Nutrient Cycling in Agroecosystems **50:** 255–264, 1998.

Uncertainty analysis in environmental modelling under a change of spatial scale

Gerard B.M. Heuvelink
The Netherlands Centre for Geo-ecological Research, Landscape and Environmental Research Group, University of Amsterdam, Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, The Netherlands

Key words: environmental modelling, geostatistics, uncertainty analysis, upscaling, change of support, error balance, validation

Abstract

Although environmental processes at large scales are to a great degree the resultant of processes at smaller scales, models representing these processes can vary considerably from scale to scale. There are three main reasons for this. Firstly, different processes dominate at different scales, and so different processes are ignored in the simplification step of the model development. Secondly, input data are often absent or of a much lower quality at larger scales, which results in a tendency to use simpler, empirical models at the larger scale. Third, the support of the inputs and outputs of a model changes with change of scale, and this affects the relationships between them. Given these reasons for using different models at different scales, application of a model developed at a specific scale to a larger scale should be treated with care. Instead, models should be modified to suit the larger scale, and for this purpose uncertainty analyses can be extremely helpful. If upscaling disturbed the balance between the contributions of input and model error to the output error, then an uncertainty analysis will show this. Uncertainty analysis will also show how to restore the balance. In practice, application of uncertainty analysis is severely hampered by difficulties in the assessment of input and model error. Knowledge of the short distance spatial variability is of paramount importance to input error assessment with a change of support, but current geographical databases rarely convey this type of information. Model error can only be estimated reliably by validation, but this is not easy because the support of model predictions and validation measurements is usually not the same.

Introduction

Much of the research effort in the environmental sciences is spent on the development of (quantitative) models. The purpose of model building is usually twofold, firstly to assist in the understanding of physical systems by providing a framework within which to analyze data, and secondly to provide a predictive tool [19]. Although the same model may be used for both purposes, one should be very aware that the second purpose puts a much higher claim on the validity of the model [19, 53]. A model can be conveniently defined as a simplified representation of the real world. Adding the adjective 'simplified' is meaningful here because it emphasizes that modelling is an activity in which one *purposely* introduces simplifications to the real world [17]. Not only is it impossible to build a model that is an exact copy of the real world (e.g., consider the example of Newtonian mechanics in [27]), but it is also naive and even illusory to assert that this should be the ultimate goal of the modeller. Instead, the modeller's goal is to build a model in which a well-considered trade-off has been made between validity and representativeness on the one hand and degree of complexity, transparency and manageability on the other.

Since a model is only an approximation of reality, and also because the inputs to the model are rarely, if ever, exactly known, the output of the model is also likely to deviate from reality. In other words, the errors or uncertainties that are contained in the model and its inputs will *propagate* to the model output. Clearly, it is important to know how large the uncertainties in the model output are, particularly when the model is used for predictive purposes. In recent years much attention has been paid to the propagation of errors in envi-

ronmental modelling, and several techniques are now readily available to carry out this type of uncertainty analysis [22, 27, 30, 61].

Another important issue in the environmental sciences, and certainly as challenging as the handling of uncertainty, concerns the issue of scale. Many of the processes that we study and for which we build models, take place at various spatial and temporal scales [25]. This workshop presented us with many examples, such as erosion [35], salinisation [45], nitrate leaching [9, 54, 59], soil nutrient depletion [51] and groundwater movement [4]. It is obvious that processes at different scales interact, but as yet "no one discipline has completely resolved the problem of translation of knowledge from smaller scales to larger, or the reverse" [56]. This is a serious problem, because it often happens that models are developed at a scale which is much smaller than the scale of relevant applications, so that some form of upscaling is required [4, 14, 26, 39].

The purpose of this review paper is to address the problem of environmental modelling at different spatial (and temporal) scales, including the accuracy assessment of model predictions at these various scales, by summarizing and commenting on the approaches and solutions offered in the papers presented at this workshop.

Model development at different spatial scales

At some stage in the development of a model describing a physical process or phenomenon, the variables that are considered relevant for the particular problem are defined and the relations between these variables are quantified, either based on physical laws or on empirical evidence. The choices that are made at this stage are crucial for the model that is finally obtained, and an important question is whether these choices would turn out differently if the same process were modelled at a different scale. If this were indeed the case, then the model would be scale-specific, and the application of the model at a different scale would have to be treated with caution.

Many papers in this workshop argue that models are indeed scale-specific, for three principal reasons: (1) different processes are important at different scales, (2) the input data availability is reduced at larger scales, and (3) the model input and model output undergo a change of support.

