SAMPLING SCALE EFFECTS IN RANDOM FIELDS AND IMPLICATIONS FOR ENVIRONMENTAL MONITORING

JON OLAV SKØIEN and GÜNTER BLÖSCHL*

*Institute for Hydraulic and Water Resources Engineering, Vienna University of Technology,
Karlsplatz 13, Vienna, Austria*
(*author for correspondence, e-mail: bloeschl@hydro.tuwien.ac.at)

Abstract. The concept of a sampling scale triplet of spacing, extent and support is used to define the spatial dimensions of a monitoring network or a field study. The spacing is the average distance between samples, the extent is the size of the domain sampled and the support is the averaging area of one sample. The aim of this paper is to examine what is the bias and the random error (uncertainty) introduced by the sampling scale triplet into estimates of the mean, the spatial variance and the integral scale of a variable in a landscape. The integral scale is a measure of the average distance over which a variable is correlated in space. A large number of two dimensional random fields are generated from which hypothetical samples, conforming to a certain sampling scale triplet, are drawn which in turn are used to estimate the sample mean, spatial variance and integral scale. The results indicate that the biases can be up to two orders of magnitude. The bias of the integral scale is positively related to the magnitude of any of the components of the scale triplet while the bias of the spatial variance is different for different components of the scale triplet. All sampling scale effects are relative to the underlying correlation length of the variable of interest which is closely related to the integral scale. The integral scale can hence be used for sampling design and data interpretation. Suggestions are given on how to adjust a monitoring network to the scales of the variables of interest and how to interpret sampling scale effects in environmental data.

Keywords: sampling, scale effects, scale triplet, variogram, integral scale, sampling bias, sampling design

1. Introduction

Scientists in numerous environmental disciplines are interested in the spatial distribution of a variable in the landscape at a given point in time. The variable of interest can be sampled at a number of locations and subsequently analysed to estimate those statistical characteristics of the variable across the landscape that are deemed relevant in a particular context. The three statistical characteristics most relevant to spatial distributions are the mean, the spatial variance and the integral scale of the variable of interest. The mean is a measure of the average behaviour within a domain. The spatial variance is a measure of how different the variables are in space. The integral scale is a measure of the average distance over which a variable is correlated in space (see, e.g., Skøien *et al.*, 2003). Small integral scales indicate that the variable varies erratically over short distances while large integral

scales indicate that the variable varies smoothly over short distances and variability starts to get significant at larger distances. This paper focuses on these three statistical characteristics.

The mean, the spatial variance and the integral scale are of interest in a number of disciplines within the environmental sciences to represent water quality parameters in estuaries (Caeiro *et al*., 2003); characteristics of biological communities (Franklin and Mills, 2003; Jurado-Expósito *et al*., 2004; Nunan *et al*., 2002; Wallace *et al*., 2000), radon concentration in soils (Oliver and Khayrat, 2001), hydraulic conductivity in aquifers (Gelhar, 1993, p. 292), and soil moisture characteristics of landscapes (Mohanty *et al*., 2000; Western *et al*., 2004), among many other examples. In these studies the three statistical characteristics are used for a range of purposes, as parameters for understanding of processes involving hypothesis testing, for quantifying a resource, for compliance, for assisting management decisions and for risk assessment.

All of these studies have in common that it is rarely possible to sample the variable of interest exhaustively and with infinite detail. There will always be some constraints imposed by the measurement procedure and the costs associated with the sampling. For example, soil moisture can be sampled by time domain reflectometry *in situ* probes that, typically, provide an estimate of soil moisture in the top 30 cm of the soil over a measurement area of 1 dm^2 (Western and Blöschl, 1999). Alternatively, remotely sensed data can be used to provide estimates of surface soil moisture over an area of, say, 1 ha or more, depending on the sensor that is used (Entin *et al*., 2000). Similar measurement areas apply to most environmental measurements. In micro meteorological studies, for example, the effective fetch is the area over which the sensors on a micrometeorological tower provide information. It changes with the meteorological conditions and can be on the order of hundreds of metres in length (Schmid, 2002). Another example are groundwater pumping tests that integrate information over the depression cone. Here, the measurement area is on the order of 1 ha, depending on the subsurface characteristics and the withdrawal rate (Anderson, 1997). All of these examples relate to the sampling scale that is variably termed grain, footprint or support depending on the discipline (Bierkens *et al*., 2000; Blöschl and Sivapalan, 1995).

There are two other scales that are important in the spatial arrangement of environmental measurements. These are the spacing and the extent. The spacing is the characteristic distance of two sample locations, i.e., it is small if the samples are narrowly spaced and large if only a few samples are taken over a given domain. Soil hydraulic conductivity measurements in dedicated studies may have spacings of meters (Zehe and Blöschl, 2004) while rain gauges in a region are typically spaced at tens of kilometres. Water quality and air quality samples, often, have larger spacings. This is because measurement costs of small spacings are usually prohibitive, so in many practical cases, the number of samples over a domain is a function of the budget available rather than based on scientific analyses of what would be needed to sample the variable of interest accurately. The extent is the
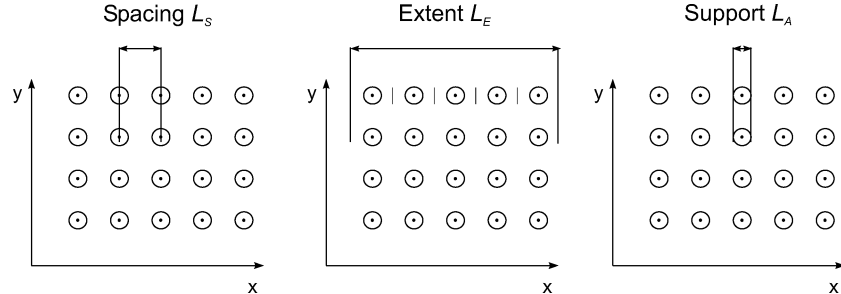
Figure 1. The sampling scale triplet for the two dimensional spatial case: spacing L_S, extent L_E and support L_A. Modified from Blöschl and Sivapalan (1995).

overall size of the domain sampled. It may range from centimetres to hundreds of kilometres, depending on the application. The extent is often selected as the area of interest but sometimes it can be much smaller.

The three scales, spacing, extent and support, have been termed the scale triplet by Blöschl and Sivapalan (1995). In the two-dimensional case, as is of importance in environmental monitoring in landscapes, Blöschl (1999) suggested to define them as characteristic length scales: spacing (L_S), extent (L_E), and support (L_A) (Figure 1).The spacing can be related to the size of the domain, A_{dom}, and the number of samples N, the extent to the size of the domain, and the support can be related to the support area A.

$$
\begin{aligned}
L_S &= \sqrt{\frac{A_{\text{dom}}}{N}} \\
L_E &= \sqrt{A_{\text{dom}}} \\
L_A &= \sqrt{A}
\end{aligned}
\tag{1}
$$

When estimating the mean, the spatial variance and the integral scale of the variable of interest, the spatial sampling scales will have some bearing on the results. The estimates from a finite number of samples, collected over a finite support area will be different from the true statistical characteristics of the variable in the landscape one is interested in because of a number of reasons. If the spacing of the samples is large relative to the underlying variability, one will not sample the small scale variability. If the extent of the samples is small relative to the underlying variability, one will, conversely, not sample the large scale variability. Finally, if the support of the samples is large relative to the underlying variability, the data will appear much smoother than the true spatial distribution of the variable of interest. These effects will result in bias (or systematic error) and uncertainty (or random error) in the estimates of the mean, the spatial variance and the integral scale.

It is not always clear that estimates may have been biased due to sampling scale effects. Franklin and Mills (2003), for example, compared a number of studies of similar microbial communities at different scales. The authors of the original

studies had suggested that different factors may have contributed to structuring those communities. Based on a discussion of Rahel (1990), Franklin and Mills (2003) however suspected that these differences were more likely to reflect viewpoints of different scales rather than differences in the way communities are organised.

While different spatial sampling strategies have attracted considerable interest in the literature (e.g. Morrissey *et al.*, 1995; Thompson, 2002; Webster and Oliver, 2001; McDonald, 2003), research on quantifying sampling scale effects on the statistical characteristics estimated from the data has been more fragmented. The main contribution probably has been from geostatistics which emerged in the 1950s focussing, among other things, on how the spatial variance of a variable (such as the ore grade) changes with the support area or sampling volume (see, e.g., Journel and Huijbregts, 1978). The methods to estimate the change in variance amount to filtering the underlying distribution by a linear filter which reduces the variance and changes the spatial correlation structure. There have been numerous applications throughout the environmental sciences from soil assessment to the estimation of extreme rainfall for catchments (e.g. Webster and Oliver, 2001; Sivapalan and Blöschl, 1998) among many other examples.

The role of the spacing of the monitoring sites has attracted less attention in the environmental sciences. Russo and Jury (1987) examined the effects of spacing on estimates of the three statistical characteristics based on Monte Carlo simulations with synthetically generated random fields. They found that large spacings will lead to overestimates (i.e. positive biases) of the integral scale and large random errors. The bias was small as long as the spacing was smaller than half the correlation length. This can be related to the Nyquist frequency in time series analysis which is the highest frequency of a signal that can be resolved by a given sampling rate (Nyquist, 1924). In a somewhat related study, Beckie (1996) analysed sampling scales associated with the resolution of a single groundwater field test and used filter functions to estimate how the unresolved small scale variability of, say, hydraulic conductivity, changes as a function of the spacing and the support of the measurements. He found that increasing spacing increases the unresolved small scale variability, while increasing support decreases it. This is because also the overall variability decreases as an effect of increasing support.

The effect of the extent of the data on the integral scale of hydraulic conductivity has been discussed by Gelhar (1993). Figure 6.5 in Gelhar (1993) suggests that the spatial scale over which hydraulic conductivity is correlated (i.e. the integral scale) tends to be on the order of 10% of the extent of the domain, irrespective of the type of underlying aquifer. This is disturbing as it would imply that the integral scales estimated from the data are mainly controlled by the sampling setup rather than by the distribution of the variable one is meaning to capture. The findings of Gelhar (1993) have been corroborated by a number of other studies such as Blöschl (1999) for the case of snow cover patterns. Di Federico and Neuman (1997) and Cintoli *et al.* (2004) suggested that the concept of truncated power variograms can be used to explain these sampling effects.

One of the important concepts common to the analysis of sampling scale effects is that sampling involves some sort of filtering, and observed properties may hence be heavily biased by the sampling scales (Cushman, 1984, 1987). While a number of aspects of sampling scale effects have been examined in the literature in various context, to our knowledge, a coherent analysis for landscapes (i.e. two dimensional fields) is still lacking. This is the purpose of this paper. Specifically, the aim of this paper is to examine what is the bias and the random error (or uncertainty) introduced by the sampling scale triplet (spacing, extent and support) into estimates of the mean, the spatial variance and the spatial integral scale of a variable in a landscape. The paper builds on an initial analysis of Western and Blöschl (1999) who used real soil moisture data to examine sampling scale effects. In contrast to Western and Blöschl (1999), however, in this paper we use a large number of synthetically generated random fields from which we take samples at various sampling scales. This allows us to analyse the biases and random errors of the mean, the variance and the integral scale in a more controlled manner than is possible with real data.

