} C(\mathbf{s}_1,\mathbf{s}_1) & \cdots & C(\mathbf{s}_1,\mathbf{s}_n) & 1 \\ \vdots & & \vdots & \vdots \\ C(\mathbf{s}_n,\mathbf{s}_1) & \cdots & C(\mathbf{s}_n,\mathbf{s}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}^{-\mathbf{1}} \cdot \begin{bmatrix} C(\mathbf{s}_0,\mathbf{s}_1) \\ \vdots \\ C(\mathbf{s}_0,\mathbf{s}_n) \\ 1 \end{bmatrix} = \begin{bmatrix} w_1(\mathbf{s}_0) \\ \vdots \\ w_n(\mathbf{s}_0) \\ \varphi \end{bmatrix} \tag{1.3.5}$$

where φ is the so-called *Langrange multiplier*.

In addition to estimation of values at new locations, a statistical spatial prediction technique offers a measure of associated uncertainty of making these estimations by using a given model. In geostatistics, this is often referred to as the **prediction variance**, i.e. the estimated variance of the prediction error. OK variance is defined as the weighted average of covariances from the new point $(\mathbf{s}_0)$ to all calibration points $(\mathbf{s}_1, .., \mathbf{s}_n)$, plus the Lagrange multiplier (Webster and Oliver, 2001, p.183):

$$\begin{aligned} \hat{\sigma}^2_{\mathtt{OK}}(\mathbf{s}_0) &= (C_0 + C_1) - \mathbf{c_0^T} \cdot \boldsymbol{\lambda_0} \\ &= C_0 + C_1 - \sum_{i=1}^{n} w_i(\mathbf{s}_0) \cdot C(\mathbf{s}_0,\mathbf{s}_i) + \varphi \end{aligned} \tag{1.3.6}$$

where $C(\mathbf{s}_0,\mathbf{s}_i)$ is the covariance between the new location and the sampled point pair, and φ is the Lagrange multiplier, as shown in Eq.(1.3.5).

As you can notice, outputs from any statistical prediction model are always two maps: (1) predictions and (2) prediction variance. The mean of the prediction variance at all location can be termed the **overall prediction variance**, and can be used as a measure of how precise is our final map: if the overall prediction variance gets close to the global variance, then the map is 100% imprecise; if the overall prediction variance tends to zero, then the map is 100% precise[14].

Note that a common practice in geostatistics is to model the variogram using a semivariance function and then, for the reasons of computational efficiency, use the **covariances**. In the case of solving the kriging weights, both the matrix of semivariances and covariances give the same results, so you should not really make a difference between the two. The relation between the covariances and semivariances is (Isaaks and Srivastava, 1989, p.289):

$$C(\mathbf{h}) = C_0 + C_1 - \gamma(\mathbf{h}) \tag{1.3.7}$$

where $C(\mathbf{h})$ is the covariance, and $\gamma(\mathbf{h})$ is the semivariance function (Fig. 1.8a). So for example, exponential model can be written in two ways:

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \text{if} \quad |\mathbf{h}| = 0 \\ C_0 + C_1 \cdot \left[1 - e^{-\left(\frac{\mathbf{h}}{R}\right)}\right] & \text{if} \quad |\mathbf{h}| > 0 \end{cases} \tag{1.3.8}$$

$$C(\mathbf{h}) = \begin{cases} C_0 + C_1 & \text{if} \quad |\mathbf{h}| = 0 \\ C_1 \cdot \left[e^{-\left(\frac{\mathbf{h}}{R}\right)}\right] & \text{if} \quad |\mathbf{h}| > 0 \end{cases} \tag{1.3.9}$$

[14] As we will see later on, the precision of mapping is measure of how well did we fit the point values. The true quality of map can only be accessed by using validation points, preferably independent from the point dataset used to make predictions.

The covariance at zero distance ($C(0)$) is by definition equal to the mean residual error (Cressie, 1993) — $C(\mathbf{h}_{11})$ also written as $C(\mathbf{s}_1, \mathbf{s}_1)$, and which is equal to $C(0) = C_0 + C_1 = Var\{z(s)\}$.

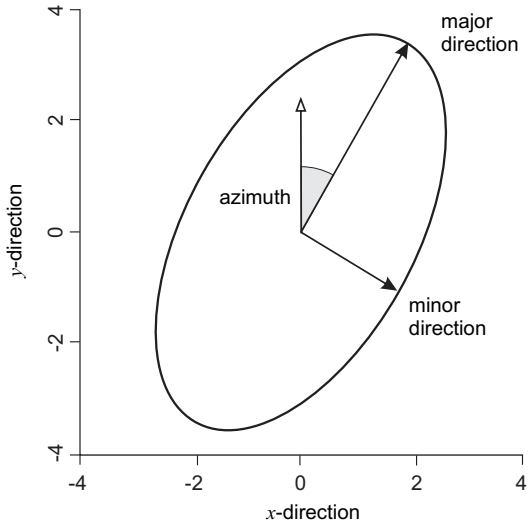

Fig. 1.9: Range ellipse for anisotropic model. After `gstat` User's manual.