Relative importance of processes

The first reason for arriving at different models at different scales is that "different processes are important at different levels" [14]. This affects the modelling, because modelling focusses on dominant processes and ignores less important ones. For instance, at the landscape scale, soil moisture is controlled by soil texture, but at the regional scale the focus is on geologic and geomorphic factors [56]. In [56], it is also observed that "pesticide sorption is a combination of equilibrium and kinetic processes, but kinetic processes are not often included in current models because they occur over a short scale of space". On a temporal scale, seasonal dynamics will be included in soil acidification models operating on a shorter time scale, but not in models operating on a large temporal scale [11]. In soil solute transport modelling, the weathering of minerals will only become a meaningful contributory process when the time scale is sufficiently large.

A word of warning should be raised here as well, because too often we will tend to believe that a certain process is important only at the scale at which it operates, whereas in reality it may well have a profound influence at other scales as well. For instance, macropore and preferential flow take place at the pedon level [48], but they remain important contributory processes at the field and region level. Similarly, photosynthesis takes place at the scale of leaves, but clearly this does not undermine its importance at the scale of forests and crops.

Reduced input data availability at the larger scale

The second reason for ending up with different models at different scales has to do with input data availability. At the small scale, data are often available through measurements. At the large scale, measurements are no longer available but inputs have to be derived from general information sources, such as general purpose soil maps, agricultural statistics or expert judgments [20, 51, 59]. Obtaining the model inputs from general information sources usually involves some kind of transformation, such as achieved by a pedo-transfer function [26, 51, 52, 63]. The use of transfer functions inevitably causes a deterioration in the quality of the input data, and this has led many model developers to simplify their models when moving from smaller to larger scales [11]. The reason is simple: what sense does it make to include sub-processes if they require

information that is either absent or very unsure? Or, as [35] puts it: "Upscaling to larger areas invariably means a loss in the precision and observation density of data used to parametrize a model. It also raises questions about the suitability of applying the model at a scale different from the one for which it was developed".

The workshop contains many examples of adjusting a model to the reduced input data availability. In [52] it is explicitly stated that "a requirement for a regional scale model is that all model parameters can be derived from commonly available soil characteristics", and the authors fulfil this requirement by building the SOACAS model, which is simpler than competitive models for tracing metal behaviour and transport in soil. Aggregation of the soil nutrient depletion model from regions to subcontinents is done using a generalized calculation based on national soil, climate and landuse data bases [51]. In [11], moving from a regional (RESAM) to a national and continental (SMART) scale, the degree of process aggregation is increased (i.e. the complexity of the model is decreased) in response to the decrease of data availability. Because necessary data are lacking at the regional scale, the elementary overlay-based erosion model in [35] is further simplified to a multiplication-based model, but "more complex models could be used where better data are available".

At this point it is useful to note that in some situations it may be very difficult to satisfy the rule that modelling should include the dominant processes and that it should also exclude processes that cannot be parametrized due to lack of adequate data. Typical examples are macropore and preferential flow in soil, which are clearly important, also at larger scales, but very hard to parametrize from general-purpose soil survey data [9, 48, 50]. Another example is given in [52], where it is recognized that ignoring the effect of city compost is a likely cause for underestimation of cadmium loads, but where this effect could not be accounted for because quantitative information on use of city compost was not available.

Change of support of model entities

The third reason for modifying a model under a change of scale is that the *support* of the model entities changes. Here ‘support’ refers to the size, shape and orientation of these entities [8, 58]. The concept of support has much in common with ‘level of aggregation’ [14] and ‘sample volume’ [56]. Moving up from the small scale 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’, i.e. the support changes. For instance, the SOACAS model used in [52] takes as inputs atmospheric deposition, fertilizers and animal manure as values averaged over 500×500 m^2 grid cells and it produces as output the grid cell averaged cadmium content.

A change of support may require an adaptation of the model because relations between variables that exist at the point support need not extend to the block support. A clear example is Darcy flow in two and three dimensions, where it has been demonstrated that when Darcy’s law is assumed valid at the measurement scale, then it cannot be valid at the block scale [3 page 65]. To illustrate the effect of a change of support, consider the example of nitrate degradation in soil, which can be described as a function of the nitrogen surpluses, the average field capacity and the site specific denitrification conditions [59]. Due to the non-linearities that are contained in the model, it is unlikely that applying the model to 3×3 km^2 grid cell averaged inputs yields the same result as averaging the model outputs computed at all points within the grid cell. The change of support effect also turns up in the simplest non-linear operations. For instance, multiplying average water flow by average solute concentration is unlikely to produce the average solute flow in a situation of spatial (or temporal) heterogeneity. Whether or not this is a substantial effect depends on the degree of heterogeneity. In some cases, the effects of spatial heterogeneity can indeed be quite large [48], but there are also cases where averaging did not cause too great a deterioration in the results [32].

