Kansas State University Libraries

New Prairie Press

SPATIAL VARIABILITY IN AGGREGATION BASED ON GEOSTATISTICAL ANALYSIS

Shoufan Fang

George Z. Gertner

Guangxing Wang

Alan B. Anderson

See next page for additional authors

Recommended Citation

Author Information

Shoufan Fang, George Z. Gertner, Guangxing Wang, and Alan B. Anderson

SPATIAL VARIABILITY IN AGGREGATION BASED ON GEOSTATISTICAL ANALYSIS

Shoufan Fang, George Z. Gertner, Guangxing Wang

Department of Natural Resources and Environmental Sciences
University of Illinois at Urbana-Champaign
W503 Turner Hall, 1102 S. Goodwin Avenue, Urbana, IL 61801, USA.

and

Alan B. Anderson

US Army Corps of Engineer, Construction Engineering Research Laboratory
9000 Research Park, Champaign, IL 61824, USA.

Abstract

This study derived the equations for computing the spatial variability in the aggregation of original maps of continuous attributes. The derivation of the equations is based on traditional statistical and geostatistical principles. The derived equations can be used to compute the variance, covariance, and spatial (auto-/cross-) covariance of the aggregated pixels and sub-areas in a given study area. Using the derived equations, the total uncertainty within a study area will not change after aggregation. For a case study, it has been shown that aggregation will reduce the values of variance/covariance and spatial covariance of the aggregated individual pixels. It was also verified that the original semivariogram models should not be used for the aggregated maps to compute spatial covariances. It is suggested to use the original scales in geostatistical analyses to produce maps and then produce courser scaled maps through aggregation.

1. Introduction

Aggregation of spatial data, that is, inferring spatial information from a finer spatial scale to a courser scale, is routinely performed in Geographic Information Systems (GIS), cartographic mapping, and image processing. There are a number of reasons for this. They include the fact that the spatial feature of an attribute or one kind of natural resource can be accurately mapped and displayed at a specific scale or spatial resolution, and thus multiple scales exist for a complex natural scene and ecologic system (Hay et al., 1997; King, 1991; Myers, 1997; Wang et al., 2001b). On the other hand, the maps at different spatial resolutions need to be overlapped to derive a product that is useful to make management plans for natural resources and ecologic systems.

Many scale-related studies have been conducted (Hay et al., 1997; King, 1991; O'Neill, 1989; Myers, 1997; Townshend and Justice, 1988; Turner et al., 1989; Turner et al., 1991; Woodcock and Strahler, 1987). These studies mainly deal with how to determine appropriate spatial resolutions and how to infer spatial data from one scale to another. For example, a widely used method for data aggregation is window averaging. In addition to estimated maps, the data aggregation from a finer to a coarser resolution will result in changes of spatial variation and spatial correlation within/among pixels/sub-areas, and corresponding uncertainties of estimates. Furthermore, these changes will affect decision-making for managing natural resources and ecological systems. However, this has only been suggested in the literature and the topic has been studied by only a few authors (Gertner et al., 2002; Wang et al., 2001a; Zhang et al., 1990).

In spatial statistics (geostatistics), spatial variability of a study area is measured using semivariograms. Theoretically, a regular or point semivariogram of an attribute should exist regardless of spatial resolution. In fact, it is usually estimated based on field observations. The semivariogram thus reflects the average spatial variability in the study area, and it varies depending on plot size of the ground observations or pixel size of images or maps (Wang et al., 2001a; Zhang et al., 1990). On the other hand, a map is often derived using either a traditional regression and classification method based on sample variation or a geostatistical method based on a sample semivariogram. The estimates used to generate the maps have uncertainty. When spatial data of a map are aggregated, the uncertainties of pixel values at the finer spatial resolution will be propagated to the estimates at the coarser resolution. Up to now, however, we have not found any discussions on the derivation of uncertainties of the estimates for the aggregation of spatial data.

This study applies traditional statistical and geostatistical principles to derive the uncertainties of map estimates when spatial data are aggregated. A case study was conducted to explore the changes of spatial variability corresponding to scales used in the aggregation.

2. Spatial Variability

At a particular point in a study area, the values of attributes of interest can be obtained from ground measurements, by remotely sensed information, or estimated using spatial statistical methods. For continuous attributes, the obtained values are usually not true values, since there are always errors due to sampling, measurement, processing, etc. In remote sensing, geostatistics, and GIS, the concept of "point" is different from its mathematical definition. In the three map-oriented disciplines, each "point" in the study area occupies a certain area, which is called a pixel, cell, or support unit. The values of the attributes of interest at a pixel are often the average values over the area of the pixel. Therefore, at each point, i.e., pixel, the attributes of interest are random variables and they have their own distributions. In the whole study area, the attributes have a spatial distribution, which is produced by the intrinsic variation of the attributes. Spatial variability is usually modeled in spatial statistical analyses. Isotropic covariance of one attribute among pixels is usually denoted as auto-covariance:

$$\text{cov}[z(u_i), z(u_i + h)] = \frac{1}{m} \sum_{i=1}^{m} \left\{ z(u_i) - E[z(u_i)] \right\} \cdot \left\{ z(u_i + h) - E[z(u_i + h)] \right\}$$

where z is the attribute of interest, u_i is a pixel location, h is the distance between a pair of pixels, $z(u_i)$ is the value of z at u_i, m is the number of pairs of pixels, and E[z] is the

Applied Statistics in Agriculture

expectation of z, which can be estimated by its sample mean. $\text{Cov}[z(u_i), z(u_i+h)]$ is expressed as $C(0)$ and $C(h)$ when h is zero and non-zero, respectively. Another statistic to indicate spatial variability is the semivariogram:

$$\gamma(h) = \frac{1}{2m} \sum_{i=1}^{m} \left\{ z(u_i) - z(u_i + h) \right\}^2$$

A semivariogram is usually used to model the relationship between the spatial correlation of attributes and the distance of the locations of attributes. Ignoring the lag effect, the relationship between spatial covariance and the semivariogram is (Chilès and Delfiner, 1999; Clark and Harper, 2001; Goovaerts, 1997; Isaaks and Srivastava, 1989; Journel and Huijbregts, 1978):

$$\gamma(h) = C(0) - C(h)$$

or

$$C(h) = C(0) - \gamma(h) \tag{1}$$

This relationship is used in estimating the value of the attribute of interest at the non-sampled pixels (regionalization).

When the attributes of interest are multivariate, spatial variability can be represented as auto- and cross-covariance:

$$\text{cov}[z_p(u_i), z_q(u_i + h)] = \frac{1}{m} \sum_{i=1}^{m} \left\{ z_p(u_i) - E[z_p(u_i)] \right\} \cdot \left\{ z_q(u_i + h) - E[z_q(u_i + h)] \right\}$$

where p and q are subscripts to indicate the individual attributes that are multivariate. When p is not equal to q, and h is not equal to zero, then $\text{cov}[z_p(u_i), z_q(u_i+h)]$ is the cross-covariance.

3. Changes of Spatial Variability in Aggregation

When the original pixels are aggregated into a larger unit, the spatial variability of the resulting courser map will be different than the original map.

Assume that the attribute z of interest at an original pixel u_i has mean $E[z(u_i)]$ and variance $\text{var}[z(u_i)]$, and a rescaled (aggregated from original pixels) pixel v_k contains n original pixels. The value of attribute z at the rescaled pixel v_k is the average of the z values that included the original pixels:

$$z(v_k) = \frac{1}{n} \sum_{i=1}^{n} z(u_i)$$

Therefore, its expected value is:

$$E[z(v_k)] = \frac{1}{n} \sum_{i=1}^{n} E[z(u_i)],$$

and its variance is:

$$\text{var}[z(v_k)] = \frac{1}{n^2} \left\{ \sum_{i=1}^{n} \text{var}[z(u_i)] + \sum_{i=1}^{n} \sum_{j \neq i}^{n} \text{cov}[z(u_i), z(u_j)] \right\} \tag{2}$$

Derivation of Eq. 2 can be found in Eq. A1 (see Appendix A). For multivariate cases,

$Z = (z_1, \cdots, z_s)'$ is a vector consisting of s attributes, its value (a vector) at the rescaled pixel v_k is still the average value (a vector) from the n original pixels:

$$Z(v_k) = \frac{1}{n} \sum_{i=1}^{n} Z(u_i)$$

Correspondingly, its covariance matrix is:

$$var[Z(v_k)] = \frac{1}{n^2} \left\{ \sum_{i=1}^{n} \Sigma_{(u_i,u_i)} + \sum_{i=1}^{n} \sum_{j \neq i}^{n} \Sigma_{(u_i,u_j)} \right\} \tag{3}$$

where

$$\Sigma_{(u_j,u_k)} = \begin{pmatrix} C_{1,1}(u_j,u_k) & C_{1,2}(u_j,u_k) & \cdots & C_{1,s}(u_j,u_k) \\ C_{2,1}(u_j,u_k) & C_{2,2}(u_j,u_k) & \cdots & \vdots \\ \vdots & \vdots & \ddots & C_{s-1,1}(u_j,u_k) \\ C_{s,1}(u_j,u_k) & \cdots & C_{s,s-1}(u_j,u_k) & C_{s,s}(u_j,u_k) \end{pmatrix}$$

and

$$C_{p,q}(u_j,u_k) = cov[z_p(u_j), z_q(u_k)]$$

Eq. A2 in Appendix A provided the derivation of Eq. 3. Eq. 3 is very similar to the variance estimator of block kriging (Chilès and Delfiner, 1999; Clark and Harper, 2001; Goovaerts, 1997; Isaaks and Srivastava, 1989; Journel and Huijbregts, 1978), although their conditions are different.

