Upscaling and Downscaling Methods for Environmental Research

Article · January 2001

Source: OAI

CITATIONS
257

READS
7,671

3 authors:

M.F.P. Bierkens
Utrecht University
349 PUBLICATIONS **29,559** CITATIONS
SEE PROFILE

Peter Finke
Ghent University
217 PUBLICATIONS **5,203** CITATIONS
SEE PROFILE

P. de Willigen
Wageningen University & Research
93 PUBLICATIONS **2,602** CITATIONS
SEE PROFILE

Contents

Preface

Environmental studies typically involve the combination of several dynamic models with many different data sources. The spatial and temporal scales at which these models operate is often very different from each other. Also, the scale at which available data have been collected is usually different from the scale required by the models. Moreover, the scale of the model output is rarely in tune with the scale at which decision makers require answers or implement environmental measures. Clearly, much environmental research suffers from a severe discrepancy of scale between data sources, models used and questions asked. To overcome this scale discrepancy a vast body of different methods has been developed to transfer data and model output from a smaller to a larger scale (upscaling or aggregation) or from a larger to a smaller scale (downscaling or disaggregation). Although many conferences have been organised on the subject of scale discrepancy, scale transfer and scale invariance, a comprehensive overview about existing methods and when they should be applied is lacking. This book is a first attempt to fill this void.

This book provides a classification of upscaling and downscaling methods that have been developed within the environmental sciences. It is the first in its kind to contain a decision support system (DSS) that advises the practitioner about the class of upscaling or downscaling methods that is most appropriate in his or her research project. The book is meant for MSc- and PhD-students, applied researchers and practitioners in soil science, hydrology, (agro-)ecology, agronomy and the environmental sciences in general.

We owe this book to the planning committee of DWK program 328 of the Netherlands Ministry of Agriculture, Nature Management and Fisheries, who asked us to make an inventory about existing methods of scale transfer and to provide clues when to apply which method. We wish to express our gratitude to Sytze De Bruin (Wageningen University), Gerard Heuvelink (University of Amsterdam), Alex McBratney (University of Sydney) and Paul Torfs (Wageningen University) for thoroughly reviewing a draft of this book. Part of the writing was funded by the above research program. The remaining part and the programming of the DSS was finished in our own time which was gracefully donated to us by the members of our families.

Wageningen, February 29, 2000

MB
PF
PdW.

Chapter 1

Introduction

1.1 The problem of scale transfer

Most research performed in environmental science could be categorised as "applied research". Applied research aims at answering specific questions raised by society's "decision makers", such as local and national government, companies and farmers. Consequently, applied research is part of a "research cycle" as depicted in *Figure 1*. Decision makers raise specific questions such as "what is the average leaching of nitrate to the groundwater for a specific parcel?" or "what is the rate of temperature change in north-western Europe due to global warming?". These questions are usually answered by making observations, building models or a combination of these two: the N-concentration of soil water is observed at a number of locations and depths on the parcel; temperatures are observed at a large number of locations or with satellites; an atmospheric general circulation model (AGCM) is built. The resulting observations and model outcomes are then translated to provide the desired answers: "the average nitrate-load leaching from this parcel is 90 kg ha^{-1} yr^{-1}"; "the expected change of temperature is 1.5 degrees (Kelvin) in the next 50 years". These answers are subsequently used to make decisions or to formulate policy: "the maximum fertiliser application allowed should be decreased to 150 kg ha^{-1} yr^{-1}"; "the emission of carbon dioxide should be decreased with 20% globally". This in turn raises new questions about the outcome of this policy, which are answered by further applied research and so on.

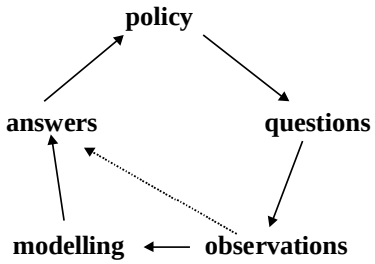

*Figure 1.*The cycle of applied research

Each component of the research cycle has its own temporal or spatial scale[1]. *Figure 2* shows the different scales involved and the possible relations between these scales. The *observations scale* is the scale for which an observation provides an average value: temperature measurements represent the average temperature for a few m^3 during a few seconds. The *model scale* is the scale at which the model provides its output: for instance, an AGCM provides the average temperature for model blocks 100x100x10 km for periods of 30 years. The *policy scale* is the scale at which the results of the research project are required in order to answer the decision makers' questions. It is also the scale at which decisions are made and policy is implemented. For instance, measures on fertiliser reduction may be implemented on farm level on a yearly basis. This means that leaching loads eventually should be given as yearly averages for each farm. Contrary to many reviews on scale, we have not distinguished the "process scale". This is the spatial and temporal scale at which a certain process is supposed to operate. Our view is that everything we can say about reality is in the form of some model: groundwater flow around a few grains of sand is described with the Navier-Stokes equations and at the level of a drilled core of sand with Darcy's law. Both equations are but *models* of groundwater flow and thus represent different *model* scales.

[1] Here, the term "scale" is used rather loosely, meaning the temporal and spatial units (i.e. intervals and areas or volumes) at which information is available or required. In section 1.3 it is defined more precisely.

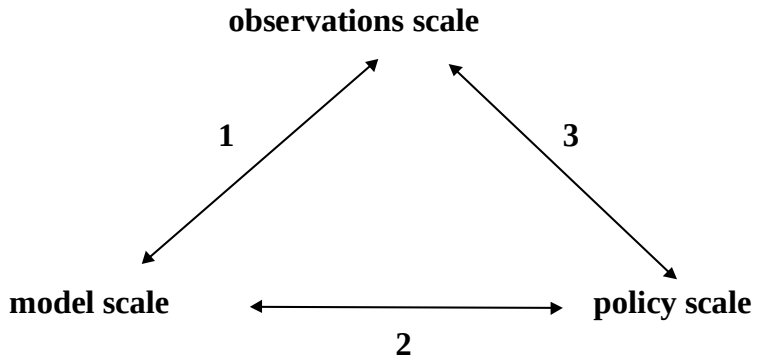

Figure 2. Three different scales involved in applied research and their relations; numbers refer to possible types of scale transfer as explained in *Figure 3*

Figure 1 and *Figure 2* show that while going through the research cycle information must be transferred from one scale to another. This so called *scale transfer* occurs between: 1) observations scale and model scale, 2) model scale and policy scale and 3) observations scale and policy scale. If the scale transfer involves transferring information from a smaller scale to a larger scale it is called *upscaling*, and if it involves going from a larger to a smaller scale it is called *downscaling*. Again, a more precise definition of these terms is given in section 1.3. *Figure 3* shows the various possibilities of scale transfer between the three scales of *Figure 2*. The various cases are exemplified below:

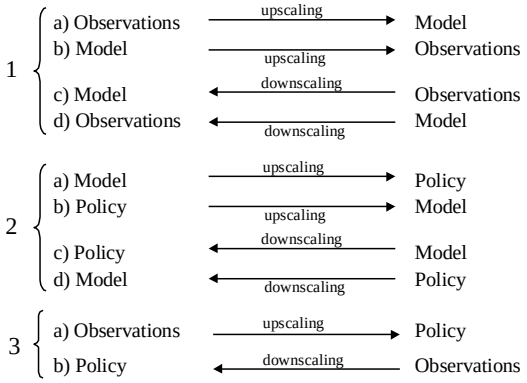

Figure 3. Types of scale transfer between the three scales involved in applied research; numbers refer to Figure 2

1. Between observations scale and model scale

Upscaling from observations scale to model scale occurs when an input variable or parameter is not available at the model scale. For instance, a numerical groundwater model requires hydraulic conductivity at the scale of the model blocks, while hydraulic conductivity is often measured in the laboratory on sediment cores which are much smaller. An example of upscaling the model to the observations scale is the derivation of Darcy's law (valid for sediment cores) from the Navier-Stokes equations (defined at the pore scale) because measurements of groundwater characteristics, such as hydraulic conductivity, are only available from sediment cores. As an example of downscaling from observations scale to model scale one can think of observing the total root uptake of water through the average moisture content of a soil profile, while having a model describing the uptake by single roots. Another example is a rainfall-runoff model that needs hourly rainfall data, while rainfall totals are available on a daily basis. Downscaling of model results to the observations scale is required if, for instance, a groundwater model predicts an average phreatic surface for an area, while the effects of changing groundwater depths on the vegetation are measured by observing plants in small plots within that area.

2. Between model scale and policy scale

One of the most common problems is upscaling the results from a model to a larger policy scale. For instance, a model calculates N-leaching from a parcel, while the government requires figures on average leaching of farms or counties. One example where information on the policy scale is upscaled to a (larger) model scale occurs when taking account of national measures to reduce CO_2 emissions in AGCMs. From this follows an obvious example of the reversed situation: downscaling the results of an AGCM to the policy scale. For instance, a local water board desires to know how much more rain can be expected in their area for some scenario of climate change predicted by the AGCM. The necessity of downscaling from the policy scale to the model scale occurs for instance if national targets of decreasing the sulphur output from factories are used in site specific models of soil acidification.

3. Between observations scale and policy scale

In this category the most common mode of scale transfer is upscaling from the observations scale to the policy scale. An obvious example occurs in soil pollution, where concentrations of a pollutant observed at small samples are used to estimate the average concentration of larger areas, such as city quarters. Downscaling from the observations scale to the policy scale may occur occasionally. For instance, the chemical load of the river Rhine is

measured just before it enters the sea. This is the average load for the whole catchment. To reduce this load, measures must be taken at smaller scale within the Rhine catchment (e.g. sub-catchments, farms, plots); hence the downscaling from observations scale to policy scale. Other types of scale transfer between observations scale and policy scale are very rare and therefore not mentioned here.

The last two decades have seen the emergence of a large body of literature about scale transfer, especially in hydrology, soil science and meteorology. Reviews and overviews can be found in for instance Hoosbeek and Bryant, 1993; Blöschl and Sivapalan, 1995; Renard and De Marsily, 1997; Van Garderen et al., 1997; Finke et al., 1998; Sposito, 1998; Wilby et al., 1998; Heuvelink and Pebesma, 1999. Before that, little literature can be found on the subject. This is not because the problem did not exist, but because it was not recognised. The reason is that the problem of scale transfer occurs mainly in the face of heterogeneity. If the parameters or variables modelled and observed were homogeneous, i.e. would not vary in space and time, they would be scale invariant. Up two 25 years ago, most of the environmental models consisted of analytical solutions to differential equations. For reasons of tractability, model parameters were often assumed to be homogenous. The advent of computer technology allowed for the development of numerical models in which spatial and temporal heterogeneity could be incorporated. Consequently, the change of parameters and variables from one scale to another is also a recent field of research.

Despite (or maybe because of) its recent emergence, the research on scale transfer in environmental science has led to an enormous amount of different methods and approaches to upscaling and downscaling information. Consequently, it is difficult for a practitioner to see where in the various steps of a research project scale transfer occurs and which methods of scale transfer are available and should preferably be used for these cases. In this book we attempt to classify the wealth of different upscaling and downscaling methods in a limited number of categories, provide examples for each of these categories and formulate decisions rules when to use which method.

1.2 Aims and scope

We have written this book to fulfil a number of goals:

1. to provide a theoretical framework for the large number of upscaling and downscaling methods used in environmental science;
2. to help the practitioner in assessing in which parts of the research project scale transfer occurs and if this is the case, which upscaling or downscaling methods are suitable to perform these instances of scale transfer;
3. to provide practical examples for each of the recognised classes of upscaling and downscaling methods.

The book is aimed at practitioners, applied researchers and students in soil science, hydrology, meteorology, agro-ecology and agronomy. In this book these fields are collectively denoted as "environmental research" or "environmental science". The book is specifically aimed at scale transformations that arise from heterogeneity in combination with differences in scale between observations, models and policy.

As this book is meant for applied research, it follows a practical approach to the problem of scale transformation, rather than a philosophical one. This means that descriptions of upscaling and downscaling methods are focused on the techniques used, not on how the methods deal with the "nature of things". Therefore, the reader will not find discussions whether variability is best described as random (disorder) or as (partly) organised. Both approaches are assumed to be alternative models of variation, each which its own merits and faults. Also, more philosophical approaches to scale, such as hierarchy (e.g. Hoosbeek and Bryant, 1992), organisation (e.g. Austin and Houze, 1972; Rodriguez-Iturbe, 1986a) and synergy (e.g. Wagenet, 1998) are not treated here.

In an excellent review paper about scale issues in hydrology, Blösch and Sivapalan (1995) distinguish two main approaches to the problem of scale transfer. The first is a model-oriented approach which focuses on the upscaling and downscaling of state variables, model parameters, input variables and model conceptualisations. The second approach, which is based on similarity concepts, tries to deal with complex processes in a much simpler fashion. Some examples of the second approach are: 1) the derivation of rules (based on geometric, kinematic or dynamic similitude) for scaling the results of scale models to real size, as applied in structural engineering and hydraulics (e.g. Kobus, 1984; Hwang and Houghthalen, 1996); 2) as used in soil physics: the practice of modelling the spatial variation of unsaturated conductivity and soil water retention functions with standard functions and a spatially variable scaling factor (Clausnitzer et al., 1992). This technique, confusingly called "scaling", is based on the concept of geometric similitude between porous media (Miller and Miller, 1956); 3) the use of fractals and multifractals (Mandelbrot, 1982; Feder 1988) to

describe the variability of heterogeneous phenomena continuously across a range of scales, or rather how certain properties of these phenomena are scale invariant. Blösch and Sivapalan (1995) rightly state that in each of the two alternative approaches considerable advances have been made. However, this book only deals with the model-oriented approach to scale transfer, because this approach has yielded more applicable methods of scale transfer and, as described in section 1.1, because a model plays a central role in most environmental research. Therefore, fractals and the technique of scaling of porous media will not be described extensively in this book, except as useful models of unknown spatial variation in some of the upscaling and downscaling methods presented.

1.3 Definitions

Throughout this book we use terms as “scale”, “support”, “model”, “parameters”, “input variables”, “output variables” etc. In this section these terms are defined. The definitions have also been collected in a glossary at the end of this book.

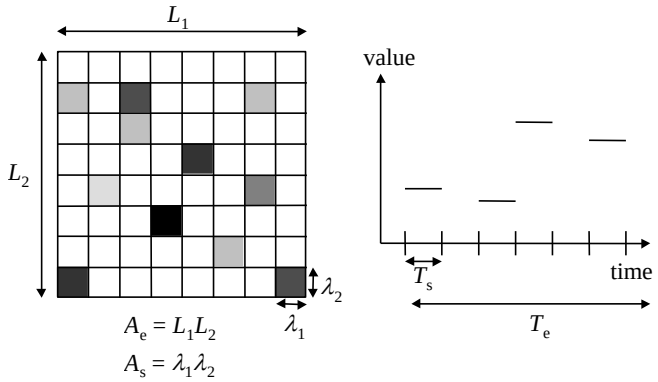

Figure 4. Figures to illustrate the concepts of support (A_s and T_s), extent (A_e and T_e) and coverage; left spatial example (coverage is 11/64); right temporal example (coverage is 4/6)

Support, extent and coverage

The following description follows largely from that of McBratney (1998). *Figure 4* shows schematically an area and a time interval. Let us assume that our research project involves the description of some property, e.g. the concentration of some pollutant, for the depicted area and for the depicted time interval. Naturally, the area does not have to have a regular form. The area A_e and the time interval T_e are called the *extent* of the research project. This is the area (or volume) or time interval over which observations are made, model outcomes are calculated or policy measures are to be made. The area and the time interval are divided into a finite number of sub-areas A_s or sub-intervals T_s. The area of these sub-areas or length of these sub-intervals are called the *support* of the research project. It is the largest area (or volume) or time interval for which the property of interest is considered homogeneous. For these areas or time intervals we only know the average value of the property considered and not its variation within. The sub-areas or sub-intervals themselves are called *support units*. It is not always possible to know the average values of the property of interest for all support units. For instance, budget often allows for a small number of support units to be observed. The ratio of the sum of areas (or time intervals or volumes) for all support units for which the average values are known and the extent is called the *coverage*[2] of the research project. Suppose that the property values of support units $A_s(i)$, i=1,..,M are known, then the coverage is given by $\left(\sum_{i=1}^{m} A_s(i)\right)/A_e$. Similarly for volume and time the coverage is given by $\left(\sum_{i=1}^{m} V_s(i)\right)/V_e$ and $\left(\sum_{i=1}^{m} T_s(i)\right)/T_e$.

[2] If the support is very small compared to the extent (i.e. it is effectively a point observation), it may be better to define the coverage as the number of observations per time interval, area or volume. If this is the case, the distance d between observations becomes an important parameter.

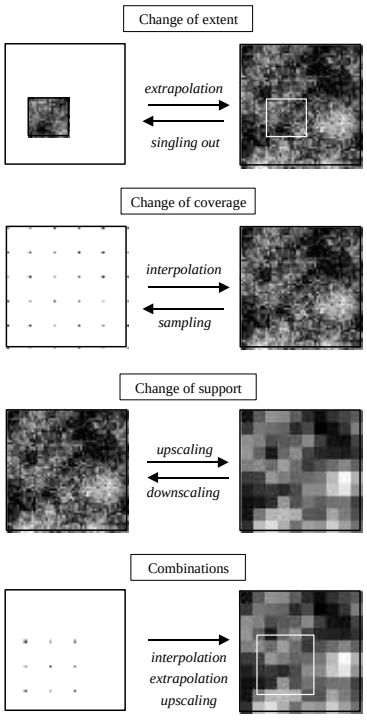

Figure 5. Basic operations involving extent, coverage and support

The top three figures of *Figure 5* depict schematically for two dimensions the three basic operations involving change of extent, change of coverage and change of support. Changing the extent of the research area usually involves going from a smaller to a larger extent. Increasing the extent is called *extrapolation*. Very rarely, the extent is decreased by *singling out* a part of the research are. *Interpolation* involves increasing the coverage of the research area as can be seen in the second figure. In this framework, *sampling*, i.e. taking a subset of support units and observing them, can be seen as the reverse of interpolation. The third figure shows the change of support. Increasing the support is called *upscaling*, while decreasing the support is called *downscaling*. The lower figure in *Figure 5* shows a combination of all three operations involving both interpolation, extrapolation and upscaling.

The term *scale* as used in this book is reserved for the concept of support. So scale transfer and scale problems in research projects pertain to changes of support. This book deals with upscaling and downscaling as defined in this section. However, in many studies changes of support also involve

changes of extent and coverage, so it is unavoidable that in some of the examples presented in this book these three basic operations are tackled in an integrated manner.

Mathematically the relationship between two scales can be described as follows: Consider a support unit with support s_2 containing N_{12} support units with support s_1 ($s_1 < s_2$). Let $z(s_1;i)$ denote the property z for support unit i at support s_1, and $z(s_2)$ the (average) value of the property for the support unit at support s_2. The value $z(s_2)$ and the values $z(s_1:i)$ are related as follows:

$$z(s_2) = \frac{1}{N_{12}} \sum_{i=1}^{N_{12}} z(s_1;i) \tag{1.1}$$

This definition of the relationship between two scales is based on two premises. First, the value of a support unit is the arithmetic average of the values of smaller support units. Other averaging methods, such as geometric or harmonic averaging, are also possible, but in this book we take Equation (1.1) as the fundamental relationship between two scales. If in this book the term "averaging" is used, we imply that values at scale s_1 are linearly combined (either weighted or non-weighted, see section 2.2) to obtain an estimate of the arithmetic average. If the term "average" is used without any adjective, the arithmetic average is implied. As will be shown later, when a model is involved, non linear types of averaging (e.g. geometric averaging) may be more appropriate. If some type of non-linear averaging is used, it is explicitly mentioned. Second, support units at a given scale are of equal size. In other scale definitions support sizes at a give scale can be different. For instance, polygons on a soil map of given map scale are of different size. In that case Equation (1.1.) would be a weighted average. In the scale definition used in this book, scale is another word for support. So by definition, support units at a given scale are of equal size.

So far, the support units were considered to be non-overlapping, as shown in *Figure 5*. In certain applications, especially in continuum mechanics (e.g. groundwater flow), it is often assumed that support units are overlapping and of equal size. *Figure 6* shows a two-dimensional spatial example and an example in time. In this representation for a given scale, a value is defined at *each* point in space and time. The value assigned to this point is the average of a time interval (in time), a circle (two spatial dimensions) or a sphere (three spatial dimensions) of which the point forms the centre. So at each point a so called moving average is defined. The support is the length of the time interval, area of the circle or the volume of the sphere. Obviously, the definition of coverage for the non-overlapping support units does not apply here. Alternative measures of coverage should

be used such as the number of observations per unit time, per unit area or per unit volume.

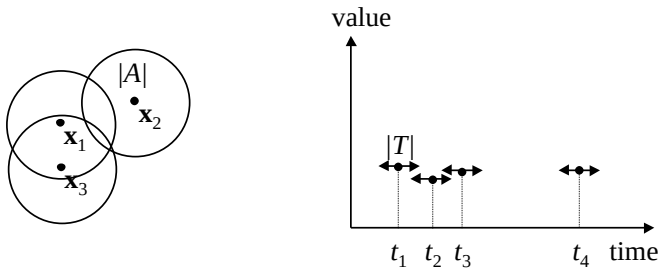

Figure 6. Figures to illustrate the concept of support in the continuum approach; left figure: the (2D) spatial domain, |A| is support, $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ are space co-ordinates; right figure: the temporal domain, |T| is support, t_1, t_2 and t_3 are time co-ordinates

In the continuum approach, the values of the support units are indicated as follows: $z(T;t)$, $z(\mathrm{A};\mathbf{x})$, $z(\mathrm{V};\mathbf{x})$, where t and $\mathbf{x}$ represent the central points in time and space of the support units ($\mathbf{x}$ is a vector containing two or three space co-ordinates) and T, A, V represent the support units: respectively the time interval, circle (area) and sphere (volume) for which the value at the central point the average property is denoted. The support is the actual size of the support units (time, area, volume) denoted respectively by $|T|$, $|A|$ and $|V|$. If in the following we do not wish to distinguish between time, 2D-space or 3D-space, we will use s for T, A, or V and denote a point (which may be in space or time) with the co-ordinate vector $\mathbf{x}$ or $\mathbf{y}$. The relation between two scales is now described as follows: Consider a support unit with support $|s_2|$ containing points with support $|s_1|$ ($|s_1| < |s_2|$). Let $z(s_1;\mathbf{x})$ denote the property for a support unit at point $\mathbf{x}$ with support $|s_1|$, and $z(s_2;\mathbf{y})$ the (average) value of a property z for the support unit at support $|s_2|$ with central point $\mathbf{y}$. The value $z(s_2;\mathbf{y})$ and the values $z(s_1;\mathbf{x})$ are related as follows:

$$z(s_2;\mathbf{y}) = \frac{1}{|s_2|}\int_{\mathbf{x}\in s_2} z(s_1;\mathbf{x})\,d\mathbf{x} \qquad (1.2)$$

Because the continuum approach to scale transfer generally leads to more complicated formulae, we will try to follow the discrete approach as much as

possible. The continuum approach is used if in the literature a certain upscaling method has been developed in the continuous domain only[3].

We conclude with mentioning some of the terms that are also used in the literature, but not in this book. The term "grain" is sometimes used to depict "support" (e.g. McBratney, 1998). "Upscaling" and "downscaling" are also referred to as "aggregation" and "disaggregation" respectively. Care should be taken when reading about or using the term "resolution" because sometimes it is used to depict something that resembles "support" (i.e. the number of pixels per area) and sometimes to depict "coverage" (e.g. in case of lattice data: the number of grid points per area). Finally, we stress that the term "scale" denoted on maps means exactly the opposite to the scale definition used here: a large map scale means a small map extent with a small support.

Model, parameters and variables

Models are often an indispensable part of the research cycle. Only in very few cases can questions asked by decision makers be answered through observations alone. In most cases these observations are used in models, while the model output is subsequently used to provide the answers. As described in section 1.1 many problems of scale transfer occur through scale discrepancies between the model and the observations on one hand and the model and the policy makers on the other hand. As will be seen in the next chapter, particularly the required method of upscaling depends on the place of the model in the research cycle and the type of model used. In this section we will define what is meant by a model and describe its various components. These definitions are used in next chapter to distinguish and describe the various methods of scale transfer.

[3] Very often a property such as porosity or hydraulic conductivity that cannot be defined at a single point (you are either in a void or on a grain) must be defined as a continuum through application of Equation (1.2) at the smallest possible overlapping volumes that result in a stable value of the property (called the representative elementary volume). In the subsequent analysis this is then called the "point support". In the next step, Equation (1.2) is used on the point support values for larger non-overlapping volumes (e.g. the cubes of a finite difference model). This results in a discrete description so that the discrete approach can be used if further upscaling is required.

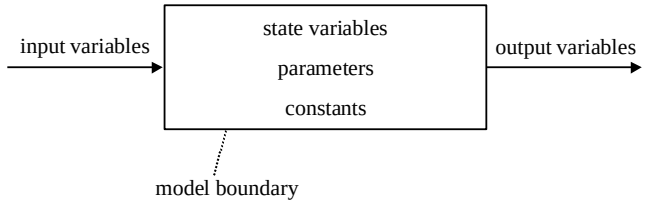

Figure 7. Model and model properties according to the systems theory

Figure 7 shows a schematic representation of a model as used in system's theory. A *model* is a simplified representation of part of reality. The *model boundary* separates the part of reality described by the model from the rest of reality. Everything that is to know about the part of reality described by the model at a certain time is contained in the *state variables*. These are variables because their values can change both in space and time. The variation of the state variables is caused by the variation of one or more *input variables*. Input variables are always observed and originate from outside the model boundary. Consequently, input variables also include boundary conditions and initial conditions such as used when solving differential equations. If the state variables are known, one or more *output variables* can be calculated. An output variable traverses the model boundary and thus influences the part of reality not described by the model. Both input variables and output variables can change in space and time. The state variables are related to the input variables and output variables through *parameters*. Parameters may change in space but are invariant in time. Because they are constant in time, parameters represent the intrinsic properties of the model. Finally, a model may have one or more *constants*. Constants are properties of a model that do not change in both space and time (within the confines of the model). Examples of such constants are the gravity constant and the viscosity of water in density independent groundwater flow at a constant temperature. Constants are the only properties of a model that are by definition scale invariant: they do not change value when transferred from one scale to another. The values of state variables, input variables and output variables can change under change of support as they may vary with space and time. The value of parameters can change under change of support because they connect two variables that are dependent on space and time.

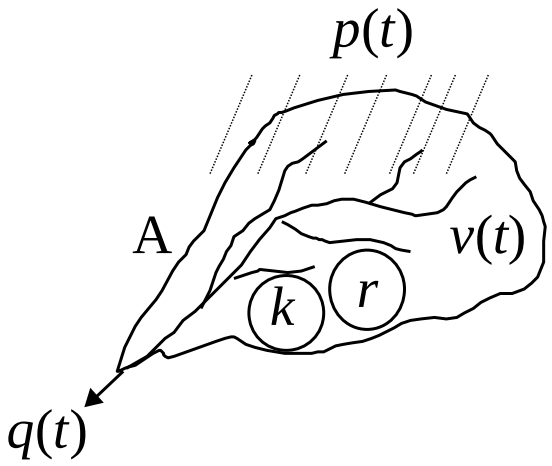

Figure 8. Illustration of model properties following systems theory with a model of a catchment; v(t): state variable, storage surface water in catchment [L^3]; q(t): output variable, surface runoff from catchment [L^3T^{-1}]; p(t): input variable, precipitation [LT^{-1}]; k : parameter, catchment parameter [T^{-1}]; r : parameter, infiltration capacity [LT^{-1}]; A: constant, area of the catchment [L^2].

Because the description above is rather abstract, we will try to illustrate it with the example shown in *Figure 8*. We consider a model describing the discharge from surface runoff q [L^3T^{-1}] from a catchment caused by the average precipitation p [LT^{-1}] observed as averages over discrete time steps Δt, i.e. $q(t)$ and $p(t)$ represent the average discharge and precipitation between t-Δt and t. The model boundary is formed by physical boundaries such as the catchment boundary (i.e. the divide), the catchment's surface and a few meters above it, and also by the virtual boundary with everything that is not described by the model such as groundwater flow, soil moisture, chemical transport etc. Obviously, rainfall is the input variable and surface runoff the output variable. The state variable of this model is the amount of water stored on the catchment's surface: v [L^3]. The state variable is modelled with the following water balance equation:

$$v(t) = v(t-1) + \left\{A \cdot [p(t) - r]^+ - q(t)\right\}\Delta t \qquad (1.3)$$

where the supserscript + is added to [$p(t) - r$] to denote that if $p(t) < r$ we have [$p(t) - r$] = 0.

The output variable s is related to the state variable v with the following equation:

$$q(t) = kv(t) \tag{1.4}$$

Through substitution of (1.4) into (1.3) we can calculate the development in time of the state variable directly from the input variable as:

$$v(t) = [1 - k\Delta t] \cdot v(t-1) + \mathrm{A} \cdot [\, p(t) - r]^{+} \Delta t \tag{1.5}$$

Two model parameters can be distinguished: the infiltration capacity of the soil r [LT^{-1}] which relates the input variable to the state variable and the catchment parameter k [T^{-1}] relating the output variable to the state variable. The constant A [L^2] is the area of the catchment.

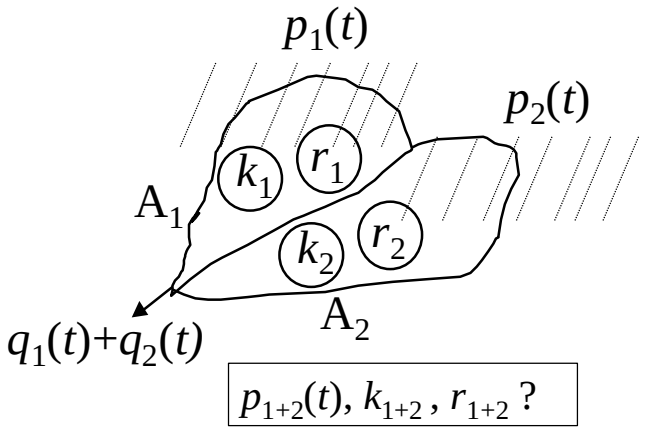

Figure 9. Two catchments to illustrate the role of model properties in an upscaling problem; for description of symbols see *Figure 8*

To give a quick example how these definitions become important for scale transitions, we consider the example of two catchments next to each other as shown in *Figure 9*. Suppose that for some reason we are interested in the total surface discharge from both catchments q_1+q_2, without having to run both models separately (this may happen in practice if the separate models are very complicated so that running them requires large amounts of computation time). In that case, one form of upscaling that can be applied (see section 2.3) is finding representative values of the infiltration capacity r_{1+2} from r_1 and r_2, representative values for the catchment parameters k_{1+2} from k_1 and k_2 and representative values for the input variable $p_{1+2}(t)$ from

$p_1(t)$ and $p_2(t)$ such that when inserted in Equations (4) and (5) we directly obtain the correct discharge $[q_1+q_2](t)$.

1.4 Contents of this book

In chapters 2 and 3 methods for upscaling and downscaling are described. The first section of each chapter starts with the introduction and description of a framework that is used to classify the wealth of methods used in the literature. By classifying the upscaling and downscaling methods, we intend to guide applied researchers and practitioners to the most suitable method for their particular problem of scale transfer. In the first section of chapter 2 and chapter 3, the reader is guided to the appropriate class of upscaling or downscaling through the answers supplied to a few questions. Each class corresponds to a section in the chapter. The sections are used to describe the classes and subclasses in more detail, and provide questions to guide the reader to subclasses. Each subclass corresponds to a subsection. In each subsection, a general description of the subclass of upscaling or downscaling methods is given as well as the conditions under which they are best applied. Each description is followed by an example from practice in which the particular (sub)class of upscaling or downscaling methods is applied. It is clear that by following this approach we aim this book to be a handbook rather than an introductory text.

Readers will notice that many examples in this book are taken from hydrology and, to a lesser extent, from soil science. We admit that this bias is partly caused by the authors' backgrounds. However, it is also true that among the environmental sciences, hydrology provides by far the most examples of upscaling and downscaling methods that actually work. We are convinced that the methods underlying these examples can be easily adapted to other fields of environmental research.

This book comes with a CDROM containing a decision support system (DSS) that can be used to pick the right (sub)class of upscaling or downscaling methods for a particular research project. The DSS, which is implemented in a DELPHI-computer program, is described in detail in chapter 4. The DSS consists of two parts. In the first part the DSS constructs through a series of questions answered by the user a flow diagram of the user's research project. An example of such a diagram is given in *Figure 10*. The diagram presents the research project as a number of steps consisting of either data collection, modelling or representation of results (to the decision makers). Interactively, the DSS identifies the locations in this diagram where scale transfer is required. In the second part of the DSS each of the instances

of scale transfer is further analysed through asking the same set of questions as in the book, and evaluating the answers by the user. The outcome of this second part consists of denoting the subsection(s) of chapters 2 and 3 where the most suitable methods of upscaling or downscaling are described, together with an appropriate example from practice. Alternatively, the appropriate sections can be read from the DSS itself.

The book ends with a list of references, a glossary of frequently used terms, a subject index and an appendix describing the basic concepts of random variables and stochastic functions, which are frequently used in this book.

When scanning the contents of chapters 2 and 3 it is clear that the classifications used for the upscaling and downscaling methods are based on different criteria. We could have used a more symmetric classification system, albeit with great difficulty. Instead, we have used a separate classification system for upscaling and downscaling, because the way a rational choice must be made between the various methods appears, at least to us, very different for upscaling and downscaling. The decision which method should be used is much less straightforward for upscaling than for downscaling. When upscaling, one should for instance consider whether a model is involved, whether this model is linear and how expensive it is to run the model many times. In case of downscaling the problem appears always practically the same: given the value (or probability distribution of values) at the larger scale, what are the values (or probability distribution of values) at a smaller scale? This problem is solved in a similar way, by assuming the property to be downscaled to vary according to some function of time or space within the support units and by fitting the parameters of this function such that the average value of support unit (or its probability distribution) is maintained. The methods only differ with the type of function chosen and how it is constructed. Choices between these methods are mostly based on the available information at the smaller scale and the familiarity of the user with the methods. We realise that the classification system used here is one of many possible systems, but it is one that appears natural to us when considering the approaches followed in the literature.

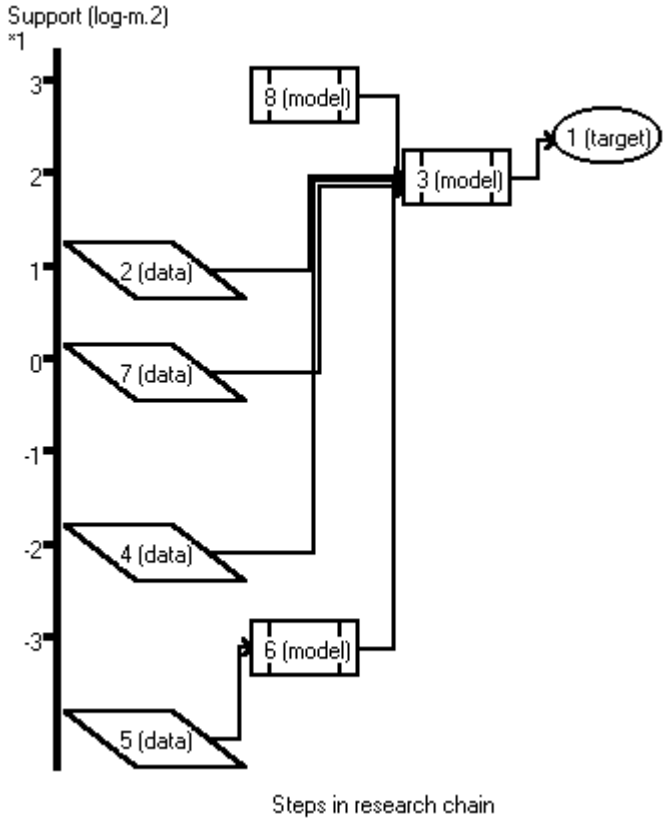

Figure 10. Example of a diagram depicting a research chain with change of support (here "data" are either observations, input variables, output variables or parameters)

Chapter 2

Upscaling

2.1 A classification of upscaling methods

We have divided the upscaling methods into four major classes on the basis of answers to the following questions:

- *Is there a model involved?*
- *Is the model linear in its input variables and parameters?*
- *Can the model be applied at many locations or time steps?*
- *Does the model have the same form at the two scales involved?*
- *Can the larger scale model be analytically derived from the smaller scale model?*

Figure 11 shows a decision tree that can be constructed from these questions. In the following each of these questions and the resulting classification is described in more detail.

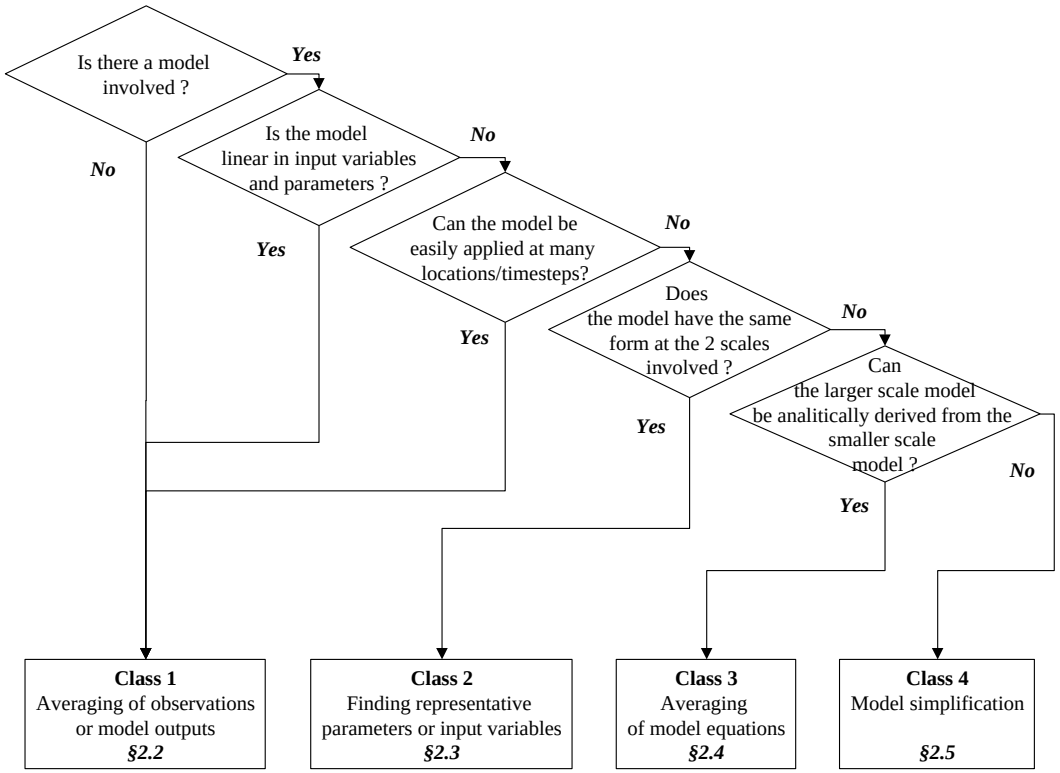

Figure 11. Decision tree to four major classes of upscaling methods

Is there a model involved?

If there is no model involved, if we have some variables observed at a number of locations at a given scale, and if we want to use these observations to calculate the value of the variables at a larger scale, the upscaling problem consists of nothing more than some averaging procedure (see section 2.2). This is schematically denoted in the diagram of *Figure 12*. Variables that are observed at a number of locations or time steps at given scale s_1 are averaged to a larger scale (= support) s_2. This averaging consists of taking the (sometimes weighted) average all observed values of support units at scale s_1 that fall within the same support unit of scale s_2. The presence of a model is a complicating factor when upscaling. If input variables or parameters at a given scale are used as input for a model that produces output at a larger scale, the upscaling becomes a non-trivial exercise.

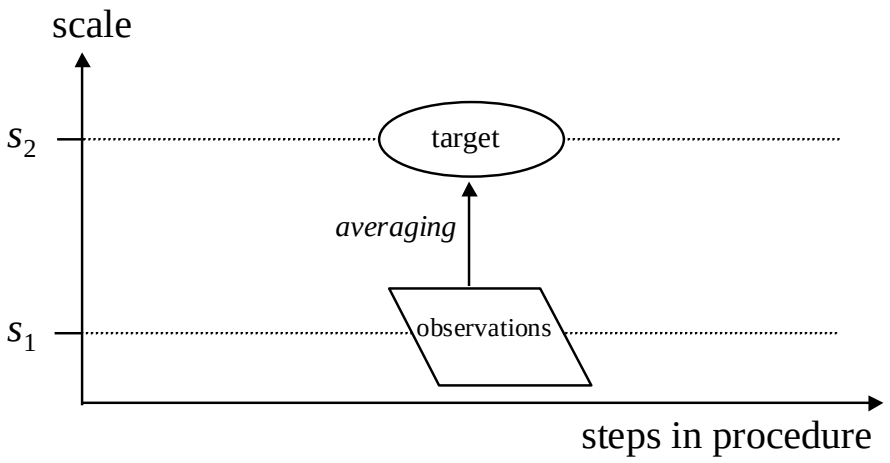

Figure 12. Upscaling by averaging observations

Is the model linear in its input variables and parameters?

If a model is involved and it is linear, the upscaling problem is still rather trivial. This can be best explained by considering *Figure 13* and *Figure 14* (see also Heuvelink and Pebesma, 1999). It shows that there are two ways to arrive from the input variables and parameters at scale s_1 to the output variable at scale s_2. The first one consists of running the model first at a large number of locations or time steps at scale s_1 and averaging the output variables at scale s_1 to s_2 (*Figure 14*). The second one is to average the input variables and parameters at scale s_1 to s_2 first, and then perform a single model run with the inputs and parameters at s_2 to produce the output variable at s_2 (*Figure 13*). In general, these two routes do not yield the same answer, and, as explained hereafter, the correct route to take is that according to *Figure 14*.

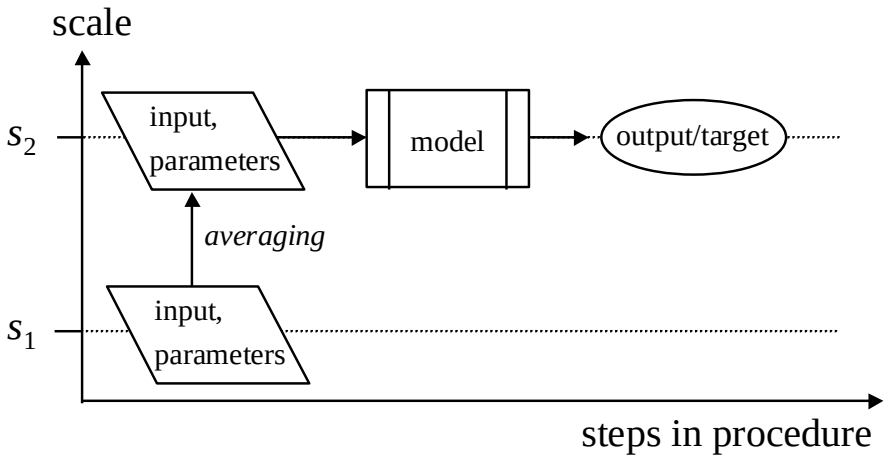

Figure 13. Upscaling model input variables or parameters

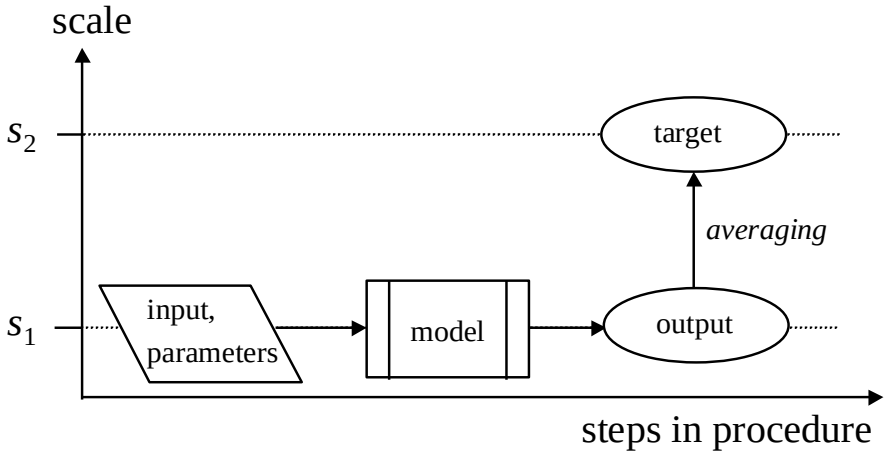

Figure 14. Upscaling model output variables

Only if the model is linear in its inputs and parameters, the two routes yield the same answer. In mathematical terms: Consider a support unit with support s_2 containing N_{12} support units with support s_1 ($s_1 < s_2$). Let $z(s_1;i)$ denote an input variable or model parameter for support unit i at scale s_1,

$y(s_1,i)$ the associated output variable and g[·] the model. So at scale s_1 we have the following relation:

$$y(s_1;i) = \mathrm{g}[\,z(s_1;i)] \tag{2.1}$$

Let $z(s_2) = < z(s_1;i)>$ denote the (average) value of a property z for the support unit at scale s_2 which is defined as

$$z(s_2) = < z(s_1;i) > = \frac{1}{N_{12}} \sum_{i=1}^{N_{12}} z(s_1;i) \tag{2.2}$$

where $y(s_2) = < y(s_1;i)> = <\mathrm{g}[z(s_1;i)]>$ is defined similarly. If the model is linear in its input variable or parameter z, i.e. $\mathrm{g}[z] = \mathrm{a} + \mathrm{b}z$, the output variable at scale s_2 can be obtained through running the model with the input variables or parameters at scale s_2:

$$y(s_2) = < \mathrm{a} + \mathrm{b}\, z(s_1;i) > = \mathrm{a} + \mathrm{b} < z(s_1;i) > = \mathrm{a} + \mathrm{b}\, z(s_2) \tag{2.3}$$

In general, if the model is non-linear, the output variable at scale s_2 is *not* obtained through running the model with the input variable or parameters at scale s_2:

$$y(s_2) \neq g[z(s_2)] \tag{2.4}$$

For example, if $\mathrm{g}[z] = \mathrm{a}z^2$, we have that

$$y(s_2) = < \mathrm{a}\,(z(s_1;i))^2 > \neq \mathrm{a}(< z(s_1;i) >)^2 \tag{2.5}$$

So if a model is involved and this model is linear in its input variables or parameters at scale s_1, the input variables and parameters at scale s_2 are obtained first through one of the averaging methods described in section 2.2. Next the input variables and parameters at scale s_2 are used in the linear model to calculate directly the output variables at scale s_2 (average first, calculate later as depicted in *Figure 13*).

Can the model be applied at many locations or time steps?

If a model is present and this model is non-linear, the procedure of *Figure 14* should be followed: the model is applied at all locations or time steps where input variables and parameters are known at scale s_1. This produces output variables at the scale s_1 at these locations or time steps. The

output variable for scale s_2 is obtained through averaging the output variables at scale s_1. This procedure, which may be called "calculate first, average later", reduces the upscaling problem to direct averaging of output variables, using the averaging methods described in section 2.2. Obviously, if we require an accurate estimate of the output variable at scale s_2, we require a sufficient number of locations or time steps at scale s_2 where the model is run. This means that the coverage should be large enough: input variables and parameters should be known at a sufficient number of support units of scale s_1, i.e. at a sufficient number of locations or time steps. Also, running the model must be computationally cheap. This means that it must not take too much time to perform a single model run at scale s_1. If the coverage at scale s_1 is insufficient (i.e. the so called data crisis) or if running the model at many locations or time steps takes too much time, it is preferable to define a model that directly produces the output at scale s_2. In that case, the upscaling becomes non-trivial and the methods required fall into the classes described in section 2.3 to 2.5.

Does the model have the same form at the two scales involved?

