ELSEVIER

Geoderma 89 (1999) 47–65

GEODERMA

# Spatial aggregation and soil process modelling

Gerard B.M. Heuvelink *, Edzer J. Pebesma

*Netherlands Centre for Geo-ecological Research, Faculty of Environmental Sciences, University of Amsterdam, Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, Netherlands*

Received 4 November 1997; accepted 6 July 1998

**Abstract**

Nonlinear soil process models that are defined and calibrated at the point support cannot at the same time be valid at the block support. This means that in the situation where model input is available at point support and where model output is required at block support, spatial aggregation should take place after the model is run. Although block kriging does both in one pass, it is sensible to separate spatial aggregation from spatial interpolation. Contrary to aggregation, interpolation should better take place before the model is run because this enables a more efficient use of the spatial distribution characteristics of individual inputs. When a model is run with interpolated inputs, it is important not to ignore the interpolation error. Substituting conditional expectations in place of probability distributions into a nonlinear model leads to bias, essentially for the same reason that aggregating inputs prior to running a model is not the same as aggregating the output after the model is run. Running a model with inputs that are probability distributions will usually call for a Monte Carlo simulation approach. This causes a substantial increase in the numerical load, but apart from eliminating bias, an important advantage is that it shows how uncertainties in model inputs propagate to the model output. Many models used in soil science suffer not only from input error but also from model error, which is support- and case-dependent. Case dependency implies that model error can only be assessed realistically through validation. A major problem in validation is that the validation data are often collected at a much smaller support than the aggregated model predictions. 

*Keywords:* soil process modelling; uncertainty analysis; change of support; kriging; Monte Carlo simulation; validation

* Corresponding author. Tel.: +31-20-525-7448; Fax: +31-20-525-7431; E-mail: g.b.m.heuvelink@frw.uva.nl

0016-7061/99/$ - see front matter 
PII: S0016-7061(98)00077-9

## 1. Introduction

The modelling of soil processes takes place at various spatial scales. For instance, models of soil acidification have been built for the scale of plots, regions and continents (de Vries et al., 1998). In a recent workshop of the International Soil Science Society (Finke et al., 1998), it was recognised that models describing the same process at different spatial scales can be quite different. From the papers presented at the workshop, it was concluded that there are three main reasons why models are scale-specific (Heuvelink, 1998a).

The first reason is that the relative importance of a subprocess depends on the scale considered. For instance, it has been observed that the dominant process governing unsaturated flow changes from matrix flow to preferential flow when moving to a larger scale (Blöschl and Sivapalan, 1995). Because modelling tends to focus on dominant processes and tends to ignore less important ones, different processes will be discarded in the simplification step of the model development. At the field scale, the modelling of nitrate leaching may focus on the influence of natural variation in the soil, but at the farm scale the variation in land use will be much more important (Addiscott and Tuck, 1996). Consequently, at the farm scale the modelling of nitrate leaching will primarily concentrate on capturing the influence of land use, and if forced to make a choice, it will do so by sacrificing the influence of the natural variation in the soil.

The second reason for ending up with different models at different scales relates to input data availability. At the small scale, data are often available through measurements. At the large scale, inputs have to be derived from general information sources, such as from general purpose soil maps, agricultural statistics or expert judgements. Obtaining input data in this way usually also involves the use of pedotransfer functions (Bouma, 1986; Wösten et al., 1995). Deriving model inputs from general information sources and through pedotransfer functions inevitably deteriorates the quality of the input data, and this has led many model developers to simplify their models when moving from smaller to larger scales (de Vries et al., 1998; Stoorvogel and Smaling, 1998; Tiktak et al., 1998). Not only does it seem intuitively right to turn to a simpler model when the more complex model requires a large amount of data that is either not available or very uncertain, it can also be theoretically justified by uncertainty analyses (Heuvelink, 1998b; Jansen, 1998).

The third reason relates to the ‘support’ (i.e., the integration volume or aggregation level) of the model variables. Moving up from the small to the large scale usually implies that the model input and output become some kind of average of ‘point’ values within the larger spatial unit or ‘block’. For example, the support of the local scale soil acidification model NUCSAM is much smaller than that of the regional scale model RESAM (de Vries et al., 1998). A change of support may give rise to a model adjustment because relationships between

model variables that exist at the point support do not necessarily extend to the block support.

In practice, model output is often required at a support much larger than that at which the soil process model was developed. The purpose of this paper is to analyse how the change of support should be carried out, taking into consideration that the model is support-dependent. Although the issue of scale and support in soil process modelling has recently received much attention (Petach et al., 1991; Hoosbeek and Bryant, 1992; Addiscott, 1993; Addiscott and Tuck, 1996; Wagenet and Hutson, 1996; Finke et al., 1998), with this paper we hope to contribute to the existing work by approaching the problem from a pedometricians perspective. We also hope to draw the attention of fellow-pedometricians to this particular problem because we feel that pedometricians have an important role to play in solving the problems associated with the change of support in soil process modelling.

A few remarks on terminology and scope of the paper are made here. Although the title indicates that the focus is on soil process models, this is not to say that static models are excluded from the analysis. For instance, empirical pedotransfer functions will be of equal interest to us. The term 'soil process models' is used here mainly to distinguish these models from soil distributional models (de Gruijter et al., 1997). Also, the use of the term aggregation in the title does not exclude disaggregation from the analysis. Disaggregation is judged less important because it is less frequently needed in practice. Note also that we have chosen to use the terms 'aggregation' and 'disaggregation' instead of the terms 'upscaling' and 'downscaling'. The latter two have a rather specific meaning, particularly in geo-hydrology, and as such they represent an area of research that is beyond the scope of this paper (we will come back to this in Section 3.3). Aggregation and disaggregation together are captured in the term 'change of support'. 'Scale' is a rather loosely defined term that we will use as a synonym to support.

