Chapter 2
Heterogeneity in Fields: Basics of Analyses

Hermann J. Heege

Abstract Sustainable and economical farming needs precise adaptation to the varying soil- and plant properties within fields. Consequently, farming operations have to be adjusted to this in a site-specific way.

An important question is, on which spatial resolution or cell size within a field these adjustments should be based. It is reasonable to expect that this depends on the spatial variations of the respective soil- or crop properties. Consequently, it is shown how the cell sizes needed can be derived from semivariances and its complement functions, the covariances.

Once thus suitable cell sizes are known, they should not be exceeded on any site-specific stage, whether this is sampling, mapping or the operations of the farm machinery.

Keywords Cell-sizes • Kriging • Resolution • Semivariance • Semivariogram

2.1 Variation and Resolution

A traveller attentive to agricultural land will notice that uniform fields are not the rule (Fig. 2.1). This becomes especially obvious to farming experts, who look more closely at soils and crops.

Many soil- and crop properties can vary within fields, such as *e.g.*

- texture (content of sand, silt, loam or clay) and pH of topsoil and subsoil
- soil content of organic matter, of water and of various minerals
- slope and orbital orientation of the soil

H.J. Heege (✉)
Department of Agricultural Systems Engineering, University of Kiel,
24098 Kiel, Germany
e-mail: hheege@ilv.uni-kiel.de

Fig. 2.1 Aerial view of a farming area in Schleswig-Holstein, Germany

- density and morphology of crops
- crop content of water and of various minerals
- infestation of crops by different weeds and by various pests.

Nowadays, many of these soil- or crop properties can be detected and recorded within a field in a site-specific way via modern sensing techniques. Yet before these techniques are dealt with in detail in subsequent chapters, the question arises, how in principle variations of these soil- or crop properties can show up.

These variations of soil- or crop properties within a field can occur in different ways. They can show up in a complete **random pattern**, *i.e.* in a similar manner as raindrop spots within a field. All locations within the field are affected by the rain in a similar way.

Yet the variations can also show up in a **nested pattern**. This is the case, if the respective property, *e.g.* the clay content of the soil, is not uniform on the whole field, but instead there are parts in various directions where it is lower and *vice versa* higher. The respective property in this case varies with the distance.

Thus the spatial variation can be uncorrelated or correlated (Fig. 2.2), depending how it presents itself in a graph with a distance scale. However, the distance scale too can change. And if it does, the appearance of variations can be different. What looks as random arrangement or noise at one scale can be recognized as structure at another scale. This is why looking through a microscope can be so fascinating.

Therefore, we have to deal with **resolutions**. What is meant is not a resolution taken by a political assembly. Here the term resolution stands for the "resolving" or the dividing up of physical properties involved, such as the area of the field, the time or the measurement units that belong to the signals of a sensor (Fig. 2.3).

