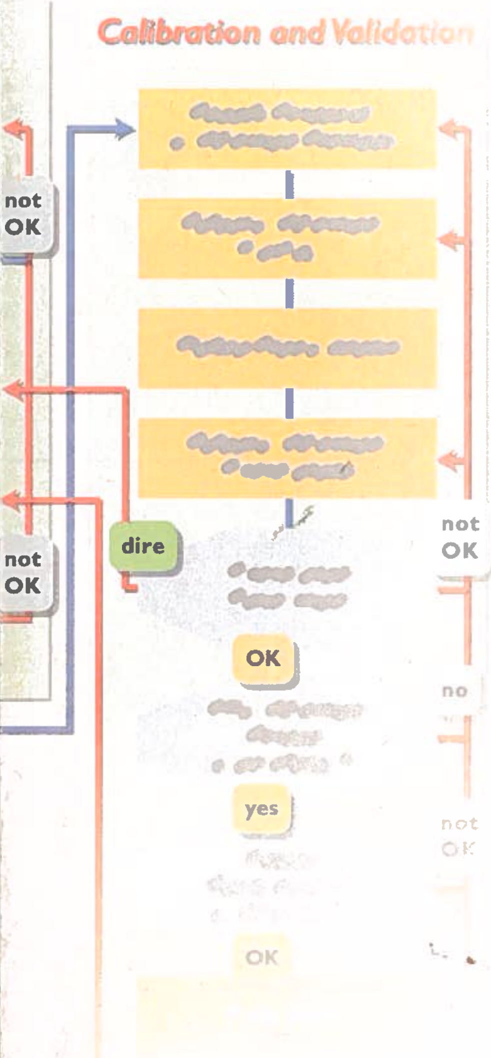
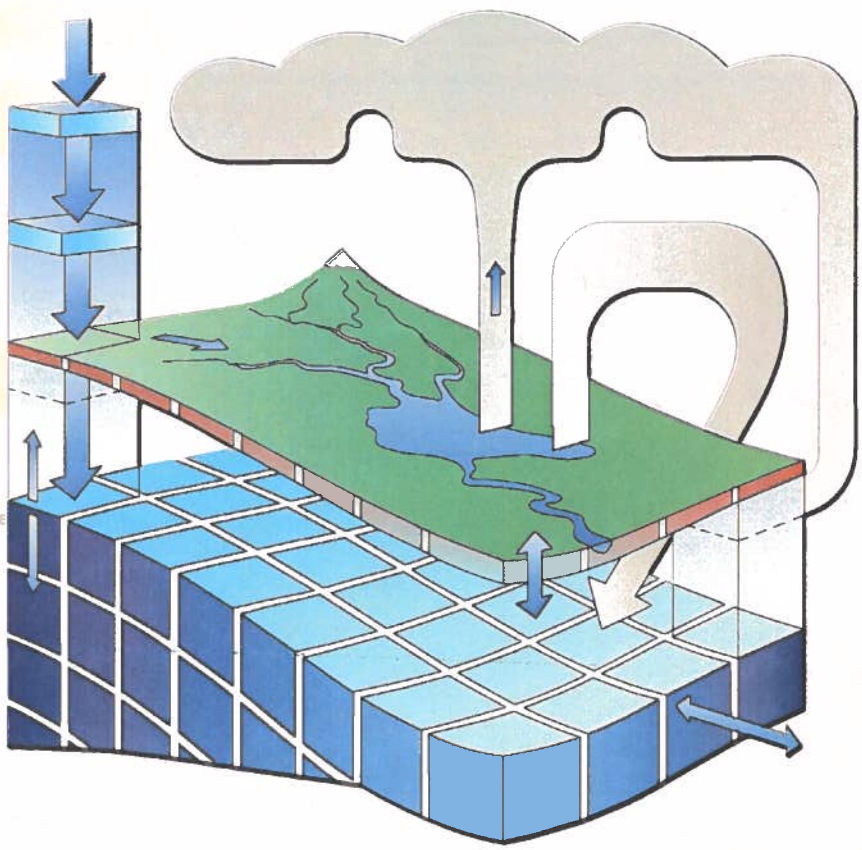
Danmarks og Grønlands Geologiske Undersøgelse — Særudgivelse 2007

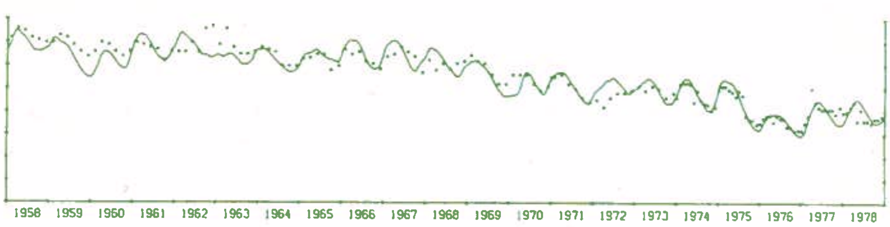
# Hydrological Modelling and River Basin Management

## Doctoral Thesis

Jens Christian Refsgaard






Geological Survey of Denmark and Greenland
Danish Ministry of the Environment

GEUS

*Danmarks og Grønlands Geologiske Undersøgelse — Særudgivelse 2007*

# Hydrological Modelling and River Basin Management

## Doctoral Thesis

*Jens Christian Refsgaard*

**Geological Survey of Denmark and Greenland**
Danish Ministry of the Environment

GEUS

Denne afhandling er af Det Naturvidenskabelige Fakultet ved Københavns Universitet antaget til offentligt at forsvares for den naturvidenskabelige doktorgrad.

København, den 5. januar, 2007

Nils O. Andersen
Dekan

Forsvaret vil finde sted fredag den 1 juni, 2007 kl $14^{00}$ i Anneksauditorium A, Studiestræde 6, Københavns Universitet

This thesis has been accepted by the Faculty of Natural Science at the University of Copenhagen for public defence in fulfilment of the degree of Doctor of Science.

Copenhagen, 5th January, 2007

Nils O Andersen
Dean

The defence will take place on Friday 1st June, 2007 at $14^{00}$ in Anneksaudiorium A, Studiestræde 6, University of Copenhagen

***Special Issue***

***Author:*** Jens Christian Refsgaard

***Illustrations***: Kristian A. Rasmussen and reproductions from existing publications

***Cover:*** Kristian A. Rasmussen

***Date:*** January 2007

The Report is available on the internet at http://www.geus.dk/

ISBN 978-87-7871-185-4

Geological Survey of Denmark and Greenland (GEUS)
Øster Voldgade 10
DK-1350 København K
Tel: +45 38142000
Fax: +45 38142050
Email: geus@geus.dk
http://www.geus.dk/

# Table of Contents

## *Appendices: Publications [1] – [15]*

[1] Refsgaard JC, Hansen E (1982) A Distributed Groundwater/Surface Water Model for the Suså Catchment. Part 1: Model Description. Nordic Hydrology, 13, 299-310.

[2] Refsgaard JC, Hansen E (1982) A Distributed Groundwater/Surface Water Model for the Suså Catchment. Part 2: Simulations of Streamflow Depletions Due to Groundwater Abstraction. Nordic Hydrology, 13, 311-322.

[3] Refsgaard JC, Christensen TH, Ammentorp HC (1991) A model for oxygen transport and consumption in the unsaturated zone. Journal of Hydrology, 129, 349-369.

[4] Refsgaard JC, Seth SM, Bathurst JC, Erlich M, Storm B, Jørgensen, GH, Chandra S (1992) Application of the SHE to catchments in India - Part 1: General results. Journal of Hydrology, 140, pp 1-23.

[5] Jain SK, Storm B, Bathurst JC, Refsgaard JC, Singh RD (1992) Application of the SHE to catchments in India - Part 2: Field experiments and simulation studies with the SHE on the Kolar subcatchment of the Narmada River. Journal of Hydrology, 140, 25-47.

[6] Refsgaard JC, Knudsen J (1996) Operational validation and intercomparison of different types of hydrological models. Water Resources Research, 32 (7), 2189-2202.

[7] Refsgaard JC (1997) Parametrisation, calibration and validation of distributed hydrological models. Journal of Hydrology, 198, 69-97.

[8] Refsgaard JC (1997) Validation and Intercomparison of Different Updating Procedures for Real-Time Forecasting. Nordic Hydrology, 28, 65-84.

[9] Refsgaard JC, Sørensen HR, Mucha I, Rodak D, Hlavaty Z, Bansky L, Klucovska J, Topolska J, Takac J, Kosc V, Enggrob HG, Engesgaard P, Jensen JK, Fiselier J, Griffioen J, Hansen S (1998) An Integrated Model for the Danubian Lowland – Methodology and Applications. Water Resources Management, 12, 433-465.

[10] Refsgaard JC, Thorsen M, Jensen JB, Kleeschulte S, Hansen S (1999) Large scale modelling of groundwater contamination from nitrogen leaching. Journal of Hydrology, 221(3-4), 117-140.

[11] Thorsen M, Refsgaard JC, Hansen S, Pebesma E, Jensen JB, Kleeschulte S (2001) Assessment of uncertainty in simulation of nitrate leaching to aquifers at catchment scale. Journal of Hydrology, 242, 210-227.

[12] Refsgaard JC, Henriksen HJ (2004) Modelling guidelines – terminology and guiding principles. Advances in Water Resources, 27(1), 71-82.

[13] Refsgaard JC, Henriksen HJ, Harrar WG, Scholten H, Kassahun A (2005) Quality assurance in model based water management – review of existing practice and outline of new approaches. Environmental Modelling & Software, 20, 1201-1215.

[14] Refsgaard JC, Nilsson B, Brown J, Klauer B, Moore R, Bech T, Vurro M, Blind M, Castilla G, Tsanis I, Biza P (2005) Harmonised techniques and representative river basin data for assessment and use of uncertainty information in integrated water management (HarmoniRiB). Environmental Science and Policy, 8, 267-277.

[15] Refsgaard JC, van der Sluijs JP, Brown J, van der Keur P (2006). A framework for dealing with uncertainty due to model structure error. Advances in Water Resources, 29, 1586-1597.

# Preface

The work presented in this thesis together with the 15 publications published between 1982 and 2006 form the material for evaluation for the degree of doctor scientiarum (dr. scient.) at the University of Copenhagen. The papers have all been published in peer reviewed international scientific journals. They are referred to by the numbers [1] to [15].

In the present report I have assembled and summarised my most important scientific contributions to catchment modelling that has been my research interest during the past three decades. In this connection I wish to thank all my co-authors for a very inspiring co-operation during the years. Research does not take place in a vacuum, and without the interactions with them my work would not have been possible.

I wish to acknowledge former and present colleagues and managements at the three organisations where I have been employed. At the Institute of Hydrodynamics and Hydraulic Engineering, Technical University of Denmark (now Environment and Resources, DTU) I was given the opportunity to explore and develop new integrated groundwater/surface water catchment models at a time when hydrological modelling was still in its infancy. This showed me the enormous potential of this new field. At Danish Hydraulic Institute (now DHI Water & Environment) I was then entrusted with further development of modelling tools and with testing them in real life applications. This taught me the limitations and difficulties we encounter and the need to be humble when applying models in water resources management. Finally, the Geological Survey of Denmark and Greenland (GEUS) has provided a very inspiring scientific environment and given me the opportunity to get involved in broader international research projects that have matured much of my previous views and allowed me to assemble this work.

A special thank goes to Kristian A. Rasmussen, GEUS, for using his magic touch to polish some of the old dusty figures from the last century to make them easier to read in this thesis.

Last, but not least, I wish to thank my family for their patience and support and for accepting that I always have been too busy with this topic.

Copenhagen, January 2007
Jens Christian Refsgaard

***"Life can only be understood backwards; but it must be lived forwards"***
***Søren Kierkegaard (1813-1855)***

# Dansk Resume

Publikationerne og materialet i denne doktorafhandling beskriver en række videnskabelige undersøgelser af hydrologisk modellering på oplandsskala i relation til vandressourceforvaltning. Hver af de 15 publikationer fokuserer på dele af det overordnede emne spændende fra udvikling af nye koncepter og modelkoder til modelanvendelser; fra punktskala til oplandsskala; fra modellering af vandstrømninger til transport af opløste og reaktive stoffer; fra fokus på planlægning til real-tids oversvømmelsesvarsling og videre til tværgående emner og protokoller for selve modelleringsprocessen.

Afhandlingens kapitel 2 præsenterer protokoller for hydrologisk modellering og en diskussion af interaktionen mellem hydrologisk modellering og vandressourceforvaltning. Endvidere forklares den terminologi og den tilgrundlæggende videnskabsfilosofiske tankegang samt den klassifikation af modeltyper, som benyttes i resten af afhandlingen. Kapitel 3 indeholder resumeer af modelstudier baseret på ni af publikationerne. Vurderingerne af disse publikationers bidrag til ny viden på det tidspunkt de blev publiceret og af emner som ikke blev behandlet i publikationerne, viser en betydelig udvikling gennem de sidste 25 år. Fx indeholder de første publikationer om udvikling af nye modelkoder, intet om verifikation af modelkode, validering af modeller mod uafhængige data eller usikkerhedsvurderinger – emner som i dag betragtes som meget væsentlige. Eksemplerne illustrerer ligeledes, hvordan generelle emner som skalaproblemer og model validering gradvis udviklede sig med baggrund i erfaringer og erkendte problemer fra modelstudier, som egentlig havde andre formål. Kapitel 4 præsenterer og diskuterer herefter fire generelle emner: (a) heterogenitet og skalering; (b) konfirmation, verifikation, kalibrering og validering af modeller; (c) usikkerhedsvurderinger; og (d) kvalitetssikring af modelleringsprocessen.

Mine væsentligste bidrag til ny videnskabelig viden har været indenfor de følgende fem områder:

- *Ny konceptuel forståelse og tilhørende kodeudvikling*. Suså modellen var baseret på en ny forståelse af interaktionen mellem overfladevand og grundvand i moræneområder og bragte ny viden om hvorledes grundvandsindvinding påvirker vandløb i sådanne oplande.
- *Validering af modeller*. Arbejdet med rigoristiske principper for validering af modeller og eksempler på anvendelser for såvel 'lumped conceptual' og 'distributed physically-based' modeller har været en grundpille gennem de sidste 15 år af min forskning. Specielt er introduktionen af begrebet 'conditional validation' ny.
- *Skalering*. Mit arbejde har ikke 'løst' skalaproblemerne, men bidrager til at tydeliggøre de principielt forskellige metoder med fokus på deres respektive forudsætninger og begrænsninger.
- *Usikkerhedsvurderinger*. En betydelig del af min forskningsaktivitet gennem de sidste 10 år har fokuseret på usikkerhedsaspekter. Mit hovedbidrag i den sammenhæng har været introduktion af bredere usikkerhedsaspekter i hele modelleringsprocessen samt arbejdet med usikkerheder på modelstruktur.
- *Protokoller for hydrologisk modellering og kvalitetssikring af modelleringsprocessen*. Den omfattende og detaljerede modelleringsprotokol, som blev udviklet i HarmoniQuA projektet er en formalisering og udmøntning af erfaring fra de foregående 25 års arbejde med hydrologisk modellering. De ny elementer heri er den fokus der lægges på (a) den interaktive dialog mellem modellør, vandressourceforvalter, reviewer, interessenter og offentligheden; (b) usikkerhedsvurderinger som et løbende element gennem hele modelleringsprocessen; (c) model validering; og (d) introduktion af erfaringer og subjektiv viden via eksterne reviews.

# Abstract

The publications and material presented in this thesis describe a series of scientific investigations on catchment modelling in relation to water resources management. Each of the 15 publications represents parts of the overall topic ranging from development of new concepts and model codes to model applications; from point scale to catchment scale; from flow modelling to transport and reactive modelling; from planning type applications to real-time forecasting and further on to crosscutting issues and protocols for the modelling process.

The thesis starts with a presentation of protocols for the hydrological modelling process together with a discussion of the interaction between the water resources planning and management process and the hydrological modelling process. This includes a definition of terminology, a discussion of the underlying scientific philosophy and a classification of hydrological models. The following chapter comprises summaries of cases of simulation models based on nine of the publications. The post evaluations of the contributions to scientific knowledge in the publications and the issues not taken into account in the earlier publications reveal significant developments over the years. For example the first publications focussing on development of new model codes did not put any emphasis on rigorous verification or validation tests nor on uncertainty assessments, which are key issues today. The cases furthermore illustrate how general issues such as scaling and model validation gradually emerged from experiences and problems encountered in catchment studies that had other primary objectives. The next chapter then provides a presentation and discussion of four general issues: (a) catchment heterogeneity and scaling; (b) confirmation, verification, calibration and model validation; (c) uncertainty assessment; and (d) quality assurance in model based water management.

My main contributions to scientific knowledge have been in the following five areas:

- *New conceptual understanding and code development*. The Suså model was based on a new conceptual understanding of the surface water/groundwater interaction in moraine catchment and brought new insight into the effect of groundwater abstraction on streamflow in catchments with such hydrogeological characteristics.
- *Model validation.* The work on rather rigorous principles for model validation and the examples of their application both for lumped conceptual and distributed physically based models is a cornerstone in my research. In particular the introduction of the term 'conditional validation' is novel.
- *Scaling.* The framework on scaling does not 'solve' the scaling problem but contributes to clarifications on applicable methodologies with focus on their respective assumptions and limitations.
- *Uncertainty assessment.* During the past decade a considerable part of my research work has focussed on uncertainty aspects. I consider my main contributions in this respect to be the introduction of the broader uncertainty aspects integrated into the modelling framework and the work with model structure uncertainty.
- *Modelling protocols and guidelines for quality assurance in the modelling process*. The comprehensive modelling protocol developed within the HarmoniQuA project is a formalisation of experience and practises that have gradually emerged over the years. The novel elements are the emphasis on (a) the interactive dialogue between modeller, water manager, reviewer, stakeholders and the public; (b) uncertainty assessments throughout the modelling process; (c) model validation; and (d) experience and subjective knowledge introduced through external model reviews.

# 1. Introduction

## 1.1 Water Resources Management and Hydrological Modelling

> "Scarcity and misuse of fresh water pose a serious and growing threat to sustainable development and protection of the environment. Human health and welfare, food security, industrial development and the ecosystems on which they depend, are all at risk, unless water and land resources are managed more effectively in the present decade and beyond than they have been in the past". (ICWE, 1992)

> "The fact that the world faces a water crises has become increasingly clear in recent years. Challenges remain widespread and reflect severe problems in the management of water resources in many parts of the world. These problems will intensify unless effective and concerted actions are taken". (WWAP, 2003)

The first of the above quotes presents the status and the future challenges facing hydrologists and water resources managers as summarised in the introductory paragraph of the Dublin Statement on Water and Sustainable Development (ICWE, 1992). The second quote is from the first chapter of the UN World Water Development Report "Water for People, Water for Life" which is a collaborative effort of 23 UN agencies and convention secretariats co-ordinated by the World Water Assessment Programme.

Thus the challenges in water resources management are enormous, both at the global scale as illustrated above and at smaller scales as for instance outlined in the vision for the European water sector recently formulated by the European Water Supply and Sanitation Technology Platform (WSSTP, 2005).

The present thesis deals with hydrological modelling. It must be emphasised that modelling in itself is not sufficient to address these challenges. Modelling only constitute one, among several, sets of tools that can be used to support water resources management. Computer based hydrological models have been developed and applied at an ever increasing rate during the past four decades. The key reasons for that are twofold: (a) improved models and methodologies are continuously emerging from the research community, and (b) the demand for improved tools increases with the increasing pressure on water resources. Overviews of the status and development trends in catchment scale hydrological modelling during this period can be found in Fleming (1975) and Singh (1995).

## 1.2 Objective and Content

The objective of this thesis is to present the contributions to scientific knowledge that has emerged from the research described in the 15 appended publications. I have structured the thesis with an aim of presenting my research contributions within a framework of catchment modelling and its application to support water resources management.

The next chapter (Chapter 2) therefore presents an overall framework of the water resources management and planning process and the modelling process and the interaction between these two processes. Here the terminology and modelling protocol are introduced and discussed. This chapter is based on publications [7], [12] and [13], i.e. mainly some of my most recent work.

Chapter 3 comprises a number of examples of simulation models ranging from point scale to catchment scale, from flow modelling to transport and reactive modelling and from planning type applications to real-time forecasting. This chapter is based on publications [1], [2], [3], [4], [5], [6], [8], [9] and [10], i.e. mainly some of my earlier work.

Chapter 4 then provides a presentation and discussion of key and cross-cutting issues in hydrological modelling such as scaling, model validation, uncertainty assessment and quality assurance. These issues that were introduced as part of the overall framework in Chapter 2 are here discussed with reference to the experience and findings made in the publications. This chapter includes ideas, views and material from all the 15 publications, but with more emphasis on some of the more general purpose publications [6], [7], [10], [11], [12], [13], [14] and [15].

Finally, Chapter 5 contains some conclusions and perspectives for future work.

Thus I have not structured the content of this report according to the chronology of my publications [1] – [15]. The reason for this is that my most recent work provides a broader and better overview of the topic and is thus better suited for providing a framework for my earlier work.

# 2 Water Resources Management and the Modelling Process

## 2.1 Modelling as Part of the Planning and Management Process

Integrated Water Resources Management (IWRM) is “a process, which promotes the co-ordinated development and management of water, land and related resources, in order to maximise the resultant economic and social welfare in an equitable manner without compromising the sustainability of vital ecosystems” (GWP, 2000). In the EU Water Framework Directive (WFD) Guidance Document on Planning Processes planning is defined as “a systematic, integrative and iterative process that is comprised of a number of steps executed over a specified time schedule” (EC, 2003b). In all new guidelines on water resources management the importance of integrated approaches, cross-sectoral planning and of public participation in the planning process are emphasised (GWP, 2000; EC, 2003b; Jønch-Clausen, 2004).

Models describing water flows, water quality, ecology and economy are being developed and used in increasing number and variety to support water management decisions. The interactions between the modelling process and the water management process are illustrated in Figs. 1 and 2. Fig. 1 shows the key actors in the water management process and the five steps that the modelling process typically may be decomposed in. The organisation that commissions a modelling study is denoted the water manager. This is often the competent authority, but can also be a stakeholder such as a water supply company. The role of the government is most often limited to providing the enabling environment such as legislation, research and information infrastructure. The typical cyclic and iterative character of the water management process, such as the WFD process, is illustrated in Fig. 2, where the interaction with the modelling process is illustrated by the large circle (water management) and the four smaller supporting circles (modelling). The WFD planning process, as most other planning processes, contains four main elements:

- *Identification* including assessment of present status, analysis of impacts and pressures and establishment of environmental objectives. Here modelling may be useful for example for supporting assessments of what are the reference conditions and what are the impacts of the various pressures (EC, 2004).
- *Designing* including the set up and analysis of programme of measures designed to be able in a cost effective way to reach the environmental objectives. Here modelling will typically be used for supporting assessments of the effects and costs of various measures under consideration.
- *Implementing* the measures. Here on-line modelling in some cases may support the operational decisions to be made.
- *Evaluation* of the effects of the measures on the environment. Here modelling may support the monitoring in order to extract maximum information from the monitoring data, e.g. by indicating errors and inadequacies in the data and by filtering out the effects of climate variability.

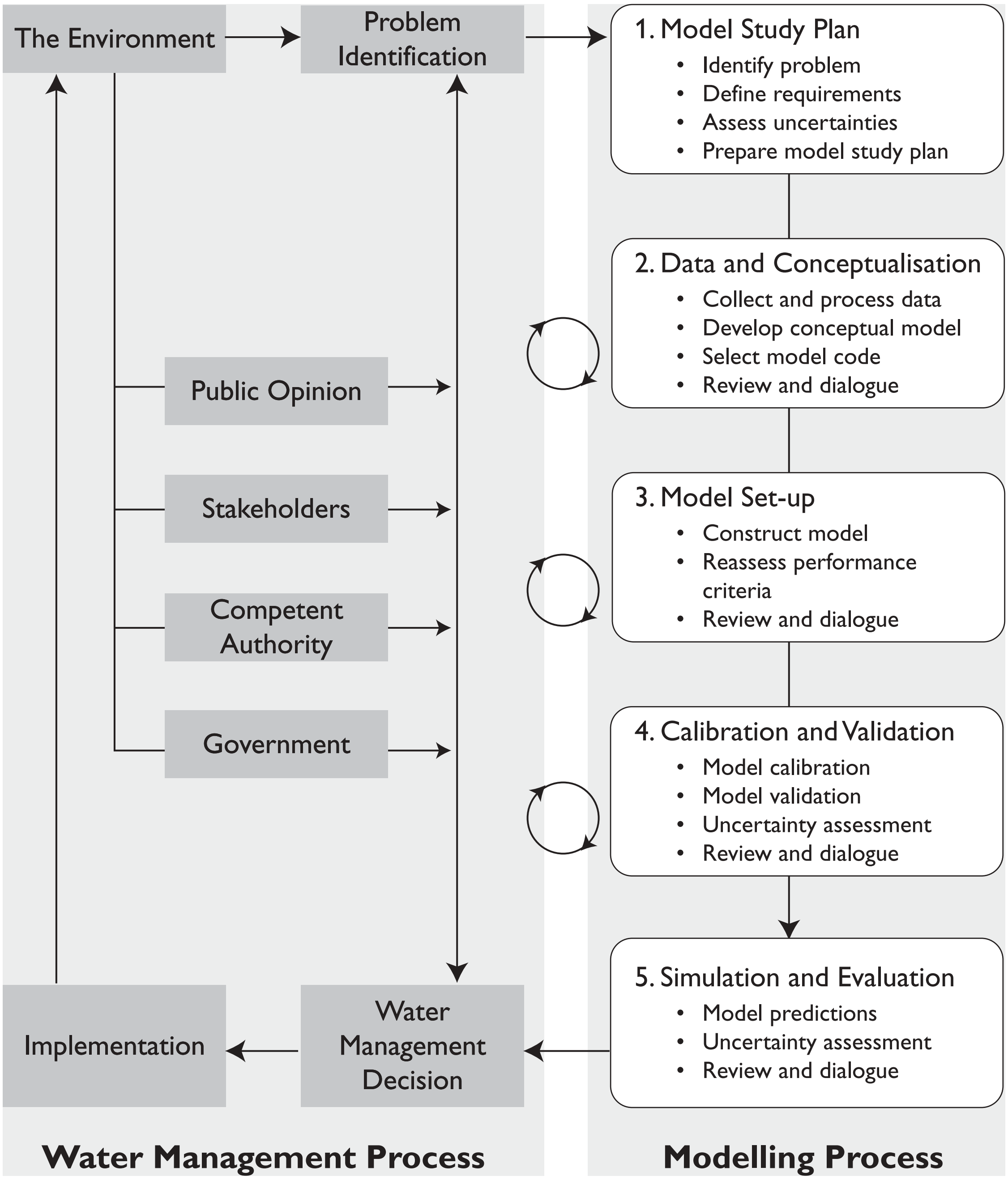



*Fig. 1 The role of the modelling process and the water management decision process (inspired from Pascual et al. (2003).*

It is important to note that the modelling studies typically do not address the entire planning and management process, but rather support certain elements of the process. Modelling is applied as a response (but usually not the only response) to an identified problem and can provide support for water management decisions. The types of interactions between the modelling process and the planning and management process are:

- The modelling process starts with a thorough *framing of the problem* to be addressed and definition of modelling objectives and requirements for the modelling study (Step 1 in Fig. 1). Water managers and stakeholders dominate this step, which basically is identical to part of the broader planning process. A participatory based assessment of the most important sources of uncertainty for the decision process should be used as a basis for prioritising the elements of the modelling study. The uncertainty assessments made at this stage will typically be qualitative.
- *The main modelling itself* is composed of steps 2, 3 and 4 of Fig. 1. Here the link with the main planning process consists of dialogue, reviews and discussions of preliminary results. The amount and type of interaction here depends on the level of public participation that may vary from case to case from providing information over consultation to active involvement (Henriksen et al., submitted).
- *The finalisation of the modelling study* (equivalent to the last step in Fig. 1), typically including *scenario simulations*. Here the water managers and the stakeholders again have a dominant role. The decisions made at the outcome of this step on the basis of modelling results are made in the context of the main planning process. Uncertainty assessment of model predictions is a crucial aspect of the modelling results and should be communicated in a way that is accessible for the stakeholders in the further water management process.

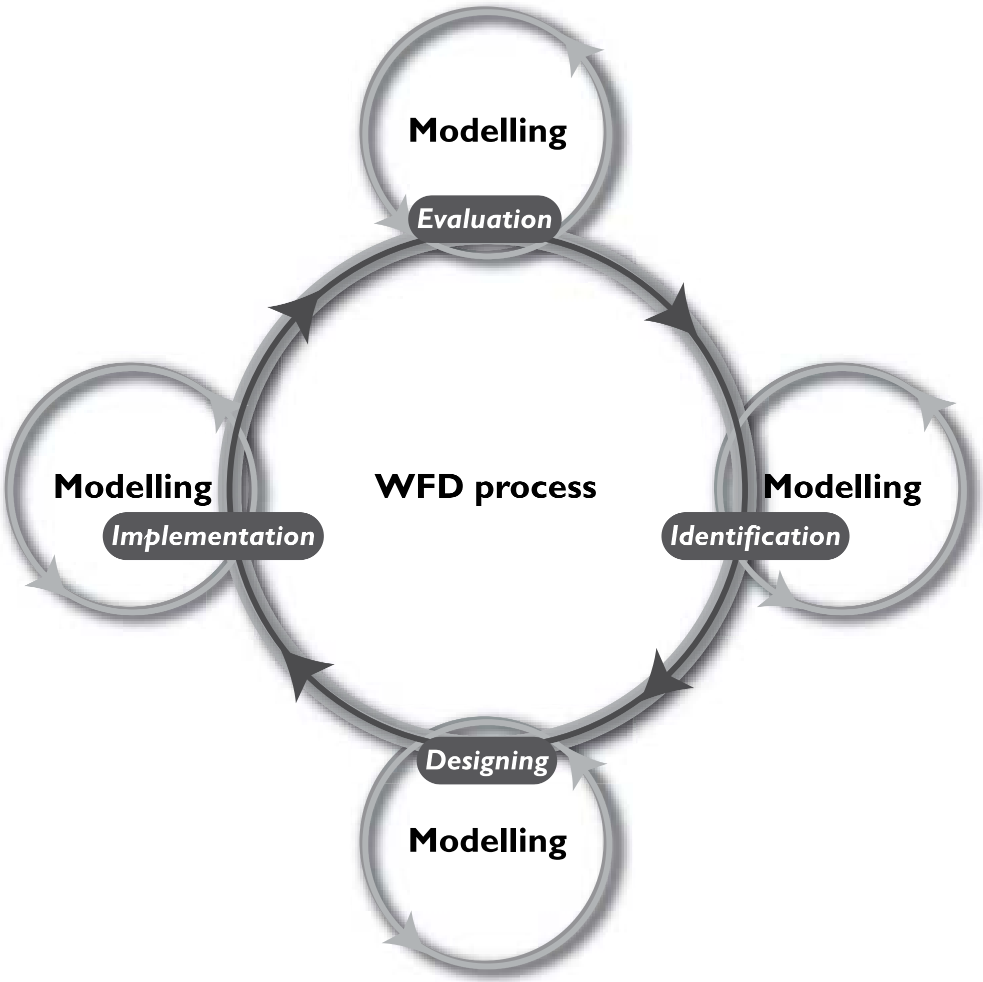



*Fig. 2 The role of modelling in the water management process within the context of the EU Water Framework Directive (WFD)*

## 2.2 Terminology and Scientific Philosophical Basis for the Modelling Process

### 2.2.1 Background

As pointed out in [12] a key problem in relation to establishment of a theoretical modelling framework is confusion on terminology. For example the terms validation and verification are used with different, and some times interchangeable, meanings by different authors. The confusion arises from both semantic and philosophical considerations (Rykiel, 1996). Another important problem is the lack of consensus related to the so far non-conclusive debate on the fundamental question concerning whether a water resources model can be validated or verified, and whether it as such can be claimed to be suitable or valid for particular applications (Konikow and Bredehoeft, 1992; De Marsily et al., 1992; Oreskes et al., 1994).

An important issue in relation to validation/verification is the distinction between open and closed systems. A system is a closed system if its true conditions can be predicted or computed exactly. This applies to mathematics and mostly to physics and chemistry. Systems where the true behaviour cannot be computed due to uncertainties and lack of knowledge on e.g. input data and parameter values are called open systems. The systems we are dealing with in water resources management, based on geosciences, biology and socio-economy, are open systems. According to Konikow and Bredehoeft (1992) and Oreskes et al. (1994) it is not possible to verify or validate models of open systems.

Finally, the principles have to reflect and be in line with the underlying philosophy of environmental modelling that have changed significantly during the past decades. In the early days many of us were focussing on the huge potentials of sophisticated models in a way that in retrospect may be characterised as rather naive enthusiasm (e.g. Freeze and Harlan (1969); Abbott, 1992). The dominant views today appears to be a much more balanced and mature view (e.g. Beven, 2002a; Beven, 2002b).

### 2.2.2 Terminology and guiding principles

According to the terminology presented in [12] the simulation environment is divided into four basic elements as shown in Fig. 3. The inner arrows describe the processes that relate the elements to each other, and the outer circle refers to the procedures that evaluate the credibility of these processes.

In general terms a model is understood as a simplified representation of the natural system it attempts to describe. However, a distinction is made between three different meanings of the general term model, namely the conceptual model, the model code and the model that here is defined as a site-specific model. The most important elements in the terminology and their interrelationships are defined as follows:

*Reality*: The natural system, understood here as the study area.

*Conceptual model*: A description of reality in terms of verbal descriptions, equations, governing relationships or 'natural laws' that purport to describe reality. This is the user's perception of the key hydrological and ecological processes in the study area (perceptual model) and the corresponding simplifications and numerical accuracy limits that are assumed acceptable in order to achieve the purpose of the modelling. A conceptual model thus includes both a mathematical description (equations) and a descriptions of flow processes, river system elements, ecological structures, geological features, etc. that are required for the particular purpose of modelling. By drawing an analogy to scientific philosophical discussion the conceptual model in other words constitutes the scientific hypothesis or theory that we assume for our particular modelling study.