2. Method

2.1. GENERATING RANDOM FIELDS

We generated a large number of two-dimensional random fields and sampled from these fields with a spatial arrangement of the samples conforming to a specified scale triplet of spacing, extent and support. The random field represents the spatial distribution of the variable of interest and each realisation is one possible outcome of the sampling. From the samples we calculated the mean, spatial variance and integral scale, and examined how similar these estimates were to the mean, spatial variance and integral scale of the underlying population of the random field. The mean, spatial variance and integral scale of the underlying population are considered to be the true values and this is what one is intending to estimate to the highest possible accuracy. Figure 2 illustrates schematically the procedure. The top charts show the underlying true distribution and the bottom charts show, schematically, the sampled values for large spacings, small extents and large supports.

The main assumptions in generating the random fields were (a) that the fields were stationary, i.e., the statistical characteristics of the population of the random field (such as the mean and the variance) were assumed not to change with spatial location. We also assumed (b) that the univariate distribution of the random field conforms to a normal distribution with population mean μ and population variance σ^2. Finally (c) we assumed that the spatial correlation structure of the population of the random field can be represented by an exponential variogram:

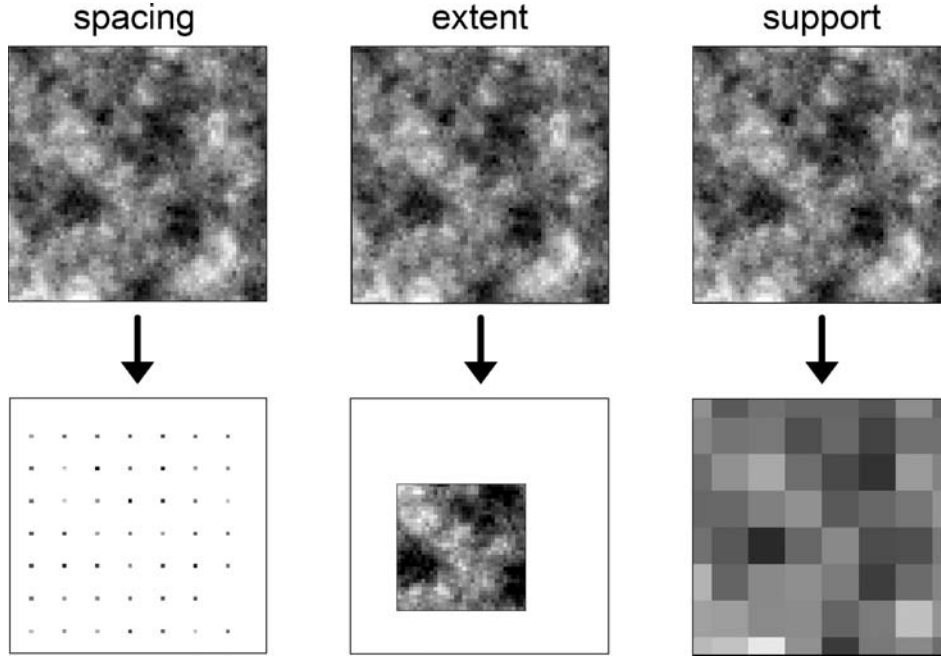$$\gamma(h) = \sigma^2(1 - \exp(-(h/\lambda))) \tag{2}$$

Figure 2. Schematic of sampling scale effects. The top row shows the underlying true distribution and the bottom row is the actual information sampled. Left: With large spacings, the small scale variability will not be sampled. Centre: With small extents, the large scale variability will not be sampled. Right: With large supports the samples will appear much smoother than the true spatial distribution.

where λ is the correlation length and h is the distance between two points in the random field (Journel and Huijbregts, 1978). The variogram describes the average variance between two points separated by distance h. The correlation length is a measure of the average distance over which a variable is correlated in space. For the variogram in Equation (2), the integral scale happens to be identical with the correlation length while for different variograms the integral scale can be smaller or larger than the correlation length, depending on the shape of the variogram. The exponential variogram is consistent with a first-order autoregressive or Markov process (Webster and Oliver, 2001) and is hence the simplest assumption one can make about the spatial variability of a random field. With these prescribed statistical characteristics we generated two dimensional random fields using the Turning Band Method (Mantoglou and Wilson, 1981, 1982) on a square grid of 1024×1024 points with grid size of Δx. For each combination of the sampling scale triplet we generated 1000 realisations.

2.2. SAMPLING

Throughout this paper we use dimensionless sampling scales L_S^*, L_E^* and L_A^*, i.e. the spacing, extent and support scaled by the true correlation length λ of the population

of the random field:

$$L_S^* = L_S/\lambda$$
$$L_E^* = L_E/\lambda \qquad (3)$$
$$L_A^* = L_A/\lambda$$

In a first analysis we varied spacing and extent jointly. From each of the 1000 realisations for each extent, we took 16, 100 and 1024 samples. For a fixed number of samples the spacing increases linearly with the extent, similar to Equation (1).

$$L_S^* = L_E^*/\sqrt{N} \qquad (4)$$

where N is the number of samples. For these analyses we set the support $L_A^* = 0$, i.e. we took point samples from the grid.

In a second analysis we varied the support. The extent was fixed as $L_E^* = 10$, with spacing being a function of the number of samples according to Equation (4). The support was changed by arithmetically averaging the point values over $(L_A/\Delta x)^2$ points. Again, we took 16,100 and 1024 samples and aggregated an increasing number of points from the generated field for each sample to represent the effect of increasing support. We assumed square support areas. In both analyses we used two monitoring schemes, sampling on a regular square grid and random sampling.

In all instances with $L_A > L_S$, measurements will be partly overlapping and the measurements closest to the border of the domain will also include values from outside what is defined as the extent of the domain.

We assumed that all sample values were error free, i.e. we only analysed the biases and random errors (uncertainties) introduced by a less than exhaustive sampling of the underlying distribution. In a real world study, instrument errors will introduce additional uncertainty.

2.3. ESTIMATION OF MOMENTS AND INTEGRAL SCALE

From the N samples $z(\mathbf{x}_i)$ at locations $\mathbf{x}_i$ we estimated the sample mean ($\bar{z}$) and spatial sample variance (s^2) as:

$$\bar{z} = \frac{1}{N} \sum_{i=1}^{N} z(\mathbf{x}_i) \qquad (5)$$

and

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (z(\mathbf{x}_i) - \bar{z})^2 \qquad (6)$$

The sample variogram we estimated by the traditional estimator of Matheron (1965):

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})]^2 \qquad (7)$$

where $h = |\mathbf{h}|$ is the spatial lag between two points. The summation over the number of pairs $n(h)$ is within bins which we chose at logarithmical intervals. The integral scale of the population is defined as (Taylor, 1921):

$$J = \int_0^\infty \left[1 - \frac{\gamma(h)}{\sigma^2} \right] dh \qquad (8)$$

The sample integral scale we estimated from the sample variogram as:

$$\hat{J} = \sum_{i=1}^{N_b} \left[1 - \frac{\hat{\gamma}(h_{i-1}) + \hat{\gamma}(h_i)}{2s^2} \right] \Delta h_i \qquad (9)$$

where N_b is the bin number where $\bar{\gamma}(h_{N_b}) = s^2$. In other words, this is the bin number for which the sample variogram first intersects a horizontal line $\bar{\gamma}(h) = s^2$. If the variogram reached the sample variance between two bins, the exact intersection point was found by linear interpolation. We assumed $\hat{\gamma}(h_0) = 0$ and $\Delta h_i = h_i - h_{i-1}$.

From each of the 1000 realisations for a given combination of the scale triplet, we estimated the statistical characteristics by Equations (5), (6) and (9). From the set of estimates we estimated the ensemble mean, i.e. the mean value of the 1000 estimates for each of the three statistical characteristics. We also estimated the variance around the ensemble mean. The variance around the ensemble mean represents the random error or uncertainty of the estimates of Equations (5), (6) and (9) one encounters if only one sample set (i.e. one replica) is available.

We plotted the estimates in Equations (5), (6) and (9) in nondimensional form scaled by the (true) population characteristics:

$$\bar{z}^* = \frac{\bar{z} - \mu}{\sigma} \qquad (10)$$

$$s^{2*} = \frac{s^2}{\sigma^2} \qquad (11)$$

$$\hat{J}^* = \frac{\hat{J}}{\lambda} \qquad (12)$$

If the ensemble mean of the nondimensional mean (Equation (10)) is zero, the estimates of the mean are unbiased. If the ensemble mean of the nondimensional variance (Equation (11)) is unity, the estimates of the variance are unbiased. Similarly, if the ensemble mean of the nondimensional integral scale (Equation (12)) is

unity, the estimates of the integral scale are unbiased. If the variance (or standard deviation) of the sample estimates around the ensemble mean are small for any of the statistical characteristics, their uncertainty is small.

In addition, we considered an alternative case, we term "single realisation case", in which we normalised the estimates differently. For each realisation independently we compared the estimated sample mean and sample variance with the mean and the variance of the entire random field of the same realisation. This is the case where we are interested in the statistical characteristics of one realisation, not the population characteristics. The equations corresponding to Equations (10-12) but for the single realisation case (denoted by subscript R) are then:

$$\bar{z}_R^* = \frac{\bar{z} - \bar{z}_{\text{dom}}}{\sigma} \tag{13}$$

$$s_R^{2*} = \frac{s^2}{s_{\text{dom}}^2} \tag{14}$$

$$\hat{J}_R^* = \frac{\hat{J}}{\hat{J}_{\text{dom}}} \tag{15}$$

where the subscript dom refers to estimates based on exhaustive sampling over the domain of size A_{dom} (Equation 1).

An application of this case is the estimation of the spatial average, variance and integral scale of precipitation for a certain event, or the concentration of a nutrient in a particular area at one point in time. For each realisation we subtracted the realisation mean from the estimated mean. The average of these deviations over the ensemble (according to Equation 5) will show if the estimates are biased compared to the realisation mean, while the variance of these deviations (according to Equation 6) will give the uncertainty of such an estimate. A similar procedure was followed for variance, except that we divided the estimated variance by the realisation variance to normalise it. The integral scale estimated from the samples we normalised by the realisation integral scale, i.e. the integral scale estimated from all points of the generated field of the same realisation.

It should be noted that, for a given sampling scale triplet, the choice of estimation methods for the statistical characteristics may affect the results to some degree. We have chosen one of the more commonly used methods. A comparison of methods is beyond the scope of this paper and some guidance on the performance of different methods is given in Cressie (1991).

2.4. ANALYTICAL EXPRESSIONS FOR BIAS AND RANDOM ERROR

For comparison with the sampling analyses from the random fields we present analytical expressions for the bias of the mean, spatial variance and integral scale as a function of the scale triplet taken from Western and Blöschl (1999). In their derivation they made a number of simplifying assumptions which can be tested

here. Among other things they suggested that it is possible to superpose the effects of the spacing, extent and support on the estimation biases. The equations used are summarised in Appendix A. Expressions for the random error or uncertainty of the estimates are given in Appendix B for the mean as a function of the spacing and extent as well as for the variance for statistically independent samples.