Groundwater modelling is one research area where one is well aware of the effect of a change of support on the relationships between model entities. Although moving up from the core scale to the local scale is not (yet) accompanied by an adjustment of model structure, the model parameters (such as hydraulic conductivity and dispersivity) do change, both in value and interpretation. In [4] it is shown how difficult the upscaling of hydraulic conductivity really is, particularly because it turns out that block conductivity is dependent on flow geometry. Model structure itself is sometimes also questioned. In the area of subsurface hydrology, it has been asserted that "it is merely assumed that the *same* small scale physical equations can be applied at the model grid scale with the *same* parameters. In doing so we make a conceptual leap" [1] and that "direct application of small-scale par-

tial differential equations to grid-scale processes gives rise to fundamental problems due to the inherent spatial variability" [31]. These problems have also been recognized in this workshop, when it was stated that "a point of concern is the use of average soil physical and soil chemical characteristics within one grid cell" [52] and "the structure of the water balance model and its parameters are oversimplifications, but if we wish to use models at a large scale on a grid with validly averaged parameters we should use models that are linear with respect to their parameters" [20]. Perhaps we should not so rigorously discard non-linear models at the larger scale, but the effect of a change of support should always be considered, including the case of application of a crop model developed for the field scale at spatial units the size of $1^\circ \times 1^\circ$ [20].

A change of scale will not always be accompanied by a change of support. For instance, in [54] a soil freezing model is applied to predict the number of frost tillable days per winter at 275 weather observation sites, using input data from these sites. In that case, upscaling in fact reduces to spatial generalization (or spatial extrapolation), because the model entities retain the same support. It is important to realize that prediction at points is fundamentally different from prediction at blocks, and therefore one should always state the support when reporting results [58 page 29].

Uncertainty analysis as a tool in upscaling environmental models

The previous section gave three main reasons why the majority of environmental models are scale-specific. The immediate consequence is that models developed at the small research scale must be adjusted to the larger scale of relevant applications. But how should such an adjustment be done? In principle it seems right to base the adjustment on the fact that processes at the larger scale are the resultant of processes at smaller scales, in other words, to make use of the idea of hierarchical systems where lower levels are part of and confined by processes at higher levels, and to take a more holistic perspective [5, 14, 56]. However, we should not forget that models are no more than simplified representations of reality, where unimportant processes are purposely ignored. Since different processes are important at different scales, it may happen that the resultant of processes at the smaller scale is relatively unimportant at the larger scale. In addition, quantification of the interaction of processes at different scales is not an easy task.

The holistic and hierarchical approach are certainly very useful as methodologies to enhance understanding of how processes at one scale affect processes at other scales, but when it comes to building predictive models that should be efficient and economic, the common pragmatic approach may still be preferred. This means that to develop or modify a model for a given scale, the dominant processes, the support and the data availability at that scale are taken as starting points. Admittedly, this yields different models at different scales and it obscures the relations that exist between processes at different scales, but it does not undermine the validity of the model at the scale for which it was built.

Unfortunately, a model will not always be sufficiently adjusted to the change of scale. Many modellers still tend to leave (parts of) their model unchanged when moving to a larger scale. This manifests itself in a tendency to use overly complex models [55]. In such cases, reduced data availability has been insufficiently incorporated in the model adjustment. It seems that therefore, there is a need for tools that can assist the modeller in making the right simplification steps.

One such a tool is uncertainty analysis. Although the main purpose of uncertainty analysis (also called error propagation analysis) is to evaluate the accuracy of a certain model at a given scale [27], it can also prove valuable in deciding in what way and to what degree to adjust or simplify a given model to the larger scale [11, 22]. But before addressing how an uncertainty analysis may be used for this purpose, let me first briefly review the main principles of an error propagation analysis.

Summary of error propagation techniques

In a crude mathematical notation, the error propagation problem can be formulated as follows. Let u be the output of a model g that operates on m inputs a_i (i=1,...,m):

$$u = g(a_1, a_2, \ldots, a_m) \qquad (1)$$

The inputs a_i may be scalars, but in many cases they will be temporally and/or spatially distributed. The inputs to the model have been measured, estimated, classified or interpolated, and so in almost all practical cases they will contain a certain amount of error. The aim of an error propagation analysis is to determine how large the error in the output u is, given the errors in the inputs a_i. This is done by making the inputs ran-

domly distributed, i.e. by defining a joint probability distribution for the inputs. Usually, the standard deviation of an individual input is interpreted as the main parameter representing its uncertainty.

Ideally, an error propagation analysis should also include the model error, because even if the inputs were exactly known, then the model output would still be in error because the model itself is only a simplified representation of reality. Model error can be included by making model parameters randomly distributed or by adding a residual noise term to the model.