Spatial (auto-/cross-) covariance of the attributes among the aggregated pixels is also aggregated from the original spatial covariance. The aggregated spatial covariance is the average of the original spatial covariance among all involved original pixels:

$$cov[z_p(v_{k1}), z_q(v_{k2})] = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n} \sum_{j=1}^{n} cov[z_p(u_i), z_q(u_j) | u_i \in v_{k1}, u_j \in v_{k2}] \tag{4}$$

where $cov[z_p(v_{k1}), z_q(v_{k2})]$ is the cross- or auto-covariance (when p equals q) of attributes z_p and z_q between aggregated pixels v_{k1} and v_{k2}, each of which contains n original pixels. Eqs. A3 to A6 in Appendix B illustrates the derivation of Eq. 4. It is also very similar to the approximate of block-to-block variance from block kriging (Goovaerts, 1997; Journel and Huijbregts, 1978).

When aggregation is based on sub-areas which contain different numbers of original pixels, the value in the larger sub-area v_k (aggregated from smaller sub-areas) is:

$$Z(v_k) = \sum_{j=1}^{k_n} \frac{n_j}{n} Z(v_j) \tag{5}$$

where k_n is the number of the smaller sub-areas, a smaller sub-area v_j contains n_j original pixels, and n is the total number of original pixels inside v_k. Therefore, the variance of an

attribute of interest in v_k can be computed using the modified form of Eqs. 2 and 3. The variance of attribute z_p in v_k is:

$$\text{var}[z_p(v_k)] = \sum_{j=1}^{k_n} \frac{n_j^2}{n^2} \text{var}[z_p(v_j)] + 2\sum_{j=1}^{k_n-1}\sum_{j1>j}^{k_n} \frac{n_j n_{j1}}{n^2} \text{cov}[z_p(v_j),z_p(v_{j1})] \qquad (6)$$

The spatial covariance of two attributes z_p and z_q between smaller sub-areas v_j and v_{j1} is defined as:

$$\text{cov}[z_p(v_j),z_q(v_{j1})] = \frac{1}{n_j \cdot n_{j1}} \sum_{i=1}^{n_j}\sum_{i1=1}^{n_{j1}} \text{cov}[z_p(u_i),z_q(u_{i1})|u_i \in v_j, u_{i1} \in v_{j1}] \qquad (7)$$

The details about the derivation of Eqs. 5, 6, and 7 can be found in Appendix C.

When the areas, instead of the number of original pixels, of sub-areas are provided from a map, then in Eqs. 5, 6, and 7, n, n_j, and n_{j1} can be replaced with s, s_j and s_{j1}, which are the areas of v_k, v_j and v_{j1}, respectively.

Eqs. 5, 6, and 7 are much more flexible and useful in GIS than the equations derived for equal-sized sub-areas. In GIS, operations are usually based on irregular and unequal-sized sub-areas. Under such situations, equations for equal-sized sub-areas could not be used to compute uncertainty propagation.

4. Case Study

Sand and silt (i.e., very fine sand) are two of essential soil components. Their proportions play an important role in determining the structure and characteristics of soil. They are often investigated in soil related research. In this study, the proportions of sand (Sand) and silt (Silt) were used as two attributes to reveal the changes of spatial variability corresponding to the changes of map scale in aggregation.

4.1 *Study Area and Data*

The study area is Fort Hood, which is located on the border of Bell and Coryell Counties in central Texas (USA). Its geographical location is between the longitude 97º55'01.1" and 97º30'01.1" W, and latitude 31º25'00.6" and 31º00'00.7" N. The soil is generally shallow to moderately deep, clayey, and underlain by limestone bedrock. The elevation ranges from 180 to 375 meters above sea level with 90 percent below 260 meters. The slopes are less than 33º, and the average slope is 2.66º. The dominant vegetation includes oak-juniper woodlands and savannah (Tazik et al., 1993).

Soil samples were collected from 192 plots over the entire area, and the percentages of sand and silt were analyzed in a laboratory. The descriptive statistics of the collected soil samples are listed in Table 1.

4.2 *Joint Sequential Simulation*

The study area was divided into 440 x 440 original pixels with a pixel size of 100 x 100 m^2. The joint sequential simulation algorithm (Almeida, 1993; Goovaerts, 1997; Zhu and Journel, 1993) (see Appendix D for details) was used for predicting the spatial distribution of Sand and Silt. At each pixel, 500 predictions of Sand and Silt were obtained from joint sequential

simulation. Across the 500 replications, the expected values and covariance matrix of Sand and Silt were estimated at each pixel.

The omni-directional auto semivariogram models measuring spatial correlation of the two attributes were derived based on the field observations using spherical semivariogram models (Table 2). Figure 1 shows the sampled and modeled standardized semivariograms of Sand and Silt. The cross semivariogram of Sand and Silt was computed using Markov-type approximation based on the standardized auto semivariogram model of Silt (see Appendix E for details).

Spatial distributions of the estimated mean of Sand and Silt are shown in Figure 2. The correlation of these two attributes can be seen very clearly in the maps. The locations which had a higher proportion of Sand had a lower proportion of Silt, and *vice versa* (Figure 2).

4.3 *Aggregation*

Besides the original pixel size (100 x 100 m², which was used in semivariogram modeling and spatial simulation), four scales were used in aggregation in order to investigate the changes of spatial variability due to map scale (Table 3). Eq. 3 was used in computing the covariance matrices of the aggregated pixels, which were used to produce the spatial variation maps of Sand and Silt at different scales. For comparison, the variance maps of Sand after aggregation were also generated using the average of the variance of the original pixels.

In order to use a common legend for the (co)variance maps with different resolutions, a conversion factor (CF) for each (co)variance map was defined as:

$$CF_{nk} = \frac{max_{nk} - min_{nk}}{max_{n1} - min_{n1}} , \quad n=1,2,3 \text{ and } k=1,2, \ldots, 5.$$

where the subscripts n and k represents the types of variation (variances of Sand and Silt, and their covariance) and scale codes (in Table 3), respectively; max and min are the maximum and minimum (co)variance, respectively. Therefore, with a smaller CF, the actual values of a map are smaller. The actual value of the (co)variance $[v(u_i)_{nk}]$ at a pixel u_i of an aggregated map can be converted using its minimum and CF and the common legend as following:

$$v(u_i)_{nk} = min_{nk} + CF_{nk} \cdot (v'(u_i)_{nk} - min_{n1}) \tag{8}$$

where CF, n, and k has the same meaning as in Eq. 8; and $v'(u_i)_{nk}$ is the value of that pixel according to the common legend. The minimum (co)variance and CF can be found at the top of each (co)variance map (Figures 3 to 6).

Since auto-/cross-covariance after aggregation can not be illustrated on the (co)variance maps, the spatial covariance had to be computed for making comparisons. In this case study, the influence of aggregation methods and resolutions to spatial covariance of proportions of sand and silt within the distance of two pixels was illustrated with graphs. Two methods were used in computing spatial covariance. One is the traditional method (see Eq. III.3, p.45 in Deutsch and Journel, 1998). The equation is:

$$C(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} \{z_p(u_i) - \hat{E}[z_p(u_i)]\} \{z_q(u_{i+h}) - \hat{E}[z_q(u_{i+h})]\} \tag{9}$$

$C(h)$ is auto- or cross-covariance when subscripts p and q are equal or not equal. In computing auto- and cross-covariance of aggregated maps, two directions were selected: along the x axis

(horizontal) and horizontal at 45º from y axis, since the longest distance of the two-pixel-unit and isotropic covariance were considered in this investigation of spatial variability. The other method used in computing spatial covariance of aggregated maps was described in Eq. 4, which considered every term of the original spatial covariance inside a pair of aggregated pixels.

5. Results and Analysis

Figures 3, 4, and 5 show variance and covariance distribution maps of Sand and Silt generated using Eq. 3. They illustrate the changes of spatial distribution of the variances and covariance of Sand and Silt when aggregation was conducted using different scales. When pixel size was larger, the ranges of variances and covariance of the aggregated maps were smaller and the patches of large/small variance/covariance were clearer. The ranges shrank from both ends. The maximal absolute value of variance/covariance became smaller as the scale became courser. In the aggregated maps, the minimal variance/covariance was larger than that in the original maps.

When the average of variances at the original pixel size was taken as the variance of a pixel in a course map, the range of the variance of the aggregated maps also shrank from both ends (Figure 6), but much less compared to that generated using Eq. 3. The patterns in the variance maps generated using the different methods were very similar (Figures 3 and 6).

