



Uppsala University
Department of Earth Sciences, Hydrology

Jan Seibert

Conceptual runoff models - fiction or representation of reality?

Seibert, J., 1999. Conceptual runoff models - fiction or representation of reality? Acta Univ. Ups., *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 436. 52 pp. Uppsala. ISBN 91-554-4402-4.

Dissertation for the Degree of Doctor of Philosophy in Hydrology presented at
Uppsala University in 1999

ABSTRACT

Seibert, J., 1999. Conceptual runoff models - fiction or representation of reality? Acta Univ. Ups., *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 436. 52 pp. Uppsala. ISBN 91-554-4402-4.

Available observations are often not sufficient as a basis for decision making in water management. Conceptual runoff models are frequently used as tools for a wide range of tasks to compensate the lack of measurements, *e.g.*, to extend runoff series, compute design floods and predict the leakage of nutrients or the effects of a climatic change. Conceptual runoff models are practical tools, especially if the reliability in their predictions can be assessed. Testing of these models is usually based solely on comparison of simulated and observed runoff, although most models also simulate other fluxes and states. Such tests do not allow thorough assessment of model-prediction reliability. In this thesis, two widespread conceptual models, the HBV model and TOPMODEL, were tested using a catalogue of methods for model validation (defined as estimation of confidence in model simulations). The worth of multi-criteria validation for evaluating model consistency was emphasised. Both models were capable to simulate runoff adequately after calibration, whereas the performance for some of the other validation tests was less satisfactory. The impossibility to identify unique parameter values caused large uncertainties in model predictions for the HBV model. The parameter uncertainty was reduced when groundwater levels were included into the calibration, whereas groundwater-level simulations were in weak agreement with observations when the model was calibrated against only runoff. The agreement of TOPMODEL simulations with spatially distributed data was weak for both groundwater levels and the distribution of saturated areas. Furthermore, validation against hydrological common sense revealed weaknesses in the TOPMODEL approach. In summary these results indicated limitations of conceptual runoff models and highlighted the need for powerful validation methods. The use of such methods enables assessment of the reliability of model predictions. It also supports the further development of models by identification of weak parts and evaluation of improvements.

Keywords: Validation, conceptual runoff models, calibration, uncertainty, TOPMODEL, HBV model, groundwater levels, spatial distribution, topography.

*Jan Seibert, Department of Earth Sciences, Hydrology, Uppsala University,
Villavägen 16, SE-752 36 Uppsala, Sweden*

© Jan Seibert 1999

ISSN 1104-232X
ISBN 91-554-4402-4

Printed in Sweden by Geo-Tryckeriet, Uppsala 1999

Begreppsmässiga avrinningsmodeller – dikt eller verklighet?

REFERAT

Seibert, J., 1999. Begreppsmässiga avrinningsmodeller – dikt eller verklighet? Acta Univ. Ups., *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 436. 52 pp. Uppsala. ISBN 91-554-4402-4.

Tillgängliga mätdata är ofta otillräckliga som beslutsunderlag i vattenresursfrågor. Begreppsmässiga avrinningsmodeller används i många sammanhang för att kompensera bristen på mätdata. Exempel på användningsområden är förlängning av avrinningsserier, beräkning av dimensionerande flöden eller prognoser av läckage av näringsämnen och effekter av en klimatförändring. Begreppsmässiga avrinningsmodeller är användbara verktyg, särskilt om tillförlitligheten av deras prognoser kan bedömas. Modellerna testas vanligtvis bara genom att jämföra simulerad och uppmätt avrinning fastän de även simulerar andra variabler. Tillförlitligheten av modellprognoserna kan inte kontrolleras ingående med denna enkla jämförelse. I denna avhandling undersöktes två avrinningsmodeller, HBV-modellen och TOPMODEL, utgående från en sammanställning av metoder för modellvalidering (definierad som skattning av tillförlitlighet i modellsimuleringar). Betydelsen av validering med hjälp av flera kriterier betonades särskilt. Båda modellerna simulerade den observerade avrinningen väl sedan de blivit kalibrerade. Resultaten av andra valideringstester var dock mindre tillfredsställande. Det var inte möjligt att bestämma entydiga parametervärden för HBV-modellen och därför var modellprognoserna behäftade med avsevärda osäkerheter. Parametersäkerheten kunde minskas genom tillägg av observerade grundvattennivåer i kalibreringen. När modellen endast kalibrerades mot avrinningen stämde de simulerade grundvattennivåerna dåligt överens med mätningarna. Överensstämmelsen av TOPMODELS simuleringar med rumsligt fördelade data var dålig för både grundvattennivåer och fördelningen av vattenmättade områden. Svagheter hos ansatsen i TOPMODEL påvisades genom att relatera delar av modellen till kända hydrologiska samband. Sammanfattningsvis visade resultaten på begränsningar av begreppsmässiga avrinningsmodeller och underströk vikten av kraftfulla valideringsmetoder. Tillämpningen av metoderna möjliggör uppskattning av tillförlitligheten hos modellprognoser. Vidare kan metoderna vara till hjälp vid modellutvecklingen genom att påvisa svaga sidor och vid utvärdering av förbättringar.

Till Petra och våra Roliga barn

Konzeptionelle Abflussmodelle – Dichtung oder Wahrheit?

ZUSAMMENFASSUNG

Seibert, J., 1999. Konzeptionelle Abflussmodelle – Dichtung oder Wahrheit? Acta Univ. Ups., *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 436. 52 pp. Uppsala. ISBN 91-554-4402-4.

Vorhandene Messdaten sind häufig keine ausreichende Grundlage für Entscheidungen in wasserwirtschaftlichen Fragen. In vielen Bereichen werden konzeptionelle Abflussmodelle als Hilfsmittel verwendet, um eine bessere Entscheidungsgrundlage zu schaffen. Beispiele hierfür sind die Verlängerung von Abflussreihen, die Berechnung von Bemessungshochwässern oder die Vorhersage von Nährstoffausträgen oder von Konsequenzen einer Klimaänderung.

Konzeptionelle Abflussmodelle sind geeignete Werkzeuge, besonders wenn die Zuverlässigkeit ihrer Vorhersagen abgeschätzt werden kann. Diese Modelle werden normalerweise nur über den Vergleich zwischen dem gemessenen und simulierten Abfluss überprüft, obwohl die meisten Modelle auch andere Variablen simulieren. Mit solchen Tests kann die Zuverlässigkeit der Modellvorhersagen nicht ausreichend kontrolliert werden. In der vorliegenden Arbeit wurden zwei weitverbreitete konzeptionelle Abflussmodelle, das HBV-Modell und TOPMODEL, untersucht. Hierbei wurde eine Zusammenstellung von Methoden zur Modellvalidierung (definiert als Abschätzung der Zuverlässigkeit von Modellsimulationen) verwendet. Die Bedeutung der Modellüberprüfung mit Hilfe verschiedener Kriterien wurde betont. Beide Modelle konnten den gemessenen Abfluss gut simulieren, nachdem sie kalibriert worden waren. Die Ergebnisse in anderen Tests waren jedoch weniger zufriedenstellend. Für das HBV-Modell war es nicht möglich eindeutige Parameterwerte zu bestimmen und die Modellvorhersagen waren daher mit erheblichen Unsicherheiten behaftet. Die Parameterunsicherheit konnte dadurch verringert werden, dass Grundwasserstandsdaten in die Kalibrierung einbezogen wurden. Als das Modell jedoch nur mit Abflussdaten kalibriert wurde, stimmten die simulierten Grundwasserstände schlecht mit den gemessenen überein. Die Übereinstimmung der TOPMODEL-Simulationen mit räumlich verteilten Daten war sowohl für Grundwasserstände als auch für die räumliche Verteilung von Sättigungsflächen gering. Der Ansatz von TOPMODEL wurde im Hinblick auf hydrologische Allgemeinkenntnisse diskutiert und Unzulänglichkeiten konnten aufgezeigt werden. Zusammenfassend gesehen deuteten diese Resultate auf Beschränkungen konzeptioneller Abflussmodelle hin und zeigten die Wichtigkeit von geeigneten Validierungsmethoden. Diese Methoden ermöglichen es die Zuverlässigkeit von Modellvorhersagen einzuschätzen. Zusätzlich unterstützen sie die Weiterentwicklung von Modellen, indem sie Schwächen aufzeigen und Verbesserungen bewerten.

Table of contents

Preface.....	7
Acknowledgements.....	8
Introduction.....	10
Background.....	11
Problems to be solved with conceptual runoff models.....	11
The term validation.....	14
Material and methods.....	17
Validation of runoff models.....	17
Description of models.....	23
Study catchments.....	24
Illustration of validation methods.....	26
Proxy-basin test.....	26
Identifiability of parameter values (paper I).....	26
Validation of simulated groundwater levels (paper II).....	28
Partial model validation: the TOPMODEL index (paper III).....	28
Multi-criteria validation of TOPMODEL (paper IV).....	30
Validation of TOPMODEL against hydrological common sense (paper V).....	32
Multi-criteria calibration to runoff and groundwater levels (paper VI).....	33
Validation based on regionalisation.....	35
Discussion.....	36
Parsimony and complexity.....	36
Limitations of conceptual models.....	37
Problems of validation.....	38
Physically-based models.....	39
Future directions.....	40
Failings in modelling studies.....	42
Conclusions.....	45
References.....	45
Appendix: Terminology.....	52

Preface

This thesis is based on the following articles, which are referred to in the text by their respective Roman numerals:

- I. Seibert, J., 1997. Estimation of parameter uncertainty in the HBV model. *Nordic Hydrology* 28: 247-262
- II. Seibert, J., Bishop, K., and Nyberg, L., 1997. A test of TOPMODEL's ability to predict spatially distributed groundwater levels. *Hydrological Processes* 11: 1131-1144
- III. Rodhe, A. and Seibert, J., 1999. Wetland occurrence in relation to topography - a test of topographic indices as moisture indicators. *Agricultural and Forest Meteorology* (accepted for publication)
- IV. Güntner, A., Uhlenbrook, S., Seibert, J. and Leibundgut, Ch., 1999. Multi-criterial validation of TOPMODEL in a mountainous catchment. *Hydrological Processes* 13 (in press)
- V. Seibert, J., 1999. On TOPMODEL's ability to simulate groundwater level dynamics. In *Regionalization in Hydrology (Proc. Conf. at Braunschweig, March 1997)* (ed. by B. Dickkrüger, M.J. Kirkby and U. Schröder), *IAHS Publication* 254 (in press)
- VI. Seibert, J., 1999. Multi-criteria calibration of a conceptual rainfall-runoff model using a genetic algorithm. Submitted to *Hydrology and Earth System Sciences*

Nordic Hydrology (paper I), John Wiley & Sons (papers II and IV), Elsevier Science (III) and IAHS Press (paper V) kindly gave permission to reprint the articles in their entirety as well as individual parts.