Analytical solutions to the error propagation problem exist only in a few special cases, such as when g is linear. For the general situation alternative techniques have to be used. Two methods will now be briefly discussed. A more detailed account is given in [27]. The idea of the *Taylor method* is to approximate g by a linear function, after which the error propagation can be analytically solved [12, 22]. The *Monte Carlo method* [38] uses an entirely different approach. The idea of the method is to compute the result of $g(a_1,...,a_m)$ repeatedly, with input values a_i that are randomly sampled from their joint distribution. The variability in the model outputs then is a measure of the output uncertainty. A random sample from the m inputs is obtained using an appropriate pseudo random number generator. Application of the Monte Carlo method to spatially distributed inputs requires the simultaneous generation of realizations from random fields. Various techniques can be used for this purpose, an attractive one being the sequential Gaussian simulation algorithm [13]. In [4], an analogous indicator simulation technique is used to generate realizations of the three-dimensional texture class distribution.

The main problem with the Taylor method is that the results are only approximate. It is far from easy to determine whether the approximations involved are acceptable, especially when g is a complex model. The Monte Carlo method does not suffer from this problem, because in principle it can reach an arbitrary level of accuracy. But high accuracies are reached only when the number of runs is sufficiently large, which may cause the method to become extremely time consuming. Another disadvantage of the Monte Carlo method is that the results do not come in an analytical form. Despite these disadvantages, in the environmental sciences Monte Carlo simulation is by far the most popular tool for tracing the propagation of errors [27], because it is also transparent, easily implemented and generally applicable.

The balance of errors

The contribution of individual error sources can be obtained by utilizing the so-called *partitioning property*, which says that the variance of the output error is approximately equal to a sum of contributions, each of which is attributable to the error of an individual input [22, 24]. This can best be demonstrated for the situation in which the input errors are uncorrelated and when the Taylor method is applied. In that case, the variance in the output u is simply given by:

$$\mathrm{Var}(u) \approx \sum_{i=1}^{m} \mathrm{Var}(a_i) \cdot (\frac{\partial g}{\partial a_i})^2 \qquad (2)$$

where $\mathrm{Var}(a_i)$ is the variance of the error in a_i and where $(\partial g/\partial a_i)$ is the partial derivative of g with respect to a_i. The partial derivative conveys the sensitivity of the model output to a change in the input. Note that equation (2) is only approximately valid, but that it can be improved by a partitioning method based on stochastic simulation [28].

The consequence of the partitioning property is that it allows one to make rational decisions on how to reduce output error. Clearly the largest reduction in output error can be achieved by reducing the error of the input that has the largest error contribution. Moreover, it is possible to calculate beforehand how much the output error reduces from the reduced input error, allowing a rational comparison of different strategies.

The partitioning property can also be used to compare the contributions of input and model error, provided the model error is quantified through randomization of its parameters or inclusion of a stochastic residual. It is clearly unwise to spend much effort on collecting data if what is gained is immediately thrown away by using a poor model. On the other hand, a simple model may be as good as a complex model if the latter needs lots of data that cannot be accurately obtained [22]. As a general rule, it is thus best to strive for a balance of errors.

Upscaling may seriously disturb the balance of errors. One important reason is that input error increases due to the reduced data availability at the larger scale. This implies that the model becomes too complex for the larger scale. Comparison of sixteen forest–soil–atmosphere models showed that complex models were not able to reproduce the field observations better than more simple models [55], and from this the authors concluded that there is a tendency to use overly complex and unbalanced models. In soil acidifica-

tion modelling, these considerations have led to the development of simpler models at larger scales [11].

Applications of uncertainty analysis

It must be said that only a few papers presented at this workshop employ an uncertainty analysis, and none of them gives a rigorous treatment. One paper executes a min-max analysis to get a worst case impression on the uncertainty in the model output from the uncertainty in organic matter [52], another performs a sensitivity analysis to space and time resolution on the TOPMODEL and a Monte Carlo analysis on the mixing model [9]. In the second case study, [4] uses stochastic upscaling in combination with stochastic spatial simulation to obtain confidence limits for the C-value of the entire confining layer. Comparison of results of the erosion model using data from the regional database and using more precise data is shown in [35].

Recent examples of more elaborate applications of uncertainty analysis in environmental modelling are [10, 15, 33, 36, 37, 40, 47, 62]. Applications of error propagation in a GIS-context are given in [16, 21, 22, 24, 49].

Assessment of input and model error

Perhaps the main obstacle against a routine application of uncertainty analysis in environmental modelling is the assessment of input and model error [23, 27]. This section addresses the problem of input and model error assessment in more detail, again with emphasis on scale-related issues.

Input error assessment

Upscaling affects the identification of input error in two meaningful ways.

Firstly, input error increases when upscaling is accompanied by a reduction of data availability. This has already been discussed and illustrated in a previous section. When the same model is used at the larger scale, the majority (or all) of the model inputs will have to be derived from general information sources, possibly through the use of transfer functions. It will be clear that because of this the inputs become less accurate, but quantification of the error remains difficult [27]. For instance, it is not yet common practice to accompany general purpose maps stored in a GIS by accuracy measures, although efforts are made to improve the current situation [7, 18, 44]. In addition, the errors inflicted by the transfer function are also rarely known [27].