The variogram models can be extended to even larger number of parameters if either (a) **anisotropy** or (b) smoothness are considered in addition to modelling of nugget and sill variation. The 2D geometric anisotropy in `gstat`, for example, is modelled by replacing the range parameter with three parameters — range in the major direction (direction of the strongest correlation), angle of the principal direction and the anisotropy ratio, e.g. (Fig. 1.9):

```
vgm(nugget=1, model="Sph", sill=10, range=2, anis=c(30,0.5))
```

where value of the angle of major direction is `30` (azimuthal direction measured in degrees clockwise), and value of the anisotropy ratio is `0.5` (range in minor direction is two times shorter).

Another sophistication of the standard 3-parameter variograms is the Matérn variogram model, which has an additional parameter to describe the smoothness (Stein, 1999; Minasny and McBratney, 2005):

$$\gamma(\mathbf{h}) = C_0 \cdot \delta(\mathbf{h}) + C_1 \cdot \left[\frac{1}{2^{v-1} \cdot \Gamma(v)} \cdot \left(\frac{\mathbf{h}}{R}\right)^v \cdot K_v \cdot \left(\frac{\mathbf{h}}{R}\right)\right] \tag{1.3.10}$$

where $\delta(\mathbf{h})$ is the Kronecker delta, K_v is the modified Bessel function, Γ is the gamma function and v is the smoothness parameter. The advantage of this model is that it can be used universally to model both short and long distance variation. In reality, variogram models with more parameters are more difficult to fit automatically because the iterative algorithms might get stuck in local minima (Minasny and McBratney, 2005). To avoid such problems, we will rely in §4 on more simple variogram models such as the Exponential model.

The fastest intuitive way to understand the principles of kriging is to use an educational program called **EZ-Kriging**, kindly provided by Dennis J.J. Walvoort from the Alterra Green World Research. The GUI of EZ-Kriging consists of three panels: (1)

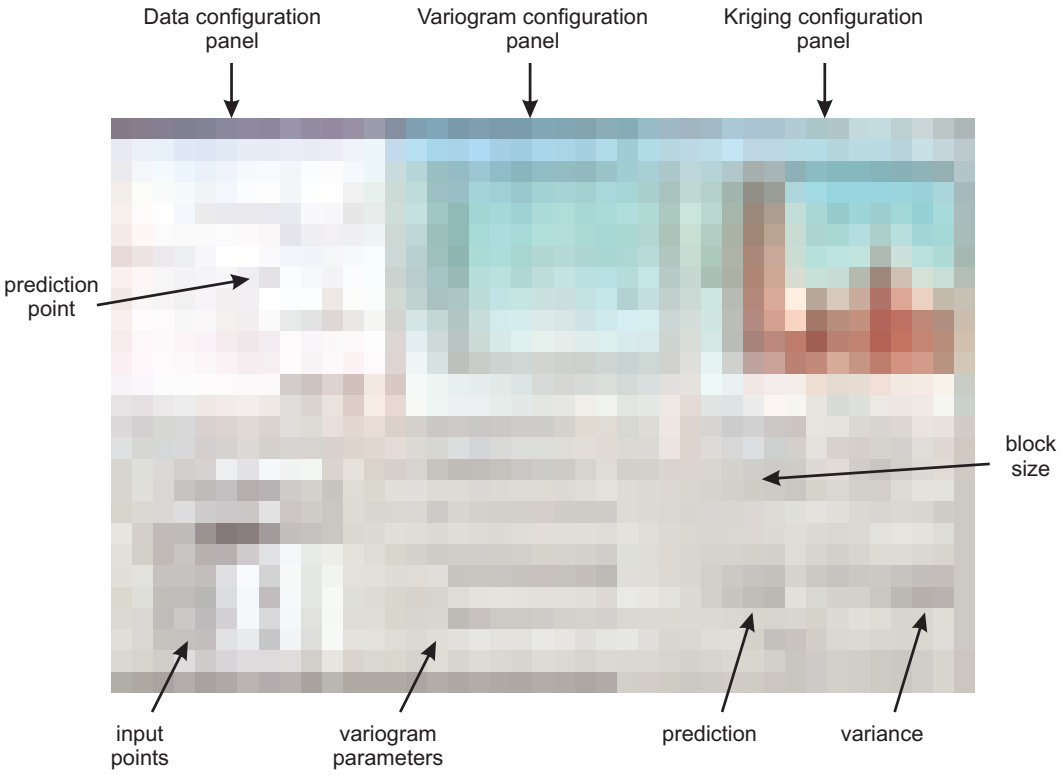

Fig. 1.10: Ordinary kriging explained: E*Z*-Kriging. Courtesy of Dennis J.J. Walvoort.