The **temporal resolution** that is required depends very much on the respective soil- or crop property. Textures and organic matter contents of soils hardly change
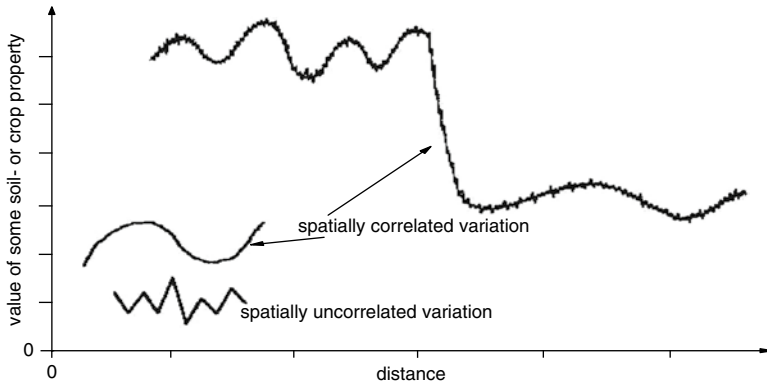
Fig. 2.2 Types of spatial variation in a dimensionless diagram (From Oliver 1999, altered)
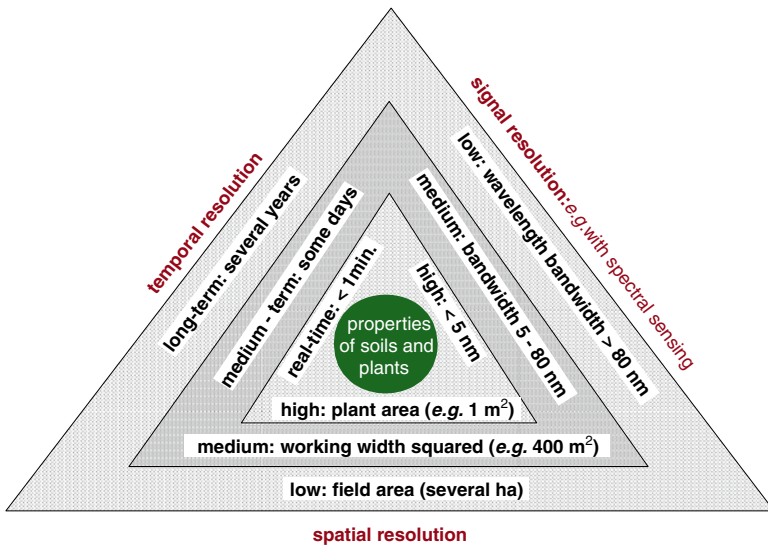

Fig. 2.3 Low-, medium- and high resolutions on a spatial-, a temporal- and a signal basis

over time. Therefore, these properties can be recorded once on a long-term basis in field maps that can be used for several years. The situation is quite different when it comes to the water-, the nitrogen- and the pesticide supply of crops in the growing season. In these cases, the best temporal resolution would be obtained with a control system that adjusts the supply in real-time, which means immediately after sensing and "on-the-go" during the application.

The **signal resolution** refers to the physical quantities that are sensed. In case of spectral sensing, the bandwidth-ranges of the light waves in the visible- or infrared region are important and can be very different (Fig. 2.3).

Finally, precision farming aims at a high **spatial resolution**. It should be noted that instead of the term "resolution", the denotation "**cell size**" often is used. This stands for the area, for which the respective farming operations are uniformly adjusted. Therefore, a low resolution means a large cell size and *vice versa*. The traditional and still rather common cell size is the individual field. Whilst the sizes and basic operations of present farm machinery are maintained, about the smallest cell size that can be realized would be an area that corresponds to the square of the working width. This approach is derived from the assumption that the basic shape of a cell would be a square. So if a working width of 20 m for fertilizing and spraying is used, a cell size of 400 m² would result. And if fields that are controlled by small robots become a reality, a high regional resolution based on treating individual plants or even leaves might become feasible.

Yet these considerations emanate from **technical possibilities** with the respective farm machinery, provided the control components are available. A better approach is to base the resolution on the respective soils and crops and to adapt the technical solutions as well as possible to these.

If theoretically the soil- and crop properties would be completely uniform within a field, no site-specific treatments would be necessary. And if on the other hand significant variations would show up within short distances, small cell sizes would be reasonable. This leads to the question, how – based on variations existing within fields – proper **cell sizes** can be deduced. Statistical indices of soil- or crop properties like averages or standard variations are no help in this respect. This is because intrinsically these indices are independent of location. What is needed are statistical indices that rely on distances within a field. The semivariance and its graph – the semivariogram – do this.

2.2   Semivariance and Semivariogram

The geostatistical concept behind semivariances and semivariograms is Matheron's (1963) regionalized variable theory. It states that the differences in the values of a spatial variable – such as a soil- or crop property – between points in a field depend on the **distance** between these points. In short, the smaller the distances, the smaller the differences.