Secondly, when a change of scale includes a change of support, then this must be included in the error assessment. This is because input error is support-dependent (an immediate consequence of this is that the results of an uncertainty analysis are meaningless if the supports of the model inputs are not the same). As an example, consider the nitrate leaching model used in [59], which uses the average precipitation within 3×3 km^2 cells as input. To carry out an uncertainty analysis in this case, the error associated with the grid cell averaged precipitation is required. But how large is this error? This question cannot easily be answered, because part of the 'point' errors within the grid cell will average out, but exactly how much depends on the spatial variability of precipitation. In any case, the error of the grid cell average will be smaller than the error at any given point within the cell. Note also that the averaging-out effect is equally valid on the time scale: prediction of the average precipitation over longer time periods is 'easier' than prediction over shorter time periods.

In geostatistics, the averaging-out effect is contained in the so-called *regularization* theory [34]. This theory describes how the variogram of a spatial attribute changes under a change of support. The general result is that the sill of the variogram decreases as one moves from point to block support, and that this effect is stronger when the nugget variance is large. A decrease of the sill means that the spatial variability decreases, and this agrees with the observation that "coefficients of variability often decrease with higher levels of spatial aggregation" [14]. Knowledge of the short distance variability is crucial to determine the variability at the block support [58], and this may be one of the reasons why so many papers in this workshop deal with the assessment of (short distance) spatial variability [26, 39, 43, 46, 48, 50].

In a spatial interpolation context, the averaging out effect causes the block kriging variance to be smaller than the point kriging variance [26, 58]. Here it is useful to note that the mere fact that the block kriging variance is smaller than the point kriging variance can never be a reason for using it, but that the only sound reason for using block kriging is that values at block support are required [26]. Note also that regularization theory and block-kriging are usually restricted to linear averaging, but that there are also situations where the goal is non-linear averaging [42]. One particular non-

linear averaging application, i.e. the derivation of the block conductivity from the hydraulic conductivity at the core scale, is extensively discussed in [4].

Of course, a change of support need not always be directed upward, but can also be directed downward. For instance, the crop model in [20] requires precipitation data on a daily basis but the available data sources provide only monthly averages. This requires temporal downscaling. In fact the crop model requires spatial downscaling as well, when irrigation water availability is to be scaled down from catchment level to individual grid cells. Downscaling means that variability should be added instead of levelled out, and this is generally considered a difficult problem [14, 39]. In [20], temporal variability is added by randomly distributing the average monthly rainfall over the average monthly number of raindays, but this may still be too little added variability. In [26], it is assumed that the weather is uniform for the area studied, which is likely to introduce more spatial dependence than is actually the case.

Model error assessment

Model error contains all errors that result from the various assumptions, discretizations and simplifications that are made to make the model manageable. Unfortunately, it is seldomly easy to quantify the model error. One important reason is that model error may vary from application to application. For instance, the linear regression erosion model developed for the Nord-Pas-de-Calais region in [35] may perform much more badly when it is applied to the Mediterranean. When model development involves calibration of model parameters, then model error will almost certainly increase when the model is extrapolated to another area or to another time period. Note that calibration does offer possibilities to estimate the errors of the calibrated parameters [2, 29], but these error measures cannot be extrapolated to other applications either.

Therefore, unless one wants to make unverifiable assumptions about the portability of model error, it can be concluded that model error must be determined anew for each application. One method of model error assessment then is to compare the model results with results that are obtained using a much more detailed model. This so-called inter-model comparison [27] has indeed been used in this workshop [11, 35]. In principle the approach is viable, but it does presuppose that the error contained in the more detailed model is negligible and it must somehow filter out the contribution of input error. Note also that the idea of inter-model comparison can be extended to inter-data comparison, where information at different quality levels is used by the same model and the results are compared [35, 52, 54].

A more reliable method to assess the model error contribution is through validation, i.e. through the comparison of model predictions with independent measurements. When input and model error are statistically independent (and it is hard to think of a reason why they should not be), the following identity holds:

$$\begin{aligned}\text{Var(output error)} = \text{Var(due to input error)} + \\ + \text{Var(due to model error)} \quad (3)\end{aligned}$$

The left hand side of (3) can be determined by validation and the first term on the right hand side through an uncertainty analysis, using only the (known) input error as the source of error. Thus the contribution of model error can be computed by subtracting the latter from the former. This yields a single lumped model error, but it can be decomposed into parts by repeating the procedure for separate sub-processes. For instance, decomposition of model error is relatively easy when the model consists of a combination of different independent submodels, such as is the case with the nitrate pollution model in [59].