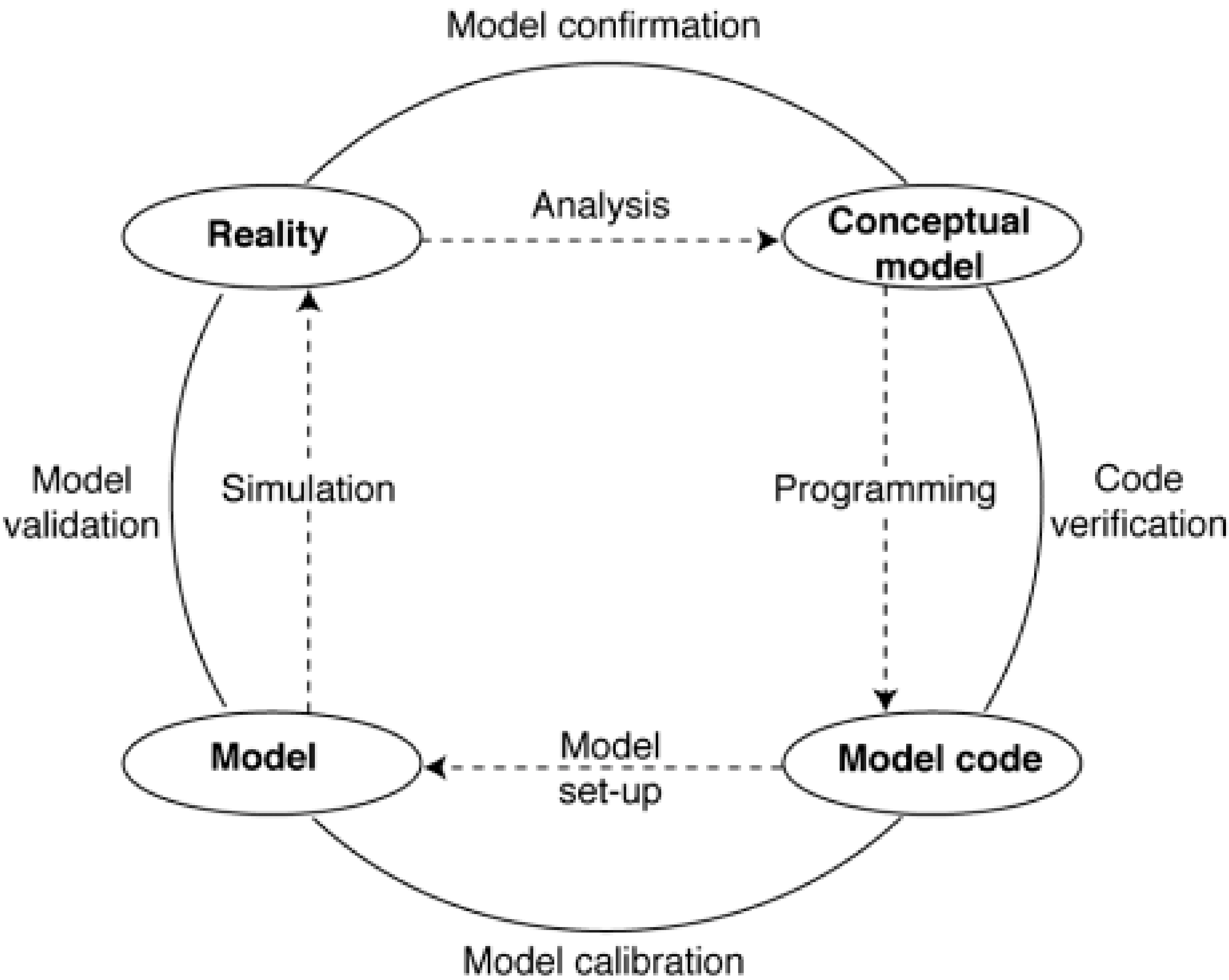



*Fig. 3 Elements of a modelling terminology [12].*

*Model code*: A mathematical formulation in the form of a computer program that is so generic that it, without program changes, can be used to establish a model with the same basic type of equations (but allowing different input variables and parameter values) for different study areas.

*Model*: A site-specific model established for a particular study area, including input data and parameter values.

*Model confirmation:* Determination of adequacy of the conceptual model to provide an acceptable level of agreement for the domain of intended application. This is in other words the scientific confirmation of the theories/hypotheses included in the conceptual model.

*Code verification*: Substantiation that a model code is in some sense a true representation of a conceptual model within certain specified limits or ranges of application and corresponding ranges of accuracy.

*Model calibration*: The procedure of adjustment of parameter values of a model to reproduce the response of reality within the range of accuracy specified in the performance criteria.

*Model validation:* Substantiation that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model.

*Model set-up*: Establishment of a site-specific model using a model code. This requires, among other things, the definition of boundary and initial conditions and parameter assessment from field and laboratory data.

*Simulation*: Use of a validated model to gain insight into reality and obtain predictions that can be used by water managers. This includes insight into how reality can be expected to respond to human interventions. In this connection uncertainty assessments of the model predictions are very important.

*Performance criteria*: Level of acceptable agreement between model and reality. The performance criteria apply both for model calibration and model validation.

*Domain of applicability (of conceptual model)*: Prescribed conditions for which the conceptual model has been tested, i.e. compared with reality to the extent possible and judged suitable for use (by model confirmation).

*Domain of applicability (of model code)*: Prescribed conditions for which the model code has been tested, i.e. compared with analytical solutions, other model codes or similar to the extent possible and judged suitable for use (by code verification).

*Domain of applicability (of model)*: Prescribed conditions for which the site-specific model has been tested, i.e. compared with reality to the extent possible and judged suitable for use (by model validation).

### 2.2.3 Scientific philosophical aspects

The credibility of the descriptions or the agreements between reality, conceptual model, model code and model are evaluated through the terms confirmation, verification, calibration and validation. Thus, the relation between reality and the scientific description of reality which is constituted by the conceptual model with its theories and equations on flow and transport processes, its interpretation of the geological system and ecosystem at hand, etc., is evaluated through the confirmation of the conceptual model. By using the term confirmation in connection with conceptual model, it is implied that it is never considered possible to prove the truth of a theory/hypothesis and as such of a conceptual model. And even if a site-specific model is eventually accepted as valid for specific conditions, this is not a proof that the conceptual model is true, because, due to non-uniqueness, the site-specific model may turn out to perform right for the wrong reasons.

The fundamental view expressed by scientific philosophers is that verification and validation of numerical models of natural systems is impossible, because natural systems are never closed and because the mapping of model results are always non-unique (Popper, 1959; Oreskes et al., 1994). I agree that it is not possible to carry out model verification or model validation, if these terms are used universally, without restriction to domains of applicability and levels of accuracy.

[12] note, however, that Popper (1959) distinguished between two kinds of universal statements: the 'strictly universal' and the 'numerical universal'. The strictly universal statements are those usually dealt with when speaking about theories or natural laws. They are a kind of 'all-statement' claiming to be true for any place and any time. In contrary, numerical universal statements refers only to a finite class of specific elements within a finite individual spatio-temporal region. A numerical universal statement is thus in fact equivalent to conjunctions of singular statements.

The restrictions in use of the terms confirmation, verification and validation imposed by the respective domains of applicability imply, according to Popper's views, that the conceptual model, model code and site-specific models can only be classified as numerical universal statements as opposed to strictly universal statements. This distinction is fundamental for the terminology described in [12] and its link to scientific philosophical theories. Consequently the terms verification and validation should never be used without qualifiers.

An important aspect of the framework outlined in [12] lies in the separation between the three different ‘versions’ of the word model, namely the conceptual model, the model code and the-site specific model. Due to this distinction it is possible, at a general level, to talk about confirmation of a theory or a hypothesis about how nature can be described using the relevant scientific method for that purpose, and, at a site-specific level, to talk about validity of a given model within certain domains of applicability and associated with specified accuracy limits.

## 2.3 Modelling Protocol

The procedure for applying a hydrological model is often denoted a modelling protocol. It comprises a series of actions to be followed in a sequential or iterative form. The modelling protocol presented in [7] for distributed catchment modelling was inspired by the groundwater community (Anderson and Woessner, 1992). It was subsequently used in the Danish Handbook for Groundwater Modelling (Henriksen et al., 2001) that has been used extensively in practise since its emergence. A more recent modelling protocol, developed within the context of the EU research project HarmoniQuA, is reported in [13] and Scholten et al. (2007). The two protocols are illustrated in Figs. 4 and 5.

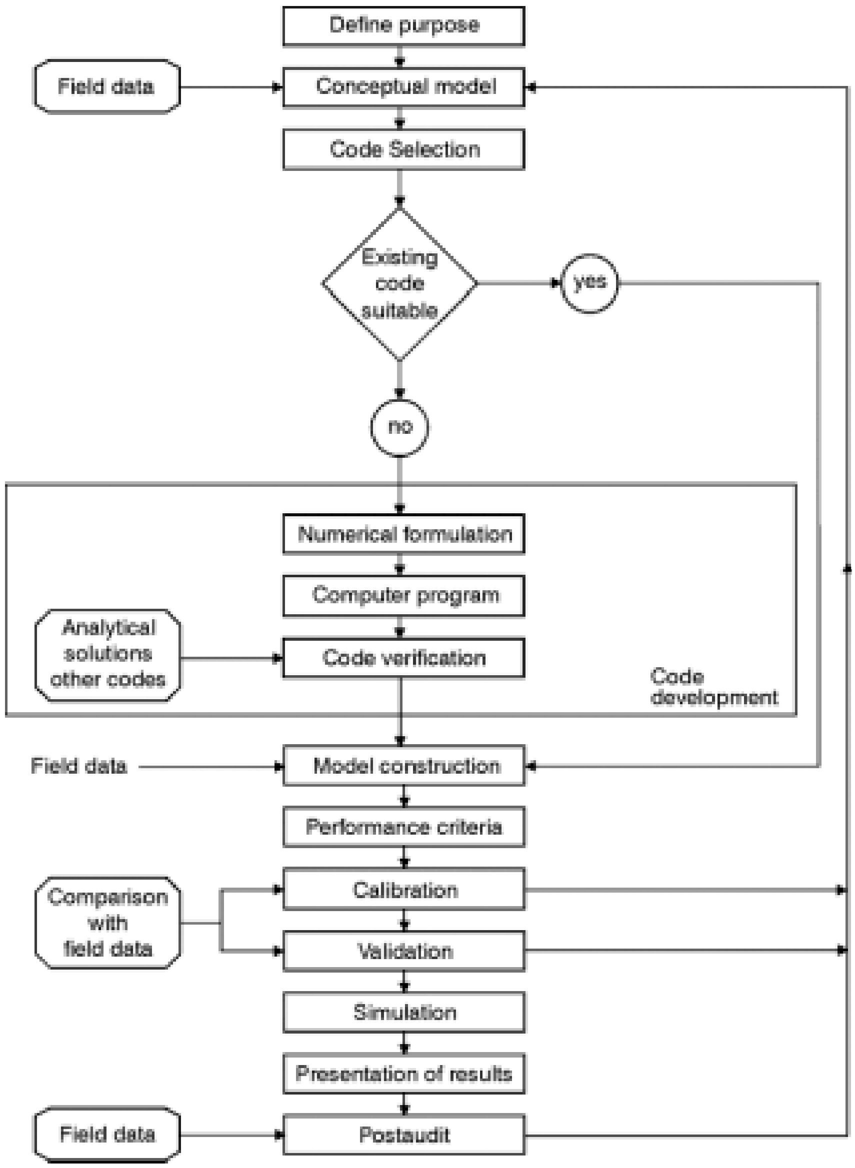



*Fig. 4 The modelling protocol from [7].*

A modelling study will involve several phases and several actors. A typical modelling study will involve the following four different types of actors:

- *The water manager*, i.e. the person or organisation responsible for the management or protection of the water resources, and thus responsible for the modelling study and the outcome (the problem owner).
- *The modeller*, i.e. a person or an organisation that works with the model conducting the modelling study. If the modeller and the water manager belong to different organisations, their roles will typically be denoted consultant and client, respectively.
- *The reviewer*, i.e. a person that is conducting some kind of external review of a modelling study. The review may be more or less comprehensive depending on the requirements of the particular case. The reviewer is typically appointed by the water manager to support the water manager to match the modelling capability of the modeller.
- *The stakeholders/public*. A stakeholder is an interested party with a stake in the water management issue, either in exploiting or protecting the resource. Stakeholders include the following different groups: (i) competent water resource authority (typically the water manager, cf. above); (ii) interest groups; and (iii) general public.

The modelling process may, according to [13], be decomposed into five major steps which again are decomposed into 48 tasks (Fig. 5). The contents of the five steps are:

- *STEP1 (Model Study Plan).* This step aims to agree on a Model Study Plan comprising answers to the questions: Why is modelling required for this particular model study? What is the overall modelling approach and which work should be carried out? Who will do the modelling work? Who should do the technical reviews? Which stakeholders/public should be involved and to what degree? What are the resources available for the project? The water manager needs to describe the problem and its context as well as the available data. A very important task is then to analyse and determine the various requirements of the modelling study in terms of the expected accuracy of modelling results. The acceptable level of accuracy will vary from case to case and must be seen in a socio-economic context. It should, therefore, be defined through a dialogue between the modeller, water manager and stakeholders/public. In this respect an analysis of the key sources of uncertainty is crucial in order to focus the study on the elements that produce most information of relevance to the problem at hand.
- *STEP 2 (Data and Conceptualisation).* In this step the modeller should gather all the relevant knowledge about the study basin and develop an overview of the processes and their interactions in order to conceptualise how the system should be modelled in sufficient detail to meet the requirements specified in the Model Study Plan. Consideration must be given to the spatial and temporal detail required of a model, to the system dynamics, to the boundary conditions and to how the model parameters can be determined from the available data. The need to model certain processes in alternative ways or to differing levels of detail in order to enable assessments of model structure uncertainty should be evaluated. The availability of existing computer codes that can address the model requirements should also be addressed.
- *STEP 3 (Model Set-up).* Model Set-up implies transforming the conceptual model into a site-specific model that can be run in the selected model code. A major task in Model Set-up is the processing of data in order to prepare the input files necessary for executing the model. Usually, the model is run within a Graphical User Interface (GUI) where many tasks have been automated. The GUI speeds

up the generation of input files, but it does not guarantee that the input files are error free. The modeller performs this work.

- *STEP 4 (Calibration and Validation).* This step is concerned with the process of analysing the model that was constructed during the previous step, first by calibrating the model, and then by validating its performance against independent field data. Finally, the reliability of model simulations for the intended domain of applicability is assessed through uncertainty analyses. The results are described so that the scope of model use and its associated limitations are documented and made explicit. The modeller performs this work.
- *STEP 5 (Simulation and Evaluation).* In this step the modeller uses the calibrated and validated model to make simulations to meet the objectives and requirements of the model study. Depending on the objectives of the study, these simulations may result in specific results that can be used in subsequent decision making (e.g. for planning or design purposes) or to improve understanding (e.g. of the hydrological/ecological regime of the study area). It is important to carry out suitable uncertainty assessments of the model predictions in order to arrive at a robust decision. As with the other steps, the quality of the results needs to be assessed through internal and external reviews.

Each of the last four steps is concluded with a reporting task followed by a review task. The review tasks include dialogues between water manager, modeller, reviewer and, often, stakeholders/public. The protocol includes many feedback possibilities (Fig. 5).

A comparison of the old protocol (Fig. 4) and the one decade younger HarmoniQuA protocol (Fig. 5) shows some interesting developments:

- The basic sequence of the prescribed activities in the protocols is the same. The HarmoniQuA protocol is much more detailed than the old one, but there are no fundamental disagreements between the two.
- The HarmoniQuA protocol puts much more emphasis on the framing of the modelling study. This is only considered in one box in Fig. 4 and not given much weight in [7], while it is one full Step comprising seven tasks in Fig 5. This implies for instance that requirements on performance criteria and uncertainty assessments are introduced rather late in the old protocol, while it is an important part of Step 1 in the HarmoniQuA protocol.
- There is much emphasis on uncertainty assessments throughout the modelling process in the HarmoniQuA protocol, while uncertainty assessments are only considered as part of model calibration and simulation in the old protocol.
- The HarmoniQuA protocol is part of a quality assurance framework with much emphasis on the role play between the various actors in the modelling process. This results in stakeholder involvement, peer reviews, focus on reporting and dialogue between water manger and modeller. In contrary to this, the old protocol only focuses on the modeller.

These developments reflect a process from guidance to the modeller only (old protocol) towards guidance to all actors involved in the modelling process (HarmoniQuA). This process has been inspired by feedbacks from introducing the old protocol to real world applications, where it was realised that a broader concept was required.

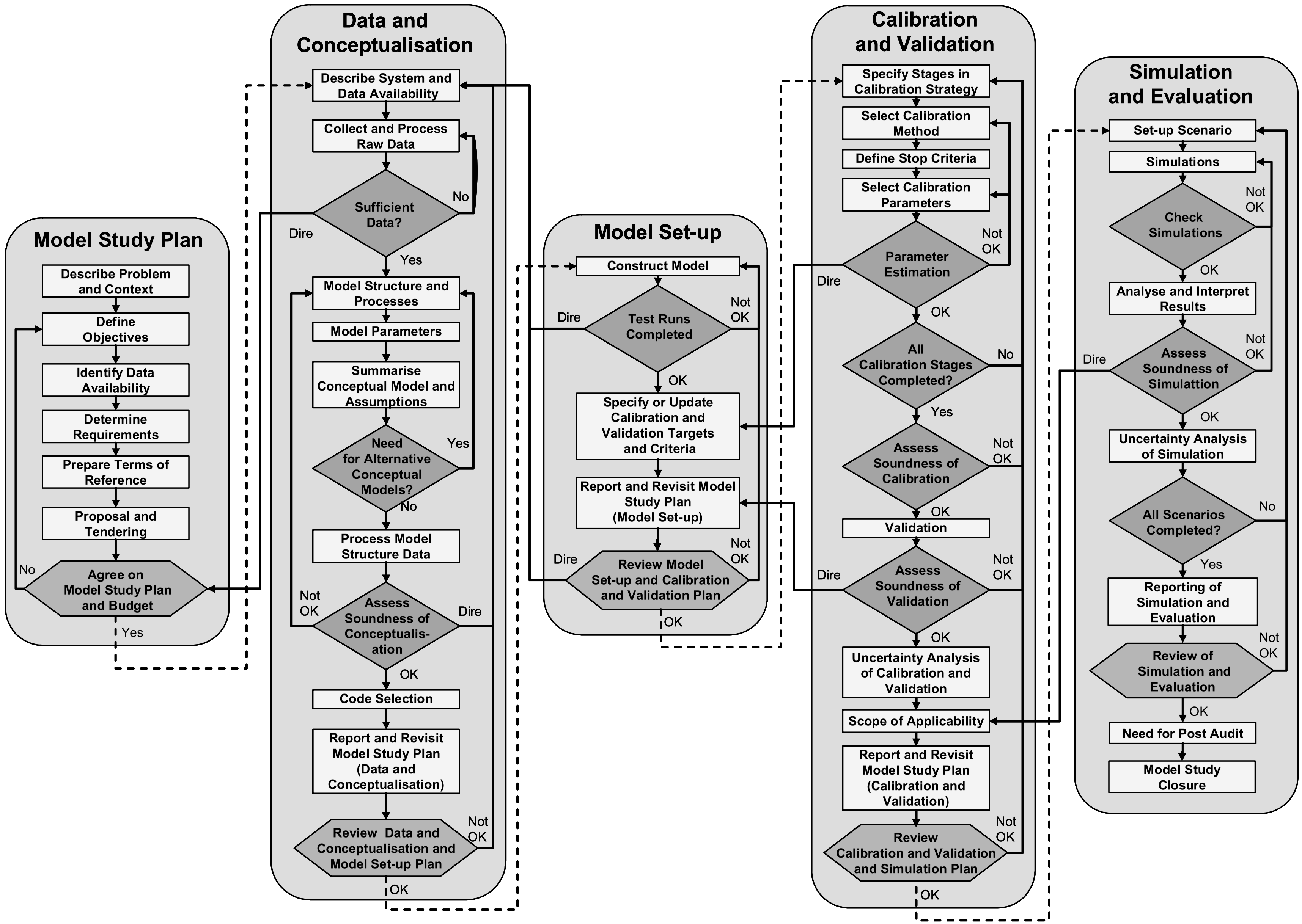



*Fig. 5 The five modelling steps and the 48 tasks in the HarmoniQuA modelling protocol. The diagram is an updated version of Fig. 5 in [13] (Refsgaard et al., 2006).*

## 2.4 Classification of Models

Many attempts have been made to classify hydrological models (or model codes). Refsgaard (1996) presented the classification shown in Fig. 6 that I have used in all papers of the present thesis. Deterministic models can be classified according to whether the model gives a lumped or a distributed description of the considered area, and whether the description of the hydrological processes is empirical, conceptual, or more physically-based. A lumped model implies that the catchment is considered as one computational unit. A distributed model, on the other hand, provides a description of catchment processes at geo-referenced computational grid points within the catchment. An intermediate approach is a semi-distributed model, which uses some kind of distribution, either in sub-catchments or in hydrological response units, where areas with the same key characteristics are aggregated to sub-units without considering their actual locations within the catchment. Examples of hydrological response units considered in semi-distributed models are elevation zones, which are relevant for snow modelling, and combinations of soil and vegetation type, which may be relevant for simulation of root zone processes such as evapotranspiration and nitrate leaching.

As most conceptual models are also lumped, and as most physically-based models are also distributed, the three main classes emerge:

- Empirical (black box)
- Lumped conceptual models (grey box)
- Distributed physically-based (white box)

The classification is discussed in some details in Refsgaard (1996). Here, the focus is on the two traditional approaches in deterministic hydrological catchment modelling, namely the lumped conceptual and the distributed physically-based ones. The fundamental difference between these two types of models lies in their process descriptions and the way spatial variability is treated. The distributed physically-based models contain equations which have originally been developed for point scales and which provide detailed descriptions of flows of water and solutes. The variability of catchment characteristics is accounted for explicitly through the variations of hydrological parameter values among the different computational grid points. This approach leaves the variability within a grid as un-accounted for, which in some cases is of minor importance but in other cases may pose a serious constraint. The lumped conceptual models uses empirical process descriptions, which have built-in accounting for the spatial variability of catchment characteristics.

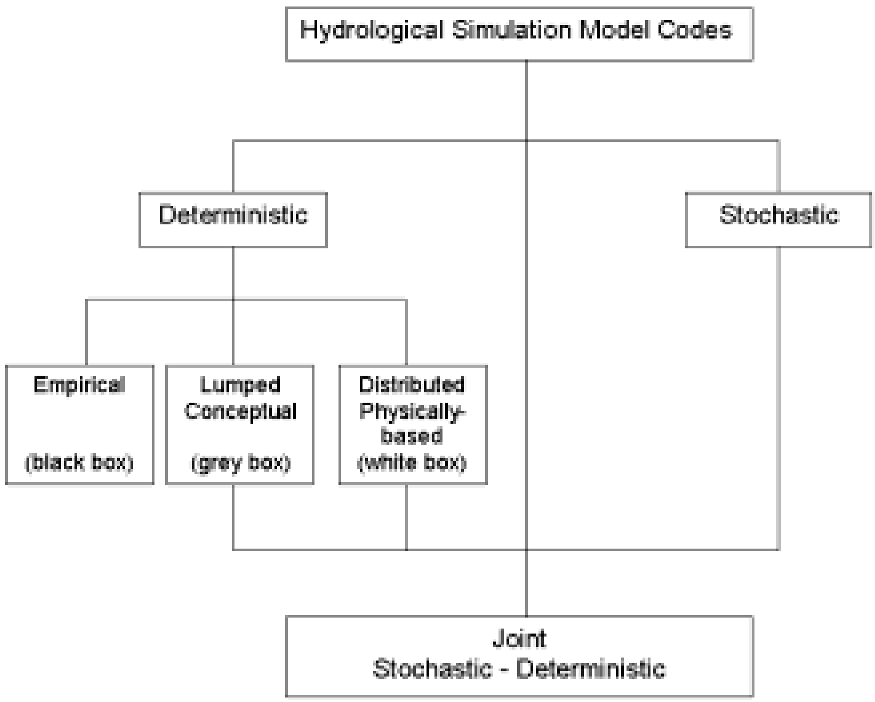



*Fig. 6 Classification of hydrological models according to process description (Refsgaard, 1996).*

Typical examples of lumped conceptual model codes are the Stanford Watershed Model (Crawford and Linsley, 1966), the Sacramento (Burnash, 1995), the HBV (Bergström, 1995) and the NAM (Nielsen and Hansen, 1973). Typical examples of distributed physically-based model codes are the MIKE SHE (Abbott et al., 1986a, b; Refsgaard and Storm, 1995) and the Thales (Grayson et al., 1992a, b). Groundwater model codes like MODFLOW belong to the distributed physically-based class.

The classification has some shortcomings that should be noted. First of all, the use of the term 'conceptual model' is unfortunate, because this is a different meaning of the term as compared to the definition given in Section 2.2 and used in the modelling protocols (Section 2.3). This can cause some confusion, but to introduce a new term completely different from what is used by almost all other scientists in the community of catchment modelling may cause even more confusion. Secondly, and more fundamental, the names of the classes should be considered as relative rather than absolute. For example Beven (1989) argued that in most applications physically-based models are used as lumped conceptual models at the grid scale. As discussed in [4] I agree that some degree of lumping and conceptualisation will always need to take place, but that in spite of this there is a fundamental difference in the functioning and, as shall also be discussed later, of the applicability of the two model types.

# 3 Simulation of Hydrological Processes at Catchment Scale

In this chapter some modelling examples from the publications are briefly summarised and discussed within the framework outlined in Chapter 2.

## 3.1 Flow modelling

### 3.1.1 Groundwater/surface water model for the Suså catchment ([1], [2])

**Summary**

The publications [1] and [2] describe a new model code and the set-up, calibration and validation of a model for a 1,000 km$^2$ area. Further details can be found in Stang (1981), Refsgaard (1981) and Refsgaard and Stang (1981). The objectives of the study were to develop a spatially distributed groundwater/surface water model code and apply it to the Suså catchment with a particular focus on the stream-aquifer interaction in a hydrogeological system consisting of confined aquifer-aquitard-phreatic aquifer and to test the model for prediction of the hydrological consequences on streamflows and hydraulic heads of groundwater abstraction.

The new model code was rather complex and computationally demanding at the time of development. Thus, standard 30 years model simulations could only be carried out as night runs at the main frame computer at DTU's computer centre.

The model area comprising the Suså and the neighbouring Køge Å catchments is located in the central and southern part of Zealand. The model area, the topographic divides and the groundwater model polygonal mesh are shown in Fig. 7. The overall structure of the model is outlined in Fig. 8. It consists of four separate components for the confined regional aquifer, the aquitard, the phreatic aquifer and the root zone. The spatial distribution and the degree of physical basis differ between the four components. The time steps in the calculations are one day in all parts of the model.

The confined aquifer is described by a two-dimensional integrated finite difference model with 112 polygons. For the phreatic aquifer consisting of till with very small transmissivities and for the aquitard each of the polygons are distributed further into four sub-polygons based on hypsographic curves (Fig. 9). Due to small scale topographic variations the flows in the aquitard in most polygons are upwards in some parts and downwards in other parts of the polygon. A correct representation of these flows between the regional aquifer and the phreatic aquifer that discharges the rivers is crucial for achieving a good description of the stream-aquifer interaction. Without such approach allowing a description of both upwards and downwards flows in the aquitard within the same polygon a much finer spatial resolution with 10-100 times as many polygons would have been required. This would have been impossible 25 years ago due to computational constraints.

The root zone component calculated the net precipitation that recharged the phreatic aquifer. The modelling area was divided into seven sub-areas with separate precipitation input and soil parameters. Further the spatial variation in vegetation was accounted for by dividing each of these seven areas into five vegetation areas based on agricultural statistics and one meadow (wetland) area. This makes the total distribution to 42 sub-areas where each sub-area is a kind of 'hydrological response unit', i.e. a semi-distributed approach. The root zone calculations were based on a box approach with four layers in the root zone.

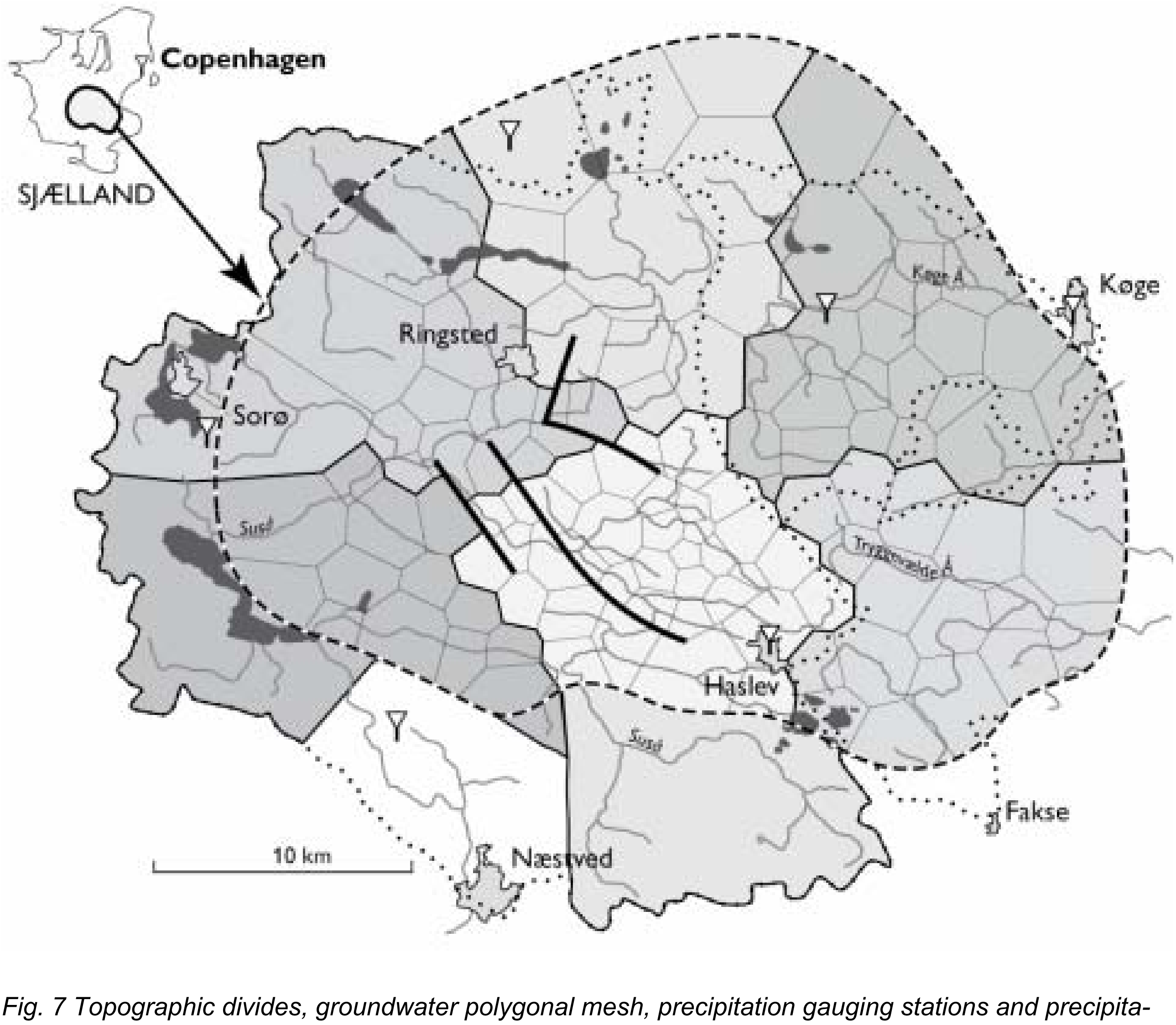



*Fig. 7 Topographic divides, groundwater polygonal mesh, precipitation gauging stations and precipitation zones of the Suså model.*

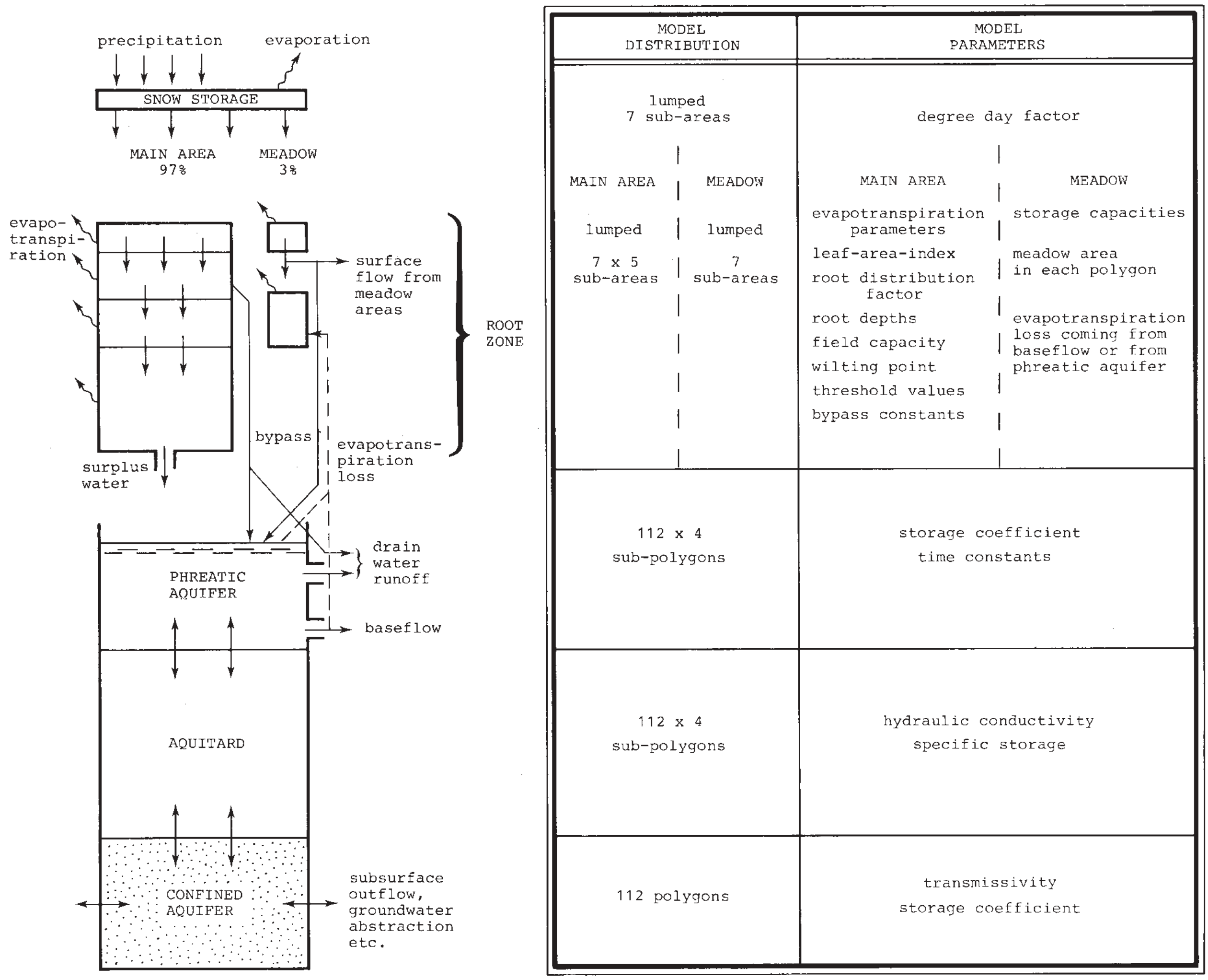



*Fig. 8 The structure of the Suså model*

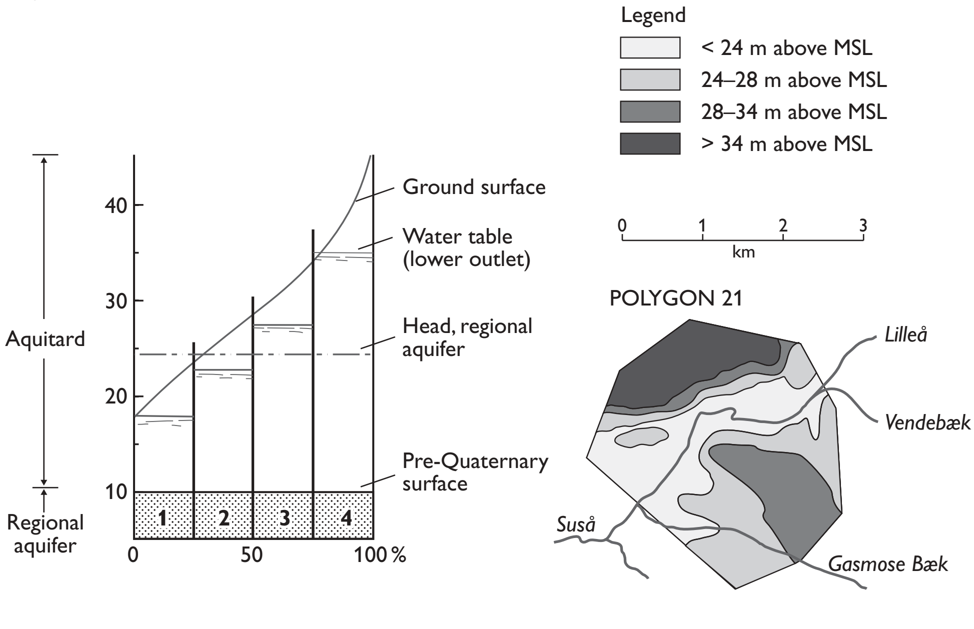



*Fig. 9 Hypsographic curve for polygon 21 and areas represented by the four sub-polygons.*