3. Results

3.1. ESTIMATION OF THE POPULATION CHARACTERISTICS

3.1.1. *Spacing*

Figure 3 shows the results for the case where spacing is large relative to the underlying correlation length λ. The monitoring scheme is gridded sampling and three cases of 16, 100, and 1024 samples are considered. The charts in the top row of Figure 3 indicate that large spacings do no introduce biases into the mean. For large
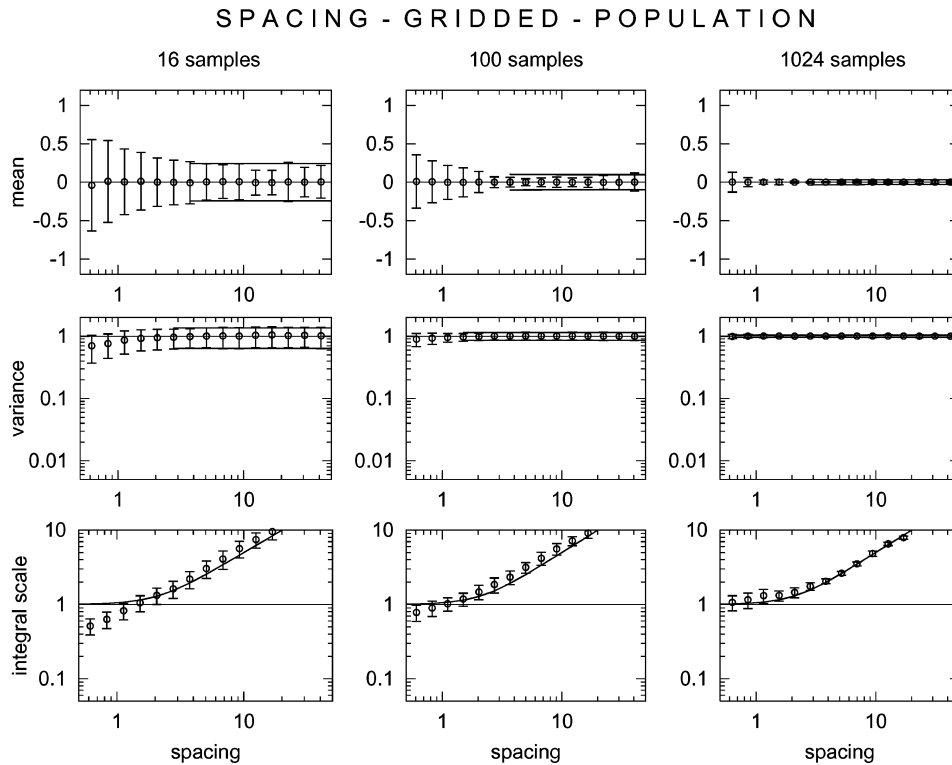

Figure 3. Effect of large spacings L_S^* on the sample mean $\bar{z}^*$, spatial variance s^{2*} and integral scale $\hat{J}^*$ for gridded sampling. The circles show the ensemble mean and the error bars represent the standard deviation around the ensemble mean. The lines in the charts show Equations (A20), (A21) and (A15).

spacings relative to the underlying correlation length λ, $L_S^* > 3$, the uncertainty of the mean (shown as error bars in the figures) does not change with spacing. This is because the samples are practically uncorrelated. The expression in Equation A20 for uncorrelated samples provides a close match to the uncertainties of the sampling analysis as shown in Figure 3. For small spacings, the uncertainties increase with decreasing spacings which is, however, related to small extents and will be discussed in Section 3.1.2. The uncertainty of the mean decreases with increasing number of samples in all cases as would be expected.

The estimates of the spatial variance (Figure 3, second row) are unbiased at large spacings, similar to the estimates of the mean and, again, the expression in Equation (A21) provides a close match. The sampling properties of uncorrelated random variables are very well defined as would be expected.

The estimates of the integral scale (Figure 3, bottom row) are biased as a function of spacing and are increasingly overestimated with increasing spacings. For 16 samples, there is only a very short range of spacings ($L_S^* \approx 2$) where the integral scales are unbiased. For a larger number of samples this range increases. The uncertainties associated with the integral scales (error bars in Figure 3) are relatively small which indicates that the integral scale can be estimated with good confidence but the biases are large. The expression in Equation (A15) closely follows the results of the sampling analysis which indicates that the scale effects of the spacing can be well predicted.

Figure 4 shows the same information as Figure 3, but here the monitoring scheme was random rather than gridded. The results for the mean and spatial variance are similar to the gridded scheme but those for the integral scales are different. The uncertainties of the estimates are considerably larger than for gridded sampling. On the other hand the biases are much smaller. For a large number of samples ($N = 1024$) the estimates of the integral scale are almost unbiased over a wide range of spacings. As indicated in Figure 4 (bottom row), Equation (A15) overestimates the biases. This is because Equation (A15) applies to gridded rather than random sampling.

3.1.2. *Extent*

Figure 5 shows the results for the case where the extent is small relative to the underlying correlation length λ. For comparison we have extended the graphs to large extents which imply large spacings for a given number of samples. The top row in Figure 5 shows that the mean is unbiased. The uncertainty, however, increases drastically with decreasing extent. In the limit of very small extents the uncertainty is unity. This is the standard deviation of the underlying field which means no information on the underlying distribution is obtained from such a monitoring scheme. On the other hand, the case for large extents shown in Figure 5 top row implies large spacings (e.g., with 100 samples, an extent of 100 implies a spacing of 10). It is interesting that the uncertainties introduced by small extents (when only a small fraction of the landscape is sampled) are much larger than those introduced
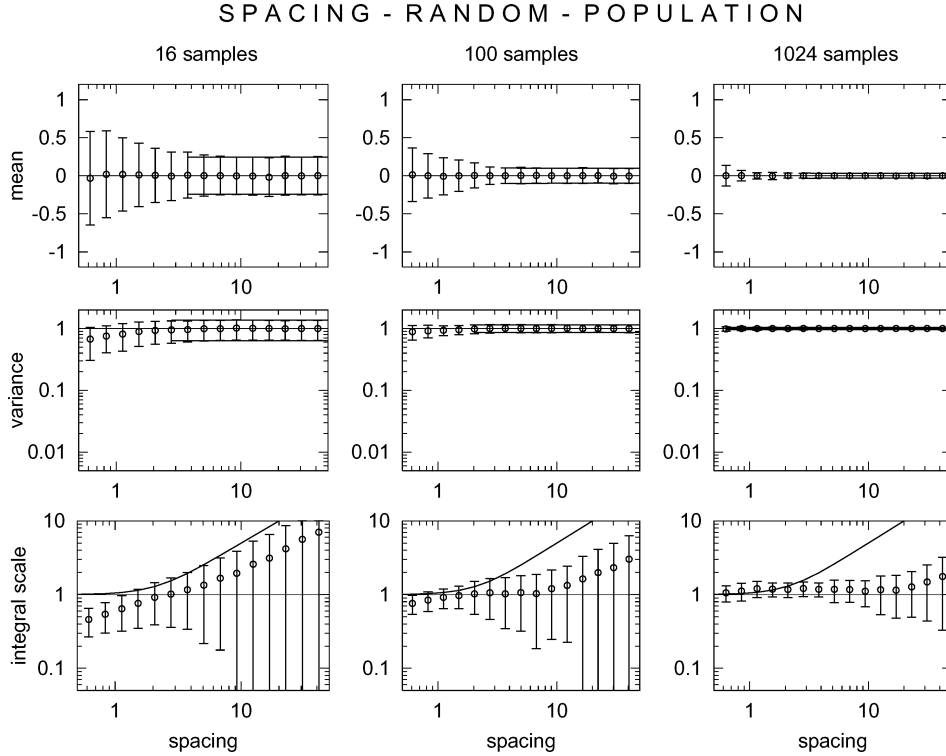
Figure 4. Same as Figure 3 but for the case of random sampling.

by large spacings (where in the limit the data are uncorrelated). For the case of small extents all samples are heavily correlated and very little is learned about the variability of the population. This behaviour is also borne out by the expressions in Equation (A18).

The second row in Figure 5 shows that the variance is significantly biased for small extents. Once $L_E^* < 10$ the variance is underestimated and this bias does not depend on the number of samples. Equation (A9) gives a good prediction of the biases although at very small extents Equation (A9) slightly underestimates the biases.

The bottom row in Figure 5 indicates that the integral scale is heavily biased for small extents. The underestimation can be almost two order of magnitudes. It is interesting that the underestimation occurs for $L_E^* < 10$. This means that the monitoring domain needs to be at least 10 times the underlying correlation length for not introducing extent biases. For larger extents there is some overestimation but this is a spacing effect as discussed above. For all cases, the uncertainty of the estimated integral scale is small in spite of the very large biases. The lines shown in the bottom row in Figure 5 are a product of Equations (A12) and (A15) to account for extent and spacing effects simultaneously. They closely represent the sampling results.

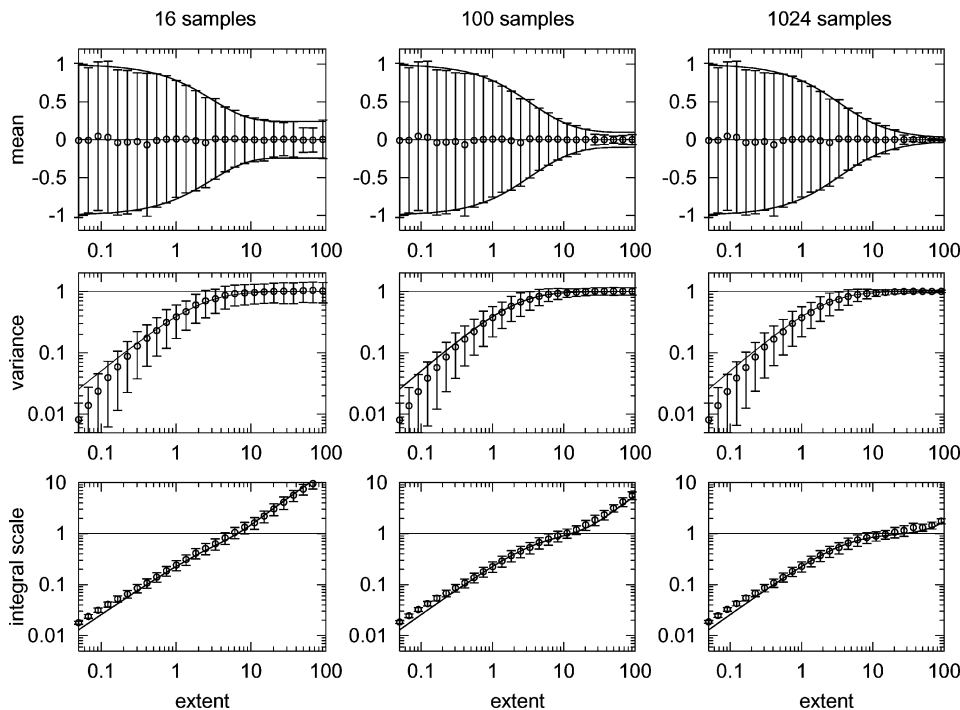E X T E N T - G R I D D E D - P O P U L A T I O N

Figure 5. Effect of small extents L_E^* on the sample mean $\bar{z}^*$, spatial variance s^{2*} and integral scale $\hat{J}^*$ for gridded sampling. The circles show the ensemble mean and the error bars represent the standard deviation around the ensemble mean. Lines in the charts show Equations (A18), (A21),(A9), and the product of Equations A12 and A15.