Figure 7 demonstrated the relationship among scale, distance, and spatial covariance in aggregated maps. Using both computational methods (Eqs. 4 and 9), spatial covariance decreased in absolute value as scale became courser. However, there was a difference in the changing rates. When Eq. 4 was used in computing spatial covariance, the effect of scales on the auto-/cross-covariance could be seen in the figures. The differences of the (co)variance at two distances were smaller when the scale was larger, especially when distance changed from 0 to 1 (Figure 7, column A). When Eq. 9 was used in computation and average variation was taken as the variation at distance of 0, the effect of scales could not be seen in the charts, and the difference of variation between distance 0 and 1 was very large (Figure 7, column B). This big change indicated losing/gaining variation during aggregation.

6. Discussion and Conclusion

This study concentrated on spatial variability of continuous attributes when their values are averaged in aggregation. Equations have been analytically derived for computing the spatial distribution of variance and covariance, and the spatial covariance after aggregation using different scales. Using the derived equations, one can easily and correctly provide the spatial variability of aggregated maps using any scale, and the uncertainty inside a study area will not change after aggregation. These equations can also be used to compute the spatial variability of sub-areas of maps, although the case study only demonstrated their usage in aggregating pixels. This makes it possible to compute the spatial variability of a map after GIS analyses.

In aggregation of maps, the change of scale causes the change of spatial variability of individual pixels in two aspects: variance/covariance inside pixels and spatial covariance among pixels. As shown in the case study, increasing pixel size will reduce variation inside and among individual pixels. This is different from the results of Woodcock and Strahler (1987). Their results showed that the variance has an approximately quadratic change when scale increases. When the scale is finer than a threshold, the variance increases as the scale becomes courser.

After that threshold, the variance tends to decrease. In the case study, the original pixel size is 100×100 m². It is much courser than most of their tested scales. So, the tested scale in this study must have been courser than the threshold. Therefore, only the decreasing trend of variation has been shown. At courser scales, Turner (1991) also found a decrease in variance.

Semivariogram models are established using observations. They can be used to express spatial correlation based on the original scale (pixel size). In aggregated maps, the pixel size is larger than the original one, which is based on the plot size of the observations. Therefore, the original semivariogram models may not be used after aggregation. Since the variance/covariance of aggregated pixels is usually smaller than that of the original map, nugget and sill values of the semivariogram models need to be modified to fit the aggregated maps. Zhang et al. (1990) mentioned that semivariogram models are different when sampling plot size varies. Comparison of the spatial covariance computed using different methods verified that finding.

This study also suggests that it might be better to use the original scale in geostatistical analysis when the unified scale is courser. After producing the original maps, multi-scale maps can be unified using aggregation techniques. Based on this suggested process, uncertainty within study areas will not be changed through unification of scales by aggregation.

Acknowledgement

References

Almeida, A.S., 1993. Joint simulation of multiple variables with a Markov-type coregionalization model, The Department of Applied Earth Sciences, Stanford University, Ph.D. Dissertation, 199 p.

Chilès, J. and P. Delfiner, 1999. Geostatistics modeling spatial uncertainty. Wiley, New York.

Clark, I. and W. V. Harper, 2001. Practical geostatistics 2000. Geostokos, Scotland.

Deutsch, C. V. and A. G. Journel, 1998. GSLIB: Geostatistical Software Library and User's Guide (2nd Edition). Oxford University Press, Inc., New York, NY, P. 45.

Gertner, G.Z., G. Wang, S. Fang, and A. B. Anderson, 2002. Error budget assessment of the effect of DEM spatial resolution in predicting topographical factor for soil loss estimation. Soil and Water Conservation, 57:164-174.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford University Press, Inc., New York, NY, p. 483.

Hay, G. J., K. O. Niemann, and D. G. Goodenough, 1997. Spatial thresholds, image-objects, and upscaling: a multiscale evaluation. Remote Sensing of Environment, Vol. 62: 1-19.

Isaaks, E. H. and R. M. Srivastava, 1989. An introduction to applied geostatistics. Oxford University Press, Inc., New York

Journel, A. G. and CH. J. Huijbregts, 1978. Mining geostatistics. Academic Press Ltd., London.

King, A. W., 1991. Translating models across scales in the landscape. In: Quantitative methods in landscape ecology: the analysis and interpretation of landscape heterogeneity (Eds. M. G. Turner and R. H. Gardner). Springer-Verlag New York Inc. PP.479-517.

Myers, D. E., 1997. Statistical models for multi-scaled analysis. In: Scale in remote sensing and GIS (Eds. D. A. Quattrochi and M. F. Goodchild). Lewis Publishers. PP. 273-293.

O'Neill, R. V., 1989. Perspectives in hierarchy and scale. In: Perspectives in ecological theory (Eds. R. M. May and S. A. Levin). Princeton University Press.

Tazik, D.J., S.D. Warren, V.E. Diersing, R.B. Shaw, R.J. Brozka, C.F. Bagley, and W.R. Whitworth, 1992. U.S. Army Land Condition Trend Analysis (LCTA) plot inventory field methods. USACERL, Technical Report N-92/03, Champaign, IL, USA, 62 p.

Townsend, J. R. G. and C. O. Justice, 1988.Selecting the spatial resolution of satellite sensors required for global monitoring of land transformations. International Journal of Remote Sensing. Vol. 9:187-236.

Turner, M. G., V. H. Dale, and R. H. Gardner, 1989. Predicting across scales: theory development and testing. Landscape Ecology. Vol. 3:245-252.

Turner, S, J., R. V. O'Neill, W. Conley, M. R. Conley, and H. C. Humphries, 1991. Pattern and scale: statistics for landscape ecology. In: Quantitative methods in landscape ecology: the analysis and interpretation of landscape heterogeneity (Eds. M. G. Turner and R. H. Gardner). Springer-Verlag New York Inc. PP. 17-49.

Wang, G., Gertner, G.Z., Parysow, P., and Anderson, A.B., 2001a. Spatial prediction and uncertainty assessment of topographic factor for RUSLE using DEM. ISPRS J. of Photogrammetry and Remote Sensing, 56(1), 65-80.

Wang, G., Gertner, G.Z., Xiao, X., Wente, S., and Anderson, A.B., 2001b, Appropriate plot size and spatial resolution for mapping multiple vegetation cover types. Photogrammetric Engineering & Remote Sensing, 67(5), 575-584.

Woodcock, C. E., and A. H. Strahler, 1987.The factor of scale in remote sensing. Remote Sensing of Environment, Vol. 21:311-332.

Zhang, R., A. W. Warrick, and D. E. Myers, 1990. Variance as a function of the sample support size. Mathematical Geology, Vol. 22(1):107-121.

Zhu, H. and A. G. Journel, 1993. Formatting and integrating soft data: stochastic image via the Markov-Bayes algorithm. In: Geostatistics Tróia 1992 (Ed. A. Soarse), PP. 1-12. Kluwer Academic Publishers.

Tables

Table 1. Descriptive statistics of the observations of sand and silt.

| Statistic	Sand (%)	Silt (%)
Mean	13.80	52.48
Minimum	1.0	30.0
Maximum	54.0	70.0
Standard deviation	11.30	7.12
Coefficient of Correlation	-0.3856	

Table 2. Parameters of the standardized semivariogram models of sand and silt. Unit of Range: 100 meters.

Parameters / Variables	Nugget	Sill	Range
Sand	0.52137	0.42768	309
Silt	0.84479	0.19349	230

Table 3. Scales used in aggregation.

Scale Code	Scale	Pixel Size
1 (original)	1	100×100 m²
2	4	200×200 m²
3	9	300×300 m²
4	25	500×500 m²
5	81	900×900 m²

Applied Statistics in Agriculture
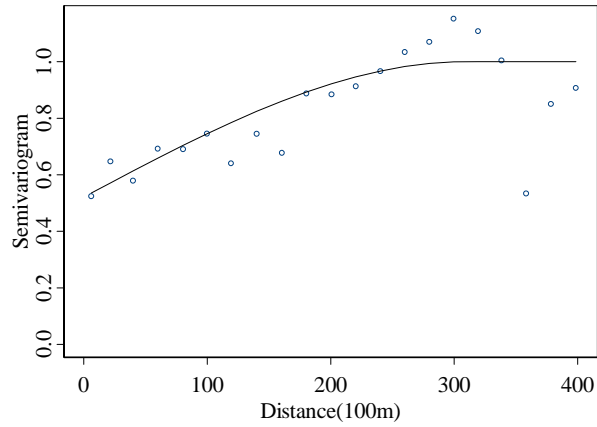
Figures

a
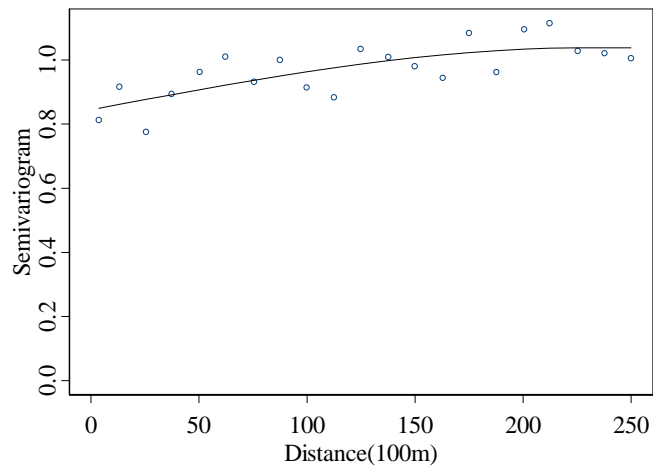
b

Figure 1. The observed and modeled standardized semivariograms of Sand (a) and Silt (b).