Although it sounds easy, there are several difficulties associated with the proposed procedure. The first is that it assumes that the input error is known. We have seen before that in practice this is not a very realistic assumption. The second problem is that it assumes that the variance of the output error can be determined exactly through validation. But in practice it can only yield an estimate of the output error, because validation is always based on a finite number of comparisons. In addition, validation often suffers from the fact that the support of the model predictions is not the same as the support of the measurements [39, 56]. The latter problem will now be discussed in more detail.

Validation of block predictions from point observations

Many models that operate on a large scale make model predictions on a support at which measurements cannot realistically be obtained. For instance, in this workshop this was the case with models of crop growth [20], groundwater movement [4], soil acidification [11] and heavy metal accumulation [52]. Direct comparison of block predictions with point measurements is not valid

because the point behaviour may importantly deviate from the block behaviour (see the initial experiment in [50] for a clear example). Such comparisons are unfair because a large scale model is not intended to be able to predict values at the point support. For instance, the supraregional nitrate flow model makes predictions for areal units of 3×3 km^2, and the "model approach does not claim to be suitable for site-related or small regional applications" [59].

A correct validation requires that the point measurements are scaled up to the block support before a comparison with model predictions is made. Let us restrict ourselves to linear upscaling. If multiple point measurements are collected within the same block using some form of probability sampling, then the block average may be estimated using a design-based approach [6]. A design-based approach has the important advantage that it makes very few assumptions, but the estimation error will be large when the number of measurements within the block is small. Although design-based methods offer possibilities to reduce the estimation accuracy by including additional information, model-based approaches are more equipped for this purpose. Examples of model-based approaches used in this workshop are block-kriging [26], which uses measurements at neighbouring locations (outside the block) as additional information, and the GAM model [52], which improves estimation through the use of external explanatory variables.

Upscaling point measurements to the block support implies that the true block average is not known but only estimated. As is remarked in [27], it is important to assess the accuracy of the estimate, because it must be included in the subsequent validation analysis. Let u^* be the model prediction of the true block averaged output u, and let $\hat{u}$ be the estimate of u based on the point validation measurements. Now the squared deviation of (u^*-$\hat{u}$) can be decomposed into:

$$(u^*-\hat{u})^2 = ((u^*-u)+(u-\hat{u}))^2 \approx (u-u^*)^2+(u-\hat{u})^2 \quad (4)$$

where the latter equality is approximate only because it ignores the cross-product of (u^*-u) and (u-$\hat{u}$). Equation (4) shows that evaluating a model on only (u^*-$\hat{u}$)^2 is unfair because it should really be judged on (u-u^*)^2, which gives a smaller value. A correction must be made, and this requires knowledge of the squared estimation error (u-$\hat{u}$)^2. Both design-based and model-based techniques allow to estimate the squared estimation error, but in environmental modelling practice this matter and its implications for validation seem not yet to have been used to its full potential.

Thorough treatments on partitioning the mean squared error of prediction are given in [41, 57, 60]. One particularly interesting result addressed in great detail in [57] is that comparison of the two terms on the right hand side of (4) gives insight into whether sufficient effort has been spent on collecting validation data.

Discussion and conclusions

Many of the models used in the environmental sciences are scale-specific. This is not surprising, because there are sound reasons that cause modellers to take different decisions at different scales. Clearly, scale-specific models should not be applied uncritically to scales for which they were not developed, but in practice, this rule is often violated. It simply is very tempting to apply an existing model to a larger scale, because that is where the relevant applications are.

One efficient way to prevent users from applying models to scales for which they were not developed is to offer them a set of alternative models, one for each scale of application. Such an approach requires that plot scale models are modified to make them suitable for larger scales. Uncertainty analysis can be used as an aid in the model adjustment.

Although uncertainty analyses are definitely very useful in adjusting models to the larger scale, several problems exist. The problems lie not so much in the error propagation analysis itself, but much more in the assessment of input and model error. At present, many data sources do not provide information about the accuracy of the data they contain. However, when given sufficient priority, we should be able to let the data and transfer functions that are stored in common data bases be routinely accompanied by accuracy measures. With regard to model error, more work is needed to develop a standard methodology for the assessment of model error through validation exercises.

One aspect that is often ignored, but that is of crucial importance in upscaling studies and uncertainty analyses, concerns the issue of support. The support has such an important influence on the assessment of input error, on the relations between model entities and on the results of validation that we cannot allow ourselves to be ambiguous about it. It is imperative that the support is always stated when reporting results,

particularly in studies making accuracy statements or involving a change of scale.