For the case of small extents, the biases and uncertainties of the estimators are independent of the spatial arrangements of the samples within the domain. Random sampling hence produces very similar results as those in Figure 5 so they are not shown here.

3.1.3. *Support*

Figure 6 shows the results for the case where the support is large relative to the underlying correlation length λ. The first row in Figure 6 indicates that the estimates of the mean are unbiased. The uncertainties of the mean do not change with the size of the support. This may seem counter intuitive as one would perhaps expect the uncertainty of the mean to decrease with increasing support. The reason for the uncertainty not to change with support is that this case examines estimation of the mean of the population rather than that of a single realisation (shown later). Most of the uncertainty comes from the variability between different realisations (what Journel and Huijbregts, 1978, p. 192, terms fluctuation variance). The uncertainty due to not exhaustively sampling a single realisation is however much smaller. A comparison of Figure 6, top row, suggests that this is indeed the case as
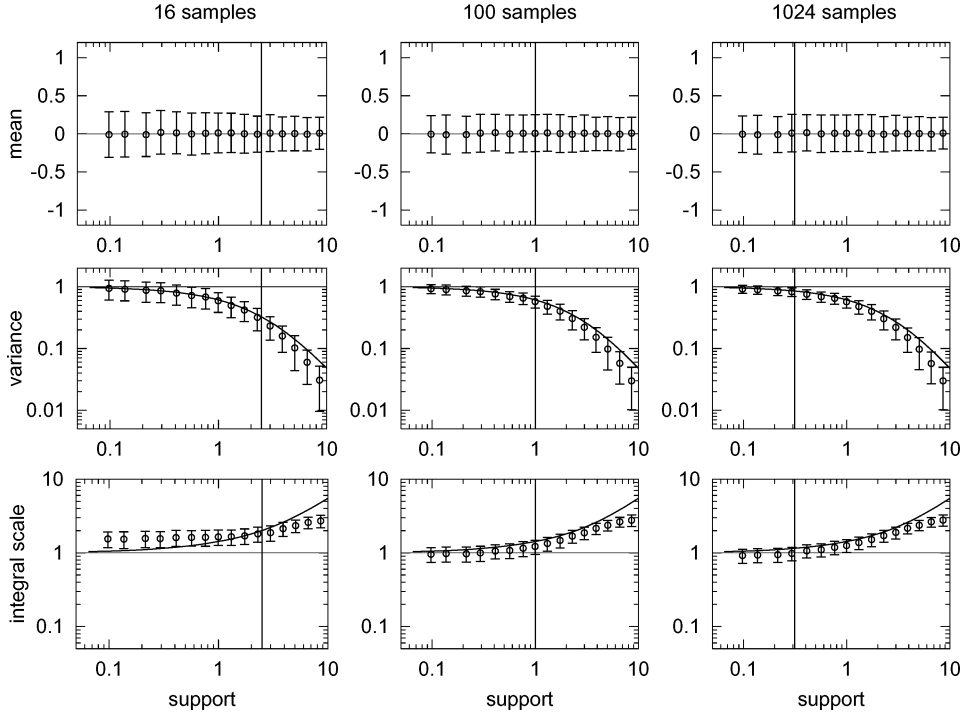
Figure 6. Effect of large supports L_A^* on the sample mean $\bar{z}^*$, spatial variance s^{2^*} and integral scale $\hat{J}^*$ for gridded sampling. The extent is $L_E^* = 10$ in all cases. The circles show the ensemble mean and the error bars represent the standard deviation around the ensemble mean. Lines in the charts show Equations (A2) and (A8). The vertical lines show $L_A^* = L_S^*$, i.e. the support where the domain is fully covered by the samples.

the uncertainty does not decrease with increasing number of samples (and hence decreasing spacing). The magnitude of the uncertainty is controlled by the extent of the data chosen for these cases ($L_E^* = 10$, see Figure 5).

The variances (Figure 6, second row) are, however, significantly biased. They are underestimated once the support exceeds about one third of the underlying correlation length. The vertical lines in Figure 6 show the support values where $L_A^* = L_S^*$, i.e. the domain is fully covered by the samples. For larger supports, the samples overlap which introduces particularly large biases. The uncertainties of the variances are not very large and there is a slight tendency for decreasing uncertainty with increasing number of samples. Equations (A2) and (A3) are close to the sampling results although they slightly underestimate the variances for large supports. This is because Equations (A2) and (A3) have been derived for non-overlapping samples.

The third row in Figure 6 shows moderate biases of the integral scale as a result of large supports. Again, once the samples overlap, the integral scales are overestimated. For 16 samples, there is a spacing effect which results in an overestimation

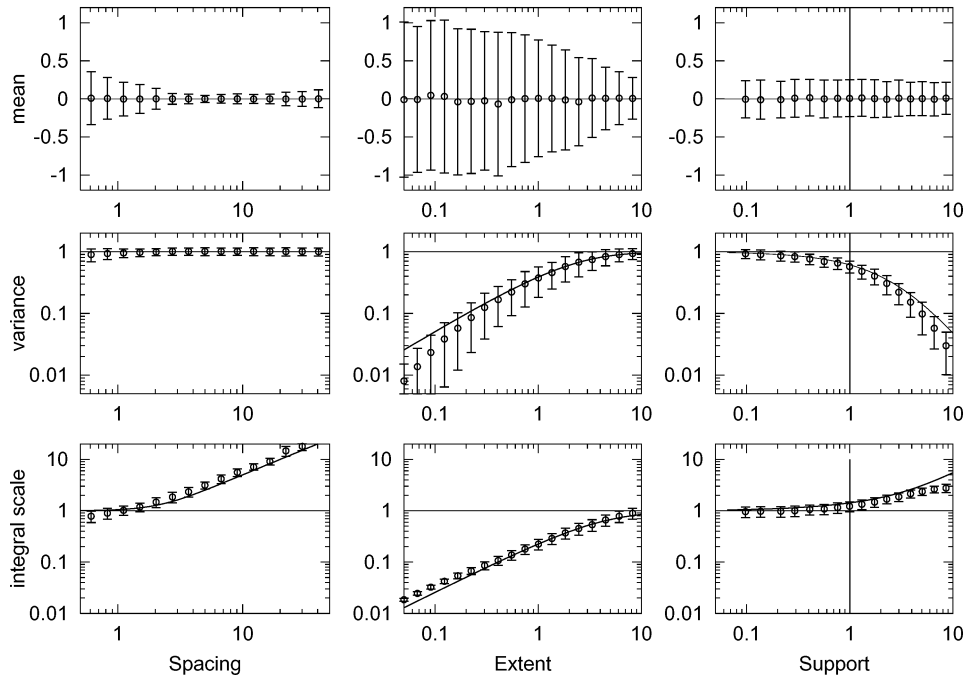S U M M A R Y - G R I D D E D - P O P U L A T I O N

Figure 7. Effect of the sampling scale triplet on the sample mean $\bar{z}^*$, spatial variance s^{2*} and integral scale $\hat{J}^*$ for 100 gridded samples. The circles show the ensemble mean and the error bars represent the standard deviation around the ensemble mean.

for all supports. Equation (A8) does not account for these spacing effects. For large supports, Equation (A8) overestimates the bias compared to the results of the sampling analysis. This is a border effect in the sampling analysis as the largest distance between the centres of the samples will be the size of the domain.

3.1.4. *Summary of Scale Triplet Effects*

For comparison the results for the sampling scale effects are summarised in Figure 7. It is interesting that the uncertainties in the mean are largest for small extents as compared to other scale effects. The biases in the variance are large both for small extents and large supports. The integral scale will always be biased and the bias is closely related to the magnitude of any of the scales of the scale triplet.

3.2. ESTIMATES OF SINGLE REALISATIONS

3.2.1. *Spacing*

Figures 8–11 present the scale effects for the analyses where we are interested in the statistical characteristics of one single realisation (Equations (13) to (15)), not
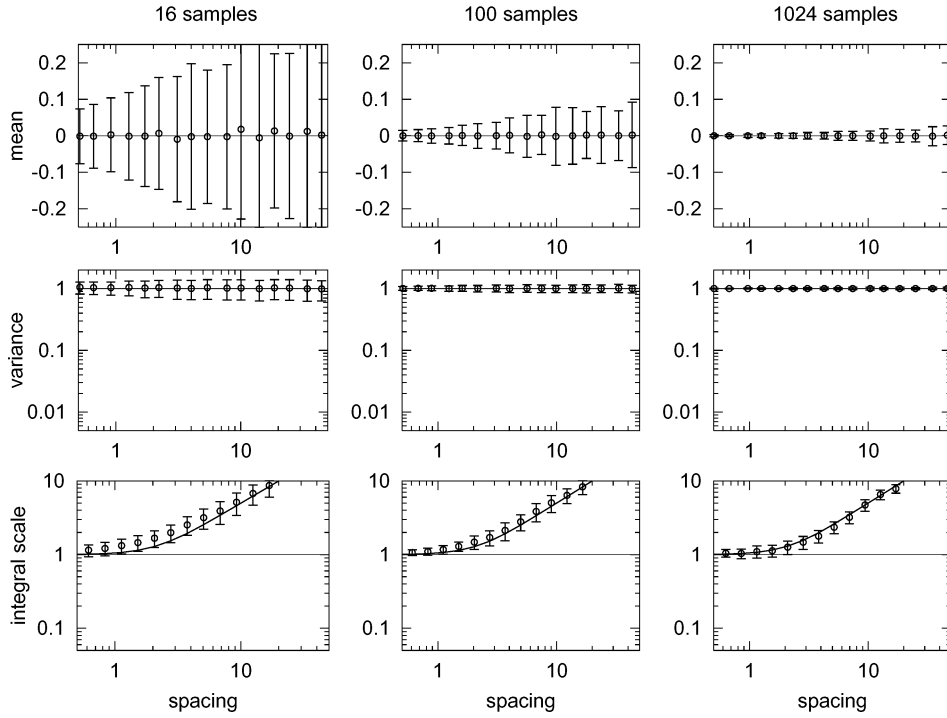
SPACING - GRIDDED - SINGLE REALISATION

Figure 8. Effect of large spacings L_S^* on the sample mean $\bar{z}_R^*$, spatial variance s_R^{2*} and integral scale $\hat{J}_R^*$ for gridded sampling. This is the case where the interest resides in the statistical characteristics of a single realisation, not the population characteristics. The circles show the ensemble mean and the error bars represent the standard deviation around the ensemble mean. Lines in the charts for the integral scale show Equation (A15).