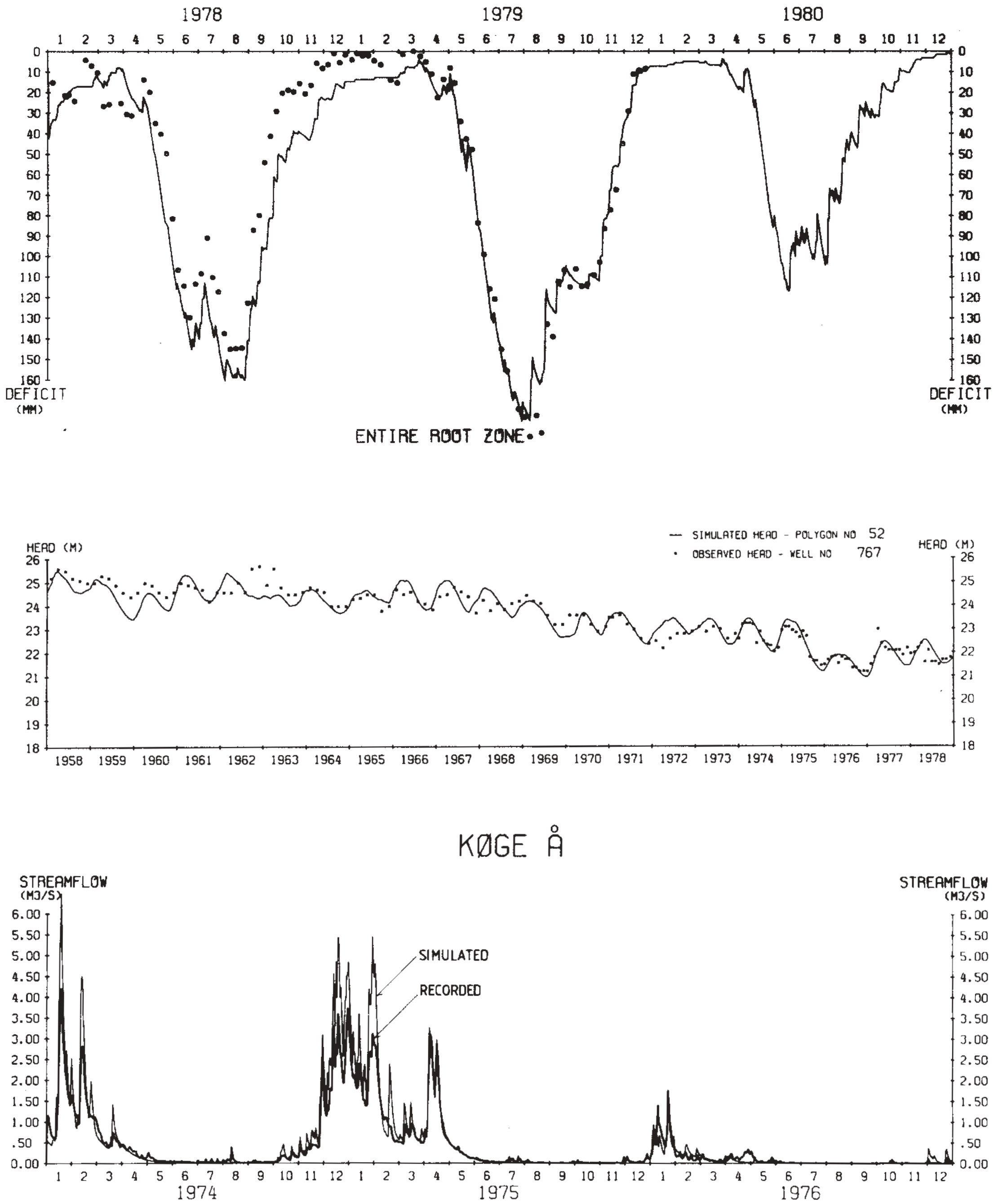



*Fig. 10 Examples of simulation results from soil moisture in root zone, hydraulic head of regional confined aquifer and river discharge.*

The model was calibrated against soil moisture data from four experimental plots, time series of hydraulic heads from 40 observation wells in the regional aquifer and streamflow from six gauging stations. Examples of simulation results from the calibration period are shown in Fig. 10 which shows excellent curve fits. The groundwater and aquitard models were calibrated, along with the code development itself, using all available hydraulic head data from the period 1950-80. Between 1964 and 1970 the groundwater abstraction to Copenhagen Water Supply from the Regnemark Waterworks in the Køge Å catchment was increased from zero to about 15 million $m^3$/year. The remaining model components were calibrated against only some of the available streamflow data, namely some of the data from the Suså catchment, while amongst others Køge Å data were not used for calibration.

While the simulation of streamflows in the Køge Å catchment in [1] was characterised as a "half-way test of the model's ability to simulate streamflow from ungauged catchments" no systematic validation tests against independent data were carried out as part of the study. Some years later the model simulations were extended with new data from the period 1981-87, where the groundwater abstractions had changed slightly. In this post audit validation study the model simulations were found to match the observations to the same degree of accuracy as during the calibration period (Jensen and Jørgensen, 1988).

The model's ability to simulate the streamflow depletion caused by a groundwater abstraction from the regional confined aquifer was tested on historical data from the Køge Å catchment. Fig. 11 shows simulated streamflow assuming actual groundwater abstraction from the Regnemark Waterworks starting in 1964, $Q_{sim}$, and assuming no abstracting from Regnemark, $Q^1_{sim}$. The recorded streamflow fits reasonably well with $Q_{sim}$. The difference $Q^1_{sim}$ - $Q_{sim}$, which is the simulated streamflow depletion caused by the increased groundwater abstraction, is seen to have a clear seasonal variation with smaller depletion during the dry summer periods and larger depletion during the wet winter season.

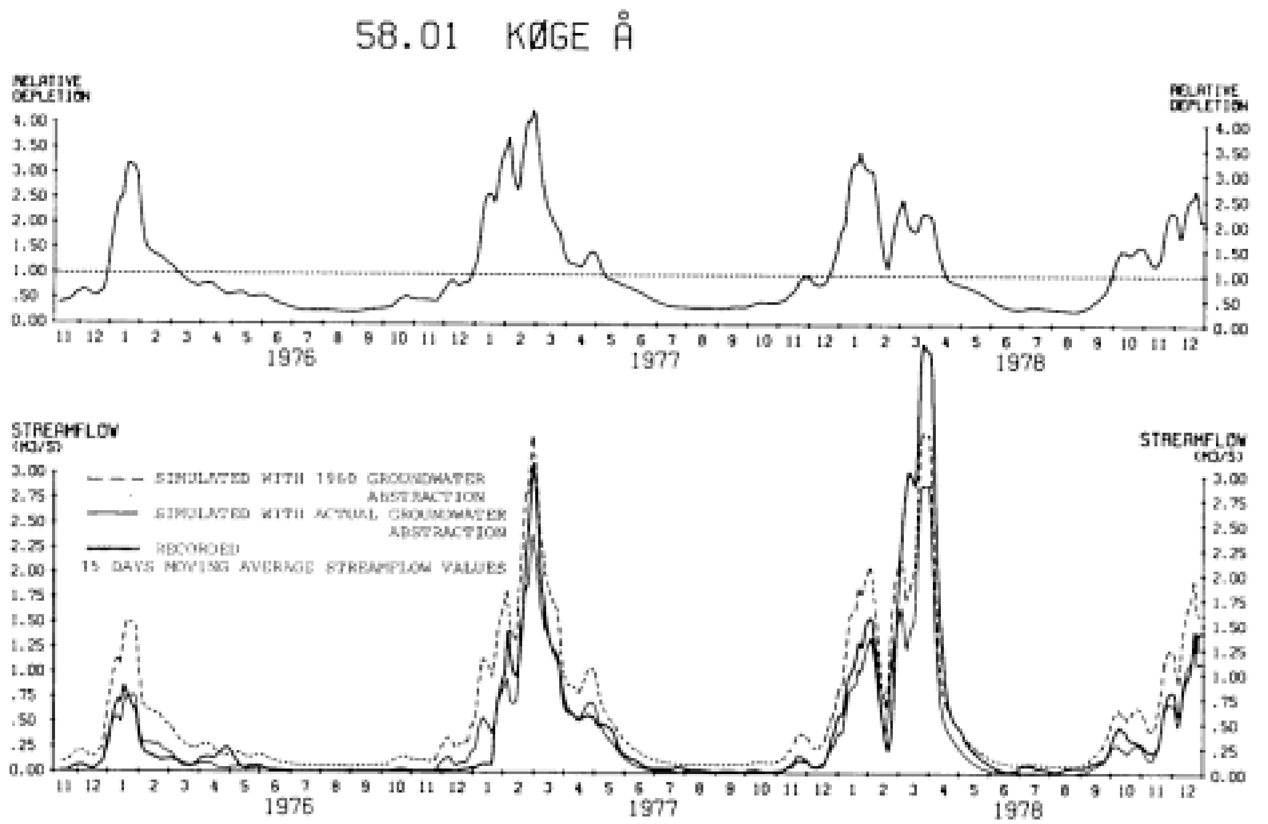



*Fig. 11 Comparison of 15 days moving average streamflows for Køge Å (lower) and the relative streamflow depletion caused by the groundwater abstraction (upper)*

**Discussion - post evaluation**
Most other catchment models existing when the Suså model code was developed were either purely rainfall runoff models of the lumped conceptual type, such as the classical Stanford Watershed Model (Crawford and Linsley, 1966), the HBV (Bergström and Forsman, 1973; Bergström, 1976) and the NAM (Nielsen and Hansen, 1973) or purely groundwater models (Prickett and Lonnquist, 1971; Thomas, 1973). A few authors had concluded that coupled groundwater/surface water modelling was essential (e.g. Luckner, 1978; Lloyd, 1980) and some had outlined specific, but not yet operational, concepts (e.g. Freeze and Harlan, 1969; Wardlaw, 1978; Jønch-Clausen, 1979). In some studies groundwater models and rainfall-runoff models were used at the same catchment, but without coupling (e.g. Weeks et al., 1974). Thus, apparently no other model had previously been used to dynamically simulate coupled groundwater/surface water conditions at catchment scale (rainfall, evapotranspiration, surface near runoff, groundwater recharge, groundwater heads, baseflow discharge from aquifers to streams).

During the decade following [1] and [2] a few model codes with integrated groundwater/surface water descriptions emerged. The most prominent of these codes was the SHE (Abbott et al., 1986a, b) and its operational daughter codes, MIKE SHE from DHI (Refsgaard and Storm, 1995) and SHETRAN from University of Newcastle (Bathurst and O'Connell, 1992), which both are used today, although in later versions. Other operational models from that period were described by Miles and Rushton (1983), Christensen (1994) and Wardlaw (1994). Miles and Rushton (1983) used a simpler root zone and surface water component than [1] together with a two-dimensional finite difference groundwater model and monthly time steps. Christensen (1994) developed a model for the Tude Å catchment (a neighbour to Suså) that conceptually was similar and a little bit simpler than [1]. Wardlaw et al. (1994) used the concepts outlined in Wardlaw (1978) coupling the Stanford Watershed Model with a finite-difference groundwater model and a channel routing model for simulation of discharge and groundwater levels in the Allen catchment in England.

During the past decade the number of integrated modelling codes has exploded. The existing codes today can be considered to fall in three classes: (a) fully integrated codes such as MIKE SHE (Graham and Butts, 2005); (b) couplings of existing groundwater codes and surface water codes such as MODFLOW and SWAT (Perkins and Sophocleous, 1999); and (c) codes based on the fully 3-dimensional Richards' equation (Panday and Hayakorn, 2004). Independent reviews of the scientific basis and practical applicability of a number of recent integrated model codes are provided by e.g. Kaiser-Hill (2001) and Tampa Bay Water (2001).

A major novelty of [1] and [2] was that the Suså model code was one of the first codes, which integrated surface water and groundwater descriptions, and the first of its kind applied operationally to moraine landscapes. The model results were unique with respect to simulation of the dynamics of the groundwater/surface water interaction, as for instance reflected by the annual hydraulic head fluctuations and the streamflow depletion due to the groundwater abstraction. Furthermore the study provided new insights and understanding on the mechanisms that governed streamflow depletion due to groundwater abstraction from confined aquifers in moraine catchments. In contrary to the traditional type curve analyses which were used extensively in hydrogeology to analyse test pumpings and to predict the effects of abstractions, [1] and [2] were based on non-stationary analysis which, as evident from the annual variations of streamflow depletion shown in Fig. 11, turns out to be crucial. The only modelling study from the following decade that considered the dynamics of the stream-aquifer interaction in moraine catchments in connection with groundwater abstraction was Christensen (1994) who basically confirmed the results of [2].

The spatial distribution and the degree of physical basis differ between the four components of the Suså model. The groundwater model can be characterised as distributed physically-based, the aquitard model as semi-distributed physically-based and the phreatic aquifer and root zone models as semi-distributed conceptual. In contrary to for instance the later SHE code (Abbott et al, 1986a, b), the Suså model code was not generic, because it could not be applied to other catchments without changes in the code. Furthermore, it was tailored to the specific hydrological conditions prevailing in the Suså catchment and could for instance not be applied to an alluvial unconfined aquifer.

In retrospect, it is interesting to observe that issues related to the credibility of model simulations were not critically analysed or discussed in [1] and [2]. First of all, aspects of code verification were not dealt with in the publications, although a major novelty of the work was the development of a completely new code. Secondly, and maybe more surprisingly, model validation and uncertainty assessments of model simulations were almost not addressed. By using all the available groundwater head data for calibration the opportunity to make split-sample validation test against parts of the data or even the unique opportunity to calibrate on data before the groundwater abstraction and validate on data after the abstraction (differential split-sample test according to Klemes (1986)) were not utilised. By not addressing the uncertainty and by not conducting rigorous validation tests the reader may be left with the, undocumented, impression that the curve fitting in Fig. 10 is supposed to reflect the predictive capability of the model. That the model proved to perform well in a subsequent post-audit validation study could not be known at the time of [1] and [2].

The other integrated groundwater/surface water modelling studies from the following decade (Miles and Rushton, 1983; Christensen, 1994; Wardlaw, 1994) had the same characteristics, i.e. only focus on calibration and model prediction but no mentioning of verification of the new model codes, no model validation tests against independent data and no uncertainty assessments. The SHE study reported by Bathurst (1986a, b) focussing on surface water hydrology did include split-sample validation testing and sensitivity analysis. For surface water (rainfall-runoff) modelling studies focusing more on model applications than code developments split-sample testing was more common (e.g. Bergström, 1976; WMO, 1975; WMO 1988) but uncertainty assessment was not systematically carried out and usually not even considered until Beven called for it (Beven, 1989; Beven and Binley, 1992). Altogether, this illustrates a very significant development in the modelling practise during these three decades.

### 3.1.2 Application of SHE to catchments in India ([4], [5])

**Summary**
The publications [4] and [5] describe the set-up, calibration and validation of the 'Système Hydrologique Européen' (SHE) code to six sub-catchments totalling about 15,000 km$^2$ of the Narmada basin in India, Fig. 12. The objective of the papers was to describe experiences from applying a distributed physically-based code like SHE to large basins with rather limited data coverage compared to previous SHE applications to research catchments. In contrary to the Suså study in [1] and [2], the India study did not include any code development, except for data processing utility software. Instead it comprised application of an existing code (Abbott et al., 1986a,b) to conditions that were far beyond the conditions for which the SHE had previously been tested in terms of catchment size, data coverage and hydrological regime (Bathurst, 1986a).

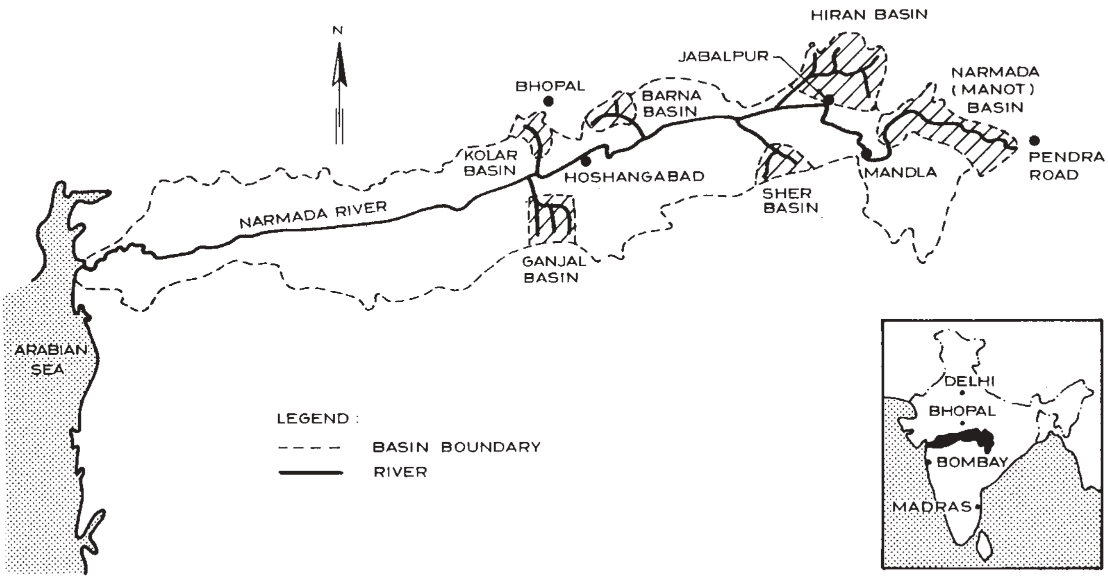



*Fig. 12 Location map for the Narmada and the six sub-catchments.*

Applicationwise, the study focused on simulation of catchment runoff, i.e. surface water aspects only. The model structure was as illustrated in Fig. 17. The groundwater zone was, however, considered only with one layer, i.e. a 2-dimensional groundwater model, and there were no data from observation wells to allow a calibration of the groundwater part of the model. The six models were set-up with a 2 km x 2 km computational grid. A split-sample approach was used with typically three years for model calibrations and other three years for the subsequent model validation.

The data requirements for a SHE based model is substantial and much larger than for a rainfall-runoff model of lumped conceptual type that previously had been applied to such types of catchments. A major challenge of the study was therefore to identify, collect and process data and to check their quality. Data were collected from more than 15 different agencies belonging to many different ministries and the data quality varied substantially.

Another challenge was how to assess parameter values in a distributed model when data, in contrary to the previous tests on small experimental catchments like in Bathurst (1986a), are scarce. Each of the grid points in a distributed model is characterised by one or more parameters. Although the parameter values in principle (as in nature) vary from grid point to grid point, it is neither feasible nor desirable to allow the parameter values to vary so freely. Instead, a given parameter should only reflect the significant and systematic variation described in the available field data. Therefore a parameterisation procedure was developed, where representative parameter values were associated to individual soil types, vegetation types, geological layers, etc. This process of defining the spatial pattern of parameter values effectively reduced the number of free parameter coefficients, which needs to be adjusted in the subsequent calibration procedure. For example, the 820 km$^2$ Kolar catchment is parameterised into three soil classes and 10 land use/soil depth classes. For the soil type classes calibration was allowed for the hydraulic conductivity in the unsaturated zone (for each soil type class the conductivity could vary among three different land uses => nine parameter values). For the land use/soil depth classes the calibration parameters comprised soil depths (10 parameters in total) and the Strictler overland flow coefficients for four land use types (four parameters in total). Further three parameters were subject to calibration (hydraulic conductivity in the saturated zone, an (empirical) by-pass coefficient and a surface retention parameter; all kept constant throughout the catchment). Although the 26 calibration parameters could not be assessed from field data alone, but had to be modified through calibration, the physical realism of the parameter values resulting from the subsequent calibration procedure could be evaluated from available field data.

The simulation results are illustrated in Fig. 13 as hydrographs for the largest sub-catchment and in Fig. 14 as annual runoff and annual peaks for all six sub-catchments. In both figures the results are for the validation periods, where results are slightly poorer as compared to the calibration periods. In [4] the rainfall-runoff simulation results were characterised as having the same degree of accuracy as would have been expected with simpler hydrological models of the lumped conceptual type. The results therefore suggested that application of complex data demanding models like the present SHE approach are not justified in cases where the modelling objective is limited to simulation of catchment runoff and where observed runoff records exist for calibration purposes. No attempts were made in the study to test the capability of a model without calibration.

After the first calibration and validation tests had been made, field investigations were carried out in the Kolar catchment during a 2½ week period to improve the parameter estimates, mainly for soil and vegetation parameters, and to evaluate the importance of additional field data. Subsequently, the Kolar model was recalibrated in such a way that rather narrow constraints were put on the range of values allowed for the key parameters. The final model, based on the additional data, produced simulation results of same quality as the preliminary model with respect to simulated hydrograph. Although it is argued in [5] that the final model is believed to give an improved physical representation of the hydrological regime, it is concluded that a good match between observed and simulated outlet hydrographs does not provide a sufficient guarantee of a hydrologically realistic process description.

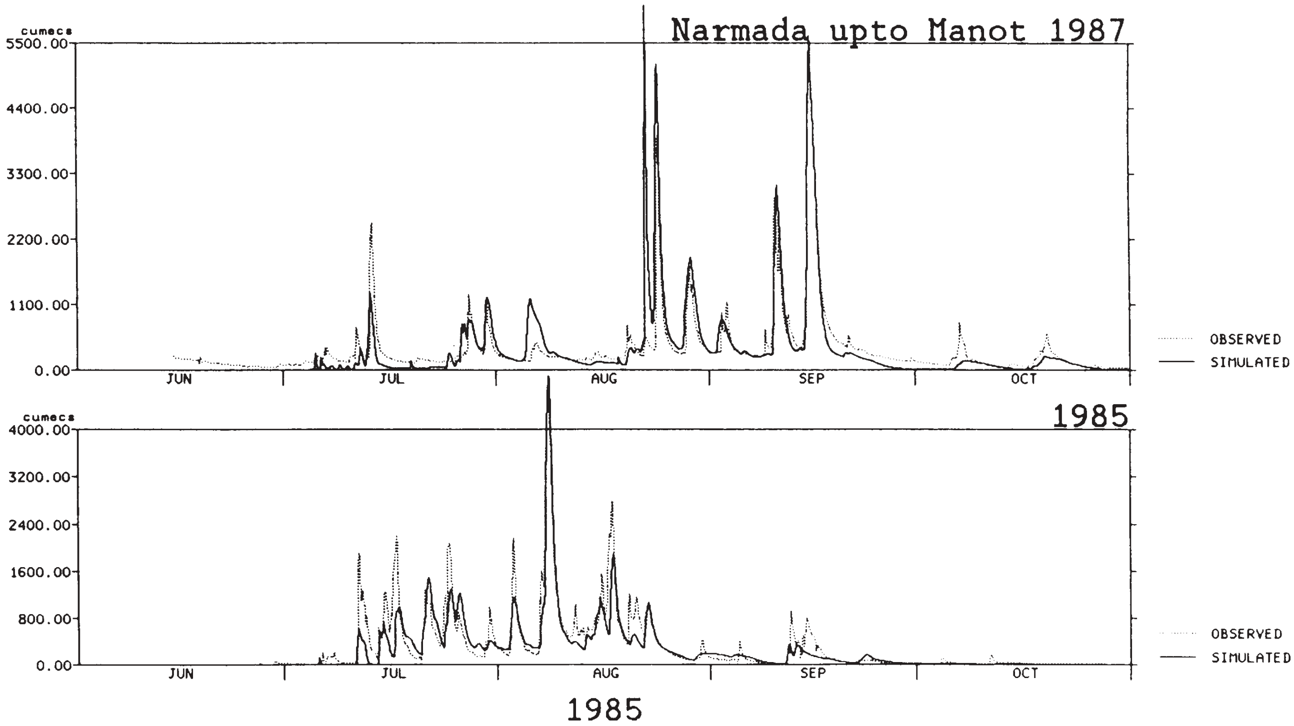



*Fig. 13 Observed and simulated hydrographs for the Narmada at Manot during the validation period 1985 and 1987.*

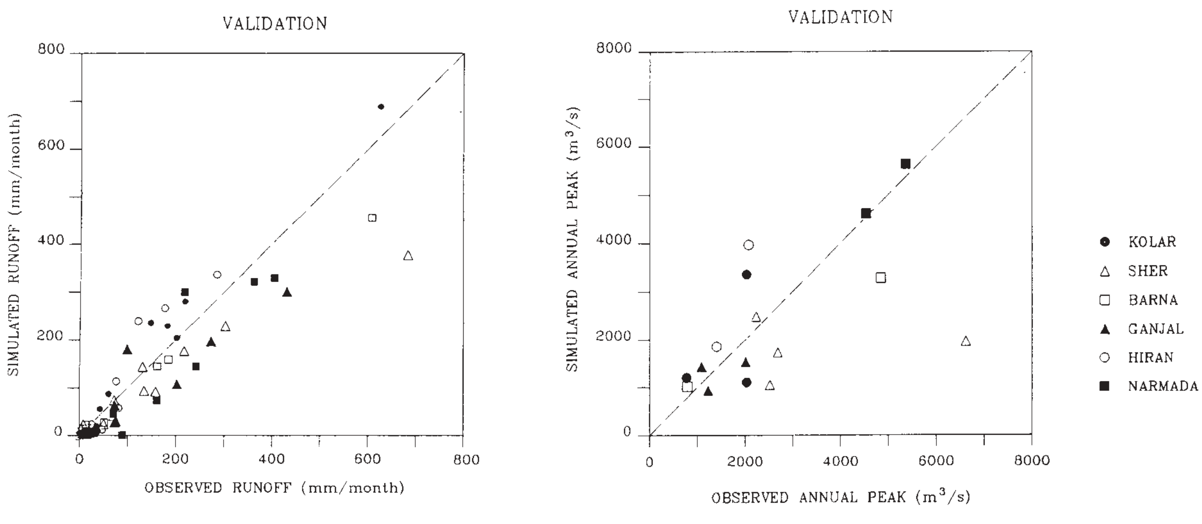



*Fig. 14 Simulated monthly runoff during monsoon season (left) and simulated annual peak discharge compared with measured values during validation periods for all six sub-catchments.*

**Discussion - post evaluation**

At the time of [4] and [5] lumped conceptual catchment model codes such as HBV (Bergström, 1992) and NAM (Jønch-Clausen and Refsgaard, 1984) had been used operationally for two decades, typically for catchments ranging from a few km$^2$ to more than 10,000 km$^2$.

At the same time distributed physically-based models had mainly been tested on flood events on small catchments that typically had very good data due to experimental instrumentation (Loague and Freeze, 1985; Bathurst 1986a; Grayson et al., 1992a,b; Troch et al., 1993). Loague and Freeze (1985) compared a quasi-physically based model with a regression model and a unit hydrograph model on three experimental catchments, the 0.1 km$^2$ R-5, Chickasha, Oklahoma, the 7.2 km$^2$ WE-38, Klingertown, Pensylvania and the 0.1 km$^2$ HB-6, West Thornton, New Hampshire. Bathurst (1986a) applied the SHE to the simulation of flood events for the 10.6 km$^2$ experimental Wye catchment in Wales. Grayson et al. (1992a,b) applied the THALES to the simulation of flood events for the 7.0 ha Wagga catchment in Australia and the 4.4 ha Lucky Hill catchment at the Walnut Gulch Experimental Area in Arizona. Troch et al. (1993) applied a model based on a 3-dimensional numerical solution to Richards' equation to the 7.2 km$^2$ WE-38 catchment and a 0.64 km$^2$ subcatchment.

To my knowledge the only examples until then of distributed physically-based model studies including applications on several hundred km$^2$ catchments and continuous simulation for periods of several years were the coupled groundwater/surface water models discussed in the previous section ([1]; [2]; Miles and Rushton, 1983; Christensen, 1994; Wardlaw et al., 1994) that all had distributed physically-based groundwater components and lumped (or semi-distributed) conceptual surface water components and some models such as WATBAL (Knudsen et al., 1986) that had semi-distributed surface water components and lumped conceptual groundwater components.

During the following few years a few additional catchment scale studies with continuous simulations of distributed physically-based models emerged. One example is Querner (1997) who applied the MOGROW to the 6.5 km$^2$ Hupselse Beek catchment simulating both discharge and groundwater head dynamics. Another example is Kutchment et al. (1996) who simulated surface water processes for the 3315 km$^2$ Ouse catchment. The study of Kutchment et al (1996) had many similarities with [4] and [5] with respect to model conceptualisation and conclusions.

The main scientific contribution of [4] and [5] was therefore as the first study to demonstrate that distributed physically-based models could be established for catchments of this size and with ordinary data availability. Previous studies reported in literature had either been tests on small research catchments or been models with major components of the lumped conceptual type. As outlined above, it is worth noting the different traditions in the communities that had dealt with (large scale) lumped conceptual models, (small scale) physically-based models and groundwater models, respectively. I believe that an important characteristic of the team who performed the present study ([4] and [5]) was that it comprised scientists who together had comprehensive experiences from all these communities.

Another key contribution was the parameterisation approach introduced. The point of departure for this approach, e.g. [1] and Bathurst (1986a), was an approach allowing parameter values to vary as required to fit the observed data during the calibration phase. This approach had been criticised by Beven (1989) to result in overparameterisation. The procedure resulted in 26 parameters to be calibrated for the Kolar catchment. Although this number is significantly less than e.g. the number of free parameters in [1], it is still very high and it is very likely that a sensitivity analysis would have shown that this number could easily be reduced without loss of model performance. It is interesting to note that similar parameterisation approaches reported for other catchments in 1997 ([7]) and 2001 (Andersen et al., 2001) resulted in 11 and 4 free parameters, respectively, implying that the parameterisation approach adopted in [4] and [5] were not yet finally developed.

Beven (1989) had provided a fundamental critique of the way physically-based models such as the SHE had been promoted by e.g. Abbott et al. (1986a) and Bathurst (1986a). His main critique was that the attitudes in these early SHE papers were not realistic with respect to the abilities and achievements of physically-based models. Beven pointed amongst others to the following key problems:

- The process equations are simplifications leading to model structure uncertainty.
- Spatial heterogeneity at subgrid scale is not included in the physically-based models. The current generation of distributed physically-based models are in reality lumped conceptual models.
- There is a great danger of overparameterisation if it is attempted to simulate all hydrological processes thought to be relevant and the related parameters against observed discharge data only.

As a conclusion Beven argued that for future applications attempts must be made to obtain realistic estimates of the uncertainty associated with their predictions, particularly in the case of evaluating future scenarios of the effects of management strategies.

[4] noted some of Beven's critique, acknowledging that the process representation at the 2 km x 2 km grid squares is causing significant violations of some of the process descriptions, that "some degree of lumping and conceptualisation has taken place at the grid scale" and that "scale problems are important". [4] stressed, however, that in spite of these acknowledged limitations "the present basin model is much more physically based and distributed than the traditional lumped conceptual model, where the entire catchment is represented in effect by one grid square, and where the process representations due to averaging over characteristics of topography, soil type and vegetation type are fundamentally different from the basic physical laws".

[4] and [5] concluded that the SHE is a suitable tool to support water management for conditions in India. In contrary to this, Beven (1989) had stated that the physically-based models "are not well suited to applications to real catchments". In retrospect, it is remarkable that [4] and [5] did not go more substantially into a dialogue with the very fundamental critique raised by Beven (1989). For instance [4] and [5] did not comment at all on Beven's main conclusion on the need for uncertainty assessment, although [5] actually used the model to study the impact of soil and land use by performing sensitivity analyses. A more comprehensive response and dialogue took place a few years later (Beven, 1996a; Refsgaard et al., 1996; Beven, 1996b).

Seen in the perspective of present protocols for good modelling practise ([12] and [13]) the approach and conclusions in [4] and [5] are especially deficient by the lacking focus on uncertainty assessment. A main reason for the lack of dialogue with Beven's critique and the lack of focus on uncertainty in [4] and [5] may be that we were too preoccupied with the real achievement as the first to setting up and running such type of model for such large catchments. Another reason may be that some of us had a background in groundwater modelling, where large scale distributed physically-based models had been successfully used to support practical water resources management for more than a decade, so we considered Beven's statement that the physically-based models "are not well suited to applications to real catchments" as a large exaggeration.

### 3.1.3 Intercomparison of different types of hydrological models ([6])

**Summary**
The research study reported in publication [6] had two objectives. The first objective was to identify a rigorous framework for the testing of model capabilities for different types of tasks. The second objective was to use this theoretical framework and conduct an intercomparison study involving application of three model codes of different complexity to a number of tasks ranging from traditional simulation of stationary, gauged catchments to simulation of ungauged catchments and of catchments with nonstationary climate conditions. Data from three catchments in Zimbabwe were used for the tests.