## 2. Role of support in soil process models

Many soil process models require parameter values and/or input data that can only be measured or obtained at a scale much smaller than the desired output scale. For example, a soil water and solute transport model such as NUCSAM (Groenenberg et al., 1995; de Vries et al., 1998) or LEACHM (Wagenet and Hutson, 1989; Petach et al., 1991) needs numerous soil physical and chemical properties that can only be measured at point support (a 'point' in this case being a volume in the order of a $dm^3$), but the scale at which the model output becomes interesting for farmers, planners and groundwater managers is the size of hectares or larger. The problem is graphically illustrated in Fig. 1. Not only do we need to run the model to move horizontally from input to

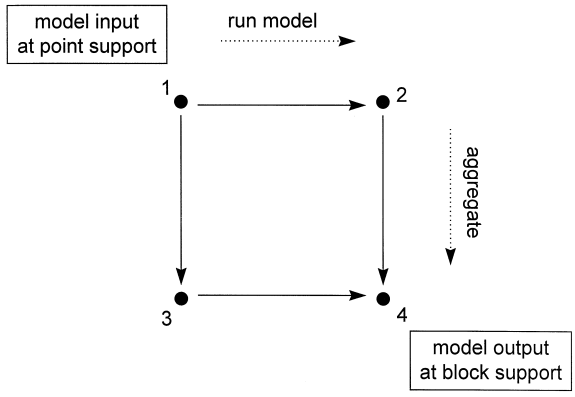


Fig. 1. Two routes from point support input to block support output.

output, we also need to *aggregate* to move vertically from point support to block support. The aggregation may be linear, such as when the block support value is the arithmetic mean of the point support values in the block, but it may also be nonlinear. For example, environmental pollution studies may be interested in the areal fraction of the block where the concentration of a pollutant exceeds a critical level, in the block median value or in the block geometric mean (Pebesma and de Kwaadsteniet, 1997). Hereafter, when the type of aggregation is not specified, aggregation will mean the computation of the block arithmetic mean.

Fig. 1 shows that there are two routes to obtain the model output at block support. One is to run the model at all points within the block first and then to aggregate the point support outputs, the other is to first aggregate the input to the block support and then to run the model at the block support. Two questions immediately emerge: will both routes yield the same result? And if not, which of the two routes should be preferred?

The answer to the first question is straightforward: the two routes will yield the same result when the model is linear and they usually will not when the model is nonlinear (Addiscott, 1993; Addiscott and Tuck, 1996; Groot et al., 1998; Heuvelink, 1998a). In mathematical terms, let $g(\cdot)$ be the model that yields as output $y = g(\boldsymbol{z})$, where the input vector $\boldsymbol{z}$ is a function of location. The difference in results is due to the fact that, in general, for nonlinear $g(\cdot)$, we have:

$$g\left(\frac{1}{|B|}\int_B \boldsymbol{z}(x)\mathrm{d}x\right) \neq \frac{1}{|B|}\int_B g(\boldsymbol{z}(x))\mathrm{d}x. \quad (1)$$

Exceptions to this rule are when $g(\cdot)$ is linear in its arguments or when the input vector $\boldsymbol{z}$ is constant within the block $B$. However, neither of these two conditions are likely to occur in real-world applications (Petach et al., 1991; Addiscott, 1993; Hosang, 1993).

### 2.1. Application to the linear Allier pedotransfer function

To illustrate that the output of a linear model is not dependent on the route that is followed, consider the linear pedotransfer function that predicts soil moisture content at wilting point ($\theta_{wp}$) from soil moisture content at field capacity ($\theta_{fc}$) and soil porosity ($\varphi$) (Heuvelink et al., 1989):

$$\theta_{wp} = -0.263 + 0.408\,\theta_{fc} + 0.491\,\varphi. \quad (2)$$

This function resulted from a linear regression on point support measurements of all three variables at twelve sites in a study area located along the Allier river in the Limagne rift valley, central France. The study area is given in Fig. 2.

Maps at point support of the two inputs to the model are given in Fig. 3. In fact, these maps were obtained by interpolating with cokriging from measurements of $\theta_{fc}$ and $\varphi$ at the 62 sampling sites indicated in Fig. 2. These maps are therefore prone to interpolation error, but we will ignore the presence of interpolation error for now. Until a later stage, we will also ignore the presence

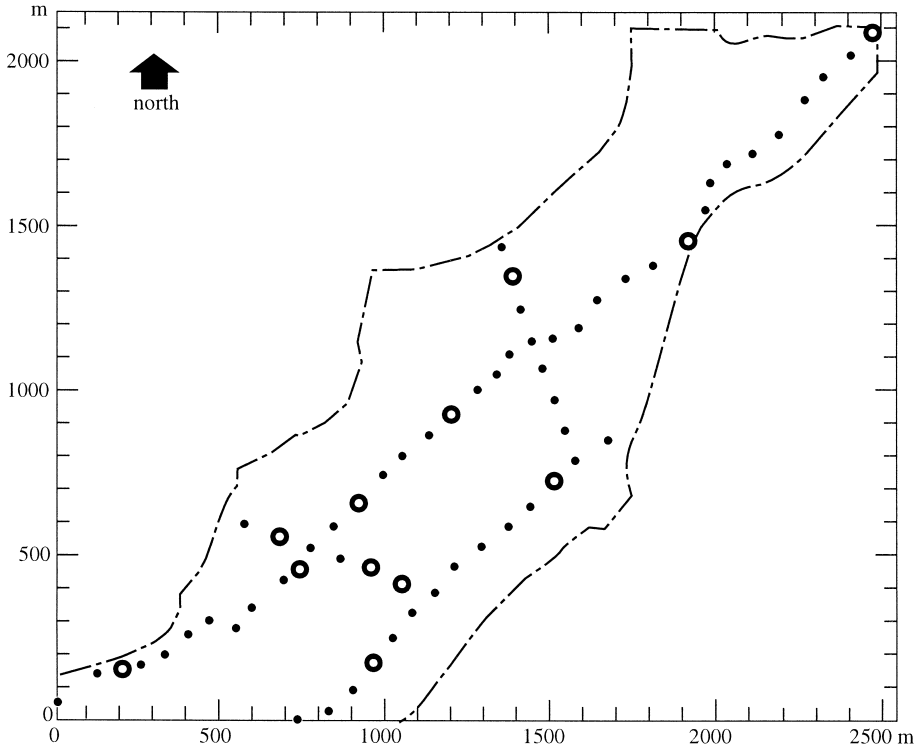


Fig. 2. The Allier study area showing sampling sites. Circled sites are those used to define the pedotransfer function.