the population characteristics. Figure 8 gives the results for the case where spacing is large relative to the underlying correlation length λ. The mean is unbiased (Figure 8 top row), similarly to the population case in Figure 3. The estimation uncertainty of the mean increases with the spacing which is the exact opposite of the population result in Figure 3, top row. For large spacings (relative to the underlying correlation length) the samples are uncorrelated, so the estimates for the single realisation are no better than the estimates for the population (in both cases 0.25 for 16 samples, for example, as given by Equation (A20)). On the other hand, as the spacing decreases (relative to the underlying correlation length) the correlation of the samples increases. While this increases the estimation uncertainty of the population (Figure 3) it decreases the estimation uncertainty of the realisation (Figure 8). In the limit, if the samples are perfectly correlated, there will be no uncertainty introduced from missing out on the locations between the samples.
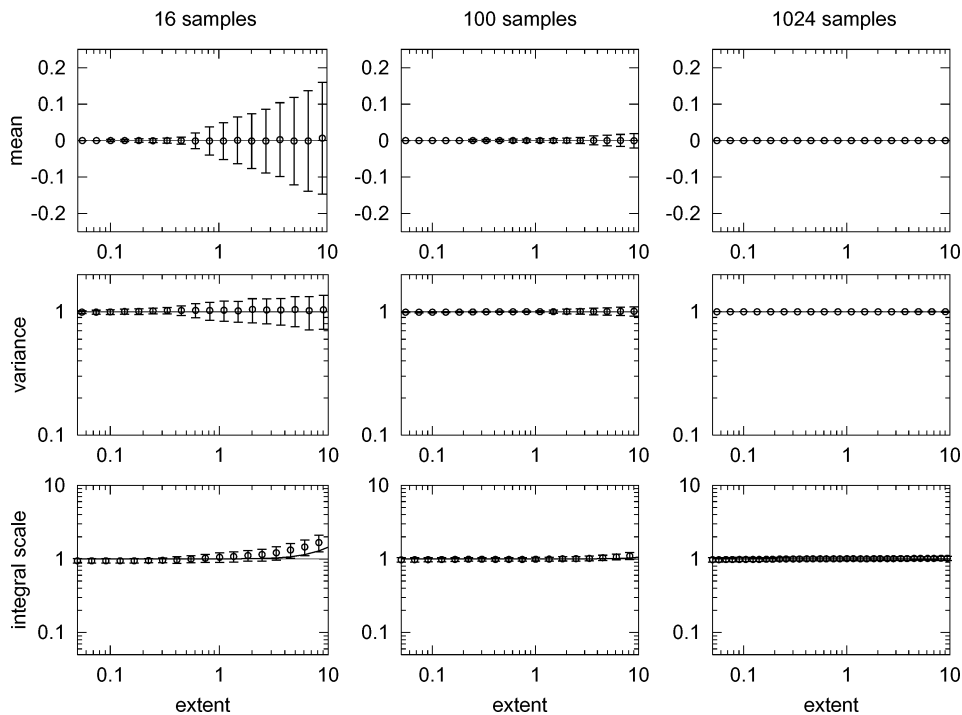
EXTENT - GRIDDED - SINGLE REALISATION

Figure 9. Same as Figure 8 but for the case of small extents L_E^*.

The second row in Figure 8 indicates that the variance is, again, unbiased. For large spacings, the samples are uncorrelated and the results are identical to those in Figure 3.

The third row in Figure 8 shows that the integral scale for the single realisation case is biased in a similar way as in the population case. For large spacings, the integral scale is substantially overestimated. The bias is predicted well by Equation 15.

3.2.2. *Extent*

Figure 9 shows the results for the case where the extent is small relative to the underlying correlation length λ, again in the single realisation mode. Small extents (relative to the underlying correlation length) imply that the samples are highly correlated or, equivalently, that the variable varies very smoothly over the domain sampled. Because of this, the difference between the sampled values and the entire underlying random field (considered as the truth here) is small in all instances. If this is the case, all statistical characteristics will be essentially unbiased (as compared to each realisation) and the uncertainty is always very small which is indeed the case in Figure 9 for small extents. In other words, in the limit of very small extents (relative to the underlying correlation length), everything is known
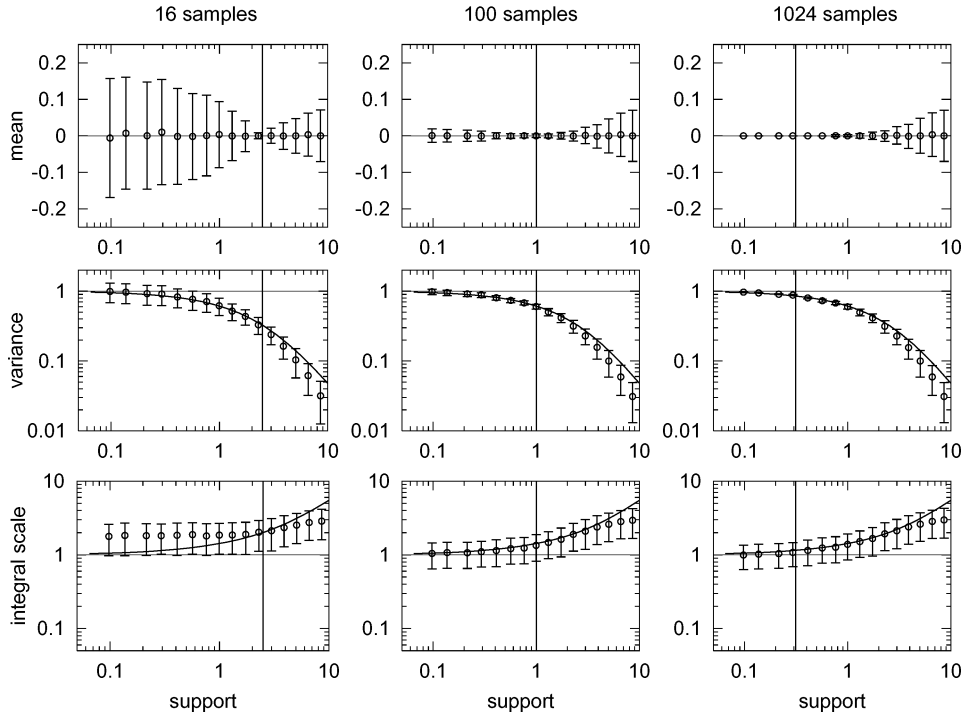
SUPPORT - GRIDDED - SINGLE REALISATION

Figure 10. Same as Figure 8 but for the case of large supports L_A^*. The extent is $L_E^* = 10$ in all cases. Lines in the charts show Equations (A2) and (A8). The vertical lines show $L_A^* = L_S^*$, i.e. the support where the domain is fully covered by the samples.

about the underlying distribution (of the one realisation) from a limited number of samples. This is in stark contrast to Figure 5 where in the limit of very small extents nothing is known about the underlying distribution (of the population) no matter how many samples are taken.

3.2.3. *Support*

Figure 10 shows the results for the case where the support is large relative to the underlying correlation length λ. The mean is unbiased (Figure 10 top row). The uncertainty of the mean decreases with the number of samples. This is a result of the spacing as the extent was held fixed to $L_E^* = 10$ in all cases, which implies smaller spacings for a larger number of samples. There is an interesting dependence of the uncertainty of the mean on the support. For small supports, the uncertainty decreases with increasing support until $L_A^* = L_S^*$. This is because an increasingly larger fraction of the domain of interest will be covered by the samples. For $L_A^* = L_S^*$ (indicated by the vertical line in Figure 10) the domain is fully covered by the samples, so there is no uncertainty involved in the estimation
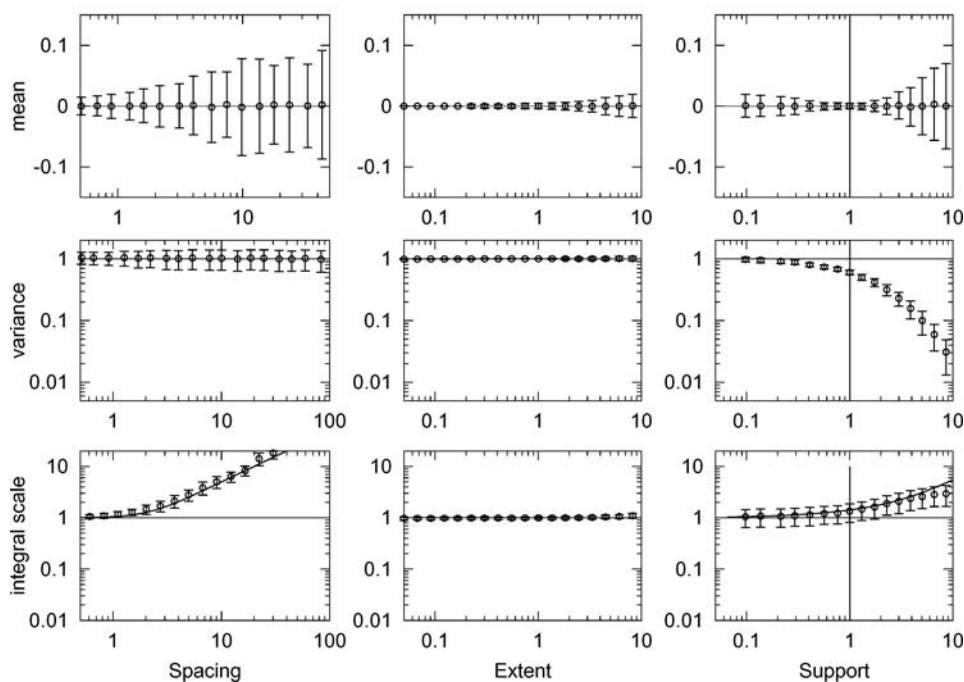
SUMMARY - GRIDDED - SINGLE REALISATION

Figure 11. Effect of the sampling scale triplet on the sample mean $\bar{z}_R^*$, spatial variance s_R^{2*} and integral scale $\hat{J}_R^*$ for 100 gridded samples. This is the case where the interest resides in the statistical characteristics of a single realisation, not the population characteristics. The circles show the ensemble mean and the error bars represent the standard deviation around the ensemble mean.

of the mean. However, as the support increases further, the samples will overlap and the uncertainty starts to increase. This is because some parts of the domain will be more frequently sampled than others and because a part of the support area will lie outside the domain A_{dom}.

The second and third rows in Figure 10 show similar results as their counterparts of the population case in Figure 6. The biases are essentially identical. However, there are minor differences in the uncertainties. The uncertainties of the variance in Figure 10 are slightly smaller than those in Figure 6. This is because of different spacing effects. On the other hand the uncertainties of the integral scale in Figure 10 are slightly larger.

3.2.4. *Summary of Scale Triplet Effects*
For comparison, the results of the sampling scale effects are summarised in Figure 11 for the analyses where we are interested in the statistical characteristics of one single realisation. The main difference to the population case in Figure 7 is that small extents neither introduce biases nor uncertainties. This is simply because in

the single realisation mode we are not interested in what happens outside the domain sampled while in the population mode we are indeed interested in the population, including the distribution outside the domain sampled. The sampling effect of large spacings on the mean is related. For large spacings, the samples are uncorrelated, so they yield increasingly smaller uncertainties in the populations case (Figure 7) but increasingly larger uncertainties if one is interested in one single realisation only (Figure 11). The spacing and support effects on both the variance and the integral scale of the single realisation case are similar to those of the population case. The biases in the variance are large for large supports. The integral scale will be biased for large spacings and large supports and the bias is closely related to the magnitude of these two scales.