If there is a model involved that cannot be applied at many locations, a model must be sought that directly gives the output variable at the required scale s_2. Again, if the model is linear, we can apply the same model with the average input and parameters to obtain the average output $y(s_2)$ for scale s_2. However, in case the model is non-linear in its input variables or its parameters, another route must be followed. There are two distinct methods of obtaining a model that directly gives the output variable $y(s_2)$:

The first method assumes that the same model can be used at scales s_1 and s_2, while seeking so called representative values for the input variables and parameters. So, if Equation (2.1) gives the model for a support unit at scale s_1, a representative value $z_R(s_2)$ for the support unit at scale s_2 is determined such that when used in the model as input variable or parameter, it produces the output variable $y(s_2)$:

$$y(s_2) = g[z_R(s_2)] \tag{2.6}$$

The upscaling problem then consists of finding representative input variables or representative parameters at scale s_2 from the input variables and parameters at scale s_1:

$$\{z(s_1;i), i = 1,..,N_{12}\} \rightarrow z_R(s_2) \tag{2.7}$$

The values of representative parameters and input variables depend on the values of these at support s_1, the model used at s_1 and the support ratio s_2/s_1. *Figure 15* shows the diagram showing the upscaling route when assuming the same model to be valid at two different scales. The methods falling in this class of upscaling are described in section 2.3. Notice that *Figure 15* closely resembles the procedure for linear models shown in *Figure 13*. The difference lies in finding the input variables and parameters at scale s_2 from those at scale s_1. For a linear model it consists of a spatial or temporal average (estimated using one of the methods described in section 2.2), while for a non-linear model more complicated procedures (e.g section 2.3) are necessary. From this comparison we can also conclude that for linear models the representative parameters and input variables at the larger scale can be found by averaging of the input variables and parameters at the smaller scale.

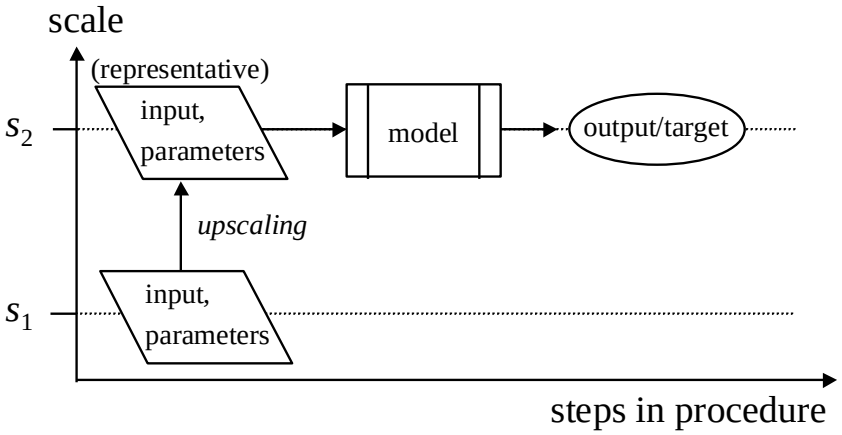

Figure 15. Upscaling to representative input variables or parameters

The second method assumes that at a different scale a different model applies. So if $g_1[\cdot]$ represents the model at scale s_1 and if $g_2[\cdot]$ the model at scale s_2, the upscaling problem consists of finding the right model $g_2[\cdot]$ such that

$$y(s_2) = < g_1[z(s_1; i)] > = g_2[z(s_2)] \tag{2.8}$$

or

$$y(s_2) = < g_1[z(s_i;i)] > = g_2[\{z(s_1;i), i = 1,..,N_{12}\}] \quad (2.9)$$

Equations (2.8) and (2.9) show that the model sought at scale s_2 can be a function of the upscaled parameters and upscaled input variables $z(s_2)$ or the parameters and input variables for the support units at scale s_1. Here, the form of the larger scale model $g_2[\cdot]$ depends on the values of the parameters and input variables at support s_1, the model used at s_1 and the support ratio s_2/s_1.

Can the larger scale model be analytically derived from the smaller scale model?

If the equations making up the smaller scale model $g_1[\cdot]$ are not too complicated, it is often possible to derive the larger scale model $g_2[\cdot]$ from the smaller scale model analytically through averaging of the model equations. So, $g_2[\cdot]$ is analytically derived from evaluating $<g_1[\cdot]>$, usually given some supplementary assumptions about the variation of the input variables and parameters $z(i;s_1)$ within the support units at scale s_2. If it is indeed possible to derive $g_2[\cdot]$ analytically from $g_1[\cdot]$, some ways to go about it are given in section 2.4. A schematic overview of the upscaling route used in section 2.4 is given in the diagram of *Figure 16*. Notice that if the model is linear, both models have the same form, i.e. $g_2[\cdot] = g_1[\cdot]$, so that the method of *Figure 16* becomes the same as that of *Figure 13*. In that case we simply have: $y(s_2) = g_1[z(s_2)]$.

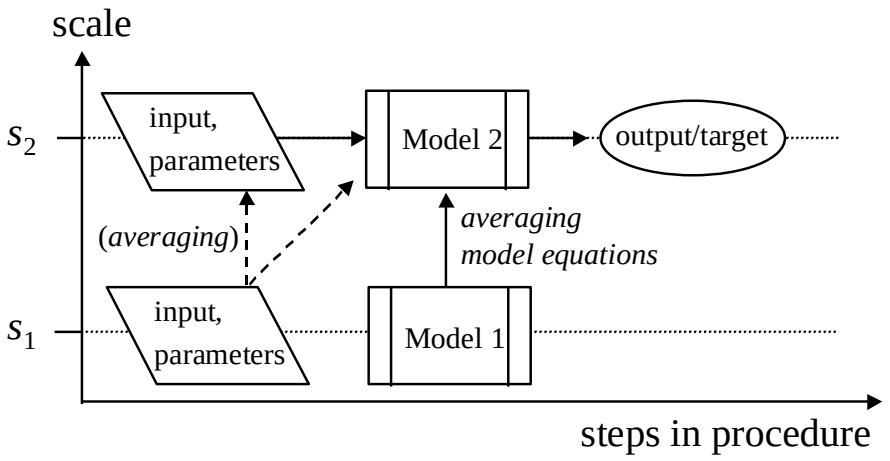

Figure 16. Upscaling the model and its input parameters

If the model $g_2[\cdot]$ is not tractable from $g_1[\cdot]$ it must be sought in a different way. One way of finding suitable models $g_2[\cdot]$ is described in section 2.5 and the diagram of *Figure 17*. First a representative sample of parameters and input variables is taken from the support units at scale s_1. The model $g_1[\cdot]$ is applied at the sample locations or sample time steps to yield the output variables at scale s_1. Using one of the averaging methods described in section 2.2, the average input variables and the average output variables for scale s_2 are estimated. The parameters may be both averaged and lumped. The latter means that some processes are omitted or represented by a single quantity. Next, a simplified model is postulated which directly relates the estimated parameters and input variables at s_2 to the output variables at s_2. Ideally this model is obtained by leaving out model parts from the (more complicated) model $g_1[\cdot]$. Usually, however, it is a lumped conceptual model postulated from expert knowledge or some regression type model with no physical meaning (called a meta-model). With the postulated model $g_2[\cdot]$ and the parameters and input variables at s_2 the output variables at s_2 can also be estimated. These results are subsequently compared with the average model outputs obtained from $g_1[\cdot]$. If these results compare well, the postulated model can then be used for all other support units at scale s_2. If the results are too different, the postulated model is calibrated through adjusting some of its parameters, or alternatively, different simplified models are postulated until a better match is obtained.

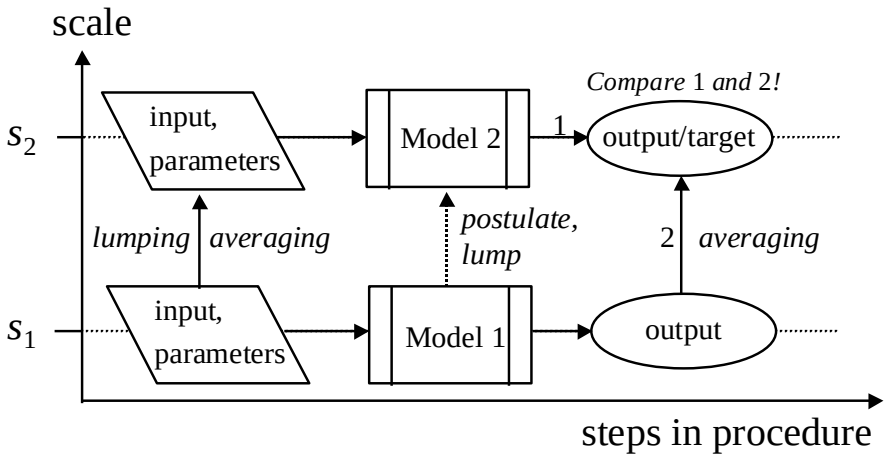

Figure 17. Upscaling the model and its input parameters when the model at s_2 cannot be analytically derived

In the following sections the four major classes of upscaling methods are described. These are further subdivided into subclasses based on additional criteria and decision trees. The subclasses are described in the subsections. Each subsection contains the list of criteria (steps in the decision trees) leading to its choice, a general description of the method and one ore more examples.

2.2 Averaging of observations or output variables

The following criteria lead to the application of the class of upscaling methods described in this section (see section 2.1):

– *No model is involved.*
The upscaling procedure is denoted schematically in *Figure 12*. Observed properties or transforms of these properties are averaged directly from scale s_1 to scale s_2.

– *A model is involved.*
– *This model is linear in its input variables and parameters.*
The upscaling procedure is denoted schematically in *Figure 13*. Input variables and parameters at scale s_1 are averaged to scale s_2 and subsequently used in the linear model to produce the output variable at scale s_2.

– *A model is involved.*
– *This model is non-linear in its input variables or parameters.*
– *The model can be applied at many locations or time steps.*
The upscaling procedure is denoted schematically in *Figure 14*. The model is applied at all locations or time steps where input variables and parameters are known at scale s_1. This produces output variables at the scale s_1 at these locations or time steps. The output variable for scale s_2 is obtained through averaging the output variables at scale s_1.

These procedures have in common that they involve the averaging of spatially or temporally variable properties (observations of some property or its transform, input variables, parameters, output variables) at scale s_1 to scale s_2. The actual upscaling thus consists of this averaging procedure only. There are various ways of estimating spatial or temporal averages. Which of these should be used depends on a number of additional questions whose answers lead to the decision tree denoted in *Figure 18*.

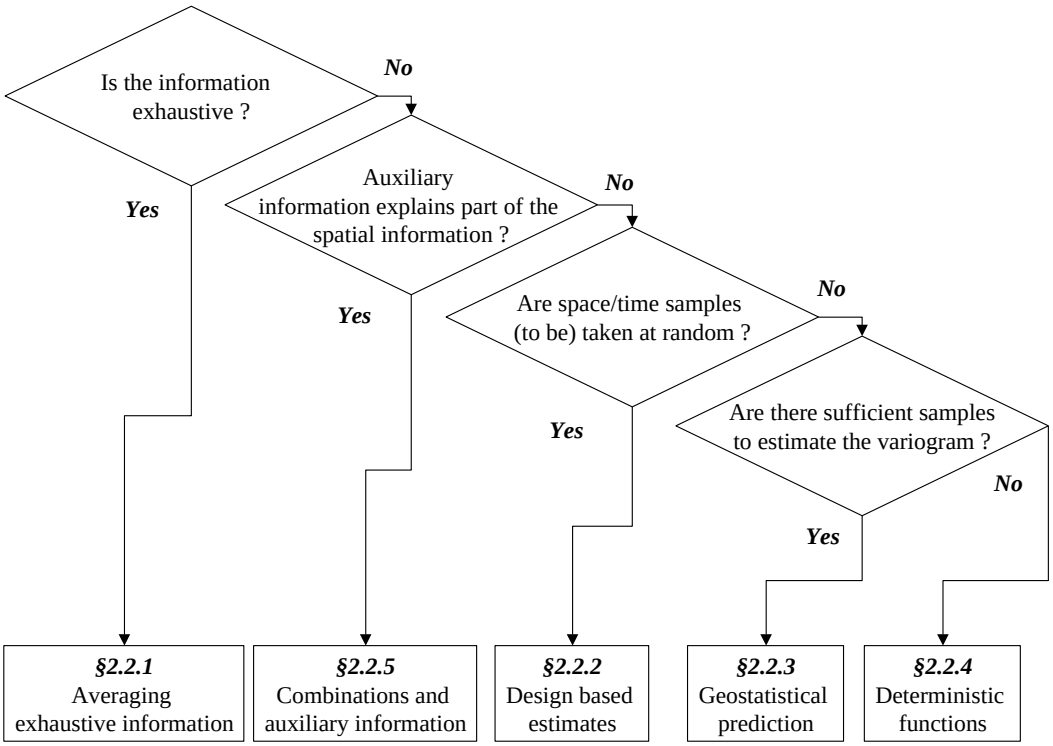

Figure 18. Decision tree to five subclasses of upscaling class 1

For each of the properties (observations of some property, input variables, parameters, output variables) involved, the questions to be answered are:

- *Is the information about the property exhaustive?*
- *Is there auxiliary information that can be used to explain some of the unknown spatial and temporal variation of the property?*
- *Are the sample locations or sample time steps at which the property at scale s_1 is determined (to be) taken randomly?*
- *Are there sufficient (>30) sample locations of the property at scale s_1 to estimate the semivariogram or the correlogram?*

These questions and the consequences of their answers are discussed in the following.

Is the information about the property exhaustive at scale s_1?

If the property is known for all support units of scale s_1 within the target support unit of scale s_2, i.e. if the coverage is complete, the average value of the property for scale s_2 is simply obtained by averaging the values of the property of the s_1 support units (see Equation 2.2). The case of exhaustive information is described in subsection 2.2.1.

Is there auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1?

If the property at scale s_1 is not known exhaustively within the s_2 support unit, a spatial or temporal average of the property must be estimated from the property values of a limited number of support units of size s_1 – this limited number is called a *sample*. In terms of the basic operations shown in *Figure 5* the operation involves interpolation, upscaling and possibly extrapolation. If the (target) property is observed at only a small number of locations or time steps, the interpolation and extrapolation of this property may be dramatically improved if another, more densely sampled (auxiliary) property is available that is strongly related to the target property. For instance, if observations of the depth of the phreatic surface are hard to obtain, the spatial variation of the phreatic surface may be largely explained by surface elevation data (high spatial coverage), while the temporal variation of the phreatic surface may be explained for a large part with observations of precipitation (high temporal coverage). There is a wealth of techniques to use auxiliary information for spatial and temporal interpolation and estimating spatial or temporal averages. These techniques are advanced cases of the three standard techniques described in subsections 2.2.2 to 2.2.4, so at least three different sections could be devoted to using auxiliary information. This would be far too detailed for this book. Therefore, in subsection 2.2.5 a more general discussion about the use of auxiliary information is given, where the example provided is only one of many available methods. For more extensive reading about the use of auxiliary information the reader is referred to Goovaerts (1997) for geostatistical methods and to Särndal et al. (1992) and Brus and De Gruijter (1997) for design based methods.

Are the sample locations or sample time steps at which the property at scale s_1 is determined (to be) taken randomly?

If the property at scale s_1 is not known exhaustively within the s_2 support unit and the sample must still be taken, it is often convenient to select the sampling locations or time steps randomly. If this is the case, a so-called "design based" estimate of the spatial or temporal average is possible, using very simple expressions. Apart for the simple expressions used, the

advantage of design based estimates is that they do not make any assumptions about the spatial or temporal variation of the property considered. As with geostatistical methods, design based methods not only provide an estimate of the spatial or temporal average, but are also able to provide an estimate of the variance of the estimation error, which can be used as a measure of accuracy. As this estimate of the error variance is not based on assumptions about the spatial or temporal variation of the property, it is an objective measure, which makes design based methods better estimation tools for deciding legal matters than geostatistical methods. Design based methods are only possible if the sample locations are selected in some random manner. If not, either a geostatistical method or a deterministic method must be used. Design based methods are treated in subsection 2.2.2.

The choice between a design based and geostatistical method for estimating spatial averages is in fact more subtle than the question whether random sampling is possible or has been performed. However, this question is sufficiently discriminating for this book. A more elaborate exposé about the choice between design based and geostatistical methods is given in Brus and De Gruijter (1997).

Are there sufficient sample locations of the property at scale s_1 to estimate the semivariogram or correlogram?

Often, sample locations are not selected randomly. A common reason is that the sample has already been taken for another purpose than estimating the spatial or temporal average. For instance, sample locations around a waste site have been selected to delineate a soil pollution. Another reason is that grid sampling has been used in order to make a map of the property through interpolation at the scale of the observations s_1. Also, it is not always possible to take a random sample because parts of the terrain are not accessible or sampling at certain time steps is prohibited (e.g. sampling chloride concentration of a river during floods). In these cases, design based methods cannot be used and some representation of the spatial or temporal variation of the property is required. This representation can be either deterministic or stochastic. A stochastic representation means that the unknown values of the properties at the non-sampled locations are seen as the outcomes of random variables or random functions (see the Appendix). Suitable stochastic methods to describe the spatial and temporal[4] variation of

[4] To describe the variation of properties with time, "time series analysis" (Box and Jenkins, 1976) is a frequently used tool. It is not promoted to estimate temporal averages here, because it requires regular intervals between observations. Also, although time series analysis can be used to estimate temporal averages, its use for this purpose is not straightforward.

properties are provided by "geostatistics" (Journel and Huijbregts, 1978). Both geostatistics and deterministic methods can be used to interpolate a property between sample locations at the same scale (increasing the coverage) s_1 as well as estimating the spatial or temporal mean for a larger area at scale s_2 from a sample (combination of increasing the coverage and upscaling). The advantage of geostatistical methods over deterministic methods is that a measure of accuracy of the estimate is provided. Also, unlike most deterministic methods, geostatistics is particularly well suited for estimating spatial and temporal averages if the sample locations are clustered or preferentially sampled. However, to apply geostatistics a "semivariogram" or a "correlogram" must be estimated. The semivariogram (as does the correlogram) provides a measure of the difference in property values at two sample locations as function of the distance between these locations (see subsection 2.2.3 and the Appendix). The semivariogram and the correlogram can only be estimated if the sample is large enough. So, the property at scale s_1 must be known at a sufficient number of locations (Webster and Oliver, 1992) to be able to use geostatistics to estimate its average value for scale s_2. If this is not the case, a deterministic method must be used, and preferably one that is able to take account of possible clustering of sample locations (e.g. Thiessen polygons). Geostatistical methods are described in subsection 2.2.3, deterministic methods in subsection 2.2.4.

2.2.1 Exhaustive Information

Criteria

Figure 12:
- *No model is involved.*
- *Information about the observed property is exhaustive at scale s_1.*

Figure 13:
- *A model is involved.*
- *This model is linear in its input variables and parameters.*
- *Information about the properties (input variables and parameters) is exhaustive at scale* s_1.

Figure 14:
- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model can be applied at many locations or time steps.*
- *Information about the properties (input variables and parameters) is exhaustive at scale s_1.*

Method

Consider a support unit with support s_2 containing N_{12} support units with support s_1 ($s_1 < s_2$). If the property is known for all support units of scale s_1 within the target support unit of scale s_2, i.e. if the coverage is complete, the average value of the property for scale s_2 is simply obtained by averaging the values of the property of the s_1 support units. In mathematical terms: Let $z(s_1;i)$ denote the property z for support unit i at scale s_1, and $z(s_2) = \langle z(s_1;i) \rangle$ the (average) value of a property z for the support unit at scale s_2. This average property is calculated from the exhaustive information as:

$$z(s_2) = \langle z(s_1;i) \rangle = \frac{1}{N_{12}} \sum_{i=1}^{N_{12}} z(s_1;i) \tag{2.10}$$

In the continuum approach the averaging of exhaustive information is described as follows: Consider a support unit with support $|s_2|$ containing points with support $|s_1|$ ($|s_1| < |s_2|$). Let $z(s_1;\mathbf{x})$ denote the property for a support unit at point $\mathbf{x}$ with support $|s_1|$, and $z(s_2;\mathbf{y}) = \langle z(s_1;\mathbf{x}) \rangle$ the (average) value of a property z for the support unit at support $|s_2|$ with central point $\mathbf{y}$. The average property $z(s_2;\mathbf{y})$ is calculated from the exhaustive information as:

$$z(s_2;\mathbf{y}) = \langle z(s_1;\mathbf{x}) \rangle = \frac{1}{|s_2|} \int_{\mathbf{x} \in s_1} z(s_1;\mathbf{x})\, d\mathbf{x} \tag{2.11}$$

Example

Examples of the upscaling of exhaustive information are trivial. An example (*Figure 19*) in the temporal domain is the upscaling of measured daily precipitation (s_1) to average precipitation for a decade (s_2). Another example would be the construction of larger support digital elevation model (DEM) from a smaller support DEM.

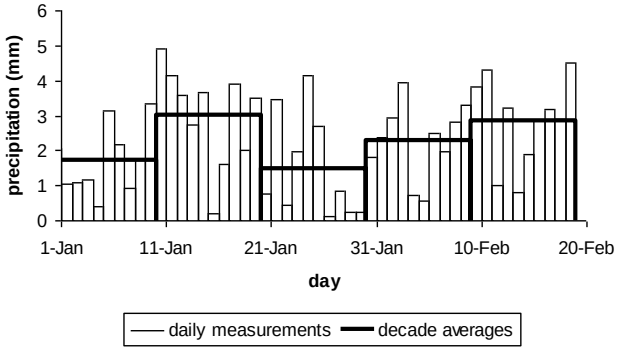

Figure 19. Fictitious example of the upscaling of daily precipitation to decade averages

2.2.2 Design based methods

Criteria

Figure 12:

- *No model is involved.*
- *Information about the observed property is not exhaustive at scale s_1.*
- *No auxiliary information is present that explains part of the variation of the property at scale s_1.*
- *The sample locations or sample time steps at which the property at scale s_1 is determined are (to be) taken randomly.*

Figure 13:

- *A model is involved.*
- *This model is linear in its input variables and parameters.*
- *Information about the properties (input variables and parameters) is not exhaustive at scale s_1.*
- *No auxiliary information is present that explains part of the variation of the property at scale s_1.*
- *The sample locations or sample time steps at which the property at scale s_1 is determined are (to be) taken randomly.*

Figure 14:

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model can be applied at many locations or time steps.*
- *Information about the properties (input variables and parameters) is not exhaustive at scale s_1.*

– *No auxiliary information is present that explains part of the variation of the property at scale s_1.*
– *The sample locations or sample time steps at which the property at scale s_1 is determined are (to be) taken randomly.*

Method

Design based methods (Särndal, 1978; Brus and De Gruijter, 1997) provide objective estimates of spatial or temporal means and also of the variance of the estimation error due to the fact that no assumptions about the variation are made. The only prerequisite is that properties at support s_1 are collected by such a design that the probability of selecting a location is known. These probabilities are then used as weights in the estimation of population means from the samples at s_1 within a support unit at scale s_2. In the general case, an estimate of the average value at scale s_2 follows from n samples at scale s_1 as:

$$\hat{z}(s_2) = \frac{\sum_{i=1}^{n} \frac{z(s_1;i)}{\pi_i}}{N} \tag{2.12}$$

where the N is the population size (i.e. the number of support units s_1 in s_2), and π_i is the inclusion probability. In case the sampling locations are selected at random with equal probabilities of selection and independently from each other (simple random sampling), π_i would be equal to the sampling fraction n/N for all i. As a consequence, the equation to estimate the average at s_2 reduces to the arithmetic average:

$$\hat{z}(s_2) = \sum_{i=1}^{n} \frac{1}{n} z(s_1;i) \tag{2.13}$$

Furthermore, from the sampling design the precision of the estimated average value can be calculated. This allows for the analysis of the efficiency of several sampling designs when the variability is quantified a priori (cf. Brus, 1994 and Domburg et al., 1994), and also allows for the testing of hypotheses (cf. the example below).

Sometimes the support unit is divided in sub-units (strata) that are being sampled separately with different sampling densities (stratified sampling), but still at random within each stratum. In this case $z(s_2)$ would estimated by the weighted mean of the stratum means:

$$\hat{z}(s_2) = \sum_{h=1}^{L} w_h \cdot \hat{\bar{z}}_h \qquad (2.14)$$

where the weights w_h follow from the fraction of the support unit s_2 occupied by stratum h and $\hat{\bar{z}}_h$ is the estimated stratum mean calculated with Equation (2.12). Stratification is often based on the use of prior information on variability patterns, which is a form of auxiliary information and is therefore addressed in subsection 2.2.5. Stratification can also be applied to ensure that the complete extent is sampled. This was the case in the example.

Example

The example is taken from Finke et al. (1996a,b) and deals with the characterisation of a Dutch soil map unit with Podsol soils in terms of its physical and chemical behaviour. The approach follows that depicted in *Figure 14*. Water and solute flow models were applied at locations s_1, each with an areal support of approximately 1/16 m^2. The models resulted in an estimation of the number of days with a good workability and with good aeration, the number of days to Chloride breakthrough, the percentage of breakthrough of Cadmium and of a herbicide, both after 1 year. These results were upscaled to the soil map unit at scale s_2 with a support area of 115 km^2.

The locations were obtained following a stratified two-stage random sampling design (Cochran, 1977, *Figure 20*). Eleven map sheets of the soil map served as strata to ensure sampling the complete extent of the map unit. In each map sheet, two map unit polygons were selected at random with replacement (stage one), where the drawing probability of a polygon was proportional to its acreage. In each polygon, four sampling locations were allocated at random (stage two). The two-stage approach was chosen because it results in clustered observations, thus reducing travel cost while maintaining a random sampling design.

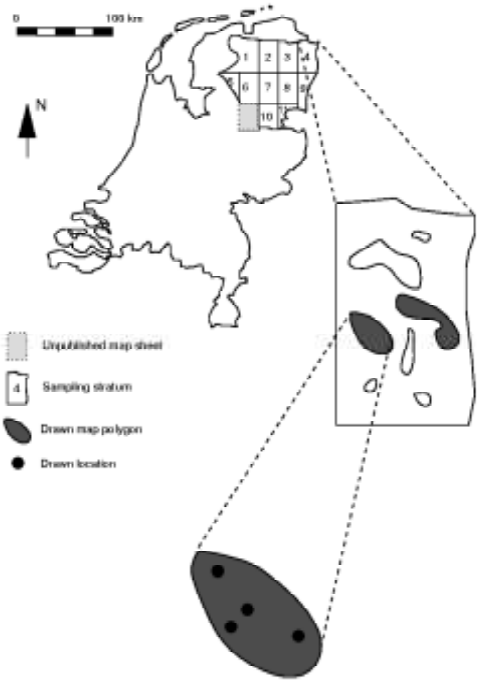

Figure 20. Stratified two-stage sampling procedure. Reproduced from Finke et al., 1996b

.

The study resulted in soil map unit averages and associated standard errors for the simulated properties. Another purpose of the study was to test whether the behaviour of a-priori selected "representative profiles" deviated significantly from the mean. The study revealed the weakness of this -still quite common- approach to estimate the average behaviour of a map unit, since it was shown that this introduced significant bias for all properties at minimally 90% confidence (Finke et al., 1996a).

Literature

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

Domburg, P., J.J. de Gruijter and D.J. Brus. 1994. A structured approach to designing soil survey schemes with prediction of sampling error from variograms. Geoderma 62: 151-164.

Finke, P.A., J.H.M. Wösten and J.G. Kroes. 1996a. Comparing two approaches of characterizing soil map unit behavior in solute transport. Soil Science Society of America Journal 60: 200-205.
Finke, P.A., J.H.M. Wösten and M.J.W. Jansen. 1996b. Effects of uncertainty in major input variables on simulated functional soil behaviour. Hydrological processes 10: 661-669.
Särndal, C.E. 1978. Design-based and model-based inference in survey sampling. Scandinavian Journal of Statistics 5: 27-52.

2.2.3 Geostatistical prediction

Criteria

Figure 12:
- *No model is involved.*
- *Information about the observed property is not exhaustive at scale s_1.*
- *No auxiliary information is present that explains part of the variation of the property at scale s_1.*
- *The sample locations or sample time steps at which the property at scale s_1 is determined are not (to be) taken randomly.*
- *There are sufficient sample locations of the property at scale s_1 to estimate the semivariogram or correlogram.*

Figure 13:
- *A model is involved.*
- *This model is linear in its input variables and parameters.*
- *Information about the properties (input variables and parameters) is not exhaustive at scale s_1.*
- *No auxiliary information is present that explains part of the variation of the property at scale s_1.*
- *The sample locations or sample time steps at which the property at scale s_1 is determined are not (to be) taken randomly.*
- *There are sufficient sample locations of the property at scale s_1 to estimate the semivariogram or correlogram .*

Figure 14:
- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model can be applied at many locations or time steps.*
- *Information about the properties (input variables and parameters) is not exhaustive at scale s_1.*
- *No auxiliary information is present that explains part of the variation of the property at scale s_1.*
- *The sample locations or sample time steps at which the property at scale s_1 is determined are not (to be) taken randomly.*

– *There are sufficient sample locations of the property at scale s_1 to estimate the semivariogram or correlogram .*

Method

Geostatistical methods describe the unknown variation of z_1 within z_2 as a realisation of a stochastic function (see the Appendix). A comprehensable introduction to geostatistics is given by Isaaks and Srivastava (1989). Here we will focus on a geostatistical method called block kriging. Block kriging is (in the spatial domain) the most commonly applied geostatistical method to change the support of observations or model outcomes from s_1 to s_2. This method can be applied in the temporal domain as well, but we will focus here on the spatial domain. The method described below (ordinary block kriging) assumes that the intrinsic hypothesis holds: (i) the statistical expectation of the mean difference between two values is zero, and (ii) for all values at a certain mutual distance **h** the variance is finite and does not depend on the location. Block kriging requires a sufficiently large number of observations at s_1 and a function called the semivariogram (Journel and Huijbregts, 1978; Burgess and Webster, 1980). The semivariogram describes the expected variance of the differences of two property values Z as a function of their distance **h** (Appendix) and is estimated from the data by:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2n(\mathbf{h})} \sum_{i=1}^{n(h)} (z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h}))^2 \qquad (2.15)$$

where $n(\mathbf{h})$ is the number of observations with mutual distance **h**.

To estimate this sample variogram properly, sufficient data are needed (cf. Webster and Oliver, 1992). The sample variogram is then modelled by fitting a so-called permitted function (e.g. an exponential model, cf. Journel and Huijbregts, 1978). Block kriging results in a predicted value at scale s_2 as well as the variance of the prediction error. This variance quantifies the accuracy of the interpolated values at the target scale and can be used to optimise sampling density. Here we will focus on the prediction itself.

The predictor Z at scale s_2 is obtained from a weighted linear combination of observations at s_1 (Burgess and Webster, 1980):

$$Z(s_2) = \sum_{i=1}^{n} \lambda_i \cdot Z(s_1; \mathbf{x}_i) \qquad (2.16)$$

The weights λ_i from the vector $\boldsymbol{\lambda}$ are obtained by solving the block kriging system:

$$\begin{bmatrix} \boldsymbol{\lambda} \\ \mu \end{bmatrix} = \mathbf{A}^{-1} \cdot \mathbf{s} \tag{2.17}$$

where μ is a Lagrange multiplier necessary to solve the system, **A** is a matrix with semivariances between the data points at scale s_1:

$$\mathbf{A} = \begin{bmatrix} \gamma(\mathbf{x}_1,\mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_n,\mathbf{x}_1) & 1 \\ \vdots & & \vdots & 1 \\ \gamma(\mathbf{x}_1,\mathbf{x}_n) & \cdots & \gamma(\mathbf{x}_n,\mathbf{x}_n) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \tag{2.18}$$

and the vector **s** contains the average semivariances between the data points at scale s_1 and all points in the support unit s_2:

$$\mathbf{s} = \begin{bmatrix} \int \gamma(\mathbf{x}_2,\mathbf{x}) \cdot p(\mathbf{x}) d\mathbf{x} \\ \vdots \\ \int \gamma(\mathbf{x}_n,\mathbf{x}) \cdot p(\mathbf{x}) d\mathbf{x} \\ 1 \end{bmatrix} \tag{2.19}$$

In the calculation of **s**, $p(\mathbf{x})$ equals the reciprocal area of s_2 if x is located inside the support unit and equals zero otherwise, where $\int p(\mathbf{x})\mathrm{d}\mathbf{x}=1$.

Example

Samra et al. (1988a,b) describe a case study on mapping the sodicity in a field with a rice-wheat rotation. The purpose of the study was (i) to investigate if observed variations in wheat growth were associated with variability in sodicity within the field, and (ii) to locate the most appropriate place to plant an orchyard with sodium-sensitive tree species. Soil samples were taken at different depths and at a grid of 12x12 m^2. Each analysed sample was a composite from 4 random samples within an area of 1x1 m^2, so the support s_1 was $1m^2$. The sodicity was mapped by block kriging the Sodium Adsorption Ration (SAR) and the Exchangeable Sodium Percentage (ESP) at supports s_2 of 12x12 m^2 but with a grid spacing of 6x6 m^2 (the planting distance of the planned orchard). Results were compared with those of kriging without change of support (punctual kriging). Maps obtained by block kriging were somewhat less erratic than those by punctual kriging due to the smoothing effect of block kriging (*Figure 21*). A zone could be delineated were sodium-sensitive trees were best planted.

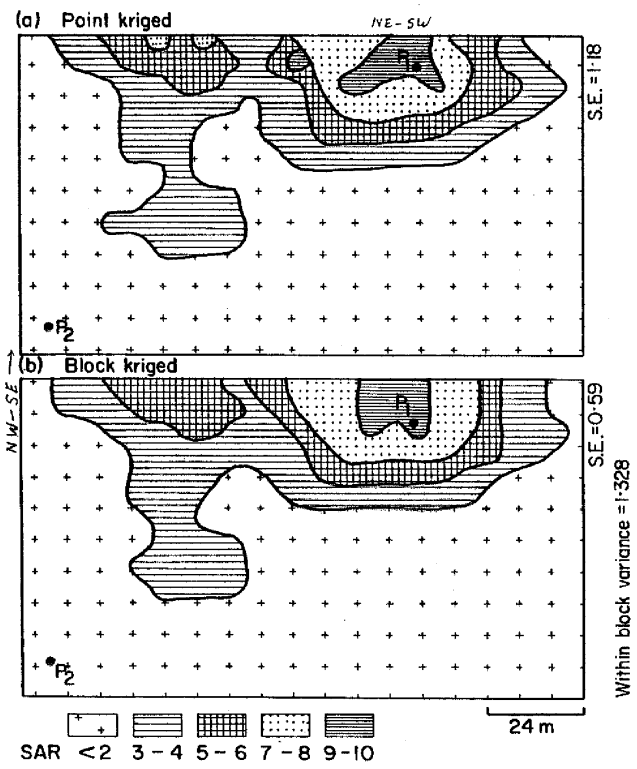

Figure 21. Maps of SAR in the layer between 0-30 cm by point and block kriging. Reproduced from Samra et al., 1988b

Literature

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

Burgess, T.M. and R. Webster. 1980. Optimal interpolation and isarithmic mapping of soil properties. II Block kriging Journal of Soil Science 31: 333-341.

Isaaks, E.H. and R.M. Srivastava. 1989. Applied geostatistics. Oxford University Press, New York, 561 pp.

Journel, A.G. and Ch. J. Huijbregts. 1978. Mining geostatistics. Academic Press, New York. 599 pp.

Samra, J.S., K.N.S. Sharma and N.K. Tyagi. 1988a. Analysis of spatial variability in sodic soils: 1. Structural analysis. Soil Science 145: 180-187.

Samra, J.S., V.P. Singh and K.N.S. Sharma. 1988b. Analysis of spatial variability in sodic soils: 2. Point - and block-kriging. Soil Science 145: 250-256.

Webster, R. and M.A. Oliver. 1992. Sample adequately to estimate variograms of soil properties. Journal of Soil Science 43: 177-192.

2.2.4 Deterministic functions

Criteria

Figure 12:

- *No model is involved.*
- *Information about the observed property is not exhaustive at scale s_1.*
- *No auxiliary information is present that explains part of the variation of the property at scale s_1.*
- *The sample locations or sample time steps at which the property at scale s_1 is determined are not (to be) taken randomly.*
- *There are not sufficient sample locations of the property at scale s_1 to estimate the semivariogram or the correlogram.*

Figure 13:

- *A model is involved.*
- *This model is linear in its input variables and parameters.*
- *Information about the properties (input variables and parameters) is not exhaustive at scale s_1.*
- *No auxiliary information is present that explains part of the variation of the property at scale s_1.*
- *The sample locations or sample time steps at which the property at scale s_1 is determined are not (to be) taken randomly.*
- *There are not sufficient sample locations of the property at scale s_1 to estimate the semivariogram or the correlogram.*

Figure 14:

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model can be applied at many locations or time steps.*
- *Information about the properties (input variables and parameters) is not exhaustive at scale s_1.*
- *No auxiliary information is present that explains part of the variation of the property at scale s_1.*
- *The sample locations or sample time steps at which the property at scale s_1 is determined are not (to be) taken randomly.*
- *There are not sufficient sample locations of the property at scale s_1 to estimate the semivariogram or the correlogram.*

Method

Deterministic methods in upscaling usually involve interpolation. The difference between geostatistical and deterministic methods is that with

deterministic methods the weights assigned to sample data depend solely on the sample configuration, whereas with geostatistical methods the weights depend on both the sample configuration and a model of spatial (or temporal) variation estimated from the datat (i.e. the semivariogram).

Example

An example of a deterministic upscaling method is the delineation of influence zones around sample locations by Thiessen polygons. Within each polygon, the value of a property is assumed constant. Thiessen polygons (*Figure 22*) are often used to assign data from meteorological stations to areas (Chow et al., 1988). Algorithms to delineate Thiessen polygons are standard routines in most GIS-packages.

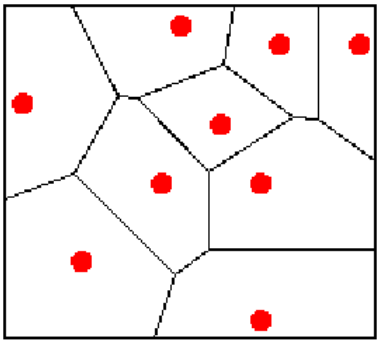

Figure 22. Thiessen polygons to assign weather stations to areas

Upscaling with Thiessen polygons is done by calculating the mean at s_2:

$$z(s_2) = \sum_{i=1}^{L} \frac{A_i}{A} \cdot z(i) \qquad (2.20)$$

where A_i is the area and z(i) the value of Thiessen polygon i within s_2 and A the total area of s_2.

Another example is the application of inverse distance weighted interpolation of a property followed by averaging over a larger support, where the weight assigned to each sample value is its reciprocal distance to the interpolation location. An example of inverse distance interpolation onto rainfall data measured on the weather stations in *Figure 22* is given in *Figure 23*.

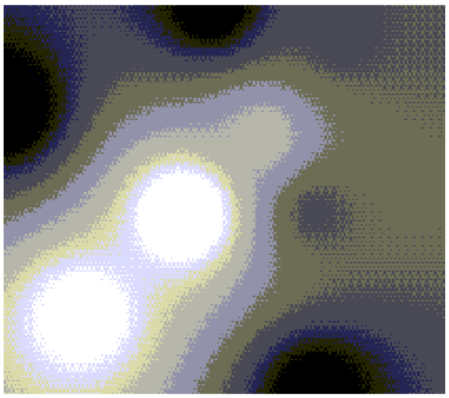

Figure 23. Inverse distance interpolation of precipitation data

Upscaling after inverse distance interpolation is done by simply calculating the mean at s_2 of the interpolated values, since all interpolated values have equal supports s_1:

$$z(s_2) = \frac{1}{n} \cdot \sum_{i=1}^{n} z(i) \tag{2.21}$$

where n is the number of interpolated values and $z(i)$ are the interpolated values.

Thiessen polygons have an advantage over inverse distance interpolations when data points are spatially clustered. With inverse distance interpolation, the clustered data will dominate the map, while Thiessen polygons around data points within a cluster will be smaller than polygons around more isolated data points. This property is used for a technique called spatial declustering (cf. Isaacs and Srivastava, 1989), where each sample location is assigned a weight according to the areal fraction occupied by its Thiessen polygon. This technique is applied in Eq. (2.20).

Literature

Isaacs, E.H. and R.M. Srivastava. 1989. Applied geostatistics. Oxford University Press, New York. 561 pp.

Chow, V.T., D.R. Maidment, L.W. Mays. 1988. Applied hydrology. McGraw-Hill, New York. 572 pp.

2.2.5 Combinations and auxiliary information

Criteria

Figure 12:
- *No model is involved.*
- *Information about the observed property is not exhaustive at scale s_1.*
- *Auxiliary information is present that explains part of the variation of the property at scale s_1.*

Figure 13:
- *A model is involved.*
- *This model is linear in its input variables and parameters.*
- *Information about the properties (input variables and parameters) is not exhaustive at scale s_1.*
- *Auxiliary information is present that explains part of the variation of the property at scale s_1.*

Figure 14:
- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model can be applied at many locations or time steps.*
- *Information about the properties (input variables and parameters) is not exhaustive at scale s_1.*
- *Auxiliary information is present that explains part of the variation of the property at scale s_1.*

Method

There are many ways in which auxiliary information can be used in scale-related studies. One distinction that could be made, is its use as either prior information or during the scale transfer itself. Auxiliary data can be used before sampling to minimise the sampling effort while maintaining an acceptable sampling error or to minimise sampling error, since part of variability is already explained by the auxiliary data. Further on in the research chain, it can be used to improve the interpolation and upscaling of a property.

Generally speaking, the use of auxiliary information will be more successful when its coverage is higher (it should have a higher coverage than the property to-be-upscaled), and when it is more firmly related to the target property (the auxiliary property should explain some of the variance of the target property). Last but not least the ability of the researcher to recognise useful auxiliary information as such is to be mentioned.

It is beyond the scope of this book to give a complete overview of methods to exploit auxiliary data. Nevertheless, a decision tree to assist with

the choice out of some methods is given in *Figure 24*. Criteria that can be applied to choose a method to exploit auxiliary information are:

Is the auxiliary variable quantitative or qualitative (thematic)?

If the auxiliary variable is of a thematic type, it can still be used for stratification during sampling and as a indicator variable during prediction. Documented examples are for instance the use of soil maps to enhance the efficiency of sampling by stratified sampling (eg. Brus, 1994), or to obtain predictions at s_1 by regression with map units as qualitative predictors (cf. Dobson, 1990) and subsequent averaging to s_2. Stein et al. (1988) provide an example on how to improve map unit estimates by stratified kriging. Bierkens (1997) provides an application of stratified residual kriging to improve predictions within map strata. Bierkens and Burrough (1993a,b) and Goovaerts and Journel (1995) illustrate the use of indicator kriging when thematic variables are available with high coverage (i.e. as maps), but it is also possible to use thematic data at lower coverages and with s_1-support. Goovaerts (1998) gives a comprehensive overview of indicator kriging based methods.

If the auxiliary variable is quantitative, it can also be used both to reduce the sampling effort of the target property and to improve the prediction of the property at scale s_2. Geostatistical techniques that can be applied are (inter alia) co-kriging, regression kriging and external drift kriging. Design based methods can be used in the form of regression estimators.

Should a design based, a geostatistical or a deterministic method be applied?

This criterion should be evaluated the same way as before: if some kind of random sampling design was applied to obtain the target property, design based methods should be preferred. Quantitative auxiliary data can then be used as regression estimators (e.g. Brus, 2000), whereas thematic auxiliary data can be used to define sampling strata. If the sampling design was not random, or unknown, geostatistical methods are available (cf. Goovaerts, 1997). If the auxiliary information is insufficient to exploit in a statistical sense, deterministic methods might be the only realistic option. Scarce auxiliary information may still be useful when applied in a deterministic way, e.g. to define spatial aggregation schemes. An example is the assignment of measurements from evapotranspiration stations with a low spatial coverage to rainfall stations with a higher spatial coverage using Thiessen polygons (Chow et al., 1988).

What is the coverage of the auxiliary information itself?

When the auxiliary variable has full (exhaustive) coverage, it can be used as external drift while block kriging the target property (Bourennane et al., 1996). Non-exhaustive auxiliary information can be used as co-variable in block co-kriging or as "soft data" in regression block kriging (Knotters et al., 1995).

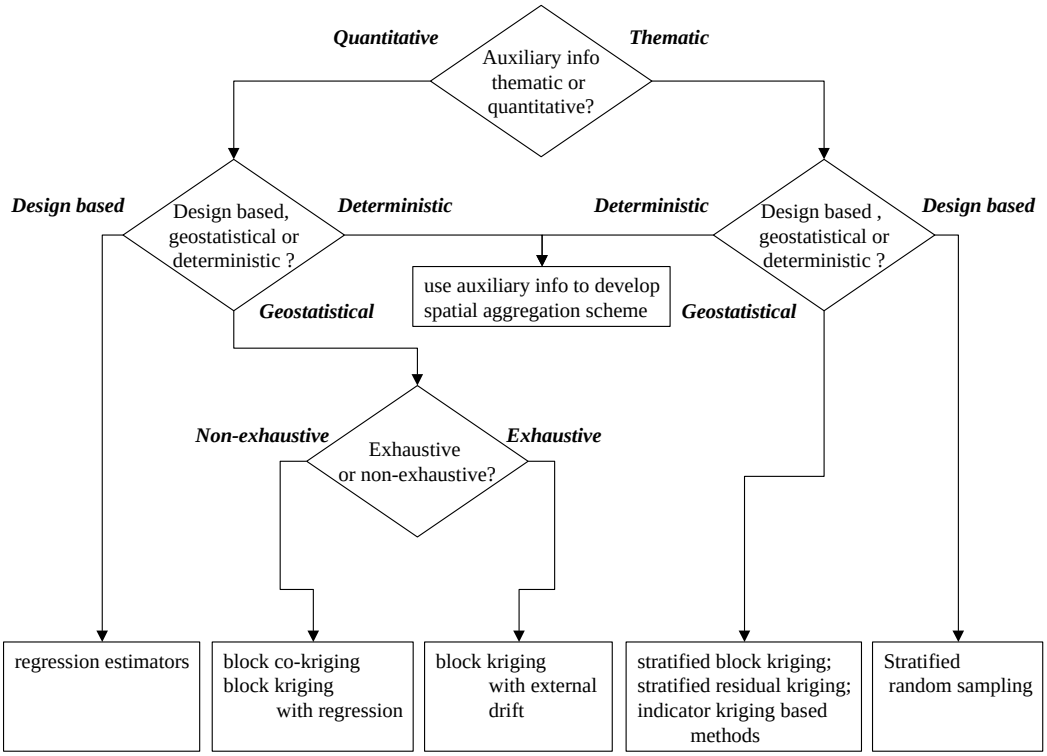

Figure 24. Decision tree to choose among techniques utilising auxiliary information when upscaling

Example

The example is taken from Pebesma and De Kwaadsteniet (1997) and concerns stratified block kriging. In this study, maps of groundwater quality variables were obtained from point scale measurements with predefined soil-landuse categories for auxiliary information. These soil-landuse categories originated from an overlay of a simplified soil map and a land use map, both with complete coverage. Thus, the auxiliary information was both thematic and exhaustive. The target property was derived from measurements at

monitoring plots that were not allocated by a random design. Thus, design based methods would not have been appropriate (*Figure 24*).

Target properties were yearly (1987-1993) measured concentrations of Ca, Cl, EC, pH, HCO_3,K, Mg, Na, NH_4-N, NO_3-N, P-tot and SO_4 in a total of 425 monitoring sites, and a total of 25 properties for the year 1991 only. These measurements were obtained from samples from groundwater wells with screen depths between 5 and 17 below ground surface, and can be considered representative for small areas only. To allow for the construction of maps with national coverage, it was decided to increase the support to blocks of 4x4 km^2 by block kriging. Semivariogram models were fitted to log-transformed concentrations within each soil-landuse category. Subsequently, block kriging was done for each stratum separately, using the stratum semivariogram and the log-transformed data values from within the stratum. The block kriging standard deviations were used to construct 95% confidence intervals as well. Block kriged values and 95% confidence limits for 4x4 km^2 blocks were backtransformed, yielding block-median values and associated 95% prediction intervals. Maps were produced indicating whether certain reference levels would fall below, inside or above the 95% confidence interval (*Figure 25*).

Pebesma and De Kwaadsteniet examined the effect of stratification by landuse and soil. They concluded that stratification was essential in terms of preserving spatial differentiation.

Literature

Bierkens, M.F.P. and P.A. Burrough. 1993a. The indicator approach to categorical soil data; part 1: Theory. Journal of Soil Science 44: 361-368.

Bierkens, M.F.P. and P.A. Burrough. 1993b. The indicator approach to categorical soil data; part 2: Application to mapping and land use suitability analysis. Journal of Soil Science 44: 369-381.

Bierkens, M.F.P. 1997. Using stratification and residual kriging to map soil pollution in urban areas. In: E.Y. Bafi and N.A. Schofield (eds.). Geostatistics Wollongong, 96: 996-1007. Kluwer Academic Publishers, Dordrecht, Netherlands.

Bourrenane, H., D. King and P. Chery. 1996. Improving the kriging of a soil variable using slope gradient as external drift. European Journal of Soil Science 47: 473-483.