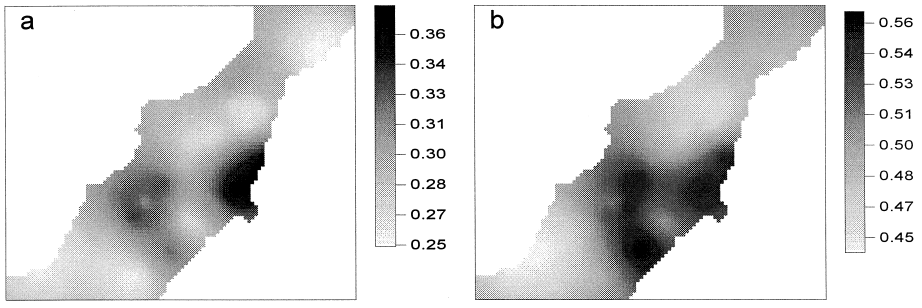


Fig. 3. Point support maps of (a) moisture content at field capacity, and (b) porosity.

of model error, and thus for now assume that Eq. (2) produces as output the true moisture content at wilting point.

Suppose we want as output the average moisture content at wilting point for 1 ha squares. Whether we first aggregate the point support inputs to the 1 ha support (Fig. 4) or first run the model at all points in the area and next aggregate the point support outputs to the 1 ha support, the result will be exactly the same (Fig. 5). In other words, although the model Eq. (2) was derived at the point support, its linearity makes it equally valid at the block support.

### *2.2. Nonlinear models*

The majority of models used in soil science are nonlinear and for these models it usually does matter which route in Fig. 1 is followed. In such a situation, it is advisable to use the route in which the model is run at point support. There are two main reasons for this. First, many models are implementations of physical laws that are established and proven 'valid' at the point support. For instance, Darcy's law and the convection–dispersion equation are partial differential equations that apply to the pedon scale but whose validity is questionable at the field scale and larger (Addiscott, 1993). Second, many soil

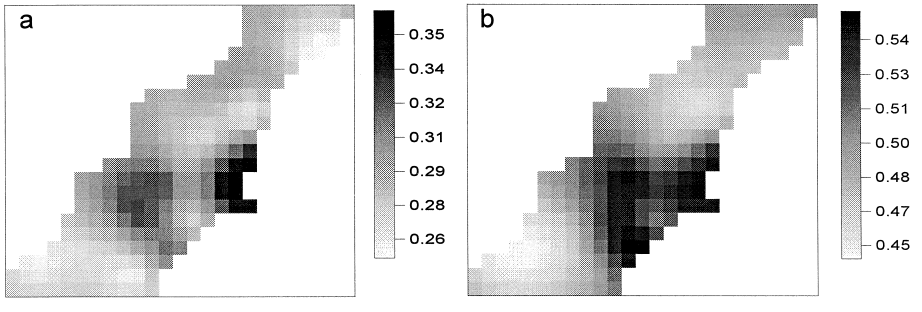


Fig. 4. Block support maps of (a) moisture content at field capacity, and (b) porosity.

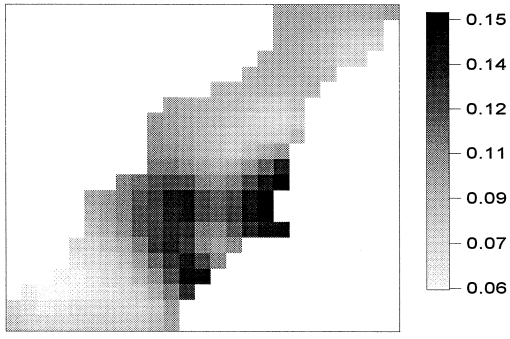


Fig. 5. Block support map of moisture content at wilting point.

process models have been calibrated at the point support because this is the support at which calibration data are obtained and experiments are carried out. Models with parameters that are calibrated using data at a particular support are only valid at the given support. For instance, a dispersion coefficient determined from a laboratory experiment is of no value as input into a model intended for field-scale application (Corwin, 1996). Likewise, a nonlinear pedotransfer function relating difficult-to-measure soil physical properties to more easily measured soil properties (Wösten and van Genuchten, 1988) is support-dependent and thus only valid at the support at which the data used to derive the pedotransfer function were collected.

## 3. Change of support

A change of support can be directed upward (i.e., aggregation) and downward (i.e., disaggregation). Aggregation means that a block value is computed from all point values within the block. Linear and nonlinear aggregation are trivial operations when all point values in the block are known. However, aggregation is no longer trivial when not all point values within the block are known exactly. Clearly, this is the common situation. Often, point values are given only at a finite set of points within the block. This means that the block value can no longer be computed but has to be estimated. Two fundamentally different methods to obtain an estimate of the block value and the associated estimation accuracy are the *design-based* and the *model-based* approach. We discuss both approaches in the usual situation where aggregation means computation of the block arithmetic mean.

### *3.1. Design-based and model-based approaches to spatial aggregation*

The design-based approach (de Gruijter and Ter Braak, 1990; Brus and de Gruijter, 1997) can only be used when at least one point value is given in the

block and when the point locations are derived from probability sampling. In the case of a simple random sample, the design-based estimate of the block average is the unweighted sample mean of the point values in the block. The efficiency of the method can be improved by utilising stratification and other techniques (Brus, 1994). An important property of the design-based approach is that it makes no assumptions about the spatial variability of the aggregation variable.

The model-based approach to spatial aggregation, otherwise known as block kriging, does assume the spatial variability of the variable of interest to be known. This enables the use of point values outside the block to predict the block average, which can be very advantageous in a situation of sparse data. Additional information in the form of explanatory variables or a stratification of the area can be handled by universal or stratified kriging (Pebesma and de Kwaadsteniet, 1997; Pebesma and Wesseling, 1998). Although a model-based approach can potentially produce more accurate predictions of the block average than its design-based counterpart, the accuracy obtained is built on the presumption that the attribute of interest satisfies the assumed spatial correlation model. The necessity to make some kind of stationarity assumption and the assumption that the variogram is known are weak spots of the model-based approach. For instance, block kriging results are highly sensitive to the nugget variance and this is typically a parameter that is often poorly estimated.