As a logical consequence, the semivariance v expresses the dissimilarity of paired property values as a function of the distances between two sampling points. The general equation of the semivariance v is:

$$v = \frac{1}{2N} \sum_{1}^{N} \left[f\left(x+h\right) - f(x) \right]^2$$

Here, x and $x+h$ stand for the vectors of areal coordinates at two locations in the field. These locations are separated by the distance h. The functions $f(x)$ and $f(x+h)$ together represent thus a pair of soil- or crop properties at these places. N is the number of location pairs that are involved.
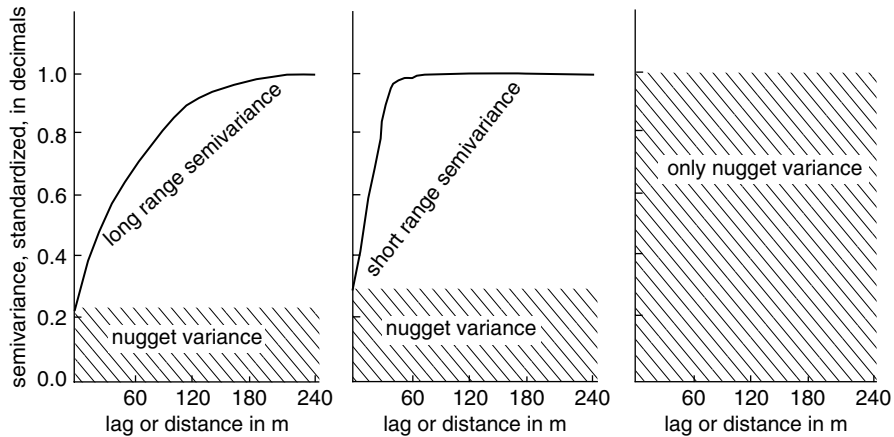
Fig. 2.4 Semivariances, its semivariograms and pure nugget variance (From suggestions by Oliver 1999, altered and redrawn)

The equation resembles the general formula for the standard variance. However, the basic data for the standard variances are not pairs that are related by distances and orientation; they just come from a common population and not more. And because of the **pairs** that are used when calculating the semivariance, the summed result is halved. This explains the denotations semivariance for the numerical values and semivariogram for the graphical description.

Most semivariance curves – the semivariograms – are bounded, which means that they reach an asymptote (Fig. 2.4). This asymptote is called the **sill variance**. But unbounded variogram curves also can occur. Note that the semivariograms are standardized to a sill variance of 1.

The distance at which the sill variance is approached is called the **range** (Fig. 2.5). Since this is an asymptotic approach, the range is arbitrarily set to be the distance needed to get 95 % of the sill variance (Tollner et al. 2002). All points that are separated by distances smaller than the range are spatially correlated. Whenever the distances are larger than the range, the points are spatially independent. This is very important: it means that any site-specific operation that is based on squared cells with sides longer than the range is useless. This is, because with cells of this size it is not possible any more to detect or catch regional differences of the respective property. So the control distance that is used for site-specific operations must be smaller than the range of the semivariogram.

Theoretically, the semivariogram curves start with zero variance and zero distance. But in reality, this seldom occurs. In practice there always is some variance already at zero distance. It is called **nugget variance** and represents variability at distances smaller than the typical sample spacing as well as measurement errors. In rare occasions, there is only nugget variance (Fig 2.4, right), which means that any site-specific treatment – at least based on the respective property – is useless.

2.3   Cell Sizes

Knowing that the cell squares should have side lengths between zero and the range alone is not sufficient. Very small cells result in high costs, and rather large cells can impair the precision. And since the number of cells that must be dealt with qua-druples with each halving of cell-side length, more precise information about the size that should be used is desirable. This information can be derived from the semi-variance (Fig. 2.5).

An interesting approach for this was developed by Russo and Bresler (1981) as well as by Han et al. (1994). It is based on the notion that as the semivariance is an indicator of **dissimilarity** of a site-specific soil- or crop property, *vice versa* the complement function to the semivariance provides information of **similarity** or **relatednes**s. For normalized situations, the semivariance plus its complement function for all respective distances or lags add up to one (1). Therefore, the complement function is the vertical mirror image of the semivariance (Fig. 2.5). It can be shown that for the pairs involved, this complement function of the semivariance is a well known statistical function – the covariance (Davis 1973; Gringarten and Deutsch 2001). In contrast to this: the sill of the semivariogram is the standard variance, which in this case stands for zero correla-tion. Therefore, the semivariance is standard variance minus covariance.