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. 2000. Using regression models in design-based estimation of spatial means of soil properties. European Journal of Soil Science 51(1): 159-172.

Chow, V.T., D.R. Maidment, L.W. Mays. 1988. Applied hydrology. McGraw-Hill, New York. 572 pp.

Dobson, A.J. 1990. An introduction to generalized linear models. Chapman and Hall. London.

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

Goovaerts, P. and A.G. Journel. 1995. Integrating soil map information in modelling the spatial variation of continuous soil properties. Europ. Journal of Soil Science 46: 397-414.

Knotters, M., D.J. Brus and J.H. Oude Voshaar. 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma 67: 227-246.

Pebesma, E.J. and J.W. De Kwaadsteniet. 1997. Mapping groundwater quality in the Netherlands. Journal of Hydrology 200: 364-386.

Stein, A., M. Hoogerwerf and J. Bouma. 1988. Use of soil-map delineations to improve (co-)kriging of point data on moisture deficits. Geoderma 43: 163-177.

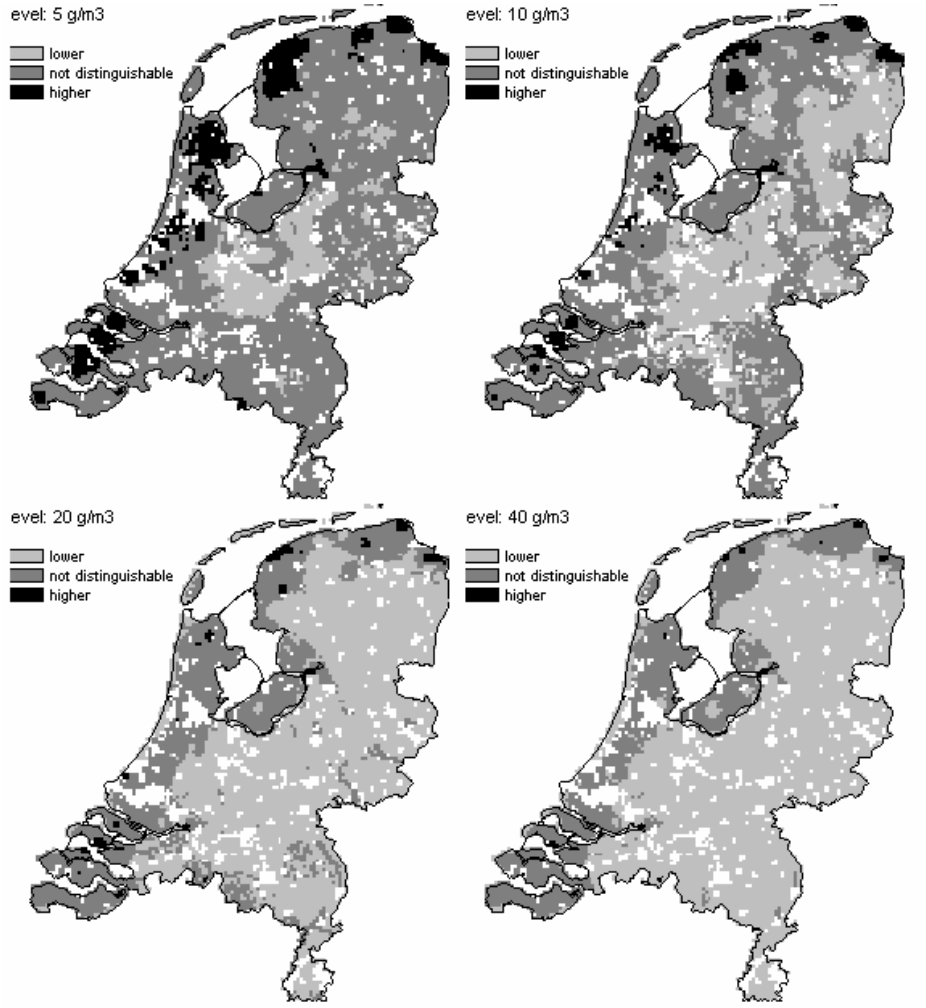

Figure 25. Map of potassium in the groundwater, indicating whether four concentration levels are lower, not distinghuishable or higher than the 95% confidence interval of 4*4 km block median values. Reproduced from Journal of Hydrology 200, Pebesma and De Kwaadsteniet, Mapping groundwater quality in the Netherlands, pp. 364-386. Copyright (1997), with permission from Elsevier Science

2.3 Finding representative parameters or input variables

Criteria

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model cannot be applied at many locations or time steps.*
- *The model has the same form at the two scales involved.*

The appropriate upscaling procedure for this case is shown in *Figure 15*. The method proceeds as follows. Suppose that at the smaller scale s_1 the following model is used to calculate the output $y(s_1;i)$ from the parameters or input variables $z(s_1;i)$:

$$y(s_1;i) = g[z(s_1;i)] \tag{2.22}$$

Our target is to calculate the spatial or temporal average $y(s_2) = \langle y(s_1;i)\rangle$, using the same model form as (2.22). So the upscaling problem consists of finding representative parameters or input variables $z_R(s_2)$ at scale s_2 such that when inserted in (2.22) this results in the output variable $y(s_2) = \langle y(s_1;i)\rangle$:

$$y(s_2) = g[z_R(s_2)] \tag{2.23}$$

where $z_R(s_2)$ is in some way calculated from the parameters or input variables of the N_{12} support units at scale s_1 within s_2:

$$\{z(s_1;i), i = 1,..,N_{12}\} \rightarrow z_R(s_2) \tag{2.24}$$

An archetypal example of upscaling through finding representative parameters is the problem of finding the representative conductivity for blocks in a numerical model of groundwater flow. With the emergence of high-resolution seismic methods and the use of geostatistical simulation (see Goovaerts, 1997), is it possible to obtain images of the hydraulic conductivity of the subsoil for very small support units s_1. However, to use the images at this scale in a numerical groundwater model would mean the calculation of groundwater head for many millions of nodes. Because of excessive computation times, this is not yet feasible for most practical applications. As a result, numerical groundwater models usually operate with much coarser grids. Consequently, it is necessary to upscale the fine scale hydraulic conductivity at s_1 to larger blocks s_2.

In the continuum approach (see section 1.2) groundwater flow is described with Darcy's law (for brevity of notation we have omitted the scale variable s_1 from $h(s_1;\mathbf{x})$, $q(s_1;\mathbf{x})$ and $k(s_1;\mathbf{x})$):

$$\mathbf{q}(\mathbf{x}) = -\mathbf{k}(\mathbf{x})\nabla h(\mathbf{x}) \tag{2.25}$$

where $\mathbf{x}$ is the spatial co-ordinate, $\mathbf{q}$ is the groundwater flux [LT^{-1}] (volume of water crossing a unit area per unit time), which is a vector with components in each direction, h the hydraulic head (pressure and gravity) [L] of the groundwater,∇h the hydraulic head gradient vector [-], e.g. in three (Cartesian) dimensions: $\nabla h = \left(\partial h/\partial x, \partial h/\partial y, \partial h/\partial z\right)^{\mathrm{T}}$, and $\mathbf{k}$ [LT^{-1}] a matrix (also called tensor) containing the hydraulic conductivities in the different directions. Often, at the smallest scale, the hydraulic conductivity is assumed to be the same in all directions, so that the matrix $\mathbf{k}$ is replaced by a single scalar parameter k. Now suppose that we need the representative conductivity $\mathbf{k}_{\mathrm{R}}$ of model block V with size $|V|$, given that the small scale conductivity $\mathbf{k}(\mathbf{x})$ is known. If inserted in (2.25), together with the average head gradient of the block, the representative conductivity should yield the block-averaged flux:

$$\frac{1}{|V|}\int_{\mathbf{x}\in V}\mathbf{q}(\mathbf{x})d\mathbf{x} = -\mathbf{k}_{\mathrm{R}}\frac{1}{|V|}\int_{\mathbf{x}\in V}\nabla h(\mathbf{x})d\mathbf{x} \tag{2.26}$$

So, the upscaling problem consists of finding $\mathbf{k}_{\mathrm{R}}$ from the fine scale hydraulic conductivities within the block V such that Equation (2.26) is satisfied.

It is worthwhile to explore this example a little further. Let $<\mathbf{k}(\mathbf{x})>$ denote the (arithmetic) average of hydraulic conductivity within the block V, i.e.:

$$<\mathbf{k}(\mathbf{x})> = \frac{1}{|V|}\int_{\mathbf{x}\in V}\mathbf{k}(\mathbf{x})d\mathbf{x} \tag{2.27}$$

and $<\nabla h(\mathbf{x})>$ and $<\mathbf{q}(\mathbf{x})>$ the averages of the gradient and the flux vector. The value of each of these variables within the block can be written as its average plus a deviation from the average: $\mathbf{k}(\mathbf{x}) = <\mathbf{k}(\mathbf{x})> + \mathbf{k}'(\mathbf{x})$, $\nabla h(\mathbf{x}) = <\nabla h(\mathbf{x})> + \nabla h'(\mathbf{x})$, and $\mathbf{q}(\mathbf{x}) = <\mathbf{q}(\mathbf{x})> + \mathbf{q}'(\mathbf{x})$, where by definition we have $<\mathbf{k}'(\mathbf{x})> = <\nabla h'(\mathbf{x})> = <\mathbf{q}'(\mathbf{x})> = 0$. If these representations are inserted in Darcy's law (2.25) we obtain:

$$< \mathbf{q}(\mathbf{x}) > + \mathbf{q}'(\mathbf{x}) = -\left(< \mathbf{k}(\mathbf{x}) > + \mathbf{k}'(\mathbf{x})\right)\left(< \nabla h(\mathbf{x}) > + \nabla h'(\mathbf{x})\right) \tag{2.28}$$

An equation describing the block average flux $<\mathbf{q}(\mathbf{x})>$ is obtained from (2.28) by taking the average of both sides of the equation:

$$\begin{aligned} < \mathbf{q}(\mathbf{x}) > &= <-\left(< \mathbf{k}(\mathbf{x}) > + \mathbf{k}'(\mathbf{x})\right)\left(< \nabla h(\mathbf{x}) > + \nabla h'(\mathbf{x})\right) > \\ &= - < \mathbf{k}(\mathbf{x}) >< \nabla h(\mathbf{x}) > - < \mathbf{k}'(\mathbf{x}) \nabla h'(\mathbf{x}) > \end{aligned} \tag{2.29}$$

where the expression in the second line is found by expanding the first line and taking account of the fact that $< \mathbf{k}'(\mathbf{x}) >=< \nabla h'(\mathbf{x}) >=< \mathbf{q}'(\mathbf{x}) >= 0$. If we look for a representative parameter for which we have $<\mathbf{q}(\mathbf{x})> = -\mathbf{k}_R <\nabla h(\mathbf{x})>$ we can conclude from (2.29) that this representative conductivity must obey the following relation:

$$\mathbf{k}_R < \nabla h(\mathbf{x}) > \; = \; < \mathbf{k}(\mathbf{x}) >< \nabla h(\mathbf{x}) > + < \mathbf{k}'(\mathbf{x}) \nabla h'(\mathbf{x}) > \tag{2.30}$$

Equation (2.30) shows that the expression for the representative conductivity consists of a term involving the arithmetic block average of conductivity and head gradient and an additional term involving the block averaged co-variation of the deviations $\mathbf{k}'(\mathbf{x})$ and $\nabla h'(\mathbf{x})$. This third term is called a "closure term" and finding a suitable expression for this term is called a "closure problem". Unfortunately, the solution to this problem not only depends on the properties of the porous medium within the block, but also on the boundary conditions (i.e. the heads at the block interfaces). However, to know these boundary conditions one needs to know the flow at the smaller scale s_1. This means that determining representative block conductivity is a vicious circle, because the reason we want to upscale is that we cannot solve the flow at scale s_1 due to computational difficulties. The dependence of the closure problem on the boundary conditions around the block is another way of saying that the closure problem not only depends on the hydraulic conductivities and hydraulic heads within the block, but also on the hydraulic conductivities and hydraulic heads outside the block. Consequently, the solution to the closure problem is said to be "non-local". To find a local solution (to break out of the vicious circle) usually a particular set of boundary conditions is assumed, such as uniform flow (Dagan, 1981) or periodic boundary conditions (Durlofski, 1991). Moreover, to find an analytical solution (i.e. a formula) to the closure problem, the porous medium is often represented by (sometimes overly) simple

analogues. Therefore, almost all solutions for representative conductivity are not only non-unique (i.e. valid for a certain set of boundary conditions) but also approximations. Extensive reviews about the various ways of calculating representative conductivity are given by Wen and Gómez-Hernández (1996) and Renard and de Marsily (1997).

The reason for the elaborate treatment of the hydraulic conductivity example is that it is typical for all upscaling problems that involve finding representative parameters for non-linear models. The same non-local closure problem will occur if the upscaling problem is analysed in its most general form. Representative parameters are almost never unique, and most (local) expressions that have been derived to calculate representative parameters at scale s_2 from the parameters at scale s_1 are approximations.

A further subdivision of upscaling by representative parameters results from the following questions that make up the decision tree of *Figure 26*:

- *Is it possible to obtain the output variables at the larger scale s_2?*
- *Is the information about the input variables and parameters exhaustive at scale s_1?*
- *Is the unknown temporal or spatial variation of input variables or parameters at scale s_1 to be described deterministically?*

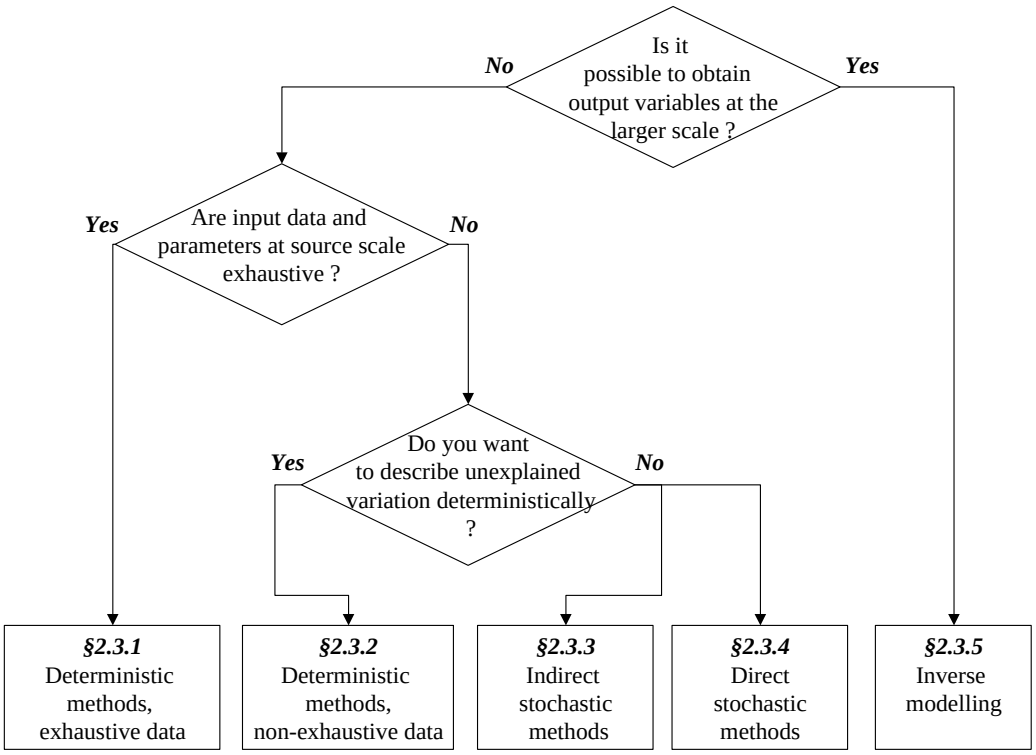

Figure 26. Decision tree to five subclasses of upscaling class 2

These questions and the results of their answers are discussed given hereafter.

Is it possible to obtain the output variables at the larger scale s_2?

If it is possible to measure the output variable directly at the scale s_2, the actual upscaling problem can be circumvented. For instance, in rainfall-runoff modelling the output variable is the catchment's discharge, which can be measured at the catchment's outlet. Suppose that we measure this discharge, say $q(s_2;t)$ at n instances in time, yielding a time series of discharges $q(s_2; k)$, $k = 1,..,n$. Suppose that the discharge depends on the input time series of precipitation and on the infiltration capacity $r(s_1;i)$ that varies with space per support unit i at scale s_1 within the catchment (which is scale s_2). Instead of finding special averaging rules to find the representative parameter $r_R(s_2)$, this parameter can also be found by minimising the following sum of squares of deviations *SSD*:

$$SSD = \sum_{k=1}^{n} \left[q(s_2;k) - \hat{q}(r_R(s_2),k) \right]^2 \qquad (2.31)$$

where $\hat{q}(r_R(s_2),k)$ is the catchment's runoff calculated with the rainfall-runoff model and the "representative" infiltration capacity $r_R(s_2)$. Finding the value of $r_R(s_2)$ such that (2.31) is minimised is referred to as solving the "inverse problem", i.e. finding a parameter or input variable from the observed output variable as opposed to the "forward problem" which involves calculating the output from input variables and parameters. There are many algorithms for solving inverse problems. A very general review of the inverse problem in terms of system's theory is given by Åström and Eykhoff (1971). Particularly in groundwater hydrology, inverse theory has been extensively applied. See Yeh (1986), Ginn and Cushman (1990) and McLaughlin and Townley (1996) for reviews. An example of an inverse problem to obtain representative parameters is given in subsection 2.3.5.

Is the information about the input variables and parameters exhaustive at scale s_1?

If the output variable is not available at the scale s_2, we have to derive the representative parameters or input variables from theoretical considerations or experience. For instance, it is known from various applications that for uniform flow, the representative conductivity $k_R(s_2)$ of a two-dimensional domain s_2 is well approximated with the geometric average of the small scale conductivities $k_R(s_1;i)$ within the block (Desbarats, 1991):

$$k_R(s_2) = \exp\left[\frac{1}{N_{12}} \sum_{i=1}^{N_{12}} \ln(k(s_1;i)) \right] \qquad (2.32)$$

In order to apply the upscaling formula (2.32) the hydraulic conductivity for all support units at scale s_1 must be known, i.e. must be exhaustive. Subsection 2.3.1 describes the case of exhaustive information.

Is the unknown temporal or spatial variation of input variables or parameters at scale s_1 to be described deterministically?

If the information about the input variables or parameters at scale s_1 is non exhaustive, i.e. is only known at a limited number of locations or points in time, representative parameters such as given in Equation (2.32) cannot be calculated directly. First, some function must be assumed that describes the (unknown) spatial or temporal variation of the property at scale s_1 within s_2. Such a function can be either deterministic or stochastic. In case of a

deterministic function, a single representation of the unknown variation of the property at scale s_1 within s_2 is constructed. As described in subsection 2.2.4, this usually involves an interpolating function passing through the observed values. This is schematically depicted for the time domain (continuum approach) in *Figure 27*. Such a function could for instance be obtained from Thiessen polygons, inverse distance weighting or spline interpolation (see for instance Burrough, 1986, pages 151-154). *Figure 27* also shows that this approach leads to a single upscaled property. However, we stress that the upscaled value is always "given the interpolating function assumed" and is therefore non-unique. So, apart from the non-uniqueness originating from the closure problem (see before), representative properties derived from non-exhaustive information are non-unique due to the choice of an interpolating function. Deterministic functions and representative parameters are described in subsection 2.3.2.

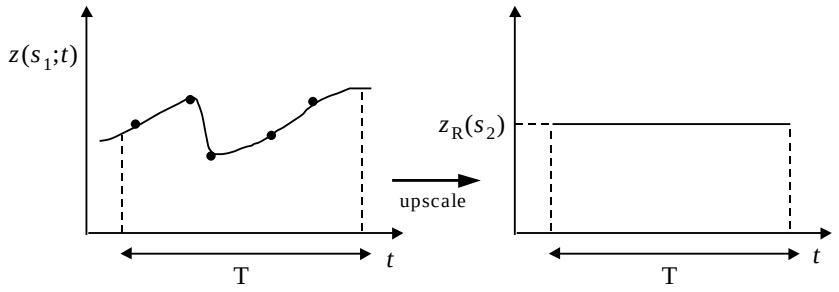

Figure 27. Upscaling if the unknown variation between variations is described by a single deterministic function

To express the uncertainty accruing from this non-exhaustiveness, the stochastic function concept can be used. Examples of using the stochastic function concept in determining representative parameters are given in subsections 2.3.3 and 2.3.4. Readers that are not familiar with the concept of a stochastic function are recommended to read the Appendix, as this will be useful in understanding how stochastic functions can be used in upscaling and downscaling and in appreciating the examples given in subsections 2.3.3, 2.3.4 and 2.4.2 and chapter 3.

Figure 28 (top-left figure) shows schematically the concept of a stochastic function. Instead of finding one function that describes the variation at scale s_1 within s_2 and passes through the observations, a family

of functions is chosen, each of which is assumed to have an equal probability of representing the true but unknown variation at scale s_1 within s_2. Also, each realisation has the similar statistical properties (e.g. mean, variance, semivariogram). The family of equally probable functions is called the stochastic function, while one particular function is called a realisation. A stochastic variable is usually denoted with a capital (e.g. $Z(s_1)$), while one particular outcome is denoted with a lower case ($z(s_1)$) (see Appendix). Similarly, a stochastic function is denoted with a capital (e.g. $Z(s_1; t)$), while one of its realisations is denoted with a lower case ($z(s_1; t)$). If a stochastic approach is chosen, the upscaling is performed for each realisation. This is shown in *Figure 28*. It can be seen that not a single upscaled value results, but a large number of upscaled values (one for each simulated realisation). The result is a probability distribution of the upscaled $Z(s_2)$. This probability distribution expresses the uncertainty in $z(s_2)$ due to the unknown variation of $z(s_1;t)$ between the observation points.

If the upscaled parameters or input variables are to be used in a model, it is better to use the stochastic approach, i.e. run this model for each of the alternative $z(s_2)$. If the model is non-linear (as is the case when using representative parameters) the average of the model outcomes is a probably a better estimate of the true model outcome then the outcome of a single interpolating function, because the latter is likely to underestimate extreme values. Moreover, the result of this procedure, which is called Monte Carlo simulation, is a probability distribution of model outcomes, which can be used for risk assessment. A deterministic method is therefore only advised if one is very sure about the correct form of the unknown spatial variation between measurement locations, such that a single interpolation function is warranted. Of course, the use of a stochastic function does not prevent subjectivity. Just as with a deterministic function, there are many types of stochastic functions that can be chosen, all with realisations that pass through the observations and all having the same statistical properties (see for instance Gómez-Hernández and Wen, 1998). If the choice of the type of stochastic function is critical, the upscaling should not only be repeated for different realisations, but also for different realisations of various types of stochastic functions.

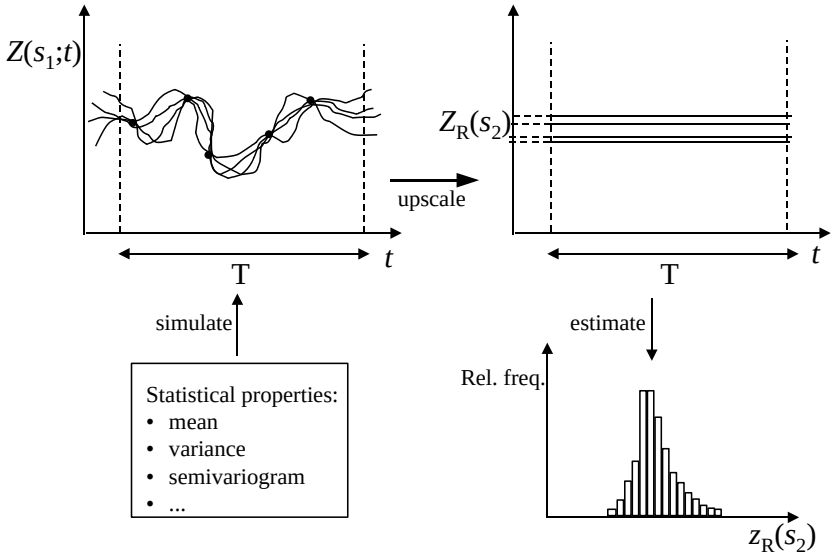

Figure 28. Upscaling if the unknown variation between variations is described by a stochastic function (a family of equally probable functions); the statistics of the upscaled property are derived indirectly.

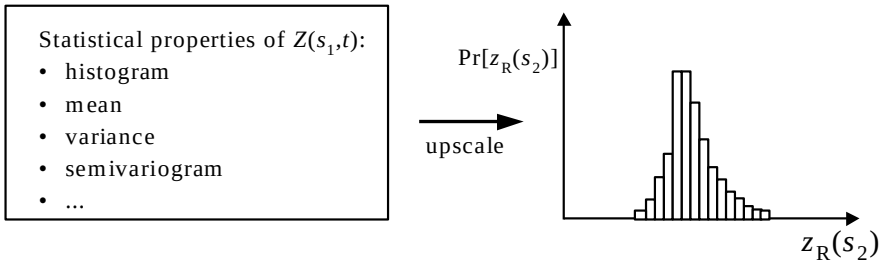

Figure 29. Upscaling if the unknown variation between variations is described by a stochastic function; the statistics of the upscaled property are derived directly.

There are two ways of using stochastic functions in upscaling by representative parameters. The first one is the most straight forward and is already denoted in *Figure 28*. After estimating relevant statistics (e.g. mean, semivariogram, extreme value statistics) from the data available at scale s_1, a

large number of realisations of the stochastic function is simulated (Deutsch and Journel, 1998). Using an upscaling formula such as (2.32) each realisation is upscaled to obtain a realisation of the representative parameter $Z_R(s_2)$. From the simulated realisations of $Z_R(s_2)$ the probability distribution (histogram) of $Z_R(s_2)$ is estimated. Because the upscaling method consists of three steps (simulation, upscaling, estimation) it is referred to as the "indirect stochastic method". *Figure 29* shows schematically the "direct stochastic method". Here, the probability distribution of $Z_R(s_2)$ is directly (and usually analytically) derived from the probability distribution (histogram) and the semivariogram of $z(s_1;t)$. We do not make a choice between direct and indirect stochastic methods here. Both can be used if one decides to calculate representative parameters with stochastic functions. The disadvantage of the direct stochastic method is that it requires rather complicated calculus and a lot of additional assumptions about the stochastic function are needed, such as $Z(s_1;t)$ being second order stationary, all $Z(s_1;t)$ (for all t within s_2) being jointly Gaussian distributed and a small variance of $Z(s_1;t)$. The disadvantage of indirect stochastic methods is that they are computationally demanding. The advantage is that a larger class of stochastic functions (e.g. with non-Gaussian probabilities) can be considered. Also, for indirect methods it is much easier to simulate realisations that are conditional, i.e. that pass through the observed values. Because the conditional realisations are a subset of the unconditional realisations, the resulting $z_R(s_2)$ will be less variable (the variance of the $Z_R(s_2)$ is smaller). The unknown $z_R(s_2)$ can thus be predicted with less uncertainty[5].

Both deterministic methods and indirect stochastic methods may benefit from the use of auxiliary information. Although we do not devote a separate subsection to finding representative parameters using auxiliary information, we emphasise that the methods described in subsections 2.3.2 and 2.3.3 can be easily augmented with any of the methods of using auxiliary information mentioned in subsection 2.2.5. To illustrate this point we will give an example of the use of auxiliary information in subsection 2.3.3.

[5] In fact, the more observations of $z(s_1;t)$ we have, the more constraints are imposed on the subset of realisations that pass through these observations. Consequently, the variation among the simulated $z_R(s_2)$ will be smaller and therefore our uncertainty about the true but unknown $z_R(s_2)$ smaller. Using this property, we can find a sampling design (number and configuration) of $z(s_1,t)$ to assure that the uncertainty about $z_R(s_2)$ remains below a certain level.

2.3.1 Exhaustive information

Criteria

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model cannot be applied at many locations or time steps.*
- *The model has the same form at the two scales involved.*
- *It is not possible to obtain the output variables at the larger scale s_2.*
- *The information about the input variables and parameters exhaustive at scale s_1.*

Method

If the information about the variable z at scale s_1 is exhaustive, i.e. we have full coverage, no assumptions need to be made about the unknown variation of $z(s_1,i)$. The representative value $z(s_2)$ can be found either by application of an explicit upscaling rule (e.g. Equation 2.32) to the $z(s_1,i)$ values or by deriving such an upscaling rule by running the model at scale s_1 for a few test areas of size s_2 for several parts of the model domain (e.g. Gómez-Hernández, 1992).

Example

The example used here is a case of finding representative hydraulic conductivity for numerical model blocks. We realise that it almost never happens that the hydraulic conductivity at scale s_1 is known exhaustively. Examples of such cases are better found in meteorology where spatially variable fluxes (such as evapotranspiration and latent heat) between the earth surface and the atmosphere as obtained by satellite images are to be upscaled to large blocks for climate models (see Dolman and Blyth (1997) for a review of current methods and results). However, the example from upscaling hydraulic conductivity can be used concurrently for subsections 2.3.1, 2.3.2 and 2.3.3. This way, the differences between the various approaches can be better elucidated.

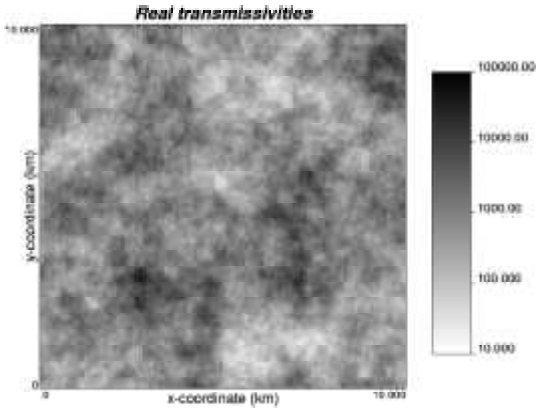

Figure 30. Reference transmissivity field (m^2d^{-1}) at scale s_1

Lets assume that *Figure 30* represents (part of) a flow domain in two dimensions in an aquifer whose thickness $D(x,y)$ varies in space. Also, let the groundwater flow be mainly uniform with a main head gradient from left to right (x-direction). If the hydraulic conductivity is isotropic (has the same value in all directions), the groundwater flow in the x-direction is described with Darcy's law as:

$$q_x = -T\frac{\partial h}{\partial x} \tag{2.33}$$

where h is the hydraulic head $\phi(x,y,z)$ averaged over the aquifer thickness:

$$h(x,y) = \frac{1}{D(x,y)}\int_0^{D(x,y)} \phi(x,y,z)dz \tag{2.34}$$

q_x the depth integral over the aquifer thickness of the horizontal flux $v_x(x,y,z)$ in the x-direction [L^2T^{-1}]:

$$q_x(x,y) = \int_0^{D(x,y)} v_x(x,y,z)dz \tag{2.35}$$

and T the transmissivity [L^2T^{-1}] which is found from the integrating the hydraulic conductivity $k(x,y,z)$ over the aquifer depth as:

$$T(x,y) = \int_0^{D(x,y)} k(x,y,z)dz \qquad (2.36)$$

Figure 30 shows the variation of $T(x,y)$ (or rather $T(s_1;i)$ as T is represented in the discrete domain here) at scale s_1 where the support for T is a block of 100×100 meter (a total of 10000 blocks for the 10×10 km flow domain). Lets assume that a groundwater model is made of this area with s_2 block support of 1000×1000 meter, resulting in a 10×10 discretisation. A reason for modelling with larger blocks may be that solving the groundwater flow problem on a computer is too computationally demanding[6]. The upscaling problem then consists of finding representative transmissivities $T_R(s_2)$ for the larger blocks containing the transmissivities of 100 smaller blocks at scale s_1.

Although the transmissivity at scale s_1 is isotropic, this is not necessarily so for the upscaled transmissivity at s_2. Generally, the upscaled transmissivity is anisotropic which means that is a 2×2 matrix (also called tensor). However, the reference field used in this example has been simulated in such a way that the upscaled transmissivities are approximately isotropic as well (see for instance Bierkens and Van Der Gaast (1998) for a description about the nature of upscaled conductivities). The theoretical way of obtaining the upscaled transmissivities in our isotropic case is building a numerical groundwater model at scale s_1 that contains the model area shown in *Figure 30* and modelling the case for uniform flow. This would yield for each small block s_1 the hydraulic head h and the flux q_x in the flow direction x. Next, for each large scale block at s_2 the average head gradient $<\partial h/\partial x>$ and average flux $<q_x>$ are numerically calculated from all the h and q_x at scale s_1 within the s_2 block. The upscaled transmissivity for each block at scale s_2 is calculated as:

$$T_R(s_2) = -\frac{<\partial h/\partial x>}{<q_x>} \qquad (2.37)$$

However, the reason that the upscaling of transmissivities is necessary is that the groundwater model cannot be run at the small scale s_1. One way of dealing with this problem is building groundwater models at scale s_1 for a limited number of blocks s_2 and determining the upscaled transmissivity for

[6] Of course, in real applications a problem of millions of nodes is reduced to a problem of tenths of thousands of nodes, instead of from 10000 to 100. However, the small sized problem introduced here serves well as an example.

the blocks with Equation (2.37). The upscaled transmissivities for the other blocks are subsequently estimated from the limited number of analysed blocks through spatial interpolation (see Gómez-Hernández, 1992). Alternatively, from theory or experience, explicit upscaling rules may be known that can be applied directly to the (exhaustive) transmissivities at scale s_1. For instance, as already mentioned before, it is known from various applications that for uniform flow, the representative conductivity $k_R(s_2)$ of a two-dimensional domain s_2 (which is large enough compared to s_1) is well approximated with the geometric average of the small scale conductivities $k_R(s_1;i)$ within the block (Desbarats, 1991). This is also true for the transmissivity. Hence, for our case an upscaling rule like Equation (2.32) can be used (100 being the number of support units at scale s_1 within s_2):

$$T_R(s_2) = \exp\left[\frac{1}{100}\sum_{i=1}^{100}\ln(T(s_1,i))\right] \tag{2.38}$$

Equation (2.38) is applied to the example in *Figure 30*. *Figure 31* shows the upscaled transmissivities. In the following examples, these upscaled transmissivities will be treated as the "real" values when compared with various way of deriving upscaled transmissivity from non-exhaustive information. Bear in mind that Equation (2.38) is itself an approximation (although a very good one), and that theoretically the real upscaled transmissivity must be obtained from a groundwater model at scale s_1 and Equation (2.37).

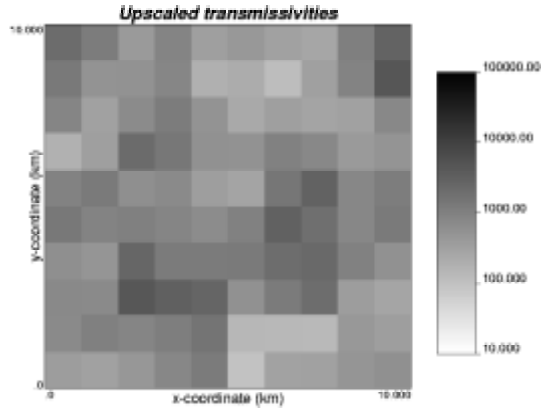

Figure 31. Reference transmissivity field (m^2d^{-1}) at scale s_2

Literature

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

Desbarats, A.J., 1991. Spatial Averaging of transmissivity. In: S Bachu (ed.), Proceedings of the 5th Canadian/American Conference on Hydrogeology: Parameter Identification and Estimation for Aquifer and Reservoir Characterization, 139-154. National Water Well Association, Dublin, Ohio.

Dolman, A.J. and E.A. Blyth, 1997. Patch scale aggregation of heterogeneous land surface cover for mesoscale meteorological model. Journal of Hydrology, 1990, 252-268.

Gómez-Hernández, J.J. 1992. Regularization of hydraulic conductivities: a numerical approach. In: A. Soares (ed.), Proceedings of the fourth geostatistical congress Troia (Portugal). Quantitative geology and geostatistics 5, 215-232. Kluwer Academic Publishers, Dordrecht, Netherlands.

2.3.2 Deterministic functions

Criteria

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model cannot be applied at many locations or time steps.*
- *The model has the same form at the two scales involved.*
- *It is not possible to obtain the output variables at the larger scale s_2.*
- *The information about the input variables and parameters is not exhaustive at scale s_1.*
- *The unknown temporal or spatial variation of input variables or parameters at scale s_1 is to be described deterministically.*

Method

The variable at scale s_1 is not exhaustively known, but only at a limited number of locations or points in time. Using some deterministic function such as a spline interpolation, inverse squared distance weighting or Thiessen polygons, the unknown variation of $z(s_1;i)$ is described resulting in a full coverage. After that, upscaling rules are used to arrive at representative values $z_R(s_2)$.

Example

Starting from the same example as used in subsection 2.3.1, suppose that transmissivity is observed at a limited number of locations, for example from pumping tests or well logs. In this example 40 random locations are chosen where the transmissivity $T(s_1;i)$ is observed. These locations are denoted in *Figure 32*. Using the values at these locations, full coverage is achieved with

an inverse squared distance interpolation between observed locations (see subsection 2.2.4). The resulting image is shown in *Figure 33*.

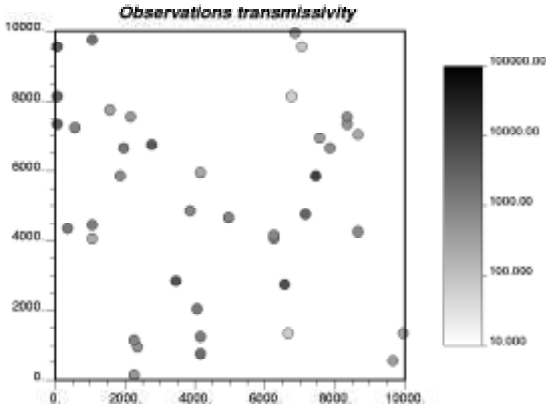

Figure 32. Observations of transmissivity (m^2d^{-1}) at scale s_1

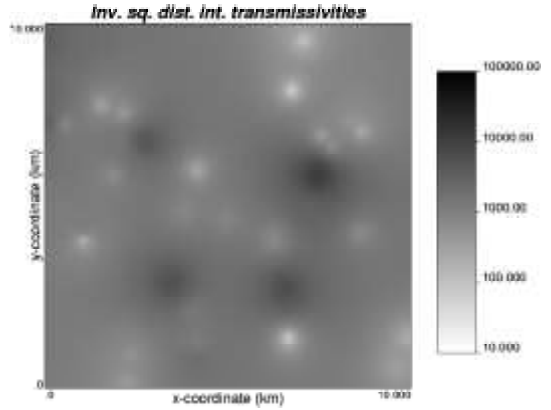

Figure 33. Transmissivities (m^2d^{-1}) at scale s_1 obtained with inverse distance interpolation

Using Equation (2.38) the upscaled transmissivities are calculated for the 1×1 km blocks. In *Figure 34* these upscaled transmissivities are compared with the "real" ones. This figure shows that the results are not particularly good. Obviously, an interpolator like inverse squared distance is too smooth to represent the right variation shown in *Figure 30*. This does not mean that deterministic functions are not suitable for interpolation in combination with finding representative parameters. First of all, there are deterministic functions that are much less smooth or even rugged like deterministic interpolating fractals (Barnsley, 1988; Puente and Obregón, 1996).

Secondly, some variables may be smoothly varying such as groundwater head in large phreatic aquifers. However, smooth interpolating functions are usually not used for upscaling flow parameters such as transmissivity. The reason is that the representative conductivity of a block is very sensitive to the way the higher and lower transmissivities at scale s_1 are spatially connected. Preferential flow paths (the dark zones in *Figure 30*) and hydraulic barriers (the light zones) are not well represented by a smooth interpolator. This can be can be seen from *Figure 34* which shows a systematic underestimation of the larger upscaled transmissivities and a systematic overestimation of the smaller ones. Therefore, in most applications of upscaling of flow parameters stochastic functions are used to describe unknown spatial variation at scale s_1.

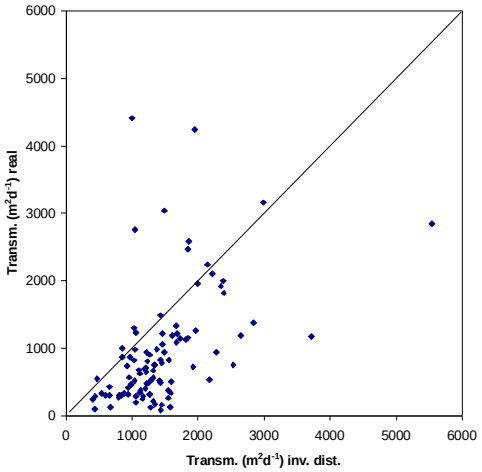

Figure 34. Scatterplot of reference transmissivities at s_2 versus upscaled transmissivities after inverse distance interpolation

Literature

Barnsley, M.F.,1988. Fractals Everywhere. Academic Press, New York, 394 pp.

Puente, C.E. and N. Obregón, 1996. A deterministic geometric representation of temporal rainfall: Results for a storm in Boston. Water Resources Research 32(9), 2825-2839.

2.3.3 Indirect stochastic methods

Criteria

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model cannot be applied at many locations or time steps.*
- *The model has the same form at the two scales involved.*
- *It is not possible to obtain the output variables at the larger scale s_2.*
- *The information about the input variables and parameters is not exhaustive at scale s_1.*
- *The unknown temporal or spatial variation of input variables or parameters at scale s_1 is to be described stochastically.*

Method

The variable at scale s_1 is non-exhaustive, i.e. only known for a limited number of locations or points in time. Instead of using a single deterministic function to describe the unknown spatial or temporal variation $z(s_1;i)$ a stochastic function $Z(s_1;i)$ is used (see the Appendix). From the observations the statistics of the stochastic function are estimated. These are usually the mean value and the semivariogram (using Equation A24). With these statistics, realisations of the stochastic function are simulated using special simulation algorithms (Deutsch and Journel, 1998) and a random number generator. These realisations have the same statistical properties as the set of observations and are conditional: they pass through the observations. Upscaling rules (e.g. Equation (2.32)) are applied to each realisation to arrive at a realisation of representative values $z_R(s_2)$. The result is not a single representative value but as many as simulated realisations. From these, a probability distribution of $z_R(s_2)$ can be estimated, reflecting the uncertainty in the upscaled values due to a limited coverage of the variable $z(s_1;i)$. Moreover, each realisation can be used to calculate a model outcome $y(s_2)=g(z_R(s_2))$, resulting in as many model outcomes as simulated realisations (Monte Carlo simulation). Again, from the alternative model outcomes a probability distribution of $y(s_2)$ can be estimated, reflecting our uncertainty about $y(s_2)$ due to a limited coverage of the input variable or parameter $z(s_1;i)$.

If the domain is covered with multiple support units s_2 (as in the example of 2.3.1) and $Z(s_1;i)$ is a stochastic function, the upscaled representative property is a stochastic function also: $Z_R(s_2; j)$, $j = 1,..,N_{23}$ (N_{23} being the number of support units s_2 in a larger domain s_3). The statistics of this stochastic function, i.e. mean, variance, covariance etc., are determined indirectly: 1) simulate realisations at s_1; 2) upscale realisations to s_2; 3) estimate the statistics of $Z_R(s_2; j)$ from the latter; hence the name "indirect

stochastic methods". Indirect stochastic methods of finding representative parameters have been particularly successful in geohydrology and petroleum engineering where they have been used to find representative conductivities for complex deposits such as sand-shale sequences (Desbarats, 1987) and cross-bedded deposits (Kasap and Lake, 1990; Bierkens and Weerts, 1994).

Example

The same example is used as in subsections 2.3.1 and 2.3.2. From the limited number of observations the statistics of transmissivity are estimated. Instead of using the transmissivity itself, one usually works with its natural logarithm: ln$T$, as experience shows that (random) samples of logconductivity and logtransmissivity are often Gaussian distributed (Freeze, 1975). The statistics estimated from the data are the mean logtransmissivity $\mu_{\ln T}(s_1)$, its variance $\sigma^2_{\ln T}(s_1)$ and the covariance function $C(s_1;\mathbf{h})$. The former two are usually assumed constant and the latter only a function of the separation distance **h** between two locations. Thus, second order stationarity is assumed. If on top of this one assumes that the elements in any set of $\ln T(s_1;\mathbf{x})$ are jointly Gaussian distributed (i.e. are multiGaussian), then these three statistics are sufficient to characterise the complete stochastic function (see also the Appendix)[7].

The "real data" (*Figure 30*) of our example were obtained as follows: First, a simulation algorithm called "sequential Gaussian simulation" (Deutsch and Journel, 1998) was first used to generate an unconditional realisation of second order stationary and multiGaussian stochastic function with $\mu_{\ln T}(s_1)$=6.908, $\sigma^2_{\ln T}(s_1)$=1.326 and the following (exponential) covariance function:

$$C(s_2;\mathbf{h}) = \sigma^2_{\ln T}(s_1)\, e^{-\left(\frac{|\mathbf{h}|}{I(s_1)}\right)} \qquad (2.39)$$

The separation distance $|\mathbf{h}|$ between two locations (x_1, y_1) and (x_2, y_2) is given by $|\mathbf{h}| = \sqrt{[(x_2-x_1)^2+(y_2-y_1)^2]}$ and $I(s_1)$ is called the "integral scale", a parameter that determines the degree of spatial dependence of the stochastic function $Z(s_1;i)$ or $Z(s_1;x)$. The larger $I(s_1)$, the larger the distance over which two random variables are still spatially dependent (Appendix). The unconditional realisation in *Figure 30* was simulated with $I(s_1)$ = 750 m. Next, a realisation of transmissivity $T(s_1;i)$ was obtained from this realisation by taking exponentials.

[7] Usually the semivariogram is estimated from the observations using (A24) and the covariance function is obtained with (A25).

As explained above, in practice the mean, variance and covariance function are estimated from the observations shown in *Figure 32*. However, in this example we know these statistics. Using these statistics realisations of $\ln T(s_2;i)$ are simulated that pass through (are conditional to) the observations. On of these conditional realisations is shown in *Figure 35*. Obviously, the structure of this realisation looks the same as the "real data" in *Figure 30*. Not only can preferential flow paths be seen in *Figure 35*, but also the locations with high and low transmissivity values seem to coincide as well. The latter is due to the observations, which are common to both images. From the realisation in *Figure 35* representative transmissivities were estimated for the 100 1×1 km blocks using Equation (2.38). *Figure 36* shows the scatterplot where these representative transmissivities are compared with the real ones. Although there is still quite a lot of scatter, which is due to uncertainty about the real $z(s_1;i)$, there is no systematic error. This is not surprising, because the interpolating function (the realisation of a stochastic function) used to describe the unknown spatial variation of $z(s_1;i)$ has the same statistical properties as a our constructed "reality".

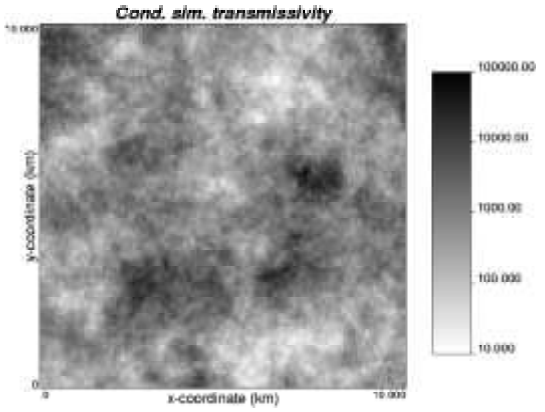

Figure 35. Realisation of transmissivity at s_1 (m^2d^{-1}) obtained with conditional simulation

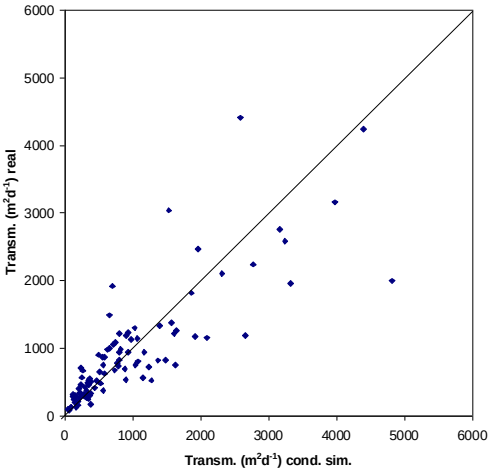

Figure 36. Scatterplot of reference transmissivities at s_2 versus upscaled transmissivities after conditional simulation

Indirect stochastic methods can easily be extended to include auxiliary information. For instance, *Figure 37* shows a map of auxiliary information about the transmissivity. Such exhaustive auxiliary information could for instance be obtained from geo-electrical measurements or seismic exploration. This information can be taken into account when generating conditional realisations of a stochastic function. One way of doing this is using a technique called "external drift kriging" in the sequential Gaussian simulation algorithm (for details see Deutsch and Journel, 1998). *Figure 38* shows a realisation of $\ln T(s_1;i)$ so obtained. The scatterplot with the resulting representative transmissivities is shown in *Figure 39*. It can be seen that the use of the auxiliary information has decreased the scatter and therefore the uncertainty. However, some overestimation of the larger transmissivities is also apparent.