In papers II and III I was responsible for computations and analyses as well as part of the writing. In paper IV I was involved in writing and in supervision of the first author's 'Diplomarbeit' (approximately corresponding to a MSc thesis) on which the publication is based.

Acknowledgements

In the summer of 1988, when I got off the train with my girlfriend Petra in Gävle in anticipation of our first biking tour through Sweden, I hardly expected that, eleven years later, I would live with a growing family here in Sweden, and have a PhD thesis ready to send to the printer.

The vague goal to live and study in Sweden became concrete when I, after another canoeing-cycling-hiking trip, came to Uppsala in 1991. Without having made any appointment in advance, I found an open door at ‘studierektor’ Lars-Christer Lundin’s office and had an interesting conversation, which persuaded me to ‘go for Uppsala’. He later turned out to be friendly L-C, always striving for smooth solutions. I certainly benefited from his confidence in me to let me take the responsibility of teaching. Before leaving Lars-Christer Lundin, he supplied me with kilos of thesis’s, reports and articles, which he pulled out from numerous corners and chambers in the cosy domicile, where the division of hydrology used to be in those “good old days”. Petra almost fainted when she saw me adding those kilos to our luggage, but already at this time she supported and shared my crazy hydroscientific behaviour. The material I got from L-C helped me put together a proposal which convinced the DAAD (Deutscher Akademischer Austauschdienst) to finance a one-year scholarship to Uppsala (Thank you, DAAD!).

Among the material I got in Uppsala was the PhD thesis of Allan Rodhe, who later became one of my supervisors. His thesis was one of the reasons for me coming to Uppsala; he as teacher, scientist and person was a main reason to stay (and to enjoy the time) as a PhD student. His contribution to this thesis actually started already long before I came to Sweden, when he installed groundwater tubes during his community service (paper VI), and his help and comments have been invaluable during the last couples of years. Due to Allan’s initiative I got involved early in the indoor hydrology at Gårdsjön and sniffed at ‘real science’ (well, actually the covered catchment at Gårdsjön looks more like a playground for grown-ups, but isn’t this what science should be like after all?). Within the Gårdsjön project I also had the pleasure of meeting Lars Nyberg and Kevin Bishop. I enjoy(ed) both, their stimulating scientific influence, as well as the delight of their company. Kevin’s interest was always encouraging and I am looking forward to joint projects in future.

When my time sponsored by the DAAD in Uppsala ended, Prof. Sven Halldin gave me the possibility to extend my visit by another five years and became my supervisor. I am thankful for his support, which, among other things, helped me to decipher the mysteries of scientific writing.

The initial plans we had for this thesis were difficult to carry out because of different reasons, and I had to modify my objectives. Therefore, much of all the field work I did during the first two years as a PhD student in ‘our small catchments’ cannot be found in this thesis, but I have learned a lot from the work out there in the forests, and it formed my view on modelling. Thanks to Nathalie, Mattias, Anna, Magnus and Ulrika for joint field days and running the stations.

The conferences in Grenoble, Lancaster, Hamburg, Akureyri, Braunschweig, Vienna and Nice were highlights and milestones during my time as a PhD student. Meeting other hydromodelists always was a pleasant and stimulating break. By mentioning no names I ensure not forgetting anybody. However, I have to mention and thank Prof. Keith Beven – he has been a source of inspiration for the work presented in this thesis.

Working together with other scientist always is stimulating. I had the pleasure of working together with a great number of colleagues and friends during the last couple of years. Thanks to all of them, especially to Chong-yu Xu for the interesting model discussions and to my roommates Meelis Mölder and Erik Kellner. I am glad to have run into Stefan Uhlenbrook; our Uppsala-Freiburg-H₂O-connection has been very beneficial. Living abroad, I of course felt homesick sometimes. Working together with Stefan, as well as Prof. Christian Leibundgut and Andreas Güntner, has not only been very pleasant and stimulating, but also gave me the chance to do science while looking at (and dreaming of) my home region, the Black Forest.

Not only scientific help is needed to succeed in doing a PhD, but also people who help to concentrate on science - and people who create distractions. In the first group Ulla Ahlinder definitely holds a top position, closely followed by Tomas Nord, Taher Mazloomian in the ‘Geotryckeriet’, Krister Lindé and his team at the ‘Geobibliotek’ as well as all the staff in the other libraries. The second group hardly can be listed completely here: all our friends (both those who did not forget us in all the years since we left them and Freiburg, and those we were happy to meet in our new hometown Uppsala), the internet and Tomas who helped to make the www useable for my obsessions, the SCF and our short-wave radio,

My family belongs to both groups. Whenever I got stuck during writing, paintings of my grandpa popped up on my computer, working not only as a screensaver, but also saving my state of mind. My parents encouraged my hydrological curiosity from the very beginning. Hikes in the Vosges were not always as far as planned, but extremely exciting for my brother Christoph and me, when we found a creek where a dam could be built. Showing me the strange water cycle of Escher, which is part of the cover picture, they “forced” me to study these phenomena in more detail – not a bad choice after all. PhD students often behave irrationally (Andreas, do you remember us sitting on a plaza in Nice and, instead of enjoying the lovely Mediterranean atmosphere, we were discussing TOPMODEL?!). Petra both showed understanding for such behaviour and managed, later supported by Ronja and Linnea, to keep part of me in ‘real life’. My first paper (paper II) was submitted eight hours before Ronja was born. She could walk and was just awaiting her sister Linnea when I got the reprints. Would it be possible to exemplify the slowness of science in a better way?

At the moment this thesis goes in to press, I do not know where the future will take us. Anyhow, this may be a good point to thank Sweden and the Swedes. Sverige är ett fantastiskt land, I appreciate the hospitality, and after all the years I still enjoy living here. As most scientists, I can only hope that the taxpayers got or will get good value for their money that financed my PhD student position.

Introduction

Everyone is curious about the future and public planning needs to know about it. An individual may want to know the weekend weather; public policy may be more concerned about the impacts of a possible global warming. Mathematical models are one of the possible tools when statements about the future are needed and in many cases public policy has to rely on such models. On the other hand, the reliability of mathematical models can be questioned and the obtained results may rather be called prophesy than prediction (Beven, 1993).

Both forecasts and predictions of the likely future states of hydrological variables are of importance for optimal operation in water management. Forecasting is mainly done in real-time and is specified in time, whereas predictions are more general and focus less on the exact timing (Klemeš, 1986a). Examples where forecasts are needed are flood warning (*e.g.*, when will a flood reach a certain town and how high will water levels rise) and operation of hydroelectric reservoirs (*e.g.*, how much water can be expected during spring flood). Predictions needed in planning are questions such as the magnitude of the probable maximum flood, the average runoff from an ungauged catchment or the hydroelectric potential. Such estimates can be achieved by different methods and mathematical models are one of them. Especially conceptual runoff models are frequently used for both forecasting and predicting.

Besides these long-standing applications, which focus mainly on the quantification of runoff, the increasing awareness of environmental problems has given additional impetus to hydrological modelling. Runoff models have to meet new requirements when they are intended to deal with problems such as acidification, soil erosion and land degradation, leaching of pollutants, irrigation, sustainable water-resource management or possible consequences of land-use or climatic changes. Linkages to geochemistry, ecology, meteorology and other sciences have to be considered explicitly and correct simulations of internal processes become essential.

Despite all efforts and progress during the last two decades (Hornberger and Boyer, 1995), hydrological modelling is faced by fundamental problems such as the need of calibration or the equifinality¹ of different model structures and parameter sets. These problems are linked to the limited data availability and the natural heterogeneity (*e.g.*, Jensen and Mantoglou, 1992; Beven, 1993; O'Connell and Todini, 1996; Bronstert, 1999). From another perspective many problems can be related to the procedures used for model testing. Traditional tests such as split-sample tests are often not sufficient to evaluate model validity and to assess the pros and cons of different model approaches, and more powerful tests are required (Kirchner *et al.*, 1996; Mroczkowski *et al.*, 1997). The need to utilise additional data in such tests has been emphasised in the recent years (de Grosbois *et al.*, 1988; Ambroise *et al.*, 1995; Refsgaard, 1997; Kuczera and Mroczkowski, 1998).

¹ Equifinality is defined as the phenomenon that equally good model simulations might be obtained in many different ways (Beven, 1993)

Testing runoff models against other variables than just catchment-outlet runoff is important for two main reasons. Firstly, in many hydrological questions and for other sciences, such as ecology, it may be of much more interest to know what happens within a catchment than at the outlet. Secondly, to have confidence in model predictions, which are often extrapolations beyond the testable conditions, it must be ensured that the model not only works, but also does so for the right reasons.

Procedures of model testing are usually called validation. This term, however, is used in different and sometimes mutually exclusive meanings (Rykiel, 1996). Both the indispensability (*e.g.*, Tsang, 1991; Mroczkowski *et al.*, 1997) and the impossibility (*e.g.*, Konikow and Bredehoeft, 1992; Oreskes *et al.*, 1994) of model validation has been emphasised.

In this thesis, the different meanings of validation are compared, a practical definition regarding conceptual runoff models is proposed and a catalogue of methods for model testing is given. The capabilities and limitations of conceptual runoff models are discussed on the basis of several such testing methods, for two conceptual runoff models, the HBV model (Bergström, 1976; 1995) and TOPMODEL (Beven and Kirkby, 1979; Beven *et al.*, 1995). In addition, some aspects of hydrological modelling are clarified that often cause confusion and, thus, hinder the scientific dialogue needed to judge the quality of different models. Definitions of some important terms are given in the appendix.

Background

Problems to be solved with conceptual runoff models

Models for extension of runoff series and data-quality assurance

Data series of climatic variables are often longer than the runoff series. In such cases a model can be calibrated with existing runoff data and the calibrated model can be used to compute runoff series from the climatic data. In a similar way, models can be used to fill gaps in runoff records.

There are many sources for deviations between simulated and observed runoff series and most of them can be connected to the model. Nevertheless, a model can be used for data-quality control. If it is impossible to fit the simulations to the observations, or if simulations of a model, which performed well for some other period, differ significantly from the observed data, then there is some indication that there might be an error in the measurements of runoff or an input variable.

Models for runoff forecasts

Conceptual models are used to obtain both short-term (a few days) and long-term (a number of weeks or months) forecasts of runoff (or other variables). For short-term forecasts, the calibrated model is run until today using observed climatic data. The differences between observed and simulated runoff might be used for updating

the simulations by changing the state or input variables. Then time series of the driving variables are generated based on a meteorological forecast and the model is run for the coming days.