Sand
4.2 - 7.6
7.6 - 11
11 - 14.4
14.4 - 17.7
17.7 - 21.1
21.1 - 24.5
24.5 - 27.9
No Data

a

N

Silt
43.6 - 45.8
45.8 - 48.1
48.1 - 50.3
50.3 - 52.5
52.5 - 54.7
54.7 - 56.9
56.9 - 59.2
No Data

7000   0   7000 14000   Meters

b
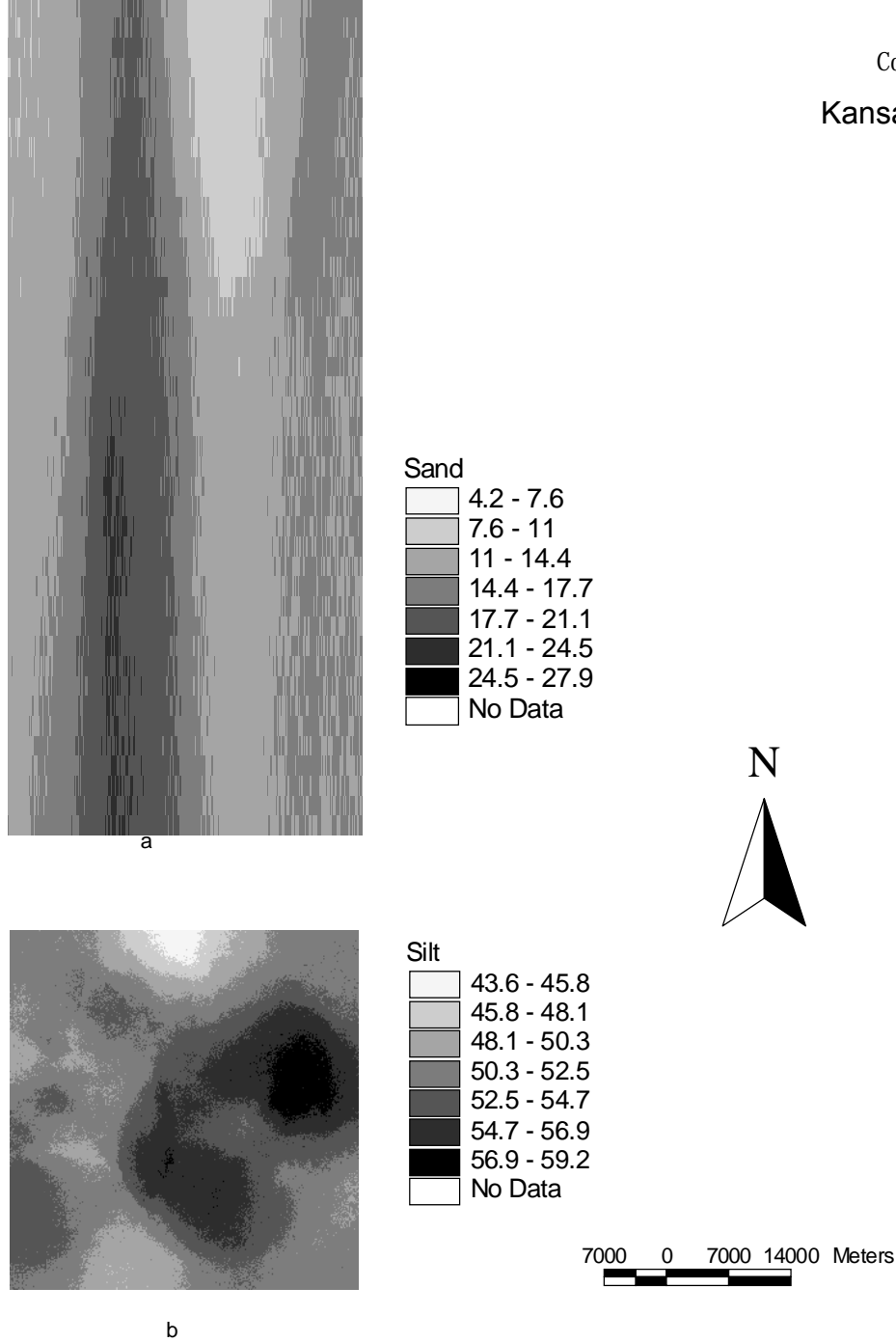
Figure 2. Spatial distribution of the expectation (%) of Sand (a) and Silt (b) in soil. Markov method was used in the spatial joint simulation.
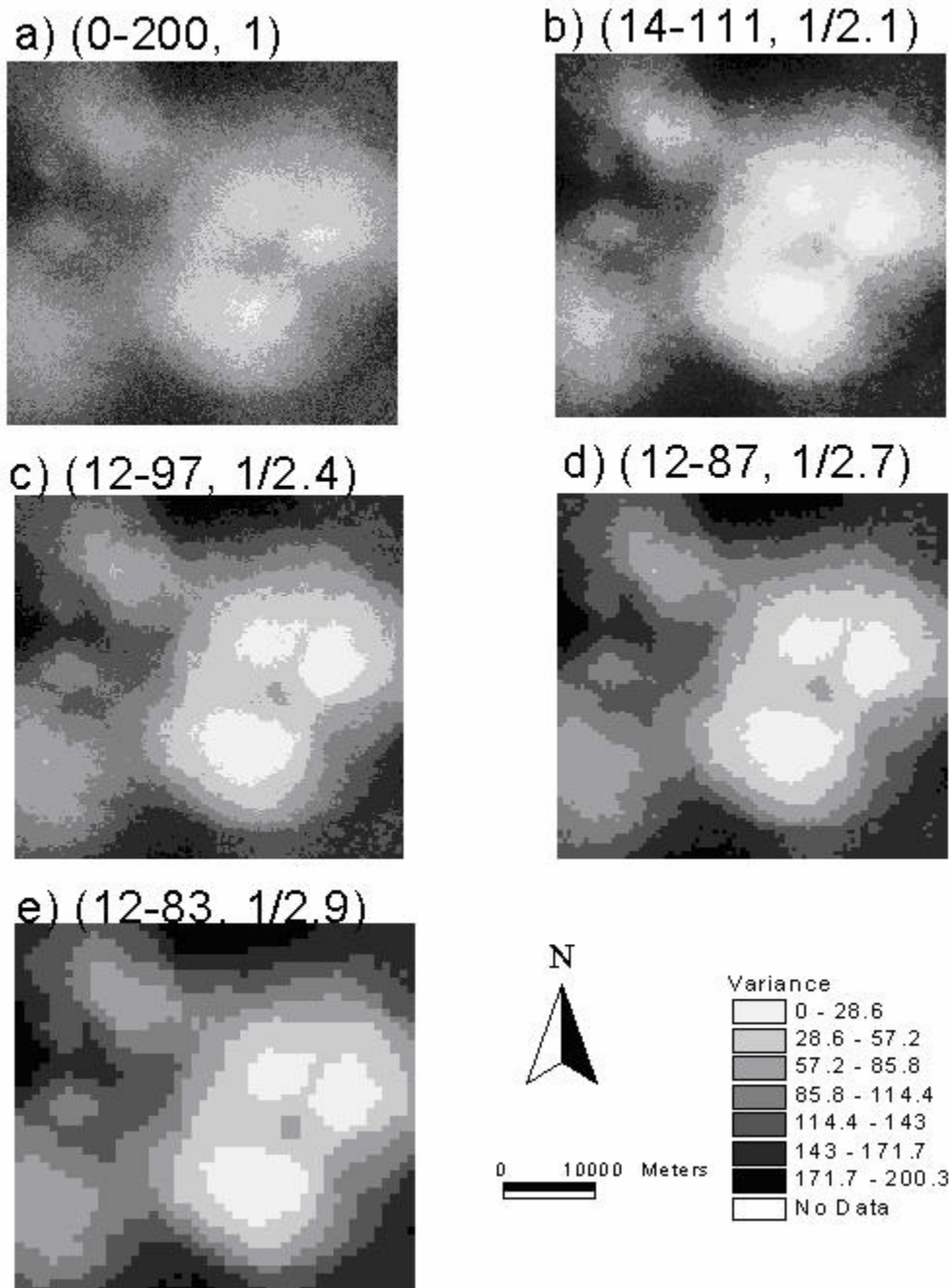
Figure 3. Spatial distribution of the variance of Sand (generated using Eq. 3). The numbers within parentheses above each map are (minimum-maximum, conversion factor). The actual values of the maps should be converted using these parameters and Eq. 8. From a to e, the pixel sizes were 100x100 m², 200x200 m², 300x300 m², 500x500 m², and 900x900 m², respectively.
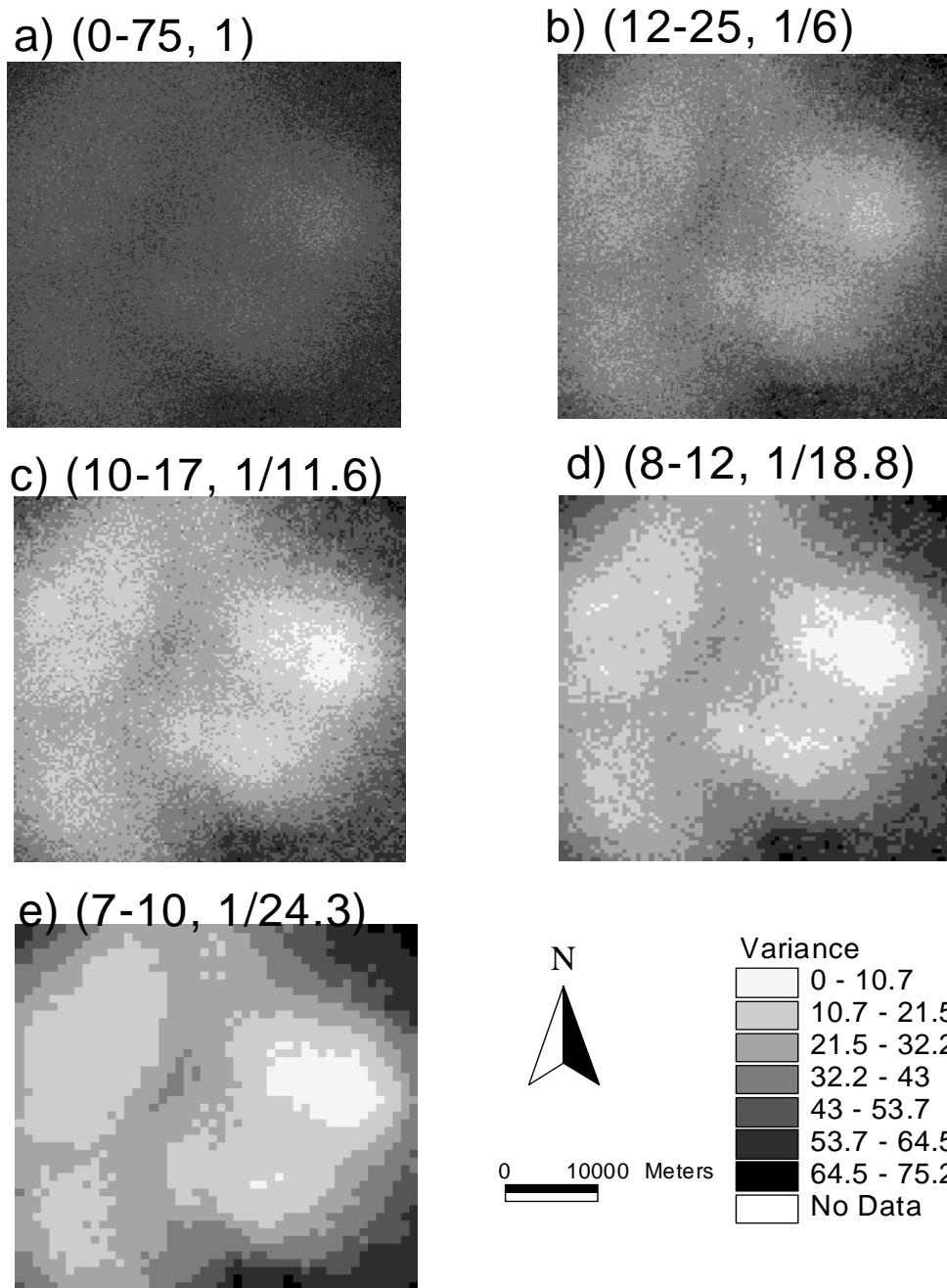
Figure 4. Spatial distribution of the variance of Silt (generated using Eq. 3). The numbers within parentheses above each map are (minimum-maximum, conversion factor). The actual