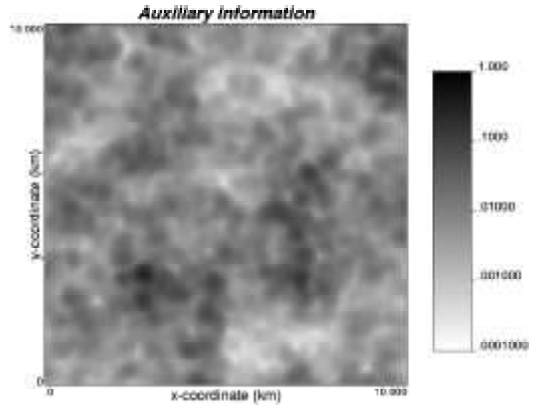

Figure 37. Exhaustive auxiliary information about transmissivity

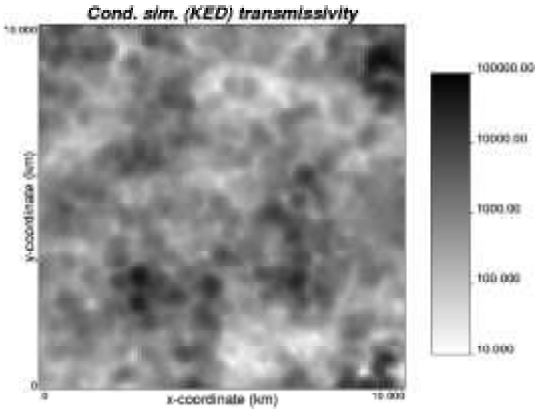

Figure 38. Realisation of transmissivity at s_1 (m^2d^{-1}) obtained with conditional simulation using exhaustive auxiliary information

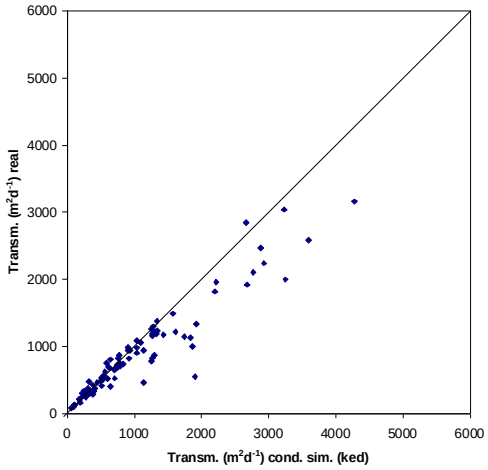

Figure 39. Scatterplot of reference transmissivities at s_2 versus upscaled transmissivities after conditional simulation using auxiliary information

Literature

Bierkens, M.F.P. and H.J.T. Weerts, 1994. Block hydraulic conductivity of cross-bedded fluvial sediments. Water Resources Research 30(10), 2665-2678.

Desbarats, A.J., 1987. Numerical estimation of effective permeability in san-shale sequences. Water Resources Research 23(2), 273-286.

Deutsch, C.V. and A.G. Journel, 1998. GSLIB, Geostatistical Software Library and User's Guide (2nd edition). Oxford University Press, New York, 369 pp.

Freeze, R.A.,1975. A stochastic-conceptual analysis of one-dimensional groundwater flow in non-uniform homogeneous media. Water Resources Research 11(5), 725-741.

Kasap, E. and L.W., 1990. Calculating the effective permeability tensor of a gridblock. SPE Formation Evaluation, June 1990, 192-200.

2.3.4 Direct stochastic methods

Criteria

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model cannot be applied at many locations or time steps.*
- *The model has the same form at the two scales involved.*
- *It is not possible to obtain the output variables at the larger scale s_2.*

– *The information about the input variables and parameters is not exhaustive at scale s_1.*
– *The unknown temporal or spatial variation of input variables or parameters at scale s_1 is to be described stochastically.*

Method

To describe the unknown spatial or temporal variation of the (non-exhaustive) input variables or parameters stochastic functions are used. As described in subsection 2.3.3, if $Z(s_1;i)$, i=1,..,N_{12} is a stochastic function and we have a modelling domain covered with multiple (say N_{23}) support units s_2, then the variation of the representative values $Z_R(s_2;j)$, j=1,.., N_{23} is also a stochastic function. In the direct stochastic method, the statistical properties (e.g. mean, covariance function) of $Z(s_1;i)$ are estimated from the observations. Next, the statistical properties of $Z_R(s_2;j)$ are derived directly from those of $Z(s_1;i)$ and the upscaling rule. With the statistical properties of the upscaled stochastic function we can directly simulate realisations of $Z_R(s_2;j)$.

Example

Much progress has been made on direct stochastic upscaling methods in geohydrology. Direct methods for obtaining the statistics of representative conductivity or transmissivity have been provided by Rubin and Goméz-Hernadez (1990), Desbarats (1991) and Indelman and Dagan (1993a,b). A thorough comparison of these methods is given by Sánchez-Villa et al. (1995). Practical applications of these methods are still rare. Two examples are the papers by Desbarats and Bachu (1994) and Bierkens (1996).

Here we will discuss the method of Rubin and Goméz-Hernandez (1990) applied to the same example as used in subsections 2.3.1 to 2.3.3. In this example we will use the continuum approach ($T(s_1;\mathbf{x})$ and $\mathbf{T}_R(s_2;\mathbf{x})$, where $\mathbf{x}$ = (x,y)). We start with the upscaling rule in two dimensions: the representative transmissivity for a block of area $|A|$ with central co-ordinates $\mathbf{x}' = (x', y')$ in an isotropic medium is that transmissivity that couples the block average flux to the block average head gradient (Darcy's law at scale s_2):

$$\frac{1}{|A|}\int_A \mathbf{q}(\mathbf{x})d\mathbf{x} = -\mathbf{T}_R(A;\mathbf{x}')\frac{1}{|A|}\int_A \nabla h(\mathbf{x})d\mathbf{x} \qquad (2.40)$$

Note that, for convenience of notation, we have omitted the s_1 scale variable from $\mathbf{q}(\mathbf{x})$ and $\nabla h(\mathbf{x})$ and replaced the general s_2 scale variable with the variable A which is the size of the model block. The $\mathbf{q}(\mathbf{x})$ and $\nabla h(\mathbf{x})$ are

vectors giving the flux and the gradient in the *x*- and *y*-directions (see the general description in section 2.3). Notice also that the upscaled transmissivity is written here as a matrix $\mathbf{T}_R(A;\mathbf{x})$, which means that it is not necessarily isotropic. The final upscaling rule follows from insertion of Darcy's law for $\mathbf{q}(\mathbf{x})$ (assuming isotropic $T(\mathbf{x})$ at scale s_1; scale variable omitted here: $T(s_1;\mathbf{x}) \rightarrow T(\mathbf{x})$):

$$\frac{1}{|A|}\int_A T(\mathbf{x})\nabla h(\mathbf{x})d\mathbf{x} = \mathbf{T}_R(A;\mathbf{x}')\frac{1}{|A|}\int_A \nabla h(\mathbf{x})d\mathbf{x} \tag{2.41}$$

The statistics of the block transmissivity $\mathbf{T}_R(A;\mathbf{x})$ are derived from those of $T(\mathbf{x})$ using the following steps (see Appendix for explanation of properties of stochastic functions):

1. Transmissivity at scale s_1 is described as a stochastic function. This means that, through Darcy's law, both the flux and the hydraulic head are stochastic functions too. We stress this by writing them as capitals, e.g. $\nabla h(\mathbf{x}) \rightarrow \nabla H(\mathbf{x})$.
2. The natural logarithm of transmissivity, $\ln T(\mathbf{x})$, is modelled as a stationary multivariate Gaussian random function. This means that it is fully characterized by its (constant) mean value $E[\ln T(\mathbf{x})] = \mu_{\ln T}$ and the covariance function: $C_{\ln T}(\mathbf{h})$.
3. The transmissivity and the hydraulic head are written as the sum of the mean value and a so called perturbation around this mean value:

$$\ln T(\mathbf{x}) = \mu_{\ln T} + \ln T'(\mathbf{x}) \tag{2.42}$$

$$H(\mathbf{x}) = E[H(\mathbf{x})] + H'(\mathbf{x}) \tag{2.43}$$

4. From (2.42) we can also express $T(\mathrm{x})$ itself in terms of perturbations as follows:

$$T(\mathbf{x}) = \exp(\mu_{\ln T} + \ln T'(\mathbf{x})) = T_G\left[1 + \ln T'(\mathbf{x}) + \frac{(\ln T'(\mathbf{x}))^2}{2} + \ldots\right] \tag{2.44}$$

where the right hand term between brackets is a Taylor expansion of the exponential of $\ln T'$ and T_G is the geometric mean transmissivity defined as $T_G = \exp(\mu_{\ln T})$.

5. Expressions (2.43) and (2.44) are substituted into the upscaling rule (2.41). From this a relation between the representative transmissivity $\mathbf{T}_R(A;\mathbf{x})$ and the smaller scale transmissivity $T(\mathbf{x})$ is derived. The expected value $E[\mathbf{T}_R(A;\mathbf{x})]$ and the covariance function $C_{\mathbf{T}_R}(A;\mathbf{h})$ can be derived analytically from this expression using five additional assumptions: a) the average flow is uniform, which means that the expected value of the hydraulic head gradient $E[\nabla H(\mathbf{x})]$ is constant; b) the block A lies far from the edges of the flow domain (no boundary effects); c) the perturbations $\ln T'$ and H' are small so that higher order terms in (2.44) can be omitted. In practice this is the case if $\sigma^2_{\ln T} = C_{\ln T}(0) < 1$, i.e. the variance of the logtransmissivity is small; d) an isotropic exponential covariance of $\ln T(\mathbf{x})$ is used: $C_{\ln T}(h) = \sigma^2_{\ln T} \exp(-h/I)$, where the parameter I is called the integral scale; e) square blocks are used.

Due to assumptions 5d) and 5e) the upscaled transmissivity is also isotropic, i.e. $\mathbf{T}_R(A;\mathbf{x}) \rightarrow T_R(A;\mathbf{x})$.The resulting expressions for the (constant) mean and the covariance function of $T_R(A;\mathbf{x})$ are as follows:

$$E[T_{\mathrm{R}}(A;\mathbf{x})] = T_{\mathrm{G}}\left\{1 + \mathrm{g}\left(\frac{\sqrt{A}}{I}\right)\sigma^2_{\ln T}\right\} \tag{2.45}$$

$$C_{T_R}(\mathbf{h}) = \frac{T_{\mathrm{G}}^2}{|A|^2}\int_A\int_{A_{+\mathbf{h}}} C_{\ln T}(\mathbf{x}-\mathbf{y})\,d\mathbf{x}d\mathbf{y} \tag{2.46}$$

where $A_{+\mathbf{h}}$ represents the block size A translated by $\mathbf{h}$. Equations (2.45) and (2.46) show that the statistics of the stochastic function $T_R(A;\mathbf{x})$ are calculated from the statistics of $\ln T(\mathbf{x})$, i.e. from the mean through $T_G=\exp(\mu_{\ln T})$ and from the covariance $C_{\ln T}(h) = \sigma^2_{\ln T}\exp(h/I)$. These equations also show that the statistics of $T_R(A;\mathbf{x})$ depend on the size of the model blocks A, in particular with respect to the integral scale I, i.e. the degree of spatial correlation in $\ln T(\mathbf{x})$.

The function g(·) in Equation (2.45) is given in Rubin and Gómez-Hernández (1990). If the block is very small compared to the integral scale, i.e. $(\sqrt{A}/I)\rightarrow 0$), then we have $\mathrm{g}(\cdot) \rightarrow 0.5$. According to (2.45), this means that the expected value of the representative conductivity becomes equal to the arithmetic mean of the smaller scale conductivity: $E[T_R(A;\mathbf{x})] \rightarrow T_G[1+0.5\sigma^2_{\ln T}] \approx E[T(\mathbf{x})]$. On the other hand, if the block size becomes much larger than the integral scale, $\mathrm{g}(\cdot) \rightarrow 0$, so that the representative transmissivity becomes equal to the geometric mean: $E[T_R(A;\mathbf{x})] \rightarrow T_G$. In the latter case we also have from (2.46) that

$C_{TR}(0) \rightarrow 0$, so the variance of the representative conductivity goes to zero. This means that for very large blocks, all block conductivities become deterministic constants, equal to the geometric mean.

Equations (2.45) and (2.46) were applied to the example described in subsections 2.3.1 to 2.3.4. For the stochastic function of the smaller scale transmissivity we have the following statistics: $E[\ln T] = 6.908$, $T_G = \exp(6.908) = 1000$ (m^2d^{-1}), $\sigma^2_{\ln T} = 1.326$ and $I = 750$ m. For block sides of 1000 m ($\sqrt{A}/I$) = 1.333. According to Rubin and Gómez-Hernández (1990) g(1.333) = 0.3. With these values we find for the expected value of the representative transmissivity $E[T_R(A;\mathbf{x})]$=1398 m^2d^{-1} and for the variance $C_{TR}(0)$=660348 m^4d^{-2}. The correlation function is shown in *Figure 40*. We also estimated these statistics from a realisation of representative transmissivities as obtained from the indirect method (see subsection 2.3.3). We simulated a very large realisation[8] of transmissivity values at scale s_1 (1000×1000 blocks of 100×100 m each) with above statistics. From this realisation we obtained a realisation of representative transmissivity values (100×100 blocks of 1000×1000 m) by applying Equation (2.38). *Figure 41* shows this realisation. The mean and variance were estimated from this realisation, as well as the correlation function which is shown in *Figure 40* along with the one according to Equation (2.46). The estimated mean value is equal to 1454 m^2d^{-1}, and the variance equal to 2077170 m^4d^{-2}. Both the mean and the correlation structure are estimated very well with Equations (2.45) and (2.46). However, the variance is severely underestimated. There are two possible causes for this. First, the geometric average applied in Equation (2.38) is an approximation to the true block transmissivity values. However, this approximation is usually very good for uniform flow conditions (Desbarats, 1991). It is therefore more likely that the discrepancy occurs because Equation (2.46) is a first order approximation for the true covariance, only valid if $\sigma^2_{\ln T} < 1$, while in our example we had $\sigma^2_{\ln T} = 1.326$.

[8] Instead of a single large sized realisation we could also have used a large number of smaller sized realisations (e.g. as used in section 2.3.1 to 2.3.3). For a stationary stochastic function $\ln T(\mathbf{x})$ this yield the same results because of the principle of ergodicity (see the Appendix).

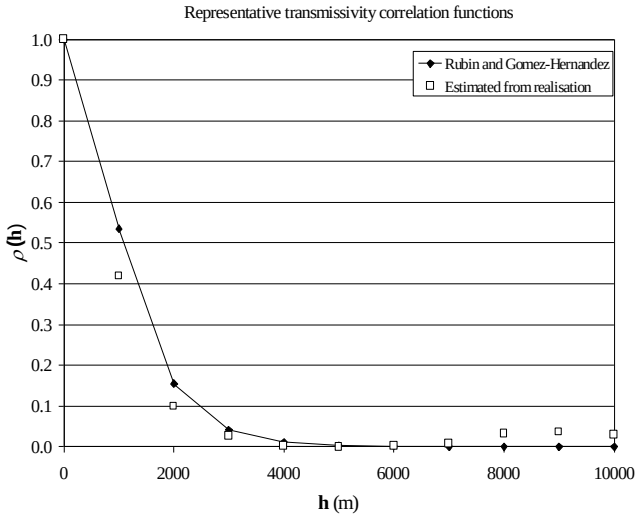

Figure 40. Correlation functions of representative transmissivities as obtained from Eq. (2.46) and estimated from simulations

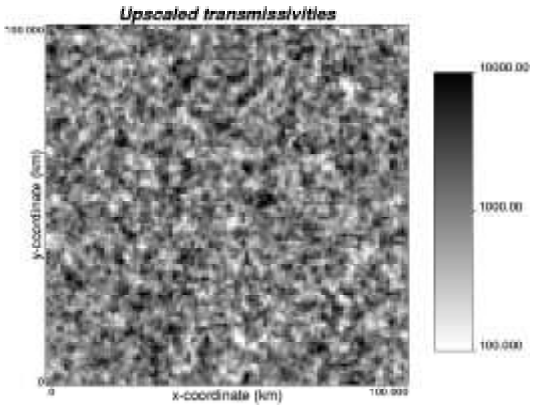

Figure 41. A large realisation of transmissivity (m^2d^{-1}) at scale s_2

Rubin and Gómez-Hernández (1990) also suggest a cokriging based approach to obtain estimates of the mean and variance of $T_R(A;\mathbf{x})$ conditional to observations of hydraulic head and transmissivity at the smaller scale s_1. This makes it also possible to generate conditional realisations of $T_R(A;\mathbf{x})$ directly.

Literature

Bierkens, M.F.P., 1996. Modeling hydraulic conductivity of a complex confining layer at various spatial scales. Water Resources Research 32(8), 2369-2382.

Desbarats, A.J., 1991. Spatial Averaging of transmissivity. In: S Bachu (ed.), Proceedings of the 5th Canadian/American Conference on Hydrogeology: Parameter Identification and Estimation for Aquifer and Reservoir Characterization, 139-154. National Water Well Association, Dublin, Ohio.

Desbarats, A.J. and S. Bachu, 1994.Geostatistical modelling of aquifer heterogeneity from the core scale to the basin scale: A case study. Water Resources Research 30(3), 673-684.

Indelman, I. and G. Dagan, 1993a. Upscaling of permeability of anisotropic heterogeneous formations, 1: The general framework. Water Resources Research 29(4), 917-923.

Indelman, I. and G. Dagan, 1993b. Upscaling of permeability of anisotropic heterogeneous formations, 2: General structure and small perturbation analysis. Water Resources Research 29(4), 925-933.

Rubin, Y. and J.J. Goméz-Hernadez, 1990.A stochastic approach to the problem of upscaling of conductivity in disordered media: Theory and unconditional numerical simulations. *Water Resources Research* 26(4), 691-701.

Sánchez-Villa, X., J.P. Girardi and J. Carrera, 1995. A synthesis of approaches to upscaling of hydraulic conductivities. Water Resources Research 31(4), 867-882.

2.3.5 Inverse modelling

Criteria

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model cannot be applied at many locations or time steps.*
- *It is possible to obtain the output variables at the larger scale s_2.*

Method

Strictly speaking, inverse modelling is not an upscaling problem. However, it is a form of finding representative parameters, so we decided to place it in this upscaling category. If the output variable at scale s_2 is known, one can - so to speak - avoid upscaling and obtain the representative parameter(s) for scale s_2 directly through model calibration or inverse modelling. A "representative" model parameter is sought such that the output of the model resembles the known values of the output variable at scale s_2 as good as possible. In this framework the model at s_1 is of no consequence, although such a lower scale could always be imagined. Therefore the criterion "*The model has the same form at the two scales involved*" is not mentioned here.

To illustrate the inverse method we present an example taken from the papers Kool *et al.* (1985) and Parker *et al*, (1985). In this example the total water outflow from a sediment core is the measured output variable at scale

s_2 and the representative parameters those that describe soil physical relationships for the entire core.

Example

Knowledge of the hydraulic properties of a soil is a prerequisite for prediction of flow of water and transport of solutes through the soil. The relevant properties are the water retention function, giving the relation between the volumetric water content θ and the pressure head h, and the conductivity function giving the relation between hydraulic conductivity k and water content or pressure head. Both are highly non-linear functions. The conventional methods in use up to the eighties were time-consuming and, especially in case of the determination of the $k(\theta)$ relationship quite tedious. A way to get around this is to use results of transient flow events, together with an assumption as to the mathematical form of the retention and conductivity curve. Kool *et al.* (1985) choose the model of van Genuchten (1980) to describe the k-θ-h relations. A normalised water content is defined as:

$$\Theta = \frac{\theta - \theta_r}{\theta_s - \theta_r} \tag{2.47}$$

where θ is the volumetric water content, θ_s the saturated water content, and θ_r the residual water content. The retention curve is modelled by (Van Genuchten, 1980):

$$\Theta = \frac{1}{\left(1 + \left|\alpha h\right|^n\right)^{1-1/n}} \tag{2.48}$$

where h is the pressure head, and α and n are parameters. The $k(\theta)$ relation is modelled as(Van Genuchten, 1980):

$$k = k_s \Theta^{1/2}\left[1 - \left(1 - \Theta^{n/(n-1)}\right)^{1-1/n}\right]^2 \tag{2.49}$$

where k_s is the saturated hydraulic conductivity. Once the form of the functions is chosen the inverse problem is in fact a non-linear optimisation problem.

The first paper (Kool et al., 1985) deals with a theoretical exercise. With a model for unsaturated flow an outflow experiment was simulated, where the cumulative flow out of a soil core under a pneumatic pressure of 1 and

10 m respectively was calculated for two different soils. It was assumed that of the five parameters of (2.47-2.49), only three (α, n and θ_r) were unknown, as the saturated conductivity k_s and saturated water content θ_s are relatively easy to measure. The problem is to find a set of parameter values which minimises the objective function, in this case the squared deviations between "measured" outflow and simulated outflow.

The second paper (Parker et al., 1985) present results from real experiments. The cumulative outflow of soil cores of four soils of different texture under a pneumatic pressure of 10 m was measured. The equilibrium water retention curve was determined by desorption. The parameters (again only α, n and θ_r) were estimated by three methods, each with its own objective function. Method I uses as in the first paper the squared deviations between measured outflow and simulated outflow, In method II the data are supplemented with the weighted squared deviation between measured and predicted water content at pressure head = -150 m, and Method III – the conventional method - employs equilibrium $\theta(h)$ values only. *Figure 42* shows some results. As is demonstrated Method I and II yield satisfactory results at least up to a pressure head of –10 m. Besides, the conventional Method III, using deviations between measured and predicted $\theta(h)$, forces possible inaccuracies into the predicted $k(h)$ relationship, whereas methods employing transient flow data distributes errors to both $\theta(h)$ and $K(h)$ in such manner that the ability of the model to describe the entire transient flow process will be optimised.

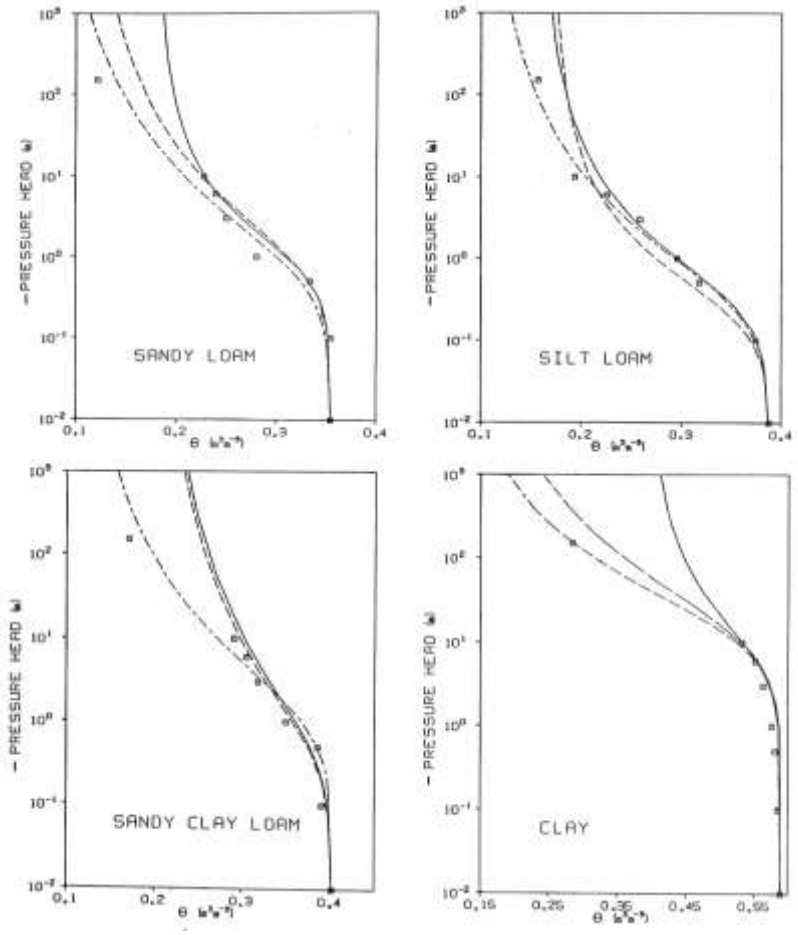

Figure 42. Measured retention curve by stepwise desorption (data points) and predicted retention curve from parameter values estimated by method I (solid line), method II (dashed line) and method III (dash-dot line). Reproduced from Parker et al., 1985.

The one-step outflow method has been extended to the multistep outflow method, where the outside air pressure is increased by several smaller steps during the experiment (Van Dam *et al.*, 1994), which causes less trouble as to the non-uniqueness of the solution.

Literature

Kool , J.B., J.C. Parker, and M. Th. Van Genuchten, 1985. Determining Soil Hydraulic Properties from One-step Outflow Experiments by Parameter Estimation: I. Theory and Numerical Studies. Soil Science Society of America Journal 49: 1348-1354.

Parker, J.C., J.B. Kool , and M. Th. Van Genuchten, 1985. Determining Soil Hydraulic Properties from One-step Outflow Experiments by Parameter Estimation: II Experimental Studies. Soil Science Society of America Journal 49: 1354-1359.

Van Dam J.C., J.N.M. Stricker and P. Droogers, 1994. Inverse Method to Determine Soil Hydraulic Functions from Multistep Outflow Experiments. Soil Science Society of America Journal 58: 647-652.

Van Genuchten, M.Th., 1980. A closed form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal 44: 892-898.

2.4 Averaging of model equations

Criteria

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model cannot be applied at many locations or time steps.*
- *The models at scale s_1 and s_2 have different forms.*
- *The model for scale s_2 can be analytically derived from the model at scale s_1.*

Figure 16 shows schematically the upscaling procedure in this case. The actual upscaling consists of upscaling the model itself, instead of the finding representative parameters or input variables. The upscaling is achieved analytically, through spatial or temporal averaging of the equations making up the model at scale s_1. Equations (2.8) and (2.9) define this upscaling method mathematically. As these equations and *Figure 16* indicate, the upscaled model can be such that it produces output at scale s_2 from either the parameters and input variables at the larger scale s_2 or the set of parameters and input variables at the smaller scale s_1.

Probably the most well known example of this approach is the derivation of Darcy's Law from the Navier-Stokes equations (e.g. Whitaker, 1986). A simpler example is the upscaling of the differential equation describing steady state groundwater flow. As in section 2.3, let **x** be the spatial co-ordinate, **q** the (vector) groundwater flux [LT^{-1}] (volume of water crossing a unit area per unit time), h the hydraulic head [L], ∇h the hydraulic head gradient vector [-], e.g. in three dimensions: $\nabla h = \left(\partial h/\partial x, \partial h/\partial y, \partial h/\partial z\right)^{\mathrm{T}}$, and k [$LT^{-1}$] a tensor containing the hydraulic conductivities in the different directions. The notation $\nabla \cdot \mathbf{q}$ denotes the inner product of the gradient vector and **q** and stands for

$$\nabla.\mathbf{q} = \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z} \tag{2.50}$$

Based on these definitions the steady state groundwater flow equation is written as (for notational reasons the scale variable s_1 is omitted from $h(s_1;\mathbf{x})$, $\mathbf{k}(s_1;\mathbf{x})$ and $p(s_1;\mathbf{x})$):

$$\nabla.[\mathbf{k}(\mathbf{x})\nabla h(\mathbf{x})] = p(\mathbf{x}) \tag{2.51}$$

where $p(\mathbf{x})$ denotes all the inputs (precipitation, well abstractions etc.) at location $\mathbf{x}$. Equation (2.51) is valid at the smaller spatial scale s_1 (continuum approach). We seek the upscaled equation s_2, for instance valid for a volume V. As in section 2.3 hydraulic conductivity and hydraulic head are written as the spatial averages over V (see Equation 2.27) and a deviation from this average: $\mathbf{k}(\mathbf{x}) = <\mathbf{k}(\mathbf{x})> + \mathbf{k}'(\mathbf{x})$, $<\nabla h(\mathbf{x}) = <\nabla h(\mathbf{x})> + \nabla h'(\mathbf{x})$. Inserting these in (2.51), expanding terms and taking the spatial average of both sides of the equation leads to the following form of the upscaled equation:

$$\nabla.[<\mathbf{k}(\mathbf{x})> \nabla <h(\mathbf{x})>] + \nabla.[<\mathbf{k}'(\mathbf{x})\nabla h'(\mathbf{x})>] = <p(\mathbf{x})> \tag{2.52}$$

The solution to this equation yields the block averaged (upscaled) hydraulic head $<\nabla h(\mathbf{x})>$. However, to solve this equation we have to find a form for the block averaged covariation of $\mathbf{k}'(\mathbf{x})$ and $\nabla h'(\mathbf{x})$. In other words, we have to solve a closure problem. As is the case with the closure term in Equation (2.30), the closure term depends on the boundary conditions around the block V. So again, the solution to the closure problem is non-unique and non-local! To find a closed form and local solution of (2.52) certain boundary conditions around the block V have to be chosen (e.g. uniform or periodic flow) and a simplified porous medium (e.g. completely random spheres) assumed. It should therefore be noted that upscaled equations are never unique and in most cases approximations.

The close resemblance of the closure problems occurring with upscaling by representative parameters and upscaling of the model equations is not coincidental. In fact, both approaches yield similar results. This can be seen by comparing equations (2.30) and (2.52). Indeed, if we adopt the form[9] (2.30) for the representative conductivity and substitute this into (2.52) we obtain a partial differential equation *with the same form* as (2.51), but with the block averaged hydraulic head as its solution and the conductivity replaced with the representative conductivity:

[9] Such an assumed form is coined a "constitutive relationship".

$$\nabla.[\mathbf{k}_R < \nabla h(\mathbf{x}) >] = < p(\mathbf{x}) > \quad (2.53)$$

The class of upscaling methods involving averaging of model equations is subdivided based on the following question only (see *Figure 43* for a trivial decision tree):

Is the temporal or spatial variation of the properties at scale s_1 described deterministically?

To apply methods similar to described above (i.e. to evaluate the spatial or temporal averages), either the variation of the input variables and parameters at scale s_1 within s_2 must be known exhaustively or some function describing unknown variation has to be chosen. If a single deterministic function is chosen, a spatial or temporal averaging procedure as described above is chosen (see the decision tree of *Figure 43*). Subsection 2.4.1 gives an example of such a method. If the unknown variation of parameters and input variables is described with stochastic functions (Appendix), the model becomes a stochastic model. This means that the output variable is stochastic also and can only be described in terms of probability distributions. The upscaling of the model equations in this case does not aim to find a model that describes the output variable at larger support units. Instead, a model is derived that describes the large scale trend of the output variable. This is achieved by so called "ensemble averaging", i.e. finding a differential equation that describes the mean of the random output[10]. Subsection 2.4.2 gives an example of ensemble averaging. Here, we will illustrate the concept a little further.

[10] Strictly speaking, ensemble averaging is not upscaling as defined in section 1.3. However, its aim is to derive large scale equations (although the concept of scale is vague), hence the technique is treated here.

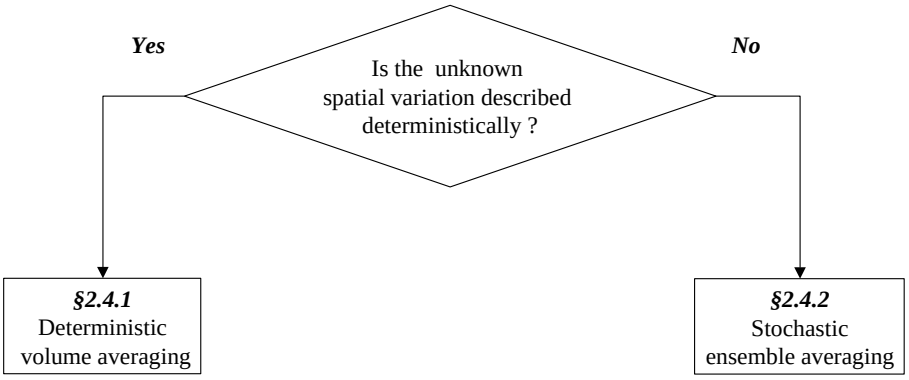

Figure 43. Trivial decision tree to two subclasses of upscaling class 3

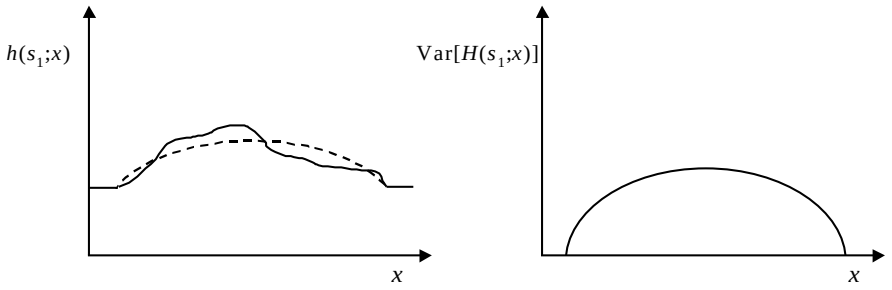

Figure 44. Example of small scale variation of the phreatic surface modelled by a stochastic function; the large (upscaled) variation is modelled by the expected value of the stochastic function (dashed line left figure); the right figure shows the variance as a measure of uncertainty.

The idea is illustrated with the example of steady state groundwater flow and *Figure 44*. The left figure shows a 1D-example of the variation of a water table between two locations where the hydraulic head is constant (e.g. channels with constant surface water levels). We see a large scale trend that is caused by the convexity of the water table due to recharge that causes flow from the groundwater body to the channels. However, we also see a small scale variation of the water table, that may for instance be caused by heterogeneity of the porous medium (hydraulic conductivity) and spatial variation of the groundwater recharge $p(\mathbf{x})$. If we would know the properties of the heterogeneous porous medium and the groundwater recharge at each

location, it would in principle be possible to solve a differential equation like (2.51) and obtain the wiggly course of the water table. However, in practice such information is not available, so that we have to be satisfied with describing the average large scale behaviour of the water table which is given by the smooth line. So the goal is to find a differential equation whose solution yields the smooth line.

The first step is to assume that the wiggly (and unknown) variation of the water table $h(\mathbf{x})$ is a realisation of a stochastic function $H(\mathbf{x})$ (see the Appendix). A good estimate of the large scale trend would then be the line representing the average of all possible realisations of the stochastic function. This average line is therefore the "mean" or "expected value" of the stochastic function $E[H(\mathbf{x})]$.

Next, the cause of the randomness of $H(\mathbf{x})$ is sought in our uncertainty about the parameters (hydraulic conductivity $\mathbf{K}(\mathbf{x})$) and input variables (groundwater recharge $P(\mathbf{x})$), which are also described as stochastic functions. The results is a stochastic differential equation of steady state groundwater flow:

$$\nabla.[\mathbf{K}(\mathbf{x})\nabla H(\mathbf{x})] = P(\mathbf{x}) \tag{2.54}$$

The solution to the stochastic differential equation consists of the probability distribution of hydraulic head for each location $\mathbf{x}$ (scale variable s_1 omitted): $f_H(h,\mathbf{x})$. Instead of the full distribution, which is often difficult to derive, one usually settles for calculating the mean $E[H(\mathbf{x})]$ and the variance $\mathrm{VAR}[H(\mathbf{x})]$ of this distribution. The example in subsection 2.4.2 shows that these statistics depend on the statistics (e.g. mean, variance, semivariogram) of $\mathbf{K}(\mathbf{x})$ and $P(\mathbf{x})$, which must be known or estimated from data. As stated before, the mean function $E[H(\mathbf{x})]$ gives the large scale average behaviour of the unknown $h(\mathbf{x})$. The variance can be seen as a measure of uncertainty about $h(\mathbf{x})$: if $E[H(\mathbf{x})]$ were used to predict the unknown $H(\mathbf{x})$, $VAR[H(\mathbf{x})]$ represents the variance of the prediction error. The right figure in *Figure 44* shows the variance of $H(\mathbf{x})$. If $H(\mathbf{x})$ is Gaussian distributed, the mean and variance are sufficient to estimate the 95% prediction interval at each location with ($E[H(\mathbf{x})]$-1.96·$STD[H(\mathbf{x})]$, $E[H(\mathbf{x})]$-1.96·$STD[H(\mathbf{x})]$), where $STD[\cdot]$ is the standard deviation or the square root of $VAR[H(\mathbf{x})]$.

In conclusion: the upscaling step consists of deriving from (2.54) the model (differential equation) describing $E[H(\mathbf{x})]$, and if needed the model describing $VAR[H(\mathbf{x})]$. Analogous to volume averaging or temporal averaging the parameters and input variables are written as a (random) deviation from the mean function: $\mathbf{K}(\mathbf{x}) = E[\mathbf{K}(\mathbf{x})] + \mathbf{K}'(\mathbf{x})$, $\nabla H(\mathbf{x}) = \nabla E[H(\mathbf{x})] + \nabla H'(\mathbf{x})$ and $P(\mathbf{x}) = E[P(\mathbf{x})] + P'(\mathbf{x})$. Substitution of these

decompositions into (2.54), expanding terms and taking the expected value of both sides of the equation yields the partial differential equation for the mean E[$H(\mathbf{x})$]:

$$\nabla.(E[\mathbf{K}(\mathbf{x})]\nabla E[H(\mathbf{x})])+\nabla.(E[\mathbf{K}'(\mathbf{x})\nabla H'(\mathbf{x})]\,) = E[P(\mathbf{x})] \quad (2.55)$$

Equation (2.55) shows that to solve for the mean the cross-covariance $E[\mathbf{K}'(\mathbf{x})\nabla H'(\mathbf{x})]$ between the conductivity and head gradient is required. To find this covariance means solving a stochastic closure problem. The solution to this problem depends on the boundary conditions around the flow domain and is therefore non-unique. Also, in general its solution is non-local: it depends not only on the statistics of **K** and ∇H at location **x** but also on the statistics of these variables at other locations in the flow domain. Again, to find a closed form solution for the cross-covariance, certain boundary conditions have to be assumed (e.g. uniform flow in an infinite flow domain) and assumptions have to be made about the statistical properties of the random functions $\mathbf{K}(\mathbf{x})$ and $P(\mathbf{x})$ (e.g. (log)Gaussian distributed, stationary, small variance).

As will be shown in subsection 2.4.2, the stochastic approach does not stop with the large scale mean, but is also concerned with obtaining the variance $E[\{H'(\mathbf{x})\}^2]$ or even the covariance $E[H'(\mathbf{x}_1)H'(\mathbf{x}_2)]$. Particularly in the field of stochastic subsurface hydrology, quite a lot of methods have been developed to derive analytical expressions for the covariance $E[H'(\mathbf{x}_1)H'(\mathbf{x}_2)]$, as well as for the cross-covariance $E[\mathbf{K}'(\mathbf{x})\nabla H'(\mathbf{x})]$that is needed to solve the mean equation (e.g. Dagan, 1989; Gelhar, 1993).

2.4.1 Deterministic: temporal or volume averaging

Criteria

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model cannot be applied at many locations or time steps.*
- *The models at scale s_1 and s_2 have different forms.*
- *The model for scale s_2 can be analytically derived from the model at scale s_1.*
- *The temporal or spatial variation of the properties at scale s_1 is described deterministically.*

Method

Consider a model at scale s_1 giving the values of a certain property as a function of, say, the spatial variables **x**:

$$z(s_1;\mathbf{x}) = f(\mathbf{x}) \tag{2.56}$$

At the scale s_2 the average value of the property is then given as:

$$z(s_2;\mathbf{x}) = \frac{1}{s_2}\int_{\mathbf{x}\in s_2} z(s_1;\mathbf{x})d\mathbf{x} = \frac{1}{s_2}\int_{\mathbf{x}\in s_2} f(\mathbf{x})d\mathbf{x} \tag{2.57}$$

where the integral at the right hand side can be evaluated analytically.

Example

During a long period of growth, a crop growing under optimum supply of nutrients takes up nutrients like nitrogen and phosphorus with constant uptake rate (de Willigen and Van Noordwijk, 1987). When the root system of the crop consists of regularly distributed roots it suffices to consider one root with a soil cylinder around it, the radius of which is a function of the root length density. The uptake potential of such a root (scale s_1) – and thus of the root system (scale s_2) - can be characterised by a time constant: the period during which uptake is in accordance with plant demand. After this period the soil cannot maintain a sufficient transport rate to fulfil the crop requirement.

Uptake of some nutrients by plant roots is so efficient that after the period of unconstrained uptake the concentration at the root surface can be assumed to be virtually zero, i.e. the root behaves as a so-called zero-sink. The convection diffusion equation describing transport of a solute, with linear adsorption characteristics, to a root can be solved analytically for a single root confined in a soil cylinder (De Willigen and Van Noordwijk 1994a and b), and so the uptake rate of a zero-sink root can be calculated. The solution, unfortunately, is quite complicated and its use for calculating the uptake rate of a root system is practically impossible. It was found, however, that a very good approximation can be deduced based on the steady-rate approach, from which it can be derived that the uptake rate is proportional to the average concentration in the soil cylinder with the proportionality constant a function of the root length density and the transport properties of the soil (the diffusion coefficient). *Figure 45* shows the time course of uptake of a single root.

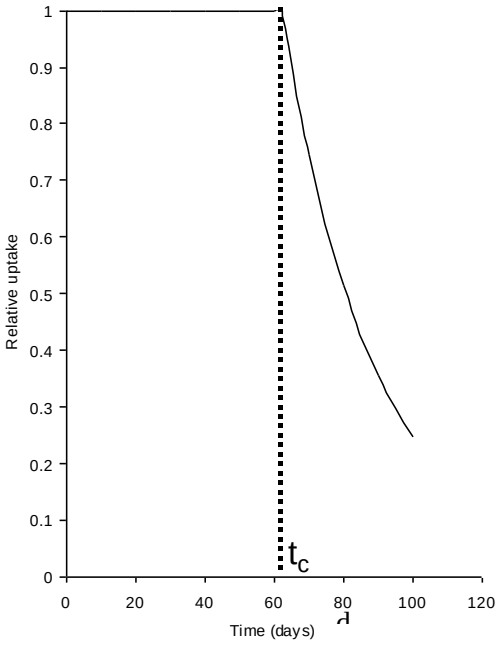

Figure 45. Time course of uptake of a single root with constant demand. Up to point t_c the demand is satisfied, thereafter transport through the soil determines possible uptake rate.

In reality of course the roots will generally not be regularly distributed: the uptake of the entire root system cannot be described by analysing a single root. To estimate the uptake potential of a non-regular distributed root system, De Willigen and van Noordwijk (1987) first constructed Thiessen polygons (see subsection 2.2.4) around each root. Next these polygons were approximated by cylinders with the same surface. In doing so the soil-root system (scale s_2) is described as a set of soil volumes each with its own root length density and thus consisting of roots within a soil cylinder of uniform radius (scale s_1). Now the roots are distributed over a number of classes according to the area allocated to them and the classes are arranged in ascending order of area. The period of unconstrained uptake of the first class ($t_{c,1}$) can be calculated straightforwardly, after this period the roots in this class behave as a zero sink, and the uptake rate of the other classes is increased so as to guarantee the required uptake rate. For the other classes this period has to be calculated recursively (i.e. $t_{c,2}$ follows from $t_{c,1}$, $t_{c,3}$ from $t_{c,1}$ and $t_{c,2}$ etc.). The period of unconstrained uptake of the last class is then the t_c of the root system. *Figure 46* shows t_c as function of buffer capacity

for a root system which is distributed randomly and for a distribution as determined in the field.

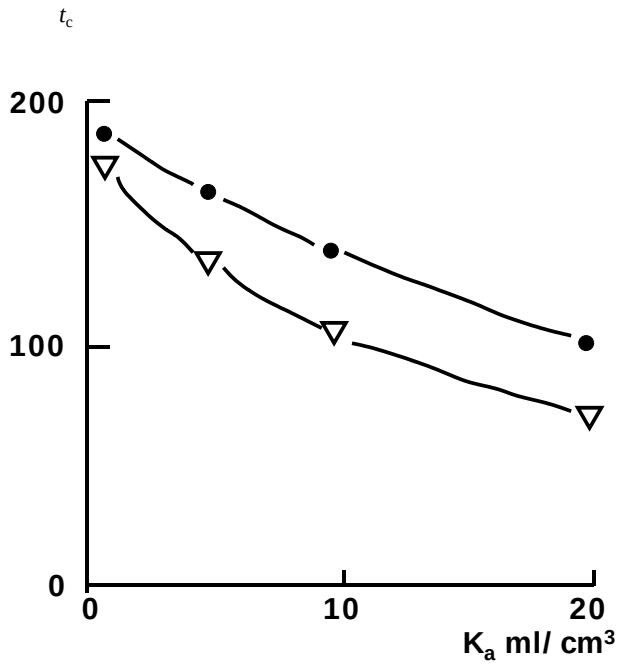

Figure 46. The period of unconstrained uptake as a function of buffer capacity K_a for two types of root distribution, a random distribution and one as observed in the plough layer of a wheat crop as a result of ploughing with a cultivator

Literature

De Willigen, P., 1990. Calculation of uptake of nutrients and water by a root system. Nota 210, Institute for Soil Fertility, Haren, The Netherlands.

De Willigen P. and M.van Noordwijk, 1991. Modelling nutrient uptake, from single roots to complete root systems. In: Penning de Vries, F.W.T., H.H.van Laar, and M.J.Kropff (eds.) Simulation and systems analysis for rice production (SARP). Selected papers from workshop. PUDOC, Wageningen, pp 277-295.

De Willigen, P. and M. van Noordwijk, 1994a. Mass flow and diffusion of nutrients to a root with constant or zero-sink uptake. I. Constant uptake. Soil Science 157:162-170.

De Willigen, P. and M. van Noordwijk, 1994b. Mass flow and diffusion of nutrients to a root with constant or zero-sink uptake. II. Zero-sink uptake. Soil Science 157:171-175.

2.4.2 Stochastic: ensemble averaging

Criteria

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model cannot be applied at many locations or time steps.*
- *The models at scale s_1 and s_2 have different forms.*
- *The model for scale s_2 can be analytically derived from the model at scale s_1.*
- *The temporal or spatial variation of the properties at scale s_1 is described stochastically.*

Method

The input variables and the parameters are described as stochastic variables or stochastic functions (see Appendix). Consequently, the output variable of the model becomes a stochastic function as well: the model becomes a stochastic model. The large scale variation (in space or time) of the output variable is described by deriving from this stochastic model a solution for the expected value of the output variable. Often, also an expression of the variance of the output variable is given, which is a measure of our uncertainty about the true value of the output variable.

Example

The example used to illustrate this class of methods involves the one-dimensional steady-state groundwater flow in a heterogeneous porous medium. It closely follows the analysis given in Gelhar (1993). Take a one-dimensional porous medium (e.g. a tube) where the conductivity k varies in space: $k(x)$, where x is the location. *Figure 47* shows a plot of spatially varying head $h(x)$ for the case that the average flow is from left to right and there is no recharge from above. The spatial variation of the head occurs due to spatial variation in the conductivity. The total flux q [LT^{-1}] is assumed constant and known. In the following we will use the continuum approach and, for convenience of notation, will omit the scale variable.

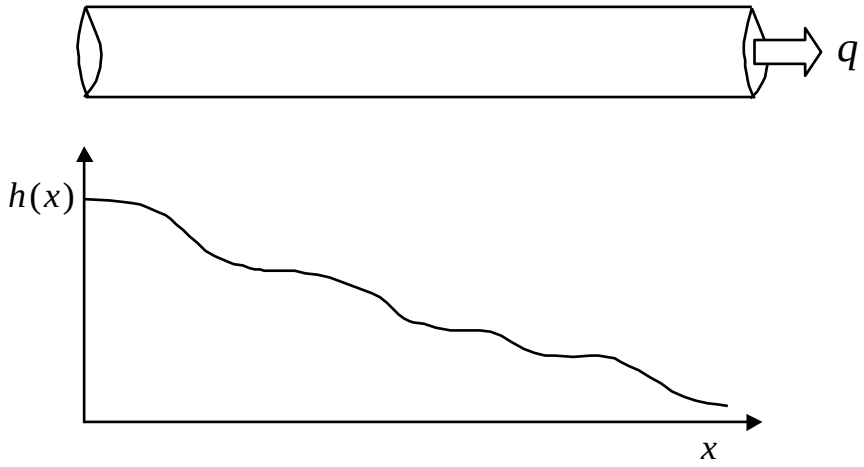

Figure 47. Example of one-dimensional flow through a heterogeneous porous medium; lower figure shows hydraulic head as a function of location.