References

1. Beven K (1989) Changing ideas in hydrology - the case of physically-based models. Journal of Hydrology 105: 157–172
2. Beven K and Binley A (1992) The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6: 279–298
3. Bierkens MFP (1994) Complex confining layers: a stochastic analysis of hydraulic properties at various scales. PhD thesis. Netherlands Geographical Studies 184
4. Bierkens MFP and Van Der Gaast JWJ (1997) Upscaling hydraulic conductivity; theory and examples from geohydrological studies (this issue)
5. Blöschl G and Sivaplan M (1995) Scale issues in hydrological modelling: a review. Hydrological Processes 9: 251–290
6. Brus DJ and De Gruijter JJ (1993) Design-based versus model-based estimates of spatial means: theory and application in environmental soil science. Environmetrics 4: 123–152
7. Burrough PA (1992) Development of intelligent geographical information systems. International Journal of GIS 6: 1–11
8. Cressie NAC (1991) Statistics for Spatial Data. New York: Wiley
9. Curmi P Durand P Gascuel-Odoux C Mérot P and Walter C (1997) Hydromorphic soils and the control of nitrate fluxes: spatial distribution and functional modeling at different scales (this issue)
10. De Roo APJ Hazelhoff L and Heuvelink GBM (1992) Estimating the effects of spatial variability of infiltration on the output of a distributed runoff and soil erosion model using Monte Carlo methods. Hydrological Processes 6: 127–143
11. De Vries W Kros J Groenenberg JE Reinds GJ and Van Der Salm C (1997) The use of upscaling procedures in the application of soil acidification models at different spatial scales (this issue)
12. Dettinger MD and Wilson JL (1981) First order analysis of uncertainty in numerical models of groundwater flow. Part I. Mathematical development. Water Resources Research 17: 149–161
13. Deutsch CV and Journel AG (1992) GSLIB: Geostatistical Software Library and User's Guide. New York: Oxford University Press
14. Dumanski J Pettapiece WW MacGregor RJ (1997) Relevance of scale dependent approaches for integrating biophysical and socio-economic information and development of agroecological indicators (this issue)
15. Finke PA Wösten JHM and Jansen MJW (1996) Effects of uncertainty in major input variables on simulated functional soil behaviour. Hydrological Processes 10: 661–669
16. Fisher PF (1991) First experiments in viewshed uncertainty: simulating fuzzy viewsheds. International Journal of GIS 5: 193–208
17. Gilchrist W (1984) Statistical Modelling. Chichester: Wiley
18. Goodchild MF and Gopal S (1989) Accuracy of Spatial Databases. London: Taylor & Francis
19. Grayson RB Moore ID and McMahon TA (1992) Physically based hydrologic modeling. 2. Is the concept realistic? Water Resources Research 28: 2659–2666
20. Groot JJR Penning De Vries FWT and Uithol PWJ (1997) Food supply capacity study at global scale (this issue)
21. Haining R and Arbia G (1993) Error propagation through map operations. Technometrics 35: 293–305
22. Heuvelink GBM (1993) Error propagation in quantitative spatial modelling: applications in geographical information systems. PhD thesis. Netherlands Geographical Studies 163
23. Heuvelink GBM (1996) Identification of field attribute error under different models of spatial variation. International Journal of GIS 10: 921–935
24. Heuvelink GBM Burrough PA and Stein A (1989) Propagation of errors in spatial modelling with GIS. International Journal of GIS 3: 303–322
25. Hoosbeek MR and Bryant RB (1992) Towards the quantitative modeling of pedogenesis - a review. Geoderma 55: 183–210
26. Hoosbeek MR Mulder M Booltink H and Bouma J (1997) Spatial and temporal variability of calculated soil quality variables (this issue)
27. Jansen MJW (1997) Prediction error through modelling concepts and uncertainty from basic data (this issue)
28. Jansen MJW Rossing WAH and Daamen RA (1994) Monte Carlo estimation of uncertainty contributions from several independent multivariate sources. In: Grasman J and Van Straten G (eds.) Predictability and Nonlinear Modelling in Natural Sciences and Economics., pp. 334–343. Dordrecht: Kluwer
29. Janssen PHM and Heuberger PSC (1995), Calibration of process-oriented models. Ecological Modelling 83: 55–66
30. Janssen PHM Heuberger PSC and Klepper O (1994) UNCSAM: a tool for automating sensitivity and uncertainty analysis. Environ. Software 9: 1–11
31. Jensen KH and Mantoglou A (1992) Future of distributed modelling. Hydrological Processes 6: 255–264
32. Jensen KH and Refsgaard JC (1991) Spatial variability of physical parameters and processes in two field soils. Part II: water flow at field scale. Nordic Hydrology 22: 303–326
33. Jones L (1989) Some results comparing Monte Carlo simulation and first order Taylor series approximation for steady groundwater flow. Stochastic Hydrology and Hydraulics 3: 179–190
34. Journel AG and Huijbregts ChJ (1989) Mining Geostatistics. London: Academic Press
35. King D Fox DM Daroussin J Le Bissonnais Y and Danneels V (1997) Upscaling a simple erosion model from small areas to a large region (this issue)
36. Kros J De Vries W Janssen PHM and Bak CI (1992) The uncertainty in forecasting trends of forest soil acidification. Water, Air and Soil Pollution 66: 29–58
37. Leenhardt D (1995) Errors in the estimation of soil water properties and their propagation through a hydrological model. Soil Use and Management 11: 15–21
38. Lewis PAW and Orav EJ (1989) Simulation Methodology for Statisticians, Operations Analysts, and Engineers, Volume I. Pacific Grove: Wadsworth & Brooks/Cole
39. Lischeid G Alewell C Bittersohl J Göttlein A Jungnickel C Lange H Manderscheid B Moritz K Ostendorf B and Sager H (1997) Investigating the impact of anthropogenic depositions on plants, soil and groundwater at different scales in a forested catchment: the Waldstein case study (this issue)
40. Loague K Yost RS Green RE and Liang TC (1989) Uncertainty in pesticide leaching assessment in Hawaii. Journal of Contaminant Hydrology 4: 139–161
41. Luis SJ and McLaughlin D (1992) A stochastic approach to model validation. Advances in Water Research 15: 15–32