values of the maps should be converted using these parameters and Eq. 8. From a to e, the pixel sizes were 100x100 m², 200x200 m², 300x300 m², 500x500 m², and 900x900 m², respectively.
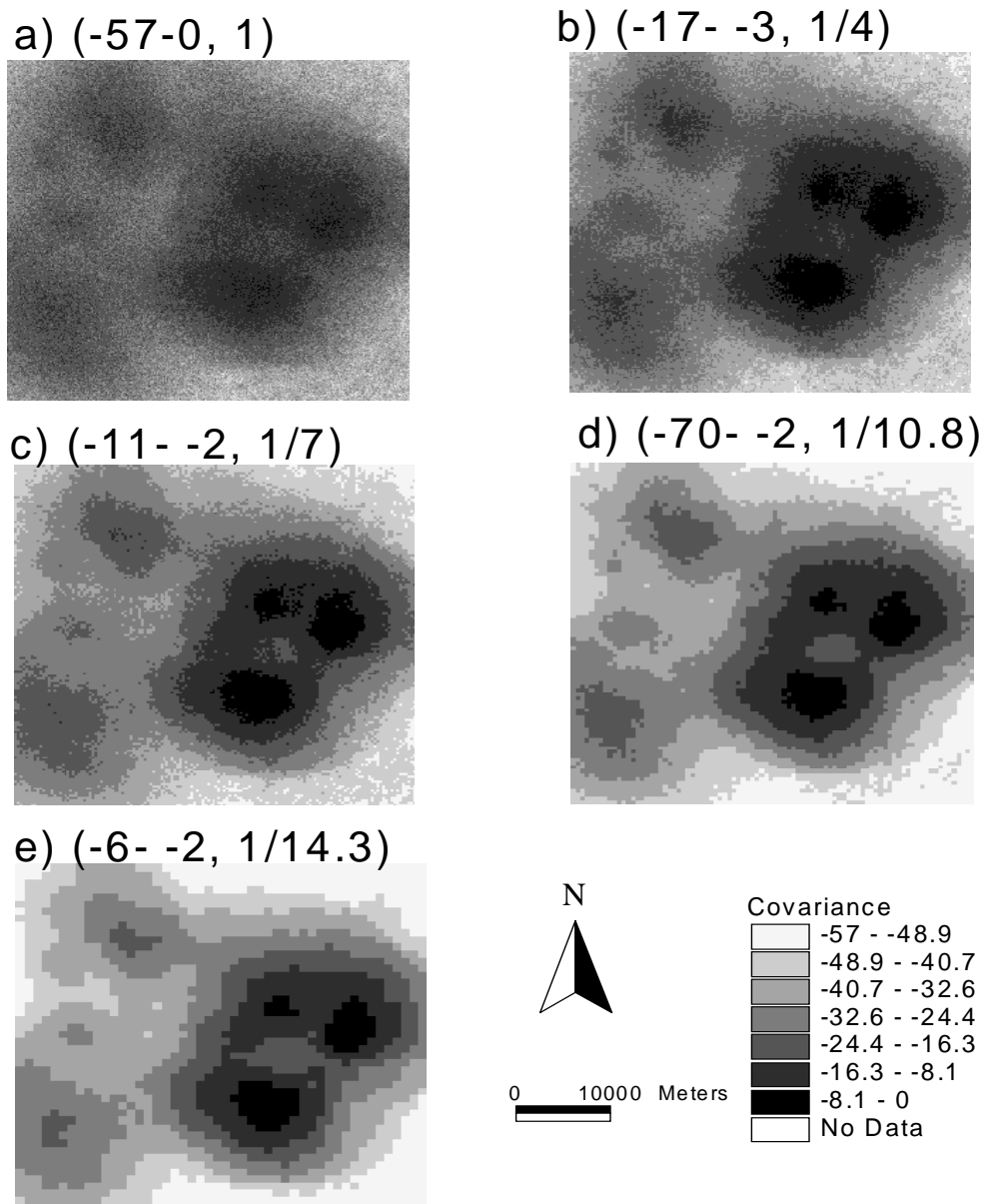
Figure 5. Spatial distribution of the covariance of Sand and Silt (generated using Eq. 3). The numbers within parentheses above each map are (minimum-maximum, conversion factor). The actual values of the maps should be converted using these parameters and Eq. 8. From a to e, the pixel sizes were 100x100 m², 200x200 m², 300x300 m², 500x500 m², and 900x900 m², respectively.
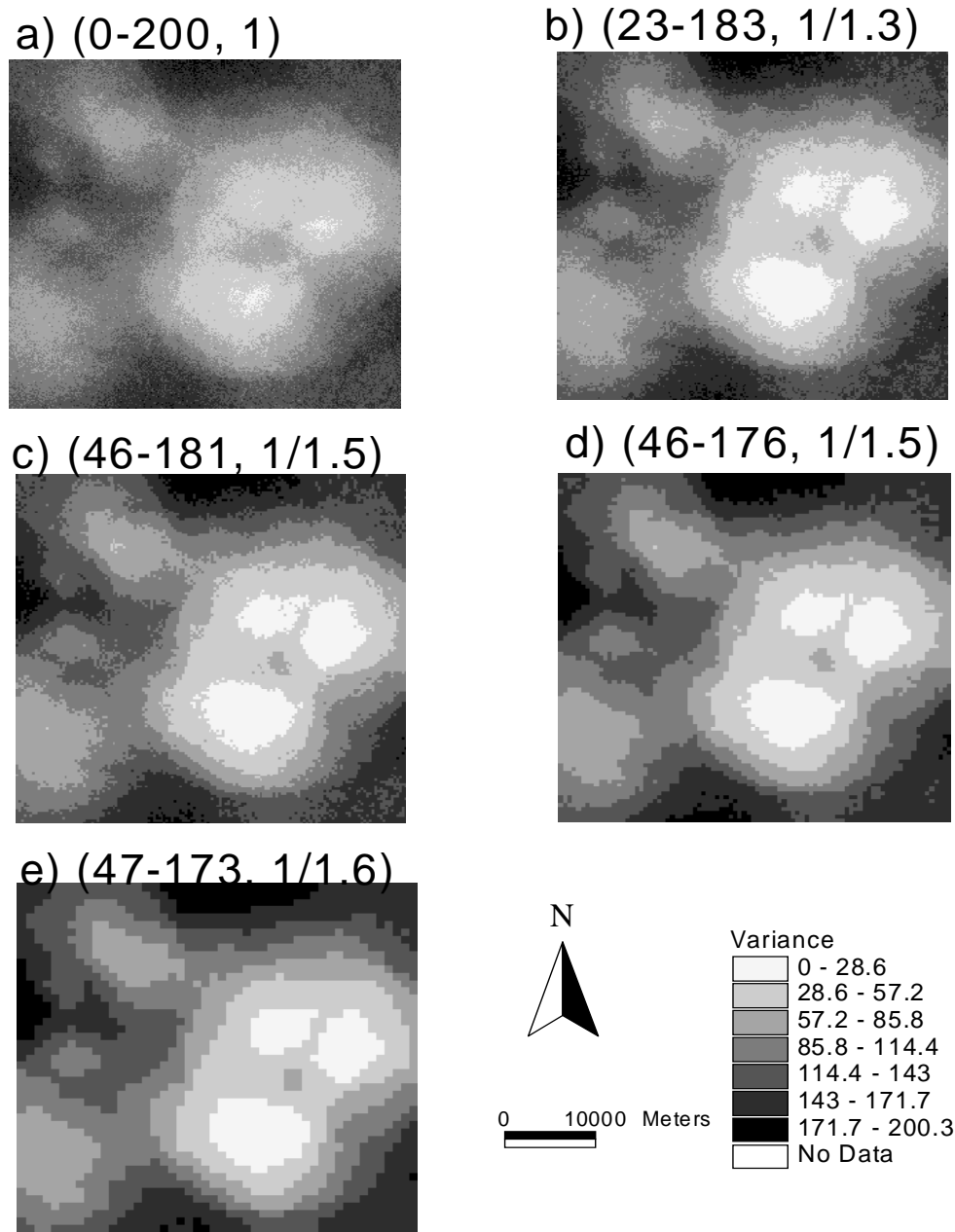
Figure 6. Spatial distribution of the variance of Sand (generated using average of variance of original pixels). The numbers within parentheses above each map were (minimum-maximum, conversion factor). The actual values of the maps should be converted using these parameters and Eq. 8. From a to e, the pixel sizes were 100x100 m², 200x200 m², 300x300 m², 500x500 m², and 900x900 m², respectively.