The unknown variation of $k(x)$ is modelled as a stochastic function. In particular, the logconductivity $\ln k(x)$ is modelled as a second order stationary stochastic function $\ln K(x)$ with mean $\mu_{\ln K}$ and variance $\sigma^2_{\ln K}$. The associated stochastic differential equation is given by:

$$\frac{d}{dx}\left[K(x)\frac{dH}{dx} \right] = 0 \tag{2.58}$$

The first step is to write (2.58) in terms of $\ln K(\mathbf{x})$. This can be achieved by first expanding (2.58) using the chain rule:

$$K(x)\frac{d^2H}{dx^2} + \frac{dK(x)}{dx}\frac{dH}{dx} = 0 \tag{2.59}$$

If we divide both terms of (2.59) by $K(x)$ and using the fact that

$$\frac{d(\ln K(x))}{dx} = \frac{1}{K(x)}\frac{dK}{dx} \tag{2.60}$$

Equation (2.59) is written as:

$$\frac{d^2H}{dx^2} + \frac{d\ln K}{dx}\frac{dH}{dx} = 0 \tag{2.61}$$

The second step is to express both the head and the logconductivity as perturbations around their mean value:

$$\ln K(\mathbf{x}) = \mu_{\ln K} + \ln K'(\mathbf{x}) \tag{2.62}$$

$$H(\mathbf{x}) = E[H(\mathbf{x})] + H'(\mathbf{x}) \tag{2.63}$$

By substitution of (2.62) and (2.63) in (2.61), assuming the perturbations to be small ($\sigma^2_{\ln K} < 1$) so that products of perturbations are omitted, and taking the expectation of the resulting equation we obtain the following equation for the expected value of the head (Gelhar, 1993):

$$\frac{d^2E[H(x)]}{dx^2} = 0 \tag{2.64}$$

Using the small perturbation approximation of Darcy's law we have from (2.64) that

$$\frac{dE[H(x)]}{dx} = -\frac{q}{E[K(x)]} \tag{2.65}$$

If $\ln K(x)$ is Gaussian distributed we can write $E[K(x)]$ in terms of $\mu_{\ln K}$ and $\sigma^2_{\ln K}$ as follows:

$$E[K(x)] = \exp\left(\mu_{\ln K} + \tfrac{1}{2}\sigma^2_{\ln K}\right) \tag{2.66}$$

So the final solution to for the expected value of the head becomes:

$$E[H(x)] = h(x_0) - qx\exp\left\{-\left(\mu_{\ln K} + \tfrac{1}{2}\sigma^2_{\ln K}\right)\right\} \tag{2.67}$$

To obtain a closed form solution to the expected value we have to assume that the hydraulic head at the boundary of the tube $h(x_0)$ is known.

If (2.64) is subtracted from the stochastic differential Equation (2.61) and the products of perturbations are omitted (small perturbation assumption) the

head perturbation follows from the solution to the following differential equation:

$$\frac{d^2H'}{dx^2} = \frac{dE[H(x)]}{dx}\frac{d\ln K'}{dx} \tag{2.68}$$

If $H'(x)$ is solved from (2.68) we can obtain an expression for the variance $VAR[H(\mathbf{x})]$ by taking the expected value. A solution for $E[H'(x)^2]$ can be found more directly from (2.68) by so called spectral representations of $\ln K'(x)$ and $H'(x)$ (see Gelhar, (1993) and De Marsily (1986)). This means that we have to assume that, next to $\ln K'(x)$ also $H'(x)$ has to be a second order stationary stochastic function. Without giving the derivation, the resulting expression for the variance is given by (Gelhar, 1993)[11]:

$$VAR[H(x)] = q^2\sigma^2_{\ln K} I^2 \tag{2.69}$$

where I is the integral scale (see Appendix) of the logconductivity covariance function. *Figure 48* shows the expected value of the head as given by (2.67) for $\mu_{\ln K}$ =0, $\sigma^2_{\ln K}$ =1, q = 1 m/d. and $h(x_0)$ = 100 m for a domain length of 100 m (i.e. the head gradient is –0.605 m/m). Also given are the 95% confidence intervals of $H(x)$ as calculated from (2.69) with integral scale I = 10 m (i.e. var[$H(x)$] = 36.78 m^2). *Figure 48* shows that, apart from being approximations, the solutions also contain an inconsistency. To obtain a closed form solution to the expected head (2.67) we need to know $h(x_0)$. So the actual confidence interval close to x_0 should decrease to zero as indicated by the dashed line in *Figure 48*. Solution (2.69) is therefore only valid for large enough x, where the influence of the boundary condition is diminished.

[11] Assuming a hole effect covariance function for ln$K$: $C_{\ln K}(h)=\sigma^2_{\ln K}[1-(h/I)]\exp(-h/l)$

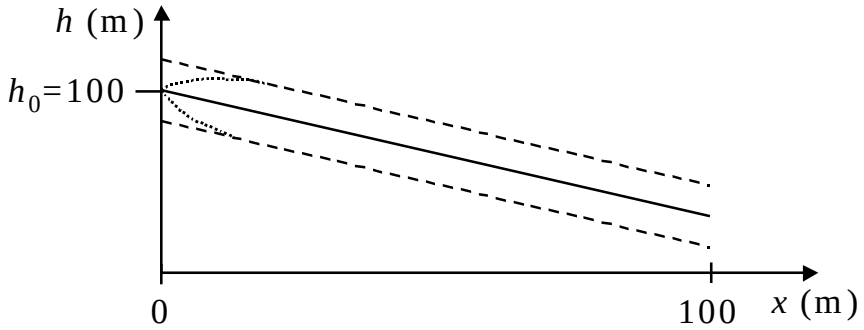

Figure 48. Expected value of hydraulic head (solid line) calculated with Equation (2.67) and the 95%-confidence interval (dashed lines) according to Equation (2.69); the true confidence interval close to x=0 is given by the dotted lines.

Literature

De Marsily, G., 1986. Quantitative Hydrogeology, Groundwater Hydrology for Engineers. Academic Press, New York, 440 pp.

Gelhar, L.W., 1993. Stochastic Subsurface Hydrology. Prentice Hall, New Jersey, 390 pp.

2.5 Model simplification

Criteria

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model cannot be applied at many locations or time steps.*
- *The models at scale s_1 and s_2 have different forms.*
- *The model for scale s_2 cannot be analytically derived from the model at scale s_1.*

As with the methods described in section 2.4 the actual upscaling consists of upscaling the model itself, instead of the finding representative parameters or input variables. As before, in the following discussion $g_1[\cdot]$ represents the model at scale s_1 and $g_2[\cdot]$ the model at scale s_2. If the model at scale s_1 consists of a simple relationship or formula or a single differential equation, the average model $g_2[\cdot]$ can be obtained analytically through spatial or temporal averaging of the model at s_1: $\langle g_1[z(s_1;i)]\rangle$. However, many environmental models consist of a complex conglomerate of sub-models,

each describing only a part of the dynamics at scale s_1. For instance, a model describing the leaching of nutrients in a soil column may consist of a hydrological sub-model describing fluctuations of water flow through the soil column, chemical sub-models for each of the relevant nutrient species and organic matter and a sub-model describing crop growth and transpiration. Even the sub-models may consist of further sub-models such as the hydrological sub-model consisting of a sub-model describing water table fluctuations in relation with surface water levels and a sub-model describing the actual unsaturated flow. Such complex models are impossible to average to a larger scale models (e.g. obtaining a leaching model for a field from that of a soil column) simply by taking temporal or spatial averages of the model equations. So, rather than being derived, the model at s_2 is postulated.

This procedure is schematically shown in *Figure 17*. A representative sample of parameters and input variables is taken from the support units at scale s_1. The model $g_1[\cdot]$ is applied at the sample locations or sample time steps to yield the output variables at scale s_1. Using one of the averaging methods described in section 2.2, the average input variables and the average output variables for scale s_2 are estimated. Next, a simplified model is postulated which directly relates the estimated parameters and input variables at s_2 to the output variables at s_2. With the postulated model $g_2[\cdot]$ and the parameters and input variables at s_2 the output variables at s_2 can also be estimated. These results are subsequently compared with the average model outputs obtained from $g_1[\cdot]$. If these results compare well, the postulated model can be used for all other support units at scale s_2. If the results are too different, the postulated model is either adjusted (calibration), or alternatively, different simplified models are postulated until a better match is obtained.

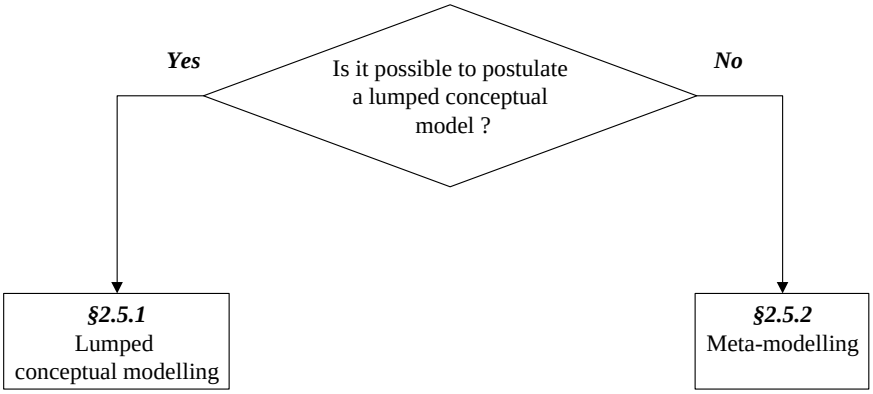

Figure 49. Trivial decision tree to two subclasses of upscaling class 4

This class of upscaling methods is further sub-divided based on the following question which determines in what manner the model $g_2[\cdot]$ is postulated (see trivial decision tree *Figure 49*).

Is it possible to postulate a lumped conceptual model at scale s_2?

The definition of a model at s_2 from a model at s_1 is not a straightforward exercise when it can not be analytically derived. What processes are put forward at s_2 and how these are formalised depends on how a research group treats a number of philosophical and practical considerations.

Philosophical considerations

Knowledge to adequately model at s_2 is often incomplete, since the knowledge on the whole system is incomplete. Nevertheless, it can be enhanced by leaning on theoretical frameworks like hierarchy theory (Weiss, 1971; O'Neill et al., 1986). Hierarchy theory sees ecosystems as open systems that are hierarchically organised. How this organisation develops in complex systems is in itself a research issue, which has resulted in theories like Chaos Theory (e.g. Gleick, 1988), the Gaia Hypothesis (Lovelock, 1979) and Entropy Theory (e.g. Runge, 1973; Addiscott, 1995). Here, we forego the differences between these theories and just assume that at a level in the hierarchy a stable system evolves, which is constrained by a higher level system and is itself constraining a lower level system. To describe such system, the holon-principle could be applied (Haigh, 1987). A holon is defined as a stable "sub-whole" in a hierarchy, and is process-based. An association of holons that are linked through the exchange of matter or

energy is termed a holarchy. Holons may be defined according to the aim of the investigation, and temporal and spatial scales are not rigidly imposed. However, the main interactions at one spatial support and temporal extent are generally within the holon. Wagenet (1998) gives an example of two interrelated holons: (A) the local processes that interact to form a stable soil aggregate through cementation of inorganic and organic particles, by wetting and drying cycles; (B) the processes that determine the soil's ability to hold water (the soil being a collection of aggregates). Holon B constrains the potential mechanisms at the collection of holons A.

From a philosophical viewpoint, the question "how to upscale a model to a lumped conceptual model" could therefore be answered by "define the holarchy related to the problem, formulate the constraints at the level above s_2 and model the processes of the holon at s_2 based on knowledge at s_1". In the case of upscaling and existing knowledge at s_1, the challenge is to resist the temptation to focus on a bottom-up approach (generalisation of s_1-model) and keep an open eye for constraints at s_2 that come from higher scale levels. Otherwise, key-controlling factors at higher levels may be overlooked. An example is the upscaling of an N-leaching model operating at the field scale and at the scale of a single growing season to the farm and crop rotation scale in order to estimate off-farm N-flows to the environment. This involves more than the aggregation of N-flow processes over multiple fields, soils, crops and fertilisation regimes. From the perspective of spatial upscaling, an estimate of the N-losses (volatilisation, leaching) from on-farm storage may be equally important. From the perspective of temporal upscaling, it is important to assess which part of the unused N in one year by one crop may still be used in the next year by a different crop.

Another philosophical consideration is that in many physical systems two laws are important at a given scale. The conservation of mass and the conservation of momentum (i.e. friction and forces). When moving from a smaller scale to a larger one, conservation of mass is usually not a problem, as long as we take account of possible sinks and sources that may appear at the larger scale (e.g. the N example above). Upscaling while conserving momentum is the real problem that cannot be solved by simple averaging of model parameters, input variables or lumping model components. While deriving a large scale model from a smaller one, we should therefore concentrate on conserving momentum.

Practical considerations

– It is often considered reasonable to refrain from modelling processes that take place at more detailed scales, especially when these are repetitive. Examples of repetitive patterns in s_1 that might be left out when modelling at s_2 are diurnal or seasonal temperature fluctuations in the

temporal domain, and water flow to individual ditches in the spatial domain.

– In some cases models have to be built around available data rather than data tailored to the "optimal" model. Examples are the use of administrative regions rather than process-relevant units (e.g. King et al., 1998) because basic data are available at the administrative level.
– In some cases (cf. the example in subsection 2.5.1), the procedure to upscale the s_1-model is to start from the assumption that this model is fully deterministic and contains all the key controlling factors at s_2 as well. This "complete" model is then analysed by sensitivity analysis for the available extent. This approach has its limitations, because it is not possible to run the model at many time steps or locations. Results from the sensitivity analysis are then used to simplify the model. Aggregated output of the s_1 model can then be used to calibrate and validate the upscaled model.

If the model at s_2 is not entirely empirical but still describes physical processes, it is sometimes referred to as a conceptual model (also "grey box" model to denote that it is not entirely black box and has some limited physical meaning), because it mimics a complex behaviour with a simplifying concept. It is also a lumped model, because it captures a number of complex processes (e.g. flow, transport and retardation) with simpler equations and less parameters. Hence, we refer to such models as "lumped conceptual models". Lumped conceptual models are described in subsection 2.5.1.

For many complex models it is not possible to know in advance what type of behaviour an averaged model at scale s_2 should exhibit. Therefore, we cannot postulate a model concept mimicking this behaviour. In that case, the $g_2[\cdot]$ can be postulated as a meta-model. The general definition of a meta-model would be: "a black-box model relating sensitive inputs of a process-based model directly to its outputs by some kind of statistical relation". This definition still allows three main methods to upscale to s_2 (*Figure 50*). The first is to construct the meta-model entirely at s_1 by fitting the relation between the input variables and parameters and the output variables of the available model at s_1. The thus obtained meta-model should then be applied at s_1 and its output variables upscaled according to some of the methods described in section 2.2. This is the most commonly applied approach. The second method is to fit a relation between inputs at s_1 and upscaled outputs at s_2. Upscaling is then implicitly done within the meta-model. This method is not commonly applied, because upscaling the outputs of the process-based model at s_1 may be hampered by limited data availability. The third method is to construct the meta-model entirely at s_2

from upscaled inputs and outputs of the s_1-model. Limited data availability may effectively prevent the application of this method as well.

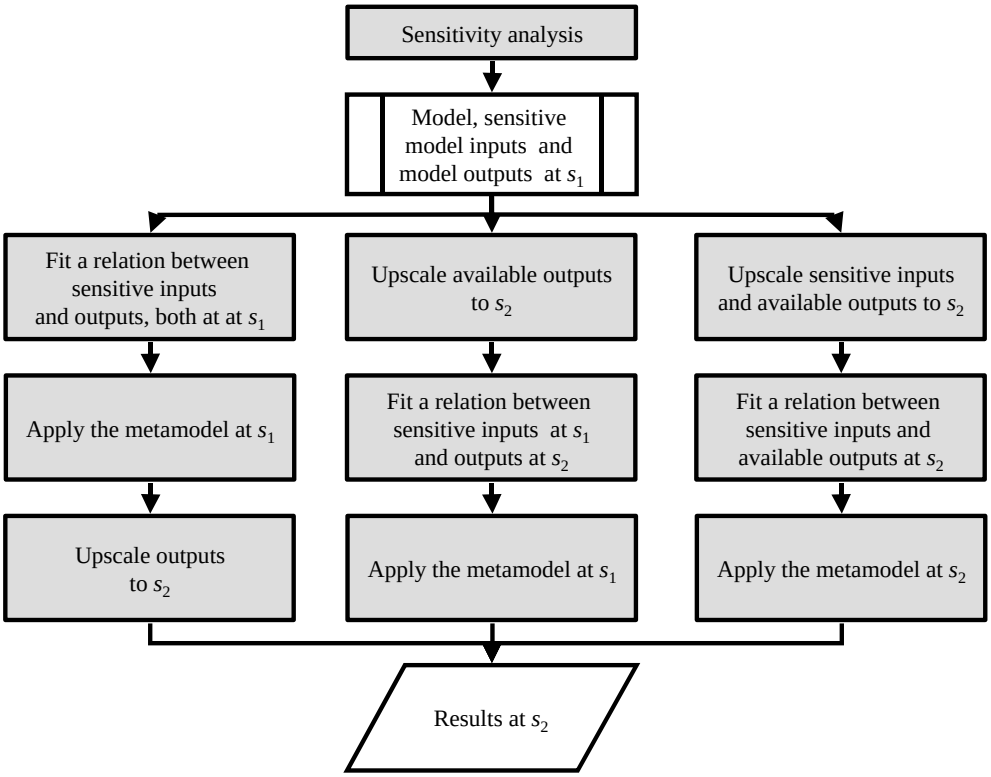

Figure 50. Three methods to construct and apply meta-models in upscaling

In any case, the meta-model should only contain the input variables and parameters that contribute information. The "best" meta-model, i.e. the one with the smallest prediction errors and the least number of regression coefficients, can for instance be selected from a large set of candidate models as the model that minimises Akaike's Information Criterion (Shibata, 1976):

$$AIC = n\left[\ln(2\pi) - \ln n + \frac{2}{n} + 1\right] + 2m + n\ln(SSD) \qquad (2.70)$$

where n is the number of "observations" used in the model fitting (i.e. the number different values of the output variable at scale s_2 used), m is the number of parameters (e.g. regression coefficients) of the meta-model and

SSD the sum of squares of deviations between the observed output variables and those calculated with the meta-model. Further examples of meta-modelling are given in subsection 2.5.2. Since meta-models require much less data and operate much faster than process models, meta-models are widely applied in the field of systems control (cf. Kleijnen, 1987).

2.5.1 Lumped conceptual modelling

Criteria

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model cannot be applied at many locations or time steps.*
- *The models at scale s_1 and s_2 have different forms.*
- *The model for scale s_2 cannot be analytically derived from the model at scale s_1.*
- *It is possible to postulate a lumped conceptual model at scale s_2.*

Method

From the considerations in section 2.5 it follows that no standard procedures exist, and, thus, probably no two research groups would end up with the same model at s_2 when starting with one model at s_1. This is not necessarily a scientific flaw but it does indicate that no standard solutions exist and the procedures may be tedious. Nevertheless, there are a number of notions and tools that can be of assistance in lumped conceptual modelling. These will be elaborated using two examples: a well-described case from soil acidification research and an example from hydrology.

Example 1: soil acidification modelling

The example is taken from De Vries et al. (1998) and concerns the upscaling of soil acidification models. The article describes a procedure that was developed based on experiences with acidification models built for different supports and extents. These models are NUCSAM (mechanistic; support: ha, day; extent: site, 16 years), RESAM (mechanistic; support: km^2, year; extent: country, 16 years) and SMART (mechanistic/empirical; support: km^2, year; extent: continent, 16 years). The upscaling procedure allows the quantification of the effect of model simplification on the reliability of model predictions at the target scale, and is summarised in *Figure 51*.

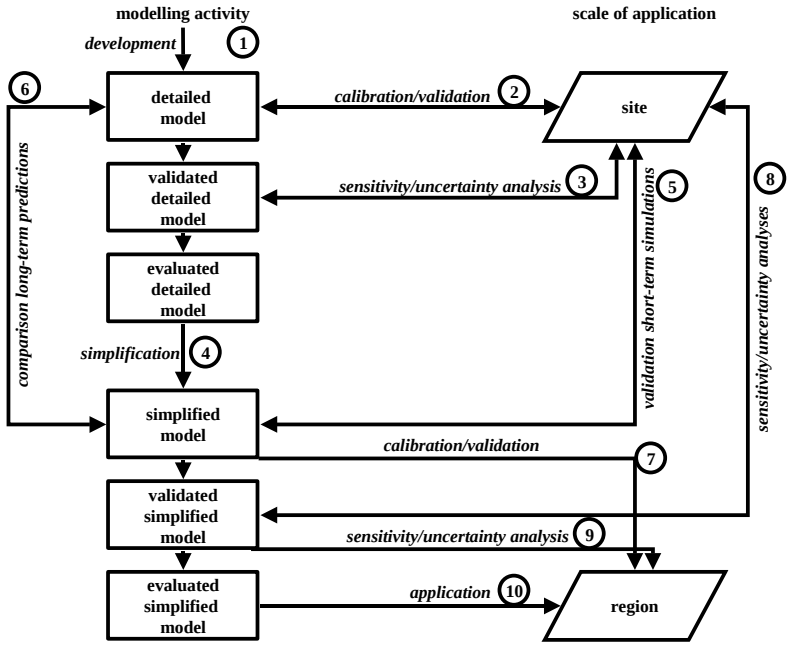

Figure 51. Flowchart for the upscaling of models from a local to a regional scale. Simplified after De Vries et al., 1998.

The numbered activities in *Figure 51* are extensively described in De Vries et al. (1998) and will be summarised below, using our terminology:

1. Start by developing a quantitative mechanistic site-scale (s_1) model with a high process resolution and detailed spatial (vertical) and temporal supports (i.e. NUCSAM).
2. Calibrate this model using data from intensively monitored sites and validate it using another detailed-support data set.
3. Perform a sensitivity analysis to determine, preferably at maximal spatial and temporal extents, to determine the most important model parameters and associated processes.
4. Simplify the site-scale model to a lumped conceptual model at s_2 by (i) a simplified process description based on the sensitivity analysis ("process aggregation"), (ii) aggregation of processes and input data based on the desired temporal and spatial support and extent ("temporal aggregation"), (iii) aggregation of soil layers ("vertical spatial aggregation"). The result would be RESAM or SMART.
5. Compare the performance of the models at s_1 and s_2 on the intensively monitored sites, using (i) the original s_1-support data for the s_1-model, and (ii) spatially and temporally aggregated data at support s_2 for the s_2-

model. In step (ii), De Vries et al. (1998) implicitly assume model linearity of the s_2-model. This is tested in step 8.

6. Compare predicted long-term trends of the models at s_1 and s_2 for actual monitored sites or a set of support unit's representative for the complete extent. This step is meant to assess the possible different effects of both models in case of extrapolation, which is what most models are built for.
7. Calibration of the model at s_2 at the maximal possible extent by minimising the difference between measurements and model results. Validation if possible on another area and/or temporal domain.
8. Assessment of the linearity of the s_2-model by evaluating if averaged input leads to the average result. The purpose is to test the assumption made in step 4.
9. Assessment of the effect of further increasing the support and extent of input data on the frequency distribution of model outputs. The purpose is to estimate the uncertainty as caused by the upscaling.
10. Application of the model at regional extent and specification of the uncertainty of model results at s_2.

The procedure given by De Vries et al. (1998) is quite generally applicable and conscientious. Since it does not prescribe what should be done in case the simplified and detailed model do not compare well (step 6), it could be extended to include an iterative definition of the simplified model.

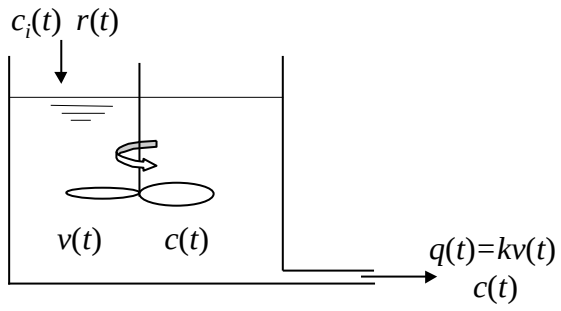

Figure 52. Conceptual model of average leaching to surface water from a parcel: *r(t):* recharge rate [LT^{-1}]; $c_i(t)$: concentration in recharge water [ML^{-3}]; *v(t)*: volume water stored [L^3]; *c(t)*: concentration of stored water and discharging water [ML^{-3}] (perfect mixing assumed); *q(t)*: discharge rate [LT^{-1}] (assumed linearly related to storage); *k:* reservoir parameter [$L^{-2}T^{-1}$].

Example 2: Modelling solute leaching

If there is expert's knowledge or experience about how a model at scale s_2 should behave, it may be possible to define a conceptual model directly as a simple set of equations and parameters. For instance, if we study the leaching from an entire field (scale s_2) to the surface waters surrounding this field, we may discover that the concentration of effluent from a field could be decribed by considering a bucket with a small hole containing water in which dissolved substances are perfectly mixed. *Figure 52* shows the model concept. The resulting model is a set of two ordinary differential equations that is able to describe the discharge from the field to the surface water $q(s_2;t)$ [LT^{-1}] and the concentration of the discharging water $c(s_2;t)$ [ML^{-3}] as a function of time series of recharge [LT^{-1}] $r(s_2;t)$ and concentration of recharge water $c_i(s_2;t)$ [ML^{-3}]:

$$\frac{dq}{dt} = k[r(s_2;t) - q(s_2;t)] \tag{2.71}$$

$$\frac{dc}{dt} = \frac{kr(s_2;t)}{q(s_2;t)}[c_i(s_2;t) - c(s_2;t)] \tag{2.72}$$

As can be seen this model has only one parameter: the reservoir parameter k. Usually it is not possible to find this parameter from expert's knowledge. It has to be determined from calibration together with the procedure denoted in *Figure 17*. For instance, a sufficient number of soil columns is sampled in the field and after determining the necessary parameters at this scale (scale s_1), the complex leaching model is applied at these locations. From these results, the average discharge to surface waters and the average concentration of the discharge water is calculated (these may also be measured directly at an outlet if the surface waters form a closed system). Using for instance a numerical solution to (2.71) and (2.72) the parameter k is calibrated such that the output of this model fits the average output of the $g_1[\cdot]$ models (or the measured discharges and concentrations) as good as possible. If the parameter k is calibrated, it can be regarded as a representative parameter obtained through inverse modelling (as in section 2.3.5). If the model $g_2[\cdot]$ must be applied for many fields for instance in a regional application, the above procedure should be repeated for a representative sample of fields. The reservoir constants obtained for these fields must then be related, for instance through regression, to other easy to come by information at field level, such as the dominant soil type and the area of the field in relation to the area of surface waters.

Literature

Addiscott, T.M. 1995. Entropy and sustainability. European Journal of Soil Science 46: 161-168.

De Vries, W., J. Kros, C. Van Der Salm, J.E. Groenenberg and G.J. Reinds. 1998. The use of upscaling procedures in the application of soil acidification models at different spatial scales. Nutrient cycling in Agroecosystems 50: 223-236.

Gleick, J. 1988. Chaos: Making an new science. Penguin Books, Harmondsworth, 352 pp.

Haigh, M.J. 1987. The holon: Hierarchy theory and landscape research. Catena supplement 10: 181-192. Catena-Verlag, Germany.

King, D., D.M. Fox, J. Daroussin, Y. Le Bissonnais and V. Danneels. 1998. Upscaling a simple erosion model from small areas to a large region. Nutrient cycling in agroecosystems 50: 143-149.

Lovelock, J.E. 1979. Gaia: A new look at life on earth. Oxford University Press, Cambridge MA 157 pp.

O'Neill, R.V., D.L. De Angelis, J.B. Waide and T.F.H. Allan. 1986. A hierarchical concept of ecosystems. Princeton University Press, Princeton, New Yersey, 253 pp.

Runge, E.C.A. 1973. Soil development sequences ands energy models. Soil Science 115: 183-193.

Wagenet, R.J. 1998. Scale issues in agroecological research chains. Nutrient cycling in agroecosystems 50: 23-34.

Weiss, P.A. 1971. The basic concepts of hierarchical systems. In: Weiss, P.A. (ed.) Hierarchically organized systems in theory and practice, pp. 1-43. Hafner publishing Co., New York.

2.5.2 Meta-modelling

Criteria

- *A model is involved.*
- *This model is non-linear in its input variables or parameters.*
- *The model cannot be applied at many locations or time steps.*
- *The models at scale s_1 and s_2 have different forms.*
- *The model for scale s_2 cannot be analytically derived from the model at scale s_1.*
- *It is not possible to postulate a lumped conceptual model at scale s_2.*

Method

A meta-model is a black box model whose parameters are calibrated through regression. Artificial neural networks (ANN) are probably the most general form of meta-models. ANNs, which can be viewed non-linear regression models, have been applied in many fields and are particularly useful if the form of the relationship between input variables and output

variables is unknown. We refer to Dowla and Rogers (1996) for a text book on the application of ANNs.

A somewhat less general form of non-linear regression is performed when using so called of Generalized Additive Models (GAM), of which multiple linear regression models are a subset.

$$g(\mu) = \alpha + \sum_{j=1}^{p} f_j(z_j) \tag{2.73}$$

where $f_j(z_j)$ are functions of all predictor (i.e. input) variables and $g(\mu)$ is the so-called link function describing the way the estimated μ_j (the output variables) are related to the right-hand side of the equation. An identity link function reduces the GAM to a multiple regression model. For example, a meta-model for the concentration of discharging water from a field can have the form of the regression model:

$$C(s_2;t) = a_0 + a_1 c_i(s_2;t) + a_2 r(s_2;t) + a_3 c_i(s_2;t) r(s_2;t) + \varepsilon \tag{2.74}$$

with a_i, the regression coefficients and ε an independent error term.

Alternatively, a meta-model may be postulated in the form of a transfer-function noise (TFN) model with discrete time steps *t*, *t*-1, *t*-2, …(Box and Jenkins, 1976). For instance, a TFN model for the concentration of the discharging water could have the form:

$$\begin{aligned} C_t &= c_t^* + N_t \\ c_t^* &= b_1 c_{t-1}^* + b_2 (c_i r)_t \\ N_t &= b_0 + b_3 \left[N_{t-1} - b_0 \right] + \varepsilon_t \end{aligned} \tag{2.75}$$

where we have omitted the scale variable s_2 for the time being to simplify the notation. The b_i are the time series parameters, c_t^* is the deterministic component, in this case driven by the input load $c_i r$ and N_t an error process which is correlated in time.

The method to construct and apply a meta-model essentially consists of the following steps:

1. Identification of the type of meta-model (from s_1 to s_1, from s_1 to s_2 or from s_2 to s_2, *Figure 50*) to be constructed. This is necessary because it may or may not involve the upscaling of the inputs or model outputs of the s_1-process model.

2. Identification of candidate models in terms of type of model (ANN, GAM, TFN model) and significantly contributing predictor variables, using information criteria like Akaike's Information Criterion.
3. Fitting the selected candidate model.
4. Application of the meta-model and, if necessary, upscaling of the results.

Example

The example is taken from Azimonti et al. (1994) who applied a meta-model for the prediction of pesticide leaching losses. The meta-model itself was developed by Van der Linden and Boesten (1989, cf. Boesten and Van der Linden, 1991). Another application is given by Van de Ven et al. (1992). The meta-model is derived from the PESTLA-model (Boesten and Van der Linden, 1991), which is a field-scale process-based model for the behaviour of pesticides in soils. After sensitivity analysis it was concluded that, within certain climatic constrains, the leaching percentage of a non-volatising pesticide is largely determined by its half-life time (DT_{50},) and its sorption coefficient (K_{OM}). By application of the process-model to a wide range of values for K_{OM} and DT_{50}, under specified climatic conditions (precipitation and temperature) and for a soil with a well-known vertical distribution of organic matter, nomograms like the one in *Figure 53* were obtained.

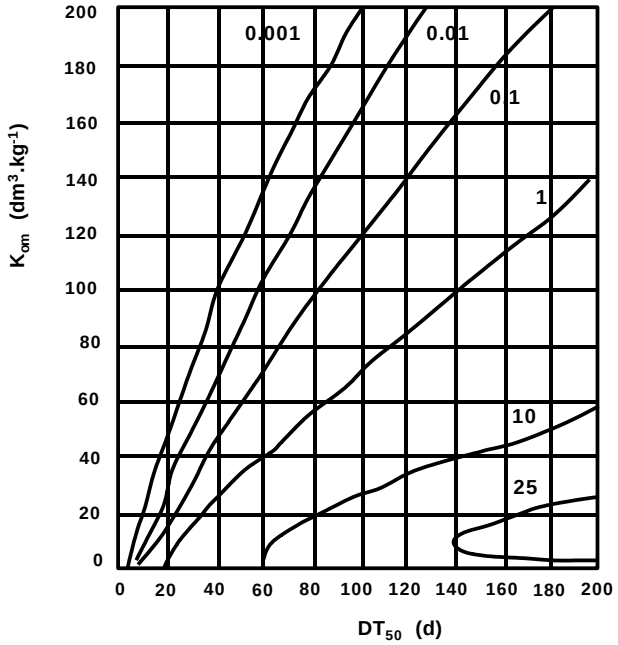

Figure 53. Nomogram predicting the leaching as percentage of the dose from the half-life time (DT_{50}) and sorption (K_{OM}). Reproduced from Van der Linden and Boesten, 1989

This nomogram is a valid metamodel for PESTLA for the standard soil. A wider application range can be obtained if for another soil the organic carbon content (*OC*) of the upper meter is known. The organic matter content (*OM*) can be calculated from *OC* by the relation $OM=OC\cdot 1.72$ (Boesten and Van der Linden, 1991). The apparent sorption constant K^*_{OM} for a pesticide, say Atrazine, is then calculated from:

$$K^*_{OM} = \mathrm{K_{OM}}\left(\frac{OM}{\mathrm{OM_{st}}}\right) \qquad (2.76)$$

where OM_{st} is the organic matter content in the upper meter of the standard soil and K_{OM} is the sorption coefficient for the pesticide as it would be put into PESTLA. From the nomogram or its mathematical equivalent, the leaching percentage for the soil with *OC* can then be derived easily. Because of its smaller data demands, the meta-model can be applied at more locations than the PESTLA-model; nevertheless, it still has a field-scale support. Upscaling is therefore still necessary. This could be done by calculating

block-averages using either geostatistical, design-based or deterministic techniques, depending on the sampling method and sample size of the *OC*-variable (see section 2.2).

The metamodel has been applied in the analysis of policy scenarios on the restriction of pesticide use in Northern Italy (Azimonti et al., 1994), and for Europe (Van de Ven et al., 1992). Upscaling was done rather poorly by applying the meta-model to a value of *OC* for the support s_2 appearing on the final maps. Implicitly, it was assumed (i) that the values for *OC* put into the meta-model were the average values for the map unit, and (ii) the meta-model is linear in its parameters (which is not the case as can be seen in *Figure 53*). Given the limited amounts of data available at continental extents this was probably the only possible approach. The resulting uncertainty was accounted for by presenting the results in fairly broad classes of leaching.

Literature

Azimonti, G., C. Riparbello, M. Maroni. 1994. Sensitivity of soil to pesticide leaching: a combined application of a leaching model and GIS. Proceedings of the Fifth European conference and exhibition on geographic information systems, EGIS'94. Utrecht, EGIS Foundation 1: 1026-1031.

Boesten, J.J.T.I. and A.M.A. Van der Linden. 1991. Modelling the influence of sorption and transformation on pesticide leaching and persistence. Journal of Environmental Quality 20: 425-435.

Dowla, F.J. and L.L.Rogers, 1996. Solving Problems in Environmental Engineering and Geosciences with Artifical Neural Networks. The MIT Press, Cambridge MA, 310 pp.

Kleijnen, J.P.C. 1987. Regression meta-model summarization of model behaviour. In: M.G. Singh (ed.) Systems and control encyclopedia: theory, technology, applications. Pergamon, Oxford.

Van de Ven, F.H.M., L. Kohsiek, J. Da Silva Costa and J.J. Fried. 1992. EC Action programme for sustainable use of groundwater. European Water Pollution Control, 294: 9-19.

Van der Linden, A.M.A. and J.J.T.I. Boesten. 1989. Calculation of the degree of leaching and accumulation of pesticides as a function of their sorption coefficient and degradation rate in plough layers (in Dutch). RIVM, rapport 728800003.

Chapter 3

Downscaling

3.1 A classification of downscaling methods

Compared to upscaling, the problem of downscaling is much more homogeneous. Every downscaling problem essentially consists of reconstructing the variation of a property at scale s_1 (and possibly singling out or sampling some of the support units at s_1), given that only the value at the larger scale s_2, which is the arithmetic average of the property values at scale s_1 within s_2, is known. *Figure 54* shows the principle of downscaling.

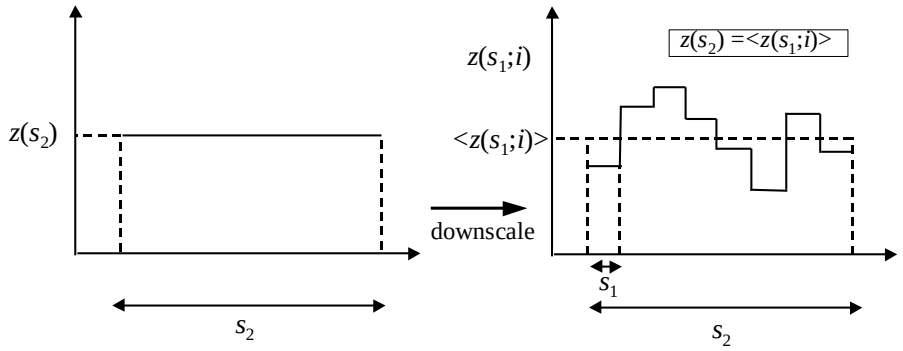

Figure 54 The principle of (deterministic) downscaling

It is important to note that downscaling is, in principle, a problem with a non-unique solution: there are an infinite number of functions describing the variation at scale s_1 within s_2 which have the same average value over the

domain s_2. Only in certain cases, when the form of the function is given, is it possible to reconstruct the variation of the property at scale s_1 within s_2.

In reality, the downscaling problem is a little more subtle than described in *Figure 54*. First, the average value of the property at scale s_2 is not always known exactly. This uncertainty about the average property value at scale s_2 is described by means of a probability distribution or probability density function (pdf). Second, one may choose a single (deterministic) function or a set of equally likely functions describing the variation of the property at scale s_1 within s_2. Consequently, the following fundamental downscaling problems can be distinguished.

Problem 1: deterministic

– The average property at scale s_2 is known exactly.
– Find a single function (deterministic) describing the temporal or spatial variation at s_1 such that the s_2-average value of this function is equal to the known average.
  This downscaling problem is denoted schematically in *Figure 54*.

Problem 2: conditional stochastic

– The average property at scale s_2 is known exactly.
– Find a set of equally probable functions describing the temporal or spatial variation at s_1 such that the s_2-average of each individual function is equal to the known average.

This problem is called stochastic because, instead of finding one function to describe the variation at scale s_1 within s_2, a family of functions is chosen, each of which is assumed to have an equal probability of representing the true but unknown variation at scale s_1 within s_2. The family of equally probably functions is called the "ensemble", or alternatively a "stochastic function". One particular equally probable function is called a "realisation" (see Appendix).

The advantage of choosing a family of functions is that in this way, we can express our uncertainty about the true variation of the property at scale s_1. This means that if the downscaled values are used for further analysis, for instance a downscaled time series of rainfall in a model for flood prediction, we do not get a single answer but a large number of answers (one for each function analysed). The result of this procedure (called Monte Carlo analysis) is that we obtain a probability distribution of outcomes, in this case of flood magnitude (i.e. flood magnitude is a random variable). The width of this distribution expresses the uncertainty about the real flood magnitude. Also, because a probability distribution is derived, the results can be used for risk assessment. The stochastic downscaling problem described here is called "conditional" stochastic, because the s_2-average of each realisation

must be equal to the known average at scale s_2. Of course, in certain applications additional constraints may be necessary. For instance, functions describing rainfall must be non-negative. A schematic presentation of this downscaling problem is given in *Figure 55*.

Problem 3: unconditional stochastic

– The average property at scale s_2 is not known exactly. Instead, the probability distribution function (pdf) of the s_2-average property is known.
– Find a set of equally probable functions describing the temporal or spatial variation at s_1 such that the pdf of the s_2-averages of these functions is equal to the known pdf of the average.

Again, the variation of the property at scale s_1 within s_2 is described with a stochastic function. However, as we do not know the exact average property at scale s_2, we do not require each realisation to have the same s_2-average. Instead, it is sufficient for the s_2-averages of all the realisations together to have the same pdf as given. This downscaling problem is denoted in *Figure 56*.

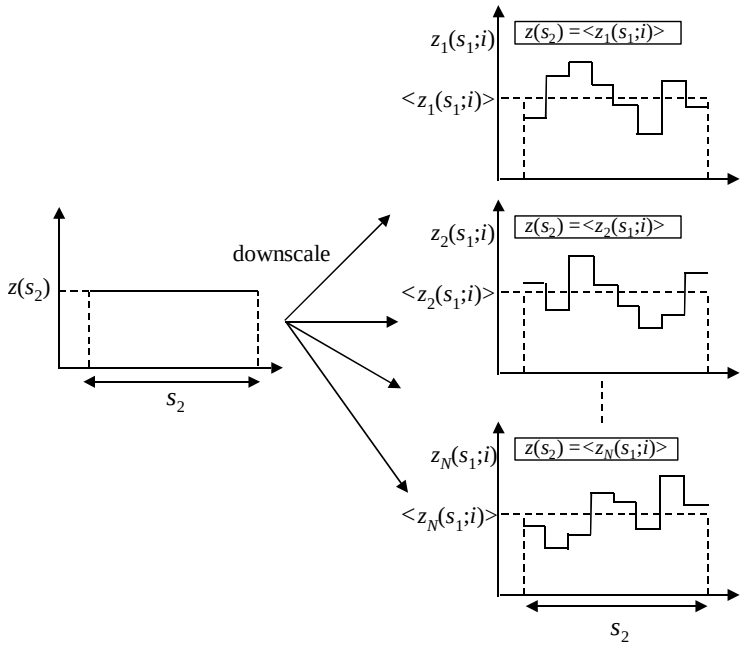

Figure 55. Conditional stochastic downscaling

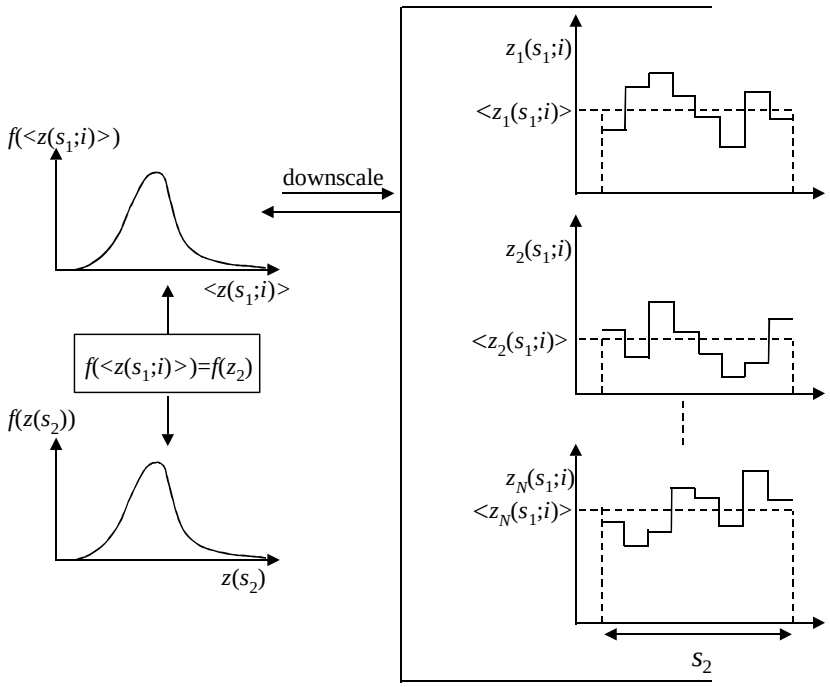

Figure 56. Unconditional stochastic downscaling

In *Figure 54* to *Figure 56* it seems that the function at scale s_1 is constructed directly. However, this is not always the case. The procedure is often as shown in *Figure 57* for the deterministic case. First, an instantaneous function (with point support as in the continuum approach) is constructed with the same s_2-average as given. After that, the function at scale s_1 is derived through averaging of the continuous function over the support units at scale s_1.

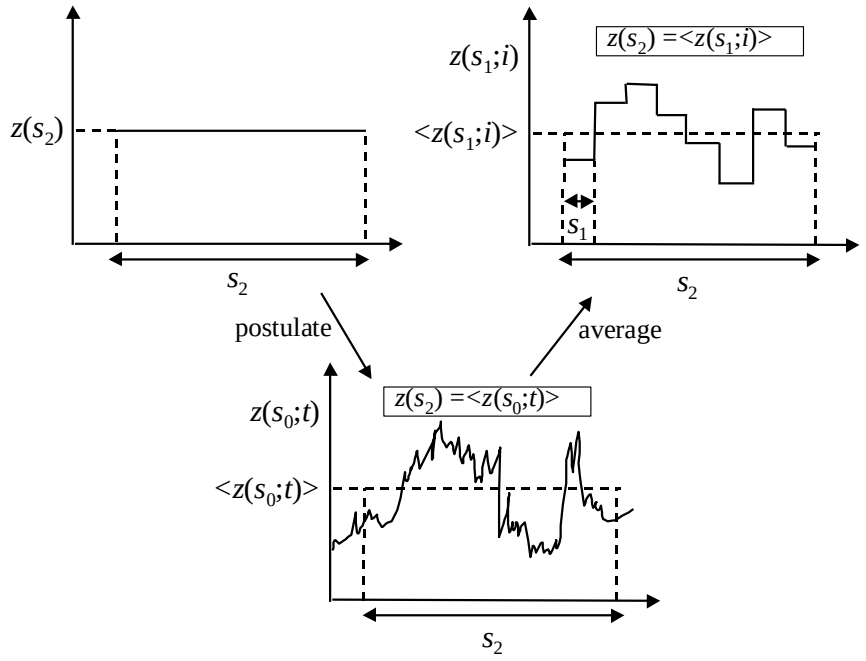

Figure 57. Downscaling using a smaller scale or instantaneous function

There are many ways of defining deterministic or stochastic functions at scale s_1 within the domain s_2. A first classification of downscaling methods is based on the type of functions: empirical, mechanistic or using auxiliary information. Therefore, the downscaling methods are divided into three major classes on the basis of answers to the following two questions:

– *Is there auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2?*
– *Is there mechanistic model describing the (unknown) temporal or spatial variation of the property at scale s_1 within s_2?*

Figure 58 shows a decision tree that can be constructed from these questions. In the following each of these questions and the resulting classification is described in more detail.

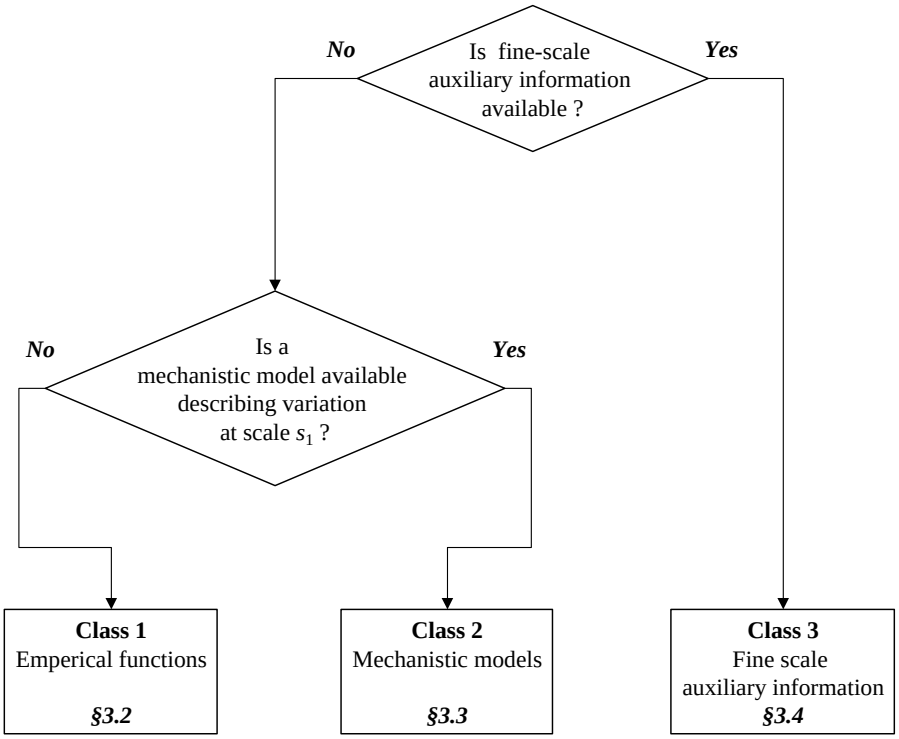

Figure 58. Decision tree to three major classes of downscaling methods

Is there auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2?