Usually there is no meteorological forecast available for hydrological long-term forecasts (*e.g.*, water availability in a river during a dry period or amount of inflow into a reservoir during spring flood). An alternative is to use the corresponding time series of the climatic variables from a number of preceding years (*e.g.*, Bergström, 1995). Runoff is then simulated for each of these series, and the forecast is deduced with some confidence interval from the simulated runoff series.

Models for runoff predictions

Extreme floods. Predictions of probabilities and magnitudes of extreme events are essential for water management. The traditional approach of fitting distribution functions to the observed extreme values and extrapolating these functions can be criticised for different reasons (Linsley, 1986; Klemeš, 1986b). The main criticism is that the distribution functions have to be extrapolated far beyond the probabilities that can be justified from the available observations.

The modelling approach is an alternative to the distribution fitting (*e.g.*, Bergström *et al.*, 1992), but this approach can be criticised in exactly the same way: the model has to be applied far beyond the conditions used for development and calibration for computation of extreme floods. The only reason why we should rely more on the model than on distribution functions is that we have confidence in the validity of the model and, thus, assume that extrapolation of the model calculations are more reliable.

Effects of land-use change. The impact of land-use changes on hydrology is another issue of common concern. Models of different complexity have been used in several studies to address this topic (*e.g.*, Hillman and Verschuren, 1988; Caspary, 1990; Brandt *et al.*, 1988; Eeles and Blackie, 1993; Calder *et al.*, 1995; Dunn and Mackay, 1995; Parkin *et al.*, 1996; Nandakumar and Mein, 1997; Lørup *et al.*, 1998).

Two different approaches are possible. The first is to apply a model to a catchment and then to change parameter values in order to mimic the land-use changes. The other possibility is to use a catchment in which the land-use has changed. Here the model is calibrated to runoff data before or after the change occurred and the calibrated model is then used to simulate the runoff for the other period. In other words, a runoff series is generated which is assumed to agree with the situation that would have been observed, if land-use had not changed. In conceptual runoff models it is usually not possible to connect a certain change of parameter values to a land-use change and, thus, the latter approach is more suitable. This approach can not be used to make any predictions about the future impacts of a possible land-use change, but allows studying the effects of land-use changes in the past.

Effects of climatic change. Hydrology plays a key role in the problem of a potential global warming (Loaiciga *et al.*, 1996). Water is both an important part in the heat balance of the earth and a resource that will be affected strongly by a climatic change. Conceptual runoff models are frequently used to predict the effects of a potential change in global climate on hydrology (*e.g.*, Nĕmec and Schaake, 1982; Gleick, 1987; Parkin *et al.*, 1996; Viney and Sivapalan, 1996; Panagoulia and Dimou, 1997; Gellens and Roulin, 1998, Xu, 1999a). A review of these modelling approaches can be found in Xu (1999b).

The standard methodology to predict the hydrological response to a potential climate change using hydrological models includes three steps. First, the hydrological model is calibrated and validated using historical data. In a second step, the historical series of climatic data are modified corresponding to climate changes, which are gathered from global-circulation-model (GCM) predictions. Finally, the model is run with these modified data series as input and the new simulations are compared to the original simulations.

Besides the uncertainties of the current GCM predictions, the important question in such modelling studies is how well a runoff model performs for nonstationary conditions, *i.e.*, how reliable model simulations are for a period where the climatic input variables and land-surface properties differ from those during calibration.

Models and environmental issues

Both climatic and land-use changes will affect not only catchment runoff, but also hydrological processes and water availability within a catchment. Runoff models can be used to assess such effects (*e.g.*, Gleick, 1987; Kite, 1993; Dunn and Mackay, 1995). The reliability of the results depends on, besides the points mentioned above, the ability of the model to simulate internal variables such as groundwater levels or soil moisture.

Hydrology plays a key role in many environmental issues. Erosion processes are closely linked to hydrological conditions (*e.g.*, Evans, 1996; Gabbard *et al.*, 1998). Nitrogen cycling depends on a combination of hydrologic and biogeochemical controls (Cirimo and McDonnell, 1997). Runoff models can provide a basis for the hydrological part in environmental models and can be extended to study environmental issues such as acidification, deterioration of aquatic ecosystems, soil erosion, solute transport, nitrogen dynamics or pesticide pollution. The TOP-MODEL approach, for instance, has been used to represent hydrology in environmental models to study carbon budgets (Band, 1993), annual net primary production (White and Running, 1994) or vegetation patterns (Moore *et al.*, 1993). When an environmental model is built upon a runoff model the environmental modeller has to rely on the ability of the hydrological part to simulate the important processes and variables.

Models as scientific and educational tools

Models are not only used for forecasts and predictions, but also as intellectual tools in research and education. Models allow compilation of existing knowledge, can serve as a language to communicate hypotheses and can be applied to gain understanding. Development of a model, discussing a model failure or a sensitivity analysis may serve as a way to reflect about theories on the functioning of natural systems. A detailed model may not be operationally applicable at larger scales, but it may allow to study the system and, thus, to develop reasonable and applicable models for larger scales.

Models can be used to examine different hypotheses about the functioning of a catchment (*e.g.*, Bathurst and Cooley, 1996). A model may help to investigate which parameter values or input data are most crucial to be estimated accurately. Blöschl (1991), for instance, found that for a snowmelt model based on the energy balance, the simulations are affected more by uncertainties in albedo than those in air temperature, and conclude that further research should concentrate on albedo.

There is a relationship between model complexity and its value for understanding and education. Very simple models do not provide much new information, whereas very complex models are not understandable. Compared to the use of a model for forecasting or for predicting, the application of a model as an intellectual tool requires less accurate numerical agreement between simulations and observations, whereas the consideration of important processes and feedback mechanisms is more important.

The idea of models as intellectual tools to gain understanding is widespread in ecological modelling. The simple predator-prey models are examples of models that are valuable for explaining and comprehending without necessarily being a suitable tool for concrete predictions. Only models that are "... understandable, manageable and able to be fully explored ..." are suitable tools for understanding, whereas neither statistical nor complex models are appropriate (Grimm, 1994, p. 642). The first can not provide explanations, while the latter is incomprehensible. A balance between complexity and simplicity is crucial for studying the relevant processes and still to understand how the model is working. Increasing the complexity of an ecological model may only be justified as a means to include new important feedback mechanisms (van Oene and Ågren, 1995).

The term validation

Bredehoeft and Konikow (1993) state in an editorial about validation that "the word validation has a clear meaning to both the scientific community and the general public" (p. 178). This statement could easily be invalidated by reading a couple of scientific publications and asking a few representatives of the public. The term is used with different meanings and much of the widely differing opinions, whether validation is possible or not, can be attributed to this ambiguity.

Runoff models

In runoff modelling the term validation usually means the test of a model with independent data. Refsgaard and Knudsen (1996), for instance, define validation "... as the process of demonstrating that a given site-specific model is capable of making accurate predictions for periods outside a calibration period" (p. 2190). The term validation has been used in a more general meaning for the assessment whether the underlying concepts of a model are adequate for a certain catchment (Iorgulescu and Jordan, 1994) and for procedures that allow "... to discriminate between good and bad model hypotheses" (Mroczkowski *et al.*, 1997, p. 2325). Other interpretations can be found in literature. The term has been used to test whether a model could be applied (fitted) to a catchment (Krysanova *et al.*, 1998) or for a sensitivity analysis of the model parameters (Tuteja and Cunnane, 1997).

Refsgaard and Knudsen (1996) argue that only a model application can be validated while a general, not site-specific model validation is not possible. On the other hand, the primary aim of a model application is often not to demonstrate that the model works for one particular application, but to demonstrate its suitability for similar problems. Often the implicit argument can be found that a model is assumed to be 'valid' because it has been successfully applied (whatever is meant by this) in a number of previous studies. The fact that several different research groups have used a model certainly expresses some common confidence in the model, although the choice of which model to use may also be based on non-scientific reasons (freely distributed, user-friendly, ...).

Groundwater models

The question whether validation is possible or not has been discussed intensively for groundwater models in connection with the use of these models for assessing the safety of underground disposal of nuclear and toxic waste. Tsang (1991) as well as Beck *et al.* (1997) emphasise the importance of model validation and present broad views of validation, including not only comparisons of model simulations with measurements but also expert knowledge and validation of model assumptions. Konikow and Bredehoeft (1992), on the other hand, assert that model validation is impossible, and that models only can be invalidated. They provide examples demonstrating the limited accuracy of model predictions and argue, that the terms verification and validation are misleading. These terms should not be used as they convey an impression of correctness, which can not be justified scientifically, to the public (Bredehoeft and Konikow, 1993). McCombie and McKinley (1993) replied that the term validation in groundwater modelling usually is used for assessing that a model is 'good enough', and that this assessment is possible. Bair (1994) reported from a court case and pointed out that also the general statement that models can not be validated may give an incorrect impression to the public, namely the inadequacy of any model for any purpose.

Against the background of groundwater modelling and models in other earth sciences, Oreskes *et al.* (1994) argue that verification and validation of models is

impossible. Models can only be confirmed by demonstrating that model simulations agree with observations. This confirmation is only partially possible. They conclude that the main benefit of models is heuristic, *i.e.*, they see models as preliminary hypotheses assisting in gaining better understanding.

Ecological models

Ecology is another research area where testing and validation of models have been discussed frequently. Rykiel (1996) reviews these discussions and conclude that much of the confusion and the mutually exclusive statements about model validation arise from varying semantic and philosophical perspectives and from different validation procedures. Mayer and Butler (1993) compare techniques for validation, which they interpret as the comparison of simulated and observed data without the specification whether this data has been used for model development or calibration or not. Brown and Kulasiri (1996, p. 129) define validation as the "... process of evaluating the level of confidence in the model's ability to represent the problem entity" and emphasise that a model can not be expected to be absolutely valid. Power (1993) surveys different definitions and finds the distinction between three types of validation, as proposed by Gaas (1983) to be useful. Replicative validity ensures that the simulations agree with the observed data already used for model development and parameter estimation. A model is considered predictively valid if it can accurately simulate a variable or time period, which has not been used in model development and calibration, and structurally valid if it reflects the main couplings and behaviour of the real system. Kirchner *et al.* (1996) ask for generally accepted standards for model testing and validating. They define validity as "... adequacy for a special purpose" (p. 36) and note that to some degree all models are unrealistic. They emphasise that parameter calibration and the use of *ad hoc* model features often make validation less rigorous, *i.e.*, even inadequate models are likely to pass the tests.