A (Eq.4)   B (Eq.9)

a.1   b.1

a.2   b.2
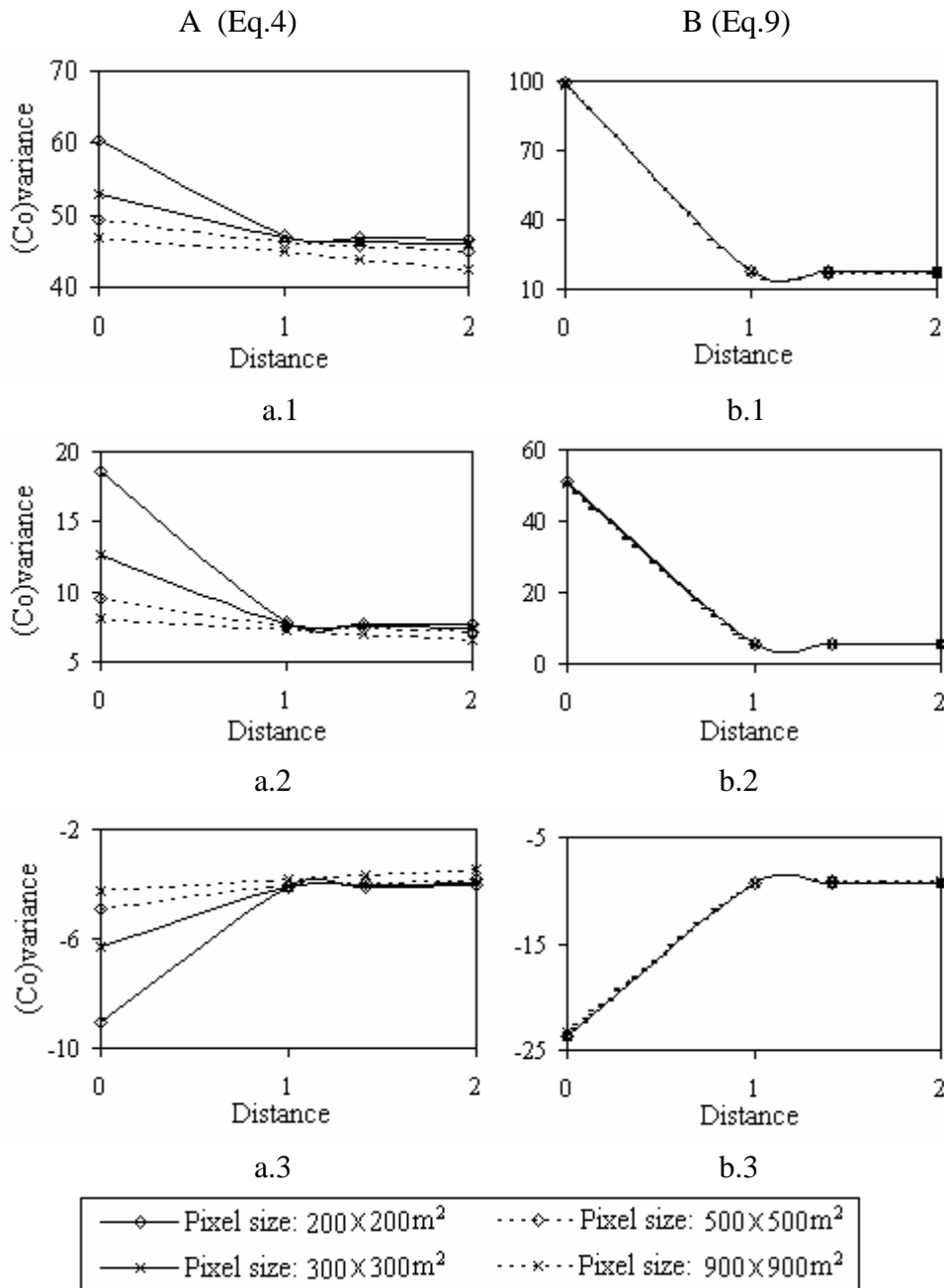

a.3   b.3

Figure 7. Spatial variability computed using Eq.4 (Column A) and Eq.9 (Column B) as distance
(unit: pixel length) and scale change. When distance was zero, (co)variance is the variance of
Sand (.1) and Silt (.2), and the covariance of Sand and Silt (.3). When distance was larger than 0,
in (.1) and (.2), it was auto-covariance of Sand and Silt, respectively. In (.3), it was cross-
covariance of Sand and Silt.

Appendix A. Variance of the aggregated pixels

Assume that the attribute of interest z at an original pixel u_i has mean $E[z(u_i)]$ and variance $var[z(u_i)]$, and a rescaled (aggregated from original pixels) pixel v_k contains n original pixels. The value of attribute z at the rescaled pixel v_k is the average value of the z values of the included original pixels:

$$z(v_k) = \frac{1}{n}\sum_{i=1}^{n} z(u_i)$$

Therefore, its mean is:

$$E[z(v_k)] = \frac{1}{n}\sum_{i=1}^{n} E[z(u_i)]$$

and its variance is:

$$var[z(v_k)] = var\left[\frac{1}{n}\sum_{i=1}^{n} z(u_i)\right]$$

$$= \frac{1}{n^2}\left\{\sum_{i=1}^{n} var[z(u_i)] + \sum_{i=1}^{n}\sum_{l\neq i} cov[z(u_i), z(u_l)]\right\}$$