The three codes used in the study were (a) NAM (Nielsen and Hansen, 1973; Havnø et al., 1995) – Fig. 15; (b) WATBAL (Knudsen et al., 1986) – Fig. 16; and (c) MIKE SHE (Abbott et al., 1986a,b; Refsgaard and Storm, 1995) – Fig. 17. The NAM and MIKE SHE can be characterised as very typical of their lumped conceptual and distributed physically-based types, respectively, while the WATBAL with its semi-distributed approach falls in between these two standard classes.

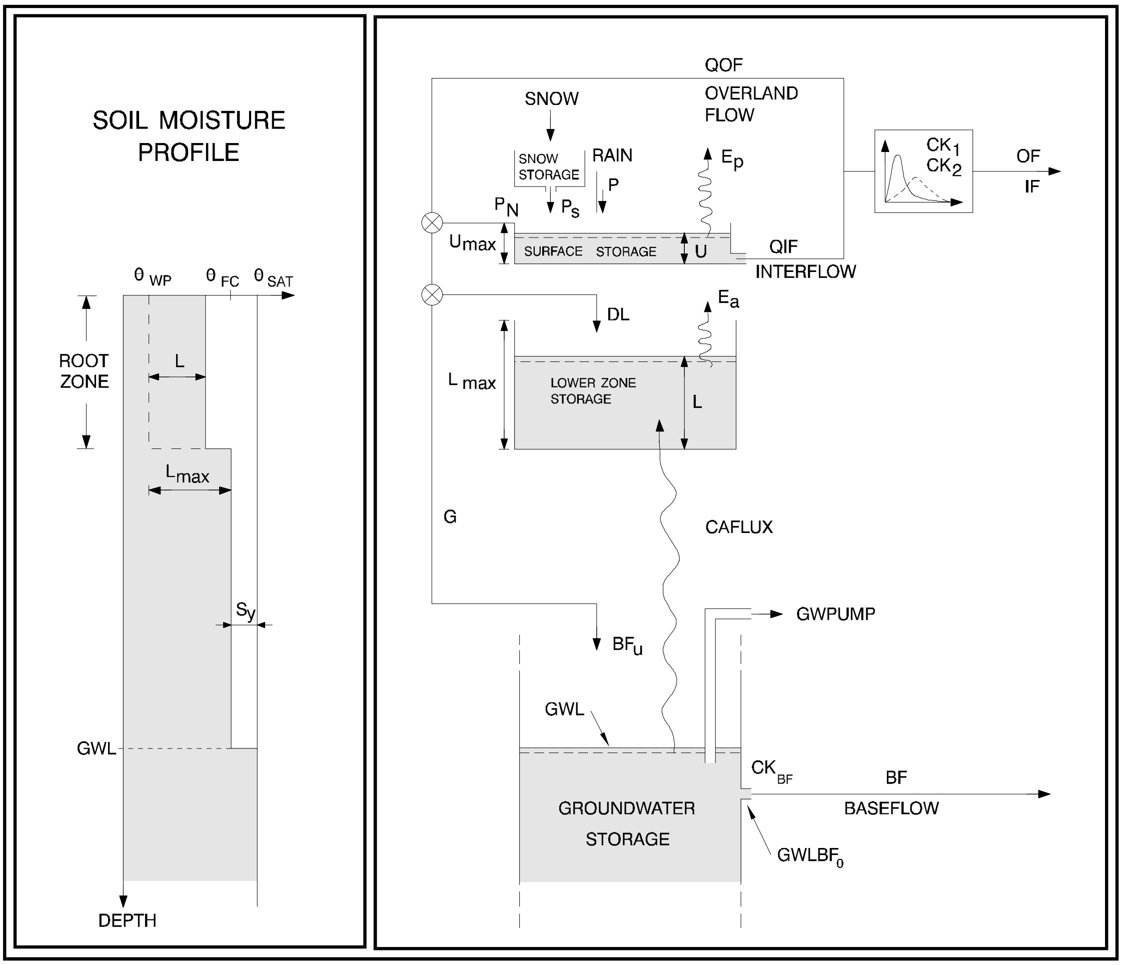



*Fig. 15 Structure of the NAM rainfall-runoff model code*

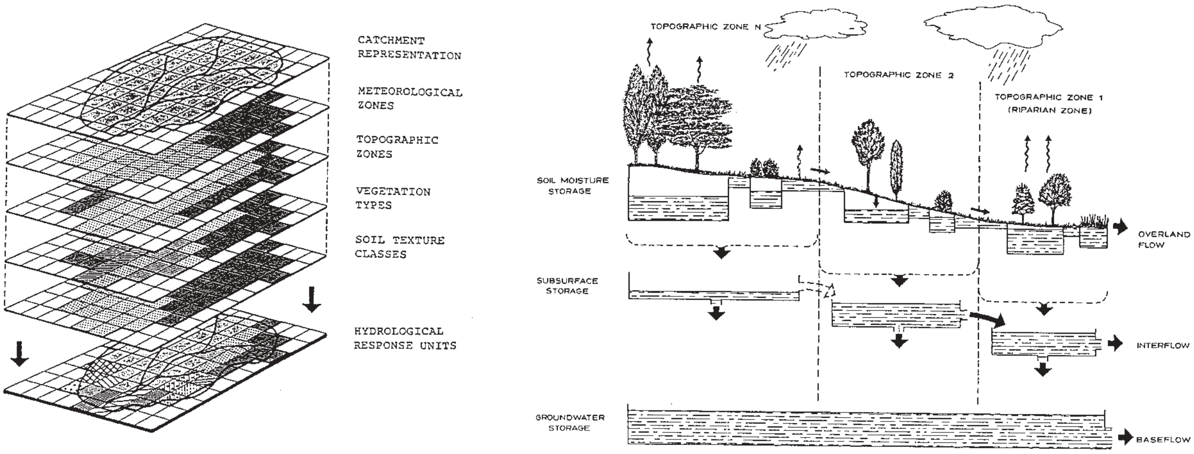



*Fig. 16 Structure of the WATBAL code.*

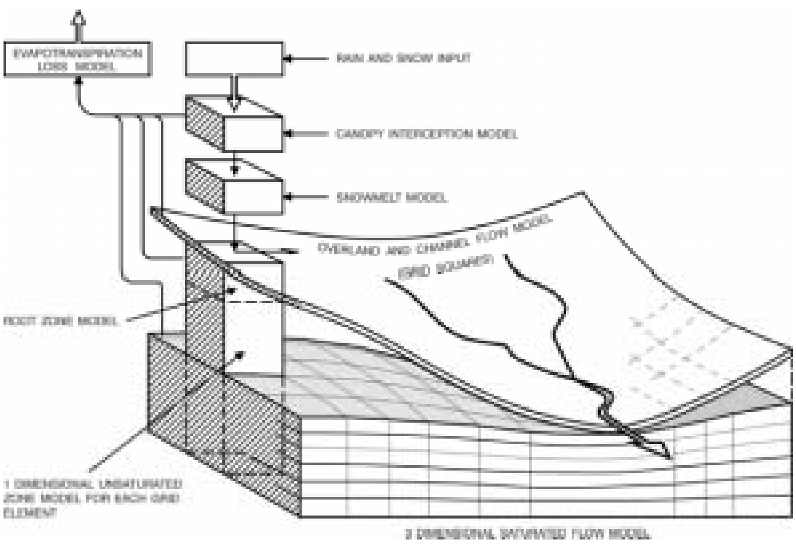



*Fig. 17 Schematic representation of the model structure of the 'Système Hydrologique Européen' (SHE) code.*

The three catchments in Zimbabwe that were selected for the tests were Ngezi-South (1090 km$^2$), Lundi (254 km$^2$) and Ngezi-North (1040 km$^2$). For two of the catchments the model simulations started with a blind simulation, i.e. a simulation where no calibration was conducted, but where model parameters were assessed directly from field data and indirectly by considering parameter values in the first catchment (proxy basin test). Then one year was made available for calibration and finally the full calibration period of 4-5 years was used. In all cases an independent period was used for validation tests (split-sample test). The hydrological regime in Zimbabwe is semi-arid and characterised by very large inter-annual variations. It was therefore possible to construct a test scheme in such a way that a model's ability to predict differences in climate input could be tested by calibrating on a dry period and validating on a wet period or vice versa (differential split-sample test).

The model performance was evaluated for annual runoff and criteria focussing on the shape of the discharge hydrograph, i.e. rainfall-runoff modelling. The modelling work was carried out by three different persons/teams that were very experienced by applying their respective model codes. A general conclusion from the study was that the performances of the three codes were surprisingly similar. Thus, the ability of WATBAL and SHE to explicitly utilise data such as topography, soil and vegetation data that the NAM could not use turned out to make no significant difference in most cases. In summary the conclusions were:

- Given a few (1–3) years of runoff measurements, a lumped model of the NAM type would be a suitable tool from the point of view of technical and economical feasibility. This applies for catchments with homogeneous climatic input as well as cases where significant variations in the exogenous input are encountered.
- For ungauged catchments, however, where accurate simulations are critical for water resources decisions, a distributed model is expected to give better results than a lumped model if appropriate information on catchment characteristics can be obtained.

**Discussion - post evaluation**

A scientific contribution of [6] was the adoption and demonstration of Klemes's model validation testing scheme, which had not been much used since the basic idea was published by Klemes (1986). This is discussed further in Section 4.2.4.

Furthermore, the results from the intercomparison contributed to the ongoing scientific discussion on which types of model codes should be recommended for which application purpose. Only a few intercomparison studies involving different model types had been reported in literature and only two studies included physically-based models (Loague and Freeze, 1985; Michaud and Sorooshian, 1994). Most of these previous studies had been conducted on small research catchments and none of them had included tests for non-stationary climate conditions as in [6].

From the emergence of the distributed physically-based models it was widely stated and believed that these new model types generally would be able to provide more accurate simulation of the hydrological cycle (Abbot et al., 1986a). In the absence of hard facts from suitable tests the scientific debate had to a very large extent been based on expectations and qualitative arguments such that the models with more physical basis in their model structure were assumed to be able to provide more accurate simulation results, or the opposite view, as e.g. advocated by Beven (1989) that such expectations to the superior performance of the physically-based models were unrealistic. In [4] we basically agreed with Beven (1989) with respect to the SHE's capability to simulate discharge for large scale catchments with ordinary data, i.e. that the rainfall-runoff simulation results were of the same degree of accuracy "as would have been expected" with simpler hydrological models of the lumped conceptual type.

With the results from [6] it was now possible to more firmly conclude that if the purpose of modelling is limited to simulation of runoff under stationary catchment conditions and if data exist for calibration purpose, there is no scientifically documented reason to go beyond lumped conceptual models. This issue has been subject to several studies since then, where the conclusions from [6] basically have been confirmed (e.g. Perrin et al., 2001; Reed et al., 2004). I believe that the only thing that may change that conclusion is the introduction of new spatial data from new airborne or satellite sensors. Whereas these new data types have proven to have great value for many hydrological purposes and for special conditions (e.g. snow cover), they have in general not yet documented that they can provide distributed models with comparative advantages in simulation of catchment runoff.

## 3.2 Reactive Transport

### 3.2.1 Oxygen transport and consumption in the unsaturated zone ([3])

**Summary**
Publication [3] describes the development of a new code for simulation of oxygen transport and consumption in the unsaturated zone. The code was linked as a sub-component to the SHE modelling system (Abbott et al., 1986a,b). The objective of the paper was to describe the new process formulation, document its applicability through two case studies and outline the perspectives in relation to its use as part of the comprehensive SHE code.

The unsaturated zone water flow calculations in SHE were based on a finite difference solution to the full Richards' equation for unsteady soil water flow. The solute transport calculations were based on the traditional convection-dispersion equation. The new code for oxygen transport and consumption was an add-on to these first two steps and used information on soil moisture content, water flows and solute concentrations and fluxes as input. Thus the spatial representation is given by the underlying flow and solute transport discretisation, implying a one-dimensional description with spatial resolution ranging from a few cm close to the terrain to 20-40 cm further down in the soil column.

The process description in [3] is based on a three-phase system (soil, water, air) and accounting for spatial heterogeneity at this small scale. Fig. 18 shows a microscale illustration of the soil. Air tends to fill the larger pores in the soil matrix whereas water is drawn into the narrow necks and finer pore spaces in aggregates, forming capillary films and wedges. The air and water coexist in the soil by occupying different geometric configurations. Oxygen movement within these different portions of the pore space can occur by: convective transport in the water, diffusion in water, convective transport in soil air, diffusion in soil air, diffusion into water-saturated soil crumbs, and consumption in free and fixed water.

Microorganisms and plant roots are generally found in the finer pores of the soil because they require close contact with the soil particles for uptake of substrate and nutrients. Transport of oxygen to these respiring sites usually occurs in the water phase of soil crumbs. It is the rate of oxygen diffusion through this fixed water in micropores that will determine the availability of oxygen for respiration and the anaerobic fraction of the soil. A soil crumb is considered to be any fully water-saturated subvolume of soil, the physical size of which is determined by the nearness of air-filled soil pores. The crumb is thus defined by the fact that oxygen transport within the crumb is primarily due to diffusion in water-filled pores. The size of the soil crumbs is dependent on the water content of the soil and the corresponding number of air-filled pores.

The relation between soil water content and size of the water crumbs is derived from the soil water retention curve that is already used in Richards' equation. The idea behind this is illustrated in Fig. 19 and described in more details in [3]. The number of air filled pores at a given soil moisture content can be calculated from the retention curve (Fig. 19b). It is furthermore assumed that the distance between two air filled pores, $d_i$, corresponds to the average diameter of a water saturated crumb (Fig. 19a).

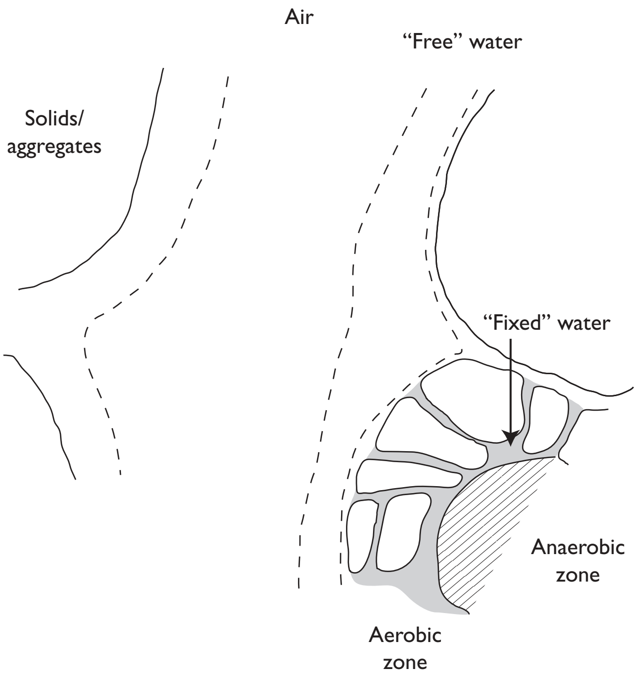



*Fig. 18 Microscale representation of the three-phase soil system with respect to oxygen transport.*

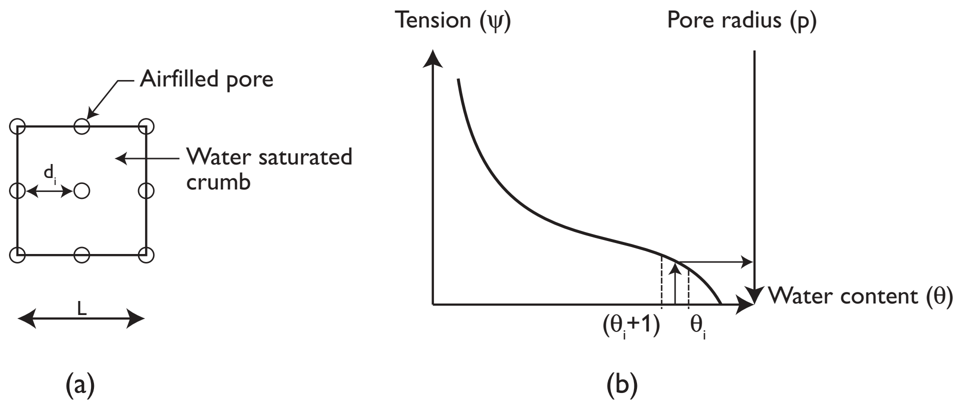



*Fig. 19 (a) The assumed pore distribution within the unit L x L. (b) Retention curve showing the relation between tension, water content and pore radius of a soil.*

The two case studies where the model code was tested and demonstrated dealt with operation of a waste water infiltration plant and assessment of anaerobic zones of importance for denitrification in agricultural soils.

**Discussion - post evaluation**

Previous research in oxygen transport processes in heterogeneous soils (e.g. Currie, 1961; Smith, 1980; Troeh et al., 1982) were based on the assumption of steady-state conditions with regard to crumb/aggregate size and aerobic-anaerobic fractions. The novel scientific contribution of this paper was the new concept of calculating the size of the water crumbs as a function of the water retention curve and the time varying soil moisture content originating from SHE calculations and the linking of this concept to the previous research in this field. In this way it became possible to calculate aerobic-anaerobic fractions dynamically.

Although the scale of consideration in this study is the smallest possible in a catchment modelling perspective, namely point or column scale, it illustrates that smaller scale phenomena (here diffusion into soil crumbs that are of mm or less in size and temporally varying) often dominate the oxygen conditions at grid (cm - dm) scale. The approach in [3] is an upscaling from grain size to computational model grid point, where the within grid heterogeneity is accounted for by developing a set of process equations that includes the effect of the smaller scale heterogeneity at the larger grid scale.

In retrospect, it is interesting to consider the issues that were not discussed in [3]. In this respect it should be noted that code verification aspects were not mentioned in [3], although a completely new code was developed. Furthermore, [3] did not discuss the issue of upscaling the present grid scale processes to application at catchment scale. Interesting issues in this regard would be evaluations of how data and parameter values could be assessed for catchment scale applications and discussions of whether it would still be the mm-scale (crumbs) processes that would be dominating when simulating at large scale, or whether larger scale heterogeneities, such as differences in crops, soil types or topography, would become more important and thus reduce the importance of the present process description.

The model code presented in [3] was developed in a ‘research version’ of the SHE code. After the completion of the study it was not upgraded to become part of the ‘commercial version’ of MIKE SHE that emerged a few years later. The oxygen model has not been used for practical purposes.

To my knowledge, process description of the same detail as in [3] has not been included in any catchment model, and not even in the most comprehensive physically-based root zone models such as DAISY (Hansen et al., 1991; Abrahamsen and Hansen, 2000). In DAISY that provides state-of-the-art descriptions of root zone processes with focus on water, plant growth and nitrogen a much simpler and more empirical process formulation is used for calculating denitrification as a function of anaerobic subsoil conditions.

### 3.2.2 An integrated model for the Danubian Lowland ([9])

**Summary**
Publication [9] is concerned with environmental assessment studies in connection with the Gabcikovo hydropower scheme along the Danube. The objective of the underlying study was to develop and apply a comprehensive integrated modelling system to support management decisions in this respect.

The Danubian Lowland (Fig. 20) in Slovakia and Hungary downstream Bratislava is an inland delta formed in the past by river sediments from the Danube. The entire area forms an alluvial aquifer, which throughout the year receives around 30 $m^3/s$ infiltration water from the Danube in the upper parts of the area and returns it to the Danube and the drainage canals in the downstream part. The aquifer is an important water resource for municipal and agricultural water supply, and the floodplain area with its alluvial forests and associated ecosystems represents a unique landscape of outstanding ecological importance.

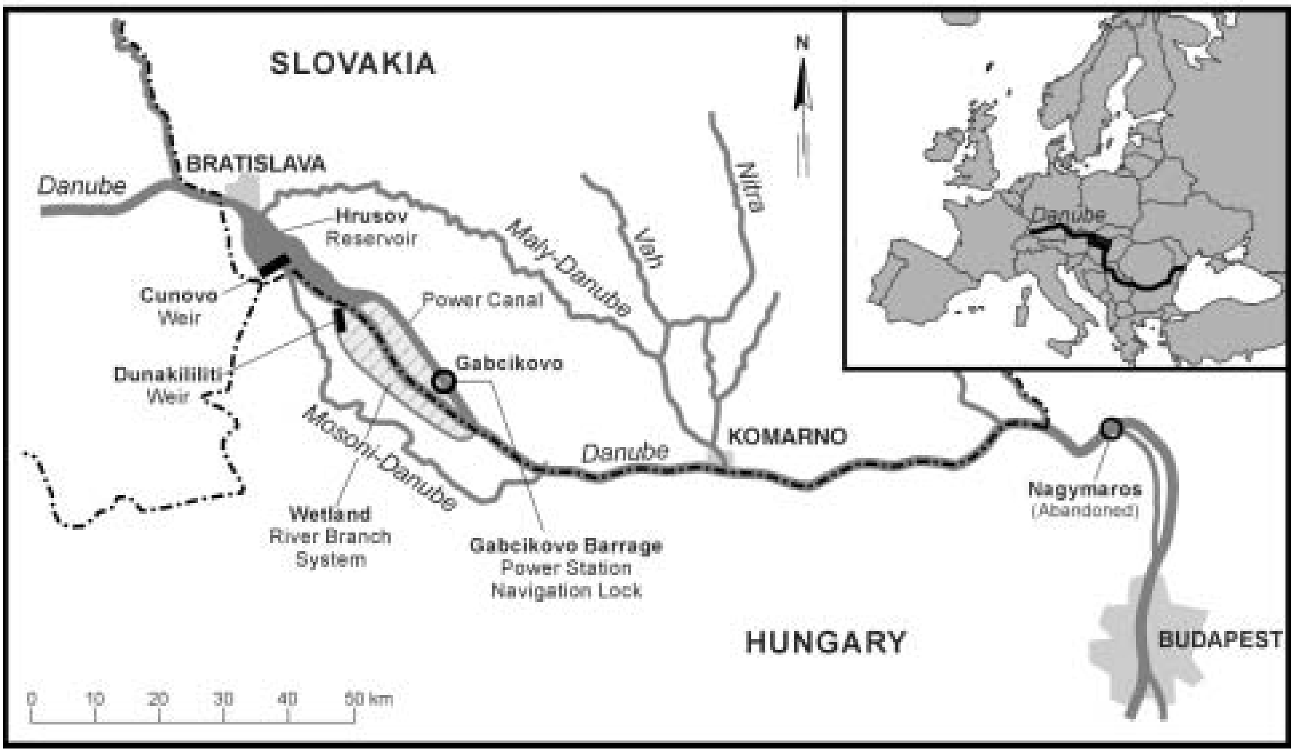



*Fig. 20 The Danubian Lowland with the new reservoir and the Gabcikovo hydropower scheme.*

The Gabcikovo hydropower scheme was put into operation in 1992. A large number of hydraulic structures was established as part of the hydropower scheme. The key structures are a system of weirs across the Danube at Cunovo 15 km downstream of Bratislava, a reservoir created by the damming at Cunovo, a 30 km long lined navigation canal, outside the floodplain area, parallel to the Danube River with intake to the hydropower plant, a hydropower plant and two ship-locks at Gabcikovo, and an intake structure at Dobrohost, 10 km downstream of Cunovo, diverting water from the new canal to the river branch system. The entire scheme has significantly affected the hydrological regime and the ecosystem of the region. The scheme was originally planned as a joint effort between former Czecho-Slovakia and Hungary, and the major parts of the construction were carried out as such on the basis of an international treaty from 1977. However, since 1989 Gabcikovo has been a major matter of controversy between Slovakia and Hungary, who have referred some disputed questions to the International Court of Justice in The Hague (ICJ, 1997).

The hydrological regime in the area is very dynamic with so many crucial links and feedback mechanisms between the various parts of the surface- and subsurface water regimes that no single existing model code was able to describe the entire regime. Therefore, the modelling system illustrated in Fig 21 was established. It integrates four model codes: (a) MIKE 21 (DHI, 1995) for describing the reservoir (2D flow, eutrophication, sediment transport); (b) MIKE 11 (Havnø et al., 1995) describing the river and river branches (1D flow including effects of hydraulic control structures, water quality, sediment transport); (c) MIKE SHE (Refsgaard and Storm, 1995) describing the ground water (3D flow, solute transport, geochemistry) and flood plain conditions (dynamics of inundation pattern, ground water and soil moisture conditions); and (d) DAISY (Hansen et al., 1991) describing agricultural aspects (crop yield, irrigation, nitrogen leaching). The interfaces between the various models were:

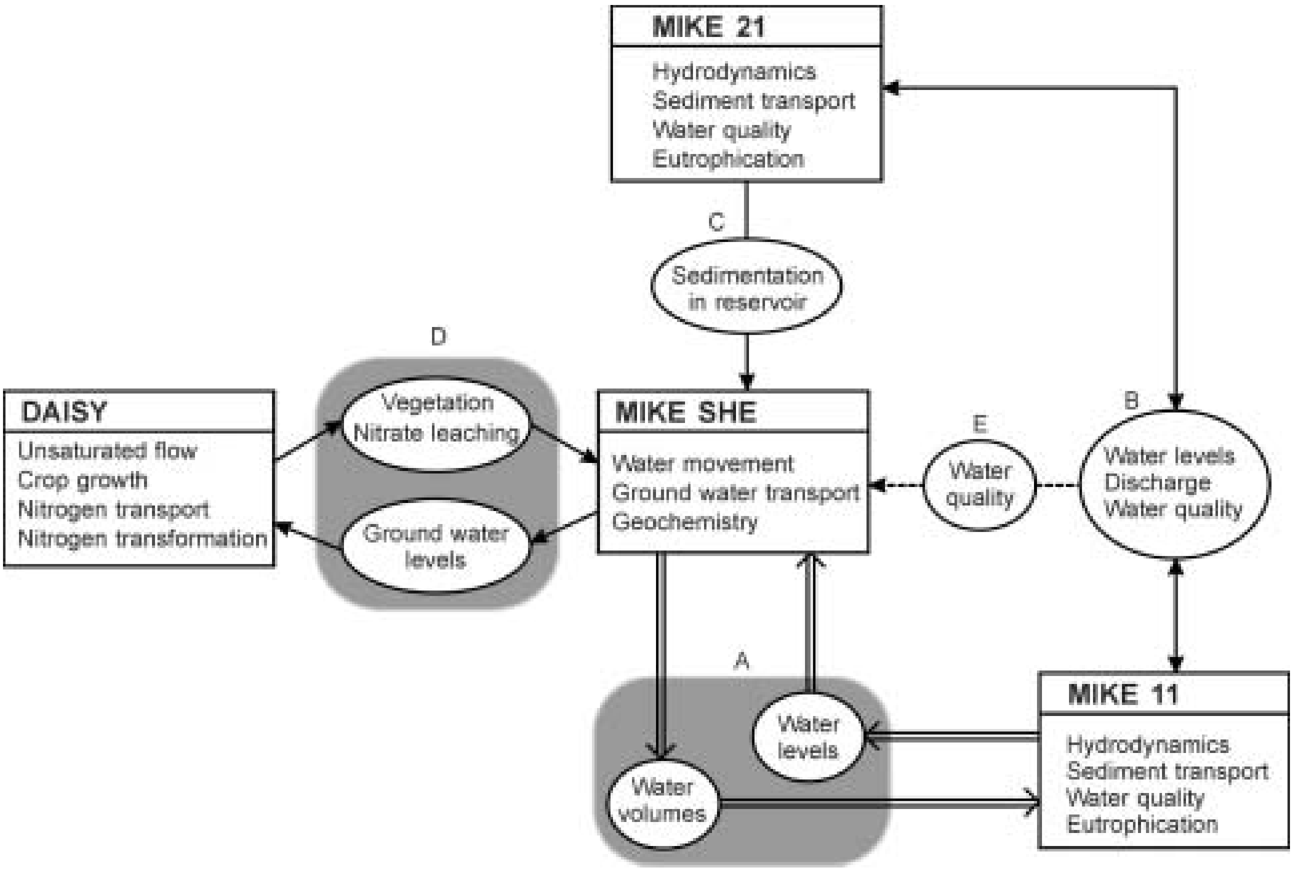



*Fig. 21 Structure of the integrated modelling system with indication of the interactions between the individual models*

A) MIKE SHE forms the core of the integrated modelling system having interfaces to all the individual modelling systems. The coupling of MIKE SHE and MIKE 11 is a fully dynamic coupling where data is exchanged within each computational time step.

B) Results of eutrophication simulations with MIKE 21 in the reservoir are used to estimate the concentration of various water quality parameters in the water that enters the Danube downstream of the reservoir. This information serves as boundary conditions for water quality simulations for the Danube using MIKE 11.

C) Sediment transport simulations in the reservoir with MIKE 21 provide information on the amount of fine sediment on the bottom of the reservoir. The simulated grain size distribution and sediment layer thickness is used to calculate leakage coefficients, which are used in ground water modelling with MIKE SHE to calculate the exchange of water between the reservoir and the aquifer.

D) DAISY simulates vegetation parameters that are used in MIKE SHE to simulate the actual evapotranspiration. Ground water levels simulated with MIKE SHE act as lower boundary conditions for DAISY unsaturated zone simulations. Consequently, this process is iterative and requires several model simulations.

E) Results from water quality simulations with MIKE 11 and MIKE 21 provide estimates of the concentration of various components/parameters in the water that infiltrates to the aquifer from the Danube and the reservoir. This can be used in the ground water quality simulations (geochemistry) with MIKE SHE.

The integrated model was established for the 3,000 km$^2$ area on the basis of a large amount of good quality data. Most of the model parameters were assessed directly from field data, and some were estimated through calibration. For most of the individual model components, traditional split-sample validation tests were carried out.

The modelling system was used in a scenario approach to assess the environmental impacts of alternative water management options. The uncertainties of the model predictions were assessed through sensitivity analyses. As an example, Figs 22 and 23 shows a characterisation of the floodplain area between the (old) main Danube river channel (western model boundary) and the power canal for pre-dam (Fig. 22) and a hypothetical post-dam condition (Fig. 23) where the major part of the water is diverted from the main Danube channel to the power canal. The classes with different ground water depths and flooding have been determined from ecological considerations according to requirements of (semi)terrestrial (floodplain) ecotopes. For the pre-dam condition (Fig. 22) the contacts between the main Danube river and the river branch system is clearly seen. Similar results for a hypothetical post-dam water management regime (Fig. 23) show significant differences in hydrological regime, e.g. many areas are characterised by high groundwater tables and small/seldom flooding, while the post-dam situation (Fig. 22) generally has deeper ground water tables and more frequent flooding. From such changes in hydrological conditions inferences can be made on possible changes in the floodplain ecosystem.

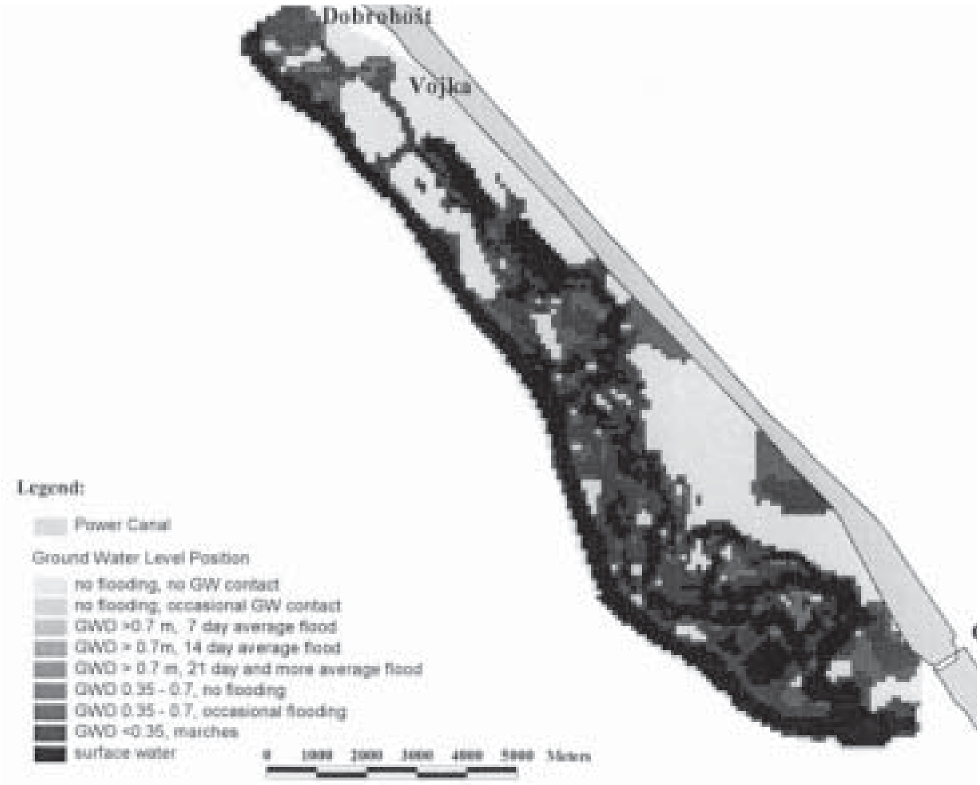



*Fig. 22 Hydrological regime in the river branch area for 1988 pre-dam conditions characterised in ecological classes*

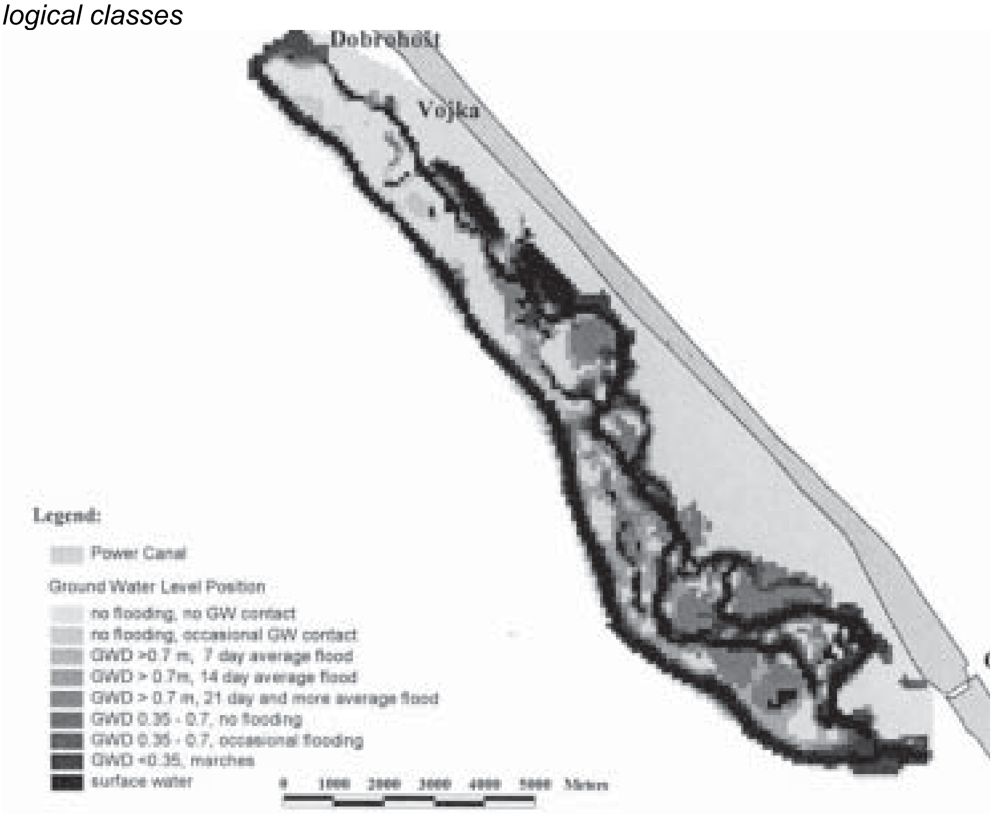



*Fig. 23 Hydrological regime in the river branch area for a post-dam water management regime characterised in ecological classes. The scenario has been simulated using 1988 observed upstream discharge data and a given hypothetical operation of the hydraulic structures.*

**Discussion - post evaluation**

The uniqueness of the established modelling system is the integration between the individual model codes, each of which providing complex distributed physically-based descriptions of the various processes. The validation tests have generally been carried out for the individual models, whereas only few tests on the integrated model were possible. Altogether, the integrated modelling system and the applications were more comprehensive and complex in terms of interactive dynamics between different components of an ecosystem than had previously been reported in the scientific literature.