Philosophy and semantics

In the fields of philosophy of science, the problems how to prove and disprove scientific theories have been disputed intensively (*e.g.*, Klemke *et al.*, 1988), whereas mathematical models seem to be a barely discussed topic (Morton, 1993). There is a fundamental difference between a theory and a model. A theory is assumed to be true, although it can not be verified but only falsified according to many philosophers of science (*e.g.*, Popper, 1934/1982; 1959/1968). For a model we know in most cases already from the beginning, that it is not true. This is definitely the case for conceptual runoff models; no real catchment consists of a number of boxes. Also physically-based, distributed models can easily be shown not to be true. Soil parameters, for instance, are never constant over a 100 m by 100 m square. Despite the fact that most models are not true, they are often needed because the governing theories, which may be true, are unmanageable (Morton, 1993). The question about absolute truth of a model is ill-posed. What can be achieved, and what is needed, is an assessment to which degree a model is an

appropriate description of the real system, and an estimation of the confidence in its predictions.

Alternative terms instead of validation have been proposed for model testing. Bredehoeft and Konikow (1993) suggest the term ‘history matching’, but this term does not distinguish between historical data used for calibration and such used for independent testing. Oreskes *et al.* (1994) propose the word confirmation. However, ‘to confirm’ is one of the words listed as explanation of ‘to validate’ in the dictionary (Allen, 1990, *The Concise Oxford Dictionary of Current English*) and ‘to confirm’ points to, among others, ‘to make definitely valid’. Thus, the term ‘confirmation’ is not less ambiguous than the term ‘validation’. Popper (1934/1982; 1959/1968) suggest the terms ‘corroboration’ (‘Bewährung’) and ‘degree of corroboration’ (‘Bewährungsgrad’) as a neutral term to describe the degree to which a hypothesis has passed tests. Similar to ‘confirmation’ the word ‘corroboration’ does not in common use express the limited and provisional acceptance better than ‘validation’ does (Rykiel, 1996).

The word ‘valid’, which is derived from the Latin word *validus* (strong, powerful), means well-grounded, sound or defensible, hence it differs from words like true or correct which are connected to the process of verification (Latin word *verus*, true). There is a need for a generally accepted vocabulary describing the qualifications of model predictions (Morton, 1993). In this thesis the term validation, although it has been criticised, is considered to be appropriate for use in connection to model testing. With reference to the false ‘impression of correctness’ it seems to be of importance to clearly state what is (not) meant by validation rather than to define a new term.

It should be mentioned that the term ‘verification’ is inappropriate for model testing although it can often be found in literature. To verify means to establish the truth, something which is hardly possible in science and absolutely not in modeling.

Material and methods

Validation of runoff models

A practical definition of validation

The use of models is faced by a paradox: models are most important for problems where a test of a simulation is not possible and less important for problems where a test is feasible. In other words, the request for model validation increases with decreasing possibilities to perform tests. The need to apply a model is, for instance, much larger for predicting a 1000-year flood than for predicting a 10-year flood. In the latter case enough data may be available to compute the flood from time series without any model.

Many authors emphasise that validation of a model is connected to a special purpose and that general validity never can be ascertained. However, a validation

for a special purpose is straightforward only in few cases. When a model is used to extend runoff series during stationary conditions in a catchment, the probable accuracy of these simulations can be assessed by simulating runoff for a period with observed data, which has not been used for calibration. In most other cases the model can not be tested under conditions that are likely to correspond to those during real applications. It is, for instance, not possible to evaluate directly how well a model predicts a 10 000-year flood or whether it is appropriate for runoff simulation after a climatic change. Under such circumstances it is only possible to roughly estimate the accuracy by predictions of a similar type (*e.g.*, a 10-year flood or runoff during periods with less/more rain). These tests should be accompanied by other tests that increase the confidence in the model more generally. The point is that, for a model that agrees with the real system in different respects (*e.g.*, with observed internal variables), extrapolation beyond the testable conditions is more reasonable than for a model that just matched runoff during some period.

In this thesis, validation is defined as the estimation of the confidence in the ability of a model to perform with a certain quality for its intended purpose. Validation is not restricted to an application in a special catchment but also includes a general assessment of the capabilities and limitations of a model. It is possible to use a well-defined methodology for validation for some purposes (*e.g.*, filling gaps in runoff series). For most purposes validation means a thorough model testing, which must consist of different tests. In this case validation is an ongoing process in which the contribution of independent research groups is of importance.

The idea behind this definition of validation is that as much as possible should be tested. It is often easy to find aspects of a model that support its validity – if one looks for such aspects. What should be looked at are those aspects that are likely to show discrepancies. In other words, the ‘risky’ simulations of a model should be studied rather than the ‘safe’ ones.

Catalogue of validation methods

Conceptual runoff models usually require calibration to estimate their parameter values. It is important to distinguish between situations where parameter values are changed to minimise the deviations between simulations and observations (calibration) and situations where such an optimisation is not performed (most usual type of validation). Parameter values may be changed in the latter case, but not based on the deviations between simulations and observations, and comparison with observations is only used to assess model performance.

The goodness of fit can be evaluated by different measures (Green and Stephenson, 1986; Servat and Dezetter, 1991; Węglarczyk, 1998). The efficiency as proposed by Nash and Sutcliffe (1970) is a dimensionless transformation of the sum of squared errors and has become one of the most widely used goodness-of-fit measures. The efficiency, here called R_{eff} , was used to evaluate model performance in most studies in this thesis. The log-efficiency and the volume error were also computed in some studies (see Tab. 1 for definitions). Other measures are needed

Table 1. *Goodness-of-fit measures*

Goodness-of-fit measure	Notation	Calculation	Value of a 'perfect' fit
Efficiency	R_{eff}	$1 - \frac{(Q_{obs} - Q_{sim})^2}{(Q_{obs} - \overline{Q_{obs}})^2}$	1
Log-efficiency	L_{eff}	$1 - \frac{(\ln Q_{obs} - \ln Q_{sim})^2}{(\ln Q_{obs} - \overline{\ln Q_{obs}})^2}$	1
Volume error	V_E	$\left \frac{(Q_{obs} - Q_{sim})}{Q_{obs}} \right $	0
Coefficient of determination	r^2	$\frac{((Q_{obs} - \overline{Q_{obs}})(Q_{sim} - \overline{Q_{sim}}))^2}{(Q_{obs} - \overline{Q_{obs}})^2 (Q_{sim} - \overline{Q_{sim}})^2}$	1

to assess model performance in particular aspects, *e.g.*, seasonality of model errors (Xu and Vandewiele, 1995), peak flow rates or low flow conditions.

Klemeš (1986a) proposed a hierarchical scheme for systematic testing of hydrological models.

- (a) Split-sample test: calibration based on one time period and validation on another period
- (b) Differential split-sample test: calibration on periods with certain conditions (climatic or land-use) and validation on periods with different conditions.
- (c) Proxy basin test: calibration of a model on data from one or several catchments and validation in another, but similar, catchment. Adjustment of parameter values according to catchment properties but no calibration is allowed.
- (d) Proxy-basin differential split-sample test: calibration of a model on data from one or several catchments and validation in another catchment with different characteristics. Adjustment of parameter values according to catchment properties but no calibration is allowed.

Examples of modelling tasks where the different tests are relevant are extension of runoff series (a), simulation of effects of climatic or land-use changes (b, d) and simulation of runoff from ungauged catchments (c, d).

In general, the scheme proposed by Klemeš (1986a) describes tests on how well a model can be transposed temporally and spatially. These tests are possible for other hydrological variables than runoff. The importance of this scheme can not be neglected, but as mentioned by Klemeš, it is a minimum standard rather than a complete catalogue of possible tests. The Klemeš testing scheme is extended below to allow more powerful model validation (Tab. 2).

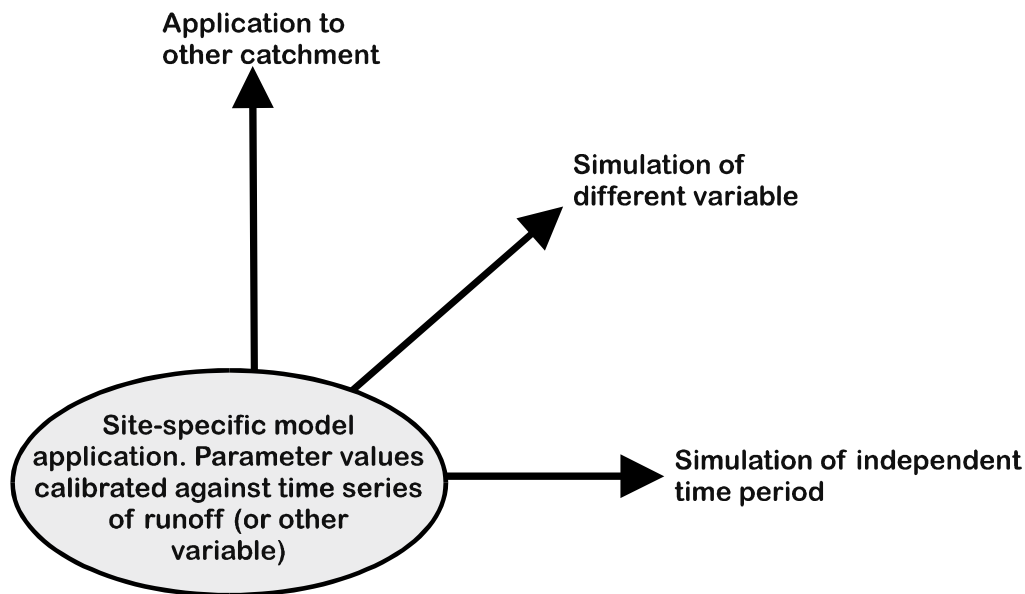


Figure 1. *Model validation by transposition*

In addition to assessment of model performance by transposing a model temporally or spatially, a model can be evaluated by changing the variable of interest. How well is one variable simulated if the model has been calibrated with respect to another variable (Fig. 1)? The following tests can be performed for all three directions of transposing (Fig. 2):

- A. Calibration against all data: is it possible to obtain acceptable results for different time periods/catchments/variables?
- B. Calibration against one part of the data, validation on the remaining data: Do parameter values give acceptable simulations for time periods/catchments/variables, which have not been used for calibration?
- C. Calibration against selected parts of the data, which differ from the rest of the data used for validation (*e.g.*, high/low flow conditions, lowland/mountainous catchments, spatially integrated/distributed data): Are parameter values suitable for simulations under conditions not considered during calibration?