### *3.2. Block kriging the output of the Allier pedotransfer function*

In Section 2.1, we assumed that soil moisture content at field capacity $\theta_{fc}$ and soil porosity $\varphi$ were known at point support everywhere in the study area. In fact this was not true. Observations of these attributes were only given at the 62 sampling sites indicated in Fig. 2. Thus the model output is only known at these 62 sites. In this case, where the majority of 1 ha blocks contain no point output values, the only option for aggregation is a model-based approach. Block kriging the 62 point outputs yields the map of Fig. 6a. The kriging error is quantified by the block kriging variance in the map of Fig. 6b.

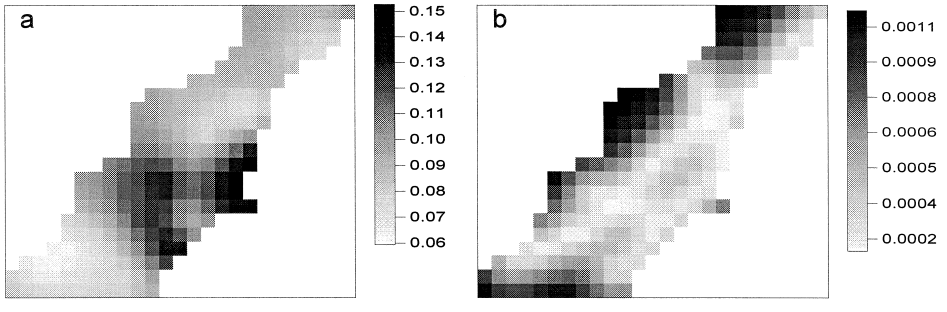


Fig. 6. Block kriging moisture content at wilting point: (a) prediction map, (b) prediction error variance map.

Although impossible to discern from visual comparison of Figs. 5 and 6a, it is worthwhile to note that these maps are not exactly the same. The reason for this will be explained in Section 4.

### 3.3. Upscaling

In groundwater modelling, many of the controlling parameters such as hydraulic conductivity and dispersivity are measured on sediment cores. However, it is well known that these measurements cannot directly be substituted in a groundwater model operating on a larger scale (Wen and Gómez-Hernández, 1996; Bierkens and van der Gaast, 1998). The point scale parameters must somehow be transformed into block scale parameters before they can be used in a regional groundwater model. This transformation process is termed upscaling.

Thus upscaling is a special case of spatial aggregation, namely one in which the objective is to transform the point parameter values into 'effective' block parameter values, such that the microscale equations in the model become valid at the macroscale (Blöschl and Sivapalan, 1995). In other words, it is a methodology that through the convenient choice of effective block parameters attempts to stretch a point scale model to the block scale. A fundamental difference between aggregation and upscaling is that whereas aggregation can be defined irrespective of a model operating on the aggregated values, upscaling must always be defined in the context of a model that uses the parameters that have been scaled up. The theory of upscaling is extensive and numerous upscaling methods have been proposed over the years. Nonetheless, due to fundamental problems, all techniques have their limitations (Wen and Gómez-Hernández, 1996).

Although the literature sometimes treats upscaling as synonymous to spatial aggregation, for reasons of clarity we have chosen to reserve this term exclusively for the transformation of point values to 'effective' block values.

### 3.4. Spatial disaggregation

A change of support need not always be directed upward. For example, when atmospheric deposition is only available at the support of grid cells of hundreds of square km, it will have to be *disaggregated* when it is to be used as input to a soil acidification model operating at the point support. Other examples are the disaggregation of satellite data (Blöschl and Sivapalan, 1995) and the temporal disaggregation of precipitation data (Groot et al., 1998). Disaggregation is more difficult than aggregation, basically because block average values do not reveal any information about the distribution of point values within the block. For instance, it is easy to derive the block average porosity (Fig. 4b) from the point porosities (Fig. 3a), but the opposite has no unique solution, implying that there will be uncertainty associated with the disaggregated values. Quantification of

this uncertainty requires that we must have knowledge about the spatial variability at the point support, which cannot be inferred from the block values alone.

## 4. Interpolate first, calculate later

### *4.1. Separating interpolation from aggregation*

We now return to the model-based approach to spatial aggregation. In Section 3.1, we introduced block kriging as a means for spatial aggregation in the usual circumstance where there are only a limited number of sampling sites in the area of interest. Closer inspection of block kriging reveals that in fact it consists of two steps. These are an interpolation step and an aggregation step, which we will now separate from each other.

Separation of the interpolation and aggregation steps means that the two-dimensional square in Fig. 1 becomes the three-dimensional cube given in Fig. 7. The objective is still to move from point support input to block support output, but, in addition, also to move from observations at sampling sites to an areal coverage of the output. The number of different routes through the cube has now become six. Which route is the most appropriate? Clearly, some routes can be discarded instantly. Following our observations in Section 2, we should run the model at point support, meaning that the optimal route should not pass through points 3 or 7. Also, aggregation before interpolation is not a real option because in order to aggregate we must know the value at all points within the block (albeit in a statistical sense). Therefore, routes through 4 must also be rejected. This leaves two possible routes: 1–2–6–8 and 1–5–6–8. The difference be-

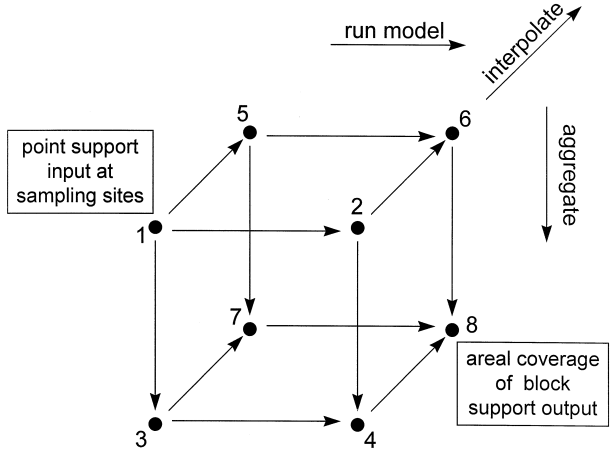


Fig. 7. From point support input at sampling sites to an areal coverage of block support output involves three separate steps.

tween these two routes is whether we first run the model and then interpolate the output or whether we first interpolate the input and then run the model.

The question raised here was also addressed by Stein et al. (1991) and Bosma et al. (1994), who referred to it as the choice between calculate first, interpolate later (CI) and interpolate first, calculate later (IC). The issue was also discussed by Addiscott and Tuck (1996), who noticed substantial differences in a case where the model was nonlinear. Here, we will focus on the comparison of CI and IC for the case of a simple linear model. This is worthwhile because it allows us to draw conclusions that go beyond the empirical results of previous studies.