4. Implications for Environmental Monitoring

4.1. Importance of sampling scale effects

The sampling analyses in this paper have shown that there may exist enormous biases in the three statistical characteristics examined here, the mean, the spatial variance and the integral scale, which result from sampling scale effects. Even in the absence of instrument error the biases can be up to two orders of magnitude for the cases considered here. Also, the uncertainty or random error can be very significant.

The findings in this paper are consistent with previous work. The biases are similar to those of Western and Blöschl (1999) of their analysis of soil moisture data in a catchment although this paper analyses the scale effects in a more controlled manner. Also, this paper gives more precise estimates of the uncertainties involved as the synthetic sampling analysis allowed multiple realisations to be drawn. Similar to Russo and Jury (1987) the results of this paper suggest that large spacings, relative to the underlying correlation length, tend to overestimate the integral scale. Similar to Gelhar (1993) the results found here suggest that small extents of the domain result in overestimates of the integral scale. The biases in the variance due to large supports are consistent with those given in the geostatistical literature. It is interesting that the biases of the integral scale are closely related to the magnitude of any of the components of the scale triplet where in most cases large sampling scales result in an overestimation of the integral scales. Intuitively, this is because of dimensional reasons as both the integral scale and the sampling scale are related to units of length. In most cases, the expressions given in Appendix A predict the biases estimated by the sampling analysis well. In examining these sampling biases, however, there are a lot of subtleties involved.

For typical monitoring scales, the results of this paper suggest that usually there is bias due to more than one of the components of the scale triplet. Each of the components produces a scale effect of its own which, for some statistical

characteristics, are similar (such as the integral scale) but for others are different, and the resulting bias is the combination of all three effects. This mixing of the effects of spacing, extent and support is not always obvious. Qi and Wu (1996), for example, examined the effect of support on three spatial autocorrelation indices, i.e., the Moran Coefficient, the Geary Ratio and the Cliff-Ord statistic based on topographic and biomass data of the Malaysian Peninsular. They observed an overall decline in spatial autocorrelation with increasing support but this was more likely an effect of the spacing of adjacent units increasing at the same time. In a number of monitoring systems, such as remote sensing data, changing (or selecting) grid size usually involves a change of both spacing and support, so the grid size effect on the statistical characteristics is a combination of spacing and support effects. The comparison of the synthetic resampling analysis with the analytical results suggested that it is indeed possible to superpose the effects of spacing, extent and support to good accuracy (e.g. Figure 5) to get an overall assessment of the total magnitude of the biases involved.

The results in this paper showed that the sampling scale effects can be very important but it is not only the sampling scale triplet that controls the biases and uncertainties it is also whether one is interested in population statistics or the statistics of one single realisation. The former case is where one is interested in the mean value of, say, concentration over a very large area, i.e. the population mean. In this case it is the small extent that introduces the largest uncertainties into estimates of the mean. For the case of small extents all samples are heavily correlated and very little is learned about the variability of the population. Gelhar (1993, p. 40) and Priestley (1981, p. 319) describe how the number of effective independent samples will be reduced when the samples are correlated. The uncertainties decrease with increasing spacing as the samples become less correlated and hence contain more information about the population. This means that the estimates are quite accurate inside the domain but if one extrapolates further away (in the case of small extents) the results will be highly uncertain as the data represent only a small fraction of the entire population. On the other hand, the second case, termed single realisation case in this paper, is where one is interested in the characteristics (e.g. the mean) of, say, precipitation for a certain event, or the concentration of a nutrient in a particular area at one point in time where the extent is the same as the size of the domain of interest. In this case very little uncertainty is introduced by extent effects and the spacing effect on the mean is the exact opposite of that in the population case. The uncertainties increase with increasing spacing as the underlying pattern becomes less smooth and hence more erratic. Because of the magnitude of the sampling scale effects there are important ramifications for sampling design and data analysis.

4.2. IMPLICATIONS FOR SAMPLING DESIGN

The most direct implication probably is the importance of adjusting the sampling design to the scale of the underlying processes one is intending to capture. Stommel

(1963) compares the process of sampling to a fish net and notes that a single net does not catch fish of all sizes. Quite obviously, if one is interested in small scale processes the sampling scales need to be adjusted to smaller scales, and the opposite is true for large scale processes. Informally, the monitoring network is often adjusted to the scales of the variables one is interested in e.g. Hatcher *et al.* (1987), and Seyfried and Wilcox 1995) but it is also possible to formalise this. The variable to gauge the scale of the underlying (true) variability used here is the correlation length. The correlation length is closely related to the integral scale. For an exponential variogram as has been used here, the integral scale and the correlation length are identical. For different variograms they differ but are usually of the same order of magnitude. In this paper, it was possible to normalise the scale effects by the correlation length, i.e. all scale effects were a function of the scale of each component of the scale triplet (spacing, extent and support) relative to the underlying correlation length. If one uses the correlation length as the main measure of the spatial variability of the underlying processes one has to design the spatial arrangement of the monitoring network relative to the correlation length. A sampling design that is commensurate with the scale of the underlying variability would then involve the following questions:

– What are the processes that are to be monitored?
– What are the spatial scales of these processes (measured by the integral scale)?
– What are the spacing, extent and support of the monitoring network that give small biases and uncertainties for the statistical characteristics one is interested in, given the constraints on costs and hence number of samples?

As a rule of thumb, the extent should be significantly larger than the correlation length (for the population case), the spacing should be smaller than the correlation length and the choice of the support depends on the statistic one is interested in. Large supports are an advantage for the single realisation case but produce significant biases in the population case.

While these steps appear logical, to some degree, this is a Catch-22 situation. One needs to know the integral scale to estimate the correlation length for the sampling design but would need the sampling results to estimate the integral scale. To break the loop one can, however, use the integral scales from other studies one expects to exhibit similar spatial variabilities, particularly, studies where a large number of samples are collected and hence are spatially representative (e.g. Skøien *et al.*, 2003).

In the planning of the monitoring network it is not only the average spacing (or number of sampling points over a given area) that is important but also the spatial sampling arrangement (see, e.g., Thompson, 2002). Depending on whether grid sampling or random sampling schemes are selected, the distribution of distances between pairs of points is different. Random sampling includes distances significantly smaller than the (average) spacing, so the small scale variability can be

better resolved. The biases associated with larger spacings in estimating the integral scale are hence smaller (Figure 4). However, this is at the cost of larger uncertainties. Nested sampling (Webster and Oliver, 2001), as an alternative, allows for still smaller distances between points, so one would expect still smaller biases, but it hinges on the assumption that the variability is spatially stationary, i.e., the small area that is sampled in more detail is representative of a larger area that is sampled more sparsely. Bellehumeur and Legendre (1998) discuss the issues of stationarity in nested sampling and Stenger *et al*. (2002) provide an example of the impacts of non-stationarity with nested sampling. From a statistical perspective, stationarity can be tested for but the issue is not a purely statistical one. Ultimately it is the environmental processes one wants to capture and they need to be representative for nested sampling to make sense.

A similar issue of representativeness applies to the case of small extents. Small extents cause large uncertainties in all statistical characteristics examined here. If one seeks to increase the generality of the results of a study, one therefore needs to increase the extent as pointed out by Hewitt *et al*. (1998). On the other hand, it is not wise to increase the extent beyond what can be assumed to be the same process. Haugen (1978) describes how the certainty of estimated moments of turbulence motion increases with the temporal extent of the measurements but will be affected by the diurnal cycle if the extent is too long. Wiens (1989) examined the temporal and spatial extents of measurements at the same time, and argued that it is easy to be overly confident in the sampling results as lack of representativeness can cause serious biases. In the case the interest resides in a single realisation (nutrients over one area at one point in time, for example) sampling design is usually such that the sampling spans the entire domain, i.e. the extent is equal to the size of the domain of interest which, in the light of the results of this paper, obviously is a prudent choice.

In selecting the support of a monitoring scheme, larger supports may be an advantage if one is interested in estimating the mean for the single realisation case. The support depends on the choice of instrument. Remote sensing sensors usually have rather large supports while in situ samples are rather small. There are, however, a large number of other factors that need to be considered when selecting an instrument, both technical (such as accuracy) and logistical, so the sampling support may not be the main consideration in choosing an instrument. A similar effect as large supports can be obtained by taking multiple point measurements over a small area and then averaging them which filters out fine-scale heterogeneity one may not be interested in Root and Schneider (1995). This may also filter out some of the measurement error, which however is beyond the scope of this paper.

The number of samples is, of course, the most important issue in spatial sampling design. The costs associated with the monitoring usually are strongly positively correlated with the number of samples and the uncertainties of the estimates are strongly negatively correlated so there are tradeoffs involved (see, e.g., Lopes, 1996; Faurès *et al*., 1995). The results of this paper suggest that for the realisation mean, spatial variance and integral scale, an increasing number of samples will in most

cases reduce the estimation uncertainty. The exception is the case of small extents, relative to the underlying variability, where the biases and uncertainties do not depend on how many samples are taken within the extent of the sampling domain. Uncertainties in extrapolating beyond the sampling domain cannot be compensated by a larger number of samples within the domain.

4.3. IMPLICATIONS FOR DATA ANALYSIS AND INTERPRETATION

As the sampling scales are likely to affect the measurements of a variable, even if the sampling design has been selected commensurate with the underlying variability, it can be difficult to compare measurements from different sources with each other. Theocharopoulos *et al.* (2001), for example, in a comparison of soil pollution sampling guidelines within the European Union pointed out that differences in the guidelines of individual countries caused problems in the objective definition of pollution, in the reproduction of the sampling by different teams, and in defining threshold values throughout the European Union. In a similar vein, Western *et al.* (1998) reviewed field studies of soil moisture around the world and concluded that the spatial statistical characteristics of soil moisture found in these studies were more closely related to the sampling scales than to the environmental conditions of the particular catchments.

It is, however, possible to account for these differences, and correct for sampling scale biases to some extent as pointed out in this paper. To illustrate the case, let us consider a hypothetical example where bias is introduced into estimates of the integral scale by large sample spacings. Let us assume that 100 samples have been collected from a field of 1000 m by 1000 m extent, i.e. the average spacing is 100 m, and that the integral scale was estimated from the data as 80 m. The equation for the expected integral scale as a function of spacing (Equation (A15)) cannot be analytically inverted, but we can find an estimate of the true correlation length by trial and error. If we assume a true correlation length of 55 m the dimensionless spacing is 1.8. Equation (A15) then gives a dimensionless integral scale of 1.5 (i.e. a bias of 50%). This is consistent with its overestimation due to large spacings (80 rather than 55 m) so the initial guess was correct. The biases can be back-calculated in a similar way for the other sampling scales. Let us consider an example of biases due to large supports where hydraulic conductivity has been sampled by groundwater pumping tests (support of 100 m). The samples gave a variance of the logarithms of hydraulic conductivity of $\sigma^2_{\ln K} = 1.2$ and an integral scale of $J = 88$ m but one is interested in the variance and the integral scale of the point process (i.e. if conductivity had been measured by laboratory core tests with a support of 10 cm). Again by iteration, we can assume that the (true) point integral scale is 40 m. The dimensionless support then is 2.5. Equation (A8) gives a dimensionless integral scale of 2.2 (i.e. a bias of 120%) which is consistent with the observed integral scale of 88 m. From Equation (A2) the dimensionless variance is 0.32, so the point variance is 3.75 rather than 1.2 what was observed for the large support.