The area under the curve of the complement function to the semivariance can be regarded as an accumulation of all relatedness or similarity of the respective
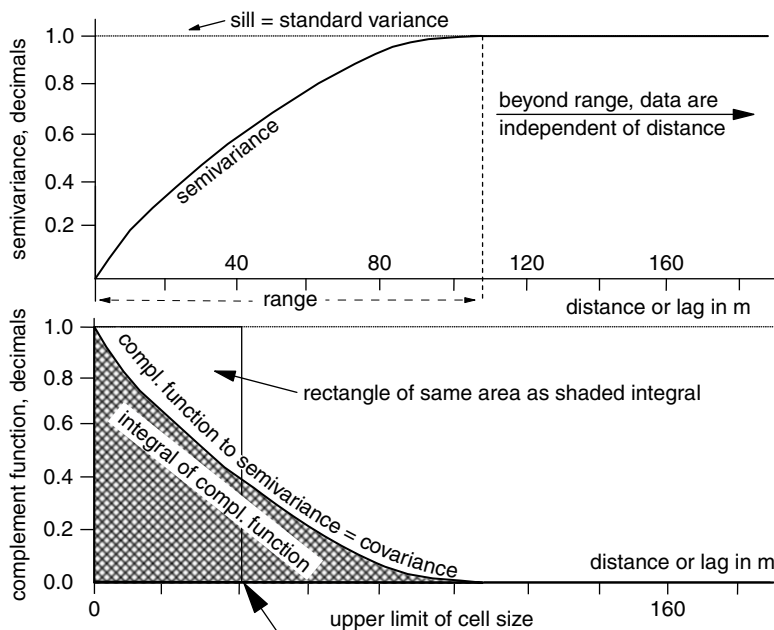

Fig. 2.5 Semivariance, its complement function and the upper limit of cell size

property. It can be computed by integrating the equation of the complement function between the range and zero. This integral of the complement function is represented by the hatched area in Fig. 2.5.

If differences based on location exist, theoretically only cells of zero size would be completely uniform. For practical purposes, a compromise in a cell size with side-length between zero and range is necessary. The proposal of Russo and Bresler (1981) as well as of Han et al. (1994) for this compromise is based on a rectangle with a height of the sill and an area that equals the integral of the complement function. This rectangle contains all the similarity or relatedness that can be obtained from the respective semivariance from which it was derived. For this reason, the length of the rectangle side along the abscissa can be regarded as an indication of the **upper limit of the cell-size** (Fig. 2.5). Any larger cell-size overlaps into areas of pure dissimilarity. It thus deteriorates the precision in site-specific management. This upper limit of cell-size represents the largest distance for which the soil or crop property is well correlated with itself.

It is also obvious that the upper limit of the cell-size depends largely on the distance of the range. Kerry and Oliver (2004, 2008) propose to use a sampling interval – or upper limit of cell size – of less than half the range. It can be seen that less than **half the range** results in approximately the same limit of cell size.

However, these determinations of cell sizes are possible only after the semivariance has been recorded. The problem is that cell-sizes for **sampling** of soil- or crop properties must be known beforehand in order to arrive at reliable resolutions for subsequent farming operations. If the sampling is based on too large cell-sizes, no detailed computation afterwards any more can result in a reliable control for site-specific farming. The resolution that is needed depends on the respective variability of the soil or crop property. It must be met at the first site-specific operation, otherwise subsequent procedures cannot be controlled with precision. And this first operation is the sampling.

Ways out of this situation are either sampling with a very high resolution from the outset so that any site-specific requirements definitely are met or alternatively the use of standardized, default cell-sizes for sampling. The first method lends itself whenever the data are recorded automatically online and on-the-go, since many modern sensors can provide for several signals per m of travel and thus for a high resolution.