If an auxiliary variable is known that is strongly related to the target property at scale s_1, it may be used to estimate the target property for s_1 (or alternatively for a smaller support as shown in *Figure 57*). Examples of using auxiliary information are given in section 3.4 for each of the fundamental downscaling problems. As explained in chapter 2, there is a great number of ways to use auxiliary information, both stochastic and deterministic, empirical and mechanistic. A further subdivision of section 3.4 could therefore be made. However, to limit the scope of this book, we decided to devote only a single section to this subject and provide a few examples.

Is there mechanistic model describing the (unknown) temporal or spatial variation of the property at scale s_1 within s_2?

If no auxiliary information is present that can be directly correlated with the variation at scale s_2, a function describing this variation (or alternatively at a smaller support as shown in *Figure 57*) must be postulated. This function may be the outcome of some mechanistic model describing the variation. For instance, the variation of the concentration of some pollutant may be described by the solution to a partial differential equation that can be derived on the basis of the conservation of mass and momentum. Of course, a mechanistic model does require some additional information, usually parameters, although not necessarily at the scale s_1. For instance, in order to solve the differential equation describing the concentration of the pollutant in groundwater, a dispersion coefficient must be known. If a mechanistic model can be used, it is preferred over empirical models, because the lack of information about the true variation of the property can be partly overcome by universal physical principles such as conservation laws. If no mechanistic model can be formulated the variation at scale s_1 (or smaller) must be described by some empirical function. Mechanistic methods are described in section 3.3, empirical methods in section 3.2.

A further subdivision of each of the three classes, and therefore of the sections, is made on the basis of a the three fundamental methods distinguished above, leading to the following questions (see *Figure 59*):

– *Is the average value of the property at scale s_2 exactly known?*
  If this is the case either a deterministic downscaling problem or a conditional stochastic downscaling problem is solved. If only the probability distribution is known, an unconditional (stochastic) downscaling problem is solved.
– *Is the temporal or spatial variation of the property at scale s_1 within s_2 described with a single deterministic function?*
  Based on this question a choice is made between using a single (deterministic) function to describe the unknown variation of the property at scale s_1 or using a family of equally probable functions (a stochastic function) to describe this variation.

As this subdivision is used for all classes of downscaling methods distinguished, and because above questions leading to this subdivision are described rather extensively above, contrary to chapter 2, no elaborate descriptions of the (sub)classes of downscaling methods are given in the following sections.

Finally, a remark must be made about downscaling and the relation of s_2 to larger scales. If there is a larger support s_3 within which z at scale s_2 varies in time or space according to some function $z(s_2;k)$, the downscaling of support unit $z(s_2;k)$ into $z(s_1;i)$ cannot be performed independently from the downscaling of a neighbouring support unit $z(s_2;l)$. If the downscaling would

be performed independently, it would yield a downscaled function $z(s_1;i)$ that shows an unnaturally large jump at the boundaries between $z(s_2;k)$ and $z(s_2;l)$. Therefore, the downscaling should somehow be performed simultaneously, such that both the averages of all $z(s_2;k)$ (or their probability distributions) are maintained, as well as yielding a function (or realisations of a stochastic function) that show no artificially large jumps at the edges of neighbouring s_2 support units. A number of examples presented have this property by construction (e.g. those of section 3.2), while other examples (e.g. the examples in sections 3.3 and 3.4) don't.

If a method is not designed for downscaling the $z(s_2;k)$ of neighbouring s_2 support units simultaneously, a way to overcome this problem could be as follows: first the downscaling method is applied separately at each s_2 support unit. Next, the downscaled functions $z(s_1;i)$ are adjusted such that the jumps at the edges of the s_2 units stay within acceptable boundaries. This could for instance be achieved by adjusting the variance of the $z(s_1;i)$ in combination with swapping the $z(s_1;i)$ around with a spatial simulated annealing technique (Deutsch, 1994). If the variance adjustment and swapping is done for each s_2 individually (the jumps at the edges of s_2 units are of course evaluated globally) the $z(s_2;k) = <z(s_1;i)>$ are automatically retained. As yet, we do not know of an example where such a technique has been applied. Therefore, for now solving this problem remains a part of future research.

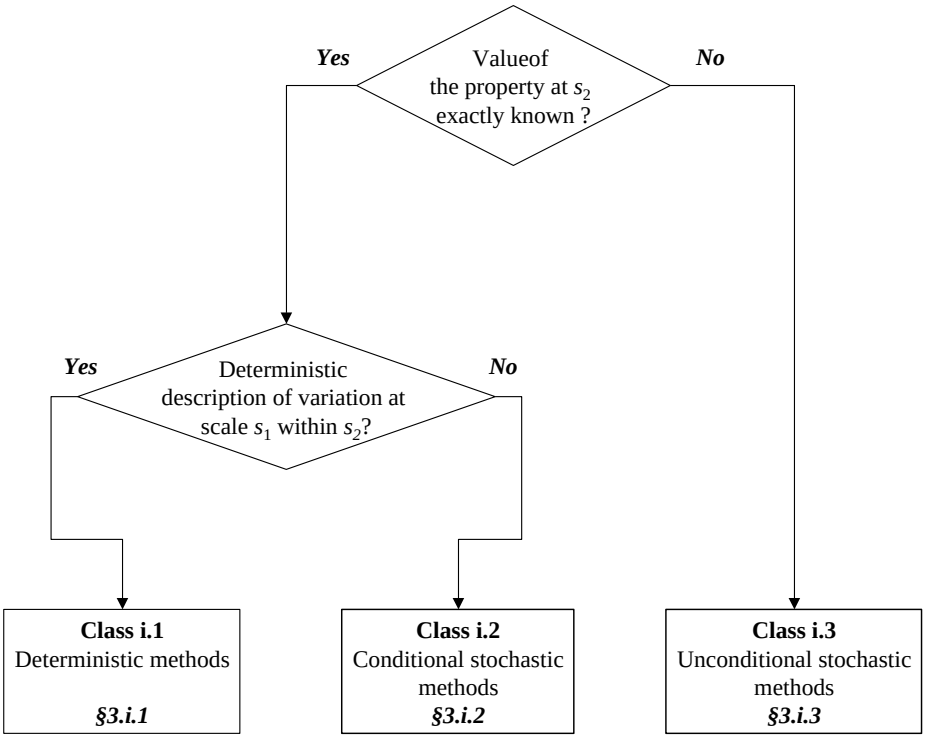

Figure 59. Decision tree to three subclasses of downscaling class 1, 2 and 3.

3.2 Empirical functions

Criteria

- *There is no auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2.*
- *There is no mechanistic model describing the (unknown) temporal or spatial variation of the property at scale s_1 within s_2.*

Further subdivision based on the following questions.

- *Is the average value of the property at scale s_2 exactly known?*
- *Is the temporal or spatial variation of the property at scale s_1 within s_2 described with a single deterministic function?*

3.2.1 Deterministic functions

Criteria

- *There is no auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2.*
- *There is no mechanistic model describing the (unknown) temporal or spatial variation of the property at scale s_1 within s_2.*
- *The average value of the property at scale s_2 is exactly known.*
- *The temporal or spatial variation of the property at scale s_1 within s_2 is described with a single deterministic function.*

Method

A deterministic empirical function (e.g. splines, linear functions, general additive models) is used to represent the variation at scale s_1 within s_2. The actual downscaling then consists of fitting this function such that its average over s_2 is equal to $z(s_2)$.

Example

When describing a soil profile in terms of the course of a certain property with depth, it usually is divided in pedological horizons, the whole depth of each is sampled and mixed for analysis. The obtained value of the soil property is assumed to correspond to the mean value of the horizon. The graph depicting the course of the value of the soil property with depth is thus discontinuous. Sometimes one wants to describe the course with depth by a smooth continuous function, and the challenge is to construct such a function from the discontinuous measurements. Ponce-Hernandez *et al.* (1986) proposed a new method to do this. They coined it the method of equal-area spline functions, because it requires that the curve describing the course within a horizon should be such that the area under the curve equals the product of the mean value and the thickness of the horizon. It is assumed that the course of the value within a certain horizon (say the *i*th) is appropriately described by a quadratic polynomial:

$$\varphi_i = a_i + b_i \cdot x + c_i \cdot x^2 \qquad \text{for} \quad x_{b,i-1} \le x \le x_{b,i} \tag{3.1}$$

where φ_i is the value of the soil property, x is the depth, $x_{b,i-1}$ and $x_{b,i}$ are the upper and lower boundary of the horizon respectively, and a_i, b_i , and c_i are the coefficients to be fitted. The values of the coefficients should be chosen such that they meet the equal-area condition:

$$<\varphi>_i \cdot\left(x_{b,i} - x_{b,i-1}\right) = \int_{x_{b,i-1}}^{x_{b,i}} a_i + b_i \cdot x + c_i \cdot x^2 dx = \\ a_i \cdot\left(x_{b,i} - x_{b,i-1}\right) + \frac{b_i}{2} \cdot\left(x_{b,i}^2 - x_{b,i-1}^2\right) + \frac{c_i}{3} \cdot\left(x_{b,i}^3 - x_{b,i-1}^3\right) \tag{3.2}$$

where $<\varphi>_i$ is the average value in the horizon. To ensure continuity at the upper and lower boundary the respective quadratics should have the same value. At the upper boundary this leads to:

$$a_i - a_{i-1} + \left(b_i - b_{i-1}\right) \cdot x_{b,i-1} + \left(c_i - c_{i-1}\right) \cdot x_{b,i-1}^2 = 0 \tag{3.3}$$

A similar equation holds at the lower boundary. In order to obtain a certain smoothness of the curve the slope at the boundaries should be the same as well. Therefore, at the upper boundary the next equation applies (taking the derivative of Equation 3.3):

$$b_i - b_{i-1} + 2 \cdot\left(c - c_{i-1}\right) \cdot x_{b,i-1} = 0 \tag{3.4}$$

and again a similar equation at the lower boundary.

For a profile with n horizons one thus obtains $3n - 2$ equations, and in order to solve for the $3n$ unknowns (the a,b, and c) two extra values should be given: either the estimates of φ or of $d\varphi/dx$ at the top and bottom of the profile. According to Ponce-Hernandez *et al.* (1986) estimates of the first derivatives at $x_{b,0}$ and $x_{b,n}$ give better fit than estimates of φ_0 and φ_n.

The equal-area splines method was compared to the more conventional method of fitting splines through the data points and was found be considerably superior. Bishop et al. (1999) modified the method of Ponce-Hernandez *et al.* (1986) to use quadratic smoothing splines. This amounts to adding a penalty term to the set of equations in order to impose a certain smoothness to the resulting profile. They did this to take into account that often the horizon averages are not known exactly. *Figure 60*, which is derived from Bishop et al. (1999) demonstrates that in certain cases the method works very well.

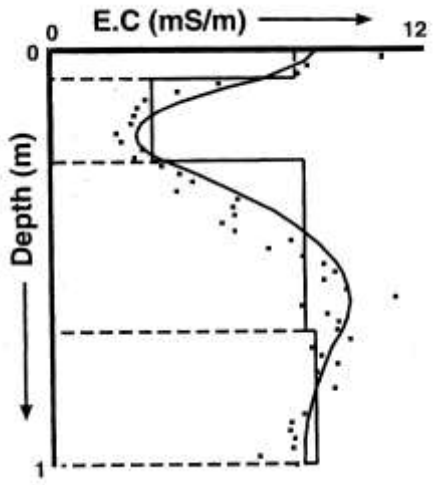

Figure 60. Application of equal-area smoothing splines to downscale EC with depth. Reproduced from Geoderma 91, Bishop et al., Modelling soil attribute depth functions with equal-area quadratic smoothing splines, pp. 27-45. Copyright (1999), with permission from Elsevier Science

Literature

Bishop, T.F.A., A.B. McBratney and G.M. Laslett, 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91: 27-45.

Ponce-Hernandez R., F.H.C.Marriott & P.H.T.Beckett, 1986. An improved method for reconstructing a soil profile from analyses of a small number of samples. Journal of Soil Science 37: 455-467.

3.2.2 Conditional stochastic functions

Criteria

- *There is no auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2.*
- *There is no mechanistic model describing the (unknown) temporal or spatial variation of the property at scale s_1 within s_2.*
- *The average value of the property at scale s_2 is exactly known.*
- *The temporal or spatial variation of the property at scale s_1 within s_2 is described with a stochastic function.*

Method

The value of the property at scale s_2 is known. The conditional stochastic method implies that a stochastic function is sought that describes variation of the property at the lower scale s_1 with the condition that each realisation of this stochastic function, when averaged over s_2 reproduces the value of the property at scale s_2 exactly. Using a stochastic function implies that many different equally probable results at the lower scale can be obtained. Here, the stochastic function used is empirical: not based on physics or auxiliary information.

Example

The example is taken from Torfs (1998) where the downscaling is done by using so-called (stochastic) wavelets. Let us consider two scales s_N and $s_{N/2}$ differing by a factor 2. The property at the coarser scale and the property at the finer scale are linked by the following equation:

$$Z(s_N;i) = \frac{Z(s_{N/2};2i-1) + Z(s_{N/2};2i)}{2} \tag{3.5}$$

For every scale s_N so called wavelet coefficients $W_N(0)$, $W_N(1)$,….., $W_N(i)$ are defined as the difference of the property at the same scale and half that scale:

$$W_N(i) = Z(s_{N/2};2i-1) - Z(s_N;i) = Z(s_N;i) - Z(s_{N/2};2i) \tag{3.6}$$

Figure 61 illustrates this definition.

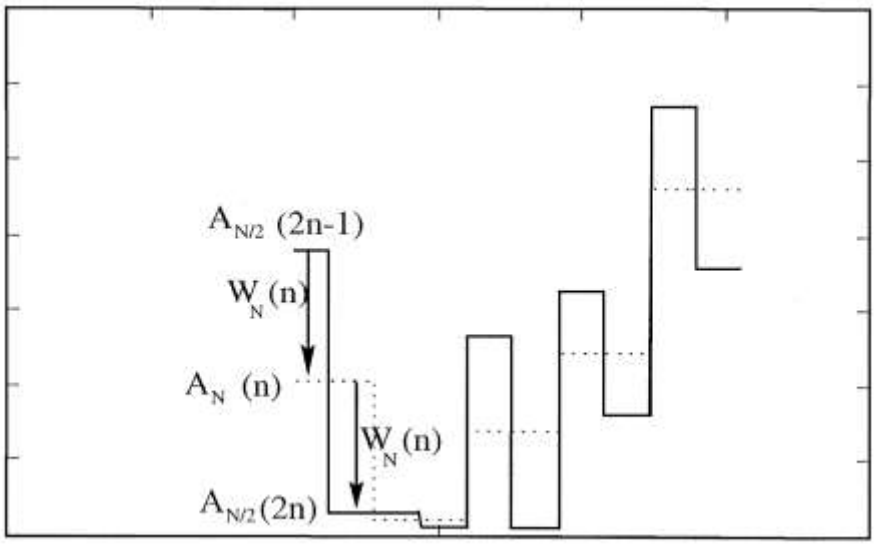

Figure 61. Downscaling by stochastic wavelets

The basic downscaling step made with the wavelet coefficients is thus given as:

$$\begin{aligned} Z(s_{N/2};2i) &= Z(s_n;i) - W_N(i) \\ Z(s_{N/2};2i-1) &= Z(s_n;i) + W_N(i) \end{aligned} \tag{3.7}$$

where Z and W are both stochastic variables and are assumed to have a stationary probability density function, which means that the statistical properties do not change with time (see Appendix). An important statistic is the variance of the wavelet coefficients (its expectation is zero), also called the marginal variance reduction function (MVRF)

$$MVRF(N) = VAR[W_N(i)] = E[W_N^2(i)] \tag{3.8}$$

The MVRF function can be constructed from the observations at the larger scale. The smaller scale values are modelled by attributing a parameterised form to the estimated MVRF values (extrapolation). The form depicted in *Figure 62* has been shown to fit a lot of MVFR-functions.

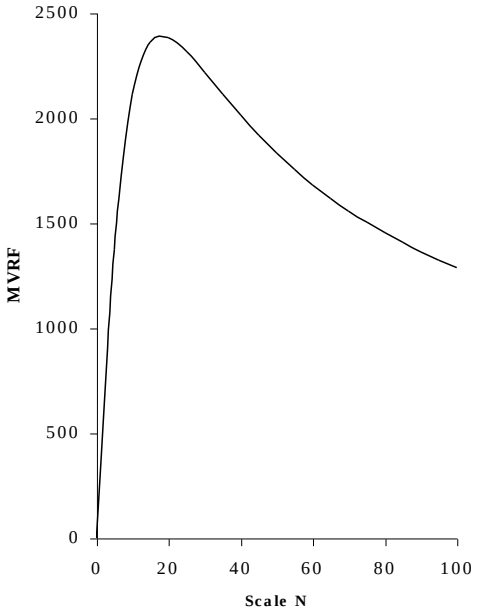

Figure 62. An example of a MVRF-function. This MRVF function pertains to winter discharges of the river Saar (Germany)**.**

The fundamental problem of downscaling can now be stated as follows: Given a realisation of the stochastic process Z_N, construct a realisation of the stochastic process $Z_{N/2}$ with the wavelet function W_N by drawing from a distribution with variance MVRF(N). By successively breaking up realisations into realisations of half their scale using the stochastic wavelet coefficients W_N, $W_{N/2}$, $W_{N/4}$,.. etc., realisations at any desired smaller scale can be obtained. By construction, each of these realisations averages to the value at the largest (observed) scale.

If the MVRF is known, the mean (zero) and the variance of W_N are known at every scale. However, this is not necessarily sufficient to determine the stochastic function W_N completely. The simplest case occurs when it is assumed that W_N is a Gaussian variable and that the autovariance of W_N and the crosscovariance between W_N and Z_N are zero for all N. In this case, equally probable realisations at the smaller scales are obtained by drawing W_N from a Gaussian distribution with mean zero and variance MVRF(N). Torfs (1998) shows that the scaling of temperature values from

the time scale in the order of a day down to that of a second can be satisfactorily performed with this method.

Another procedure assumes again that W_N is Gaussian, but now its standard deviation is taken to be proportional to the value of Z_N, where the proportionality constant equals the ratio of the square root of MVRF and the standard deviation of Z_N. With this procedure it could be shown that downscaling of river discharge could be done successfully, whereas the first method failed to do so.

In a study linking flood frequency to rainfall properties and catchment response, Robinson and Sivapalan (1997) used another conditional stochastic approach to obtain downscaled rainfall data. The method they used is called random cascades and is similar in nature to the random wavelet method presented in this subsection.

Literature

Torfs, P.J.J.F. 1998. Statistical disaggregation of hydrological time series using wavelets. In: Hydrology in a Changing environment Volume I: 81-89. British Hydrology Society.

Robinson, J.S. and M. Sivapalan, 1997. Temporal scales and hydrological regimes: Implications for flood frequency scaling. Water Resources Research, 33(12): 2981-2999.

3.2.3 Unconditional stochastic functions

Criteria

- *There is no auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2.*
- *There is no mechanistic model describing the (unknown) temporal or spatial variation of the property at scale s_1 within s_2.*
- *The average value of the property at scale s_2 is not exactly known; only its probability distribution is known.*
- *The temporal or spatial variation of the property at scale s_1 within s_2 is described with a stochastic function.*

Method

We only know the probability distribution of the property at scale s_2. The unknown variation at scale s_1 is described with a stochastic function. In a numerical framework the downscaling proceeds as follows (see *Figure 56*): the parameters of the stochastic function (e.g. semivariogram, mean,

variance) are adjusted such that when realisations of this stochastic function are simulated and these realisations are averaged over support s_2, the histogram of the simulated s_2-averages resembles the probability distribution of the property at scale s_2. Sometimes it is possible to perform the downscaling analytically. The procedure proceeds as follows. First a stochastic function is defined that describes the spatial or temporal variation of the property at the "point scale" (as in *Figure 57*). This stochastic function is such that its statistics, such as its mean, variance and correlation function, are dependent on a few supposedly scale-invariant coefficients. The stochastic function is averaged to the scale s_2 and expressions are derived for the mean, variance and correlation function at scale s_2. These expressions contain the scale-invariant coefficients, as well as a variable s denoting the support size. By equating these expressions to the estimated statistics at scale s_2, the scale-invariant coefficients of the stochastic function can be estimated. Next, the statistics at any scale s_i can be derived by adjusting the scale variable s in the fitted expressions. An example of the latter method is given below.

Example

Suppose that we need to design a reservoir at the outlet of a small catchment in order to alleviate flash floods and to serve as water source in dry periods. Suppose that we have a single rain gauge in this catchment with a historical record of daily rainfall observations. However, because our catchment is small, we expect the discharge peak after a storm to occur within a few hours. Clearly, to design the reservoir properly, we need hourly rainfall data, instead of daily averages. In a probabilistic framework we are looking for ways to simulate realisations of hourly rainfall data whose statistics, when daily averaged, resemble those estimated from the daily data.

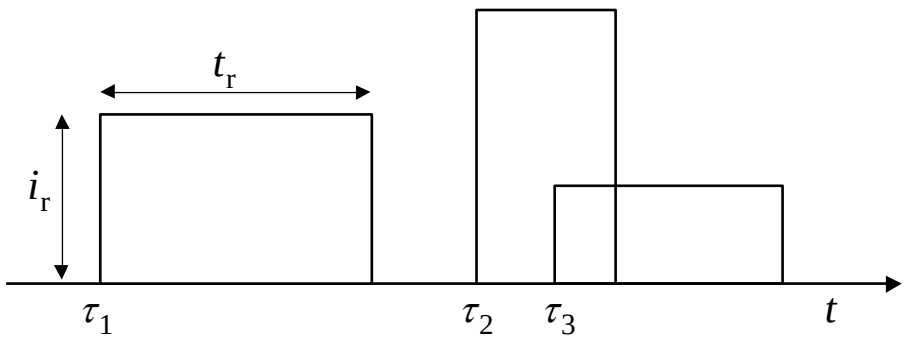

Figure 63. Schematic representation of a Poisson rectangular pulses process; τ_1, τ_2, etc.: (random) arrival times of rainfall cells; i_r: random cell intensity; t_r: random cell duration.

Let $X(t)$ be the "instantaneous" (i.e. point) rainfall intensity (LT^{-1}) as a function of time t, modelled by a stochastic function (or stochastic process). This stochastic process is constructed as follows (see *Figure 63*): Rainfall events (storm cells) arrive according to a Poisson process. This means that for a fixed time interval t-t_0 the number of storm cells N arrived is drawn from a Poisson distribution with parameter $\lambda \cdot (t\text{-}t_0)$ (λ is the average number of arrivals per time unit [T^{-1}]) and that the arrival times τ_1, τ_2,.., τ_N are selected randomly over the interval t-t_0. Each arrival of a storm cell gives rise to a random cell intensity i_r and a random cell duration t_r, each of which are independent of each other and independent from these properties of other cells and drawn from negative exponential probability distributions. The total rainfall process $X(t)$ is then constructed by adding all intensities of the active storm cells at time t:

$$X(t) = \sum_{k=1}^{N(t)} h(t, \tau_k, i_{r,k}, t_{r,k}) \tag{3.9}$$

where $h(t,\tau,i_r, t_r)$ is the contribution at time t of a storm cell arriving at time τ, with intensity i_r, and duration t_r, represented mathematically by the following function:

$$h(t, \tau, i_r, t_r) = \begin{cases} i_r & \tau < t < \tau + t_r \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

Finally, if the rainfall process is averaged over disjunct intervals of size T the T-averaged stochastic process is defined as:

$$X(T;i) = \frac{1}{T}\int_{(i-1)T}^{T} X(t)dt \qquad i = 1,2,.. \tag{3.11}$$

The mean, variance and covariance of $X(T;i)$ can be derived analytically as (Rodríguez-Iturbe, 1986b):

$$E[X(T;i)] = \lambda \bar{i}_r \bar{t}_r \tag{3.12}$$

$$\mathrm{Var}[X(T;i)] = \frac{4\lambda \bar{i}_r^{\,2} \bar{t}_r^{\,3}}{T^2}\left[\frac{T}{\bar{t}_r} - 1 + e^{-T/\bar{t}_r}\right] \tag{3.13}$$

$$\mathrm{Cov}[X(T;1), X(T;n)] = \frac{2\lambda \bar{i}_r^{\,2} \bar{t}_r^{\,3}}{T^2}\left[1 - e^{-T/\bar{t}_r}\right] e^{-(n-2)T/\bar{t}_r} \qquad n \geq 2 \tag{3.14}$$

where $\bar{i}_r$ and $\bar{t}_r$ are the average intensity and the average duration of the storm cells. This stochastic model has three (supposedly) scale-invariant coefficients: $\lambda, \bar{i}_r$ and $\bar{t}_r$.

By estimating the mean, the variance and the lag-one covariance from the daily rainfall data and equating these statistics to expressions (3.12), (3.13) and (3.14) for T=24 hours we have three equations from which the unknown $\lambda, \bar{i}_r$ and $\bar{t}_r$ can be uniquely solved. With these coefficients known, the statistics of the one-hour process can be estimated by setting T=1 hour in Equations (3.13) and (3.14). Also, realisations of the (instantaneous) process $X(t)$ can be simulated, which can then be aggregated to T=1 hour to serve as input rainfall series for a probabilistic assessment of reservoir performance.

The stochastic rainfall model explained here is called the Poisson Rectangular Pulses model (PRP-model). Table 1 shows the 1-hour statistics estimated from 27 year of hourly observations of rainfall near Denver Colorado for the period May 15 to June 16, 1945-1976. Rodríguez-Iturbe et al. (1987) applied the PRP-model for this data set. The second line in Table 1 gives the statistics reproduced from the PRP-model whose coefficients $\lambda, \bar{i}_r$ and $\bar{t}_r$ where estimated from 24 hour data. It can be seen that the variance of the one-hour process is underestimated and that the correlation function decays too slowly when compared to the observed one-hour

statistics. Rodríguez-Iturbe et al. (1987) concluded that this may be caused by the fact that actual storm cells do not arrive randomly in time, but rather as clusters. Therefore, they applied a clustered point process model called the Neyman-Scott Rectangular Pulses model (NSRP-model) to the same data. As shown in Table 1 this model performs much better than the PRP model[12]. Puente et al. (1993) obtained similar results for rainfall data from Hupsel catchment (Netherlands) and the Ceibas catchment (Colombia).

Table 1 Historical statistics of Denver hourly rainfall observations for the period May 15 to June 16, 1945 to 1976 and statistics reproduced from the PRP-model (fitted to 24-hour averaged data) and the NSRP-model (taken from Rodríguez-Iturbe et al., 1987); μ, σ^2, γ, $\rho(k)$ respectively mean, variance, skewness coefficient and lag-k correlation.

Source	μ (mmh^{-1})	σ^2 (mm^{2}h^{-2})	γ	ρ (1)	ρ (2)	ρ (3)	P[rain=0]
Observations	0.089	0.40	10.96	0.48	0.32	0.27	0.94
PRP	0.089	0.18	5.46	0.85	0.76	0.64	0.91
NSRP	0.089	0.42	11.86	0.51	0.24	0.18	0.96

Similar approaches to the one described here have been used for modelling and downscaling of runoff data (Bierkens and Puente, 1990; Puente et al, 1993), spatio-temporal rainfall data (Rodriguez-Iturbe, 1986b) and the output of general circulation models (Wilby et al, 1998).

Literature

Bierkens, M.F.P. and C.E. Puente, 1990. Analytically derived runoff models based on rainfall point processes. Water Resources Research 26(11): 2653-2659.

Puente, C.E., M.F.P. Bierkens, M.A. Diaz-Granados, P.E. Dik and M.M. Lopez, 1993. Practical use of analytically derived runoff models based on rainfall point processes. Water Resources Research 29(10): 3551-3560.

Rodríguez-Iturbe, I., 1986b. Scale of fluctuation of rainfall models. Water Resources Research 22(9): 15S-37S.

Rodríguez-Iturbe, I., B. Febres de Power and J.B. valdes, 1987. Rectangular pulses point process models for rainfall: analysis of empirical data. Journal of Geophysical Research 92(D8): 9645-9656.

Wilby, R.L., M.L. Wigley, D. Conway, P.D. Jones, B.C. Hewitson, J. Main and D.S. Wilks, 1998. Statistical downscaling of general circulation model output: a comparison of methods. Water Resources Research 34(11): 2995-3008.

[12]It must be noted that in Rodríguez-Iturbe et al (1987) the NSRP-model was not fitted to the 24-hour averaged data alone, but to several combinations of statistics from different averaging levels, among which the 24 hour data. However, every combination yielded almost the same values for the scale-invariant coefficients, so that it can be safely assumed that the results in Table 1 for the NSRP model are similar to those that would have been obtained when the NSRP-model was fitted to 24-hour data only.

3.3 Mechanistic models

Criteria

– *There is no auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2.*
– *There is a mechanistic model describing the (unknown) temporal or spatial variation of the property at scale s_1 within s_2.*

Further subdivision made based on the following questions:

– *Is the average value of the property at scale s_2 exactly known?*
– *Is the temporal or spatial variation of the property at scale s_1 within s_2 described with a single deterministic function?*

3.3.1 Deterministic functions

Criteria

– *There is no auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2.*
– *There is a mechanistic model describing the (unknown) temporal or spatial variation of the property at scale s_1 within s_2.*
– *The average value of the property at scale s_2 is exactly known.*
– *The temporal or spatial variation of the property at scale s_1 within s_2 is described with a single deterministic function.*

Method

At scale s_2 the value of a property $z(s_2)$ is known and one wants to estimate values at the smaller scale s_1. There exists a mechanistic model g(.) describing the variation of $z(s_1;i)$ within s_2 such that the s_2-average of $z(s_1;i)$ equals $z(s_2)$. This is usually achieved by adjusting the parameter values or boundary conditions of the mechanistic model.

Example

The example used here concerns soil temperature. Suppose that at some location we need for each month (s_1) the average soil temperature at 10 cm depth to be used as input for a model simulating grass growth for a typical year (s_2). If no soil temperature has been measured, the only information about temperature we have is that the average soil temperature of a typical year is equal to the long term average temperature of the lower atmosphere.

In this example this is taken as 10 ^{0}C (degrees Celcius). For a homogeneous soil with transport of heat in the vertical direction only, the partial differential equation describing the temperature T as a function of time t and soil depth z is given by a linear diffusion equation (Van Wijk, 1963):

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial z^2} \tag{3.15}$$

where α is the diffusivity [L^2T^{-1}] which depends on the heat conductance and heat capacity of the soil.

If we imagine the temperature of the lower atmosphere to be described by a periodic function (i.e. a sinus) with a period of one year and an average of 10 ^{0}C a solution to (3.15) for α constant in time would be:

$$T(z,t) = < T > + T_A e^{-z/D} \sin(\omega t + \phi_0 - z/D) \tag{3.16}$$

where T_A is the amplitude of the atmospheric temperature, ϕ_0 the phase shift, ω the frequency and $D = \sqrt{(2\alpha/\omega)}$ [L] a soil parameter determining the dampening exp[-z/D] and phase shift z/D of the temperature wave as it travels downwards into the soil.

Figure 64 shows an example of the atmospheric temperature and the soil temperature at 10 cm depth described by Equation (3.16) with <T> = 10 ^{0}C, T_A= 5 ^{0}C, $\omega = (2\pi/365) \approx 0.0172$, ϕ_0 = -120ω (minimum temperature of lower atmosphere around March 1) and α = 3 cm^{2}d^{-1} which amounts to a value for D of 18.7 cm. This value of D is comparable to that of a sandy soil with porosity 40% and a water content of 20% (see Van Wijk, 1963).

By construction the yearly average temperature of the soil is equal to <T>. Although Equation (3.16) would be a poor model to describe the soil temperature for a given day in a given year, it would perform adequately if the goal is to obtain the monthly average soil temperature for a typical year (a year with an average soil temperature of 10 ^{0}C). For instance, if we are interested in the average temperature of April because this is the month that young seedlings start to grow, from *Figure 64* we deduce an average April temperature of 5.9 ^{0}C for a typical year.

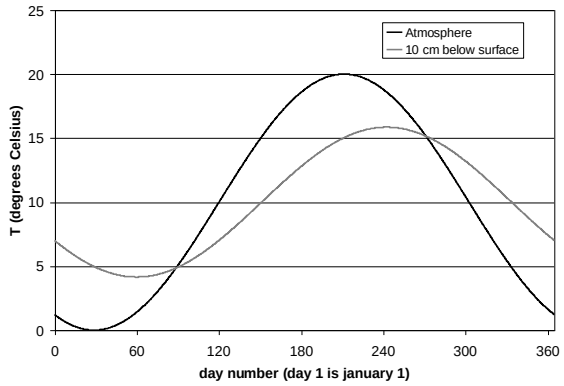

Figure 64. Variation of temperature in time for the lower atmosphere and the soil at 10 cm depth as described with Equation (3.16); $\langle T \rangle = 10\ ^0$C, $T_A = 5\ ^0$C, $\omega = (2\pi/365) \approx 0.0172$, $\phi_0 = -120\omega$ and $\alpha = 3$ cm^{2}d^{-1}

Literature

Van Wijk, W.R. (ed.), 1963. Physics of Plant Environment. North Holland, Amsterdam, 382 pp.

3.3.2 Conditional stochastic functions

Criteria

- *There is no auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2.*
- *There is a mechanistic model describing the (unknown) temporal or spatial variation of the property at scale s_1 within s_2.*
- *The average value of the property at scale s_2 is exactly known.*
- *The temporal or spatial variation of the property at scale s_1 within s_2 is described with a stochastic function.*

Method

The value of z at scale s_2 is known. Instead of using the mechanistic model to construct a single function describing the variation at scale s_1 within s_2, it is used to construct equally probable realisations of a stochastic

function. This is done either by adding a noise component to the model or assuming one or more of its parameters to be stochastic variables or stochastic functions. This way, the results of the mechanistic model is a stochastic function too. If the value $z(s_2)$ is known exactly, this stochastic function should be conditional: the s_2-average of each of its realisations should be equal to $z(s_2)$.

Example

The same example as in section 3.3.1 is used. Again, the goal is to obtain monthly average (s_1) temperatures for a typical year (s_2) where the average temperature of such a year is 10 ^{0}C. Solution (3.16) is only valid if α is constant with time. This is a strong assumption because α depends on the heat conductance and the heat capacity of the soil, which are in turn determined by the soil moisture content. We know that the soil moisture content changes with time following alternating periods of precipitation and evaporation. Therefore, an improvement would be if in Equation (3.15) the temporal variation $\alpha(t)$ is included. However, to include $\alpha(t)$ we would either have to measure the soil moisture content with time or model it using a soil moisture flow model, in which case we would need time series of precipitation and evaporation and additional soil physical parameters such as soil moisture retention and unsaturated conductivity relationships.

If these data are not present, the next best thing is to account for the uncertainty accruing from the unknown varying $\alpha(t)$[13]. To do this we have modelled $\alpha(t)$ as the following discrete stochastic process with time steps of Δt = 1 day:

$$\begin{aligned} \alpha_t &= 3\exp(\xi_t) \\ \xi_t &= 0.995\xi_{t-1} + 0.007\chi_t \end{aligned} \qquad (3.17)$$

[13] Of course other sources of uncertainty could be included to in this analysis such as unknown soil heterogeneity (α also varying with depth: $\alpha(z,t)$) and the temperature of the lower atmosphere varying more erratically than a nice periodic function. Here only the uncertainty due to $\alpha(t)$ is analysed. Readers familiar with the theory of soil heat conduction have probably noted that, if α is function of time, Equation (3.15) only applies if the heat conductance λ is a function of time and the heat capacity C is constant: $\alpha(t) = \lambda(t)/C$. Given that both λ and C depend on the soil moisture which varies with time (Van Wijk, 1963), this is a rather strong assumption. However, to keep our example simple we have used (3.15) as the governing equation.

where χ_t is a standard Gaussian deviate. From (3.17) it follows that α_t has a lognormal distribution with geometric mean equal to 3. With (3.17) realisations of α_t can be simulated. Next, using a finite difference scheme for Equation (3.15) (time step 1 day), for each realisation α_t a realisation of the T_t (the temperature with time steps of 1 day) is obtained. Each realisation is an equally valid representation of the temperature variation with time, given the uncertainty about $\alpha(t)$. As we want to know the temperature variation for a given year the average of each realisation should be equal to 10 ^{0}C. This can be easily achieved by calculating the yearly average of each realisation an correcting al of its simulated values for the difference with 10 ^{0}C.

Figure 65 shows five realisations of the temperature variation in time simulated with Equation (3.15) and (3.17). The analytical solution according to Equation (3.16) for $\alpha = 3$ (cm^{2}d^{-1}) is also shown. The average of each realisation is 10 ^{0}C. If we are interested in the average monthly temperature for April for a typical year (typical for the temperature, not for the soil moisture), we find from the five realisations that this may vary between 5.5 and 7.1 ^{0}C. This variation is caused by the (unknown) variation in soil moisture content.

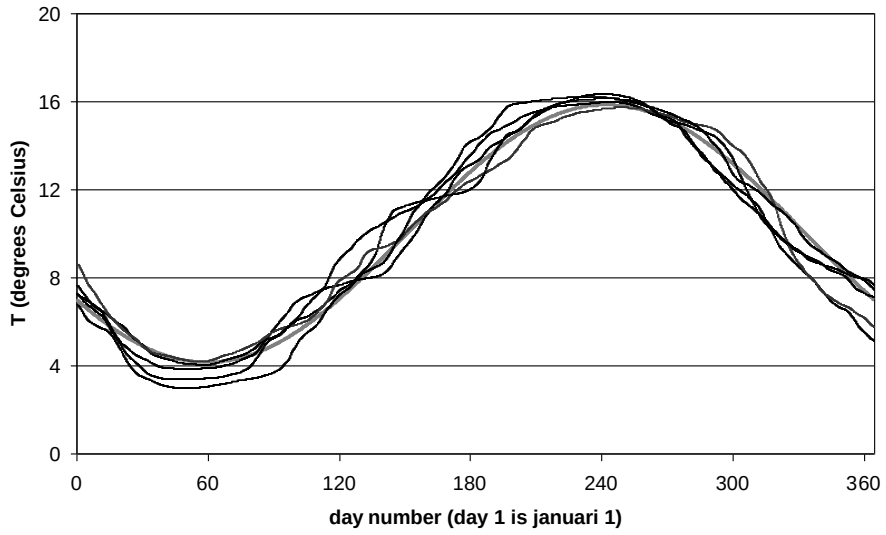

Figure 65. Five realisations of the variation of soil temperature with time at 10 cm depth obtained through Equation (3.15) and (3.17). Each realisation has a yearly average

temperature of 10 ^{0}C. The smooth thick line is the analytical solution obtained with Equation (3.16) for $\alpha = 3$ cm^{2}d^{-1}. All other parameters the same as in *Figure 64*

3.3.3 Unconditional stochastic functions

Criteria

- *There is no auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2.*
- *There is a mechanistic model describing the (unknown) temporal or spatial variation of the property at scale s_1 within s_2.*
- *The average value of the property at scale s_2 is not exactly known; only its probability distribution is known.*
- *The temporal or spatial variation of the property at scale s_1 within s_2 is described with a stochastic function.*

Method

The value of $z(s_2)$ is not known exactly. Instead, $z(s_2)$ is modelled as a stochastic variable $Z(s_2)$. The probability distribution (pdf) of this stochastic variable is assumed to be known. Using a mechanistic model with added noise or with random coefficients, realisations of a random function are simulated. Each of these realisations represents an equally probable representation of the variation of $z(s_1;i)$ within s_2. If a large number of such realisations are simulated, and the s_2-average of each of these realisations is taken, then the histogram of these s_2-averages should resemble the pdf of $Z(s_2)$.

Example

Once more the example of section 3.3.1 is used. Here, the method of 3.3.2 is used to simulate realisations of the temperature variation in time. Instead of providing the realisations for a typical year ($<T> = 10$ ^{0}C), we are now interested in obtaining downscaled realisations for any given future year. So, instead of knowing the yearly average value $<T>$ we have a probability distribution (pdf) of $<T>$. If we assume that the pdf of yearly average soil temperature is the same as that of the lower atmosphere, we can obtain this pdf from air temperature measurements of a nearby meteo station. Suppose that we know that $<T>$ has a mean of 10 ^{0}C and a standard

deviation of 0.5 ^{0}C. To simulate suitable realisations of temperature variation whose yearly averages have the proper mean and standard deviation we have to adjust the coefficients in the stochastic process of Equation (3.17) accordingly. Of course in our example we have already chosen the values (0.995 and 0.007) such that this is indeed the case.

Figure 66 shows five realisations obtained with Equations (3.15) and (3.17) that describe the possible variation of time of the soil temperature at 10 cm depth for any given year. Among these five realisations the yearly average temperature varies between 9.3 and 10.5 ^{0}C, while the average temperature of April varies between 5.2 and 6.7 ^{0}C. This variation not only reflects the unknown variation of soil moisture content within a typical year, but also between the years.

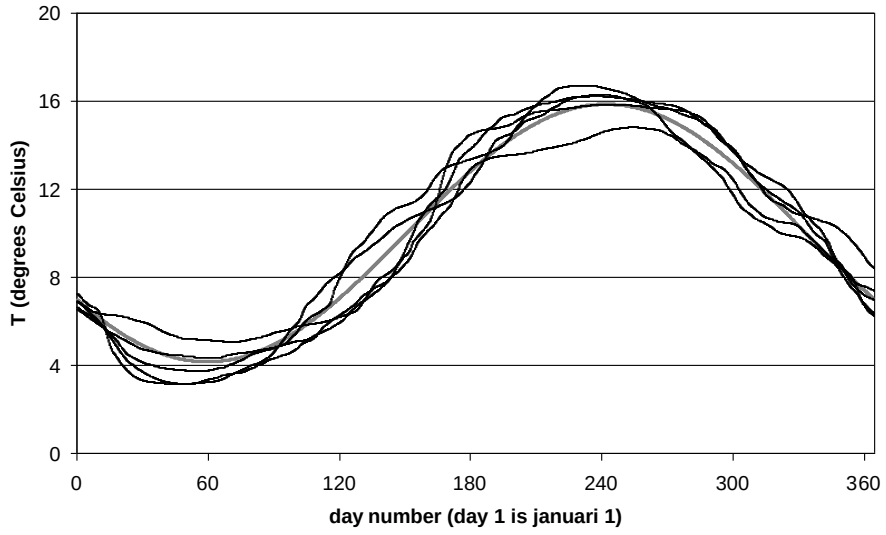

Figure 66. Five realisations of the variation of soil temperature with time at 10 cm depth obtained through Equation (3.15) and (3.17). The smooth thick line is the analytical solution obtained with Equation (3.16) for $\alpha = 3$ cm^{2}d^{-1}. All other parameters the same as in *Figure 64*.

3.4 Fine scale auxiliary information

Criteria

– *There is auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2.*

Further subdivision is made based on the following questions:

– *Is the average value of the property at scale s_2 exactly known?*
– *Is the temporal or spatial variation of the property at scale s_1 within s_2 described with a single deterministic function?*

3.4.1 Deterministic functions

Criteria

– *There is auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2.*
– *The average value of the property at scale s_2 is exactly known.*
– *The temporal or spatial variation of the property at scale s_1 within s_2 is described with a single deterministic function.*

Method

The method to use auxiliary information in a deterministic sense to downscale a property consists of the following steps:

1. Formalise the relation between the auxiliary information and the property, both at the detailed scale s_1. This step involves the identification of a suitable relation, and, possibly, the fitting of some parameters.
2. Derive a relation from the one obtained in step 1 to ensure that mass is preserved. Usually, this is done by combining the relation from step 1 with a relation that shows how the averaged values at the detailed scale equal the known average at the coarse scale.
3. Apply the relation from step 2 at the detailed scale.

Example

The example is based on a study by Sivapalan (1993), but applied to an artificial data set to demonstrate the approach. Sivapalan had the objective to determine the fine-scale variability of water storage in a sloping landscape, using fine-scale topographic data (for instance in the form of a digital

elevation model) and a value for the over-all water storage. The over-all water storage could for instance be obtained by interpreting satellite images or by a coarse scale model.

Analogous to step 1 in the method described above, Sivapalan started by identifying and formalising a suitable relation. Therefore, a topographic parameter was sought that would explain much of the variability in steady state perched water tables among hillslopes. Sivapalan obtained the following relation[14]:

$$\frac{c_i}{a_i} = f_0 + f_1\left(\frac{1}{\tan\beta_i}\right)^{\delta} \tag{3.18}$$

where c_i is the steady state storage for the i^{th} hillslope with area a_i and average slope tan β_i. In the entire section 3.4 we will denote the ratio c/a as the storage fraction (-), i.e. the storage over the area and assuming a fixed profile depth. The parameters f_0, f_1 and δ depend on the hydrogeology and the geometry of the hillslope.

In analogy to step 2, Eq. (3.18) was combined with a mass-preservation relation:

$$C = \sum_{all\ i} c_i = f_0 \sum_{all\ i} a_i + f_1 \sum_{all\ i} a_i \left(\frac{1}{\tan\beta_i}\right)^{\delta} \tag{3.19}$$

By combination of Eq. (3.18) and (3.19), f_0 can be eliminated, and the sought relation is obtained:

$$\frac{c_i}{a_i} = \frac{C}{A} + f_1\left\{\left(\frac{1}{\tan\beta_i}\right)^{\delta} - \mu\right\} \tag{3.20}$$

[14] Sivapalan (1993) obtained this relation through regression of relief properties and the outcome of many simulations with a fine scale water balance and runoff model. Strictly speaking, the relation is therefore only valid for hillslopes and the particular type of catchments it was derived for. Other catchment types or areas would possibly require the derivation of another relationship using the results of fine scale hydrological modelling. However, in this example Equation (3.18) is merely used to demonstrate the principle. The method applied here is in fact very general. If other factors are identified that control the variation of the property within s_2, e.g. soil type or distance to the nearest water course, these can be easily incorporated in Equation (3.18) as additional regression terms: e.g. $f_2\{soil\} + f_3\{distance\}$ etc, as long as the mass preservation is ensured through Equation (3.19).

in which C/A is the storage fraction at the coarse scale s_2, and μ is the area weighted average of the topographic index, calculated as.

$$\mu = \sum_{all\ i} \frac{a_i}{A} \left(\frac{1}{\tan \beta_i} \right)^{\delta} \tag{3.21}$$

The final step is, to apply Eq. (3.20) to each hillslope i. This is easy after a_i and tan β_i have been calculated from the digital elevation model at the detailed scale s_1. *Figure 67* shows the resulting downscaled storage fractions obtained with 3 values for f_1, δ=0.75 and C/A=0.5 m^3/m^3.