Other tests may be more appropriate for certain studies. A blind test is in better agreement with real situations where no measured data is available (*e.g.*, ungauged catchments or effects of land-use change) (Bathurst and O'Connell, 1992; Ewen and Parkin, 1996). Here the task for the modeller is to perform simulations without having any access to the observed data. These data are only used for comparison with model predictions after the modelling work is completed. The quality of, for instance, simulated runoff series for other, ungauged catchments can be estimated from these comparisons. It must be noted that in this case not only the model but also the hydrological expertise of its user is evaluated.

Another way to assess model quality is addressing the problem of parameter uncertainty. The main reason for this test is to evaluate the effects of parameter uncertainty on predictions. If the effects of the parameter uncertainty are signifi-

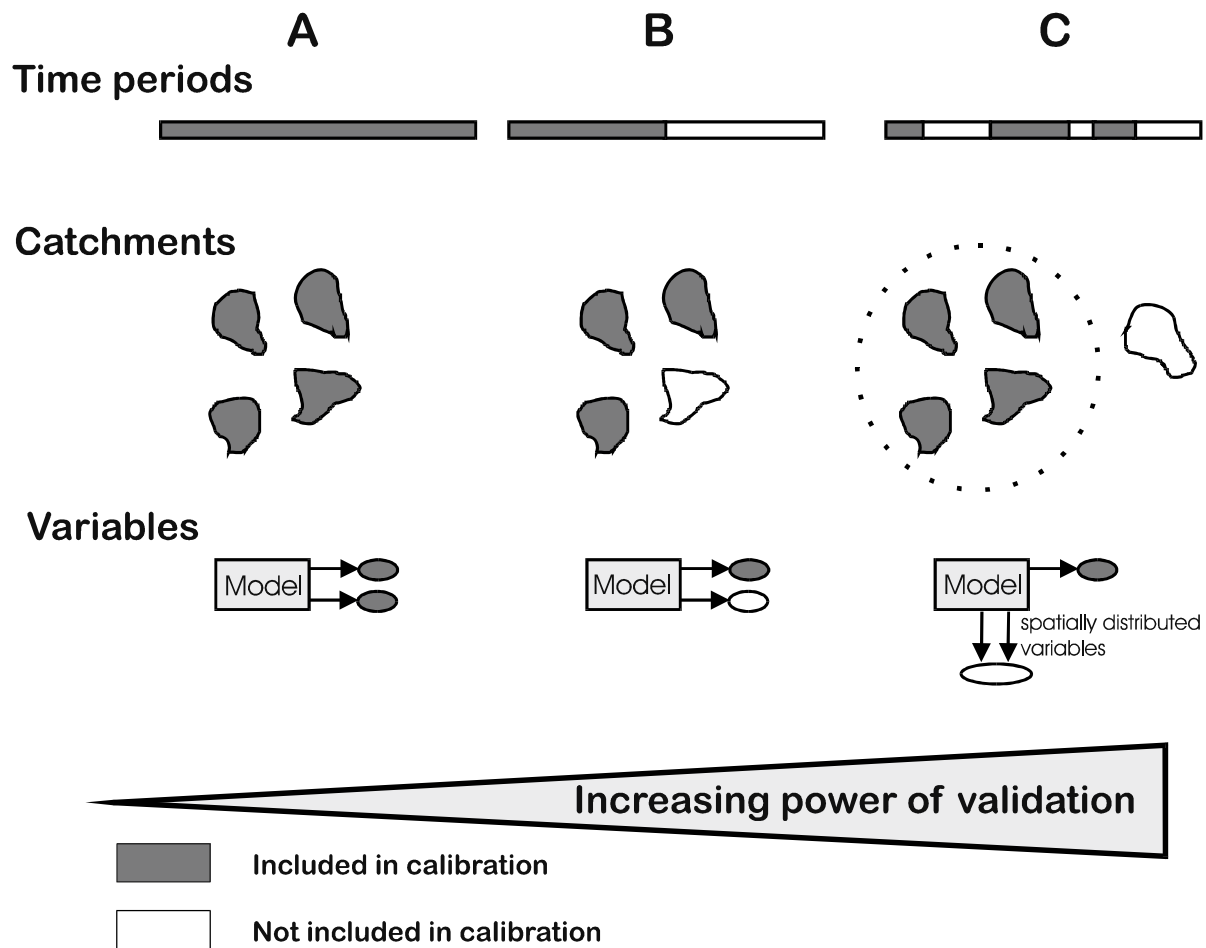


Figure 2. *Different ways of model testing with transposition of simulated time period, catchment or variable. A: Calibration against all data, B: Calibration against one part of the data, validation on the remaining data, C: Calibration for validation (e.g., high/low flow conditions, lowland/mountainous catchments, spatially integrated/distributed data)*

cant, the modeller may try to reduce the parameter uncertainty using, for instance, multi-criteria calibration.

Besides testing procedures, which focus on the validity of a model for a special application, the question of a more general validity of a model can be addressed with a validation based on regionalisation or testing the model against hydrological common sense (Tab. 2). In the first case calibrated parameter values and catchment characteristics are related to assess the physical soundness of a model. Some relationships can be expected from physical reasoning. Consequently, the existence of these relationships with objectively optimised parameter values would support the physical soundness of the model. On the contrary, relationships that can not be explained physically indicate weaknesses in the model structure.

Testing a model against hydrological common sense is subjective. Nevertheless, it allows the judgement of how reasonable the structure, the underlying assumptions and the behaviour of a model are. This is of value to assess the validity of a model in more general terms.

Table 2. Methods for model validation (from tests of a site-specific model application towards tests of the model in general)

Type of model test	Explanation	Purpose / comment
Split-sample test	Calibration based on one time period and validation on another period	How good are model simulations for independent periods?
Differential split-sample test	Calibration on periods with certain conditions (climatic or land-use) and validation on periods with different conditions	How good are model simulations for conditions beyond those available for calibration?
Identifiability of parameter values	(1) Can a unique set of parameter values be identified?	Evaluation of the effects of parameter uncertainty on predictions
	(2) Do calibrated parameter values depend on the chosen goodness-of-fit measure?	Parameter values should not depend on the goodness-of-fit measure, if the model is assumed to be a valid depiction of a catchment
Comparison with other methods or models	Comparison with simpler/more complex models, analytical solutions, alternative computation methods, ...	Comparison is no validation method in the strict meaning, but it allows assessing the benefits and shortcomings of a model
Partial model validation	Only a part of the model (e.g., one routine) or its underlying assumptions is tested.	Allows assessment of a part of the model more thoroughly
Multi-criteria calibration	Is it possible to find parameter values that are acceptable for simulation of different variables?	Is the model structure adequate?
Multi-criteria validation	Calibration against one variable and validation against other variables	How good are simulations of different variables when the model is calibrated with respect to another variable?
Proxy-basin test	Calibration of a model against data from one or several catchments and validation in another, but similar, catchment. Adjustment of parameter values according to catchment properties but no calibration is allowed	How good are model simulations for an independent but similar catchment?
Proxy-basin differential split-sample test	Calibration of a model against data from one or several catchments and validation in another catchment with different characteristics. Adjustment of parameter values according to catchment properties but no calibration is allowed	How good are model simulations for an independent catchment?
Blind test	Simulation without having access to the observed data	Corresponds to real situations where no measured data is available (e.g., ungauged catchments or land-use change), provides an indication about reliability in model predictions for other catchments

Table 2. *Continued*

Validation based on regionalisation	Relating optimised parameter values to catchment characteristics	Discussion of the physical soundness of a model: the existence of relationships that can be expected from physical reasoning supports the physical soundness of the model, relationships that can not be explained physically indicate weaknesses in the model structure.
Test against hydrological common sense	The judgement about how reasonable the structure and the behaviour of a model is, also called 'face validity'	Assessment of model qualities in general terms. Subjective test.

Description of models

The models used in this thesis were the HBV model (Bergström, 1976; 1995) and TOPMODEL (Beven and Kirkby, 1979; Beven *et al.*, 1995). Both models are examples of conceptual runoff models and have been applied in numerous studies in different geographical regions during the last two decades. Short descriptions for each model are given below. More detailed descriptions can be found in the literature (HBV: Bergström, 1992; 1995; Lindström *et al.*, 1997; paper I; TOPMODEL: Beven *et al.*, 1995; Ambroise *et al.*, 1996; paper II, paper V).

The HBV model. The HBV model is a conceptual model that simulates daily discharge using daily rainfall and temperature, and monthly estimates of potential evaporation as input. The model consists of different routines, where snowmelt is computed by a degree-day method, groundwater recharge and actual evaporation are functions of actual water storage in a soil box, runoff formation is represented by three linear reservoir equations and channel routing is simulated by a triangular weighting function (Fig. 3).

The HBV model has been used for different hydrological tasks, *e.g.*, to compute spillway design floods or flood forecasting (Bergström *et al.*, 1992), for water-balance modelling at large scales (Bergström and Graham, 1998), simulation of groundwater levels (Bergström and Sandberg, 1983) and to study the effects of changes in climate (Saelthun, 1996) and land use (Brandt *et al.*, 1988). Furthermore, the model has been modified to simulate the transport of solutes (Bergström *et al.*, 1985; Lindström and Rodhe, 1986; Arheimer and Brandt, 1998).