The use of standardized, default cell-sizes for sampling is advisable if manual sampling and processing is needed as with *e.g.* the conventional and traditional collection of data about soil texture, -nutrients or crop properties. Because of the high amount of labour involved, in these cases sampling with a very high resolution from the outset as with online and on-the-go methods cannot be practised. Therefore, the information about the sampling cell-sizes that are needed must be obtained from previously made semivariograms, which are based on data of soils or crops under conditions that are similar. Information about such standardized semivariograms has been published (Kerry and Oliver 2004, 2008; McBratney and Pringle 1999). These **standardized semivariograms** are based on sampling with a high resolution. They thus provide the default cell-sizes for subsequent site-specific sampling.

It should be noted that reliable semivariograms cannot be obtained from fewer than 100 data. The nugget to sill ratios of the standardized default semivariograms should match those of the respective fields (Kerry and Oliver 2008).

The term "standardized semivariogram" should not conceal the fact that in reality the curves for the semivariances are rather unique and different for every field. Yet precise sampling and site-specific management relies on the semivariograms for information about the cell-size that should be used. If the cell-size is oriented at less than half the range (Kerry and Oliver 2004, 2008) as outlined above, what ranges for soil- and crop properties do actually occur?

Actually, the **ranges** that have shown up in research with semivariograms vary immensely – as do the local conditions around the globe. In extreme cases, ranges on the one hand as short as 1 m (Solie et al. 1992) and on the other hand as long as 26 km (Cemek et al. 2007) have been recorded. However, the vast majority of ranges for soil- and crop properties is between 20 and 110 m (McBratney and Pringle 1999). So for most cell-sizes, the upper limits of side lengths should be between 10 and 55 m. This is still a rather wide span. Therefore, to define the actual need more closely, a careful deduction of required cell-sizes from suitable standardized semivariograms is necessary.

2.4   Processing and Adjusting the Resolution

Finding the appropriate upper limit of cell-sizes deserves some effort. This is because knowing about it is important on several stages of site-specific farming, first when sampling, then for mapping and finally when the machinery operates in the field. It is obvious that the sampling must occur within the distance limits defined by about half the range. But the same holds for techniques used to make the maps and finally for the cell-sizes, on which the farm machinery works in a site-specific way. If on any of these stages the distance limits that are defined by less than half the range of the respective semivariogram are exceeded, the precision of site-specific farming is impaired.

The question is, at which stage – sampling, mapping or machine operations – striving for small cell-sizes is most difficult. As long as sampling of soil properties and of nutrient contents is done in a manual way, this will be the sampling stage. In the long run, however, manual sampling will be more and more replaced by online and on-the-go sensing methods. Many of these methods will allow for sensing of small cell-sizes. As a consequence, then the cell-sizes that can be realized with wide farm machinery become important.

It is not recommended to directly combine fine grid spacings for sampling or sensing on the one hand with a much coarser resolution for the machinery operations on the other hand without any signal corrections. This is because the control of the machinery is less erratic and is more stable if **averages** of highly resolved signals are used. If the machinery is controlled via online and on-the-go sensing, this averaging step can easily be implemented into the processing computer program. In case a field
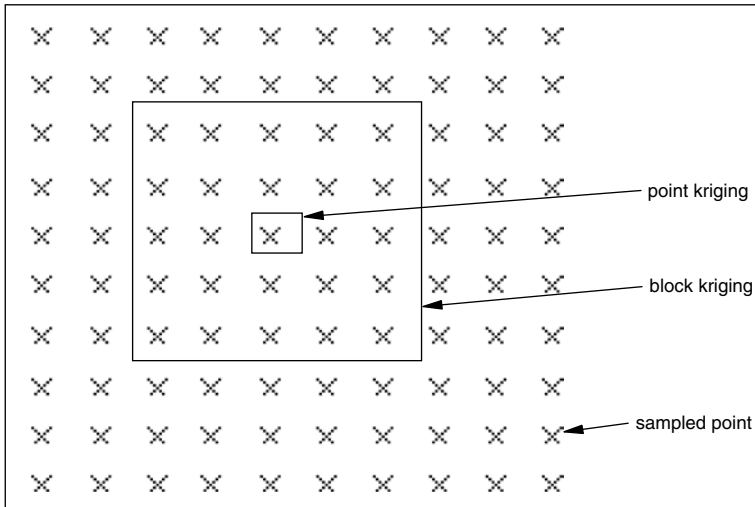
Fig. 2.6 Principles of block-kriging and point-kriging (From Whelan et al. 2002, altered)

map for later use is made from the sampled or sensed data, the averaging can occur during mapping.