In the years following [9] a few comprehensive large scale studies with coupled models emerged. The most comprehensive of those was probably Wolf et al. (2003) who developed the STONE for calculating nutrient emissions from agriculture in The Netherlands. Although based on different codes the STONE resembles the integrated modelling system in [9] in terms of number of codes and complexity of process descriptions. One main difference, however, was that STONE consists of a chain of models without the feedback couplings that characterise [9]. Simpler, although still comprehensive, modelling systems were presented by Birkinshaw and Ewen (2000) as the SHETRAN code with a built-in nitrate transformation component and Conan et al. (2003) with a coupling of SWAT, MODFLOW and MT3DMS also focusing on nitrate fate at catchment scale.

The complexity of the modelling studies in [9] may be compared to coupled modelling studies in neighbouring fields. The hydrology related field with the strongest modelling traditions is no doubt the atmospheric science. Here very comprehensive coupled models have been used in connection with hydrology oriented climate change studies. An example of a sequentially coupled atmospheric-hydrological model from that period is Graham (1999) who used the ECHAM4 regional atmospheric model coupled with the HBV hydrological model to simulate discharge for the entire 1.6 $10^6$ $km^2$ Baltic Sea basin. The atmospheric modelling component is in itself more demanding in terms of computer power than comprehensive hydrological modelling such as [9], and the complexity of the atmospheric modelling is maybe larger than the complexity of the individual process model codes in [9]. Otherwise the complexity of the coupled atmospheric-hydrological studies with respect to feedback couplings between process descriptions, data requirements, different scales for different processes, etc., may be considered comparable to the complexity of [9].

In retrospect it is interesting to evaluate how much this comprehensive modelling system actually was used as part of the political decision process? Were the full potential of the models utilised by the decision makers? In the following my personal perception of these aspects are presented. The application of the integrated modelling and information system in practise may be categorised in three principally different functions: (a) to assist in design of structures and details of water management regimes, (b) to assist in policy analysis by assessing the environmental impacts of alternative water management regimes, and (c) to assist in resolving different views between interest groups on environmental assessments.

The use of models to assist in designs is the classical "engineering" way of using such models. There were a number of such applications. The best example of this is the final design in 1993 of the guiding structures of the Cunovo reservoir that was based on model simulations. Such model use was possible, because the objectives of the decision-makers were clear and there was an urgent need for the results before the construction works actually started.

Use of models to assess the environmental impacts of alternative water management regimes was one of the primary reasons for establishing the modelling systems. There were several examples of such model applications. A key example was a combined field and modelling study of the geochemical conditions in the aquifer to assess whether the changed boundary conditions with the new reservoir would affect the redox conditions and hence the groundwater quality in the aquifer that forms the basis for the water supply of Bratislava. Another example is a combined field and modelling study of the eutrophication conditions in the reservoir. Such studies were conducted in close dialogue with the decision-makers in order to assist in their policy formulation.

Finally, the modelling system was an invaluable tool in connection with the international attempts made to assist in resolving some of the issues that were disputed between Slovakia and Hungary. Many of the arguments brought forward on these highly controversial issues were mixtures of scientifically based facts and politically based views, but they were often claimed as purely scientifically based. It is very natural and fully legitimate that all parties have political interests and do their best to pursue them. However, the mixing of scientific facts and political interest makes the whole scene less transparent and may be an obstacle for arriving at rationale decisions. The role the modelling system had in this context was that it made it possible at some occasions to help distinguish between facts and fiction with respect to the scientific arguments. In this way the modelling tools assisted in separating scientific and political problems. Thus, the modelling system was often used as an important tool in resolving technical disagreements between the Slovakian and Hungarian delegations in the international expert groups (EC, 1992, 1993a, 1993b). Similarly, it is my impression that the modelling results played a significant role for the International Court of Justice when dealing with the question of whether the ecological situation could be characterised as a catastrophe justifying the use of the legal principle of "the ecological state of necessity" as done when Hungary stopped the construction works on the Gabcikovo scheme in 1989 (ICJ, 1997).

However, there were also clear limitations to the application of the modelling tools. These limitations occurred when the political objectives were not clearly defined. It was for instance imagined that the modelling tools should be used to identify the optimal solution for the water management regime in the river branch system. This unique area is, however, subject to considerable interest from different sectors such as commercial forestry, fishery, tourism and natural conservation. The requirements of these different sectoral interests are not common and in some cases even contradictory with respect to how the water regime should be. Thus, until the balance of interests between these different stakeholders has been decided in terms of clear political goals from the government, an optimal solution does not exist. Another example of lack of clear political goals was related to the overall sharing of water between hydropower and the environment.

### 3.2.3 Large scale modelling of groundwater contamination ([10])

**Summary**

Publication [10] describes results from an EU research project on groundwater pollution from non-point sources. The rationale outlined in [10] is that physically based models for describing nitrate due to better process descriptions may be expected to have better predictive capabilities than simpler empirical models for certain applications related to assessing the impacts of changes in agricultural management practise. Such models were well proven for simulation of nitrate contamination at small scale with good data availability. Two of the main constraints for using such models operationally were that (a) the databases existing at national or European scale had not previously been tested as input for such models; and (b) almost no tests had been conducted for such models at large scale. The objectives of the paper were therefore to study the data availability at the large scale and develop methodologies for model upscaling/aggregation to represent conditions at larger scale. The theoretical aspects on scaling included in [10] are dealt with in Section 4.1. Here some key results from one of the two catchments (Karup) are discussed.

The modelling system used was MIKE SHE (Refsgaard and Storm, 1995) coupled with the DAISY root zone model (Hansen et al., 1991). Two Danish catchments of about 500 km$^2$ each, Karup and Odense, were used for the tests.

The principles used for collecting input data and assessing values of model parameters were:

- The data must be easily accessible. This implied that most of the data were aggregated data from national or European databases.
- No model calibration is carried out. Instead parameter values are estimated from generic transfer functions.

Data were collected from the following sources:

- Topography: 1 km grid data downloadable from USGS and GISCO (Geographical Information System of the European Commission)
- Catchment boundaries and river network: generated from the topographical data using standard GIS functionality.
- River cross-sections: derived from a special GIS application where the cross-section was estimated based on upstream catchment area, slope and a characteristic discharge.
- Soil type: GISCO soil map.
- Soil organic matter: experience values.
- Vegetation: EEA CORINE land cover map.
- Agricultural management practise: Agricultural statistics and government prescribed norms
- Geology and groundwater abstraction: EC report
- Climatic variables and discharge data: national data

The MIKE SHE models were run with 1, 2 and 4 km grids. For describing the nitrate leaching from the root zone, 17 crop rotation schemes were established by use of DAISY. The crop rotations were based on the statistical information on crop type and livestock densities. The 17 schemes were distributed randomly over the catchment in such a way that the statistical distribution was in accordance with the agricultural statistics. As an alternative, all the agricultural area was described by one representative crop instead of 17 cropping patterns. These two approaches are denoted 'Distributed' and 'Uniform' in Figs. 24 and 25 below.

The Karup model was validated by comparison of model simulations and field data on annual water balances, discharge hydrographs (Fig. 24) and nitrate concentrations in the upper groundwater layer from 35 observation wells (Fig. 25). The results of the validation tests were characterised as follows:

- The annual water balance was simulated remarkably well with only 2% difference as average value over the five years validation period. The variation over the year (Fig. 24) is less well described.
- The simulated nitrate concentrations (Fig. 25) match the observed data remarkably well both with respect to average concentrations and statistical distribution of concentrations within the catchment.
- The simulations are clearly affected by various scale effects (1, 2, 4 km grid and Distributed/Uniform). This is addressed further in Section 4.1 below.

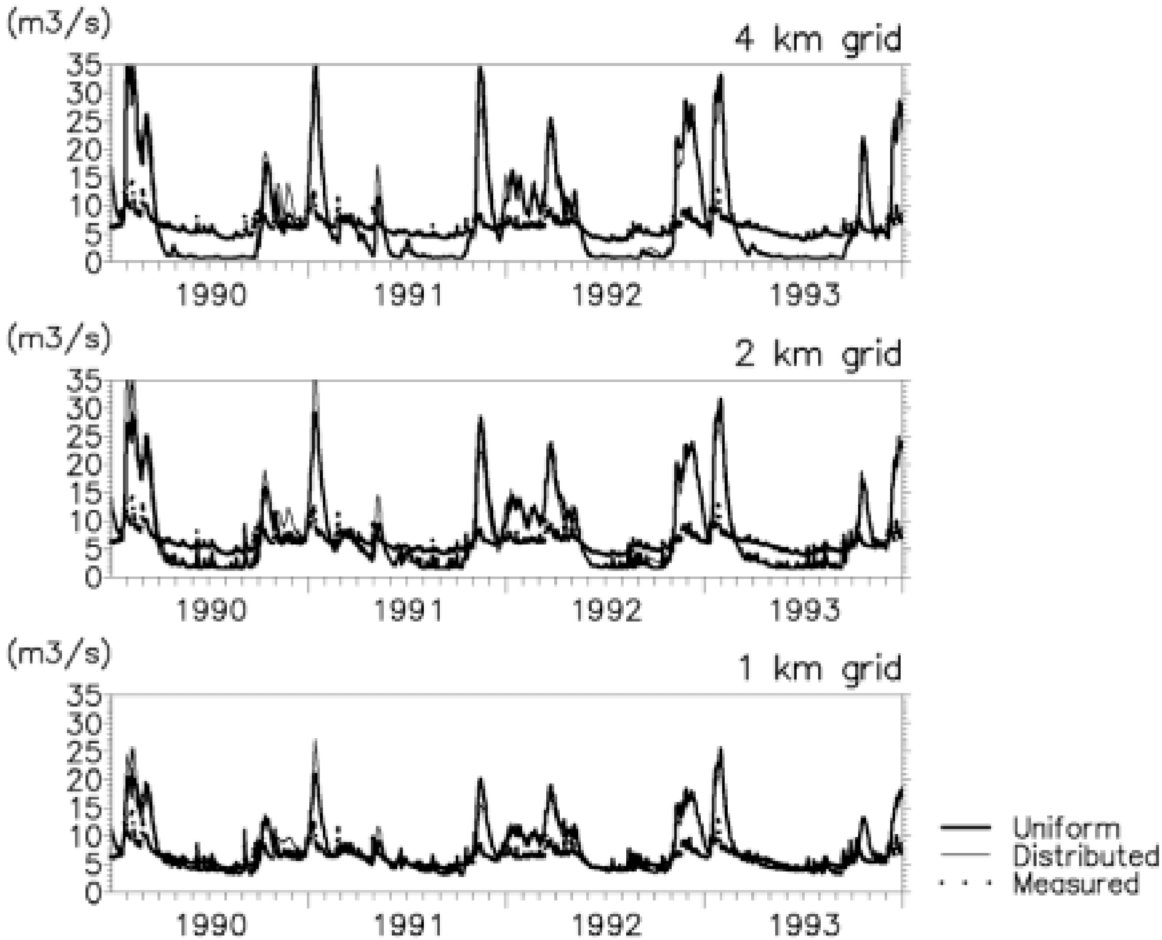



*Fig. 24 Comparison of the recorded discharge hydrograph for the Karup catchment with simulations based on 1, 2 and 4 km grids. The two simulated curves correspond to the combined upscaling/aggregation procedure (Distributed) and the simpler upscaling procedure (Uniform).*

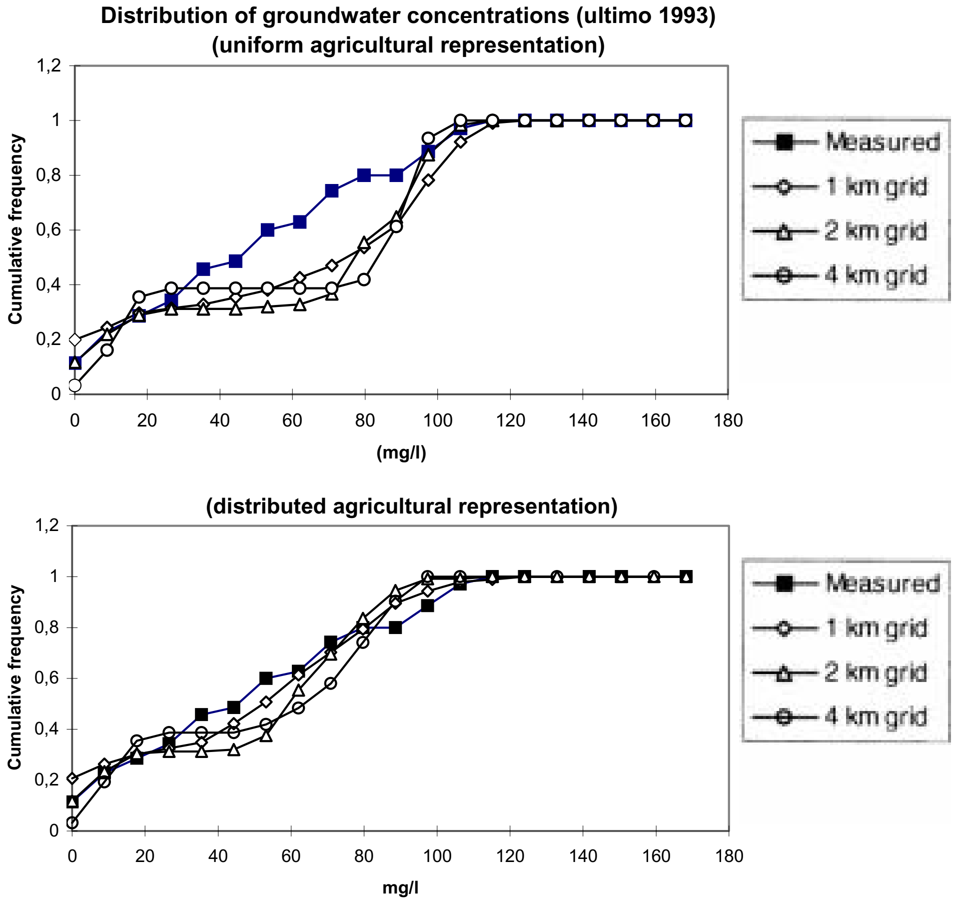



*Fig. 25 Comparison of statistical distribution of nitrate concentrations in groundwater for the Karup catchment by the model with 1, 2 and 4 km grids and observed in 35 wells. The lower figure corresponds to the upscaling procedure resulting in a distributed representation of agricultural crops, while the upper figure is from the run with the upscaling procedure, where all agricultural area is represented by one uniform crop.*

**Discussion - post evaluation**

The model codes used in [10] were well known and previously used in one of the catchments (Styczen and Storm, 1993a, b). The scientific contributions of [10] relate partly to scaling issues, which are dealt with in Section 4.1 below, and partly to testing the performance of nitrate catchment models when scarce data are used and when no model calibration is carried out. The most important finding with respect to data availability is probably that aggregated data in many cases can provide sufficient input to perform useful model simulations. This message is similar to the output from the first large scale application of SHE to catchments in India with scarce data ([4] and [5]), namely that an apparent lack of primary data should not always prevent you from using a model.

With regard to data availability at large scale it was concluded that the most critical data that may cause problems for large scale applications are the geological data for which no suitable global or European digital database exist. In this respect the development of a national hydrological model in Denmark (Henriksen et al., 2003) that is based on comprehensive geological data from the very large national geological database is an important development.

The study showed that one of the strengths of physically-based models is the possibility to assess many parameter values from standard values, achieved from experience through a number of other applications. It also showed some of the limitations in this respect. While the key results in terms of annual runoff and nitrogen concentration distributions are encouraging, the discharge hydrographs clearly illustrate that it would be very easy to obtain a better hydrograph fit through calibration of a couple of parameter values. When parameters are assessed in this way they are subject to considerable uncertainty, which will generate significant uncertainty in model predictions. This aspect is addressed in ([11]) which is discussed in Section 4.3 below.

The attempt to assess parameter values directly from data without any model calibration can be seen as the extreme end of the development starting with hundreds of free parameters in the Suså model ([1]), over 26 parameters in the Kolar basin in India ([5]), to 11 free parameters in a previous Karup study ([7]). The results from the present study showed some obvious shortcomings of this approach, and in a later study of the Senegal basin (Andersen et al., 2001) we used 4 free parameters for calibration.

## 3.3 Real-time Flood Forecasting

### 3.3.1 Intercomparison of updating procedures for real-time forecasting ([8])

**Summary**
Publication [8] presents a classification of updating procedures used in real-time flood forecasting modelling and a review of the results from the WMO project ‘Simulated Real-Time Intercomparison of Hydrological Models’ (WMO, 1992) comprising more than 10 commonly used hydrological model codes and a variety of different updating procedures. The objective of the paper was to analyse the performance of different types of updating procedures and to assess what is more important, the simulation model or the updating procedure.

In the context of real-time forecasting a hydrological catchment model, as those in the remaining part of this thesis, may be denoted a *process model* (Fig. 26). A process model consists of a model structure including process equations, model parameters that are constant throughout a model run and state variables. The transformation from input to output by the process model is called simulation, in accordance with the terminology defined in Section 2.2 above. Process models that operate in real-time may take into consideration the measured discharge/water level at the time of preparing the forecast. This feedback process of assimilating the measured data into the forecasting procedure is referred to as *updating*, or data assimilation. Updating procedures can be classified according to four different methodologies (Fig. 26):

1. Updating of input variables, typically by adjusting precipitation.
2. Updating of state variables, e.g. the soil moisture content.
3. Updating of model parameters.
4. Updating of output variables (error prediction).

The core of the WMO project was a workshop held in Vancouver during the period July 30 – August 8, 1987, where 15 models from 14 different organisations were run in a simulated real-time environment. Data from three catchments with significantly different hydrological characteristics were used for the tests. Before the workshop the modellers had received historical data for several years for calibration and validation and two ‘warm up’ flood events. During the workshop four additional flood events were forecasted as blind tests, each with seven forecasts at consecutive times. Each event was forecasted within one workshop day, often under considerable time pressure.

I participated in the workshop with two models that differed both with respect to process model and updating procedure:

∉ NAMS11 comprising the NAM as catchment model, St. Venant river routing and an error prediction model as updating procedure. This is basically identical to what later became known as the flood forecasting module of MIKE 11 (Havnø et al., 1995).

- NAMKAL comprising the NAM formulated in a state-space form and build into an extended Kalman filter for updating. This version had no separate river routing but relied on the linear reservoirs in NAM.

The two models were tested on the 104 km$^2$ Orgeval catchment (France) and the 2,344 km$^2$ Bird Creek catchment (United States). The models were not tested on the third, snow-dominated catchment.

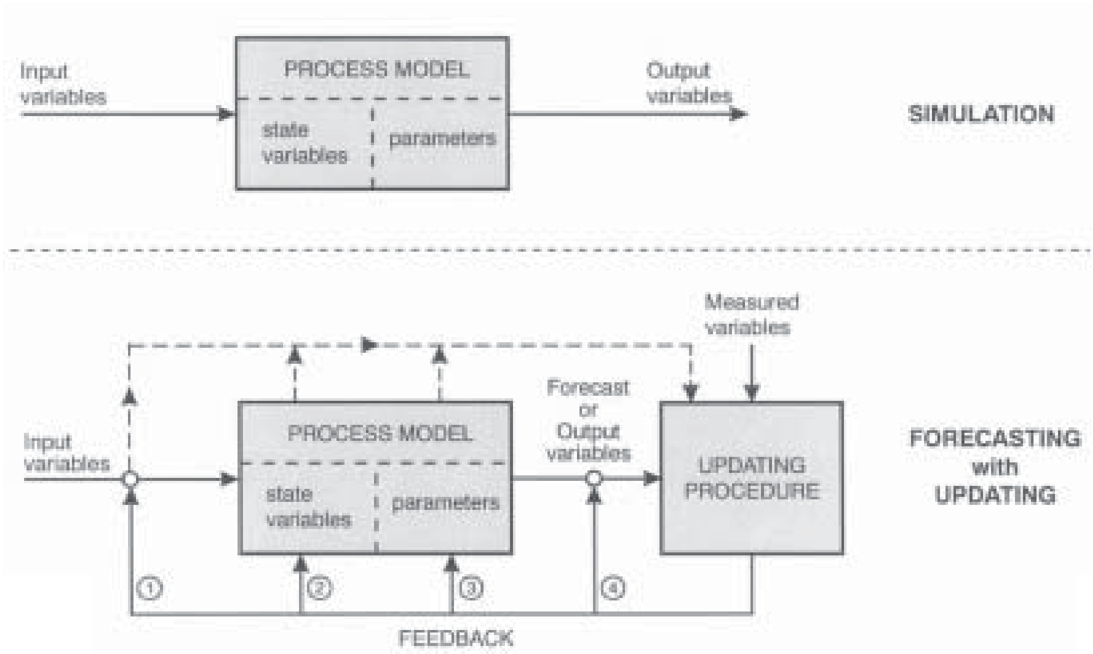



*Fig. 26 Schematic diagram of simulation and forecasting with illustration of four different updating methodologies), [8].*

Summary results from the two catchments are shown in Fig. 27 as root mean square errors (RMSE) as a function of forecast lead time (lag). As can be seen from the figure the intercomparison test turned out to be a very close 'race' with at least one third of the models performing almost equally well. Depending on the selected criteria for comparison (which catchment, priority to short, medium or long lead times, etc.) several of these could claim to be the 'best model'. What is maybe more interesting is some of the general findings:

- The process models belonged to two of the classes shown in Fig. 6, namely empirical (black box) models and lumped conceptual models. From the results it was not possible to clearly distinguish which model type performed better.
- All four types of updating procedures were represented, both among the models with the best performance and among the models with the poorest performance. This indicates that the selection of a specific updating methodology is only one out of several important factors.
- The forecast error (RMSE) generally increases with forecast lead time. This shows that updating procedures most often significantly improve the performance of hydrological models for short-range forecasting.
- In most cases the models with the best performance for short lead times were also those with the best results for the long lead times. This indicates that the goodness of the basic simulation (by the

process model) is crucial to forecast accuracy, or in other words that a good updating procedure can not compensate for a poor process model.

**Discussion - post evaluation**

Real-time forecasting is the toughest field I have experienced in hydrological modelling with respect to model validation, because the results of the model forecasts are continuously confronted with observations. In many studies involving model simulations for planning purposes it is often not possible to conduct a validation test that exactly fits the conditions for which model simulations of future conditions are needed. Therefore, the validation test results will often have many qualifiers and be considered together with other arguments. In real-time flood forecasting there is no need for such qualifiers and arguments ('no nonsense') and therefore only the hard facts are considered.

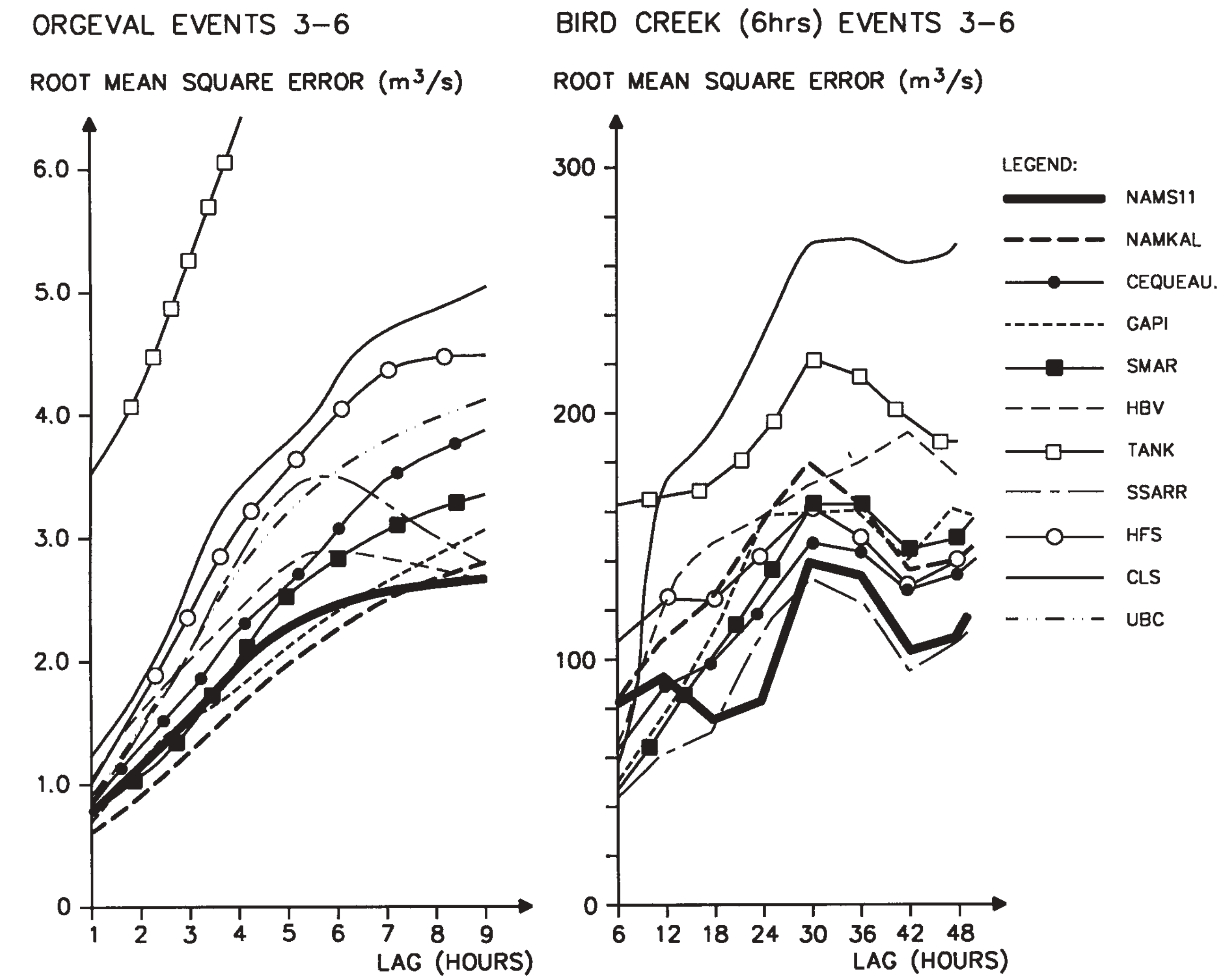



*Fig. 27 Root Mean Square Errors (RMSE) as a function of forecast lead time for all models participating in the Orgeval and Bird Creek catchments. The RMSE values are averaged over the four forecasted flood events with blind tests (events 3-6), [8].*

The main scientific contribution of [8] was the analysis of the performance of different types of process models and updating procedures and combinations hereof. Our motivations to participate in this unique WMO intercomparison project were (a) to test DHI's code NAMS11 (now MIKE 11), which was used operationally in India at that time, in an intercomparison with some of the internationally leading codes and modellers; and (b) to test whether an extended Kalman filter could provide a better updating routine than the more commonly used and simpler error prediction routine. In addition to noting that the NAMS11 performed very well and that the extended Kalman filter under ideal conditions could perform marginally better than the standard updating procedure, the analysis lead to the following interesting findings:

- It was not possible to conclude which model type, black box or lumped conceptual, is better suited for simulation of runoff. This is in good agreement with [6] and later studies such as Reed et al. (2004), which concluded that lumped conceptual and distributed physically-based models performed equally well for split-sample tests. Thus it may be argued that all three model types described in Section 2.4 in many cases can be expected to be able to perform equally well in rainfall-runoff modelling.
- It turned out that the personal factor is maybe the most important aspect of hydrological modelling. It was clear after the workshop that the difference in model performances between the participating codes could often not be explained by differences in model codes. Personal factors such as the modeller's ability to make a good model calibration, experience from working in hydrological regimes different from the regime you see in your home office, ability to work under extreme stress, level of preparation beforehand and random luck also played important roles. The personal factor is most often overlooked in natural science, maybe because it is subjective of nature and therefore does not fit well into the methods usually adopted in natural science. The ultimate consequence of this finding is that good quality of modelling results requires both use of good scientifically based methodologies and adoption of sound practises by competent professionals. This consequence was not derived in [6] but is central for recent work on quality assurance guidelines in the modelling process ([13]).

Most of the model codes that participated in the intercomparison study were state-of-the-art hydrological model codes such as Sacramento (Burnash, 1995), HBV (Bergström, 1995) and MIKE 11 (NAMS11) with comprehensive experience in operational flood forecasting. These codes are still among the most commonly used today. The updating techniques tested in [8] are also still the basic techniques used operationally today, although more sophisticated developments and improvements have taken place, e.g. a combination of the Kalman filtering and the error prediction procedure (Madsen and Skotner, 2005).

# 4. Key Issues in Catchment Scale Hydrological Modelling

## 4.1 Scaling

This section provides a discussion of catchment heterogeneity and upscaling in relation to catchment modelling based partly on the publications in the present thesis (most importantly [7] and [10]) and partly on other previous work such as Refsgaard (1981), the foundation of [1] and [2], and Refsgaard and Butts (1999) that was heavily inspired by the EU research project behind [10] and [11].

Hydrological modelling is being carried out at spatial scales ranging from pore scale to global scale and a variety of scaling theories has been developed, see e.g. Blöschl and Sivapalan (1995) and Beven (1995). Many of the scaling theories consider different spatial scales for single processes. For catchment modelling it is necessary to include several processes and their linkages.

### 4.1.1 Catchment heterogeneity

Catchment properties exhibit spatial variability. For almost all properties this heterogeneity is very large and dominates the behaviour of the catchment. Scaling is basically a question of how to handle heterogeneity at different spatial scales. Different model types do this fundamentally different. Let us illustrate this by two examples.

As the first example, let us consider an idealised description of flow through the root zone (Fig. 28). If a soil column, initially dry, is supplied with a certain amount of water it will retain water, until it is filled to a certain level, the field capacity $\theta'_F$, whereupon all the supplied water will pass through. This is illustrated in Fig. 28 A,B,C, where also the frequency and the distribution of $\theta_F$ are shown. If we then consider a catchment with a spatial variability in soil physical properties, the frequency and the distribution of the field capacity are illustrated in Fig. 28 D and E respectively. If the root zone of this catchment, initially dry, is being supplied with water, not all of the area will contribute to throughflow at the same time, as $\theta_F$ varies in the catchment. When, for instance, the rainfall has supplied the water amount $\theta'_{F,m}$, it is seen from Fig. 28 E that field capacity has been reached in one half of the catchment, thus contributing to throughflow, while the other half of the catchment still retains the rain in its root zone.

In a lumped model, such as NAM, such spatial variability is taken into account by using semi-empirical relations as e.g. the dashed line in Fig. 28 F, where $\theta'_1$ and $\theta'_2$ typically have to be estimated from calibration. The difference between $\theta'_1$ and $\theta'_2$ can be seen as a measure of the heterogeneity of the catchment, or of the catchment input that is also assumed homogeneously distributed in a lumped approach. This way of accounting for the spatial variability in the process equations can be considered the heart of lumped models and also explains why the process equations in lumped models are fundamentally different from point scale physical process equations.

In a distributed model the spatial variability is taken into account by dividing the catchment into several smaller elements, which are then usually treated as homogeneous units, i.e. as a column in Fig. 28.

However, the spatial variability of soil physical properties comprise both variability between different soil types and variability within the same soil type as illustrated in Fig. 29. It has been demonstrated in several studies (Nielsen et al., 1973; Jensen and Refsgaard, 1991a,b,c; Djurhus et al. 1999) that the spatial variability of e.g. soil properties within one standard soil type at field scale is very high and can significantly influence the water balance and solute transport at this scale.

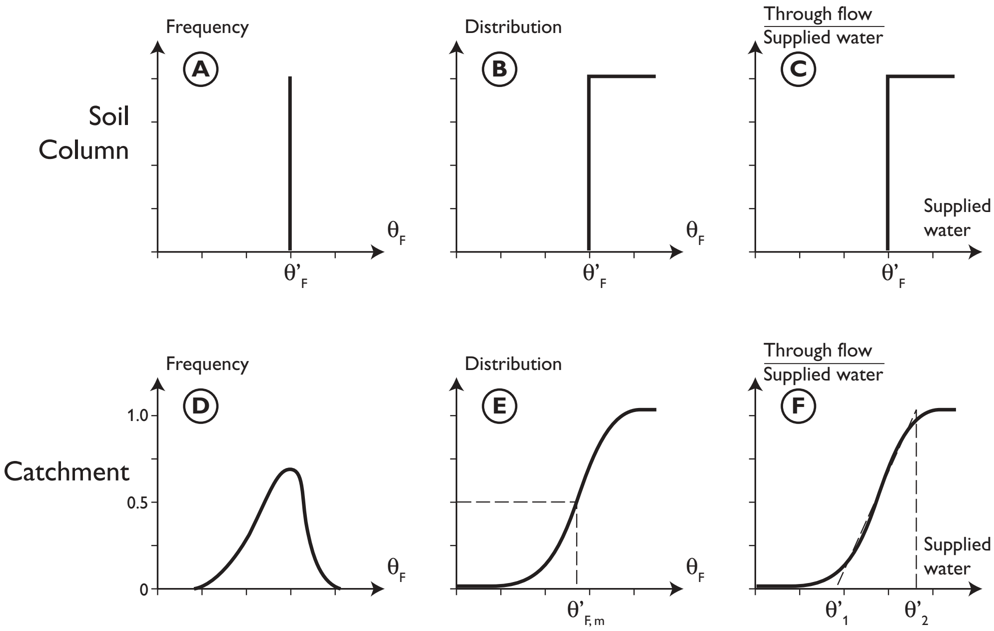



*Fig. 28 Idealised description of the variation of field capacity, èF, and its effect on flow through the root zone in a soil column and in a catchment (Refsgaard, 1981).*

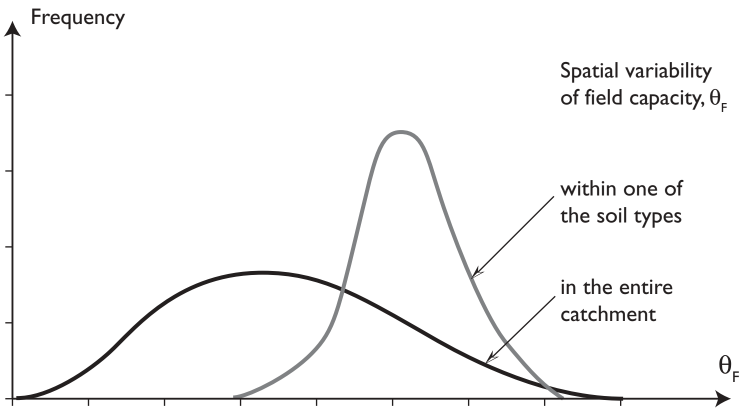



*Fig. 29 The principle of spatial variability of a soil physical property within a single soil type and within a catchment containing more than one soil type (Refsgaard, 1981).*

Let us then turn to another example focusing on the limitation of a distributed model to resolve key features of a catchment. Fig. 30 shows the topography and river network for two models that are identical except for differences in spatial discretisation. It is clearly seen that the 500 m grid provides a much better resolution of the topography and the river network, and also of other catchment characteristics as explained in [7]. In the 2000 m grid the river valley cannot be described well and many of the smaller streams have to be omitted, where the distance between neighbouring streams are smaller than the model grid size. This significantly affects the stream-aquifer interaction and in this way the simulation of both river discharge and groundwater heads. As discussed in [7] a change in scale (grid size) in this way changes the model simulations. This can in some cases be compensated by adjusting parameter values. But it implies that parameter values are scale dependent and that the physical basis is reduced if the grid size is increased.