This equation can be verified by using samples of $z(u_i)$. Assume that there are m observations of attribute z at each of the n original pixels inside of the rescaled pixel v_k. The estimated variance of $z(v_k)$ is $\widehat{var}[z(v_k)]$:

$$\widehat{var}[z(v_k)] = E\left\{\langle z(v_k) - E[z(v_k)]\rangle^2\right\}$$

$$= E\left\{\left\langle\left[\frac{1}{n}\sum_{i=1}^{n} z(u_i)\right] - E\left[\frac{1}{n}\sum_{i=1}^{n} z(u_i)\right]\right\rangle^2\right\}$$

$$= \frac{1}{n^2}\cdot\frac{1}{m}\sum_{j=1}^{m}\left\{\left\langle\left[\sum_{i=1}^{n} z_j(u_i)\right] - E\left[\sum_{i=1}^{n} z(u_i)\right]\right\rangle^2\right\}$$

$$= \frac{1}{n^2}\cdot\frac{1}{m}\sum_{j=1}^{m}\left\{\left\langle\sum_{i=1}^{n}[z_j(u_i) - E[z(u_i)]]\right\rangle^2\right\}$$

$$= \frac{1}{n^2}\cdot\frac{1}{m}\sum_{j=1}^{m}\left\{\sum_{i=1}^{n}\langle z_j(u_i) - E[z(u_i)]\rangle^2 + \sum_{i=1}^{n}\sum_{l\neq i}\langle z_j(u_i) - E[z(u_i)]\rangle\cdot\langle z_j(u_l) - E[z(u_l)]\rangle\right\}$$

$$= \frac{1}{n^2}\left\{\sum_{i=1}^{n}\frac{1}{m}\sum_{j=1}^{m}\langle z_j(u_i) - E[z(u_i)]\rangle^2 + \right.$$

Applied Statistics in Agriculture

$$+\sum_{i=1}^{n}\sum_{l\neq i}\frac{1}{m}\sum_{j=1}^{m}\left\langle z_{j}(u_{i})-E[z(u_{i})]\right\rangle \cdot \left\langle z_{j}(u_{l})-E[z(u_{l})]\right\rangle \Bigg\}$$

$$=\frac{1}{n^{2}}\left\{\sum_{i=1}^{n}E\left\langle \llbracket z(u_{i})-E[z(u_{i})]\rrbracket^{2}\right\rangle +\sum_{i=1}^{n}\sum_{l\neq i}E\left\langle \llbracket z(u_{i})-E[z(u_{i})]\rrbracket \cdot \llbracket z(u_{l})-E[z(u_{l})]\rrbracket \right\rangle \right\}$$

$$=\frac{1}{n^{2}}\left\{\sum_{i=1}^{n}\widehat{var}[z(u_{i})]+\sum_{i=1}^{n}\sum_{l\neq i}\widehat{cov}[z(u_{i}),z(u_{l})]\right\}$$

$$(A1)$$

Assume that the distance between a pair of original pixels inside of the rescaled pixel is h. Under the assumption of homogeneity in geostatistics, then, the estimated covariance of the attribute z between a pair of original pixels in Eq. A1 becomes:

$$\widehat{cov}[z(u_{i}),z(u_{l})]=\frac{1}{m}\sum_{j=1}^{m}\left\langle z_{j}(u_{i})-E[z(u_{i})]\right\rangle \cdot \left\langle z_{j}(u_{l})-E[z(u_{l})]\right\rangle$$

$$=\frac{1}{m}\sum_{j=1}^{m}\left\langle z_{j}(u_{i})-E[z(u_{i})]\right\rangle \cdot \left\langle z_{j}(u_{i}+h)-E[z(u_{i}+h)]\right\rangle =C(h)$$

where $C(h)$ is the auto-covariance of the attribute z at the distance between u_{i} and u_{l}.

When $Z=(z_{1},\cdots,z_{s})'$ is a vector consisting of s attributes, its value (a vector) at the rescaled pixel v_{k} is still the average value (a vector) from the n original pixels:

$$Z(v_{k})=\frac{1}{n}\sum_{i=1}^{n}Z(u_{i})$$

Its mean vector is:

$$E[Z(v_{k})]=\frac{1}{n}\sum_{i=1}^{n}E[Z(u_{i})]$$

and its covariance matrix is:

$$var[Z(v_{k})]=var\left[\frac{1}{n}\sum_{i=1}^{n}Z(u_{i})\right]$$

$$=\frac{1}{n^{2}}\left\{\sum_{i=1}^{n}\Sigma_{(u_{i},u_{i})}+\sum_{i=1}^{n}\sum_{l\neq i}\Sigma_{(u_{i},u_{l})}\right\} \qquad (A2)$$

where

$$\Sigma_{(u_{j},u_{k})}=\begin{pmatrix} C_{1,1}(u_{j},u_{k}) & C_{1,2}(u_{j},u_{k}) & \cdots & C_{1,s}(u_{j},u_{k}) \\ C_{2,1}(u_{j},u_{k}) & C_{2,2}(u_{j},u_{k}) & \cdots & \vdots \\ \vdots & \vdots & \ddots & C_{s-1,1}(u_{j},u_{k}) \\ C_{s,1}(u_{j},u_{k}) & \cdots & C_{s,s-1}(u_{j},u_{k}) & C_{s,s}(u_{j},u_{k}) \end{pmatrix}$$

where

$$C_{p,q}(u_j, u_k) = cov[z_p(u_j), z_q(u_k)]$$

Similar to Eq. A1, Eq. A2 can also be derived:

$$\widehat{var}[Z(v_k)] = E\left\{\left\langle Z(v_k) - E[Z(v_k)]\right\rangle\left\langle Z(v_k) - E[Z(v_k)]\right\rangle'\right\}$$

$$= \frac{1}{n^2} \cdot \frac{1}{m} \sum_{j=1}^{m}\left\{\left\langle\left[\sum_{i=1}^{n}Z_j(u_i)\right] - E\left[\sum_{i=1}^{n}Z(u_i)\right]\right\rangle\left\langle\left[\sum_{i=1}^{n}Z_j(u_i)\right] - E\left[\sum_{i=1}^{n}Z(u_i)\right]\right\rangle^T\right\}$$

$$= \frac{1}{n^2} \cdot \frac{1}{m} \sum_{j=1}^{m}\left\{\begin{bmatrix}\sum_{i=1}^{n}g_{i,j,1,1} & \sum_{i=1}^{n}g_{i,j,1,2} & \cdots & \sum_{i=1}^{n}g_{i,j,1,s} \\ \sum_{i=1}^{n}g_{i,j,2,1} & \sum_{i=1}^{n}g_{i,j,2,2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \sum_{i=1}^{n}g_{i,j,s-1,1} \\ \sum_{i=1}^{n}g_{i,j,s,1} & \cdots & \sum_{i=1}^{n}g_{i,j,s,s-1} & \sum_{i=1}^{n}g_{i,j,s,s}\end{bmatrix}\right\} +$$

$$+ \frac{1}{n^2} \cdot \frac{1}{m} \sum_{j=1}^{m}\left\{\begin{bmatrix}\sum_{i=1}^{n}\sum_{l\neq i}g'_{i,l,j,1,1} & \sum_{i=1}^{n}\sum_{l\neq i}g'_{i,l,j,1,2} & \cdots & \sum_{i=1}^{n}\sum_{l\neq i}g'_{i,l,j,1,s} \\ \sum_{i=1}^{n}\sum_{l\neq i}g'_{i,l,j,2,1} & \sum_{i=1}^{n}\sum_{l\neq i}g'_{i,l,j,2,2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \sum_{i=1}^{n}\sum_{l\neq i}g'_{i,l,j,s-1,s} \\ \sum_{i=1}^{n}\sum_{l\neq i}g'_{i,l,j,s,1} & \cdots & \sum_{i=1}^{n}\sum_{l\neq i}g'_{i,l,j,s,s-1} & \sum_{i=1}^{n}\sum_{l\neq i}g'_{i,l,j,s,s}\end{bmatrix}\right\}$$

where

$$g_{i,j,p,q} = \{z_{j,p}(u_i) - E[z_p(u_i)]\} \cdot \{z_{j,q}(u_i) - E[z_q(u_i)]\}$$

and

$$g'_{i,l,j,p,q} = \{z_{j,p}(u_i) - E[z_p(u_i)]\} \cdot \{z_{j,q}(u_l) - E[z_q(u_l)]\}$$

Change the order of summations as in Eq. A1, the relationship described in Eq. A2 holds.