The sampling or sensing of the data usually occurs on a **punctual basis**. Hence for mapping of a continuous surface, some interpolation is needed. This **interpolation** of the punctual signals is called **kriging** in honour of D.G. Krige, a South African geologist, who was a pioneer in this processing of data for mapping. Today, many kriging methods are available. These methods are based on the assumption that geological properties in close proximity to each other are more likely to be similar than those separated by longer distances. So the basic premise – the concept of spatial dependence – is the same as for semivariances (Sect. 2.2).

Kriging can be subdivided into either block- or point methods, depending on the size of the respective map area, for which the property is estimated by interpolation. In block kriging, the average value of the property for a block within the map is calculated, whereas in point kriging the target is just a point within the area. This means that point kriging can be considered as the limiting case of block kriging when the block size approaches the sampling- or sensing size (Fig. 2.6). For details to interpolation methods and to computer programs for this see Hengl (2007), Webster and Oliver (2007) and Whelan et al. (2002).

An important feature of block-kriging is that its estimate – the block mean – may gain in reliability as the block area increases. This follows from basic features of means. This averaging advantage of **block-kriging** is effective as long as the size of the blocks does not exceed the cell-size that is effective in the machinery operations. But when the sizes of the kriged blocks get larger than the cells of the machines, this averaging gets detrimental. This is because then another averaging feature – its leveling effect – has consequences. This leveling effect deteriorates the objectives of
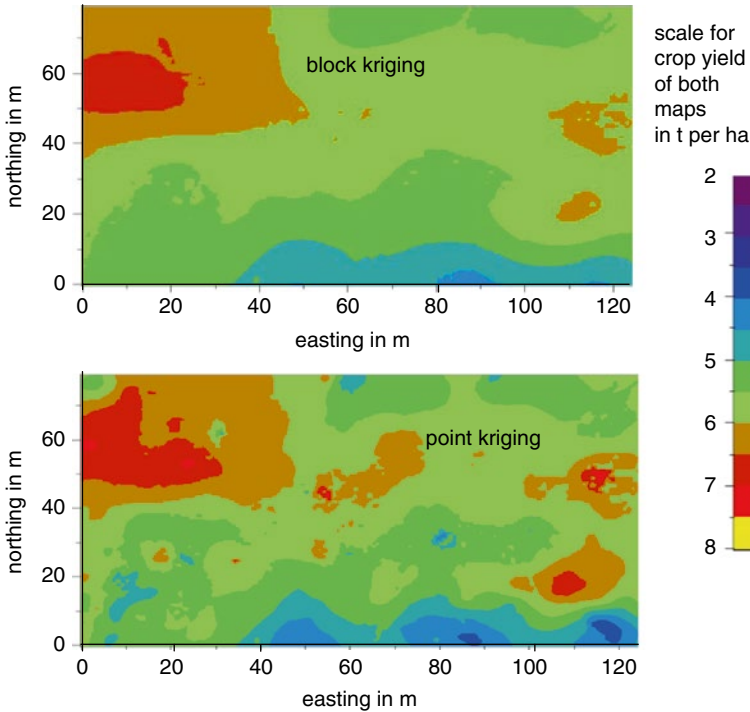
Fig. 2.7 Maps of wheat yields from site-specific harvesting after block- and point-kriging. The smoothing effect of block-kriging is obvious (From Whelan et al. 2002, rearranged and altered)

site-specific farming, since it artificially erases local differences and thus eliminates possibilities to react on them. Yet as long as the kriged blocks are smaller than the cells of the machinery, any leveling effects of averaging are unavoidable – just because the size of the machinery does this anyway.