42. McBratney AB (1997) Some methodologies for the aggregation and disaggregation of soil information (this issue)
43. Meshalkina JL Belousova NI and Vasenev II (1997) Soil absorbing complex properties of Russian boreal soils and its dependence on the spatial scale of study area (this issue)
44. Mowrer HT Czaplewski RL and Hamre RH (1996) Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Fort Collins: USDA Forest Service, Report RM-GTR-277
45. Odeh IOA Todd AJ Triantafilis J and McBratney AB (1997) Status and trends of soil salinity at different scales: the case for the irrigated cotton growing areas of eastern Australia (this issue)
46. Radcliffe DE Gupte SM and Box JE (1997) Solute transport at the pedon and polypedon scales (this issue)
47. Scavia D Powers WF Canale RP and Moody JL (1981) Comparison of first-order error analysis and Monte Carlo simulation in time-dependent lake eutrophication models. Water Resources Research 17: 1051–1059
48. Simmonds LP and Nortcliff S (1997) Small scale variability in the flow of water and solutes, and implications for lysimeter studies of solute leaching (this issue)
49. Stanislawski LV Dewitt BA and Shrestha RL (1996) Estimating positional accuracy of data layers within a GIS through error propagation. Photogrammetric Engineering & Remote Sensing 62: 429–433
50. Steenhuis T Vandenheuvel K Weiler KW Boll J Daliparthy J Herbert S and Kung K-JS (1997) Mapping and interpreting soil textural layers to assess agri-chemical movement at several scales along the eastern seaboard (USA) (this issue)
51. Stoorvogel JJ and Smaling EMA (1997) Research on soil fertility decline in tropical environments: integration of spatial scales (this issue)
52. Tiktak A Alkemade R Van Grinsven H and Makaske K (1997) Modelling cadmium accumulation on a regional scale in the Netherlands (this issue)
53. Tiktak A and Van Grinsven HJM (1995) Review of sixteen forest-soil-atmosphere models. Ecological Modelling 83: 35–53
54. Van Es HM DeGaetano AT Wilks DS and Mikhailova EA (1997) Upscaling plot-based research information: frost tillage and nitrate leaching (this issue)
55. Van Grinsven HJM Driscoll CT and Tiktak A (1995) Workshop on comparison of forest-soil-atmosphere models: preface. Ecological Modelling 83: 1–6
56. Wagenet RJ (1997) Scale issues in agroecological research chains (this issue)
57. Wallach D and Goffinet B (1987) Mean squared error of prediction in models for studying ecological and agronomic systems. Biometrics 43: 561–573
58. Webster R and Oliver MA (1990) Statistical Methods in Soil and Land Resource Survey. Oxford: Oxford University Press
59. Wendland F Bach M and Kunkel R (1997) The influence of nitrate reduction strategies on the temporal development of the nitrate pollution of soil and groundwater throughout Germany - a regionally differentiated case study (this issue)
60. Whitmore AP (1991) A method for assessing the goodness of computer simulation of soil processes. Journal of Soil Science 42: 289–299
61. Wingle WL Poeter EP and McKenna SA (1996) UNCERT: a geostatistical uncertainty analysis package applied to groundwater and contaminant transport modeling. http://www.mines.edu/fs_home/wwingle/uncert/
62. Woldt W Goderya F Dahab M and Bogardi I (1996) Consideration of spatial variability in the management of non-point source pollution to groundwater. In: Mowrer HT Czaplewski RL and Hamre RH (eds.) Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, pp. 49–56. Fort Collins: USDA Forest Service, Report RM-GTR-277
63. Wösten JHM Finke PA and Jansen MJW (1995) Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. Geoderma 66: 227–237