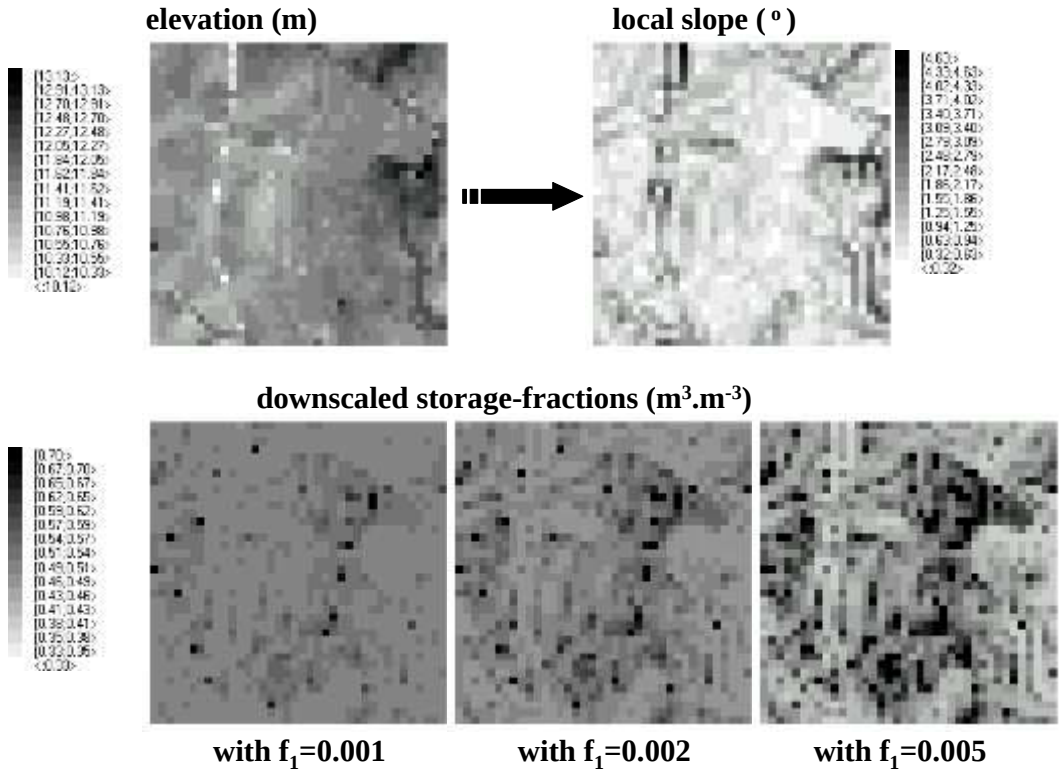

Figure 67. Downscaled storage fractions using a digital elevation model and a deterministic function (Eq.3.20) with three values of parameter f_1.

Literature

Sivapalan, M. 1993. Linking hydrologic parameterizations across a range of scales: hillslope to catchment to region. In: Exchange processes at the land surface for a range of space and time scales. Proceedings of the Yokohama symposium, IAHS Publ. 212, 1993: 115-123.

3.4.2 Conditional stochastic functions

Criteria

- *There is auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2.*
- *The average value of the property at scale s_2 is exactly known.*
- *The temporal or spatial variation of the property at scale s_1 within s_2 is described with a stochastic function.*

Method

The method to be applied here is quite comparable to that in section 3.4.1, with an extension to allow for the incorporation of the ensemble of equiprobable functions instead of only one deterministic function. Subsequent activities are:

1. Formalise a set (ensemble) of equiprobable relations between the auxiliary information and the property, both at the detailed scale s_1. This step involves the identification of suitable relations, and, possibly, the fitting of some parameters.
2. Derive a set (ensemble) of equiprobable relations from the one obtained in step 1 to ensure that mass is preserved. Usually, this is done by combining each relation from step 1 with a relation that shows how the averaged values at the detailed scale equal the known average at the coarse scale.
3. Apply the set of equiprobable relations from step 2 at the detailed scale.
4. Assess the distribution at each support unit at the detailed scale s_1 from the set of results obtained in step 3.

Example

The example is an extension of that in section 3.4.1. The only modification is that we consider the hydrogeological parameter f_1 uncertain. *Figure 67* already gives three possible results of the downscaled storage fractions for three values of f_1. The example for conditional stochastic

downscaling was constructed by assuming that f_1 is normally distributed with mean 0.002 and standard deviation 0.001. We drew 100 values of f_1 from its probability distribution and subsequently calculated 100 maps with downscaled storage fractions. Due to the mass-preserving properties of Equation (3.20) the average of each map is equal to *C*. For each support unit (here: grid cell) at scale s_1 we would then obtain a distribution of resulting storage fractions. *Figure 68* gives, as a result, a map with median values for the storage fraction in each grid cell.

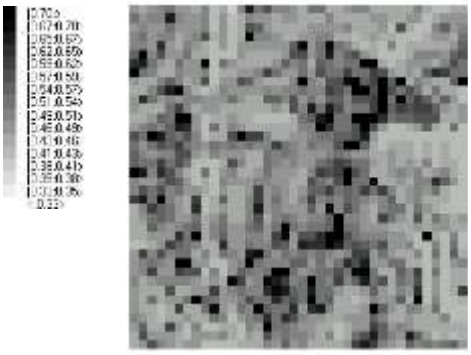

Figure 68. Median values of downscaled storage fractions obtained by applying 100 drawn values of f_1 to each cell of the digital elevation model, using Eq. (3.20).

Literature

Sivapalan, M. 1993. Linking hydrologic parameterizations across a range of scales: hillslope to catchment to region. In: Exchange processes at the land surface for a range of space and time scales. Proceedings of the Yokohama symposium, IAHS Publ. 212, 1993: 115-123.

3.4.3 Unconditional stochastic functions

Criteria

- *There is auxiliary information that can be used to explain some of the unknown temporal or spatial variation of the property at scale s_1 within s_2.*
- *The average value of the property at scale s_2 is not exactly known; only its probability distribution is known.*

– *The temporal or spatial variation of the property at scale s_1 within s_2 is described with a stochastic function.*

Method

Two methods can be applied to downscale with unconditional stochastic functions while using auxiliary data. One method would be to determine the probability density function at the detailed scale s_1 directly from that at s_2. The other would be (i) to generate realisations of the downscaled field, drawing from the probability distribution at s_2, and (ii) subsequently calculate probability distributions (histograms) at each support unit at s_1.

In summary, the second method is as follows:

1. Formalise the relation between the auxiliary information and the property, both at the detailed scale s_1. This step involves the identification of a suitable relation, and, possibly, the fitting of some parameters.
2. Derive a relation from the one obtained in step 1 to ensure that mass is preserved. Usually, this is done by combining the relation from step 1 with a relation that shows how the averaged values at the detailed scale equal the known average at the coarse scale.
3. Draw a value of the property from its distribution at s_2 and apply the relation from step 2 at the detailed scale.

Repeat step 3 a sufficient number of times to adequately sample the distribution at s_2.

4. For each support unit at s_1, determine the probability distribution from the results of step 3.

Example

The example is an extension of that in section 3.4.1. The only modification is that we consider the probability distribution of the storage fraction *C*/*A* at scale s_2. We constructed the example for unconditional stochastic downscaling by assuming an empirical probability distribution of *C*/*A* (the thick line in the probability graphs in *Figure 69*. We drew 100 values of *C*/*A* from this probability distribution and subsequently calculated 100 maps with downscaled storage fractions. For each support unit at scale s_1 we then obtained a distribution of resulting storage fractions. *Figure 69* gives, as a result, a map with median values for the storage fraction in each grid cell, and, additionally, shows distributions of *c*/*a* in three support units at s_1. It is clear from the lower probability graph, that local probability distributions can substantially deviate from the distribution at s_2.

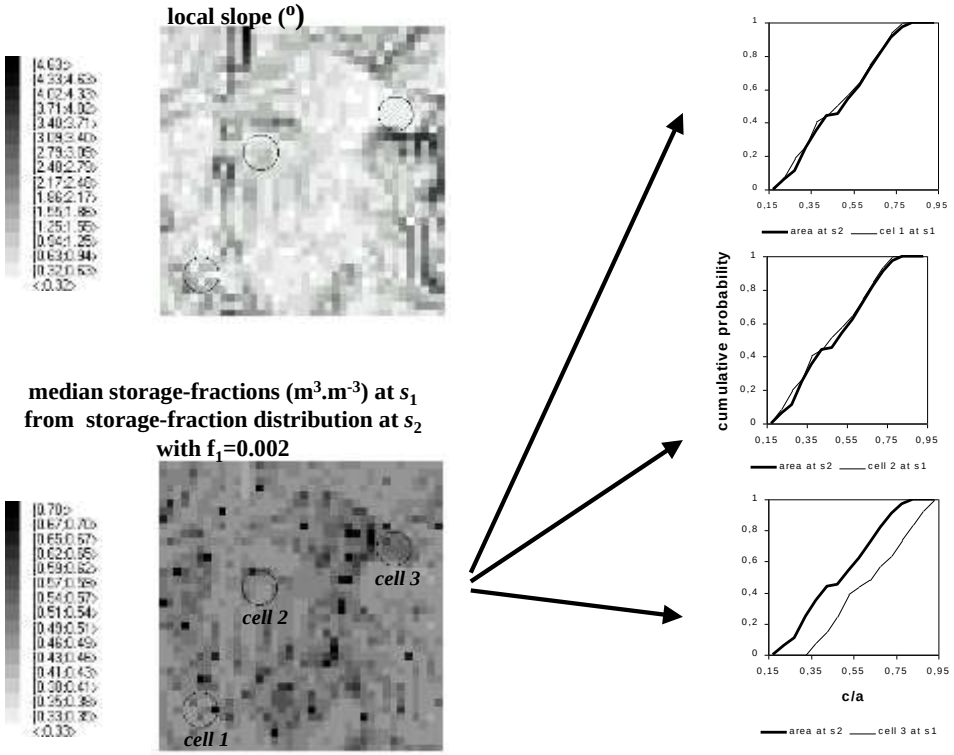

Figure 69. Median values of downscaled storage fractions obtained by applying 100 drawn values of the probability distribution of C/A to the digital elevation model, using Eq. (3.20), with three examples of resulting probability distributions of c/a .

Literature

Sivapalan, M. 1993. Linking hydrologic parameterizations across a range of scales: hillslope to catchment to region. In: Exchange processes at the land surface for a range of space and time scales. Proceedings of the Yokohama symposium, IAHS Publ. 212, 1993: 115-123.

Chapter 4

A simple DSS for upscaling and downscaling

4.1 Purpose and philosophy of the DSS

The purpose of the Decision Support System (DSS) is to provide assistance in choosing the most appropriate upscaling or downscaling method. The DSS is based on the same sequence of questions asked in chapters 2 and 3 in this book (i.e. the decision trees) and leads the user to the most relevant class and subclass of upscaling or downscaling methods. A help-functionality is added to the DSS, from which further reading on the appropriate upscaling or downscaling subclass is easily done. The DSS thus provides an alternative way of navigating through the text of this book.

From a practical viewpoint, the first question to be asked is, whether upscaling or downscaling should at all be part of the research. This question can be answered only if the researcher is fully aware of the scales involved in his or her research project. Thereto, a method is added to the DSS that forces the user to identify the spatial and temporal scales (support) associated with the different objects in the research chain. Here, a research chain is defined as the course data take from sampling (through models) to the target property, whereby possibly change of support occurs. An object in the research chain is either a model, or data (an input variable, an input parameter or an output variable). So the first step a user of the DSS has to take, is to identify the following:

1. What is the target property of the research project, and at what temporal and spatial support size do we want to know this target property?
2. What model(s) is necessary to obtain this target property, and for what temporal and spatial support size was it developed and tested?
3. Which input variables and parameters enter the model(s), and what is their temporal and spatial support size.

4. How are the models and the data objects linked in sequence.

If the issues above are identified, they can be entered into the DSS. From the support sizes of the objects of the research chain and the linkage between these objects the DSS can then evaluate whether upscaling or downscaling is necessary between two adjacent objects in the chain. Note that within this DSS, scale transfer between two objects is always exclusive the starting object and inclusive the ending object. It is necessary to formalise this, because the actual scale transfer often lies hidden inside models. After the research chain is so defined, a sequence of questions leads to a subclass of upscaling methods or downscaling methods.

4.2 Functionality and options at startup

The DSS has the following basic functions:

a) Definition and visualisation of research chains over different scales;
b) Identification of scale transfer steps;
c) Identification of the most appropriate upscaling or downscaling method.

Each of these functions is associated with a separate form in the DSS. At function a), the DSS shows the way the research chain is defined over space and time scales in diagrams. At function b), the DSS provides a tool to select subsequent objects from the research chain. At function c), the DSS shows the decision trees being navigated through by the sequence of questions. Each of the diagrams can be saved at any time by right-clicking it. A defined research chain can be saved to and loaded from a formatted (ASCII) file via the menu or a speedbutton. The contents of this file are described in *Table 2*.

Table 2. Contents of the file with the research chain (file header and four objects). The file header contains the space and time units of support delimited by a white space. Each descriptor takes 15 positions, so each value starts at position 16. The grayshaded field gives a description

Descriptor		**Value**	**Description**
m.2 year			Space and time units of support (user input)
[nodeno]	:	0	Number of object (assigned automatically)
[xcenter]	:	850	X-coordinate in diagram (assigned automatically through position in research chain)
[ycenter]	:	4	Y-coordinate ($_{10}$log of spatial multiplier, user input)
[ycenterT]	:	0	Y-coordinate ($_{10}$log of temporal multiplier, user input)
[typenode]	:	Target	Type of object

Descriptor		Value	Description
			(target/model/parameters, user input)
[namenode]	:	Nitrate leaching concentration	Name of object (user input)
[fromnode]	:	-1	Preceding object (assigned automatically for target object)
[tonode]	:	-1	Succeeding object (assigned automatically for target object)
[nodeno]	:	2	Number of object
[xcenter]	:	500	X-coordinate in diagram
[ycenter]	:	0	Y-coordinate spatial scale
[ycenterT]	:	-2	Y-coordinate temporal scale
[typenode]	:	Model	Type of object
[namenode]	:	ANIMO-model	Name of object
[fromnode]	:	3	Preceding object (user input)
[tonode]	:	0	Succeeding object (user input)
[nodeno]	:	3	Number of object
[xcenter]	:	150	X-coordinate in diagram
[ycenter]	:	5	Y-coordinate spatial scale
[ycenterT]	:	4	Y-coordinate temporal scale
[typenode]	:	Parameters	Type of object
[namenode]	:	Soil Layer Sequence from Soil Map	Name of object
[fromnode]	:	-1	Preceding object (user input)
[tonode]	:	2	Succeeding object (user input)
[nodeno]	:	4	Number of object
[xcenter]	:	150	X-coordinate in diagram
[ycenter]	:	4	Y-coordinate spatial scale
[ycenterT]	:	-2	Y-coordinate temporal scale
[typenode]	:	Parameters	Type of object
[namenode]	:	Fertilizer addition	Name of object
[fromnode]	:	-1	Preceding object (user input)
[tonode]	:	2	Succeeding object (user input)

At startup, the DSS starts with its main form, used for the definition and visualisation of research chains, which is then empty (*Figure 70*).

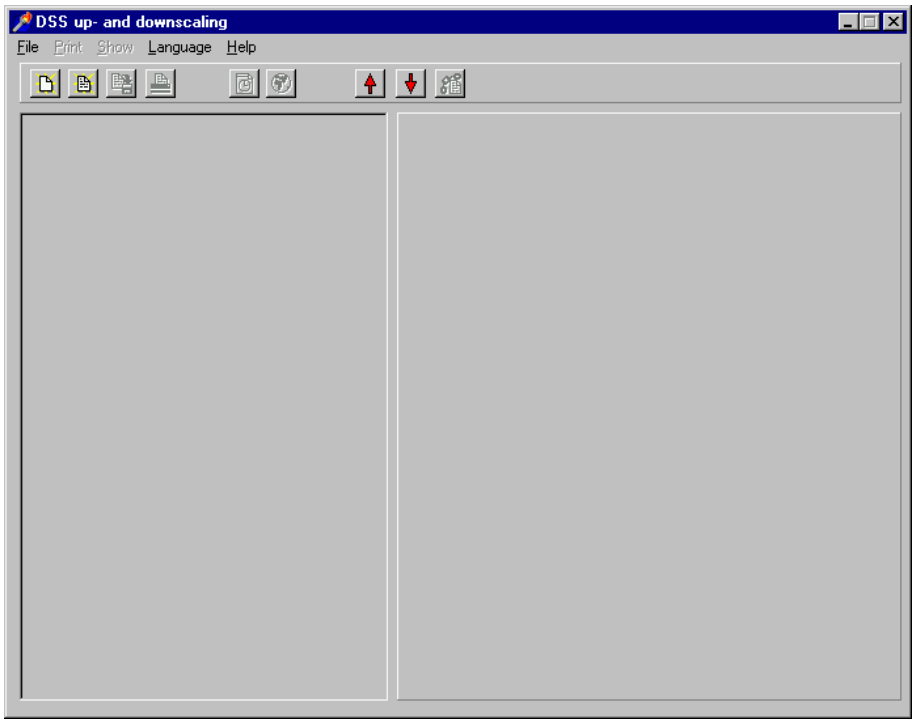

Figure 70. Startup form

There are now three options available to start a session:

1. Start from scratch by defining a new “scale project”. To do this, choose File|New from the pulldown menu or press the “New projectfile” speedbutton. The next actions to be performed are described in section 4.3.1.
2. Start by importing a previously saved scale project. In this case, choose File|Load from the pulldown menu or press the “Load projectfile”speedbutton. If parts of the research chain have to be modified, the actions in section 4.3.2 have to be performed.
3. Start by going directly to either the DSS upscaling or the DSS downscaling by pressing either the speedbutton with the upward or with the downward arrow. By using one of these buttons, the stage of defining a research chain is bypassed. This poses a risk if the user is not fully aware of what is inside and outside the scale transfer step he is evaluating. The next actions to be performed are described in section 4.5.

4.3 Definition of the research chain over the scales

4.3.1 Define a new research chain

Though it may seem unlogical, the most objective way to define a new research chain is to start at the end (the target variable) and then define the chain in a backward direction by identifying the objects (models or data) used to obtain this target variable. The approach to be taken in the DSS consists of three steps that are explained hereafter.

The first step in the definition of the research chain is the characterisation of the target variable. This involves typing a name and selecting the appropriate space (L) and time (T) support of this variable. If the target variable would be "nitrate leaching concentration", the space support might be the field scale (L-support of m^2, multiplier **10000* for fields with an area of 1 hectare), and the time support might be the year scale (T-support of *year*, multiplier **1*). Note that multipliers may be typed in directly as well. If the target variable is defined, pressing the OK button allows the user to proceed with the definition of the other objects in the research chain. An example of the result of these actions, after having pressed the "Show diagram (spatial scales)" speedbutton is given in *Figure 71*.

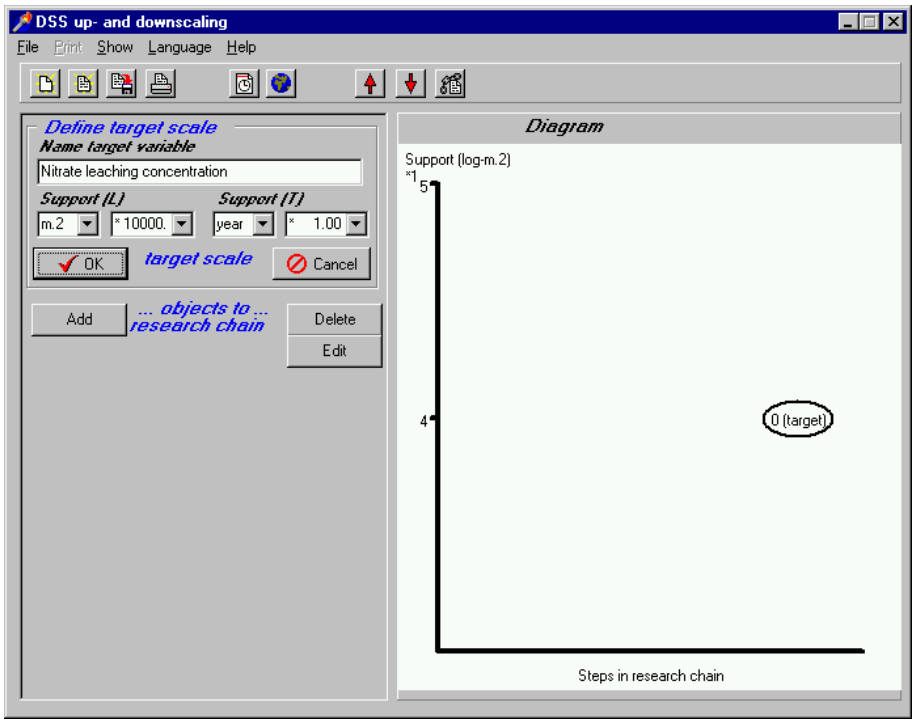

Figure 71. Example of main form after step 1 of the definition of the research chain

The second step is to press the ADD button to add a new object to the research chain. Each new object has to be given a name and its space and time support has to be set as was the case with the target variable. Note that the dimensions of the support are equal to those of the target variable, but that the multipliers can be changed. Furthermore, the type of each object has to be defined: it can either be a model or data (data stands for either variable or parameter). To define the mutual relations between the objects, the link to the preceding object and the succeeding object in the chain has to be identified. Since we work in backward direction, the succeeding object is known and can be selected from the list, but the preceding object is still unknown. Therefore we select the option "-1: Outside system" for the moment. Later, we can modify this easily. An example of the main form at this stage is given in *Figure 72*.

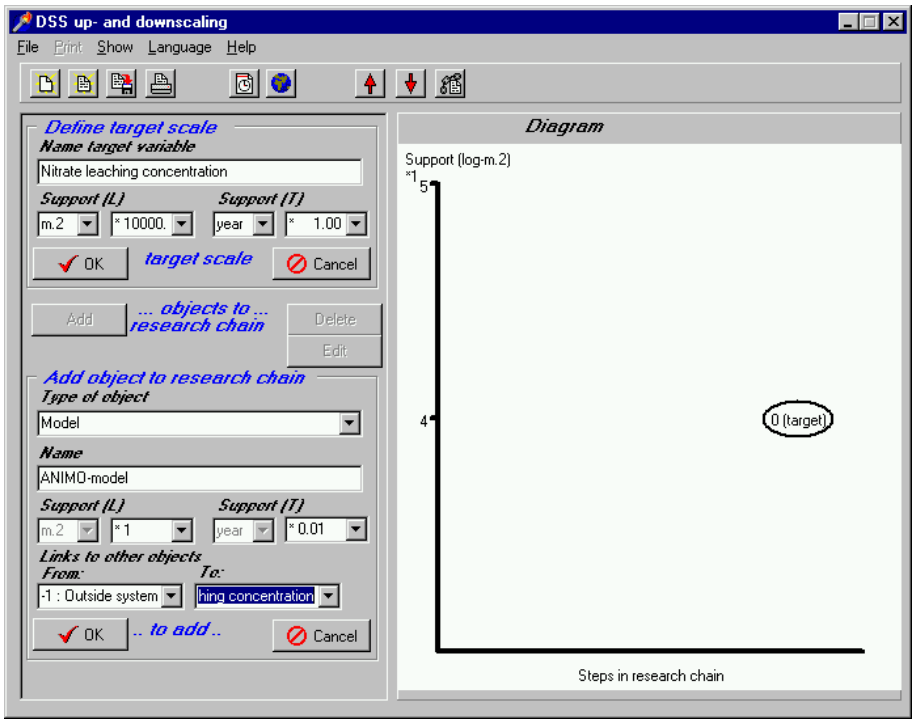

Figure 72. Example of main form while adding a model object to the research chain

If the object is defined in a satisfactory way, pressing the OK-button at the lower part of the screen brings us back to the beginning of the second step. The stage of definition of the research chain can now be shown again by pressing the speedbuttons. The second step can thereafter be repeated as many times as needed to define a research chain. Errors can be corrected by pressing the CANCEL button (after which the object has to be redefined entirely). An entire object can be removed from the research chain by pressing the DELETE button and selecting the object to be deleted by its type, spatial and temporal multipliers and its "from" and "to" objects.

The third step is to reconnect each of the objects to one of their predecessors if appropriate. For instance, if a model has been defined and it is using two kinds of data that are separately defined, the "from" object should be selected from one of the two data objects. If an object has no predecessor in the research chain (e.g. in case the object contains "basic data"), the "-1: Outside system" option can remain selected. An object can be reconnected to another object by pressing the EDIT button and subsequently changing the selection in the "from" field. Note that any of the

object properties can be changed at this time, beside the space and time dimensions.

The current stage of definition of a research stage can be visualised after pressing the OK-button and either of the two “Show diagram” speedbuttons. An example is given in *Figure 73*.

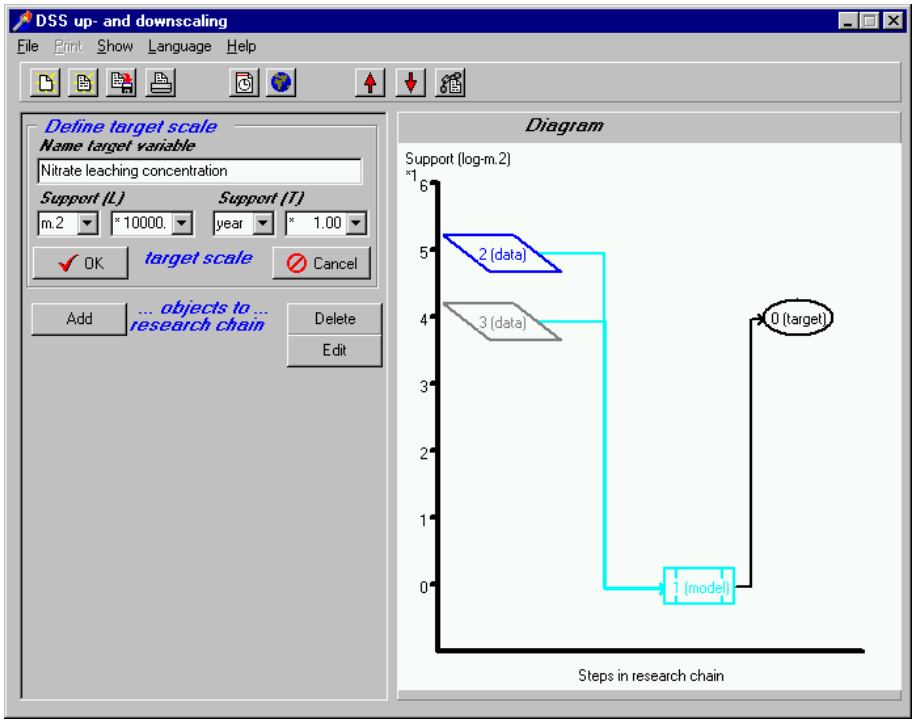

Figure 73. Example of a research chain with 2 data objects, 1 model object and a target variable

4.3.2 Modify parts of research chain

By pressing the EDIT button any object in the research chain can be modified. The first action to be taken is to select the object to be edited (*Figure 74*) and to press the OK button.

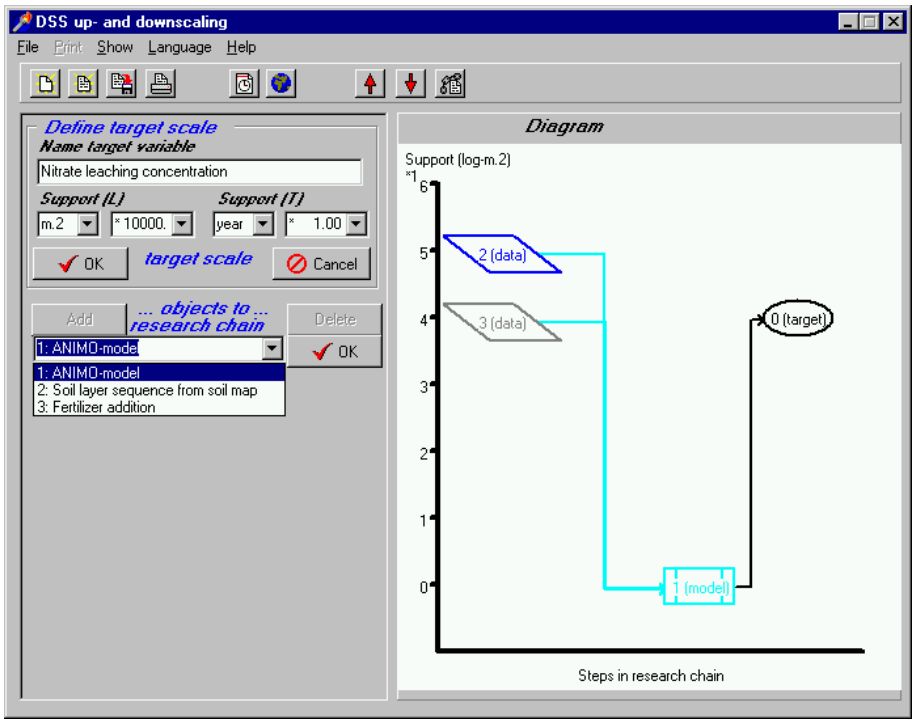

Figure 74. Example of selecting an object from the research chain to be edited

Next, the fields that characterise the object appear, and the object can be redefined easily (*Figure 75*). The editing of an object ends when the OK button at the lower part of the form is pressed (modifications are completed), or when the CANCEL button is pressed (no modifications are brought about).

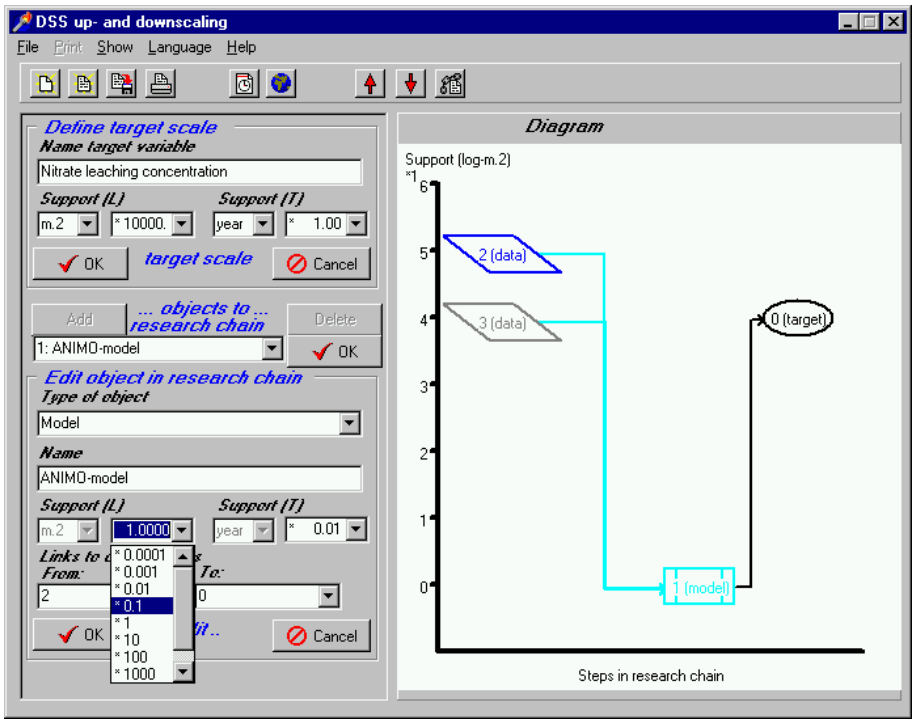

Figure 75. Example of editing an object from the research chain

4.4 Enter the DSS from the research chain

When the research chain is defined properly, the speedbutton “To DSS from research chain” can be pressed. A small form appears, on which the starting point of a section of the research chain can be selected from the available objects. Next, a succeeding object can be chosen from the list of available succeeding objects. Furthermore, choice can be made between the spatial and the temporal domain (*Figure 76*). Note, that the DSS considers any scaling step exclusive the starting (preceding) object, and inclusive the ending (succeeding) object.

Figure 76. Selecting a scaling step in the research chain

When the SHOW DIAGRAM button is pressed, the selected scaling step is shown with a dotted line for the selected space or time domain in the research chain diagram. Pressing the OK button results in the identification by the DSS whether upscaling or downscaling is the issue, and subsequently brings up the upscaling or downscaling form (section 4.5).

4.5 DSS Upscaling and Downscaling

The forms of the DSS upscaling and downscaling are similar, except for the decision trees shown. The diagram shows the current part of the decision tree. The position of the "active question" in the decision tree is always shown in yellow, an explanatory text with the active question is always added in a yellow-coloured memobox. The part of the decision tree that has been navigated through is indicated in green.

Possible actions at any time are to restart the series of questions, and to go back to the main form. Also, the decision tree can be saved to a file by right-clicking on the diagram.

For the possible answers to any active question, buttons for YES and NO are available next to the memobox. Finally, if a class or a subclass of upscaling or downscaling has been chosen by the DSS, a button FURTHER READING appears that allows the user to read the corresponding part of this book. If a main class of upscaling or downscaling has been selected by the DSS, a button appears with which the user can enter the decision tree towards a subclass.

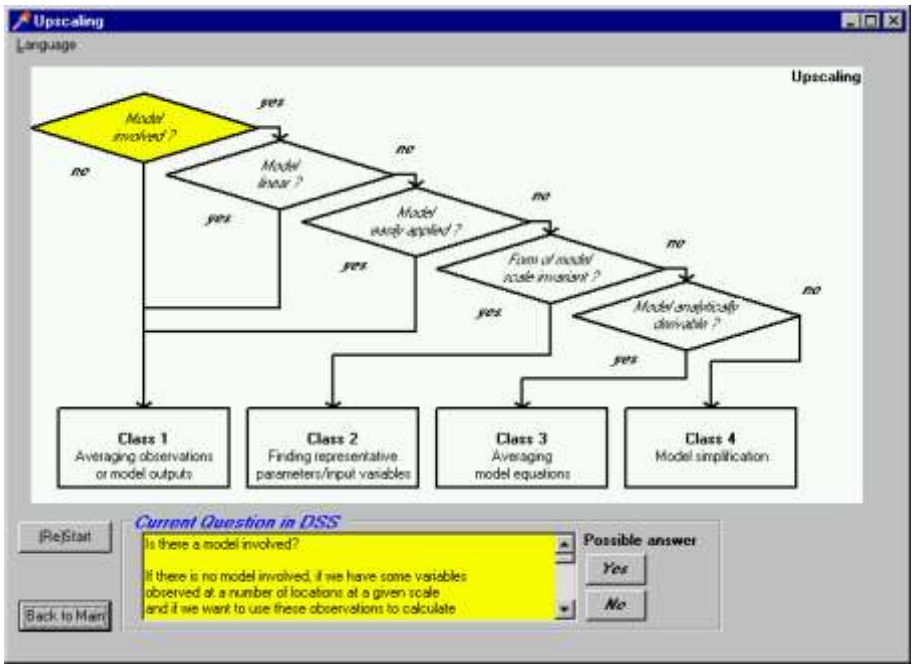

Figure 77. DSS upscaling at startup

There is one difference between launching the DSS directly from the main form or indirectly from a selected step in the research chain. In the indirect case with upscaling, the DSS knows whether a model is involved, and starts off with the second question or with a classification into a main class of upscaling.

Appendix

Random Variables and Stochastic Functions

Next follows an exposé about random variables and stochastic functions. Readers not familiar with these subjects are recommended to read this appendix, as this will be useful in understanding how stochastic functions can be used in upscaling and downscaling and in appreciating the examples given in subsections 2.3.3, 2.3.4 and 2.4.2 and chapter 3.

Random variables

Before we explain the concept of a stochastic function, we start with explaining the concepts and properties of a random (=stochastic) variable. A random variable is a variable that can have a set of different values generated by some probabilistic mechanism. We do not know the value of a stochastic variable, but we do know the probability with which a certain value can occur. For instance, the outcome of throwing a die is not known beforehand. We do however know the probability that the outcome is 3. This probability is 1/6 (if the die is not tampered with). So the outcome of throwing a dice is a random variable. The same goes for the outcome of throwing two dice. The probability of the outcome being 3 is now 1/18. A random variable is usually written as a capital (e.g. D for the unknown outcome of throwing two dice) and an actual outcome (after the dice have been thrown) with a lower case (e.g. d). The actual outcome is also called a "realisation" of the random variable.

The "expected value" or "mean" of a random variable can be calculated if we know which values the random variable can take and with which probability. If D is the result of throwing two dice, the probability distribution $\Pr(d)$ is given in the following table:

Table A1 Probabilities of outcomes of throwing two dice

D	2	3	4	5	6	7	8	9	10	11	12
Pr(d)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

The mean or expected value is calculated as (N_d the number of possible outcomes and d_i outcome i):

$$E[D] = \sum_{i=1}^{N_d} d_i \Pr[d_i] = 2 \cdot 1/36 + 3 \cdot 2/36 + ... + 12 \cdot 1/36 = 7 \qquad \text{(A1)}$$

That the expected value equals 7 means that if we were to throw the two dice a very large number of times and calculate the average outcome of all these throws we would end up with a number very close to 7. This means that we could take a sample of n outcomes d_j of a random variable D and estimate its mean as:

$$\hat{E}[D] = \frac{1}{n} \sum_{j=1}^{n} d_j \qquad \text{(A2)}$$

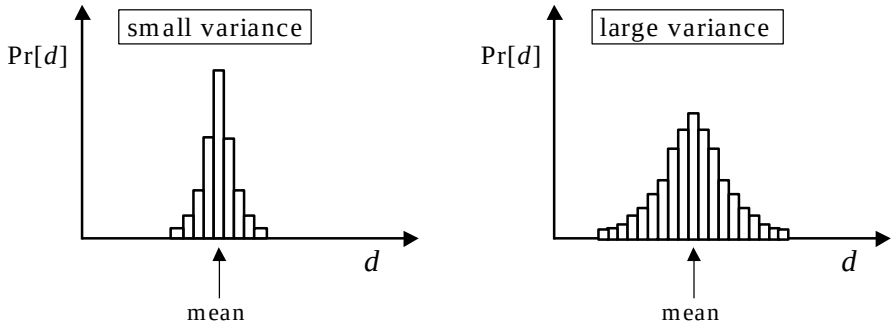

Figure A1 Two probability distributions with the same mean but with different variances

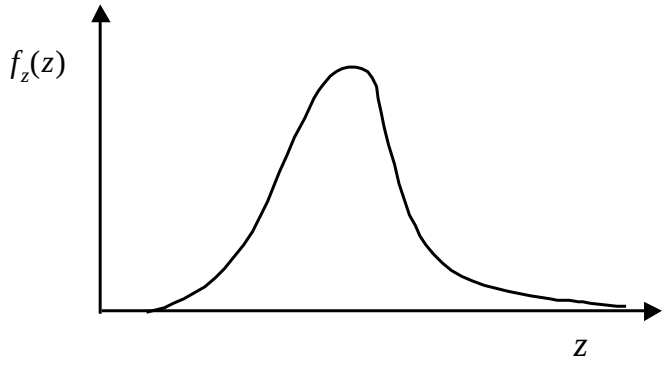

Figure A2. A probability density function

The mean is the centre of mass of the probability distribution and tells us what would be the average of generating many outcomes (see Figure A1). The variance is a measure of spread. It tells us something about the width of the probability distribution. In other words, it tells us how different the various generated outcomes (throws of the dice) are. A larger variance means that the probability distribution is wide, the variation among outcomes is large and therefore we are more uncertain about the outcome of the stochastic variable. Figure A1 shows two probability distributions with the same mean, but with different variances. The variance of a random variable is calculated from the probability distribution as:

$$\begin{aligned} VAR[D] &= E[(D - E[D])^2] = \sum_{i=1}^{N_d} (d_i - E[D])^2 \Pr[d_i] \\ &= (2-7)^2 \cdot 1/36 + (3-7)^2 \cdot 2/36 + ... + (12-7)^2 \cdot 1/36 \\ &= 5.8333 \end{aligned} \qquad \text{(A3)}$$

The variance can be estimated from a sample of n outcomes (n throws of two dice) d_j as

$$\hat{VAR}[D] = \frac{1}{n-1} \sum_{j=1}^{n} (d_j - \hat{E}[D])^2 \qquad \text{(A4)}$$

Instead of the variance, one often uses its square root as a measure of spread. This square root is called the standard deviation. Greek symbols used for the mean, variance and standard deviation are μ, σ^2 and σ respectively.

The concept of a random variable is used to express uncertainty. If we are uncertain about the actual value of some property (e.g. the concentration of a pollutant or the number of individuals in a population), this property is "modelled" as a random variable. The more uncertain we are about the actual but unknown value, the larger the variance of the probability distribution of this random variable.

The outcome of throwing dice is a discrete property. It can only take a limited number of countable values. If the property is continuous it can take any real value between certain bounds (e.g. altitude, hydraulic conductivity, concentration). To describe the probability of a certain outcome of real valued random variable *Z*, instead of a (discrete) probability distribution, a continuous function called the probability density function $f_Z(z)$ is used (see Figure A2). The probability density itself provides not a probability. For instance, we *cannot* say that $\Pr[Z = z_1] = f_Z(z_1)$. Instead, the probability density gives the probability mass per unit *z*. So, the probability density function *can* be used to calculate the probability that *Z* is between certain boundaries as follows:

$$\Pr[z_1 < Z \le z_2] = \int_{z_1}^{z_2} f_Z(z)dz \tag{A5}$$

The probability that *Z* is smaller than or equal to a certain value *z* is given by the cumulative probability distribution function (cpdf) $F_Z(z)$, which is related to the probability density as:

$$F_Z(z) = \Pr[Z \le z] = \int_{-\infty}^{z} f_Z(z')dz' \tag{A6}$$

In pollution problems one is often interested in calculating the probability that a certain critical threshold z_c is exceeded. This can be calculated from the pdf or cpdf as follows:

$$\Pr[Z > z_c] = \int_{z_c}^{\infty} f_Z(z)dz = 1 - F_Z(z_c) \tag{A7}$$

Finally, the area under the pdf is equal to one:

$$\int_{-\infty}^{\infty} f_Z(z)dz = 1 \tag{A8}$$

The mean or expected value of a continuous random variable Z is given by an integral instead of a sum:

$$\mu_Z = \mathrm{E}[Z] = \int_{-\infty}^{\infty} z\, f_Z(z)dz \tag{A9}$$

and the variance is defined as

$$\sigma_Z^2 = E[(Z-\mu)^2] = \int_{-\infty}^{\infty} (z-\mu)^2\, f_Z(z)dz \tag{A10}$$

The estimators of the mean and the variance are the same as in Equations (A2) and (A4) with d_j replaced with z_j.

As stated above, a random variable Z (and its pdf) is a model to express our uncertainty about the real but unknown value of some property z. One particular pdf that is often used to model uncertainty is the Gaussian or normal distribution. A Gaussian random variable has the following pdf:

$$f_Z(z) = \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{\frac{-(z-\mu_z)^2}{2\sigma_z^2}} \tag{A11}$$

The Gaussian pdf is often used because many phenomena that are the result of random processes can be described by a Gaussian distribution. Also, the Gaussian pdf is fully described with only two parameters, the mean μ_Z and the variance σ_Z^2. Moreover, the "central limit theorem" states that a linear combination (e.g. a sum or an average) of many identically distributed random variables is approximately Gaussian distributed. The Gaussian random variable χ with mean $\mu_Z = 0$ and the variance $\sigma_Z^2 = 1$ is called a standard normal deviate. Any Gaussian stochastic variable Z with mean μ_Z and variance σ_Z^2 can be obtained from χ as

$$Z = \mu_Z + \sigma_Z \chi \tag{A12}$$

Finally, a useful property is that any linear combination (e.g. sum, average, linear regression) of Gaussian stochastic variables is also Gaussian distributed.

The following rules apply to the mean and the variance of a stochastic variable Z (a and b are deterministic constants):

$$E[a+bZ] = a + bE[Z] \tag{A13}$$

$$VAR[a+bZ] = b^2 VAR[Z] \tag{A14}$$

In case we are studying two random variables, it is important to know whether these are statistically dependent. For instance, we may be uncertain about the actual values of the concentration of lead and zinc in a soil column, but we may know that higher lead concentrations generally go together with higher zinc concentrations. A measure of (linear) statistical dependence between random variables Y and Z is the covariance:

$$COV[Y,Z] = E[(Y-\mu_Y)(Z-\mu_Z)] \tag{A15}$$

The following relationship exists between variance and covariance (a and b constants):

$$VAR[aY+bZ] = a^2 VAR[Y] + b^2 VAR[Z] + 2abCOV[Y,Z] \tag{A16}$$

Often the correlation coefficient is used as a measure of linear statistical dependence:

$$\rho_{YZ} = \frac{COV[Y,Z]}{\sigma_Y \sigma_Z} \tag{A17}$$

Figure A3 shows scatter plots of a limited number of outcomes of random variables under different levels of statistical dependence. If $\rho_{YZ} > 0$, Y and Z are positively correlated, if $\rho_{YZ} < 0$ they are negatively correlated (large values of y go together with small values of z and vice versa). Values of $\rho_{YZ} = 1$ and $\rho_{YZ} = -1$ stand for perfect positive and negative correlation respectively. If $\rho_{YZ} = 0$ Y and Z are uncorrelated.

Two random variables Y and Z are said to be statistically independent if their joint probability density function $f_{YZ}(y,z)$ (giving the probability that

both *Y* and Z have values between certain boundaries) is equal to the product of the individual probability densities of *Y* and *Z*:

$$f_{YZ}(y,z) = f_Y(y)\, f_Z(z)$$

If two random variables are statistically independent they are also uncorrelated: COV[*Y*,*Z*]=0 and ρ_{YZ} = 0. However a zero correlation coefficient does not necessarily mean that *Y* and *Z* are statistically independent. The covariance and correlation coefficient only measure *linear* statistical dependence. If a non-linear relation exists, the correlation may be 0 but the two variables may still be statistically dependent, as is shown in the lower right figure of Figure A3.

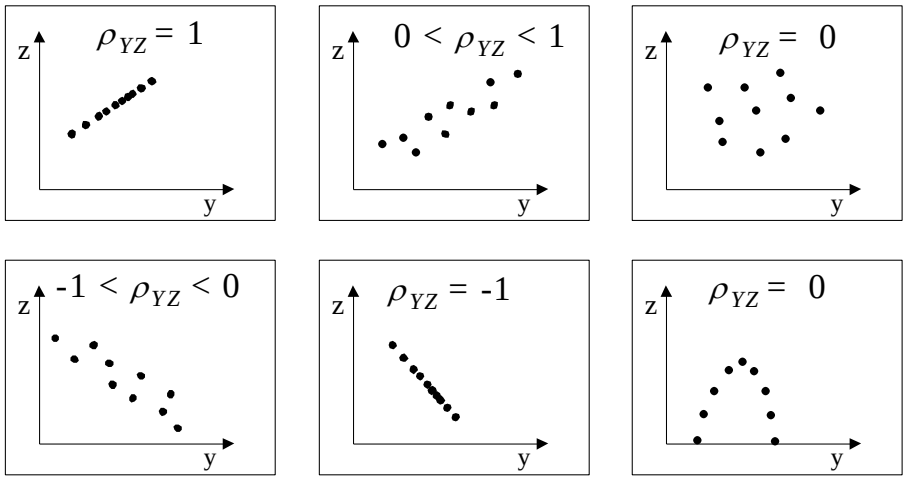

Figure A3. Scatter plots of outcomes of two random variables under various forms of statistical dependence

Stochastic functions

Figure A4 shows schematically the concept of a stochastic function in the temporal domain. Instead of finding one function that describes the unknown variation at scale s_1 within s_2, a family of functions is chosen, each of which is assumed to have an equal probability of representing the true but unknown variation at scale s_1 within s_2. The family of equally probably functions is called the "ensemble", or alternatively a "stochastic function" (other names: stochastic process, random function, random process and, in case the domain is two dimensions or more, a random field). As with a random variable, a

stochastic function (i.e. ensemble) is usually denoted with a capital, e.g. $Z(s_1;t)$[15] for the continuum approach in the time domain. One particular function out of the many is called a "realisation" and is denoted with the lower case $z(s_1;t)$. Isaaks and Srivastava (1989) provide a very basic and excellent introduction to stochastic functions. Standard textbooks on the subject are for instance Papoulis (1991) and Vanmarcke (1983).