TOPMODEL. The idea of the TOPMODEL approach is to represent topographic effects on hydrology by a topographic index. This TOPMODEL index is defined as $I = \ln(a/\tan\beta)$, where a is the local upslope catchment area per unit contour length and β is the slope angle of the ground surface. The index describes the tendency of water to accumulate (a) and to be moved downslope by gravitational forces ($\tan\beta$). For steep slopes at the edge of a catchment a is small and $\tan\beta$ is large which yields a small value for the topographic index. High index values are found in areas with a large upslope area and a small slope, *e.g.*, valley bottoms. There are two central equations derived by the TOPMODEL theory. The first

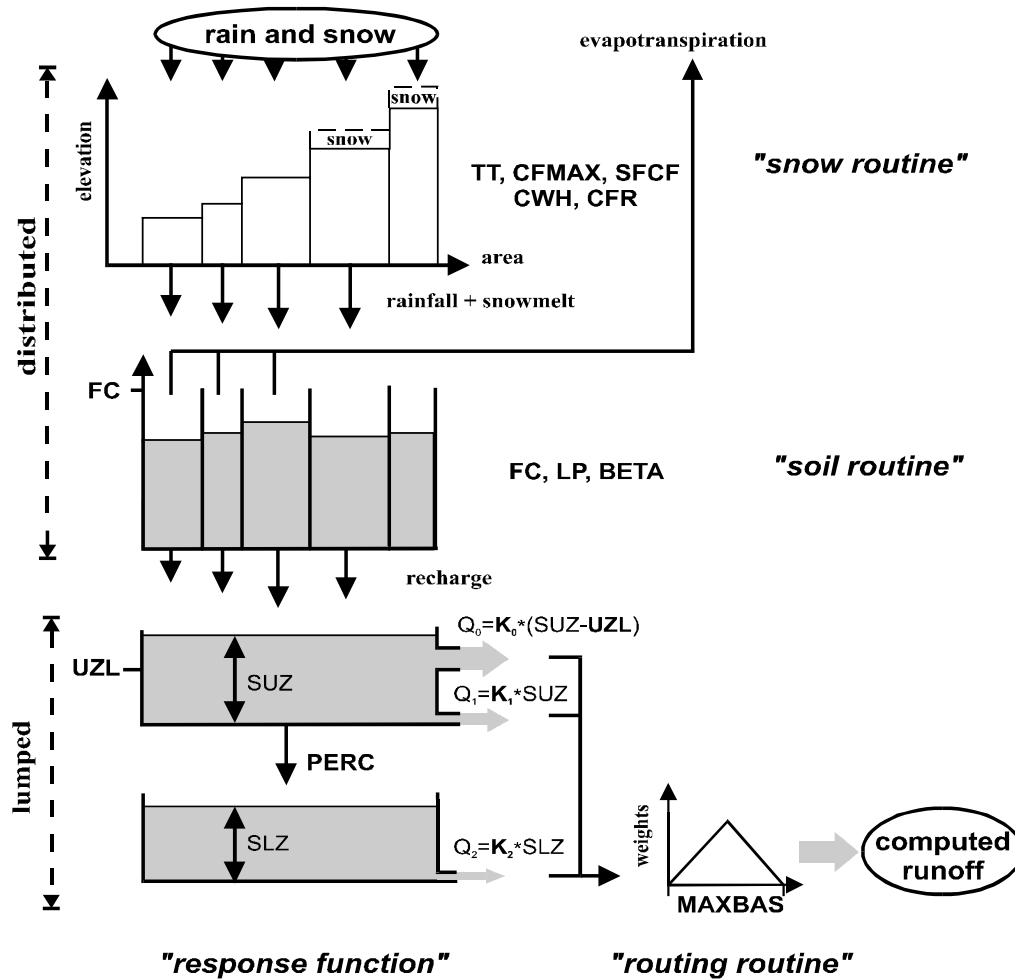


Figure 3. Structure of the HBV model (parameters in bold capitals)

relates the mean groundwater level within the catchment to the local groundwater levels at any location within the catchment. The second links the catchment runoff from the saturated zone to the mean groundwater level.

The distribution of wetness states over a catchment can thus be simulated in an easy way with low computational demands as topography is taken into account by a distribution function of indices. This has made the model very popular, especially since digital-elevation models (DEMs) have become easily available in the last decade. Apart from runoff modelling the TOPMODEL approach has been used for geochemical modelling (Robson *et al.*, 1992), to represent the hydrological part in ecological models (Band *et al.*, 1993; White and Running, 1994), and to aggregate soil-vegetation-atmosphere transfer (SVAT) models to larger scales (Famiglietti and Wood, 1994a,b)

Study catchments

The different tests were performed in various catchments located in four different regions in Sweden and Germany (Tab. 3).

NOPEX research area (papers I, III and VI). The NOPEX (Halldin *et al.*, 1998) research area is a region of roughly 50 km by 100 km situated north-west of

Uppsala. A characteristic of this landscape is its flatness with elevations between 30 and 70 m a.s.l. for the main part of the region. The area is covered by a mixture of boreal forests, agricultural lands, bogs and lakes. The predominating soils are till and clay. See Bergqvist (1971) (Nåsten) and Seibert (1994) (Sävaån, Svartån and Tärnsjö) for further descriptions of the respective catchments.

Gårdsjön, G1 ROOF (paper II). The G1 ROOF catchment is a small (6300 m²), forested headwater catchment located on the west coast of Sweden close to Lake Gårdsjön (Andersson *et al.*, 1998). The topography is characterised by a central valley with steep side slopes. The bedrock is covered by a till soil of varying depths (0 - 1.40 m). In a de-acidification experiment (Bishop and Hultberg, 1995) the catchment was covered by a transparent plastic roof constructed below the tree crowns and water input to the catchment was simulated by an irrigation system beneath the roof.

Kassjöån (papers III and VI). The former International Hydrological Decade representative basin Kassjöån (Waldenström, 1977) is located in central Sweden, 50 km NW of the city of Sundsvall. The landscape is moderately hilly and characterised by slope lengths of the large-scale topography being 0.5-2 km with height differences of 50-150 m. The area is mostly forested, the soil cover is thin and till soils are prevailing. The Lilla Tivsjön catchment (paper VI) is one subbasin.

Black Forest, Brugga (paper IV). The Brugga basin (40 km²) is located in the Southern Black Forest in south-west Germany. It is a mountainous catchment with elevation ranging from 450 to 1500 m a.s.l. and a nival runoff regime (Lindenlaub *et al.*, 1997). The area can be classed into three topographically different units: steep valley sides (75 % of the catchment area), hilly uplands (20 %) and narrow valley floors (5 %). The bedrock consists of gneiss, covered by soils of varying depth (0.5-10 m). 75 % of the basin is forested and the remaining part is used as pasture; urban land use is below 2 %.

Table 3. Catchments characteristics

Characteristic	Lilla Tivsjön	Tärnsjö	Svartån	Sävaån	Nåsten	G1 ROOF	Brugga
Region	Kassjöån	NOPEX	NOPEX	NOPEX	NOPEX	Gårdsjön	Black Forest
SMHI ^a station number	42-1920	54-2299	61-2216	61-2247	61-1742	-	-
Area [km ²]	12.8	14	730	198	6.6	0.0063	40
Forest percentage [%]	88	85	69	66	87	100	75
Lake percentage [%]	2.7	1.8	4.0	0.9	0	0	0
Range of elevations [m a.s.l.]	246-440	55-105	25-215	15-105	18-55	123-143	450-1500
Mean annual precipitation [mm]	586 ^b	729 ^c	733 ^c	734 ^c	693 ^c	1020 ^c	1750 ^b
Mean annual runoff [mm]	262	266	276	194	235	330	1200

^a Swedish Meteorological and Hydrological Institute

^b Uncorrected values

^c Corrected for systematic measurement errors (correction increased yearly amounts by 15-25 %)

Illustration of validation methods

Proxy-basin test

Seibert *et al.* (1999) applied the HBV model to three nested catchments located in the Black Forest. The catchments were similar but varied in size (15, 40 and 257 km²). The test was no proxy-basin test in the strict meaning since runoff series from nested catchments are not independent (the direct dependence was minor because in all cases the area of the nested catchment was small compared to the entire catchment area). On the other hand, the test was assumed to be harder than a usual proxy-basin test since the parameter sets had to be transferred between catchments of significantly different sizes.

Using a parameter set optimised in one catchment in the two other catchments gave acceptable results in terms of the model efficiency (on average 0.76). The efficiency was higher when using the parameter set calibrated in the respective catchment (0.84) and when using a single parameter set that was optimised with respect to all three catchments (0.81). Transferring the series of specific runoff directly (instead of the calibrated parameter values) yielded much poorer runoff estimates (R_{eff} on average 0.51).

Results were less favourable for the HBV model in the study by Seibert (1999). Regionalised parameter values, which had been derived from 11 catchments within the NOPEX region, were used to simulate runoff in independent catchments. The agreement with observed runoff was weak (R_{eff} on average 0.6) and the simulations were only slightly better than the direct use of the mean runoff series from the 11 catchments.

Identifiability of parameter values (paper I)

The HBV model was applied to two catchments in the NOPEX region (Sävån and Svartån) and a Monte-Carlo procedure was used to evaluate parameter uncertainty. Wide ranges of possible values were set for each parameter based on ranges of calibrated values from other model applications (*e.g.*, Bergström, 1990; Braun and Renner, 1992). 500 000 parameter sets were generated for each catchment using random numbers from a uniform distribution within these ranges for each parameter. The model was run for each parameter set and the efficiency as well as a new goodness-of-fit measure, which combined the efficiency, the log-efficiency and the volume error, were computed.

High efficiency values could be obtained for most parameters with values varying over wide ranges. The combination of the efficiency with other goodness-of-fit measures helped to reduce the parameter uncertainty for a few parameters. Simulations of a large spring flood and a summer period with low flow indicated significant effects of the parameter uncertainty on model predictions. Considerably different hydrographs were simulated with the different parameter sets, which had performed equally well for the 10-year calibration period (Fig. 4).

Uhlenbrook *et al.* (1999) obtained similar results in a following study in the Brugga catchment. Again, only few parameters were well defined and for most parameters good simulations were found with values varying over wide ranges. The effects of the parameter uncertainty on model predictions were evaluated by the computation of design floods and of low flows. These flow predictions were found to vary considerably. The peak discharge of a flood generated by a rainfall sequence with a probability of 0.01 yr^{-1} , for instance, varied from 40 to almost 60 mm d^{-1} (Fig. 5).

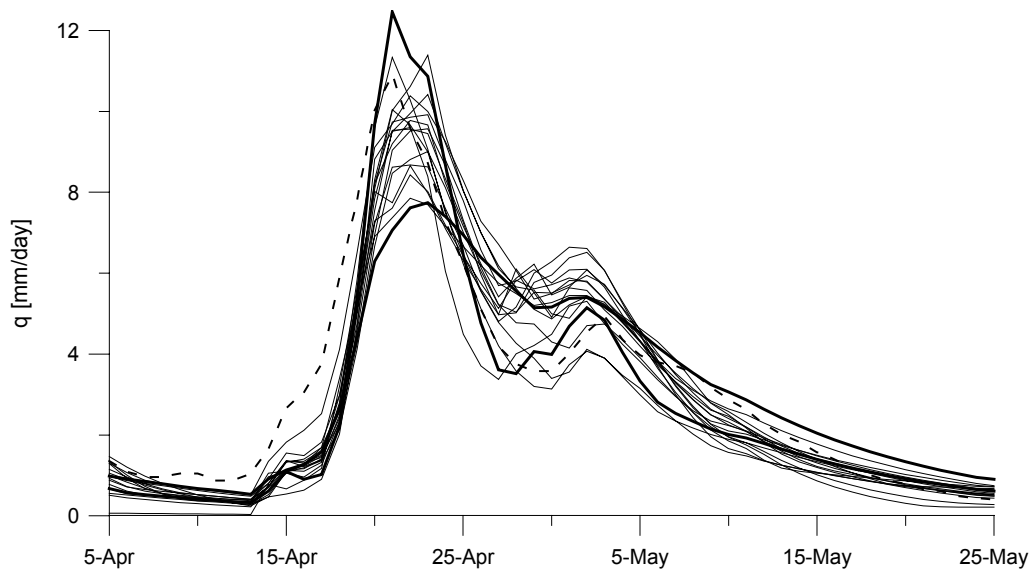


Figure 4. Spring flood 1985 (Sävaån) simulated with parameter sets that gave a fit with R_{eff} within 0.02 from the maximal value of R_{eff} . The simulations with the lowest and highest peak discharge are shown with thick lines, the observed hydrograph is shown with the dashed line. (from paper I)