While this back-calculation is useful to assess potential biases, in many cases, the uncertainty with this will be large. This is particularly true of the integral scale if a small number of samples is available. Figure 5, for example, shows that, if only 16 samples are available, the estimated integral scale becomes a function of extent only and is essentially independent of the underlying correlation length. It is on the order of 10 % of the extent of the domain. This is exactly the observation of Gelhar (1993, p. 292) and Blöschl (1999) in comparisons of case studies of aquifer hydraulic conductivity and snow cover variability, respectively, and implies that the integral scale is essentially unobservable if only a small number of samples is available. As the number of samples increases, the integral scale gets better defined. Also, random sampling decreases the bias over grid sampling. It is clear that a large number of samples is a prerequisite for reliably identifying the integral scales if they are to be used in either sampling design or data interpretation.

It should be noted that in this study a number of simplifying assumptions have been made. Most importantly, the random fields were assumed to be stationary with one single scale, the integral scale, representing the spatial variability. While, often, this is a reasonable assumption for a certain scale range, there are also examples where the underlying variability possesses more than one scale or a suite of nested scales (see, e.g., McBratney, 1992 for a soils example). In these cases a power law variogram may be a better choice than an exponential variogram. This is then termed fractal variability. There is a large body of literature on fractal random fields in the environmental sciences (see e.g. Chilès and Delfiner, 1999) and more research needs to be done on the sampling scale effects of this type of processes. Qualitatively, one would assume that the biases for fractal processes are similar to those found here (see, e.g., Blöschl, 1999), but the uncertainties most likely will be different.

Another assumption made here was that the instrument errors were assumed negligible. It is clear that, as the instrument errors increase, so will the uncertainties in the estimates of the statistical characteristics.

We have taken a simple statistical approach here that intends to capture the first order sampling scale effects on the mean, the spatial variance and the integral scale of an environmental variable. There are applications where the interest resides in different characteristics of a variable such as trends or patterns. While for these characteristics the scale effects will be different, the analysis in this paper can give some guidance on assessing their order of magnitude.

Acknowledgments

Appendix A: Bias of Estimates

BIAS DUE TO LARGE SUPPORT

The assumption here is that the variable aggregates linearly, i.e.

$$z_A = \frac{1}{A} \int_A Z(x)ds \tag{A1}$$

where z_A is the average value of a random field Z within an area A where $L_A = \sqrt{A}$ is termed the support in terms of units length. The variance of the aggregated field, $\sigma_{L_A}^2$, then is the difference between the total variance, σ^2, and the variance within the support of the measurement, $\sigma_{WL_A}^2$ (Journel and Huijbregts, 1978):

$$\sigma_{L_A}^2 = \sigma^2 - \sigma_{WL_A}^2 \tag{A2}$$

with

$$\sigma_{WL_A}^2 = \int_0^{R_{\max}} \gamma_T(r) \cdot f_1(r \mid L_A) dr \tag{A3}$$

where γ_T denotes the true variogram. r is the distance between two randomly chosen points in the aggregated region, $R_{\max}$ is the maximum distance in the region and $f_1(r \mid L_A)$ is the probability density function (pdf) of distances r within the region (Rodríguez-Iturbe and Mejía, 1974). For a square region, f_1 has been derived by Ghosh (1951) as:

$$f_1(r) = \frac{4r}{L_{L_A}^4} \phi_1(r) \tag{A4}$$

where

$$\phi_1(r) = \frac{1}{2}\pi L_A^2 - 2L_A r + \frac{1}{2}r^2 \quad \text{for } 0 \le r \le L_A \tag{A5a}$$

$$\phi_1(r) = L_A^2 \left\{ \sin^{-1} \frac{L_A}{r} - \cos^{-1} \frac{L_A}{r} \right\}$$
$$+ 2L_A\sqrt{r^2 - L_A^2} - \frac{r^2 + 2L_A^2}{2} \quad \text{for } L_A \le h \le \sqrt{2}L_A \tag{A5b}$$

Equation (A3) can then be evaluated numerically. The variogram of the averaged process can be found in a similar way (Journel and Huijbregts, 1978):

$$\gamma_{L_A}(h) = \int_0^{R_{\max}} \gamma_T(r) \cdot f_2(r \mid (h, L_A)) \, dr - \sigma_{L_A W}^2 \tag{A6}$$

where f_2 is the pdf of distances of two points, each of them is located randomly in one of two regions. The two regions are separated by a centre-to-centre distance h and have side lengths (support) L_A. For square regions and $h \geq L_A$, f_2 has been derived by Sivapalan (1986) as:

$$f_2(r) = \frac{2r}{L_A^4}\phi_2(r)$$

where

$$\phi_2(r) = L_A\sqrt{r^2 - e^2} - \frac{(r-e)^2}{2} - eL_A \cos^{-1}\left(\frac{e}{r}\right) \quad \text{for } e \leq r\sqrt{L_A^2 + e^2}$$

$$\phi_2(r) = \frac{L_A^2}{2} + e\left(r - \sqrt{r^2 - L_A^2}\right) - eL_A \sin^{-1}\left(\frac{L_A}{r}\right)$$

$$\text{for } \sqrt{L_A^2 + e^2} \leq r \leq h$$

$$\phi_2(r) = L_A g \cos^{-1}\left(\frac{h}{r}\right) - eL_A\left[\sin^{-1}\left(\frac{L_A}{r}\right) - \cos^{-1}\left(\frac{h}{r}\right)\right]$$

$$- e\sqrt{r^2 - L_A^2} - 2L_A\sqrt{r^2 - h^2} + \frac{L_A^2}{2} + h^2 + r^2 - gr \qquad \text{(A7)}$$

$$\text{for } h \leq r \leq \sqrt{L_A^2 + h^2}$$

$$\phi_2(r) = L_A g \sin^{-1}\left(\frac{L_A}{r}\right) - L_A^2 + g\sqrt{r^2 - L_A^2} - gr$$

$$\text{for } \sqrt{L_A^2 + h^2} \leq r \leq g$$

$$\phi_2(r) = L_A g\left[\sin^{-1}\left(\frac{L_A}{r}\right) - \cos^{-1}\left(\frac{g}{r}\right)\right] + g\sqrt{r^2 - L_A^2}$$

$$+ L_A\sqrt{r^2 - g^2} - \frac{L_A^2 + r^2 + g^2}{2} \quad \text{for } g \leq r \leq \sqrt{L_A^2 + g^2}$$

with $e = h - L_A$ and $g = h + L_A$

and it can be evaluated numerically if the regions overlap (i.e. $h < L_A$). The expected integral scale can then be found by numerical integration similar to Equation (6):

$$J_{L_A} = \int_0^\infty \left(1 - \frac{\gamma_{L_A}(h)}{\sigma_{L_A}^2}\right) dh \qquad \text{(A8)}$$

BIAS DUE TO SMALL EXTENT

The variance $\sigma_{L_E}^2$ of a random field within a domain of extent L_E can be found similar to Equation (A3) as:

$$\sigma_{L_E}^2 = \int_0^{R_{\max}} \gamma_T(r) \cdot f_1(r \mid L_E)\,dr \tag{A9}$$

Western and Blöschl (1999) suggested that the expected variogram of a random field with a small extent can be assumed to be similar to the true variogram for lags that are small compared to the extent L_E, and that it will be equal to $\sigma_{L_E}^2$ for lags that are large compared to the extent with a sharp crossover at a certain lag distance B where the true variogram reaches $\sigma_{L_E}^2$, i.e.

$$\gamma_{L_E}(h) = \gamma(h) \quad \text{for } 0 \leq h \leq B \tag{A10a}$$
$$\gamma_{L_E}(h) = \sigma_{L_E}^2 \quad \text{for } h > B \tag{A10b}$$

For an exponential variogram (Equation (2)) this gives:

$$\gamma_T(B) = \sigma_{L_E}^2 \quad \text{and} \quad B = -\lambda \cdot \ln\left(1 - \frac{\sigma_{L_E}^2}{\sigma_T^2}\right) \tag{A11}$$

where λ is the true correlation length of the underlying process. This gives the expectation of the integral scale:

$$J_{L_E} = \frac{\lambda}{\sigma_{L_E}^2} \cdot \left[\sigma_{L_E}^2 + \left(\sigma_T^2 - \sigma_{L_E}^2\right) \cdot \ln\left(1 - \frac{\sigma_{L_E}^2}{\sigma_T^2}\right)\right] \tag{A12}$$

BIAS DUE TO LARGE SPACING

For any large spacing L_S, the expectation of the sample variance $\sigma_{L_S}^2$ will be equal to the true variance:

$$\sigma_{L_S}^2 = \sigma^2 \tag{A13}$$

For lags larger than the average spacing, L_S, Western and Blöschl (1999) suggested that the sample variogram can be approximated by the true variogram and for shorter lags it can be assumed to be linear between the origin $\gamma(0) = 0$ and $\gamma(L_S)$, i.e.

$$\gamma_{L_S}(h) = \frac{h}{L_S}\gamma_T(L_S) \quad \text{for } h \leq L_S \tag{A14a}$$
$$\gamma_{L_S}(h) = \gamma_T(h) \quad \text{for } h > L_S \tag{A14b}$$

The expectation of the integral scale for an exponential variogram then is:

$$J_{L_S} = \lambda \left[\exp\left(\frac{-L_S}{\lambda}\right) + \frac{L_S}{2\lambda} \cdot \left(1 + \exp\left(\frac{-L_S}{\lambda}\right)\right) \right] \tag{A15}$$

which approaches the true correlation length λ for small spacings.

Appendix B: Error Boundaries of Estimates of Mean

For grid sampling, the uncertainty (in terms of the error variance Var) of the estimates of the mean $\overline{z_{L_E}}$ as a function of extent L_E is (Rodríguez-Iturbe and Mejía, 1974):

$$Var(\overline{z_{L_E}}) = \frac{1}{N^2} E\left[\sum_{i=1}^{N} Cov(z_i, z_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} Cov(z_i, z_j) \right]$$

$$Var(\overline{z_{L_E}}) = \frac{\sigma_{L_E}^2}{N^2} \left[N + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho(h_{ij}) \right] \tag{A16}$$

where $\sigma_{L_E}^2$ is the variance of the random field within the region with extent L_E. For brevity, we use here the covariance $Cov = \sigma_{L_E}^2 - \gamma_{L_E}$ and the correlogram $\rho = Cov/\sigma_{L_E}^2$ of the random variables in points i and j with a distance h_{ij}, instead of the variogram. A16 can be solved numerically, given the number of samples, N, and a correlogram.

For random sampling, the distance becomes a random variable, and Rodríguez-Iturbe and Mejía (1974) replaced the distance with the expectation of distance:

$$Var(\overline{z_{L_E}}) = \frac{1}{N^2} E\left[\sum_{i=1}^{N} Cov(z_i, z_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} Cov(z_i, z_j) \right] \tag{A17}$$

$$Var(\overline{z_{L_E}}) = \frac{\sigma_{L_E}^2}{N^2} \{N + N(N-1)E[\rho(h_{ij} \mid L_E)]\} \tag{A18}$$

$E\left[\rho(h_{ij} \mid L_E)\right]$ represents the expected value of the correlation between two points randomly located in a square of side length L_E and can be expressed in a similar way to Equation (A9):

$$E\left[\rho(h_{ij} \mid L_E)\right] = \int_0^{R_{max}} \rho(r) f_1(r \mid L_E)\, dr \tag{A19}$$

where f_1 is taken from Equation (A4).