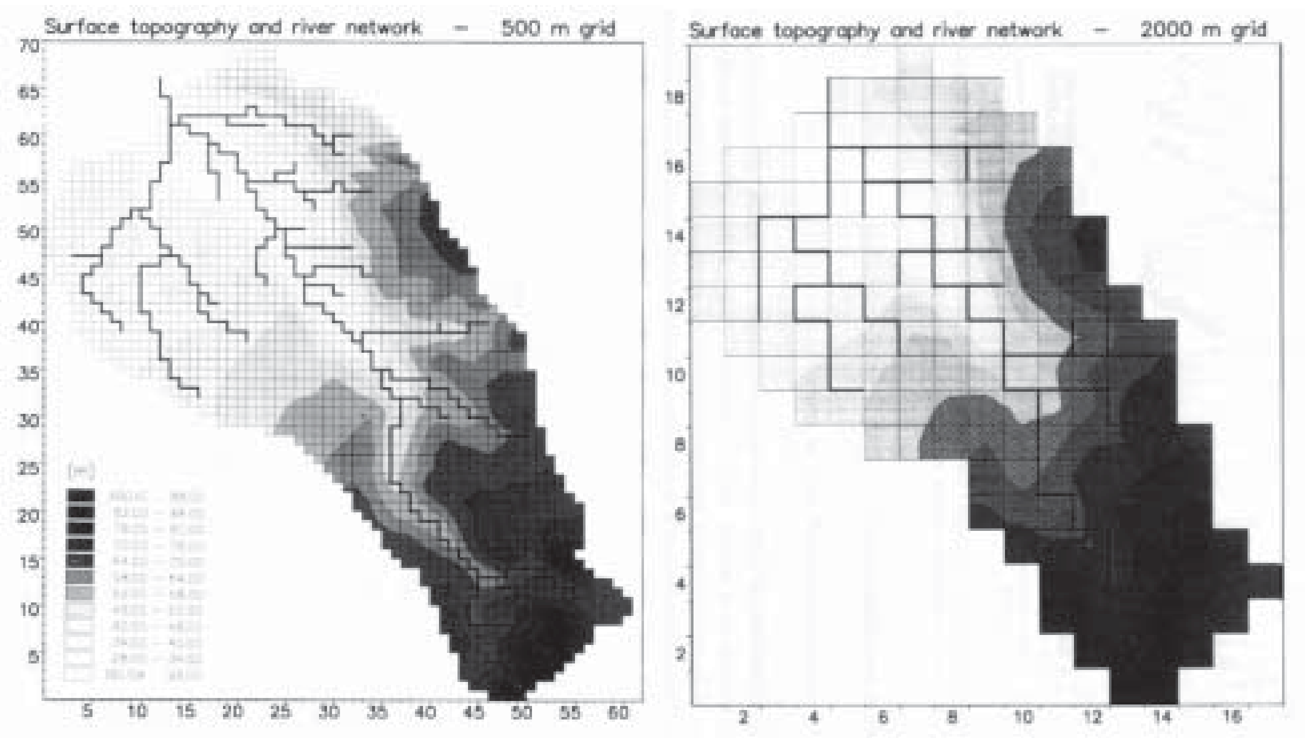



*Fig. 30 Topography, river network and model grid for two models with discretisations of 500 m and 2000 m [7].*

This example focussed on river discharges and hydraulic heads at some given observational locations for which [7] argues that a 500 m resolution provides an adequate description. If we instead had focussed on other processes such as reactive transport in aquifers or in river valleys, we would have needed to account for geological and geomorphological heterogeneity of much smaller scale than 500 m. This line of argument can continue down to pore scale processes such as those described in [3]. The point is that, no matter which resolution a model has, it is always possible to find processes that require a smaller scale in order to provide a physically based description. Consequently, the ultimate distributed physically based model where everything is described can never be achieved. This implies that any distributed model needs to provide a kind of lumped conceptual representation at its scale of operation. An excellent example of this is the traditional advection dispersion equation with its associated dispersivities, where the dispersivities show the well known scale dependence (Gelhar, 1986). The process description of oxygen transport and consumption given in [3] is another example. Although meant for inclusion as a submodel in a distributed physically based model, [3] incorporates spatial heterogeneity of processes at pore scale (mm) to a process equation assumed valid at its scale of operation (grid points with 10-40 cm distance). This process equation can therefore be considered a lumped conceptual description at this scale.

### 4.1.2 A scaling framework

In this section we only consider the case of moving from the smaller to the larger scale, which is often denoted upscaling. When moving to larger scales the spatial variability of physical parameters and variables have to be taken into account. This can in principle be done in two ways, either by aggregation or upscaling (Heuvelink and Pebesma, 1999):

- *Upscaling* means that the process equations and the associated parameters that basically constitute the model in principle are modified or substituted when moving from the smaller scale to the larger scale.
- *Aggregation* means that the process equations are applied at the smaller scale (where they were derived) and the large-scale results are obtained by aggregating the small-scale results at the larger scale.

Hence, in order not to confuse the terminology with two different meanings of the term upscaling the term *scaling* will in the following be used for the case of moving from modelling at the smaller scale to modelling at the larger scale. Thus, the term upscaling is reserved to the specific approach of scaling defined above.

The differences between upscaling and aggregation are illustrated in Fig. 31 and some key characteristics are summarised in Table 1. At the smaller scale, the hydrological processes can be described by smaller scale equations and associated smaller scale parameters. If the aggregation approach is adopted for large-scale modelling, then the model is operated at the smaller scale units with smaller scale equations and parameters and the model output valid for the larger scale emerges after aggregation of the results. The aggregation consists of estimating the spatial mean and in some cases also the statistical distribution of the model outputs. If the model is linear or the parameters and variables are spatially constant, computational time may be saved by averaging of model parameters and input before running the model; otherwise the models runs must be made before the aggregation step.

*Table 1. Characteristics of different scaling procedures when moving from a smaller scale (SS) to a larger scale (LS).*

| | **Aggregation** | **Upscaling** | | |
|---|---|---|---|---|
| | | SS equations used at LS | Large-scale PDE | LS equations developed |
| **Basis of process descriptions** | Smaller scale | Smaller scale | Smaller scale | Larger scale |
| **Computational unit** | Smaller scale | Larger scale | Larger scale | Larger scale |
| **Parameter estimation possible from field data?** | Yes | No, some values need calibration | Yes | No, some values need calibration |

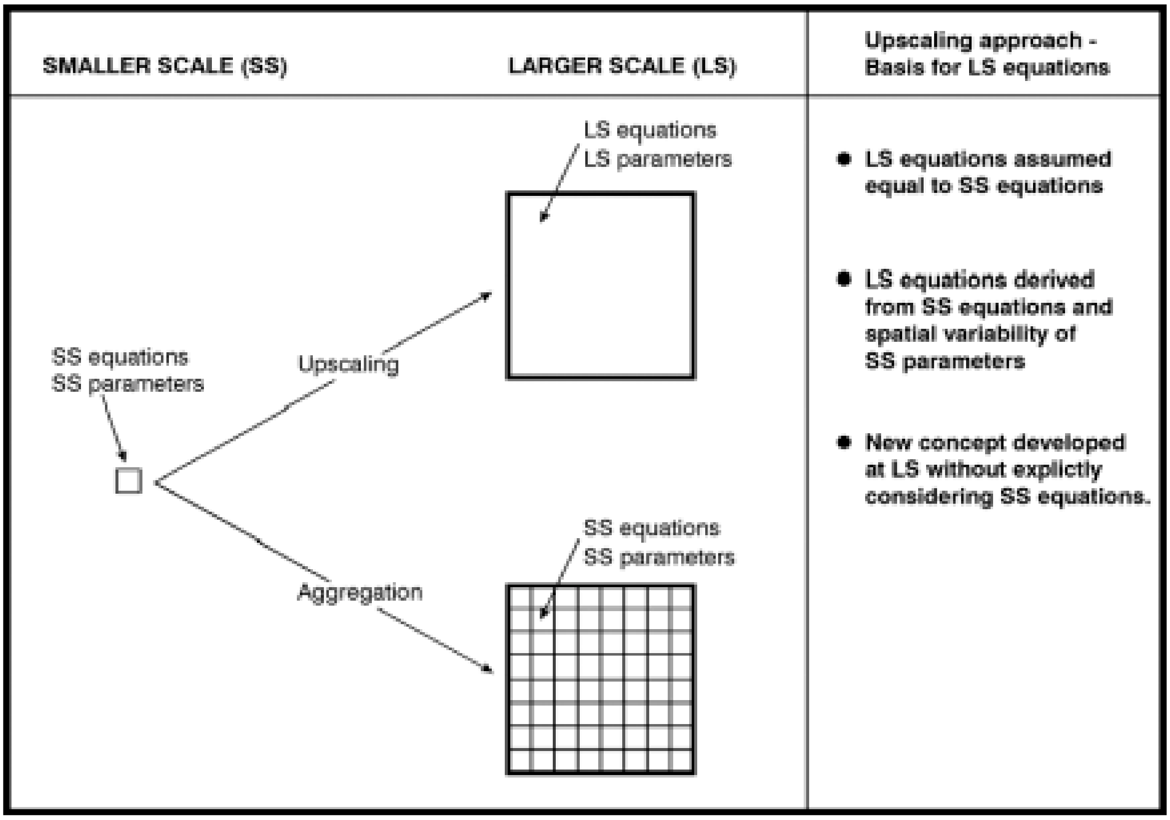



*Fig. 31 Upscaling and aggregation methods for extending hydrological processes from small-scale (SS) to large-scale (LS) models (Refsgaard and Butts, 1999).*

If the upscaling approach is adopted for the large-scale modelling, the smaller scale equations and parameters are in principle substituted by larger scale ones. The upscaling approach can be carried out in three different ways:

- The *smaller scale equations are assumed valid also at the larger scale*. In this case the parameter values have to be estimated as effective parameters corresponding to the larger scale computational unit. Effective parameters are single values, similar to point scale parameters, but somehow reproduce the bulk behaviour of a heterogeneous medium. The estimation of parameter values is in such case often done by calibration, at least for a handful of the key parameters. An example of this approach is given in [5] describing an application of the SHE to a large catchment in India using spatial grid sizes of 2 km x 2 km.
- The equations at the larger scale are derived in a *theoretical framework* from a set of deterministic partial differential equations (PDE) assumed valid at the smaller scale and assumptions on the spatial variability of key parameters and/or input data. This is often carried out in a stochastic framework where quantities such as the average value and higher order statistical moments of the desired model output variables can be assessed. An example of this approach is Jensen and Mantouglou (1992) who consider the spatial variability of soil hydraulic parameters in field scale modelling. In this case the parameter values may be assessed directly on the basis of smaller scale information.
- The *equations at the larger scale are developed at the larger scale* using a concept, which does not explicitly consider the smaller scale equations, i.e. the formulation of laws that apply at the large scale. Examples of this approach are the conceptual rainfall-runoff models such as the NAM (Niel-

sen and Hansen, 1973; [6]; [8]), cf Fig. 28 and the discussion above. The oxygen model described in [3] is also an example of this approach, although smaller scale and larger scale here refer to mm and dm scales and not to catchment scale. As a result of the larger scale concepts such codes are often not adequate also for smaller scale application and can most often not assess parameters directly from small scale information.

### 4.1.3 Scaling - an example

The above four scaling approaches each have their advantages and limitations and the specific approach to use in particular applications will depend on many factors such as the purpose of a given study, the dominating processes in the particular hydrological regime and the data availability. Thus, no unique approach can be claimed superior in all cases. As illustrated below, scaling procedures are in practise often based on combinations of the above approaches.

The example outlines the scaling methodologies adopted under an EU research project dealing with uncertainties of assessing non-point pollution to aquifers at the European scale (Refsgaard et al, 1998; [10]). During this project two model codes were used:

- SMART2 for studying leaching to groundwater of nitrate and aluminium from natural areas due to atmospheric deposition. SMART2 is a relatively simple dynamic model operating in vertical columns with annual time steps (Kros et al., 1995).
- MIKE SHE/DAISY for studying groundwater contamination from agricultural areas. Both MIKE SHE (Refsgaard and Storm, 1995) and DAISY (Hansen et al., 1991) are physically-based model codes with detailed process descriptions and typically hourly time steps.

The objective of the project was to assess the uncertainty in model predictions when applied at the European scale. As both codes had been developed for and previously mainly been applied at much smaller scales a scaling procedure had to be adopted. The two scaling procedures, illustrated in Fig. 32, show significant differences:

SMART 2 is operating at a 1 km grid scale. It was developed on the basis of experience with the NUCSAM code (Groenenberg et al., 1995) which is a detailed physically-based code operating at point scale. Thus, SMART2 can be considered as an upscaling of NUCSAM with new equations and parameters applicable at the 1 km scale, equivalent to the upscaling procedure of the conceptual hydrological models described above. For use for the Netherlands the SMART2 model results were aggregated to 5 km x 5 km grid by selecting the median value among the 25 grids of 1 km x 1 km size. The parameters were assessed by pedotransfer functions from field data without prior model calibration. The scaling procedure from point scale to national or European scales thus consists of a combination of an upscaling and an aggregation step.

MIKE SHE/DAISY, on the other hand, is in this case run with equations and parameter values in each model grid point representing field scale conditions. The field scale is characterised by 'effective' soil and vegetation parameters, but assuming only one soil type and one cropping pattern. The smallest horizontal discretisation in the model is the grid scale (1-5 km) that is larger than the field scale. This implies that all the variations between categories of soil type and crop type within the area of each grid can not be resolved and described at the grid level. Input data, whose variations are not included in the grid scale representation, are distributed randomly at the catchment scale so that their statistical distributions are preserved at that scale. The results from the grid scale modelling are then aggregated to catchment scale (10-50 km) and the statistical properties of model output and field data are then compared at catchment scale (Hansen et al., 1999; [10]). Thus the scaling procedure from point scale to catchment scale is again a combination of an upscaling step and an aggregation step. In contrary to the NUCSAM-SMART2 case the upscaling step here is simply the (important) assumption that the point scale equations are valid at field scale. The aggregation step highlights a key issue from the concept of Representative Elementary Area, REA (Wood et al., 1988), namely that variability can be explicitly represented only at scales larger than the model grid size.

Validation tests against field data suggested that the two different scaling procedures basically could be assumed valid for their respective cases, although important limitations were also identified. An important question regarding the differences between the two upscaling methods is, why it apparently was possible to make the large upscaling step from the smaller scale NUCSAM to the larger scale SMART 2 code, while a similar step was not judged possible for the MIKE SHE/DAISY code. The answer may be that the nitrogen leaching in agricultural fields is a highly non-linear and dynamic process that depends on cropping pattern and agricultural management practise, which can not be lumped to a larger scale description, while the geochemical processes below natural lands, where no management practise is interfering, more easily can be represented by long term average simulations focussing on the gradual reduction of the chemical buffer capacities due to the acids in the atmospheric deposition.

An inherent limitation of the scaling methodologies illustrated in this example is that they do not preserve the georeferenced location of simulated concentrations, but only their statistical distribution over the catchment area (e.g. Fig. 25). Therefore, comparisons with field data make no sense on a well by well or subcatchment by subcatchment basis, and no information on the actual location of the simulated ‘hot spots’ within the catchment is provided. If it from a management point of view is required with a more detailed spatial resolution of the model predictions, then the same scaling method has to be carried out at a finer scale with all the statistical input data being supplied on a subcatchment basis. This is in principle straightforward, but in reality it may often be limited by data availability.

### 4.1.4 Discussion – post evaluation

The issue of scaling represents both a major scientific challenge and a practical problem in water resources management. Scaling is dealt with as a key issue in two of the publications in this thesis ([7], [10]). As the studies behind the other publications operate on scales ranging from point scale ([3]) to thousands of $km^2$ ([4], [5], [9]) catchment heterogeneity and scaling are dealt with and discussed in many of the publications.

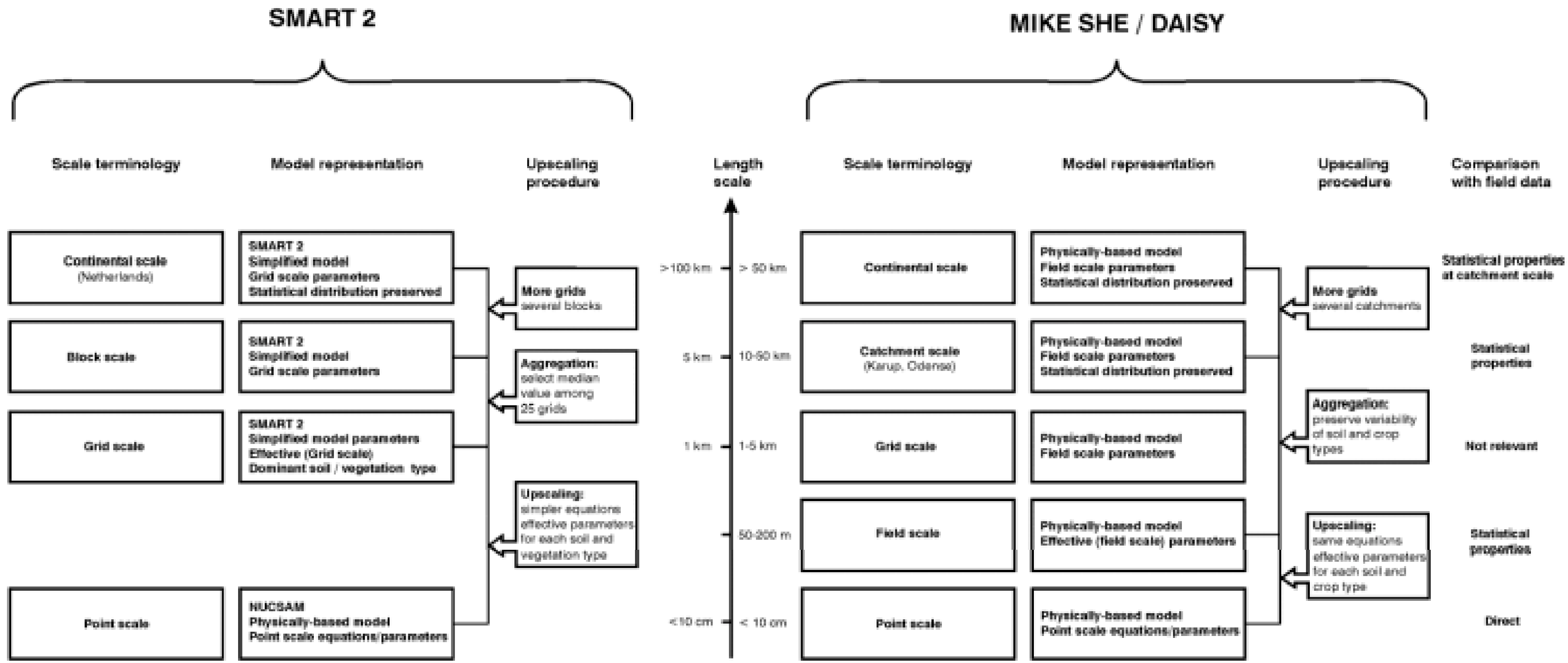



*Fig. 32 Scaling methodology adopted by the SMART2 and MIKE SHE/DAISY models in the UNCERSDSS project (Refsgaard and Butts, 1999).*

In the beginning of my career I had the rather naive view that it might be possible to develop a universal model code and a methodology that could be used to address most problems in hydrological management. This is reflected in the dualism of statements of the MIKE SHE description in Refsgaard and Storm (1995), where it on the one hand is stated that "MIKE SHE is applicable on spatial scales ranging from a single soil profile to a large regions", while it on the other hand is acknowledged that "there are a number of fundamental scale problems which need to be carefully considered in the model applications". I do not believe any longer that a universally applicable code and modelling methodology is theoretically realistic, and certainly it is not feasible in practise. The main reason for this is the scaling problems. Because scaling is interlinked with modelling concepts, I therefore do not believe it will ever be possible to derive a universal scaling theory of practical applicability.

Scaling implies to take spatial heterogeneity into account. In catchment modelling it is furthermore complicated by the need to include and link several processes, such as subsurface processes (Dagan, 1986; Gelhar, 1986; Wen and Gómez-Hernández, 1996), root zone processes including land surface-atmosphere interaction (Michaud and Shuttelworth, 1997); and surface water processes including stream-aquifer interaction (Saulnier et al., 1997; [7]).

Many researchers have expressed doubts whether it is feasible to use the same model process descriptions at different scales. For instance Beven (1995) states that "… the aggregation approach towards macroscale hydrological modelling, in which it is assumed that a model applicable at small scales can be applied at larger scales using 'effective' parameter values, is an inadequate approach to the scale problem. It is also unlikely in the future that any general scaling theory can be developed due to the dependence of hydrological systems on historical and geological perturbations."

Beven's view can be considered a universal and fundamental statement to which it is difficult to disagree. A more pragmatic, but not necessarily conflicting, view is expressed by Grayson and Blöschl (2000): "As modellers, we are often left with little choice but to use the effective parameter approach, but we must recognise that effective parameters may have a narrow range of application and an effective parameter value that "works" for one process may not be valid for another process." The scaling framework presented above should be seen in this context. It is not a fundamental theory but rather a collection of different methods and an emphasis on their respective assumptions and associated costs in terms of lost information. These methods or building blocks can then be used in composing specific scaling methodologies depending on the purposes of the particular modelling studies. In this respect it is crucial that the modeller is aware of the limitations of the scaling methodology chosen in a particular study.

## 4.2 Confirmation, Verification, Calibration and Validation

As illustrated in Fig. 3 the credibility of the descriptions or the agreements between reality, conceptual model, model code and model are evaluated through confirmation of the conceptual model, verification of the code, model calibration and model validation. These four terms are addressed in this section.

### 4.2.1 Confirmation of conceptual model

The conceptual model, with its selection of process descriptions, equations, etc., is the foundation for the model structure. Therefore a good conceptual model is most often a prerequisite for obtaining trustworthy model results. In groundwater modelling, establishment of the conceptual model is often considered the most important part of the entire modelling process (Middlemis, 2000). Evaluation of conceptual models is an important part in assessing uncertainty due to model structure error (Section 4.3 below and [15]).

Methods for conceptual model confirmation should follow the standard procedures for confirmation of scientific theories. This implies that conceptual models should be confronted with actual field data and be subject to critical peer reviews. Furthermore, the feedback from the calibration and validation process may also serve as a means by which one or a number of alternative conceptual models may be either confirmed or falsified.

As Beven (2002b) argues we need to distinguish between our qualitative understanding (perceptual model) and the practical implementation of that understanding in our conceptual model. As a conceptual model is defined in [12] as combination of a perceptual model and the simplifications acceptable for a particular model study a conceptual model becomes site-specific and even case specific. For example a conceptual model of a groundwater aquifer may be described as two-dimensional for a study focussing on regional groundwater heads, while it may need to include more complex three-dimensional geological structures for a study requiring detailed solute transport simulations.

### 4.2.2 Code verification

The ability of a given model code to adequately describe the theory and equations defined in the conceptual model by use of numerical algorithms is evaluated through the verification of the model code. Use of the term verification in this respect is in accordance with Oreskes et al. (1994), because mathematical equations are closed systems. The methodologies used for code verification include comparing a numerical solution with an analytical solution or with a numerical solution from other verified codes. However, some programme errors only appear under circumstances that do not routinely occur, and may not have been anticipated. Furthermore, for complex codes it is virtually impossible to verify that the code is universally accurate and error-free. Therefore, the term code verification must be qualified in terms of specified ranges of application and corresponding ranges of accuracy.

Code verification is not an activity that is carried out from scratch in every modelling study. In a particular study it has to be ascertained that the domain of applicability for which the selected model code has been verified covers the conditions specified in the actual conceptual model. If that is not the case, additional verification tests have to be conducted. Otherwise, the code explicitly must be classified as not verified for this particular study, and the subsequent simulation results therefore have to be considered with extra caution.

### 4.2.3 Model calibration

The application of a model code to be used for setting up a site-specific model is usually associated with model calibration. The model performance during calibration depends on the quantity and quality of the available input and observation data as well as on the conceptual model. If sufficient accuracy cannot be achieved either the conceptual model and/or the data have to be re-evaluated.

Many of the publications ([1], [4], [5], [6], [7], [8], [9]) have involved model calibration. This was in all cases done manually. Today automatic calibration (inverse modelling) is state-of-the-art (Duan et al., 1994; Hill, 1998; Doherty, 2003), also as part of the calibration process for rather complex distributed physically-based models (Sonnenborg et al., 2003; Henriksen et al., 2003).

A key issue related to calibration of distributed models with potentially hundreds or thousands of parameter values is a rigorous parameterisation procedure, where the spatial pattern of the parameter values are defined and the number of free parameters adjustable through calibration is reduced as much as possible. A methodology for this is presented in [7], and this issue is further discussed in [4], [5], [10] and Andersen et al. (2001).

### 4.2.4 Model validation

Often the model performance during calibration is used as a measure of the predictive capability of a model. This is a fundamental error. Many studies (e.g. [4]; [6]; Andersen et al., 2001) have demonstrated that the model performance against independent data not used for calibration is generally poorer than the performance achieved in the calibration situation. Therefore, the credibility of a site-specific model's capability to make predictions about reality must be evaluated against independent data. This process is denoted model validation.

In designing suitable model validation tests a guiding principle should be that a model should be tested to show how well it can perform the kind of task for which it is specifically intended (Klemes, 1986). Klemes proposed the following scheme comprising four types of test corresponding to different situations with regard to whether data are available for calibration and whether the catchment conditions are stationary or the impact of some kind of intervention has to be simulated:

- The *split-sample test* is the classical test, being applicable to cases where there is sufficient data for calibration and where the catchment conditions are stationary. The available data record is divided into two parts. A calibration is carried out on one part and then a validation on the other part. Both the calibration and validation exercises should give acceptable results.
- The *proxy-basin test* should be applied when there is not sufficient data for a calibration of the catchment in question. If, for example, streamflow has to be predicted in an ungauged catchment Z, two gauged catchments X and Y within the region should be selected. The model should be calibrated on catchment X and validated on catchment Y and vice versa. Only if the two validation results are

acceptable and similar can the model command a basic level of credibility with regard to its ability to simulate the streamflow in catchment Z adequately.

- The *differential split-sample test* should be applied whenever a model is to be used to simulate flows, soil moisture patterns and other variables in a given gauged catchment under conditions different from those corresponding to the available data. The test may have several variants depending on the specific nature of the modelling study. If for example a simulation of the effects of a change in climate is intended, the test should have the following form. Two periods with different values of the climate variables of interest should be identified in the historical record, such as one with a high average precipitation and the other with a low average precipitation. If the model is intended to simulate streamflow for a wet climate scenario, then it should be calibrated on a dry segment of the historical record and validated on a wet segment. Similar test variants can be defined for the prediction of changes in land use, effects of groundwater abstraction and other such changes. In general, the model should demonstrate an ability to perform through the required transition regime.
- The *proxy-basin differential split-sample test* is the most difficult test for a hydrological model, because it deals with cases where there is no data available for calibration and where the model is directed to predicting non-stationary conditions. An example of a case that requires such a test is simulation of hydrological conditions for a future period with a change in climate and for a catchment, where no calibration data presently exist. The test is a combination of the two previous tests.

The above test types are very general and needs to be translated to specific tests in each case depending on data availability, hydrological regime and purpose of the modelling study. Except for the situations, where the split-sample test is sufficient, rather limited work has been carried out so far on validation test schemes.

From a theoretical point of view the procedures outlined by Klemes (1986) for the proxy-basin and the differential split-sample tests, where tests have to be carried out using data from similar catchments, are weaker than the usual split-sample test, where data from the specific catchment are available. However, no obviously better testing schemes exist.

It must be realised that the validation test schemes proposed above are so demanding that many applications today would fail to meet them. Thus, for many cases where either proxy-basin or differential split-sample tests are required, suitable test data simply do not exist. This is for example the case for prediction of regional scale transport of potential contamination from underground radionuclide deposits over the next thousands of years. In such case model validation is not possible. This does not imply that these modelling studies are not useful, only that their output should be recognised to be somewhat more uncertain than is often stated and that the term 'validated model' should not be used. Thus, a model's validity will always be confined in terms of space, time, boundary conditions, types of application, etc.

### 4.2.5 Discussion – post evaluation

Relative to confirmation, verification and calibration, the main scientific contributions in my publications [1] – [15] are on the model validation issue. The motivation for this research was twofold: First of all, there were too many undocumented claims (over-selling) in the modelling community on model capabilities during the years following the development of many comprehensive model codes such as MIKE

SHE. This over-selling was most obvious in practical studies conducted by consultants, but it was also common in large parts of the scientific community, e.g. Abbot et al. (1986a,b) and many others. Secondly, dominant parts of the hydrological scientific community advocated that model validation was not possible (Konikow and Bredehoeft, 1992; Beven, 1996a). This left the practising world in a vacuum without scientifically based methodologies to test and document the degree of credibility of particular model predictions. The methodologies described in [6] and [7] should be seen as pragmatic approaches to help filling this vacuum and the discussions in [12] should be seen as an attempt to provide a scientific basis for adopting rigorous model validation schemes as part of a good modelling practise.

The principles and schemes proposed by Klemes have been extensively used in the last 12 of the publications ([4] – [15]). Thus, the intercomparison study in [6] was based on a rigorous use of all four types of tests. Furthermore, [7] 'translated' Klemes' principles that were developed with lumped conceptual models in mind to use in distributed modelling. After demonstrating that a distributed model that was validated for simulating catchment response often performs much poorer for internal sites, [7] emphasised that a model should only be assumed valid with respect to the outputs that have been directly validated. This implies e.g. that multi-site validation is needed if predictions of spatial patterns are required. Furthermore, a model which is validated against catchment runoff can not automatically be assumed valid also for simulation of erosion on a hillslope within the catchment, because smaller scale processes may dominate here; it will need specific validation against hillslope soil erosion data. Furthermore, systematic split-sample tests were made in [4], [5] and [9], and proxy- basin tests were conducted in [10]. Finally, the validation requirements are emphasised in the publications related to quality assurance [12] and [13].

[6] and [7] were not the first studies to use Klemes' principles for validation. For example Quinn and Beven (1993) used split sample-tests, proxy-basin tests and differential split-sample tests (wet/dry periods) to analyse TOPMODEL's predictive capabilities for the Plynlimon catchment in Wales. The key contribution of [7] and [12] in this respect was the integration of Klemes' principles as core elements of a protocol for good modelling practise.

The principles outlined in [7] and consolidated in [12] that a model should never be considered universally validated, but can only be conditionally validated restricted by the availability of data and specifically performed validation tests are well in line with Lane and Richards (2001) who argue that "evidence of a successful prediction in observed spaces and times (conventional validation) cannot provide a sufficient basis for use of a model beyond the set of situations for which the model has been empirically tested". The principles are also in accordance with the new coherent philosophy for modelling of the environment proposed by Beven (2002b) where he argues that it is required to be able to "define those areas of the model space where behavioural models occur".

## 4.3 Uncertainty Assessment

This section presents a broad framework originating from Refsgaard et al. (2005) and [14] followed by a discussion on data uncertainty (including [14]), parameter uncertainty (including [11]) and model structure uncertainty (including [15]) and how they affect model output uncertainty.

### 4.3.1 Modelling uncertainty in a water resources management context

**Definitions and Taxonomy**

Uncertainty and associated terms such as error, risk and ignorance are defined and interpreted differently by different authors (see Walker et al. (2003) for a review). The different definitions reflect, among other factors, the different scientific disciplines and philosophies of the authors involved, as well as the intended audience. In addition they vary depending on their purpose. Here I will use the terminology used in Refsgaard et al. (2005) and [14] that has emerged after discussions between social scientists and natural scientists specifically aiming at applications in model based water management (Klauer and Brown, 2003). It is based on a subjective interpretation of uncertainty in which the *degree of confidence* that a decision maker has about possible outcomes and/or probabilities of these outcomes is the central focus. Thus, according to this definition *a person is uncertain if s/he lacks confidence about the specific outcomes of an event. Reasons for this lack of confidence might include a judgement that the information is incomplete, blurred, inaccurate, imprecise or potentially false.* Similarly, a person is certain if s/he is confident about the outcome of an event. It is possible that a person feels certain but has misjudged the situation (i.e. s/he is wrong).

There are many different (decision) situations, with different possibilities for characterising of what we know or do not know and of what we are certain or uncertain. A first distinction is between ignorance as a lack of awareness about imperfect knowledge and uncertainty as a state of confidence about knowledge (which includes the act of ignoring). Our state of confidence may range from being certain to admitting that we know nothing (of use), and uncertainty may be expressed at a number of levels in between. Regardless of our confidence in what we know, ignorance implies that we can still be wrong ('in error'). In this respect Brown (2004) has defined a taxonomy of imperfect knowledge illustrated in Fig. 33.