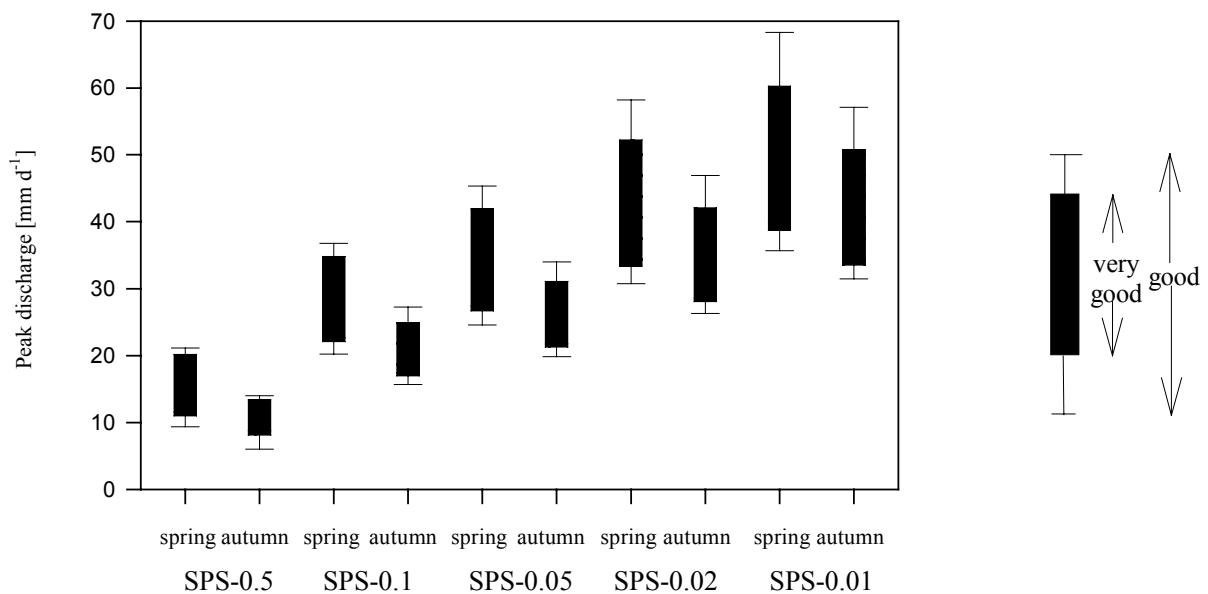


Figure 5. The range of predicted peak discharge generated by synthetic rainfall sequences (SPS) of different probabilities applied in spring and autumn 1980 simulated with very good ($R_{eff} > 0.860$) and good ($R_{eff} > 0.850$) parameter sets (the highest R_{eff} value was 0.867) for the Brugga catchment. (from Uhlenbrook *et al.*, 1999)

Validation of simulated groundwater levels (paper II)

TOPMODEL was applied to the small G1 ROOF catchment near Lake Gårdsjön. Simulated hourly runoff series agreed reasonably well with the observations during both calibration ($R_{eff} = 0.77$) and validation ($R_{eff} = 0.69$) periods. The simulation of local groundwater levels agreed poorly with observations. For three groundwater tubes with continuous measurements the general trends of the fluctuations of the simulations corresponded with the observations (*i.e.*, both did go up and down at the same time), but both amplitudes and mean values differed considerably. Manual measurements at 34 additional groundwater tubes allowed the study of the two deviations for a larger number of locations. According to TOPMODEL there should be a linear relationship between local groundwater levels and TOPMODEL indices. The correlation was weak (Fig. 6, light symbols) and the r^2 values were between 0.1 and 0.35 for 80 % of the 32 investigated occasions. Besides the large scatter, the simulated levels differed systematically from the observations, when using the parameter values calibrated against runoff.

Groundwater-level observations at a single point in time were used to replace the local topographic index values computed from the topography by calibrated topographic-soil index values. The simulations could be improved significantly by this means (Fig. 6). The practical use of this method is limited because it is only applicable for locations where at least one groundwater-level measurement is available.

The main result of this TOPMODEL application was that groundwater-level simulations were not reliable when the model had been calibrated only to runoff observations.

Partial model validation: the TOPMODEL index (paper III)

The TOPMODEL index allows mapping of wetness distribution within the landscape from topographic data. Testing the relationship between index values and local wetness is important to assess the suitability of the TOPMODEL index as a wetness indicator. With current measurement techniques it is practically impossible to measure soil moisture or groundwater levels with a spatial coverage that allows such tests over larger areas. The occurrence of mires, which were assumed to delineate the wettest parts of the landscape, was used as alternative field data in paper III. The possibility to predict the distribution of mires in a catchment from topographic data using the TOPMODEL index was investigated for two areas with contrasting topography: the Nåsten catchment in the flat NOPEX region and the hilly Kassjöån basin.

The index values for mire and non-mire areas were similar in Nåsten. The failure to predict mires, and probably also other wetness classes, in Nåsten could be explained by the spatial resolution of the DEM used for the index calculations. Typical topographic features in this catchment had a length scale of only a few tenths of metres and were not captured by the 50 m by 50 m DEM.

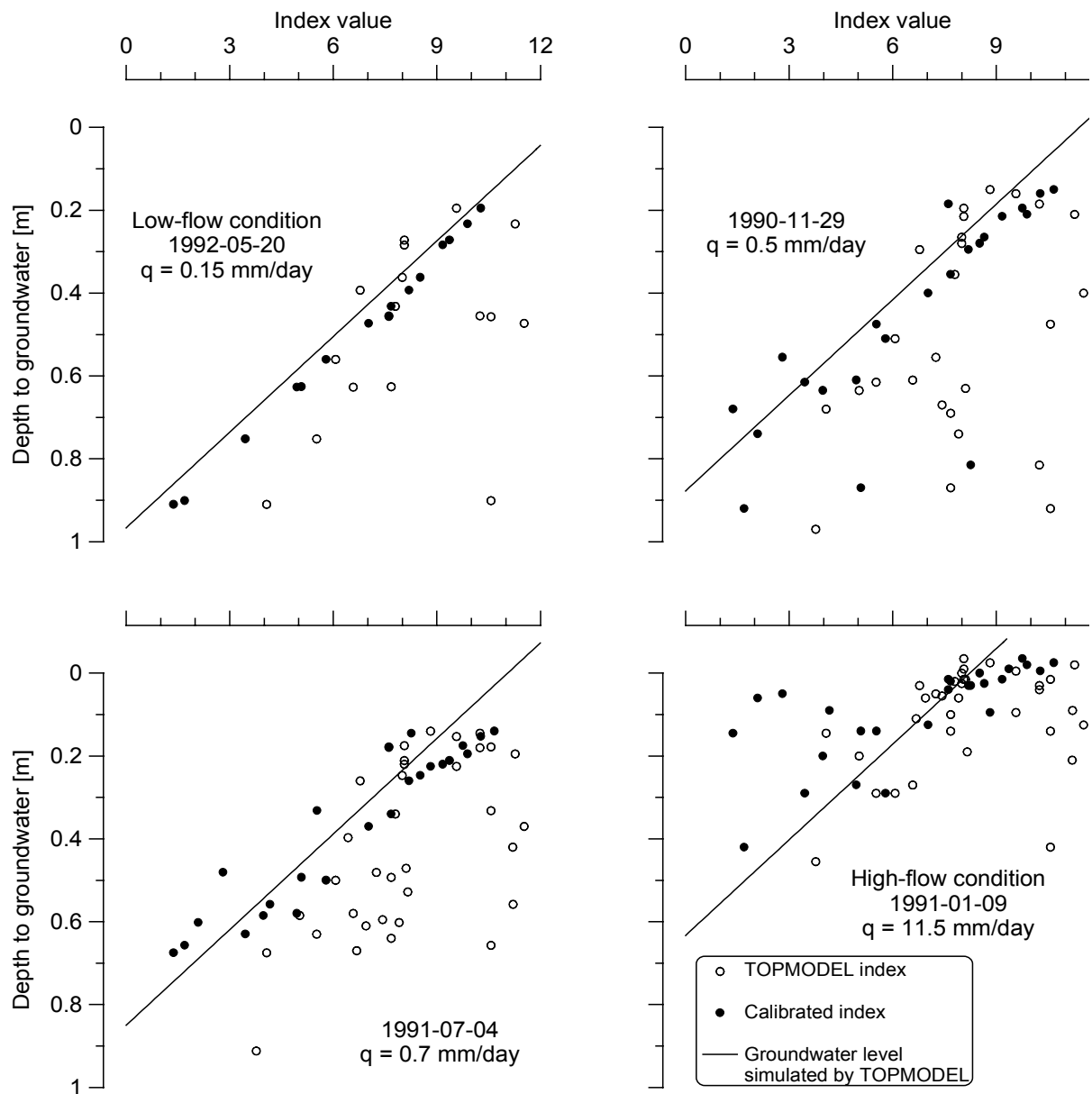


Figure 6. Observed groundwater levels plotted against the TOPMODEL index (light symbols) and the index that was calibrated based on groundwater measurements on 4 April 1991 (solid symbols) (G1 ROOF catchment). (modified from naner II)

In Kassjöån, mires had on average higher index values and the frequency distributions of topographic indices for mire and non-mire areas were clearly different (Fig. 7). On the other hand, there was a large overlapping between the two distributions and, thus, the predictive power of the TOPMODEL index was limited. The areal pattern of high index values and mires agreed roughly, but a quantitative measure showed poor results with only 40 % of the observed mire area being predicted by the index. The deviations put the validity of the TOPMODEL index as a wetness indicator into question but could also be related to other problems: scale and spatial resolution, the methods of index calculation, and the suitability of the mapped mires to test the index. A tentative field control indicated that some of deviations between mapped mires and mires predicted by high index values could