With **point-kriging**, there is no processing error at the actual sampling- or sensing point. However, this theoretical advantage of point-kriging can only improve the overall precision, if the farm machinery too can do punctual work. As long as this is not the case, block kriging will primarily be the choice for creating maps.

The objective is that maps should only show the spatial distribution of the respective soil- or crop property. However, maps always too reflect the influence of the sampling- or sensing techniques and in addition of the kriging method that was used (Fig. 2.7). Maps from the same basic data therefore can look very different.

The averaging and estimating that is connected with kriging should be fine-tuning of the spatial resolution as it is needed for the respective site-specific operation. A fine-tuning by averaging that is oriented at the cell-size of the machine operations should be the objective. This adjusting of the resolution can take place either during online and on-the-go control or when maps are processed.

Maps can act as precious **time-bridges** between sampling or sensing on the one hand and machine operations on the other hand. These time-bridges are essential or useful with soil- or crop properties that have a low temporal resolution, thus remain constant over a rather long time. Prime examples for this are topography, texture, organic matter and pH of the soil.

However, there are also soil- or crop properties that vary rather fast over time, thus have a high temporal resolution. Examples for this are water and nitrate in the soil or some crop properties. The temporal variance in these cases may in fact be more important than the spatial variance (McBratney and Whelan 1999), and consequently maps should then be viewed and used with caution.

References

Cemek B, Güler M, Kilic K, Demir Y, Arslan H (2007) Assessment of spatial variability in some soil properties as related to soil salinity and alkalinity in Bafra plain in Northern Turkey. Environ Monit Assess 124:223–234

Davis JC (1973) Statistics and data analysis in geology. Wiley, New York

Gringarten E, Deutsch CV (2001) Teacher's aide. Variogram interpretation and modelling. Math Geol 33(4):507–534

Han S, Hummel JW, Goering CE, Cahn MD (1994) Cell size selection for site-specific crop management. Trans Am Soc Agric Eng 37(1):19–26

Hengl T (2007) A practical guide to geostatistical mapping of environmental variables. JRC scientific and technical reports. European Commission. JRC Ispra, Ispra. http://ies.jrc.ec.europa.eu

Kerry R, Oliver MA (2004) Average variograms to guide soil sampling. Int J Appl Earth Obs Geoinform 5:307–325

Kerry R, Oliver MA (2008) Determining nugget: sill ratios of standardized variograms from aerial photographs to krige sparse soil data. Precis Agric 9:33–56

Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246–1266

McBratney AB, Pringle MJ (1999) Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. Precis Agric 1:125–152

McBratney AB, Whelan BM (1999) The "null hypothesis" of precision agriculture. In: Stafford JV (ed) Precision agriculture '99. Part 2. Sheffield Academic Press, Sheffield, pp 947–957

Oliver MA (1999) Exploring soil spatial variation geostatistically. In: Stafford JV (ed) Precision agriculture '99. Part 1. Sheffield Academic Press, Sheffield, pp 3–17

Russo D, Bresler E (1981) Soil hydraulic properties as stochastic processes: I. An analysis of field spatial variability. Soil Sci Soc Am J 45:682–687

Solie JB, Raun WR, Whitney RW, Stone ML, Ringer JD (1992) Optical sensor based field element size and sensing strategy for nitrogen application. Trans Am Soc Agric Eng 39(6):1983–1992

Tollner EW, Schafer RL, Hamrita TK (2002) Sensors and controllers for primary drivers and soil engaging implements. In: Upadhyaya SK et al (eds) Advances in soil dynamics, vol 2. ASAE, St. Joseph, p 182

Webster R, Oliver MA (2007) Geostatistics for environmental scientists, 2nd edn. Wiley, Chichester

Whelan BM, McBratney AB, Minasny B (2002) VESPER 1.5 – Spatial prediction software for precision agriculture. In: Robert PC, Rust RH, Larson WE (eds) Proceedings of the 6th international conference on precision agriculture, Madison