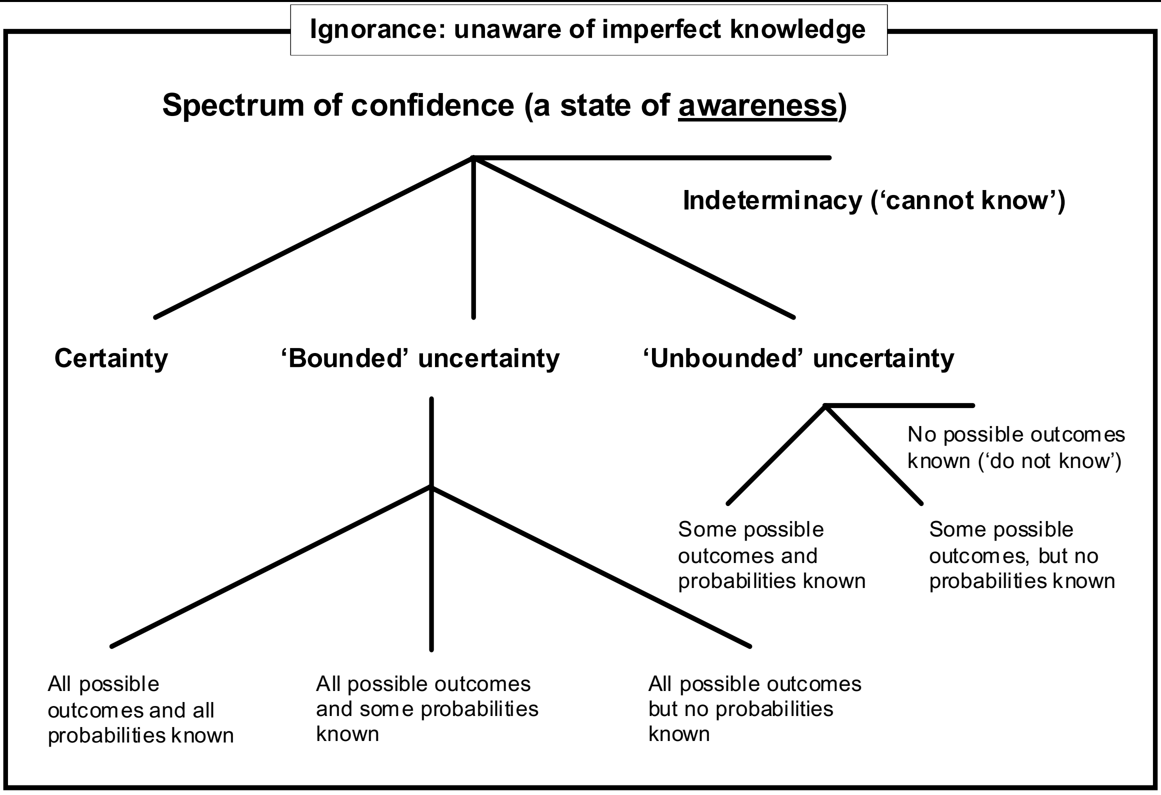



*Fig. 33 Taxonomy of imperfect knowledge resulting in different uncertainty situations (Brown, 2004)*

In evaluating uncertainty, it is useful to distinguish between uncertainty that can be quantified e.g. by probabilities and uncertainty that can only be qualitatively described e.g. by scenarios. If one throws a balanced die, the precise outcome is uncertain, but the 'attractor' of a perfect die is certain: we know precisely the probability for each of the 6 outcomes, each being 1/6. This is what we mean with 'uncertainty in terms of probability'. However, the estimates for the probability of each outcome can also be uncertain. If a model study says: "there is a 30% probability that this area will flood two times in the next year", there is not only 'uncertainty in terms of probability' but also uncertainty regarding whether the estimate of 30% is a reliable estimate.

Secondly, it is useful to distinguish between bounded uncertainty, where all possible outcomes have been identified and unbounded uncertainty, where the known outcomes are considered incomplete. Since quantitative probabilities require 'all possible outcomes' of an uncertain event and each of their individual probabilities to be known, they can only be defined for 'bounded uncertainties'. If probabilities cannot be quantified in any undisputed way, we often can still qualify the available body of evidence for the possibility of various outcomes.

The bounded uncertainty where all probabilities are deemed known (Fig. 33) is often denoted 'statistical uncertainty' (e.g. Walker et al., 2003). This is the case traditionally addressed in model based uncertainty assessment. It is important to note that this case constitutes one of many decision situations outlined in Fig. 33, and in other situations the main uncertainty in a decision situation cannot be characterised statistically.

**Sources of uncertainty**

Walker et al. (2003) describes the uncertainty as manifesting itself at different locations in the model based water management process. These locations, or sources, may be characterised as follows:

- *Context*, i.e. at the boundaries of the system to be modelled. The model context is typically determined at the initial stage of the study where the problem is identified and the focus of the model study selected as a confined part of the overall problem. This includes, for example, the external economic, environmental, political, social and technological circumstances that form the context of problem.
- *Input uncertainty* in terms of external driving forces (within or outside the control of the water manager) and system data that drive the model such as land use maps, pollution sources and climate data.
- *Model structure uncertainty* is the conceptual uncertainty due to incomplete understanding and simplified descriptions of processes as compared to nature.
- *Parameter uncertainty*, i.e. the uncertainties related to parameter values.
- *Model technical uncertainty* is the uncertainty arising from computer implementation of the model, e.g. due to numerical approximations and bugs in the software.
- *Model output uncertainty*, i.e. the total uncertainty on the model simulations taken all the above sources into account, e.g. by uncertainty propagation.

**Nature of uncertainty**

Many authors (e.g. Walker et al., 2003) categorise the *nature* of uncertainty into:

- *Epistemic uncertainty*, i.e. the uncertainty due to imperfect knowledge.
- *Stochastic uncertainty*, i.e. uncertainty due to inherent variability, e.g. climate variability.

Epistemic uncertainty is reducible by more studies: e.g. research or data collection. Stochastic uncertainty is non-reducible.

Often the uncertainty on a certain event includes both epistemic and stochastic uncertainty. An example is the uncertainty of the 100 year flood at a given site. This flood event can be estimated: e.g. by use of standard flood frequency analysis on the basis of existing flow data. The *(epistemic)* uncertainty may be reduced by improving the data analysis, by making additional monitoring (longer time series) or by a deepening our understanding of how the modelled system works. However, no matter how much we improve our knowledge, there will always be some (*stochastic*) uncertainty inherent to the natural system, related to the stochastic and chaotic nature of several natural phenomena, such as weather. Perfect knowledge on these phenomena cannot give us a deterministic prediction, but would have the form of a perfect characterisation of the natural variability; for example, a probability density function for rainfall in a month of the year.

**The uncertainty matrix**

The uncertainty matrix in Table 2 can be used as a tool to get an overview of the various sources of uncertainty in a modelling study. The matrix is modified after Walker et al. (2003) in such a way that it matches Fig. 33 and so that the taxonomy now gives ‘uncertainty type’ in descriptions that indicates in what terms uncertainty can best be described. The vertical axis identifies the source of uncertainty while the horizontal axis covers the level and nature of uncertainty. It is noticed that the matrix is in reality three-dimensional (source, type, nature), because the categories Type and Nature are not mutually exclusive

*Table 2 The uncertainty matrix (modified after Walker et al., 2003).*

| | | **Taxonomy (types of uncertainty)** | | | | **Nature** | |
|---|---|---|---|---|---|---|---|
| **Source of uncertainty** | | Statistical uncertainty | Scenario uncertainty | Qualitative uncertainty | Recognised ignorance | Epistemic uncertainty | Stochastic uncertainty |
| Context | Natural, technological, economic, social, political | | | | | | |
| Inputs | System data | | | | | | |
| | Driving forces | | | | | | |
| Model | Model structure | | | | | | |
| | Technical | | | | | | |
| | Parameters | | | | | | |
| Model outputs | | | | | | | |

**Methodologies for assessing uncertainty**

A list of the most common methodologies applicable for addressing different types of uncertainty has been compiled and briefly described in Refsgaard et al. (2005). Table 3 provides an overview.

*Table 3 Applicability of different methodologies to address different types and sources of uncertainty (modified after Refsgaard et al., 2005).*

| Source of uncertainty | | Taxonomy (types of uncertainty) | | | |
|---|---|---|---|---|---|
| | | Statistical uncertainty | Scenario uncertainty | Qualitative uncertainty | Recognised ignorance |
| Context | Natural, technological, economic, social, political | EE | EE, SC, SI | EE, EPR, NUSAP, SI, UM | EE, EPR, NUSAP, SI, UM |
| Inputs | System data | DA, EPE, EE, MCA, SA | DA, EE, SC | DA, EE | DA, EE |
| | Driving forces | DA, EPE, EE, MCA, SA | DA, EE, SC | DA, EE, EPR | DA, EE, EPR |
| Model | Model structure | EE, MMS, QA | EE, MMS, SC, QA | EE, NUSAP, QA | EA, NUSAP, QA |
| | Technical | QA | QA | QA | QA |
| | Parameters | EE, IN-PA, SA | EE, IN-PA, SA | EE | EE |
| Model outputs | | EPE, EE, IN-UN, MCA, MMS, SA | EE, IN-UN, MMS, SA | EE, NUSAP | EE, NUSAP |

*Abbreviations of methodologies:*

| | |
|---|---|
| DA | Data Uncertainty |
| EPE | Error Propagation Equations |
| EE | Expert Elicitation |
| EPR | Extended Peer Review (review by stakeholders) |
| IN-PA | Inverse modelling (parameter estimation) |
| IN-UN | Inverse modelling (predictive uncertainty) |
| MCA | Monte Carlo Analysis |
| MMS | Multiple Model Simulation |
| NUSAP | NUSAP |
| QA | Quality Assurance |
| SC | Scenario Analysis |
| SA | Sensitivity Analysis |
| SI | Stakeholder Involvement |
| UM | Uncertainty Matrix |

### 4.3.2 Data uncertainty

Uncertainty in data is a major source of uncertainty when assessing uncertainty of model outputs. It is also an uncertainty source that is very visible for people outside the modelling community. One of the scientific contributions of the HarmoniRiB project ([14]) is to address data uncertainty. This has been done in three steps:

- A methodology has been developed for characterising uncertainty in different types of data (Brown et al., 2005).
- A software tool (Data Uncertainty Engine – DUE) for supporting the assessment of data uncertainty has been developed (Brown and Heuvelink, 2005).
- Reviews with results on data uncertainty reported in the literature have been compiled into a guideline report for assessing uncertainty in various types of data originating from meteorology, soil physics and geochemistry, hydrogeology, land cover, topography, discharge, surface water quality, ecology and socio-economics (Van Loon and Refsgaard, 2005).

The categorisation of data types distinguishes 13 categories (Table 4) for each of which a conceptual data uncertainty model is developed. By considering measurement scale, it becomes possible to quickly limit the relevant uncertainty models for a certain variable. On a discrete measurement scale, for example, it is only relevant to consider discrete probability distribution functions, whereas continuous density functions are required for continuous numerical data. In addition, the use of space and time variability determines the need for autocorrelation functions alongside a probability density function (*pdf)*. Each data category is associated with a range of uncertainty models, for which more specific pdfs may be developed with different simplifying assumptions (e.g. Gaussian; second-order stationarity; degree of temporal and spatial autocorrelation).

*Table 4 The subdivision of uncertainty categories, along the 'axes' of space-time variability and measurement scale (Brown et al., 2005).*

| **Space-time variability** | **Measurement scale** | | | |
|---|---|---|---|---|
| | Continuous numerical | Discrete numerical | Categorical | Narrative |
| Constant in space and time | **A1** | **A2** | **A3** | **4** |
| Varies in time, not in space | **B1** | **B2** | **B3** | |
| Varies in space, not in time | **C1** | **C2** | **C3** | |
| Varies in time and space | **D1** | **D2** | **D3** | |

### 4.3.3 Parameter uncertainty

In addition to data uncertainty, uncertainty of parameter values is the most commonly considered source of uncertainty in hydrological modelling. The scientifically soundest way of assessing parameter uncertainty is through inverse modelling (Duan et al., 1994; Hill, 1998; Doherty, 2003). These techniques have the benefit that they, in addition to optimal parameter values, also produce calibration statistics in terms of parameter- and observation sensitivities, parameter correlation and parameter uncertainties.

When parameter uncertainties are assessed they can be propagated through the model to infer about model output uncertainty. A serious constraint in this respect is the interdependence between model parameters and model structure as discussed under model structure uncertainty below.

[11] describe an example of how (input) data uncertainty and parameter uncertainty are propagated through a model to assess uncertainty in model simulation of nitrate concentrations in groundwater. The assessment of data and parameter values were done by expert judgement and a Monte Carlo technique with Latin hypercube sampling was used for the uncertainty propagation. The simulated uncertainty band around the deterministic model simulation in Fig. 25 is shown in Fig. 34 based on 25 Monte Carlo realisations. The uncertainty is seen to be considerable, e.g. with the estimate of the areal fraction of the aquifer having concentrations less than 50 mg $NO_3$/l ranging between 30% and 80%.

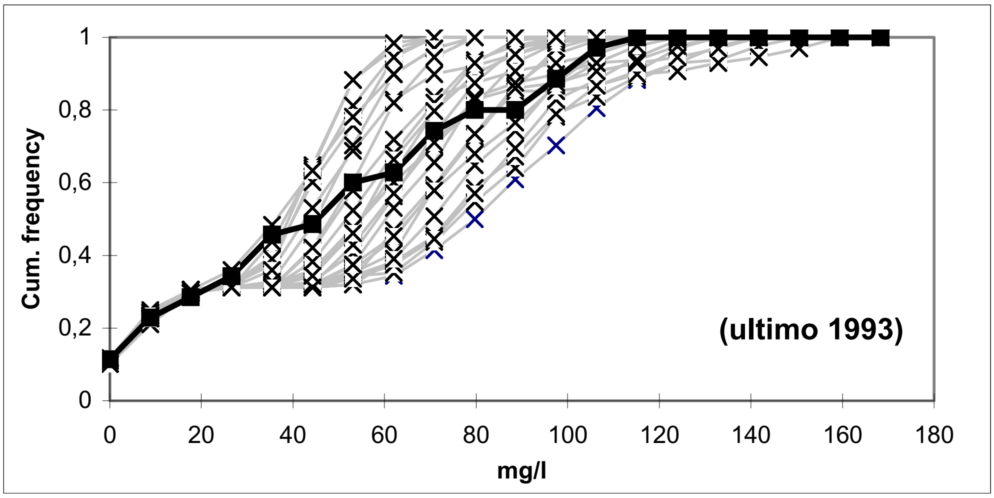



*Fig. 34 Measured (Â) and simulated (Δ) areal distribution of $NO_3$ concentrations in groundwater at a point in time. Measured values are based on 35 groundwater observations. [11].*

As noted in [11] a fundamental limitation of the approach adopted in [11] is that the errors due to incorrect model structure are neglected. As discussed also below one approach to assess such model structure error is through comparison of predicted and observed values. In the present case (Figs 25 and 34) the deviation between observed and simulated values is so small that this term may be neglected. This is, however, by no means a proof of a correct model structure. It only shows that the particular model performs without apparent model errors for this particular application.

### 4.3.4 Model structure uncertainty

**Existing approaches and new framework**

Any model is an abstraction, simplification and interpretation of reality. The incompleteness of a model structure and the mismatch between the real causal structure of a system and the assumed causal structure as represented in a model will therefore always result in uncertainty about model predictions. The importance of the model structure for predictions is well recognised, even for situations where predictions are made on output variables, such as discharge, for which field data are available (Franchini and Pacciani, 1992; Butts et al., 2004). The considerable challenge faced in many applications of environmental models is that predictions are required beyond the range of available observations, either in time or in space, e.g. to make extrapolations towards unobservable futures (Babendreier, 2003) or to make predictions for natural systems, such as ecosystems, that are likely to undergo structural changes (Beck, 2005). In such cases, uncertainty in model structure is recognised by many authors to be the main source of uncertainty in model predictions (Dubus et al., 2003; Neumann and Wierenga, 2003; Linkov and Burmistrov, 2003).

The existing strategies for assessing uncertainty due to incomplete or inadequate model structure may be grouped into the categories shown in Fig. 35. The most important distinction is whether data exist that makes it possible to infer directly on the model structure uncertainty. This requires that data are available for the output variable of predictive interest and for conditions similar to those in the predictive situation. In other words it is a distinction between whether the model predictions can be considered as interpolations or extrapolations relative to the calibration situation.

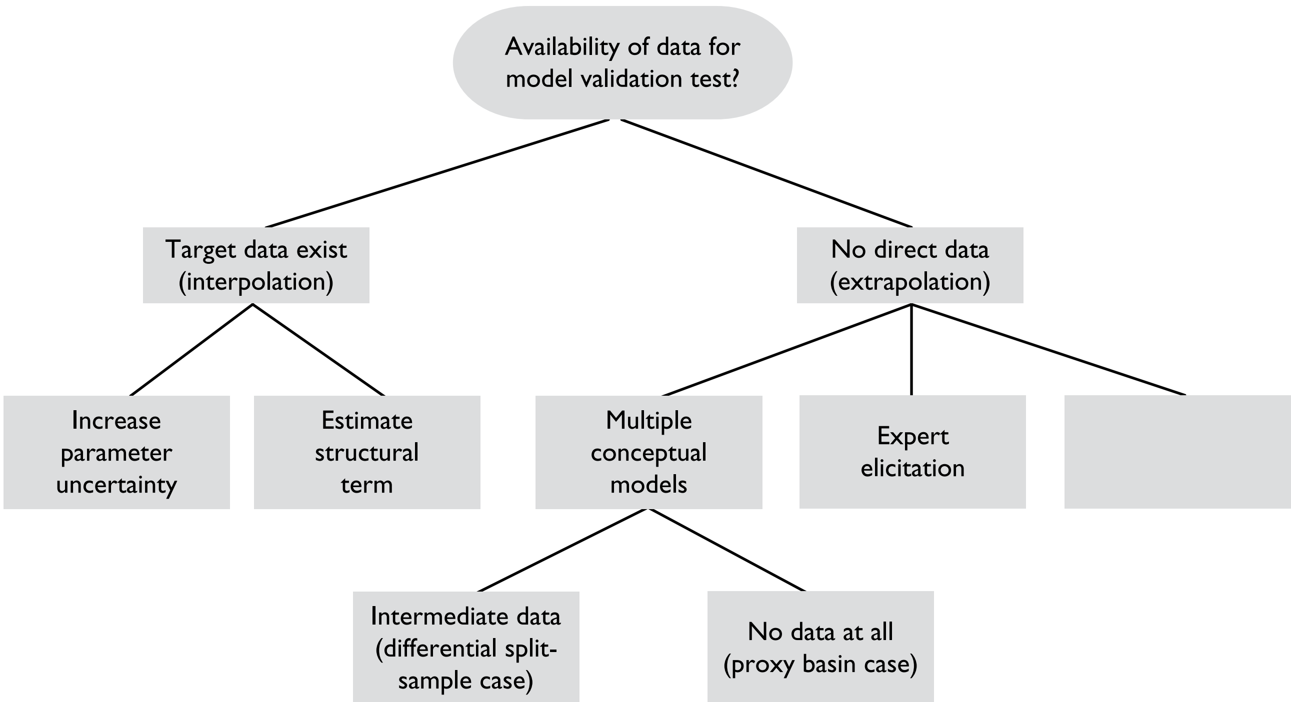



*Fig. 35 Classification of existing strategies for assessing conceptual model uncertainty [15].*

The two main categories are thus equivalent to different situations with respect to model validation tests. According to Klemes' classical hierarchical test scheme (Klemes, 1986; see Section 4.2 above), the interpolation case corresponds to situations where the traditional split-sample test is suitable, while the extrapolation case corresponds to situations where no data exist for the concerned output variable (proxy-basin test) or where the basin characteristics are considered non-stationary, e.g. for predictions of effects of climate change or effects of land use change (differential split-sample test).

The strategies used in 'interpolation', i.e. for situations that are similar to the calibration situation with respect to variables of interest and conditions of the natural system, have the advantage that they can be based directly on field data (e.g. Radwan et al., 2004; van Griensven and Meixner, 2004; and Vrugt et al., 2005). A fundamental weakness is that field data are themselves uncertain. Nevertheless, in many cases, they can be expected to provide relatively accurate estimates of, at least, the total predictive uncertainty for the specific measured variable and for the same conditions as those in the calibration and validation situation. A more serious limitation of the strategies depending on observed data is that they are only applicable for situations where the output variables of interest are measured. While relevant field data are often available for variables such as water levels and water flows, this is usually not the case for concentrations, or when predictions are desired for scenarios involving catchment change, such as land use change or climate change. Another serious limitation stems from an assumption that the underlying system does not undergo structural changes, such as changes in ecosystem processes due to climate change.

The strategy that uses multiple conceptual models benefits from an explicit analysis of the effects of alternative model structures, e.g. IPCC (2001), Harrar et al. (2003), Troldborg (2004), Poeter and Anderson (2005) and Højberg and Refsgaard (2005). The multiple conceptual model strategy makes it possible to include expert knowledge on plausible model structures. This strategy is strongly advocated by Neuman and Wierenga (2003) and Poeter and Anderson (2005). They characterise the traditional approach of relying on a single conceptual model as one in which plausible conceptual models are rejected (in this case by omission). They conclude that the bias and uncertainty that results from reliance on an inadequate conceptual model are typically much larger than those introduced through an inadequate choice of model parameter values. This view is consistent with Beven (2002b) who outlines a new philosophy for modelling of environmental systems. The basic aim of his approach is to extend traditional schemes with a more realistic account of uncertainty, rejecting the idea that a single optimal model exists for any given case. Instead, environmental models may be non-unique in their accuracy of both reproduction of observations and prediction (i.e. unidentifiable or equifinal), and subject to only a conditional confirmation, due to e.g. errors in model structure, calibration of parameters and period of data used for evaluation.

A weakness of the multiple modelling strategy, is the absence of quantitative information about the extent to which each model is plausible. Furthermore, it may be difficult to sample from the full range of plausible conceptual models. In this respect, expert knowledge on which the formulations of multiple conceptual models are based, is an important and unavoidable subjective element.

The framework presented in [15] for assessing the predictive uncertainties of environmental models used for extrapolation includes a combination of use of multiple conceptual models and assessment by use of the pedigree approach of their credibility as well as a reflection on the extent to which the sampled models adequately represent the space of plausible models.

**The role of model calibration**
Some of the existing strategies used in 'interpolation' cannot differentiate how the total predictive uncertainty originates from model input, model parameter and model structure uncertainty. Other methods attempt to do so, but as discussed in [15] this is problematic. In the case of uncalibrated models, the parameter uncertainty is very difficult to assess quantitatively, and wrong estimates of model parameter uncertainty will influence the estimates of model structure uncertainty. In the case of calibrated models, estimates of model parameter uncertainty can often be derived from autocalibration routines. An inadequate model structure will, however, be compensated by biased parameter values to optimise the model fit with field data during calibration. Hence, the uncertainty due to model structure will be underestimated in this case.

The importance of model calibration can be illustrated by the example described in Højberg and Refsgaard (2005). They use three different conceptual models, based on three alternative geological interpretations, for a multi-aquifer system in Denmark. Each of the models was calibrated against piezometric head data using inverse technique. The three models provided equally good and very similar predictions of groundwater heads, including well field capture zones. However, when using the models to extrapolate beyond the calibration data to predictions of flow pathways and travel times the three models differed dramatically. When assessing the uncertainty contributed by the model parameter values, the overlap of uncertainty ranges between the three models significantly decreased when moving from groundwater heads to capture zones and travel times. They conclude that the larger the degree of extrapolation, the more the underlying conceptual model dominates over the parameter uncertainty and the effect of calibration.

This diminishing effect of calibration as the prediction situation is extrapolated further and further away from the calibration base resembles the conclusion on the effects of updating relative to the underlying process model, when forecast lead times are increased in real-time forecasting (Fig. 27, Section 3.3). Here the effect of updating is reduced and the forecast error therefore increases as the forecast lead time (= degree of extrapolation) increases.

### 4.3.5 Discussion – post evaluation

Uncertainty is a key, and crosscutting, issue that I consider a useful platform or catalyst for establishing a common understanding in hydrological modelling and water resources management. By this I mean both a common understanding within the natural science based modelling issues such as scaling and validation and between people from the modelling and the monitoring communities as well as a broader dialogue between modellers and stakeholders on issues such as when is a model accurate and credible enough for its purpose of application, see Subsection 4.4.4 below.

In the publications on developing the Suså model ([1], [2]) and the oxygen module ([3]) no explicit consideration is given to the goodness of the model structure and uncertainty assessment was not an issue at all. In the later work on catchment modelling in India ([4], [5]), where some twisting was done of the physical realism of the model due to scaling problems, it was noted that the model results might be 'right for the wrong reasons', and the limitations of model applicability were emphasised in this respect, but no uncertainty assessments were made. In the paper describing a methodology for parameterisation, calibration and validation of distributed hydrological models ([7]) uncertainty is also neglected. In the publications [6], [8], [9] and [10] uncertainty is discussed, but as a secondary issue only.

Although examples of model prediction uncertainty assessments had been reported previously from different modelling disciplines (e.g. Refsgaard et al., 1983; Beck, 1987), the fist to emphasise the need to systematically perform uncertainty assessments related to catchment model predictions was probably Beven (1989). This was followed by Binley et al. (1991) who used Monte Carlo analysis to assess the predictive uncertainty for the Institute of Hydrology Distributed Model and by the introduction of the Generalised Likelihood Uncertainty Estimation (GLUE) methodology (Beven and Binley, 1992) after which uncertainty in catchment modelling was high on the agenda in the scientific community.

My main scientific contributions on uncertainty are the publications [11], [14] and [15] and the link of uncertainty to principles and protocols for good modelling practise in [12] and [13]. Although reported 10 years later than Binley et al. (1991), [11] was one of the first studies with uncertainty propagation through a complex, coupled distributed physically based catchment model with a focus on water quality. A key contribution of [14] and Refsgaard et al. (2005) is the broad framework for characterising uncertainty. This framework provides the link to uncertainty in the quality assurance work ([12], [13]). This broad framework is inspired by research in social science (Pahl-Wostl, 2002; van Asselt and Rotmans, 2002; Dewulf et al., 2005). The main difference between the traditions in social science and natural science is that social scientists emphasise participatory processes including consultation and involvement of users, also on uncertainty aspects, right from the beginning of a study, while natural scientists often talk about users as someone to which uncertainty results should be communicated, e.g. Pappenberger and Beven (2006).

The most difficult uncertainty problem (in natural science) to handle today is the model structure uncertainty, and the most important and novel contribution is probably the efforts made in this respect, primarily the new framework outlined in [15] but also the inclusion of options for evaluating multiple conceptual models in the HarmoniQuA modelling protocol ([13] and Fig. 5). The approach suggested in [15] of using multiple conceptual models (model structures) is not new (IPCC, 2001; Beven, 2002a; Neuman and Wierenga, 2003) and the use of pedigree analysis to qualitatively assess the credibility of something is not new either (van der Sluijs et al., 2005). The novelty lies in the combination of the two approaches that originate from different disciplines.

## 4.4 Quality Assurance in Model based Water Management

### 4.4.1 Background

During the last decade many problems have emerged in river basin modelling projects, including poor quality of modelling, unrealistic expectations, and lack of credibility of modelling results. Some of the reasons for this lack of quality can be evaluated ([13]; Scholten et al., 2007) as the effect of:

- Ambiguous terminology and a lack of understanding between key-players (modellers, clients, reviewers, stakeholders and concerned members of the public)
- Bad practice (careless handling of input data, inadequate model set-up, insufficient calibration/validation and model use outside of its scope)
- Lack of data or poor quality of available data
- Insufficient knowledge on the processes
- Poor communication between modellers and end-users on the possibilities and limitations of the modelling project and overselling of model capabilities
- Confusion on how to use model results in decision making
- Lack of documentation and clarity on the modelling process, leading to results that are difficult to audit or reproduce
- Insufficient consideration of economic, institutional and political issues and a lack of integrated modelling.

In the water resources management community many different guidelines on good modelling practice have been developed, see [13] for a review. One, if not the most, comprehensive example of a modelling guideline has been developed in The Netherlands (Van Waveren et al., 2000) as a result of a process involving all the main players in the Dutch water management field. The background for this was a perceived need to improve the quality of modelling (Scholten et al., 2000). Similarly, modelling guidelines for the Murray-Darling Basin in Australia were developed due to the perception among end-users that model capabilities may have been 'over-sold', and that there was a lack of consistency in approaches, communication and understanding among and between the modellers and the water managers, which often resulted in considerable uncertainty for decision making (Middlemis, 2000).

### 4.4.2 The HarmoniQuA approach

A software tool, MoST, with its associated knowledge base (KB), has been developed by the HarmoniQuA project ([13]; Scholten et al., 2007) to provide QA in modelling through guidance, monitoring and reporting. As defined in HarmoniQuA: *"Quality Assurance (QA) is the procedural and operational framework used by an organisation managing the modelling study to build consensus among the organisations concerned in its implementation, to assure technically and scientifically adequate execution of all tasks included in the study, and to assure that all modelling-based analysis is reproducible and justifiable".* This modification of the older NRC (1990) definition includes the organisational, technical and scientific aspects, but also the need to build consensus among the organisations concerned in accordance with the discussion in Section 2.1 above.

Guidelines for good modelling practise are included in the Knowledge Base (KB) of MoST. The modelling process has been decomposed into five steps, see the flowchart in Fig. 5. Each step includes several tasks. Each task has an internal structure i.e. name, definition, explanation, interrelations with other tasks, activities, activity related methods, references, sensitivity/pitfalls, task inputs and outputs.

The KB contains knowledge specific to seven domains (groundwater, precipitation-runoff, river hydrodynamics, flood forecasting, water quality, ecology and socio-economics), and forms the heart of the tool. A computer based journal is produced within MoST where the water manager and modelling team record the progress and decisions made during a model study according to the tasks in the flowchart. This record can be used when reviewing the model study to judge its quality.

The most important QA principles incorporated in the KB are:

- The five modelling steps conclude with a formal *dialogue* between the modeller and manager, where activities and results from the present step are reported, and details of plans for the next step (a revised work plan) are discussed.
- *External reviews* are prescribed as the key mechanism of ensuring that the knowledge and experience of other independent modellers are used.
- The KB provides *public interactive guidelines* to facilitate dialogue between modellers and the water manager, with options to include auditors (reviewers), stakeholders and the public.
- There are many *feed back loops*, some technical involving only the modeller, and others that may require a decision before doing costly additional work.
- The KB allows *performance and accuracy criteria* to be updated during the modelling process. In the first step the water manager's objectives and requirements are translated into performance criteria that may include qualitative and quantitative measures. These criteria may be modified during the formal reviews of subsequent steps.
- Emphasis is put on *validation schemes*, i.e. tests of model performance against data that have not been used for model calibration.
- *Uncertainties* must be explicitly recognised and assessed (qualitatively and/or quantitatively) throughout the modelling process.

MoST supports multi-domain studies and working in teams of different user types (water managers, modellers, auditors, stakeholders and members of the public). It contains an interactive glossary that is accessible via hyperlinked text. The key functionality of MoST is to:

- *Guide*, to ensure a model has been properly applied. This is based on the Knowledge Base.
- *Monitor*, to record decisions, methods and data used in the modelling work and in this way enable transparency and reproducibility of the modelling process.
- *Report*, to provide suitable reports of what has been done for managers/clients, modellers, auditors, stakeholders and the general public.

### 4.4.3 Organisational requirements for QA guidelines to be effective

Modelling studies involve several parties with different responsibilities. The key players are modellers and water managers, but often reviewers, stakeholders and the general public are also involved. To a large extent the quality of the modelling study is determined by the expertise, attitudes and motivation of the teams involved in the modelling and QA process.

QA will only be successful if all parties actively support its use. The attitude of the modellers is important. NRC (1990) characterises this as follows: “most modellers enjoy the modelling process but find less satisfaction in the process of documentation and quality assurance”. Scholten and Groot (2002) describe the main problem with the Dutch Handbook on Good Modelling Practice as “they all *like* it, but only a few *use* it”. The water manager, however, has a particular responsibility, because he/she has the power to request and pay for adequate QA in modelling studies. Therefore, QA guidelines can only be expected to be used in practice if the water manager prescribes their use. It is therefore very important that the water manager has the technical capacity to organise the QA process. Often, water managers do not have individuals available with the appropriate training to understand and use models. An external modelling expert should then be sought to help with the QA process. However, this requires that the manager is aware of the problem and the need.

### 4.4.4 Performance criteria and uncertainty – when is a model good enough?

A critical issue is how to define the performance criteria. We agree with Beven (2002b) that any conceptual model is known to be wrong and hence any model will be falsified if we investigate it in sufficient detail and specify very high performance criteria. Clearly, if one attempts to establish a model that should simulate the truth it would always be falsified. However, this is not very useful information. Therefore, we are using the conditional validation, or the validation restricted to domain of applicability (or numerical universal as opposed to strictly universal in Popperian terms). The good question is then what is good enough? Or in other words what are the criteria? How do we select them?

A good reference for model performance is to compare it with uncertainties of the available field observations. If the model performance is within this uncertainty range we often characterise the model as good enough. However, usually it is not so simple. How wide confidence bands do we accept on observational uncertainties – ranges corresponding to 65%, 95% or 99%? Do we always then reject a model if it cannot perform within the observational uncertainty range? In many cases even results from less accurate models may be useful.

Therefore, the decision on what is good enough generally must be taken in a socio-economic context. For instance, the accuracy requirements to a model to be used for an initial screening of alternative options for location of a new small well field for a small water supply will be much smaller than the requirements to a model that is intended to be used for the final design of a large well field for a major water supply in an area with potential damaging effects on precious nature and other significant conflicts of interests. Thus, the accuracy criteria can not be decided universally by modellers or researchers, but must be different from case to case depending on how much is at stake in the decision to depend on the support from model predictions. This implies that the performance criteria must be discussed and agreed between the manager and the modeller beforehand.

Accuracy requirements and uncertainty assessments of model simulations are two sides of the same coin, just seen from two different perspectives, namely the water manager and the modeller. As all uncertainty can not be characterised as statistical uncertainty (see Fig. 33 and Tables 2 and 3 in Subsection 4.3.1) it is also required to characterise accuracy requirements in qualitative terms. Furthermore the risk perception of the water manager and the stakeholders/public has to be considered. Therefore, involvement of stakeholders and public are most often required as an integrated part of this process (see also Section 2.1 and Figs. 1-2). According to the HarmoniQuA methodology stakeholder/public involvement is crucial at the beginning of a modelling project to frame the problem, define the requirements and assess the uncertainties (Henriksen et al., submitted).

This way of thinking is well in line with the principles behind some of the Water Framework Directive Guidance Documents. For example the Guidance Document on Monitoring (EC, 2003a) does not specify the levels of precision and confidence required from the monitoring programmes, but rather states that the precision and confidence level should be sufficient to enable a meaningful assessment of for instance the status of the environment and should be sufficient to achieve an acceptable risk of making the wrong decision. This obviously calls for uncertainty assessments and public participation to have a central role in the entire process, which pave the road towards making adaptive management an important part of the river basin management process (Pahl-Wostl, 2002).