Appendix B. Spatial covariance among the aggregated pixels

Based on Eq. A1, Spatial (auto-/cross-) covariance of the attributes among the aggregated pixels can also be derived. Assume that there are k_n aggregated pixels and each one contains n_k original pixels. Pixel v_k is aggregated from the k_n aggregated pixels and contains a total of n ($= n_k \cdot k_n$) original pixels. Therefore, the variance of an attribute z at rescaled pixel v_k can be calculated using the k_n aggregated pixels:

Applied Statistics in Agriculture

$$\text{var}[z(v_k)] = \frac{1}{k_n^2}\left\{\sum_{j=1}^{k_n}\text{var}[z(v_j)] + 2\sum_{j=1}^{k_n-1}\sum_{j1>j}^{k_n}\text{cov}[z(v_j),z(v_{j1})]\right\} \tag{A3}$$

or directly using the n original pixels:

$$\text{var}[z(v_k)] = \frac{1}{n^2}\left\{\sum_{i=1}^{n}\text{var}[z(u_i)] + 2\sum_{i=1}^{n-1}\sum_{l>i}^{n}\text{cov}[z(u_i),z(u_l)]\right\} \tag{A4}$$

The variances converted using Eqs. A3 and A4 have to be consistent. Therefore, the spatial covariance among the aggregated pixels should be defined to satisfy this restriction. When the auto-covariance of the aggregated pixels is defined as:

$$\text{cov}[z(v_{k1}),z(v_{k2})] = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\text{cov}[z(u_i),z(u_j)|u_i \in v_{k1}, u_j \in v_{k2}] \tag{A5}$$

the variances converted from Eqs. A3 and A4 are equal:

$$\text{var}[z(v_k)] = \frac{1}{k_n^2}\left\{\sum_{j=1}^{k_n}\text{var}[z(v_j)] + 2\sum_{j=1}^{k_n-1}\sum_{j1>j}^{k_n}\text{cov}[z(v_j),z(v_{j1})]\right\}$$

$$= \frac{1}{n^2}\left\{n_k^2\sum_{j=1}^{k_n}\text{var}[z(v_j)] + 2n_k^2\sum_{j=1}^{k_n-1}\sum_{j1>j}^{k_n}\text{cov}[z(v_j),z(v_{j1})]\right\}$$

$$= \frac{1}{n^2}\left\{\sum_{j=1}^{k_n}\left\langle\sum_{i=1}^{n_k}\text{var}[z(u_{i,j})] + 2\sum_{i=1}^{n_k-1}\sum_{l>i}^{n_k}\text{cov}[z(u_{i,j}),z(u_{l,j})]\right\rangle + \right.$$

$$\left. +2\sum_{j=1}^{k_n-1}\sum_{j1>j}^{k_n}\sum_{i=1}^{n_k}\sum_{l=1}^{n_k}\text{cov}[z(u_{i,j}),z(u_{l,j1})]\right\}$$

$$= \frac{1}{n^2}\left\{\sum_{j=1}^{k_n}\sum_{i=1}^{n_k}\text{var}[z(u_{i,j})] + 2\sum_{j=1}^{k_n}\sum_{i=1}^{n_k-1}\sum_{l>i}^{n_k}\text{cov}[z(u_{i,j}),z(u_{l,j})] + \right.$$

$$\left. +2\sum_{j=1}^{k_n-1}\sum_{j1>j}^{k_n}\sum_{i=1}^{n_k}\sum_{l=1}^{n_k}\text{cov}[z(u_{i,j}),z(u_{l,j1})]\right\}$$

$$= \frac{1}{n^2}\left\{\sum_{i=1}^{n}\text{var}[z(u_i)] + 2\sum_{i=1}^{n-1}\sum_{l>i}^{n}\text{cov}[z(u_i),z(u_l)]\right\}$$

In multivariate case, it can be verified that Eq. A5 becomes more general:

$$\text{cov}[z_p(v_{k1}),z_q(v_{k2})] = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\text{cov}[z_p(u_i),z_q(u_j)|u_i \in v_{k1}, u_j \in v_{k2}] \tag{A6}$$

where $\text{cov}[z_p(v_{k1}),z_q(v_{k2})]$ is the auto-/cross-covariance of attribute(s) z_p and z_q (p could equal to q) between aggregated pixels v_{k1} and v_{k2}, each of which contains n original pixels.

Appendix C. Spatial variability of aggregated sub-areas

When aggregation is based on sub-areas which contain different numbers of original pixels, the larger sub-area (aggregated from the smaller sub-areas) v_k has its mean:

$$E[z(v_k)] = \frac{1}{n}\sum_{i=1}^{n}E[z(u_i)] = \sum_{j=1}^{k_n}\frac{n_j}{n}E[z(v_j)]$$

where n is the total number of original pixels contained in the larger sub-area v_k, k_n is the number of the smaller sub-areas, and n_j is the number of original pixels contained inside smaller sub-area v_j. The variance of $z(v_k)$ is:

$$\text{var}[z(v_k)] = \sum_{j=1}^{k_n}\frac{n_j^2}{n^2}\text{var}[z(v_j)] + 2\sum_{j=1}^{k_n-1}\sum_{j1>j}^{k_n}\frac{n_j n_{j1}}{n^2}\text{cov}[z(v_j),z(v_{j1})] \qquad (A7)$$

where the auto-covariance of attribute z between smaller sub-areas j and j1 is defined as:

$$\text{cov}[z(v_j),z(v_{j1})] = \frac{1}{n_j}\sum_{i=1}^{n_j}\frac{1}{n_{j1}}\sum_{i1=1}^{n_{j1}}\text{cov}[z(u_i),z(u_{i1})|u_i \in v_j, u_{i1} \in v_{j1}] \qquad (A8)$$

Based on Eqs. A3, A5, A7, and A8, the variances converted using the original pixels and smaller sub-areas are consistent:

$$\text{var}[z(v_k)] = \frac{1}{n^2}\left\{\sum_{i=1}^{n}\text{var}[z(u_i)] + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n}\text{cov}[z(u_i),z(u_j)]\right\}$$

$$= \frac{1}{n^2}\left\{\sum_{j=1}^{k_n}n_j^2\text{var}[z(v_j)] + 2\sum_{j=1}^{k_n-1}\sum_{j1>j}^{k_n}n_j n_{j1}\text{cov}[z(v_j),z(v_{j1})]\right\}$$

$$= \sum_{j=1}^{k_n}\frac{n_j^2}{n^2}\text{var}[z(v_j)] + 2\sum_{j=1}^{k_n-1}\sum_{j1>j}^{k_n}\frac{n_j n_{j1}}{n^2}\text{cov}[z(v_j),z(v_{j1})]$$

For multivariate cases, it can be verified that Eqs. A7 and A8 become:

$$\text{var}[z_p(v_k)] = \sum_{j=1}^{k_n}\frac{n_j^2}{n^2}\text{var}[z_p(v_j)] + 2\sum_{j=1}^{k_n-1}\sum_{j1>j}^{k_n}\frac{n_j n_{j1}}{n^2}\text{cov}[z_p(v_j),z_p(v_{j1})] \qquad (A9)$$

and

$$\text{cov}[z_p(v_j),z_q(v_{j1})] = \frac{1}{n_j}\sum_{i=1}^{n_j}\frac{1}{n_{j1}}\sum_{i1=1}^{n_{j1}}\text{cov}[z_p(u_i),z_q(u_{i1})|u_i \in v_j, u_{i1} \in v_{j1}] \qquad (A10)$$

When the area, instead of the number of original pixels, of the smaller sub-areas are provided in a map, the numbers of original pixels, n and n_j, of v_k and v_j in Eqs. A7 to A10 can be replaced with their corresponding areas.

Applied Statistics in Agriculture

Appendix D. Joint sequential simulation

The joint sequential simulation algorithm is based on Bayes' conditional probability axiom to generate L joint realizations of P variables (attributes) from their Cumulative Distribution Functions (CDFs) (Almeida, 1993). A realization implies that each cell of the grid is provided with an estimation vector of the variables, derived from the vectors of neighboring field plots and the vector of the secondary variables at the location to be estimated. Theoretically, a joint P-variable CDF characterizing the P random events can be decomposed into a product of (P-1) univariate conditional CDFs and a marginal CDF according to Bayes' axiom for conditional probability. From the decomposition, a general sequential simulation algorithm can be developed to jointly simulate the P dependent variables by drawing from the sequence of univariate conditional CDFs.

A joint sequential simulation is completed when all the cells are visited and provided with simulated values. Repeating the joint sequential simulation process many times with probable different visiting paths leads to a set of estimates at each location for each variable. Finally, an expected vector and co-variance matrix for P variables at each location is calculated.

Appendix E. Calculation of spatial covariance

Using the standardized semivariogram models, the auto covariance of attributes can be computed and the cross covariance of a pair of attributes can be approximated (Goovaerts, 1997). The auto covariance of an attribute, $C_{1,1}(h)$, is calculated as:

$$C_{1,1}(h) = C_{1,1}(0) - \gamma_{1,1}(h) \cdot C_{1,1}(0)$$
$$= C_{1,1}(0) \cdot [1 - \gamma_{1,1}(h)] \tag{A11}$$

where $\gamma_{1,1}(h)$ is the standardized auto semivariogram given a distance h. When a pair of attributes are correlated, their correlogram, $\rho_{1,2}(h)$, is approximately:

$$\rho_{1,2}(h) \simeq \rho_{1,2}(0) \cdot [1 - \gamma_{1,1}(h)]$$

where $\rho_{1,2}(0)$ is the coefficient of correlation of that pair of attributes. Their cross covariance, $C_{1,2}(h)$, is approximately:

$$C_{1,2}(h) \simeq \rho_{1,2}(h) \cdot \sqrt{C_{1,1}(0) \cdot C_{2,2}(0)}$$
$$\simeq [1 - \gamma_{1,1}(h)] \cdot \rho_{1,2}(0) \cdot \sqrt{C_{1,1}(0) \cdot C_{2,2}(0)}$$
$$= [1 - \gamma_{1,1}(h)] \cdot \frac{C_{1,1}(0)}{C_{1,1}(0)} \cdot C_{1,2}(0)$$
$$= \frac{C_{1,2}(0)}{C_{1,1}(0)} \cdot C_{1,1}(h) \tag{A12}$$

In the case study of this paper, Silt was taken as the dominant attribute in calculation of the cross covariance of Sand and Silt.