data configuration panel, (2) variogram panel, and (3) kriging panel (Fig. 1.10). This allows you to zoom into ordinary kriging and explore its main characterises and behaviour: how do weights change for different variogram models, how data values affect the weights, how does block size affects the kriging results etc. For example, if you study how model shape, nugget, sill and range affect the kriging results, you will notice that, assuming some standard variogram model (zero nugget, sill at global variance and practical range at 10% of the largest distance), the weights will decrease exponentially[15]. This is an important characteristic of kriging because it allows us to limit the search window to speed up the calculation and put more emphasise on fitting the semivariances at shorter distances. Note also that, although it commonly leads to smoothing of the values, kriging is an exact and non-convex interpolator. It is exact in the sense that the kriging estimates are equal to input values at sampling locations, and it is non-convex because its predictions can be outside the data range, e.g. we can produce negative concentrations.

Another important aspect of using kriging is the issue of the support size. In geostatistics, one can control the support size of the outputs by averaging multiple (randomized) point predictions over regular blocks of land. This is known as **block prediction** (Heuvelink and Pebesma, 1999). A problem is that we can sample elevation at point locations, and then interpolate them for blocks of e.g. 10×10 m, but we could also take composite samples and interpolate them at point locations. This often confuses GIS users because as well as using point measurements to interpolate values at regular point

[15]In practice, often >95% of weights will be explained by the nearest 30–50 points. Only if the variogram is close to the pure nugget model, the more distant points will receive more importance, but then the technique will produce poor predictions anyhow.

locations (e.g. by point kriging), and then display them using a raster map (see Fig. 1.6), we can also make spatial predictions for blocks of land (block kriging) and display them using the same raster model (Bishop and McBratney, 2001). For simplicity, in the case of block-kriging, one should always use the cell size that corresponds to the support size.

1.3.2 Environmental correlation

If some exhaustively-sampled auxiliary variables or **covariates** are available in the area of interest and if they are significantly correlated with our target variable (spatial cross-correlation), and assuming that the point-values are not spatially auto-correlated, predictions can be obtained by focusing only on the deterministic part of variation:

$$Z(\mathbf{s}) = f\left\{q_k(\mathbf{s})\right\} + \varepsilon \tag{1.3.11}$$

where q_k are the auxiliary predictors that can be used to explain the deterministic part of spatial variation. This approach to spatial prediction has a strong physical interpretation. Consider Rowe and Barnes (1994) observation that earth surface energy-moisture regimes at all scales/sizes are the dynamic driving variables of functional ecosystems at all scales/sizes. The concept of vegetation/soil-environment relationships has frequently been presented in terms of an equation with six key **environmental factors** as:

$$V \times S[x,y,\tilde{t}] = f\left\{\begin{array}{l} s[x,y,\tilde{t}]\ c[x,y,\tilde{t}]\ o[x,y,\tilde{t}] \\ r[x,y,\tilde{t}]\ p[x,y,\tilde{t}]\ a[x,y,\tilde{t}] \end{array}\right. \tag{1.3.12}$$

where V stands for vegetation, S for soil, c stands for climate, o for organisms (including humans), r is relief, p is parent material or geology, a is age of the system, x, y are the coordinates and t is time dimension. This means that the predictors which are available over entire areas of interest can be used to predict the value of an environmental variable at unvisited locations — first by modelling the relationship between the target and auxiliary environmental predictors at sample locations, and then by applying it to unvisited locations using the known value of the auxiliary variables at those locations. Common auxiliary environmental predictors used to map environmental variables are land surface parameters, remote sensing images, and geological, soil and land-use maps (McKenzie and Ryan, 1999). Because many auxiliary predictors (see further Table 3.2) are now also available at low or no cost, it makes this approach to spatial prediction ever more important (Hengl et al., 2007b).

Functional relations between environmental variables and factors are in general unknown and the correlation coefficients can differ for different study areas, different seasons and different scales. However, in many cases, relations with the environmental predictors often reflect causal linkage: deeper and more developed soils occur at places of higher potential accumulation and lower slope; different type of forests can be found at different expositions and elevations; soils with more organic matter can be found where the climate is cooler and wetter etc. This makes this technique especially suitable for natural resource inventory teams because it allows them to validate their empirical knowledge about the variation of the target features in the area of interest.

There are (at least) four groups of statistical models that have been used to make spatial predictions with the help of environmental factors (Chambers and Hastie, 1992; McBratney et al., 2003; Bishop and Minasny, 2005):

Classification-based models — Classification models are primarily developed and used when we are dealing with discrete target variables (e.g. land cover or soil