For large spacings, $\rho(h_{ij} \mid L_E) = 0$, $i \neq j$. The expression then reduces to:

$$Var(\overline{z_{L_E}}) = \frac{\sigma_{L_E}^2}{N} \tag{A20}$$

which is the normal expression for the expected random error of the mean for uncorrelated samples. The uncertainty of the estimates of variance for uncorrelated samples can be found as:

$$Var(s^2) = \frac{2\sigma_{L_E}^4}{N} \tag{A21}$$

References

Anderson, M. P.: 1997, 'Characterization of Geological Heterogeneity', in: G. Dagan and S. P. Neuman (eds), *Subsurface Flow and Transport: A Stochastic Approach. International Hydrology Series*, University Press, Cambridge, pp. 23–43.

Beckie, R.: 1996, 'Sampling scale, network sampling scale, and groundwater model parameters', *Water Resources Res.* **32**(1), 65–76.

Bellehumeur, C. and Legendre, P.: 1998, 'Multiscale sources of variation in ecological variables: Modeling spatial dispersion, elaborating sampling designs', *Landscape Ecol.* **13**(1), 15–25.

Bierkens, M. F. P., Finke, P. A. and de Willigen, P.: 2000, *Upscaling and Downscaling Methods for Environmental Research*. Kluwer Academic Press, Dordrecht, 190 pp.

Blöschl, G. and Sivapalan, M.: 1995, 'Scale issues in hydrological modelling – a review', *Hydrol. Processes* **9**(3/4), 251–290.

Blöschl, G.: 1999, 'Scaling issues in snow hydrology', *Hydrol. Processes* **13**(14–15), 2149–2175.

Caeiro, S., Painho, M., Goovaerts, P., Costa, H. and Sousa, S.: 2003, 'Spatial sampling design for sediment quality assessment in estuaries', *Environ. Modell. Software* **18**(10), 853–859.

Chilès, J.-P. and Delfiner, P.: 1999, *Geostatistics*. Modelling Spatial Uncertainty. Wiley, NY, 695 pp.

Cintoli, S., Neuman, S. P. and Di Federico, V.: 2004, Variograms of Fractional Brownian Motion on Finite Domains, geoENV 2004, Neuchâtel, Pre-Proceedings, pp. 347–357.

Cressie, N.: 1991, *Statistics for Spatial Data*. John Wiley, New York, 900 pp.

Cushman, J. H.: 1984, 'On unifying the concepts of scale, instrumentation, and stochastics in the development of multiphase transport theory', *Water Resources Res.* **20**(11), 1668–1676.

Cushman, J. H.: 1987, 'More on stochastic models', *Water Resources Res.* **23**, 750–752.

Di Federico, V. and Neuman, S. P.: 1997, 'Scaling of random fields by means of truncated power variograms and associated spectra', *Water Resources Res.* **33**(5), 1075–1085.

Entin, J. K., Robock, A., Vinnikov, K. Y., Hollinger, S. E., Liu, S. and Namkhai, A.: 2000, 'Temporal and spatial scales of observed soil moisture variations in the extratropics', *J. Geophys. Res.* **105**(D9), 11865–11877.

Faures, J.-M., Goodrich, D. C., Woolhiser, D. A. and Sorooshian, S.: 1995, 'Impact of small-scale spatial rainfall variability on runoff modeling', *J. Hydrology* **173**(1–4), 309–326.

Franklin, R. B. and Mills, A. L.: 2003, 'Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field', *FEMS Microbiol. Ecol.* **44**, 335–346.

Gelhar, L. W.: 1993, *Stochastic Subsurface Hydrology*. Prentice-Hall, Englewood Cliffs, N. J., 390 pp.

Ghosh, B.: 1951, 'Random distances within a rectangle and between two rectangles', *Bull. Calcutta Math. Soc.* **43**, 17–24.

Hatcher, B. G., Imberger, J. and Smith, S. V. 1987, 'Scaling analysis of coral reef systems: An approach to problems of scale', *Coral Reefs* **5**, 171–181.

Haugen, D. A.: 1978, 'Effects of Sampling Rates and Averaging Periods on Meteorological Measurements', in: *Proceedings of the 4th Symposium on Meteorological Observations and Instrumentation*, American Meteorological Society, Boston, pp. 15–18.

Hewitt, J. E., Thrush, S. F., Cummings, V. J. and Turner, S. J.: 1998, 'The effect of changing sampling scales on our ability to detect effects of large-scale processes on communities', *J. Experimental Marine Biology and Ecology* **227**(2), 251–264.

Journel, A. G. and Huijbregts, C. J.: 1978, *Mining Geostatistics*. Academic Press, London, 600 pp.

Jurado-Expósito, M., López-Granados, F., González-Andújar, J. L. and García-Torres, L.: 2004, 'Spatial and temporal analysis of *Convolvulus arvensis* L. populations over four growing seasons', *Eur. J. Agronomy* **I21**(3), 287–296.

Lopes, V. L.: 1996, 'On the effect of uncertainty in spatial distribution of rainfall on catchment modelling', *Catena* **28**, 107–119.

Mantoglou, A. and Wilson, J. L.: 1981, Simulation of Random Fields with the Turning Bands Method. *Report No. 264*, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Mantoglou, A. and Wilson, J. L.: 1982, 'The turning bands method for simulation of random fields using line generation by a spectral method', *Water Resources Res.* **18**(5), 1379–1394.

Matheron, G.: 1965, Les variables regionalisées et leur estimation. Masson, Paris, 305 pp.

McBratney, A. B.: 1992, 'On variation, uncertainty and informatics in environmental soil management', *Austr. J. Soil Res.* **30**, 913–935.

McDonald, T. L.: 2003, 'Review of environmental monitoring methods: Survey designs', *Environ. Monit. Assess.* **85**, 277–292.

Mohanty, B. P., Famiglietti, J. S. and Skaggs, T. H.: 2000, 'Evolution of soil moisture spatial structure in a mixed vegetation pixel during the Southern Great Plains 1997 (SGP97) Hydrology Experiment', *Water Resources Res.* **36**(12), 3675–3686.

Morrissey, M. L., Maliekal, J. A., Greene, J. S. and Wang, J.: 1995, 'The uncertainty of simple spatial averages using rain gauge networks', *Water Resources Res.* **31**(8), 2011–2017.

Nunan, N., Wu, K., Young, I. M., Crawford, J. W. and Ritz, K.: 2002, 'In situ spatial patterns of soil bacterial populations, mapped at multiple scales, in an arable soil', *Microbial Ecol.* **44**, 296–305.

Nyquist, H.: 1924, 'Certain factors affecting telegraph speed', *Bell Systems Techn. J.* **3**.

Oliver, M. A. and Khayrat, A. L.: 2001, 'A geostatistical investigation of the spatial variation of radon in soil', *Computers Geosciences* **27**, 939–957.

Priestley, M. B.: 1981, 'Spectral Analysis and Time Series', in: *Probability and Mathematical Statistics*. Academic Press Ltd, London, 890 pp.

Qi, Y. and Wu, J.: 1996, 'Effects of changing spatial resolution on the results of landscape pattern analysis using spatial autocorrelation indices', *Landscape Ecol.* **11**(1), 39–49.

Rahel, F. J.: 1990, 'The hierarchical nature of community persistence: A problem of scale', *The American Naturalist* **136**(3), 328–344.

Rodríguez-Iturbe, I. and Mejía, J. M.: 1974, 'The design of rainfall networks in time and space', *Water Resources Res.* **10**(4), 713–728.

Root, T. L. and Schneider, S. H.: 1995, 'Ecology and climate: Research strategies and implications', *Science* **269**, 334–340.

Russo, D. and Jury, W. A.: 1987, 'A theoretical study of the estimation of the correlation scale in spatially variable fields 1. stationary fields', *Water Resources Res.* **23**(7), 1257–1268.

Schmid, H. P.: 2002, 'Footprint modeling for vegetation atmosphere exchange studies: A review and perspective', *Agricult. Forest Meteor.* **113**, 159–183.

Seyfried, M. S. and Wilcox, B. P.: 1995, 'Scale and the nature of spatial variability: Field examples having implications for hydrological modeling', *Water Resources Res.* **31**(1), 173–184.

Sivapalan, M. and Blöschl, G.: 1998, 'Transformation of point rainfall to areal rainfall: Intensity-duration-frequency curves', *J. Hydrology* **204**(1–4), 150–167.

Sivapalan, M.: 1986, Scale Problems in Rainfall, Infiltration and Runoff Production. *PhD Dissertation Thesis*, Dept. of Civil Engng, Princeton University, 271 pp.

Skøien, J. O., Blöschl, G. and Western, A. W.: 2003, 'Characteristic space-time scales in hydrology', *Water Resources Res.* **39**(10): 1304, doi: 10.1029/2002WR001736.

Stenger, R., Priesack, E. and Beese, F.: 2002, 'Spatial variation of nitrate-N and related soil properties at the plot-scale', *Geoderma* **105**(3–4), 259–275.

Stommel, H.: 1963, 'Varieties of oceanographic experience', *Science* **139**, 572–576.

Taylor, G. I.: 1921, 'Diffusion by continuous movements', *Proc. London Math. Soc.* **20**, 196–211.

Theocharopoulos, S. P., Wagner, G., Sprengart, J., Mohr, M.-E., Desaules, A., Muntau, H., Christou, M. and Quevauviller, P.: 2001, 'European soil sampling guidelines for soil pollution studies', *Sci. Total Environ.* **264**(1–2), 51–62.

Thompson, S. K.: 2002, *Sampling. Wiley Series in Probability and Statistics*. Wiley, New York, USA, 367 pp.

Wallace, C. S. A., Watts, J. M. and Yool, S. R.: 2000, 'Characterizing the spatial structure of vegetation communities in the Mojave Desert using geostatistical techniques', *Computers Geosciences* **26**(4), 397–410.

Webster, R. and Oliver, M. A.: 2001, *Geostatistics for Environmental Scientists. Statistics in Practice*. Wiley, Chichester, England.

Western, A. W. and Blöschl, G.: 1999, 'On the spatial scaling of soil moisture', *J. Hydrology* **217**, 203–224.

Western, A. W., Zhou, S.-L., Grayson, R. B., McMahon, T. A., Blöschl G. and Wilson D. J.: 2004, 'Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes', *J. Hydrology* **286**(1–4), 113–134.

Western, A. W., Blöschl G. and Grayson R. B.: 1998, 'Geostatistical characterisation of soil moisture patterns in the Tarrawarra catchment', *J. Hydrology* **205**, 20–37.

Wiens, J. A.: 1989, 'Spatial scaling in ecology', *Functional Ecol.* **3**, 385–397.

Zehe, E. and Blöschl, G.: 2004, 'Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions', *Water Resources Res.* **40**, W10202, doi:10.1029/2003WR002869.