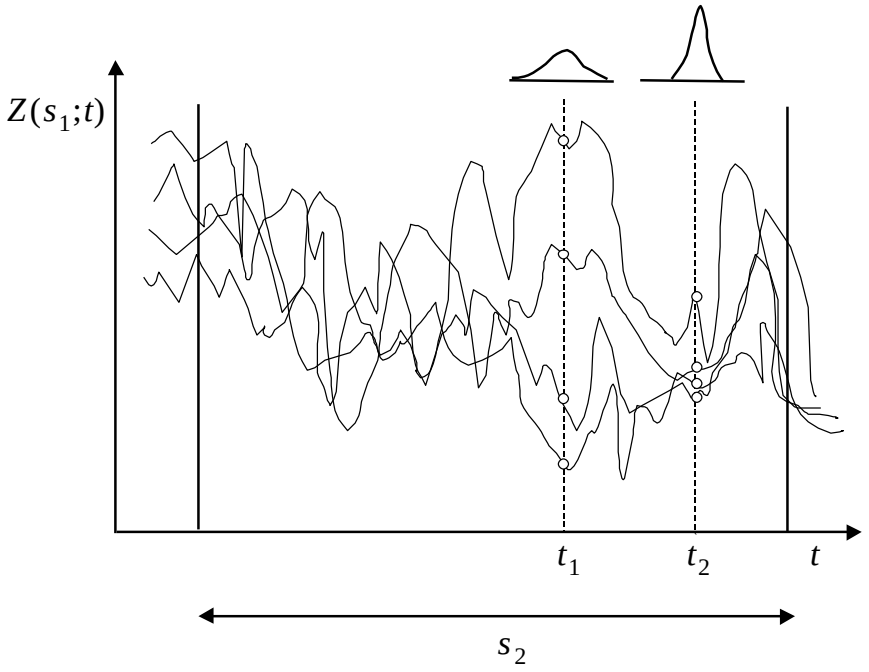

Figure A4. Schematic representation of a stochastic function used to describe unknown variation of $z(s_2;t)$ within s_2

Another way of viewing a stochastic function is as a collection of random variables (one at every location in space or point in time) that are all statistically dependent of each other. In Figure A4: at every point in time t a random variable $Z(s_1;t)$ with support s_1 (continuum approach[16]) is defined. At each t the random variable $Z(s_1;t)$ is described with a probability density

[15] In order to explain the concept of a stochastic function, the scale variable in $Z(s_1;t)$ is not really necessary. In most textbooks it is omitted, i.e. $Z(t)$ is used and some averaging scale is implicitly assumed. We denote the scale variable here to be consistent with the rest of the book. The drawback is that the notation becomes slightly more elaborate.

[16] In the discrete approach we have N_{12} random mutually dependent random variables $z(s_1;i)$ (one for each support unit at scale s_1) within the support unit s_2.

function (pdf) $f(s_1;z;t)$ which not only depends on the value of z and the support size s_1, but also on the value of t. This pdf could be estimated by sampling all realisations at a certain point (say t_1) in time or location and calculating the histogram of the samples. In the example of Figure A4 the variation among the realisations is larger at point t_1 than at point t_2, leading to a probability density function at t_1 with a larger variance than at t_2. So, we are less certain about the unknown value at t_1 than we are at t_2.

At each point in time (or location) we can calculate the time- (or location-) dependent mean and variance:

$$\mu(s_1;t) = E[Z(s_1;t)] = \int_{-\infty}^{\infty} z\, f(s_1;z;t)dz \qquad \text{(A18)}$$

$$\sigma^2(s_1;t) = E[\{Z(s_1;t) - \mu(s_1;t)\}^2] = \int_{-\infty}^{\infty} \{z - \mu(s_1;t)\}^2\, f(s_1;z;t)dz \qquad \text{(A19)}$$

Also, the random variables are usually dependent in time (or in space for spatial stochastic functions):

$$COV[Z(s_1;t_1), Z(s_1;t_2)] \neq 0 \qquad \text{if } t_1 \neq t_2$$

The covariance is generally smaller when random variables are considered at locations further apart. For $t_1 = t_2$ the covariance equals the variance (A19).

For a set of N discrete random variables the joint or multivariate probability distribution $P[d_1,d_2,\ldots,d_N]$ describes the probability that N random variables have a certain value, i.e. that $D_1 = d_1$ and $D_2 = d_2$ and ….and $D_N = d_N$. Similarly, for N continuous random variables the multivariate probability density $f(z_1,z_2,\ldots. z_N)$ is the probability that N random Z_i, i=1,..,N having values between certain boundaries. The analogue for a stochastic function is the probability that a realisation of $Z(s_1;t)$ at a set of N locations t_i, i = 1,..,N has the values $z(s_1;t_1) = z_1$, $z(s_1;t_2) = z_2$, …., $z(s_1;t_N) = z_N$ between certain boundaries. The associated multivariate probability density function (pdf) is denoted as $f(s_1; z_1,z_2,.., z_N ; t_1,t_2,\ldots. t_N)$. Because it relates to random variables at different points or locations, the multivariate pdf of a stochastic function is sometimes referred to as "multipoint" pdf. Theoretically, a stochastic function is fully characterised (we know all there is to know about it) if the multivariate probability distribution for any finite set of points is known.

A frequently used model for the multivariate pdf is the multivariate Gaussian pdf. Consider a vector $\mathbf{Z} = [Z(s_1;t_1), Z(s_1;t_2),.., Z(s_1;t_N)]^T$ of random variables obtained by considering the stochastic function $Z(s_1;t)$ at N points in time (or at N locations). We define $\mathbf{m_Z} = [\mu(s_1;t_1), \mu(s_1;t_2),.., \mu(s_1;t_N)]^T$ as the vector of mean values and $\mathbf{C_{ZZ}}$ as the $N \times N$ covariance matrix, where element i,j is formed by the covariance $COV[Z(s_1;t_i), Z(s_1;t_j)]$. If $\mathbf{Z}$ is multivariate Gaussian distributed, its joint or multivariate pdf is given by

$$f_{\mathbf{Z}}(\mathbf{z}) = (2\pi)^{-N/2} \mid \mathbf{C_{ZZ}} \mid^{-1/2} \exp[-\frac{1}{2}(\mathbf{z}-\mathbf{m_Z})^{\mathrm{T}} \mathbf{C}_{\mathbf{ZZ}}^{-1}(\mathbf{z}-\mathbf{m_Z})] \quad \text{(A20)}$$

where $|\mathbf{C_{ZZ}}|$ is the determinant of the covariance matrix. If the multivariate distribution of a stochastic function obeys Equation (A20) it is said to be multivariate Gaussian. Just as the Gaussian distribution (A11) is fully characterised by the mean and the variance, so is the multivariate Gaussian pdf of a given set of points fully characterised by the mean vector and the covariance matrix.

Strict stationarity, isotropy, ergodicity

A stochastic function is called "strict stationary" if its multivariate pdf is invariant under translation. So for any set of N locations and for any translation t' we have that

$$\begin{aligned} f(s_1;\, z_1, z_2, \ldots,\, z_N \,;\, t_1, t_2, \ldots .\, t_N) = \\ f(s_1;\, z_1, z_2, \ldots, z_N \,;\, t_1 + t', t_2 + t', \ldots,\, t_N + t') \end{aligned} \quad \text{(A21)}$$

So we can have any configuration of points on the time axis and move this configuration (the whole configuration, not one point at the time) of points forward and backwards in time and have the same multivariate pdf. For the spatial domain we have to stress that strict stationarity means an invariant pdf under translation only, not rotation. So for a strict stationary stochastic function in two dimensions, the two sets of locations in the left figure of A5 have the same multivariate pdf, but those in the right figure not necessarily so. A stochastic function whose multivariate pdf is invariant under rotation is called a "statistically isotropic" stochastic function.

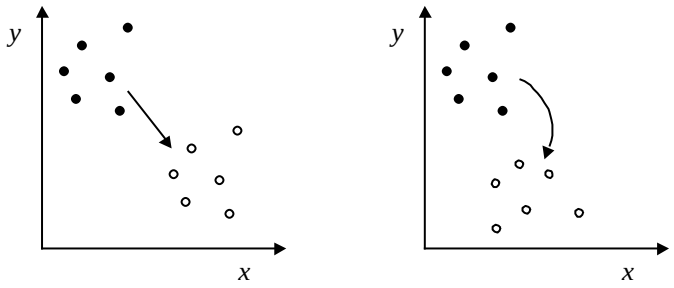

Figure A5. Translation of a configuration of points without rotation (left figure) and with rotation (right figure)

One could ask why the property of stationarity is so important. The reason lies in determining the statistical properties of a stochastic function such as the mean, the variance and the covariance. In case of a random variable, such as the outcome of throwing dice, we can do a series of random experiments (actually throwing the dice) and estimate the mean and variance from the results of these experiments using Equations (A2) and (A4). Not so with a stochastic function. To estimate the statistical properties of a stochastic function, we should be able to draw a large number of realisations as shown in Figure A4. However, in practice, we only have one realisation of the stochastic function, namely reality itself. So we must be able to estimate all the relevant statistics of the stochastic function from a single realisation. It turns out that this is possible for strict stationary stochastic functions. The reason is that strict stationarity actually says that all statistical properties are the same, no matter where you are. For instance, suppose we want to estimate the mean $\mu(s_1;t_1)$ at a certain point t_1. The normal procedure would be to take the average of many realisations at point t_1, which is impossible because we only have one realisation (reality). However, if the stochastic function is stationary the pdf $f(s_1;z;t)$ at any location is the same and therefore also the mean. This also means that *within any single* realisation we have at every location a sample from the *same* pdf $f(s_1;z;t)$. So, the mean can also be estimated if we take a sufficient number of samples from a single realisation, such as our reality. This is illustrated in Figure A6. This property of a stochastic function, i.e. being able to estimate statistical properties of a stochastic function from a large number of samples of a single realisation is called "ergodicity". Apart from the stochastic field being strict stationary, there is another condition necessary for ergodicity to apply. The samples from the single realisation should be taken from a large enough period of time or, in the spatial case, large enough area. How large this period of time or area must be will be explained hereafter.

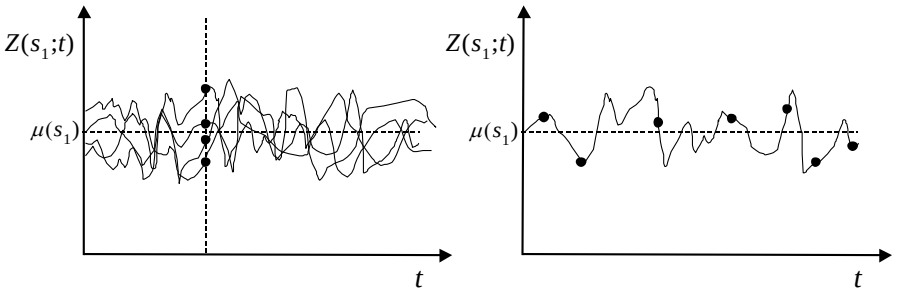

Figure A6. Ergodicity: the average of (observations from) a single realisation is the same as the average of many realisations at a given location

Second order stationarity, covariance function, semivariogram

The number of observations available is usually only sufficient to estimate the mean, variance and covariances of the stochastic function. So in practice, we require only ergodicity and therefore only stationarity for the mean, variance and covariances. Hence, a milder form of stationarity is usually assumed which is called "second order stationarity". For second order stationary stochastic functions the mean and variance do not depend on t (or $\mathbf{x}$) ($\mu(s_1;t) = \mu(s_1)$ and $\sigma^2(s_1;t) = \sigma^2(s_1)$) and the covariance depends only on the separation distance between two points in time (or space): $\mathrm{COV}[Z(s_1;t_1),Z(s_1;t_2)] = C(s_1;t_2-t_1)$. The graph describing the covariance as a function of the separation distance t_2-t_1 (also called "lag") is called the covariance function. In case second order stationarity, the covariance function for $t_2=t_1$ (or $\mathbf{x}_2=\mathbf{x}_1$) is equal to the variance and decreases to zero when the distances $|t_2-t_1|$ (or $|\mathbf{x}_2-\mathbf{x}_1|$) becomes larger. This means that random variables at sufficiently large distances are not correlated. Dividing the covariance function by the variance yield the correlation function:

$$\rho(s_1;t_2-t_1) = \frac{C(s_1;t_2-t_1)}{\sigma^2(s_1)} \qquad \text{(A21)}$$

and by calculating the area under the correlation function we obtain the integral scale:

$$I(s_1) = \int_0^{\infty} \rho(s_1;\tau)d\tau \qquad \text{(A22)}$$

The integral scale (dimensions [T] or [L]) is a measure for the time span or distance over which the values of the stochastic function remain correlated. For the integral scale to be finite, the stochastic function must be second order stationary and the covariance function most go to zero fast enough for large lags.

A strict stationary stochastic function is also second order stationary, but not necessarily the other way around. However, if a stochastic function is second order stationary and multivariate Gaussian (Equation A20), it is also a strict stationary. More importantly, a second order stationary stochastic function that is multivariate Gaussian (and thus also strict stationary) is completely characterised by only a few statistics: the mean $\mu(s_1;t) = \mu(s_1)$ and the covariance function $C(s_1;t_2\text{-}t_1)$; in Equation (A20), for any set of points the constant mean $\mu(s_1)$ is used for all elements of the mean vector $\mathbf{m_Z}$, whereas the covariance matrix $\mathbf{C_{ZZ}}$ can be build from the covariances calculated with $C(s_1;t_2\text{-}t_1)$. This is the reason why the multivariate Gaussian model is so popular in many stochastic analyses involving stochastic functions. The reader should however be aware that the multivariate Gaussian model is very restrictive and therefore a poor model for certain natural phenomena. For instance, hydraulic conductivity often shows long and narrow connected zones of high values (preferential flow paths) that cannot be described with the multivariate Gaussian model (see Gómez-Hernández and Wen, 1997). Rainfall fields show higher peaks than can be explained by a Gaussian distribution. In these cases one either has to resort to non-linear transformation of the multivariate Gaussian stochastic function or use other stochastic functions such as indicator stochastic functions (Journel, 1983) or stochastic functions generated by point processes (e.g. Rodriguez-Iturbe, 1986b).

Another measure of spatial or temporal correlation that is often used (especially in the spatial domain) is called the semivariance (see also 2.2.3). The semivariance is defined as:

$$\gamma(s_1;t_1,t_2) = \tfrac{1}{2} E[\{Z(s_1;t_2) - Z(s_1;t_1)\}^2] \qquad \text{(A23)}$$

In case of a second order stationary stochastic function the semivariance depends only on separation distance $t_2\text{-}t_1$ (also called "lag"). The function relating the semivariance to the lag is called the "semivariogram" and can be estimated from observations as (similarly in the spatial domain):

$$\hat{\gamma}(s_1;h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} \{z(s_1;t_i) - z(s_1;t_i + h)\}^2 \qquad \text{(A24)}$$

where $h=t_2-t_1$ is the separation distance between two locations and $n(h)$ is the number of observations a distance h apart. The semivariogram and the covariance function of a second order stationary stochastic function are related as follows (similar relation in the spatial domain):

$$\gamma(s_1;t_2 - t_1) = \sigma^2(s_1) - C(s_1;t_2 - t_1) \qquad \text{(A25)}$$

This also means that where the covariance function becomes zero for large enough separation distances, the semivariogram will reach a plateau (called the "sill" of the semivariogram) that is equal to the variance. The distance at which this occurs (called the "range" of the semivariogram) is the distance beyond which values on the stochastic function are uncorrelated. So, next to the integral scale, the semivariogram range can be used a measure of temporal or spatial correlation. As stated before, apart from assuming stationarity, samples should be taken from a large enough period of time or large enough area, if ergodicity is to apply. In practice, this means that observations should be taken from a period of time (or from an area) the extent of which is a few times larger than the semivariogram range.

In the spatial domain, the range does not have to be the same in all directions. After all, a property can be less variable (values are more alike) in one direction than in another. A stochastic function with different ranges in different directions is "statistically anisotropic".

Figure A7 gives an example of a covariance function and a semivariogram of a second order stationary stochastic function. The figure also shows that both functions make a little jump close to lag zero. At lag zero the semivariance is zero and the covariance function is equal to the variance. At an infinitesimally small lag the functions make the jump. The magnitude of this jump is called the "nugget" variance and represents the combined effect of short distance variation (i.e. between locations that are only a very short distance apart) and observation errors.

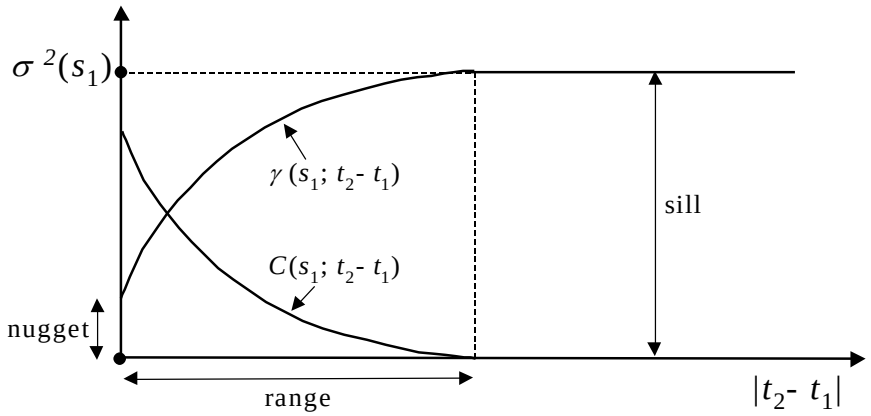

Figure A7. Covariance function and semivariogram for a second order stationary stochastic function

Stochastic simulation

As stated above, if certain assumptions about its behaviour are made (e.g. second order stationarity and a Gaussian multivariate pdf) the stochastic function can be parameterised with only a limited number of statistical measures, such as the mean, and the semivariogram (or the covariance function). Assuming ergodicity, these statistics can be estimated from the observations at scale s_1. With the estimated statistics as input certain algorithms can be used to randomly generate a large number of realisations from the stochastic function. The generation of realisations from a stochastic function is called "stochastic simulation" or (in the spatial domain) "geostatistical simulation".

On may choose to generate from the set of all possible realisations *without* looking at the actual observed values. This is called "unconditional simulation". It is also possible to generate only those realisations that pass through the observations (as in Figure A8). This is called "conditional simulation". If observations at scale s_1 are present, the latter method is preferred, because it yields representations of the unknown variation at scale s_1 that agree not only with the statistics of the observations, but also with the observed values themselves.

Conditional stochastic simulation can be used for uncertainty analysis by means of the "Monte Carlo method" as follows. Consider an air pollution model calculating the concentration plume of some pollutant at some spatial grid. One of the inputs of such a model would be surface roughness that

must be known at each grid location. Suppose that the surface roughness has been determined at a limited number of locations only, which means that it must be estimated at most of the grid locations. The uncertainty about the true concentration plume caused by this lack of knowledge can be analysed with stochastic simulation. A large number of realisations of surface roughness are conditionally simulated (conditional to the roughness observations). The necessary statistics (e.g. mean and semivariogram) can be estimated from the observations as well. By running the air pollution model with each of the simulated realisations a large number of model outcomes (concentration plumes) are obtained. The probability distribution estimated from these model outcomes reflects the uncertainty about the true but unknown concentration plume caused by the lack of information about the surface roughness. The more observations, the more alike the realisations are, because each realisation must pass through all observations. This reduces the variation among the calculated pollution plumes and therefore the uncertainty about the true but unknown plume.

To simulate realisations of stochastic functions in time we can use time series models (Box and Jenkins, 1976) for discrete time steps and solutions to stochastic differential equations (e.g. Unny, 1984; Newton, 1991) in the continuous time domain. Quite a lot of different algorithms have been developed for simulating realisations of multivariate Gaussian stochastic functions in the spatial domain. The most straightforward of these algorithms is sequential Gaussian simulation, which can be used for both conditional and unconditional simulation. Again, one should be aware that many properties are poorly described by multivariate Gaussian stochastic functions. Other types of simulation algorithms that are able to simulate unconditional and conditional realisations of non-Gaussian stochastic functions are sequential indicator simulation and simulated annealing. We refer to Goovaerts (1997) and Deutsch and Journel (1998) for theoretical and practical descriptions of the spatial simulation methods mentioned here.

Conditional probability and kriging

In the same way as shown in Figure A4, one could estimate the probability distributions at each point (or location) by sampling at this point (or location) the values of all conditional realisations (Figure A8). These probability distributions are called "conditional" or "posterior" probability distributions because they give the probability of $Z(s_1;t)$ *given* the n observations $z_i(s_1)$, $i= 1,..,n$. The conditional pdf of $Z(s_1;t)$ is denoted as $f_Z(s_1;z,t \mid z_i(s_1), i= 1,..,n)$. The pdf $f_Z(s_1;z,t)$ of $Z(s_1;t)$ without conditioning to observations is called the "marginal" or "prior" pdf. Figure A8 shows two conditional pdfs at points t_1 and t_2. It can be seen that uncertainty is larger further from an observation (t_1) than close to an observation (t_2), which is

intuitively correct. This is the case, provided the stochastic function is second order stationary, such that the a priori distribution $f_Z(s_1;z,t)$ is the same for all t.

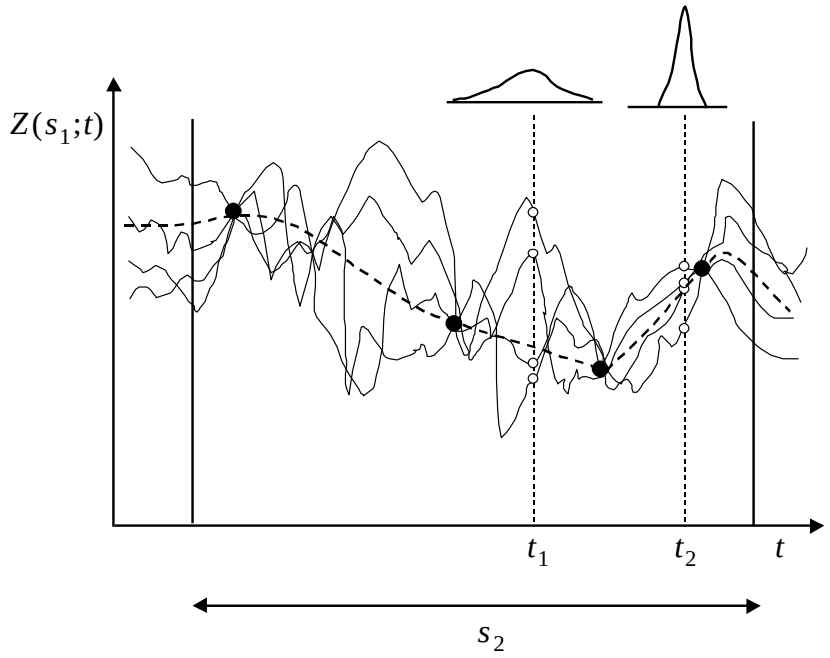

Figure A8. Realisations of a stochastic function that is conditional to a number of observations; dashed line is the conditional mean that can be obtained directly through simple kriging if $Z(s_2;t)$ is multivariate Gaussian and stationary

To obtain the conditional distribution of $Z(s_1;t)$, or its mean value and variance, it is not necessary to perform conditional simulation and estimate it from the realisations. There are geostatistical techniques called kriging (see also section 2.2.3) that can be applied to estimate for every point or location the conditional distribution (or its mean and variance) directly, using the observations and the statistics of the stochastic function used. These statistics (e.g. mean and semivariogram) must be estimated from the observations first. The most basic technique, called simple kriging, provides estimates of the conditional mean (dashed line in Figure A8) and the conditional variance, provided that the stochastic function is multivariate Gaussian and stationary. As in this case the conditional distribution is also Gaussian, it is fully characterised by the conditional mean and the conditional variance. More advanced kriging methods, such as indicator kriging, can be used to estimate the complete conditional pdf or cpdf for each point or location, even

if the stochastic function is non-Gaussian. We refer to Isaaks and Srivastava (1989) for an introduction to kriging methods and Goovaerts (1997) and Deutsch and Journel (1998) for advanced kriging algorithms.

Change of support

In the context of this book it is worthwhile to consider what happens to a stochastic function under change of support, in particular when upscaling. For the continuum approach, the fundamental relation between a stochastic function $Z(s_1,t)$ at scale s_1 and a stochastic function at a larger scale s_2 is obtained through Equation (1.2), i.e.

$$Z(s_2;t) = \frac{1}{|s_2|}\int_{t' \in s_2} Z(s_1;t')dt' \qquad \text{(A26)}$$

If $Z(s_1;t)$ is a second order stationary stochastic function $Z(s_2;t)$ is second order stationary too. In this case, the following can be said about he statistics of $Z(s_2;t)$ when compared to $Z(s_1;t)$:

1. the mean remains the same: $E[Z(s_2;t)] = E[Z(s_1;t)] = \mu(s_1)$
2. if the semivariogram of $Z(s_1;t)$ has a nugget variance, this nugget variance becomes zero in the semivariogram of $Z(s_2;t)$; this will happen regardless of the size of s_2, as long as s_2 is larger than s_1;
3. the semivariogram range and the integral scale of the semivariogram of $Z(s_2;t)$ become larger;
4. the variance of $Z(s_2;t)$ (the sill of the semivariogram) becomes smaller.

How much smaller the variance becomes depends on the ratio $I(s_1)/|s_2|$, i.e. the ratio of the integral scale of $Z(s_1;t)$ and the size of s_2. In the limit, if the size of s_2 becomes very large compared to $I(s_1)$, the variance goes to zero and $Z(s_2;t)$ becomes a deterministic constant $z(s_2;t)$ equal to the mean $\mu(s_1)$. Figure A9 shows an example of three semivariograms of second order stationary stochastic functions with increasing support. Vanmarcke (1983) provides an elaborate treatment of the statistical properties of stochastic functions under changing support.

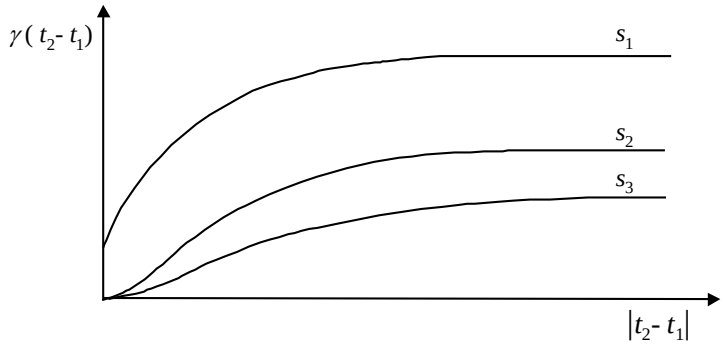

Figure A9. Example of the change in semivariograms under change of support of a stochastic function.

The remarks made above pertain to upscaling of the stochastic function itself, without conditioning to observations. However, in certain applications we may be interested in obtaining the probability $f_Z(s_2;z,t \mid z_i(s_1), i= 1,..,n)$, which is the conditional probability of $Z(s_2;t)$ *given* the n observations $z_i(s_1)$, $i= 1,..,n$ available at scale s_1. For instance, when cleaning up a polluted site by removal of the soil, the total volume of soil is often partitioned into a finite number of blocks of equal size. For each block as a whole it is decided whether it should be removed or not by comparing the average concentration of the block to some critical threshold. As the sample volumes are usually much smaller than the volume of these blocks, upscaling is necessary. Moreover, as the samples are only taken at a limited number of locations, both change of support (upscaling) and change of coverage (interpolation) is required. By using a stochastic function to describe the unknown spatial variation of $z(s_1,t)$ within the blocks s_2, we take the uncertainty of the limited sampling into account as follows. Using the stochastic function concept, for each block the conditional probability $f_Z(s_2;z,t \mid z_i(s_1), i= 1,..,n)$ is estimated. Using this probability distribution, it is then possible to estimate the probability that the average concentration of a given block is above the critical threshold. If this probability is large, the block is removed, otherwise it is left. If the probability is close to 0.5, more samples may be necessary to come to a definite decision.

How does one obtain the conditional distribution $f_Z(s_2;z,t| z_i(s_1), i= 1,..,n)$? If the stochastic function $Z(s_1;t)$ is multivariate Gaussian and stationary and $Z(s_2;t)$ is indeed the arithmetic average of $Z(s_1;t)$ as in Equation (A26), we can use simple block kriging, which is similar to the method of section 2.2.3

(see Goovaerts, 1997). Simple block kriging cannot be used if $Z(s_1;t)$ is not multivariate Gaussian or if we are interested in a non linear average:

$$Y(s_2;t) = \mathrm{g}_2\left[\frac{1}{|s_2|}\int_{t'\in s_2} \mathrm{g}_1\left[Z(s_1;t')\right]dt'\right] \quad \text{(A27)}$$

where either $\mathrm{g}_1[\cdot]$ or $\mathrm{g}_2[\cdot]$ are non-linear functions. In these cases one can either resort to approximate methods, or better, use conditional stochastic simulation. The latter method is equivalent to the upscaling method described in section 2.3.3, i.e. the "indirect stochastic method":

1. using a simulation algorithm, a large number (say N) conditional realisations of $Z(s_1;t)$ are simulated, given the observed values $z_i(s_1)$, i= 1,..,n;
2. for each realisation the required linear average $z(s_2;t)$ or non-linear average $y(s_2;t)$ is calculated for all the support units s_2.
3. From the N simulated $z(s_2;t)$ or $y(s_2;t)$ the conditional probability distribution of $Z(s_2;t)$ or $Y(s_2;t)$ can be estimated for each support unit s_2.

Glossary

Constant	Property of a model that does not change in both space and time within the confines of the model, e.g. the gravity constant. Constants are scale invariant.
Coverage	The ratio of the sum of intervals, areas or volumes of all support units for which the average values are known and the extent
Downscaling	Decreasing the support of the research area (also called "disaggregation")
Extent	The time interval [T], area [L^2] or volume [L^3] or over which observations are made, model outcomes are calculated or policy measures are to be made
Extrapolation	Increasing the extent of the research area
Input variable	Variable originating from outside the model boundary, changes with time and causes the variation in time of the state variable. Input variables also include boundary conditions and initial conditions such as used when solving differential equations.
Interpolation	Increasing the coverage of the research area
Model	A simplified representation of part of reality

Model boundary Separates the part of reality described by the model from the rest of reality

Output variable Variable related to the state variable that changes with time, traverses the model boundary and thus influences the part of reality not described by the model

Parameter Property relating the input variables to the state variable and the state variable to the output variable. Parameters may change in space but are invariant in time. Because they are constant in time, parameters represent the intrinsic properties of the model.

Representative Representative parameters or representative input variables for a support s_2 are those parameters and input variables that, if used in a model that is defined at a smaller support s_1, produce model outputs at scale s_2. The values of representative parameters and input variables depend on the values of these at support s_1, the model used at s_1 and the support ratio s_2 / s_1.

Sample A subset of n support units which have been selected from the N support units making up an area and whose properties have been observed

Sampling Selecting and observing a subset of n support units from the N support units making up an area

Scale Support

Scale transfer Change of support

Singling out Decreasing the extent of the research area: selecting an area of smaller extent

State variable Variable that contains all information about the part of reality that is described by the model at a certain time t_1. To calculate the future behaviour of the model from time on t_1, it is sufficient to know the state at time t_1 as initial condition. The values of the state variables for $t < t_1$ are not important, as they are all contained in the state variables at time t_1.

Support	The largest time interval [T], area [L^2] or volume [L^3] for which the property of interest is considered homogeneous (also called "grain"). For these time intervals, areas or volumes we only know the average value of the property considered and not its variation within.
Support unit	Sub-interval, sub-area or sub-volume with respectively the same time interval, area or volume as the support
Upscaling	Increasing the support of the research area (also called "aggregation")

References

Addiscott, T.M. 1995. Entropy and sustainability. European. Journal of Soil Science 46: 161-168.

Armstrong, M. and P.A. Dowd (eds.), Geostatistical Simulations. Kluwer Academic Publishers, Dordrecht, Netherlands.

Åström, K.J. and P. Eykhoff, 1971. System identification – A survey. Automatica 7: 123-162.

Austin, P.M. and R.A. Houze, 1972. Analysis of the structure of precipitation patterns in New England. Journal of Applied Meteorology 11: 926-935.

Azimonti, G., C. Riparbello, M. Maroni. 1994. Sensitivity of soil to pesticide leaching: a combined application of a leaching model and GIS. Proceedings of the Fifth European conference and exhibition on geographic information systems, EGIS'94. Utrecht, EGIS Foundation 1: 1026-1031.

Barnsley, M.F.,1988. Fractals Everywhere. Academic Press, New York, 394 pp.

Bierkens, M.F.P., 1996. Modeling hydraulic conductivity of a complex confining layer at various spatial scales. Water Resources Research 32(8): 2369-2382.

Bierkens, M.F.P. 1997. Using stratification and residual kriging to map soil pollution in urban areas. In: E.Y. Bafi and N.A. Schofield (eds.). Geostatistics Wollongong, 96: 996-1007. Kluwer Academic Publishers, Dordrecht, Netherlands.

Bierkens, M.F.P. and C.E. Puente, 1990. Analytically derived runoff models based on rainfall point processes. Water Resources Research 26(11): 2653-2659.

Bierkens, M.F.P. and P.A. Burrough. 1993a. The indicator approach to categorical soil data; part 1: Theory. Journal of Soil Science 44: 361-368.

Bierkens, M.F.P. and P.A. Burrough. 1993b. The indicator approach to categorical soil data; part 2: Application to mapping and land use suitability analysis. Journal of Soil Science 44: 369-381.

Bierkens, M.F.P. and H.J.T. Weerts, 1994. Block hydraulic conductivity of cross-bedded fluvial sediments. Water resources Research 30(10): 2665-2678.

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

Bishop, T.F.A., A.B. McBratney and G.M. Laslett, 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91: 27-45.

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

Boesten, J.J.T.I. and A.M.A. Van der Linden. 1991. Modelling the influence of sorption and transformation on pesticide leaching and persistence. Journal of Environmental Quality 20: 425-435.

Bourrenane, H., D. King and P. Chery. 1996. Improving the kriging of a soil variable using slope gradient as external drift. European Journal of Soil Science 47: 473-483.

Box, G.E.P. and G.M. Jenkins, 1976. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, 575 pp.

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

Brus, D.J. 2000. Using regression models in design-based estimation of spatial means of soil properties. European Journal of Soil Science 51 (1): 159-172.

Burgess, T.M. and R. Webster. 1980. Optimal interpolation and isarithmic mapping of soil properties. II Block kriging Journal of Soil Science 31: 333-341.

Clausnitzer, V., J.W. Hopmans and D.R. Nielsen, 1992. Simultaneous scaling of soil water retention and hydraulic conductivity curves. Water Resources Research 28(1): 19-31.

Chow, V.T., D.R. Maidment, L.W. Mays. 1988. Applied hydrology. McGraw-Hill, New York. 572 pp.

Dagan, G., 1981. Analysis of flow through heterogeneous formations by the method of embedding matrix 1: Steady flow. Water Resources Research 17(1): 107-122.

Dagan, 1989. Flow and Transport in Porous Formations. Springer Verlag, Berlin, 465 pp.

De Marsily, G., 1986. Quantitative Hydrogeology, Groundwater Hydrology for Engineers. Academic Press, New York, 440 pp.

Desbarats, A.J., 1987. Numerical estimation of effective permeability in san-shale sequences. Water Resources Research 23(2): 273-286.

Desbarats, A.J., 1991. Spatial Averaging of transmissivity. In: S Bachu (ed.), Proceedings of the 5th Canadian/American Conference on Hydrogeology: Parameter Identification and Estimation for Aquifer and Reservoir Characterization: 139-154. National Water Well Association, Dublin, Ohio.

Desbarats, A.J. and S. Bachu, 1994.Geostatistical modelling of aquifer heterogeneity from the core scale to the basin scale: A case study. Water Resources Research 30(3): 673-684.

Deutsch, C.V. 1994. Algorithmically defined random function models. In: R Dimitrakopoulos (ed.). Geostatistics for the Next Century: 422-435. Kluwer, Dordrecht.

Deutsch, C.V. and A.G. Journel, 1998. GSLIB, Geostatistical Software Library and User's Guide (2nd edition). Oxford University Press, New York: 369 pp.

De Vries, W., J. Kros, C. Van Der Salm, J.E. Groenenberg and G.J. Reinds. 1998. The use of upscaling procedures in the application of soil acidification models at different spatial scales. Nutrient cycling in Agroecosystems 50: 223-236.

De Willigen, P., 1990. Calculation of uptake of nutrients and water by a root system. Nota 210, Institute for Soil Fertility, Haren, The Netherlands.

De Willigen P. and M.van Noordwijk, 1991. Modelling nutrient uptake, from single roots to complete root systems. In: Penning de Vries, F.W.T., H.H.van Laar, and M.J.Kropff (eds.) Simulation and systems analysis for rice production (SARP). Selected papers from workshop: 277-295. PUDOC, Wageningen.

De Willigen, P. and M. van Noordwijk, 1994a. Mass flow and diffusion of nutrients to a root with constant or zero-sink uptake. I. Constant uptake. Soil Science 157:162-170.

De Willigen, P. and M. van Noordwijk, 1994b. Mass flow and diffusion of nutrients to a root with constant or zero-sink uptake. II. Zero-sink uptake. Soil Science 157:171-175.

Dobson, A.J. 1990. An introduction to generalized linear models. Chapman and Hall. London.

Dolman, A.J. and E.A. Blyth, 1997. Patch scale aggregation of heterogeneous land surface cover for mesoscale meteorological model. Journal of Hydrology 190: 252-268.

Domburg, P., J.J. de Gruijter and D.J. Brus. 1994. A structured approach to designing soil survey schemes with prediction of sampling error from variograms. Geoderma 62: 151-164.

Dowla, F.J. and L.L.Rogers, 1996. Solving Problems in Environmental Engineering and Geosciences with Artifical Neural Networks. The MIT Press, Cambridge MA, 310 pp.

Durlofski, L.J., 1991. Numerical calculation of equivalent grid block permeability tensors for heterogeneous porous media. Water Resources Research 27(5): 699-708.

Feder, J., 1988. Fractals, Plenum Press, New York.

Finke, P.A., J.H.M. Wösten and J.G. Kroes. 1996a. Comparing two approaches of characterizing soil map unit behavior in solute transport. Soil Science Society of America Journal 60: 200-205.

Finke, P.A., J.H.M. Wösten and M.J.W. Jansen. 1996b. Effects of uncertainty in major input variables on simulated functional soil behaviour. Hydrological processes 10: 661-669.

Finke, P.A., J. Bouma and M.R. Hoosbeek (eds.), 1998. Soil and Water Quality at Different Scales. Developments in Plant and Soil Sciences 80, Kluwer Academic Publishers, Dordrecht.

Freeze, R.A.,1975. A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. Water Resources Research 11(5): 725-741.

Gelhar, L.W., 1993. Stochastic Subsurface Hydrology. Prentice Hall, New Jersey, 390 pp.

Ginn, T.R. and J.H. Cushman, 1990. Inverse Methods for subsurface flow: A critical review of stochastic techniques. Stochastic Hydrology and Hydraulics 4: 1-26.

Gleick, J. 1988. Chaos: Making an new science. Penguin Books, Harmondsworth, 352 pp.

Gómez-Hernández, J.J. 1992. Regularization of hydraulic conductivities: a numerical approach. In: A. Soares (ed.), Proceedings of the fourth geostatistical congress Troia (Portugal). Quantitative geology and geostatistics 5: 215-232. Kluwer Academic Publishers, Dordrecht, Netherlands.

Gómez-Hernández, J.J. and X.-H. Wen, 1997. To be or not to be multi-Gaussian? A reflection on stochastic hydrology. Advances in Water Resources 21(1): 47-61.

Goovaerts, P. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, 481 pp.

Goovaerts, P. and A.G. Journel. 1995. Integrating soil map information in modelling the spatial variation of continuous soil properties. European Journal of Soil Science 46: 397-414.

Haigh, M.J. 1987. The holon: Hierarchy theory and landscape research. Catena supplement 10: 181-192. Catena-Verlag, Germany.

Heuvelink, G.B.M. and E.J. Pebesma, 1999. Spatial aggregation and soil process modelling. Geoderma 89: 47-65.

Hoosbeek, M.R. and R. Bryant, 1992. Towards the quantitative modelling of pedogenesis: a review. Geoderma 55: 183-210.

Hwang, N.H.C. and R.J. Houghthalen, 1996. Fundamentals of Hydraulic Engineering Systems (3rd edition). Prentice Hall, Upper Saddle River, New Jersey, 416 p.

Indelman, I. and G. Dagan, 1993a. Upscaling of permeability of anisotropic heterogeneous formations, 1: The general framework. Water Resources Research 29(4): 917-923.

Indelman, I. and G. Dagan, 1993b. Upscaling of permeability of anisotropic heterogeneous formations, 2: General structure and small perturbation analysis. Water Resources Research 29(4): 925-933.

Isaaks, E.H. and R.H. Srivastava, 1989. An introduction to Applied Geostatistics. Oxford University Press, New York, 561 pp.

Journel, A.G., 1983. Nonparametric estimation of spatial distributions. Mathematical Geology 15(3): 445-468.

Journel, A.G. and C.J. Huijbregts, 1978. Mining Geostatistics. Academic Press, London, 599 pp.

Kasap, E. and L.W. Lake, 1990. Calculating the effective permeability tensor of a gridblock. SPE Formation Evaluation, June 1990: 192-200

King, D., D.M. Fox, J. Daroussin, Y. Le Bissonnais and V. Danneels. 1998. Upscaling a simple erosion model from small areas to a large region. Nutrient cycling in agroecosystems 50: 143-149.

Kleijnen, J.P.C. 1987. Regression meta-model summarization of model behaviour. In: M.G. Singh (ed.) Systems and control encyclopedia: theory, technology, applications. Pergamon, Oxford.

Knotters, M., D.J. Brus and J.H. Oude Voshaar. 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma 67: 227-246.

Kobus, H. (ed.), 1984. Symposium on Scale Effects in Modelling Hydraulic Structures. Technische Akademie Esslingen.

Kool, J.B., J.C. Parker, and M. Th. Van Genuchten, 1985. Determining Soil Hydraulic Properties from One-step Outflow Experiments by Parameter Estimation: I. Theory and Numerical Studies. Soil Science Society of America Journal 49: 1348-1354.

Lovelock, J.E. 1979. Gaia: A new look at life on earth. Oxford University Press, Cambridge MA, 157 pp.

Mandelbrot , B.B., 1982. The Fractal Geometry of Nature. W.H. Freeman and Co., San Francisco, 468 pp.

McBratney, A.B., 1998. Some considerations on methods for spatially aggregating and disaggregating soil information. Nutrient Cycling in Agroecosystems 50, Special Issue: Finke, P.A., M.R. Hoosbeek and J. Bouma (eds.), Soil and Water Quality at Different Scales: 51-62.

McLaughlin, D and LR. Townley, 1996. A reassessment of the groundwater inverse problem. Water Resources Research 32(5): 1131-1161.

Miller, E.E. and R.D. Miller, 1956. Physical theory for capillary flow phenomena. Journal of Applied Physics 27, 324-332.

Newton, N.J., 1991. Asymptotically efficient Runge-Kutta methods for a class of Itô and Stratonovich equations. SIAM Journal of Applied Mathematics51(2): 542-567.

O'Neill, R.V., D.L. De Angelis, J.B. Waide and T.F.H. Allan. 1986. A hierarchical concept of ecosystems. Princeton University Press, Princeton, New Yersey, 253 pp.

Papoulis, A., 1991. Probability, Random Variables, and Stochastic Processes (third edition). McGraw-Hill Inc., New York, 666 pp.

Parker, J.C., J.B. Kool , and M. Th. Van Genuchten, 1985. Determining Soil Hydraulic Properties from One-step Outflow Experiments by Parameter Estimation: II Experimental Studies. Soil Science Society of America Journal 49: 1354-1359.

Pebesma, E.J. and J.W. De Kwaadsteniet. 1997. Mapping groundwater quality in the Netherlands. Journal of Hydrology 200: 364-386.

Ponce-Hernandez R., F.H.C.Marriott & P.H.T.Beckett, 1986. An improved method for reconstructing a soil profile from analyses of a small number of samples. Journal of Soil Science 37: 455-467.

Puente, C.E., M.F.P. Bierkens, M.A. Diaz-Granados, P.E. Dik and M.M. Lopez, 1993. Practical use of analytically derived runoff models based on rainfall point processes. Water Resources Research 29(10): 3551-3560.

Puente, C.E. and N. Obregón, 1996. A deterministic geometric representation of temporal rainfall: Results for a storm in Boston. Water Resources Research 32(9): 2825-2839.

Renard, Ph. And G. de Marsily, 1997. Calculating equivalent permeability: a review. Advances in Water Resources 20(5-6): 253-278.

Robinson, J.S. and M. Sivapalan, 1997. Temporal scales and hydrological regimes: Implications for flood frequency scaling. Water Resources Research, 33(12): 2981-2999.

Rodriguez-Iturbe, I., 1986a. Scale problems in hydrological processes. In: Shen, H.W., J.T.B. Obeysekera, V. Yevyevich and D.G. Decoursy (eds.), Multivariate Analysis of Hydrological Processes: 51-65. Colorado State University.

Rodríguez-Iturbe, I., 1986b. Scale of fluctuation of rainfall models. Water Resources Research 22(9): 15S-37S.

Rodríguez-Iturbe, I., B. Febres de Power and J.B. valdes, 1987. Rectangular pulses point process models for rainfall: analysis of empirical data. Journal of Geophysical Research 92(D8): 9645-9656.

Rubin, Y. and J.J. Goméz-Hernadez, 1990. A stochastic approach to the problem of upscaling of conductivity in disordered media: Theory and unconditional numerical simulations. Water Resources Research 26(4): 691-701.

Runge, E.C.A. 1973. Soil development sequences ands energy models. Soil Science 115: 183-193.

Samra, J.S., K.N.S. Sharma and N.K. Tyagi. 1988a. Analysis of spatial variability in sodic soils: 1. Structural analysis. Soil Science 145: 180-187.

Samra, J.S., V.P. Singh and K.N.S. Sharma. 1988b. Analysis of spatial variability in sodic soils: 2. Point - and block-kriging. Soil Science 145: 250-256.

Sánchez-Villa, X., J.P. Girardi and J. Carrera, 1995. A synthesis of approaches to upscaling of hydraulic conductivities. Water Resources Research 31(4): 867-882.

Särndal, C.E., B. Swensson and J. Wretman, 1992. Model Assisted Survey Sampling. Springer-Verlag, New York, 694 pp.

Särndal, C.E. 1978. Design-based and model-based inference in survey sampling. Scandinavian Journal of Statistics 5: 27-52.

Shibata, R., 1976. Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika 63: 117-126.

Sivapalan, M. 1993. Linking hydrologic parameterizations across a range of scales: hillslope to catchment to region. In: Exchange processes at the land surface for a range of space and time scales. Proceedings of the Yokohama symposium, IAHS Publ. 212, 1993: 115-123.

Sposito, G. 1998. Scale dependence and scale invariance in hydrology. Cambridge University Press. 423 p.

Stein, A., M. Hoogerwerf and J. Bouma. 1988. Use of soil-map delineations to improve (co-)kriging of point data on moisture deficits. Geoderma 43: 163-177.

Torfs, P.J.J.F. 1998. Statistical disaggregation of hydrological time series using wavelets. In: Hydrology in a Changing environment Volume I: 81-89. British Hydrology Society.

Unny, T.E., 1984. Numerical integration of stochastic differential equations in catchment modelling. Water Resources Research 20(3): 360-368.

Van Dam J.C., J.N.M. Stricker and P. Droogers, 1994. Inverse Method to Determine Soil Hydraulic Functions from Multistep Outflow Experiments. Soil Science Society of America Journal 58: 647-652.

Van der Linden, A.M.A. and J.J.T.I. Boesten. 1989. Calculation of the degree of leaching and accumulation of pesticides as a function of their sorption coefficient and degradation rate in plough layers (in Dutch). RIVM, rapport 728800003.

Van de Ven, F.H.M., L. Kohsiek, J. Da Silva Costa and J.J. Fried. 1992. EC Action programme for sustainable use of groundwater. European Water Pollution Control 294: 9-19.

Van Gardingen, P.R., G.M. Foody and P.J. Curran (eds.), 1997. Scaling-up from Cell to Landscape. Cambridge University Press, 397 pp.

Van Genuchten, M.Th., 1980. A closed form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal 44: 892-898.

Vanmarcke, E. 1983. Random Fields: Analysis and Synthesis. The MIT Press, Cambridge MA, 382 pp.

Van Wijk, W.R. (ed.), 1963. Physics of Plant Environment. North Holland, Amsterdam, 382 pp.

Wagenet, R.J., 1998. Scale issues in agroecological research. Nutrient Cycling in Agroecosystems 50, Special Issue: Finke, P.A., M.R. Hoosbeek and J. Bouma (eds.), Soil and Water Quality at Different Scales: 51-62.

Webster, R. and M.A. Oliver. 1992. Sample adequately to estimate variograms of soil properties. Journal of Soil Science 43: 177-192.

Weiss, P.A. 1971. The basic concepts of hierarchical systems. In: Weiss, P.A. (ed.) Hierarchically organized systems in theory and practice: 1-43. Hafner publishing Co., New York.

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

Whitaker, S. 1986. Flow in porous media I: A theoretical derivation of Darcy's Law. Transport in porous media 1: 3-25.

Wilby, R.L., M.L. Wigley, D.Conway, P.D. Jones, B.C. Hewitson, J. Main and D.S. Wilks, 1998. Statistical downscaling of general circulation model output: a comparison of methods. Water Resources Research 34(11): 2995-3008.

Yeh, W.W.-G., 1986. Review of parameter identification procedures in groundwater hydrology: The inverse problem. Water Resources Research 22(2): 95-108.

Index