### *4.2. Comparison of CI and IC for the Allier pedotransfer function*

The variograms and cross-variogram of the two inputs, $\theta_{fc}$ and $\varphi$, to the Allier pedotransfer function Eq. (2) were determined from the 62 observations using the linear model of coregionalisation (Journel and Huijbregts, 1978). The parameters of the fitted model are given in Table 1. Thus, each variogram is a weighted sum of a nugget variogram, a spherical variogram with a range of 500 m and a spherical variogram with a range of 800 m. Together with the pedotransfer function these variograms completely specify the variogram of the model output $\theta_{wp}$, also given in Table 1.

The CI procedure is done by first computing the value of $\theta_{wp}$ at each of the 62 sampling sites using Eq. (2), followed by interpolation of $\theta_{wp}$. Here, we use ordinary kriging to interpolate $\theta_{wp}$ (Journel and Huijbregts, 1978). This yields maps of kriging predictions $\hat{\theta}_{wp}^{CI}$ and kriging prediction error variances $\sigma_{CI}^2$. The IC procedure first interpolates $\theta_{fc}$ and $\varphi$ using ordinary cokriging and then substitutes the interpolated maps in Eq. (2). The prediction error variance associated with the so-obtained prediction $\hat{\theta}_{wp}^{IC}$ is:

$$\sigma_{IC}^2 = \mathrm{var}\left(\theta_{wp} - \hat{\theta}_{wp}^{IC}\right) = 0.408^2\ \mathrm{var}\left(\theta_{fc} - \hat{\theta}_{fc}\right) + 0.491^2\ \mathrm{var}\left(\varphi - \hat{\varphi}\right) + 2 \cdot 0.408 \cdot 0.491\ \mathrm{cov}\left(\theta_{fc} - \hat{\theta}_{fc}, \varphi - \hat{\varphi}\right), \quad (3)$$

where the kriging variances and the covariance term at the right hand side of

Table 1
Variogram parameters of $\theta_{fc}$, $\varphi$ and $\theta_{wp}$

| | Nugget | Spherical variogram with range 500 m | Spherical variogram with range 800 m |
|---|---|---|---|
| Variogram of $\theta_{fc}$ | 0.0023 | 0.0015 | 0.0003 |
| Variogram of $\varphi$ | 0.0020 | 0.0003 | 0.0017 |
| Cross-variogram of $\theta_{fc}$ and $\varphi$ | −0.0005 | 0.0005 | 0.0006 |
| Variogram of $\theta_{wp}$ | 0.00066 | 0.00052 | 0.00070 |

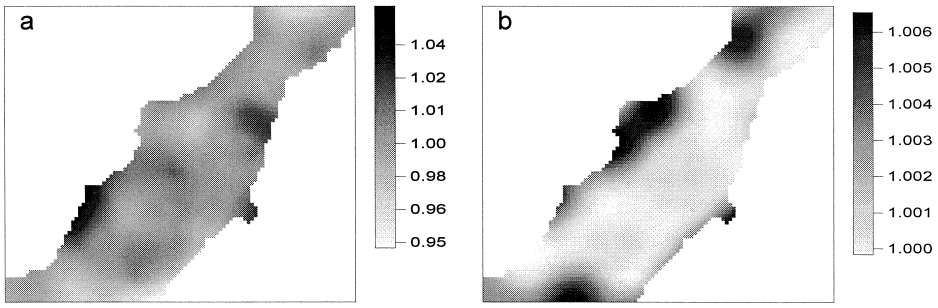


Fig. 8. CI vs. IC: (a) quotient of CI predictions and IC predictions, (b) quotient of CI prediction error variances and IC prediction error variances.

Eq. (3) are produced as output by the gstat geostatistical program (Pebesma and Wesseling, 1998).

Fig. 8 gives the quotients of the CI and IC predictions and prediction error variances. Although differences are small, apparently even with linear models, it does matter whether CI or IC is used. The reason that IC yields slightly smaller prediction error variances is that the kriging step of CI does not fully exploit all available information. The two inputs, $\theta_{fc}$ and $\varphi$, each have their own specific spatial correlation structure. This is taken into account by IC when it kriges $\theta_{fc}$ and $\varphi$. But CI cannot make use of the differences in spatial correlation structure because it interpolates $\theta_{wp}$, which is a composite of $\theta_{fc}$ and $\varphi$. In other words, information is lost when prior to the interpolation, the variables $\theta_{fc}$ and $\varphi$ are translated into $\theta_{wp}$ through Eq. (2). The only way to exploit all information would be to include $\theta_{fc}$ and $\varphi$ as covariates when kriging $\theta_{wp}$, but for complicated models, this will not be a realistic option. Note also that CI and IC

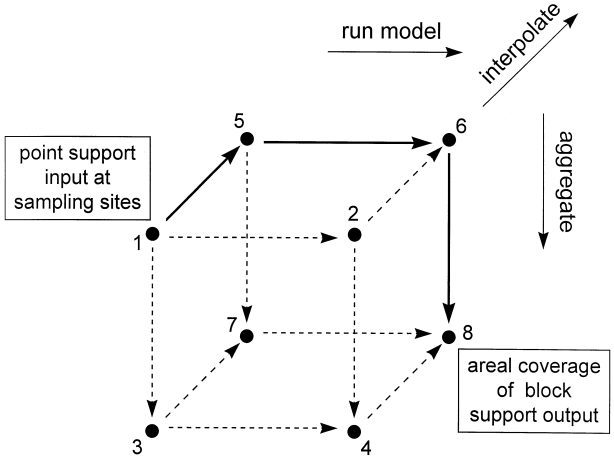


Fig. 9. The preferred route from point support input at sampling sites to an areal coverage of block support output is to interpolate first, then to run the model and last to aggregate.

would have been identical if the spatial correlation structure of $\theta_{fc}$ and $\varphi$ would have been the same, i.e., if they would have been autokrigeable (Wackernagel, 1994).