### 4.4.5 Discussion – post evaluation

The ideas and concepts behind the HarmoniQuA guidelines ([12], [13]) summarised above have been inspired from previous QA guidelines. The novel contributions have been inspired both from previous research activities (including [4], [5], [6], [7], [9], [11]) and from participation in a large range of national and international consultancy projects. Without having been in this crossroad between the research world and the practical world for more than two decades this would not have been possible. I consider my most important contributions in this respect to be:

- The terminology and guiding principles behind the guidelines [12] are novel in their attempt to formulate a coherent approach that on the one hand has a solid scientifically philosophical foundation and on the other hand can be useful for practitioners. In the very controversial issue of model validation, where there has been almost a deadlock between different schools with respect to whether validation at all is possible, the philosophy of *conditional validation* is novel.
- The major novelty of the HarmoniQuA approach does not lie in its guidance on model technical issues, but on its emphasis and more elaborate focus on the dialogue between modeller, water manager, reviewer, stakeholders and the public. In addition, there are novel elements on the large emphasis on uncertainty assessments throughout the modelling process and model validation. Finally, the emphasis on model reviews allows bringing in subjective knowledge and experience in the QA process.

Both the HarmoniQuA guidelines and other recent good modelling practise guidelines have been deeply rooted both in the scientific community and among practitioners ([13]). As a comparison, ideas originating alone from the natural science community, such as the suggested Code of Practise on performing uncertainty analysis by Pappenberger and Beven (2006), are typically limited to valuable contributions on model technical issues, while they often do not consider the broader aspects of the modelling process such as the involvement of water managers and stakeholders.

# 5 Conclusions and Perspectives for Future Work

## 5.1 Summary of Main Scientific Contributions

The contributions to scientific knowledge in the papers of the present thesis are discussed in the previous chapters. The main contributions have been in the following five areas:

- *New conceptual understanding and code development*. The Suså model ([1], [2]) was based on a new conceptual understanding of the surface water/groundwater interaction in moraine catchment. The code and its application brought new insight regarding the effect of groundwater abstraction on streamflow in catchments with such hydrogeological characteristics.

- *Model validation.* The adoption and adaptation of rather rigorous principles for model validation and the examples of their application both for lumped conceptual and distributed physically based models is a cornerstone in my research. This work was first published in [6] and [7] and later brought into a broader modelling framework in [12] and [13]. In particular the introduction of the term ‘conditional validation’ in [7] and the outline of its scientific philosophical basis in [12] is novel.

- *Scaling.* The publications focussing on scaling ([7], [10]) presents ideas crystallised from work with scaling problems in many modelling studies ranging from point scale to thousands of km$^2$. The later framework, outlined in Section 4.1 above does not in any way ‘solve’ the scaling problem but contributes to clarifications on applicable methodologies with focus on their respective assumptions and limitations.

- *Uncertainty assessment.* During the past decade a considerable part of my research work has focussed on uncertainty aspects. I consider my main contributions in this respect to be the introduction of the broader uncertainty framework integrated into the modelling framework ([13], [14]) and the work with model structure uncertainty ([15]).

- *Modelling protocols and guidelines for quality assurance in the modelling process*. The modelling protocol in [7] and the later and more comprehensive one presented as part of the guidelines for quality assurance in the modelling framework in [13] are a formalisation of experience and practises that have gradually emerged over the years. The novel elements in [13] are the emphasis on (a) the interactive dialogue between modeller, water manager, reviewer, stakeholders and the public; (b) uncertainty assessments throughout the modelling process; (c) model validation; and (d) experience and subjective knowledge introduced through external model reviews.

These main contributions to scientific knowledge would, however, not have been possible without the experience and insight gained in modelling studies ranging from point scale ([3]) to large catchments ([4], [5], [8], [9], [11]).

## 5.2 Modelling Issues for Future Research

Hydrological modelling has developed significantly during the three decades I have worked in this field. I started with editing punch cards and could only run one simulation per day (overnight) using model codes that today are considered small and simple. Since then, comprehensive new knowledge has been build into model codes and into the methodologies used in the modelling process.

During the process of writing this thesis, where I had to review my older publications, it was interesting to note the gradual change in research focus. The first decade my research focused on development of new codes. During the second decade more general methodological problem areas such as scaling and model validation were addressed. Towards the end of the third decade the emphasis is now on the broader issues such as uncertainty assessment and quality assurance frameworks for the entire modelling process, and the interaction between the modelling and the water management processes. While this no doubt is affected by personal and career developments, it also reflects a general trend. We are no longer satisfied with being able to produce beautiful simulations with sophisticated new model codes; we also want to evaluate the credibility of such simulations and to apply them in real-world water management decisions.

Certainly I did not foresee this development three decades ago. On this background it is therefore not wise to make long range forecasts on what we can expect as the key issues for future modelling research. Hence, the following list should not pretend to cover all the most important research issues for modelling during the coming many years. It rather presents a list of issues which I, seen from the perspective dealt with in the present thesis, consider the presently most important and fundamental problems requiring more research during the coming years.

- *Improved representation of heterogeneity in reactive transport modelling*. There will always be a need to improve our conceptual understanding of hydrological processes. It appears that, whereas we have had some success with prediction of flows and hydraulic heads, the existing paradigms in hydrological modelling are not good enough to simulate concentrations of conservative and reactive contaminants. Flows and hydraulic heads are much less depending on heterogeneity than concentrations, and it will be necessary to include heterogeneity much more explicitly in the modelling than done until now. Examples of areas, where this is important, include simulation of transport and fate of contaminants in aquifers and simulation of the stream-aquifer interaction governed by processes in river valleys.

- *Utilisation of new data types*. Whenever possible we should try to make use of new data types. New techniques for collecting satellite data on surface conditions and geophysical data on subsurface features are promising and have not been fully exploited yet. We can hope and expect that better techniques will be developed during the coming years. Thus, it is not unrealistic in some years to have improved data providing both a much better spatial resolution of catchment/aquifer properties and on-line information on state variables. The improved spatial resolution can help us give a better representation of heterogeneities in models (see above), while on-line information provide interesting potentials for improved management. In order to utilise on-line data optimally new and improved data assimilation (updating) techniques will be required.

- *Model structure error*. Probably the most important single issue related to uncertainty of model predictions is how to assess uncertainty caused by model structure error. It is important, because the most interesting fields of model applications deal with assessments of the effects on the ecosystem of human activities. And it is at the same time fundamentally difficult, because we in such situations are using models beyond the situations, where we can test the model performance against field data. I consider the framework based on multiple conceptual models ([15]) only to be a very first beginning in this respect.

- *Uncertainty and credibility of modelling in relation to water resources management*. Uncertainty assessments of model predictions are crucial for a sound use of models in water resources management in practise. Model predictions without uncertainty assessments correspond to only presenting a (minor) part of the available information. Uncertainty in relation to water resources management in practise is not confined to statistical uncertainty. It is also required to include aspects of qualitative uncertainty and ignorance. Furthermore, uncertainty must be seen in a broad socio-economic context where stakeholder and policy views are taken into account. There are many future challenges on this multi-disciplinary road. How do we ensure that models incorporate the best available information and adequately address the issues and the priorities set by water managers and stakeholders? How should we translate objectives and requirements formulated in qualitative language by water managers and stakeholders to accuracy criteria for a modelling study? And how should we compile and present uncertainties from a modelling study in a way that is understandable by non-modellers? Some of these questions are likely to be answered within the context of new water management paradigms such as adaptive management.

# 6 References

Abbott MB (1992) The theory of the hydrological model, or: the struggle for the soul of hydrology. In: O'Kane JP (Ed.) Advances in theoretical hydrology, Elsevier, 237-254.

Abbott MB, Bathurst JC, Cunge JA, O'Connel PE, Rasmussen J (1986a) An introduction to the European Hydrological System - Systeme Hydrologique Européen "SHE", 1: History and philosophy of a physically-based distributed modelling system. Journal of Hydrology, 87, 45-59.

Abbott MB, Bathurst JC, Cunge JA, O'Connel PE, Rasmussen J (1986b) An introduction to the European Hydrological System - Systeme Hydrologique Européen "SHE", 2: Structure of a physically-based distributed modelling system. Journal of Hydrology, 87, 61-77.

Abrahamsen P, Hansen S (2000) Daisy: an open soil-crop-atmosphere system model. Environmental Modelling & Software, 15, 313-330.

Andersen J, Refsgaard JC, Jensen KH (2001) Distributed hydrological modelling of the Senegal River Basin – model construction and validation. Journal of Hydrology, 247, 200-214.

Anderson MP, Woessner WW (1992) The role of postaudit in model validation. Advances in Water Resources, 15, 167-173.

Babendreier JE (2003) National-scale multimedia risk assessment for hazardous waste disposal. International Workshop on Uncertainty, Sensitivity and Parameter Estimation for Multimedia Environmental Modelling held at U.S Nuclear Regulatory Commission, Rockville, Maryland, August 19-21, 2003. Proceedings, 103-109.

Bathurst JC (1986a) Physically-based distributed modelling of an upland catchment using the Systeme Hydrologique Européen. Journal of Hydrology, 87, 79-102.

Bathurst JC (1986b) Sensitivity analysis of the Systeme Hydrologique Européen for an upland catchment. Journal of Hydrology, 87, 103-123.

Beck MB (1987) Water quality modelling: a review of the analysis of uncertainty. Water Resources Research, 23(8), 1393-1442.

Beck MB (2005) Environmental foresight and structural change. Environmental Modelling & Software, 20, 651-670.

Bergström (1976) Development and application of a conceptual runoff model for Scandinavian catchments. PhD Thesis, University of Lund, Bulletin Series A No 52.

Bergström S (1992) The HBV model – its structure and applications. SMHI RH No 4. Norrköping.

Bergström S (1995) The HBV model. In: Singh VP (Ed) Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado, 443-476.

Bergström S, Forsman A (1973) Development of a conceptual deterministic rainfall-runoff model. Nordic Hydrology, 4, 147-170.

Beven K (1989) Changing ideas in hydrology – the case of physically based models. Journal of Hydrology, 105, 157-172.

Beven K (1995) Linking parameters across scales: Subgrid parameterization and scale dependent hydrological models. Hydrological Processes, 9, 507-525.

Beven K (1996a) A discussion of distributed hydrological modelling. In: Abbott MB, Refsgaard JC (Eds): Distributed Hydrological Modelling, Kluwer Academic Publishers, 255-278.

Beven K (1996b) Response to comments on 'A discussion of distributed hydrological modelling'. In: Abbott MB, Refsgaard JC (Eds): Distributed Hydrological Modelling, Kluwer Academic Publishers, 289-295.

Beven K (2001) How far can we go in distributed hydrological modelling? Hydrology and Earth System Sciences, 5(1), 1-12.

Beven K (2002a) Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. Hydrological Processes, 16(2), 189-206.

Beven K (2002b) Towards a coherent philosophy for modelling the environment. Proceedings of the Royal Society of London, A, 458 (2026), 2465-2484.

Beven K, Binley AM (1992) The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes, 6, 279-298.

Binley AM, Beven KJ, Calver A, Watts LG (1981) Changing Responses in Hydrology: Assessing the Uncertainty in Physically Based Model Predictions. Water Resources Research, 27(6), 1253-1261.

Birkinshaw SJ, Ewen J (2000) Nitrogen transformation component for SHETRAN catchment nitrate transport modelling. Journal of Hydrology, 230, 1-17.

Blöschl G, Sivapalan M (1995) Scale issues in hydrological modelling: A review. Hydrological Processes, 9, 251-290.

Brown JD (2004) Knowledge, uncertainty and physical geography: towards the development of methodologies for questioning belief. Transactions of the Institute of British Geographers 29(3), 367-381.

Brown JD, Heuvelink GBM, Refsgaard JC (2005) An integrated framework for assessing and recording uncertainties about environmental data. Water Science and Technology, 52(6), 153-160.

Brown JD, Heuvelink GBM (2005) Data Uncertainty Engine (DUE) User's Manual. University of Amsterdam. http://www.harmonirib.com.

Butts MB, Payne JT, Kristensen M, Madsen H (2004) An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow prediction. Journal of Hydrology, 298, 242-266.

Burnash RJC (1995) The NWS river forecast system - catchment modelling. In: Singh VP (Ed): Computer Models of Watershed Hydrology, Water Resources Publications, 311-366.

Christensen S (1994) Hydrological Model for the Tude Å Catchment. Nordic Hydrology, 25, 145-166.

Conan C, Bouraoui F, Turpin N, de Marsily G, Bidoglio G (2003) Modelling Flow and Nitrate Fate at Catchment Scale in Brittany (France). Journal of Environmental Quality, 32, 2026-2032.

Crawford NH, Linsley RK (1966) Digital simulation in hydrology, Stanford Watershed Model IV, Department of Civil Engineering, Stanford University, Technical Report 39.

Currie JA (1961) Gaseous diffusion in the aeration of aggregated soils. Soil Science, 92, 40-45.

Dagan G (1986) Statistical theory of groundwater flow and transport: pore to laboratory, laboratory to formation and formation to regional scale. Water Resources Research, 22(9), 120-134.

De Marsily, G Combes P, Goblet P (1992) Comments on 'Ground-water models cannot be validated', by Konikow LF, Bredehoeft, JD, Advances in Water Resources, 15, 367-369.

Dewulf A, Craps M, Bouwen R, Pahl-Wostl C (2005) Integrated management of natural resources dealing with ambiguous issues, multiple actors and diverging frames. Water Science and Technology, 52(6), 115-124.

DHI (1995) MIKE 21 Short Description. Danish Hydraulic Institute, Hørsholm, Denmark.

Djuurhus J, Hansen S, Schelde K, Jacobsen OH (1999) Modelling mean nitrate leaching from spatially variable fields using effective parameters. Geoderma, 87,261-279.

Doherty J (2003) Ground water model calibration using pilot points and regularization. Ground Water, 41(2), 170-177.

Duan Q, Sorooshian S, Gupta VK (1994) Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of Hydrology 158, 265–284.

Dubus, IG, Brown CD, Beulke S (2003) Sources of uncertainty in pesticide fate modelling. The Science of the Total Environment, 317, 53-72.

EC (1992) Working Group of Independent Experts on Variant C of the Gabcikovo-Nagymaros Project, Working Group Report, Commission of the European Communities, Czech and Slovak Federative Republic, Republic of Hungary, Budapest November 23, 1992.

EC (1993a) Working Group of Monitoring and Water Management Experts for the Gabcikovo System of Locks - Data Report, Commission of the European Communities, Republic of Hungary, Slovak Republic, Budapest November 2, 1993.

EC (1993b) Working Group of Monitoring and Water Management Experts for the Gabcikovo System of Locks - Report on Temporary Water Management Regime, Commission of the European Communities, Republic of Hungary, Slovak Republic, Bratislava, December 1, 1993.

EC (2003a) Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance Document No. 7. Monitoring under the Water Framework Directive. Working Group 2.7. Office for the Official Publications of the European Communities, Luxembourg.

EC (2003b) Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance Document No. 11. Planning Processes. Working Group 2.9. Office for the Official Publications of the European Communities, Luxembourg.

EC (2004) Common Implementation Strategy for the Water Framework Directive (2000/60/EC) Guidance Document No 3, pressures and impacts, IMPRESS. Working Group 2.3. Office for the Official Publications of the European Communities, Luxembourg.

Fleming G (1975) Computer simulation techniques in hydrology. Elsevier, New York.

Franchini M, Pacciani M (1992) Comparative analysis of several conceptual rainfall-runoff models. Journal of Hydrology, 122, 161-219.

Freeze RA, Harlan RL (1969) Blueprint for a physically-based digitally-simulated hydrologic response model. Journal of Hydrology, 9, 237-258.

Gelhar LW (1986) Stochastic subsurface hydrology. From theory to application. Water Resources Research, 22(9), 135-145.

Graham DN, Butts MB (2005) Flexible integrated watershed modelling with MIKE SHE. In: Singh VP, Frevert DK (Eds) Watershed Models. CRC Press, Chapter 10.

Graham LP (1999) Modelling runoff to the Baltic Sea, Ambio, 28, 328-334.

Grayson RB, Moore ID, McHahon TA (1992a) Physically based hydrologic modelling, 1. A terrain-based model for investigative purposes. Water Resources Research, 28(10), 2639-2658.

Grayson RB, Moore ID, McHahon TA (1992b) Physically based hydrologic modelling, 2. Is the concept realistic ? Water Resources Research, 28(10), 2639-2658.

Grayson R, Blöschl G (2000) Spatial Modelling of Catchment Dynamics. In: Grayson R, Blöschl G (Eds.) Spatial Patterns in Catchment Hydrology: Observations and Modelling. Cambridge University Press, UK.

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

GWP (2000) Integrated Water Resources Management. TAC Background Papers No. 4. Global Water Partnership, Stockholm.

Hansen S, Jensen HE, Nielsen NE, Svendsen H (1991) Simulation of nitrogen dynamics and biomass production in winter wheat using the Danish simulation model DAISY. Fertilizer Research, 27, 245-259.

Hansen S, Thorsen M, Pebesma E, Kleeschulte S, Svendsen H (1999) Uncertainty in simulated leaching due to uncertainty in input data. A case study. Soil Use and Management, 15, 167-175.

Harrar WG, Sonnenborg TO, Henriksen HJ (2003) Capture zone, travel time and solute transport predictions using inverse modelling and different geological models. Hydrogeology Journal, 11(5), 536-548.

Havnø K, Madsen MN, Dørge J (1995) MIKE 11 - A Generalized River Modelling Package. In: Singh VP (Ed) Computer Models of Watershed Hydrology, Water Resources Publications, Highlands Ranch, Colorado, 733-782.

Henriksen HJ, Refsgaard JC, Sonnenborg TO, Gravesen P, Brun A, Refsgaard A, Jensen KH (2001) STÅBI i grundvandsmodellering (Handbook in groundwater modelling). Danmarks og Grønlands Geologiske Undersøgelse, Rapport 2001/56. (In Danish)

Henriksen HJ, Troldborg L, Nyegaard P, Sonnenborg TO, Refsgaard JC, Madsen B (2003) Methodology for construction, calibration and validation of a national hydrological model for Denmark. Journal of Hydrology 280, 52-71.

Henriksen HJ, Refsgaard JC, Højberg AL, Ferrand N, Gijsbers P, Scholten H (submitted) Public participation in relation to quality assurance of water resources modelling (HarmoniQuA).

Heuvelink GBM, Pebesma EJ (1999) Spatial aggregation and soil process modelling. Geoderma, 89, 47-65.

Hill MC (1998) Methods and guidelines for effective model calibration. U.S. Geological Survey, Water-Resources Investigations Report 98-4005. Denver CO.

Højberg AL, Refsgaard JC (2005) Model Uncertainty - Parameter uncertainty versus conceptual models. Water Science and Technology, 52(6), 177-186.

ICJ (1997) Case Concerning Gabcikovo-Nagymaros project (Hungary/Slovakia). Summary of the Judgement of 25 September 1997. International Court of Justice, The Hague.

ICWE (1992) The Dublin Statement and report of the conference. International Conference on Water and the Environment: Development issues for the 21st century. 26-31 January 1992, Dublin, Ireland.

IPCC (2001) Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel of Climate Change [Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K and Johnson CA (eds)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, 881 pp.

Jensen KH, Mantoglou A (1992) Application of stochastic unsaturated flow theory, numerical simulations, and comparisons to field observations. Water Resources Research, 28, 269-284.

Jensen RA, Jørgensen GH (1988) Hydrologisk overfladevands/grundvands model (Hydrological surface water/groundwater model). Technical report prepared by Danish Hydraulic Institute for the County of Storstrøm and the County of Vestsjælland. (in Danish)

Jensen KH, Refsgaard JC (1991a) Spatial variability of physical parameters and processes in two field soils. Part I: Water Flow and Solute Transport at Local Scale. Nordic Hydrology, 22, 275-302.

Jensen KH, Refsgaard JC (1991b) Spatial variability of physical parameters and processes in two field soils. Part II: Water flow at field scale. Nordic Hydrology, 22, 303-326.

Jensen KH, Refsgaard JC (1991c) Spatial variability of physical parameters and processes in two field soils. Part III: Solute Transport at Field Scale. Nordic Hydrology, 22, 327-340.

Jønch-Clausen T (1979) SHE. Système Hydrologiique Européen. A short description. Danish Hydraulic Institute, Hørsholm, Denmark.

Jønch-Clausen T (2004) Integrated Water Resources Management (IWRM) and Water Efficiency Plans by 2005. Why, What and How? Global Water Partnership, TEC Background Papers No. 10, Stockholm.

Jønch-Clausen T, Refsgaard JC (1984) A Mathematical Modelling System for Flood Forecasting. Nordic Hydrology, 15, 307-318.

Kaiser-Hill (2001) Model Code and Scenario Selection Report Site-Wide Water Balance Rocky Flats Environmental Technology Site. Report 01-RF-00337. Kaiser-Hill Company LLC.

Klauer B, Brown JD (2003) Conceptualising imperfect knowledge in public decision making: ignorance, uncertainty, error and 'risk situations'. Environmental Research, Engineering and Management.

Klemes V (1986) Operational testing of hydrological simulation models. Hydrological Sciences Journal, 31, 13-24.

Knudsen J, Thomsen A, Refsgaard JC (1986) WATBAL: A semi-distributed, physically based hydrological modelling system. Nordic Hydrology, 17, 347-362.

Konikow LF, Bredehoeft JD (1992) Ground-water models cannot be validated. Advances in Water Resources, 15, 75-83.

Kros J, Reinds GJ, de Vries W, Latour JB, Bollen M (1995) Modelling of soil acidity and nitrogen availability in natural ecosystems in response to changes in acid deposition and hydrology. Report 95, DLO Winand Staring Centre, Wageningen.

Kutchment LS, Demidov VN, Naden PS, Cooper DM, Broadhurst P (1996) Rainfall-runoff modelling of the Ouse basin, North Yorkshire: an application of a physically based distributed model. Journal of Hydrology, 181, 323-342.

Lane SA, Richards KS (2001) The 'Validation' of Hydrodynamic Models: Some Critical Perspectives. In: Anderson MG, Bates PD (Eds) Model Validation perspectives in Hydrological Science, 413-438. John Wiley & Sons, Ltd.

Linkov I, Burmistrov D (2003) Model Uncertainty and Choices Made by Modelers: Lessons Learned from the International Atomic Energy Model Intercomparisons. Risk Analysis, 23(6), 1297-1308.

Lloyd JW (1980) The importance of drift deposit influences on the hydrogeology of major British aquifers. Institution of Water Engineers and Scientists, Journal, 34, 346-356.

Loague KM, Freeze RA (1985) A Comparison of Rainfall-Runoff Modelling Techniques on Small Upland Catchments. Water Resources Research, 21(2), 1985.

Luckner L (1978) Gekoppelte Grundwasser-Oberflächenwassermodelle (A coupled groundwater-surface water model). Wasserwirtschaft-Wassertechnik, 1978, 276-278 (In German).

Madsen H, Skotner C (2005) Adaptive state updating in real-time river flow forecasting – a combined filtering and error forecasting procedure. Journal of Hydrology, 308, 302-312.

Michaud J, Sorooshian S (1994) Comparison of simple versus complex distributed runoff models on a mid-sized semiarid watershed. Water Resources Research, 30(3), 593-605.

Michaud JD, Shuttelworth WJ (1997) Executive summary of the Tuczon aggregation workshop. Journal of Hydrology, 190, 176-181.

Middlemis H (2000) Murray-Darling Basin Commission. Groundwater flow modelling guideline. Aquaterra Consulting Pty Ltd., South Perth. Western Australia. Project no. 125.

Miles JC, Rushton KR (1983) A coupled surface water and groundwater catchment model. Journal of Hydrology, 62, 159-177.

Neuman SP, Wierenga PJ (2003) A comprehensive strategy of hydrogeologic modeling and uncertainty analysis for nuclear facilities and sites. University of Arizona, Report NUREG/CR-6805.

Nielsen DR, Bigger JW, Erk KT (1973) Spatial variability of field measured soil water properties. Hilgardia, 42, 215-259.

Nielsen SA, Hansen E (1973) Numerical simulation of the rainfall-runoff process on a daily basis. Nordic Hydrology, 4, 171-190.

NRC (1990) Ground Water Models: Scientific and Regulatory Applications. National Research Council, National Academy Press, Washington, D.C.

Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation and confirmation of numerical models in the earth sciences. Science, 264, 641-646.

Pahl-Wostl C (2002) Towards sustainability in the water sector – The importance of human actors and processes of social learning. Aquatic Sciences, 64, 394-411.

Panday S, Hayakorn PS (2004) A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. Advances in Water Resources, 27, 361-382.

Pappenberger F, Beven KJ (2006) Ignorance in bliss: Or seven reasons not to use uncertainty analysis. Water Resources Research 42, W05302, doi:10.1029/2005WR004820.

Pascual P, Steiber N, Sunderland E (2003) Draft guidance on development, evaluation and application of regulatory environmental models. The Council for Regulatory Environmental Modeling. Officie of Science Policy, Office of Research and Development. US Environmental Protection Agency, Washington D.C. 60 pp.

Perkins SP, Sophocleous M (1999) Development of a Comprehensive Watershed Model Applied to Study Stream Yield under Drought Conditions. Ground Water, 37(3), 418-426.

Perrin C, Michel C, Andréassian V (2001) Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. Journal of Hydrology, 242, 275-301.

Poeter E, Anderson D (2005) Multiple Ranking and Inference in Ground Water Modeling. Ground Water, 43(4), 597-605.

Popper KR (1959) The logic of scientific discovery. Hutchingson & Co, London.

Prickett TA, Lonnquist CG (1971) Selected digital computer techniques for groundwater resource evaluation. Illinois State Water Survey, Bulletin 55.

Querner EP (1997) Description and application of the combined surface and groundwater flow model MOGROW. Journal of Hydrology, 192, 158-188.

Quinn PF, Beven KJ (1993) Spatial and temporal predictions of soil moisture dynamics, runoff, variable source areas and evapotranspiration for Plynlimon, Mid-Wales. Hydrological Processes, 7, 425-448.

Radwan M, Willems P, Berlamont J (2004) Sensivity and uncertainty analysis for river quality modelling. Journal of Hydroinformatics, 6, 83-99.

Reed S, Koren V, Smith M, Zhang Z, Moreda F, Seo D-J (2004) Overall distributed model intercomparison project results. Journal of Hydrology, 298, 27-60.

Refsgaard JC (1981) The surface water component of an integrated hydrological model. Danish Committee for Hydrology. Suså Report No. H12.

Refsgaard JC (1996) Terminology, modelling protocol and classification of hydrological model codes. In: Abbott MB, Refsgaard JC (Eds): Distributed Hydrological Modelling, Kluwer Academic Publishers, 17-39.

Refsgaard JC, Stang O (1981) An integrated groundwater/surface water hydrological model. Danish Committee for Hydrology. Suså Report No. H13.

Refsgaard JC, Rosbjerg D, Markussen LM (1983) Application of Kalman filter to real-time operation and to uncertainty analyses in hydrological modelling. IAHS Publication No 147, 273-282.

Refsgaard JC, Storm B (1995) MIKE SHE. In: Singh VP (Ed) Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado, 809-846.

Refsgaard JC, Storm B, Abbott MB (1996) Comments on 'A discussion of distributed hydrological modelling'. In: Abbott MB, Refsgaard JC (Eds): Distributed Hydrological Modelling, Kluwer Academic Publishers, 279-287.

Refsgaard JC, Ramaekers D, Heuvelink GBM, Schreurs V, Kros H, Rosén L, Hansen S (1998) Assessment of 'cumulative' uncertainty in spatial decision support systems: Application to examine the contamination of groundwater from diffuse sources (UNCERSDSS). Presented at the European Climate Science Conference, Vienna, 19-23 October 1998.

Refsgaard JC, Butts MB (1999) Determination of grid scale parameters in catchment modelling by upscaling local scale parameters. Key note presentation. Proceedings of the EurAgEng International Workshop on Modelling of transport processes in soils at various scales in time and space, 24-26 November 1999, Leuven, Belgium.

Refsgaard JC, van der Sluijs JP, Højberg AL, Vanrolleghem P (2005) Harmoni-CA Guidance Uncertainty Analysis. Guidance 1. 46 pp. www.harmoni-ca.info.

Rykiel ER (1996) Testing ecological models: the meaning of validation. Ecological Modelling, 90, 229-244.

Saulnier GM, Beven K, Obled C (1997) Digital elevation analysis for distributed hydrological modelling: Reducing scale dependence in effective hydraulic conductivity values. Water Resources Research, 33(9), 2097-2101.

Scholten H, Van Waveren RH, Groot S, Van Geer FC, Wösten JHM, Koeze RD, Noort JJ (2000) Good Modelling Practice in water management. Paper presented on Hydroinformatics 2000, Cedar Rapids, IA, USA.

Scholten H, Groot S (2002) Dutch guidelines. In: Refsgaard, JC (Ed) State-of-the-Art Report on Quality Assurance in modelling related to river basin management. Chapter 12, Geological Survey of Denmark and Greenland, Copenhagen. www.harmoniqua.org.

Scholten H, Kassahun A, Refsgaard JC, Kargas T, Gavardinas C, Beulens AJM (2007) A methodology to support multidisciplinary model-based water management. Environmental Modelling & Software, 22, 743-759.

Singh VP (Ed) (1995) Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado.

Smith KA (1980) A model of the extent of anaerobic zones in aggregated soils and its potential application to estimates of denitrification. Journal of Soil Science, 31, 263-277.

Sonnenborg TO, Christensen BSB, Nyegaard P, Henriksen HJ, Refsgaard JC (2003) Transient modelling of regional groundwater flow using parameter estimates from steady-state automatic calibration. Journal of Hydrology, 273, 188-204.

Stang O (1981) A regional groundwater model for the Suså area. Danish Committee for Hydrology. Suså Report No. H9.

Styczen M, Storm B (1993a) Modelling of N-movements on catchment scale – a tool for analysis and decision-making. 1. Model description. Fertilizer Research, 36, 1-6.

Styczen M, Storm B (1993b) Modelling of N-movements on catchment scale – a tool for analysis and decision-making. 2. A case study. Fertilizer Research, 36, 7-17.

Tampa Bay Water (2001) Scientific review of integrated hydrologic model ISGW/CNTB121. Prepared by West Consultants, Gartner Lee Ltd and AQUA TERRA Consultants for Tampa Bay Water, Florida.

Thomas RG (1973) Groundwater models. FAO, Irrigation and Drainage Paper 21, Rome.

Troch PA, Mancini M, Paniconni C, Wood EF (1993) Evaluation of a Distributed Catchment Scale Water Balance Model. Water Resources Research, 29(6), 1805-1817.

Troeh FR, Jabro JD, Kirkham D (1982) Gaseous diffusion equations for porous materials. Geoderma, 27, 239-253.

Troldborg L (2004) The influence of conceptual geological models on the simulation of flow and transport in Quaternary aquifer systems. PhD Thesis. Geological Survey of Denmark and Greenland, Report 2004/107.

Van Asselt MBA, Rotmans J (2002) Uncertainty in Integrated Assessment Modelling. From Positivism to Pluralism. Climatic Change, 54: 75-105.

Van der Sluijs JP, Craye M, Funtowicz SO, Kloprogge P, Ravetz J, Risbey JS (2005) Combining Quantitative and Qualitative Measures of Uncertainty in Model based Foresight Studies: the NUSAP System. Risk Analysis, 25(2), 481-492.

Van Griensven A, Meixner T (2004) Dealing with unidentifiable sources of uncertainty within environmental models. In: Pahl C, Schmidt S, Jakeman T. (Eds.), iEMSs 2004 International Congress: "Complexity and Integrated Resources Management". International Environmental Modelling and Software Society, Osnabrück, Germany, June 2004.

Van Loon E, Refsgaard JC (eds.) (2005) Guidelines for assessing data uncertainty in hydrological studies. HarmoniRiB Report. Geological Survey of Denmark and Greenland. http://www.harmonirib.com.

Van Waveren RH, Groot S, Scholten H, Van Geer FC, Wösten JHM, Koeze RD, Noort JJ (2000) Good Modelling Practice Handbook, STOWA Report 99-05, Utrecht, RWS-RIZA, Lelystad, The Netherlands, http://waterland.net/riza/aquest/

Vrugt J, Diks CGH, Gupta HV (2005) Improved treatment of uncertainty in hydrologic modelling: Combining the strengths of global optimization and data assimilation. Water Resources Research, 41, W01017, doi:10.1029/2004WR003059.

Walker WE, Harremoës P, Rotmans J, Van der Sluijs JP, Van Asselt MBA, Janssen P, Krayer von Krauss MP (2003) Defining Uncertainty A Conceptual Basis for Uncertainty Management in Model-Based Decision Support, Integrated Assessment, 4(1), 5-17.

Wardlaw RB (1978) The development of a deterministic integrated surface/subsurface hydrological response model. PhD Thesis, University of Stratchclyde, Glasgow.

Wardlaw RB, Wyness A, Rippon P (1994) Integrated catchment modelling. Surveys in Geophysics, 15, 311-330.

Weeks JB (1974) Simulated effects of oil-shale development on the hydrology of the Piceance basin, Colorado. US Geological Survey, Professional Paper 908.

Wen X-H, Gómez-Hernández JJ (1996) Upscaling hydraulic conductivities in heterogeneous media: An overview. Journal of Hydrology, 183, ix-xxxii.

WMO (1975) Intercomparison of conceptual models used in operational hydrological forecasting. WMO Operational Hydrology Report No 7, WMO No 429, World Meteorological Organisation, Geneva.

WMO (1988) Intercomparison of models for snowmelt runoff. WMO Operational Hydrology Report No 23, WMO No 646, World Meteorological Organisation, Geneva.

WMO (1992) Simulated real-time intercomparison of hydrological models. WMO Operational Hydrology Report No 38, WMO No 779, World Meteorological Organisation, Geneva.

Wolf J, Beusen AHW, Groenendijk P, Kroon T, Rötter R, van Zeijts H (2003) The integrated modelling system STONE for calculating nutrient emissions from agriculture in the Netherlands. Environmental Modelling & Software, 18, 597-617.

Wood EF, Sivapalan M, Beven KJ, Band L (1988) Effects of spatial variability and scale with implications to hydrologic modelling. Journal of Hydrology, 102, 29-47.

WSSTP (2005) Water safe strong and sustainable. A European vision for water supply and sanitation in 2030. Water Supply and Sanitation Technology Platform. October 2005. http://www.wsstp.org

WWAP (2003) Water for People, Water for Life. UN World Water Development Report. Prepared as a collaborative effort of 23 UN agencies and convention secretariats co-ordinated by the World Water Assessment Programme. UNESCO, Paris. http://www.unesco.org/water/wwap/index.shtml