Although in this example the improvement of IC over CI is only marginal, in the general situation, the differences between IC and CI may become more important. A recent example where the differences were indeed nonnegligible is given by Sinowski et al. (1997). Pronounced differences can also be expected when some of the inputs strongly depend on soil type or land use where others do not, or when the temporal variability of inputs and outputs differs strongly (de Gruijter et al., 1997). Thus, in the general situation where soil process models operate on many inputs, each with their own specific spatial correlation structure, IC is expected to yield more accurate results than CI.

In summary, the optimal route through the cube of Fig. 7 is to first interpolate the model inputs, second to run the model at all points in the block, and third to aggregate the point support model outputs. This is graphically illustrated in Fig. 9.

## 5. Input uncertainty

Interpolation errors cause uncertainty in the interpolated model inputs. If, in the case of a nonlinear model, we ignore this fact and substitute the kriging prediction (a deterministic value) instead of the probability distribution associated with the kriging prediction (a random variable) then we may introduce a bias. This is because for nonlinear $g(\cdot)$ we will usually have:

$$g\big(E[Z(x)]\big) \neq E\big[g(Z(x))\big]. \tag{4}$$

It is interesting to examine the analogy between Eqs. (4) and (1). Both equations tell us that with a nonlinear model, it matters whether we first run the model or first do the averaging. However, in Eq. (1) averaging is done in the spatial domain, whereas in Eq. (4), averaging takes place in probability space.

Given the probability distribution associated with the interpolated value, $Z(x)$, which characterises our uncertainty about it (Goovaerts, 1997, Chap. 7; Heuvelink, 1998b), we can compute the probability distribution of the model output, $g(Z(x))$. Although an analytical approach may sometimes be used for this (Eq. (3) is an example), usually one will turn to a Monte Carlo simulation approach (Heuvelink, 1998b). Apart from eliminating bias, an additional advantage of treating model inputs as stochastic is that it reveals how model input uncertainties propagate to the model output (Burrough, 1992; Heuvelink, 1998b).

It is important to distinguish between the necessity of repeatedly running the model at a single location as required by the Monte Carlo step and the necessity of repeatedly running the model at multiple locations in the block as required by the aggregation step. Both steps will yield variability in the point support output, but the source of the variability is not the same. The first is caused by

*interpolation uncertainty* in the model input, the second by *spatial variability* of the input. Clearly, only the first source of variability will cause uncertainty in the aggregated output.

To illustrate the difference between the two sources of variability, consider Fig. 10. This figure gives the cumulative distribution of the point support outputs within a single block. If there were no input uncertainty, then the cumulative distribution would be a single nondecreasing graph. However, due to input uncertainty, there will be a population of graphs since each Monte Carlo run produces a different graph. The population of graphs is represented here by the median and the upper and lower limit of a prediction interval. Starting at the horizontal axis, the figure shows which fraction of the block has a value below a certain threshold. It does so both in terms of a central value and a prediction interval. Starting from the vertical axis, the figure shows above or below which model output value a given fraction of the block lies. Again, it does so both in terms of a central value and a prediction interval.

An interesting property of Fig. 10 is that it neatly separates spatial variability from uncertainty. Fig. 10a is an example where the spatial variability is small because the cumulative distribution is a steep graph, but where the uncertainty is large because the prediction interval is wide. Fig. 10b is an example of the opposite: the spatial variability is large because the cumulative distribution is a flatter curve, but the uncertainty is small because the prediction interval is narrow. Note also that, when everything else is kept the same, increasing the block size moves one from a situation such as in Fig. 10a to a situation as in Fig. 10b (Pebesma et al., in preparation).

In parameterising input uncertainties, it is extremely important to quantify the spatial correlation realistically. For instance, ignoring spatial correlation, i.e., assuming that all spatial variation is pure nugget or white noise, leads to the

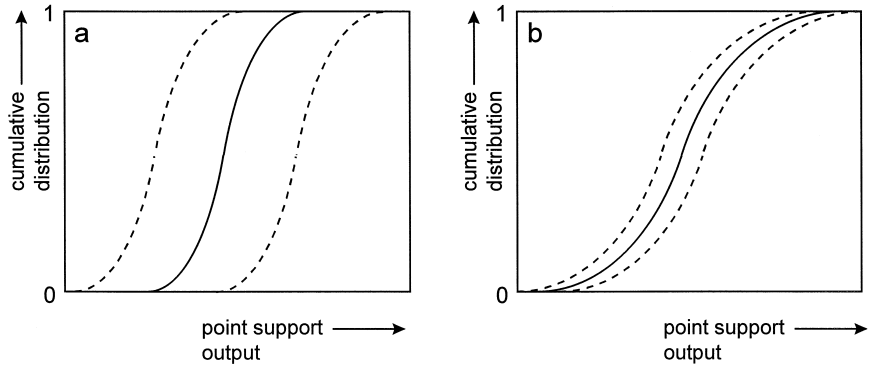


Fig. 10. Cumulative distribution of point support output in a single block: (a) small within-block spatial variability and large uncertainty, (b) large within-block spatial variability and small uncertainty. Solid lines represent median, dashed lines represent upper and lower limit of the associated prediction interval.

situation where uncertainty in block-support model output vanishes completely because it will be absorbed by the within-block spatial variability (Pebesma et al., in preparation).

## 6. Model uncertainty

Up to now, we have assumed that the soil process model itself is error-free. But clearly, the various assumptions, discretisations and simplifications that are made to make the model manageable will cause the model output to deviate from reality, even when the model inputs are free of error. This is also true of the empirical pedotransfer function Eq. (2) because it is simply not true that $\theta_{wp}$ exactly and linearly depends on $\theta_{fc}$ and $\varphi$ only. The purpose of this section is not to dwell on handling model error, but to single out how the change of support affects model error assessment.

Arguably, the only satisfactory way to assess model error is through validation. Note that validation is used here in the interpretation of confronting model predictions with independent measurements, not in its meaning of showing whether the model is 'right' (Konikow and Bredehoeft, 1992). When validation is used to assess model error, we will face a change of support problem because the model output and the validation measurements are usually not at the same support. One approach to deal with this problem is to aggregate the validation measurements to the block support using the methods described in Section 3. This brings about an uncertainty in the aggregated validation data, and the validation procedure will have to take this into account.

Let $\tilde{y}_B$ be the model prediction of the true block support output $y_B$, and let $\hat{y}_B$ be the estimate of $y_B$ based on the point validation measurements. Now the squared deviation $(\tilde{y}_B - \hat{y}_B)^2$ can be decomposed into (Heuvelink, 1998a):

$$(\tilde{y}_B - \hat{y}_B)^2 = ((\tilde{y}_B - y_B) + (y_B - \hat{y}_B))^2 \approx (y_B - \tilde{y}_B)^2 + (y_B - \hat{y}_B)^2, \qquad (5)$$

where the latter equality holds when the cross-product of $(y_B - \tilde{y}_B)$ and $(y_B - \hat{y}_B)$ is zero. Assuming this is true, Eq. (5) shows that evaluating a model on only $(\tilde{y}_B - \hat{y}_B)^2$ is unfair because it should really be judged on $(\tilde{y}_B - y_B)^2$, which gives a smaller value. Comparison of the magnitude of the two terms on the right hand side of Eq. (5) gives insight into whether sufficient effort has been spent on collecting validation data (Wallach and Goffinet, 1987).

Aggregation of the point validation measurements can be done using a design-based and a model-based approach. As mentioned, for efficiency reasons one will often favour a model-based approach. This may be done using straightforward ordinary block kriging but also using universal block kriging. By incorporating explanatory information such as land use and soil type, universal kriging is likely to improve the aggregation. However, when universal kriging

uses the same ancillary information as the soil process model, and when the ancillary information is subject to error, this may yield similarities between the aggregated validation data and process model output that are unsupported by reality. This means that the cross-product of $(y_B - \tilde{y}_B)$ and $(y_B - \hat{y}_B)$ is not zero but positive. In fact, any model-based approach to aggregating point validation measurements should be critically examined on the possibility of inducing a positive cross-product. This is important because ignoring a positive cross-product and judging the performance of the soil process model on a direct comparison with the aggregated validation data may yield over-optimistic results.

## 7. Discussion and conclusions

In general, nonlinear soil process models are only valid at the support at which they were developed. Usually, this will be the point support because at this support the underlying physical equations are proven valid, and calibration data have been collected at this support. To obtain spatially aggregated values for model output, we advocate to run the model at the point scale at multiple point locations within the block, and then to aggregate the model output, thereby avoiding application of the model at a larger scale.

The representation of a block by a limited number of points results in an error, and this error can be made very small by running the model at many points within a block. When computational burden limits the number of points, a compromise should be sought—too few points may not represent the block too well. General guidelines for the number of discretising points that should be used are hard to give since the approximation error typically increases with increasing within-block (spatial) variability of the model inputs, and with the degree of nonlinearity of the aggregation performed.

When model input is uncertain, Monte Carlo simulations using realisations of the input are necessary to avoid bias and to obtain subsequent model output uncertainties. This multiplies the computational load with the number of Monte Carlo runs. Given a maximum feasible number of individual (point support) model runs, block discretisation density and Monte Carlo sample size should be balanced.

The price of this computationally intensive procedure may be high, but the main advantages—no upscaling of the point support model and unbiased predictions of model outputs and their uncertainties—are clear. Other advantages of the procedure are: (i) since point support output is available, any linear or nonlinear aggregation may be used (e.g., block mean, block medians and other percentiles, areal fractions exceeding a threshold); (ii) any block size or shape may be chosen; (iii) the procedure yields for each block the within-block distribution of the point support model output for each Monte Carlo run, thereby nicely separating spatial variability from uncertainty. This may help in choosing the block size.

Advocating that a model should be run at the point support raises the question what the 'point' support really is. On a regional scale, volumes of 1 $dm^3$ and 1 $m^3$ are both considered as 'points', but clearly these supports are not the same, and again a nonlinear model will usually not be valid at both 'point' supports. The solution to this problem is simply that the model should be run at the support for which it was developed. Thus, we place the responsibility with the model builder because that is where we believe it should be. It is the task of the model builder to specify what the support of the various model inputs is and to clearly convey this information to the user. A model is simply incomplete as long as this is left open. Note that this is true for both the spatial and the temporal support of the model inputs. Although the focus in this paper was on spatial aggregation, most of the ideas presented here apply to temporal aggregation as well.

A particular situation arises when the 'block' contains a large number of sampling sites which were allocated using probability sampling. In such a situation it may be attractive to skip the interpolation step altogether by estimating the block support output from the point support outputs at the sampling sites, using a design-based approach. The advantages are that no model-based assumptions need to be made and that the entire procedure from point support inputs at sampling sites to an areal coverage of block support outputs is very much simplified. The disadvantage is a possible loss of efficiency, which is likely to be large when the number of sampling sites in the block is small, or when relevant information at unsampled sites is ignored. Therefore in practice, this procedure is only a feasible option with large blocks, such as when the block is a soil mapping unit or the entire study area (Brus, 1994).

Another alternative procedure only briefly addressed in this paper is to make the point support model suitable to the block support, by upscaling its parameters. This is probably the only realistic option for models involving spatial interactions (e.g., lateral flow). For example, regional application of a groundwater model with model entities that are at the core support is simply not practically feasible. This is particularly so because the spatial interactions that are implicit in the model impose that *all* cores in the block are considered in the analysis. With (one-dimensional) soil process models there is no need to run the model at all points in the block. Given the fundamental problems of upscaling, one should be critical about the use of upscaling as a solution method for change of support problems in soil process modelling.

## Acknowledgements

This work was partly funded by the European Commission, through the Environment and Climate Research and Technological Development Programme

(contract number ENV4-CT95-0070). We thank Dr. J. Bouma and Dr. A. Stein of the Department of Soil Science and Geology, Agricultural University Wageningen for their permission to use the Allier data set. We thank Dr. G. Blöschl, Dr. D.L. Corwin and Dr. J.J. de Gruijter for helpful comments.

## References

Addiscott, T.M., 1993. Simulation modelling and soil behaviour. Geoderma 60, 15–40.

Addiscott, T.M., Tuck, G., 1996. Sensitivity analysis for regional-scale solute transport modeling. In: Corwin, D.L., Loague, K. (Eds.), Applications of GIS to the Modeling of Non-Point Source Pollutants in the Vadose Zone. SSSA Special Publication Number 48. SSSA, Madison, pp. 153–162.

Bierkens, M.F.P., van der Gaast, J.W.J., 1998. Upscaling hydraulic conductivity: theory and examples from geohydrological studies. Nutrient Cycling in Agroecosystems 50, 193–207.

Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling: a review. Hydrological Processes 9, 251–290.

Bosma, W.J.P., Marinussen, M.P.J.C., van der Zee, S.E.A.T.M., 1994. Simulation and areal interpolation of reactive solute transport. Geoderma 62, 217–231.

Bouma, J., 1986. Using soil survey information to characterize the soil–water state. Journal of Soil Science 37, 1–7.

Brus, D.J., 1994. Improving design-based estimation of spatial means by soil map stratification. A case study of phosphate saturation. Geoderma 62, 233–246.

Brus, D.J., de Gruijter, J.J., 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma 80, 1–59.

Burrough, P.A., 1992. Development of intelligent geographical information systems. International Journal of GIS 6, 1–11.

Corwin, D.L., 1996. GIS applications of deterministic solute transport models for regional-scale assessment of non-point source pollutants in the vadose zone. In: Corwin, D.L., Loague, K. (Eds.), Applications of GIS to the Modeling of Non-point Source Pollutants in the Vadose Zone. SSSA Special Publication Number 48. SSSA, Madison, pp. 69–100.

de Gruijter, J.J., Ter Braak, C.J.F., 1990. Model-free estimation from spatial samples: a reappraisal of classical sampling theory. Mathematical Geology 22, 407–415.

de Gruijter, J.J., Walvoort, D.J.J., van Gaans, P.F.M., 1997. Continuous soil maps—a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma 77, 169–195.

de Vries, W., Kros, J., van der Salm, C., Groenenberg, J.E., Reinds, G.J., 1998. The use of upscaling procedures in the application of soil acidification models at different spatial scales. Nutrient Cycling in Agroecosystems 50, 223–236.

Finke, P.A., Bouma, J., Hoosbeek, M.R. (Eds.), 1998. Soil and Water Quality at Different Scales. Kluwer, Dordrecht, 324 pp.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford Univ. Press, New York, 483 pp.

Groenenberg, B.-J., Kros, J., van der Salm, C., de Vries, W., 1995. Application of the model NUCSAM to the Solling spruce site. Ecological Modelling 83, 97–107.

Groot, J.J.R., Penning de Vries, F.W.T., Uithol, P.W.J., 1998. Food supply capacity study at global scale. Nutrient Cycling in Agroecosystems 50, 181–189.

Heuvelink, G.B.M., 1998a. Uncertainty analysis in environmental modelling under a change of spatial scale. Nutrient Cycling in Agroecosystems 50, 255–264.

Heuvelink, G.B.M., 1998b. Error Propagation in Environmental Modelling with GIS. Taylor and Francis, London, 127 pp.

Heuvelink, G.B.M., Burrough, P.A., Stein, A., 1989. Propagation of errors in spatial modelling with GIS. International Journal of GIS 3, 303–322.

Hoosbeek, M.R., Bryant, R.B., 1992. Towards the quantitative modeling of pedogenesis—a review. Geoderma 55, 183–210.

Hosang, J., 1993. Field-scale solute transport—spatially interpolating the parameters of a transfer function model. Geoderma 60, 119–133.

Jansen, M.J.W., 1998. Prediction error through modelling concepts and uncertainty from basic data. Nutrient Cycling in Agroecosystems 50, 247–253.

Journel, A.G., Huijbregts, Ch.J., 1978. Mining Geostatistics. Academic Press, London, 600 pp.

Konikow, L.F., Bredehoeft, J.D., 1992. Ground-water models cannot be validated. Advances in Water Resources 15, 75–83.

Pebesma, E.J., de Kwaadsteniet, J.W., 1997. Mapping groundwater quality in the Netherlands. Journal of Hydrology 200, 364–386.

Pebesma, E.J., Wesseling, C.G., 1998. Gstat, a program for geostatistical modelling, prediction and simulation. Computers and Geosciences 24, 17–31.

Pebesma, E.J., Kros, J., Heuvelink, G.B.M., in preparation. Uncertainty in modelling soil acidification at a European scale.

Petach, M.C., Wagenet, R.J., DeGloria, S.D., 1991. Regional water flow and pesticide leaching using simulations with spatially distributed data. Geoderma 48, 245–269.

Sinowski, W., Scheinost, A.C., Auerswald, K., 1997. Regionalization of soil water retention curves in a highly variable soilscape: II. Comparison of regionalization procedures using a pedotransfer function. Geoderma 78, 145–159.

Stein, A., Staritsky, I.G., Bouma, J., van Eijnsbergen, A.C., Bregt, A.K., 1991. Simulation of moisture deficits and areal interpolation by universal cokriging. Water Resources Research 27, 1963–1973.

Stoorvogel, J.J., Smaling, E.M.A., 1998. Research on soil fertility decline in tropical environments: integration of spatial scales. Nutrient Cycling in Agroecosystems 50, 151–158.

Tiktak, A., Alkemade, J.R.M., van Grinsven, J.J.M., Makaske, G.B., 1998. Modelling cadmium accumulation at a regional scale in the Netherlands. Nutrient Cycling in Agroecosystems 50, 209–222.

Wackernagel, H., 1994. Cokriging versus kriging in regionalized multivariate data analysis. Geoderma 62, 83–92.

Wagenet, R.J., Hutson, J.L., 1989. LEACHM Leaching Estimation and Chemistry Model. A process-based model of water and solute movements, transformations, plant uptake and chemical reactions in the unsaturated zone. Continuum Vol. 2 (Version 2.0). Water Resources Institute, Center for Environmental Research, Cornell University.

Wagenet, R.J., Hutson, J.L., 1996. Scale-dependency of solute transport modeling/GIS applications. Journal of Environmental Quality 25, 499–510.

Wallach, D., Goffinet, B., 1987. Mean squared error of prediction in models for studying ecological and agronomic systems. Biometrics 43, 561–573.

Wen, X.H., Gómez-Hernández, J.J., 1996. Upscaling hydraulic conductivities in heterogeneous media: an overview. Journal of Hydrology 183, ix–xxxii.

Wösten, J.H.M., van Genuchten, M.Th., 1988. Using texture and other soil properties to predict the unsaturated soil hydraulic functions. Soil Science Society of America Journal 52, 1762–1770.

Wösten, J.H.M., Finke, P.A., Jansen, M.J.W., 1995. Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. Geoderma 66, 227–237.