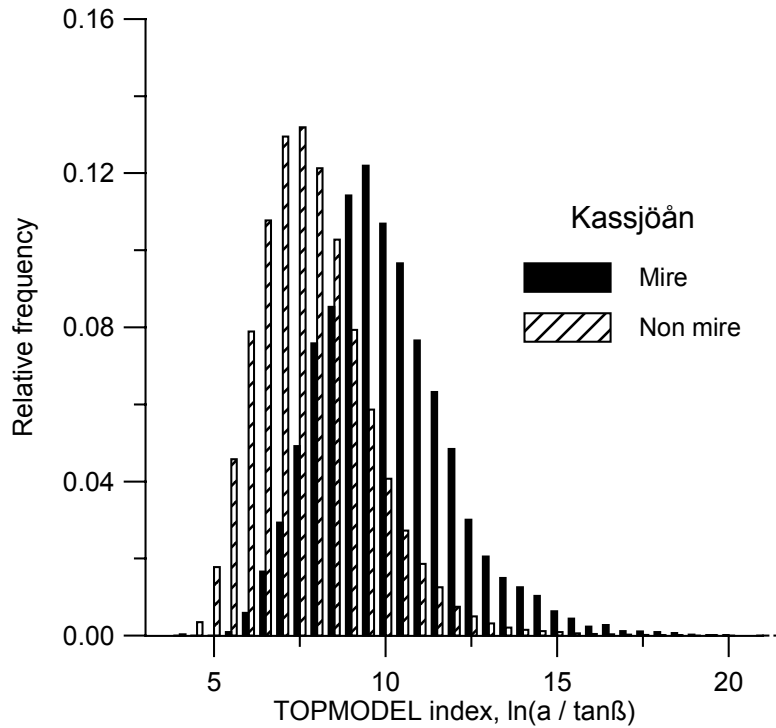


Figure 7. Frequency distributions of the TOPMODEL index for mire and non-mire cells (Kassjöån basin).

be attributed to erratic mapping and to the assumption that only mapped mires were the wettest areas.

The partly successful prediction of mires in Kassjöån supports the possibility that topographic data can be used also for prediction of the spatial wetness distribution, but at the same time it demonstrated the need to investigate this possibility by more detailed field studies. An example with two neighbouring mountain areas demonstrated the importance of bedrock geology for mire occurrence and illustrated the fact that a simple relation between topography and wetness, as the TOPMODEL index, has to be used with great care.

Multi-criteria validation of TOPMODEL (paper IV)

TOPMODEL was applied to the Brugga catchment and its performance was tested in several ways. The split-sample test gave good results for the runoff simulations during both calibration ($R_{eff} = 0.85$) and validation (0.93) period. The calibrated parameter values for the maximum capacity of soil storage and transmissivity were within the ranges obtained from experimental data. The ranges of reasonable values were rather large for both parameters and, thus, these did not offer any rigorous validation criteria.

Güntner *et al.* (1999) mapped saturated areas based on pedological and geobotanical characteristics. The mean simulated percentage of saturated areas (5.5 %) corresponded well to the mapped saturated area (6.2 %), whereas comparison of the spatial distribution of mapped and simulated saturated areas indicated a poor agreement. Only 34 percent of the simulated saturated areas were also mapped as saturated area (Fig. 8). Systematic deviations could be seen besides random errors

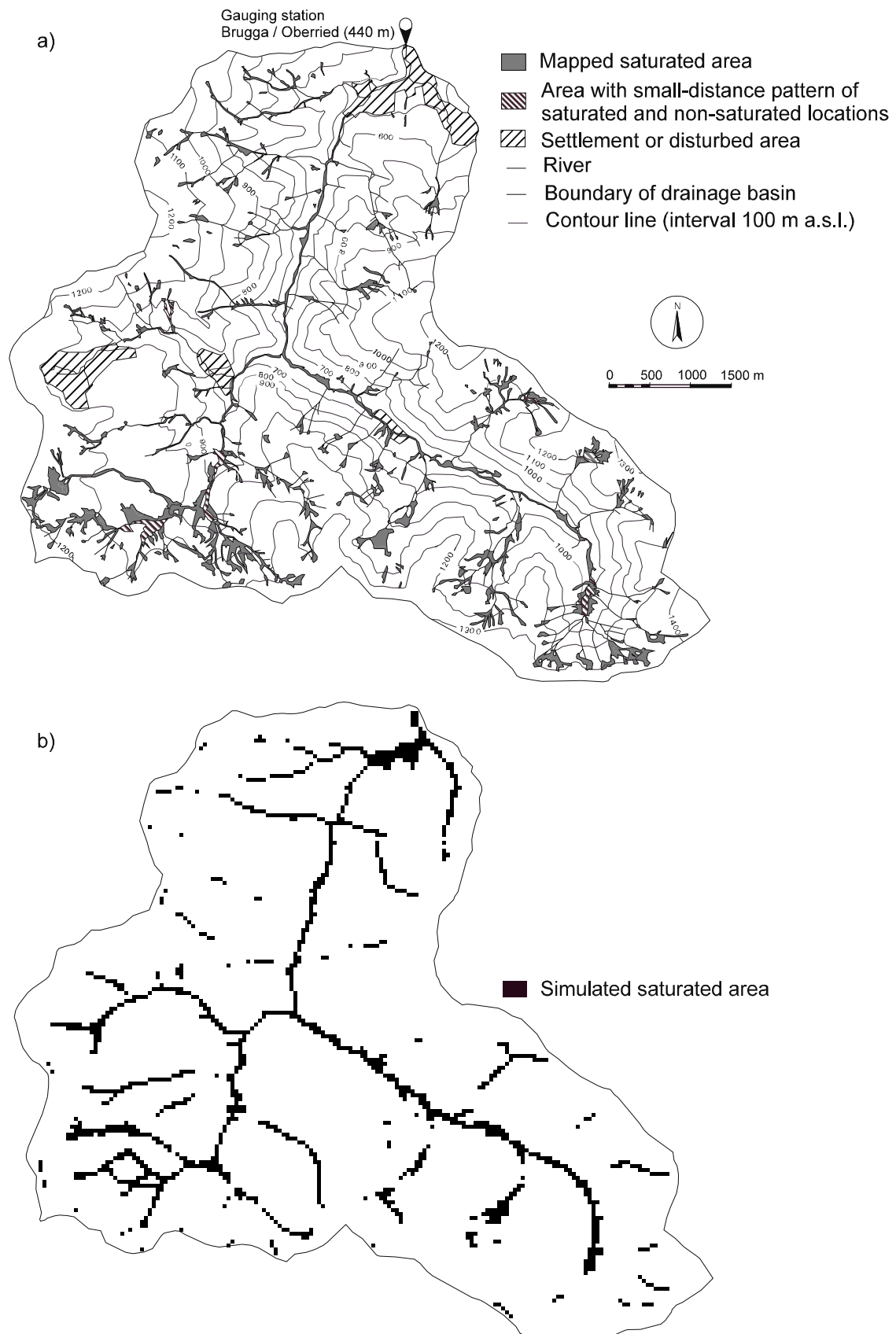


Figure 8. Spatial distribution of saturated areas in the Brugga catchment. (a) Mapped. (b) Predicted by the TOPMODEL index (the threshold value of the index, by which saturated areas were distinguished from non-saturated areas, was chosen so that the portion of the saturated areas equalled the mapped percentage, which was 6.2 % of the catchment area). (from paper IV)

and errors caused by the too coarse resolution of the digital elevation model (50 m). The model did not represent mapped saturated areas on steep slopes and close to the top of valley sides. Furthermore, the simulated percentage of saturated areas was very variable with time with a maximum close to 20 percent during high flow periods. This was in contrast to field observations, which indicated saturated areas to be less dynamical. A percentage higher than 10 % was not reasonable in the study area, except for extreme situations which did not occur during the study period.

The simulated runoff components were compared to those derived from a hydrograph separation based on electric conductivity. Significant differences could be recognised between modelled saturation-excess-overland flow (Q_{sat}) and the separated runoff component 'event water' (Q_{event}): (1) both first appearance and peak of Q_{sat} was ahead of those of Q_{event} , (2) the contribution during peak runoff of Q_{sat} was larger than that of Q_{event} , (3) the contribution of Q_{sat} ended soon after peak runoff while that of Q_{event} continued, (4) the volume of Q_{sat} was smaller than that of Q_{event} . Based on descriptions of TOPMODEL (*e.g.*, Beven and Kirkby, 1979; Beven *et al.*, 1984; Beven *et al.*, 1995), where it is said that "any rain falling upon the saturated areas is taken to be runoff" (Beven *et al.*, 1995, p. 633), the modelled saturation-excess-overland flow should be interpreted as pure event water. Consequently, Q_{sat} always should be less than or equal to Q_{event} and the first two differences indicate a weakness of the model. These differences may partly be explained by the use of the electrical conductivity as tracer, which causes an underestimation of the event component because of the crude assumption that event water retains its electrical conductivity of rainwater on the way to the catchment outlet.

Furthermore, it may be argued that there is an exchange of water between the simulated flow components and that the rain falling on areas simulated to be saturated gives only the amount but not the source and flow paths of a fast flow component. The results obtained in the Brugga catchment suggest such an interpretation. TOPMODEL cannot be validated against environmental tracer data with this vague definition of runoff components and much of the idea to distinguish between two different flow components is lost.

In summary, although the runoff simulations were satisfying major inadequacies of the model could be identified by the different validation methods. These inadequacies could be related to the concept of runoff generation and the simulation of spatially-distributed saturated areas.

Validation of TOPMODEL against hydrological common sense (paper V)

The capability to simulate spatial variations of groundwater levels (or surface wetness) is one of the most attractive features of TOPMODEL, but validation against measured groundwater levels has often not been successful (*e.g.*, Iorgulescu and Jordan, 1994; paper II, Lamb *et al.*, 1997). One might argue that these failures were specific for the different test catchments and that they do not generally invalidate the predictions of local groundwater levels by TOPMODEL.

In paper V the question on TOPMODEL's ability to predict local groundwater levels was deliberated more in principle. The underpinning assumptions of the TOPMODEL theory, their reasonableness and the errors generated by these assumptions were discussed. The most problematic assumptions were those of steady-state flow rates and spatially uniform recharge to the groundwater.

The assumption of a spatially uniform recharge to the saturated zone is needed to derive the simple relationship between local and mean groundwater level, but it may be unreasonable in many situations. Looking at the situation during and shortly after a rainfall event, one should expect the recharge to increase with decreasing depth to the groundwater for two reasons: the vertical path through the unsaturated zone is shorter and there is less storage per unit depth possible in the unsaturated zone above the groundwater. For longer time intervals, local recharge depends on evaporation, rainfall and snowmelt which all can be expected to vary spatially. In most TOPMODEL applications spatially variable recharge rates are computed, although this is inconsistent with the underlying assumption of a spatially uniform recharge. In paper V an example was used to illustrate that the use of spatially variable recharge rates, even though these may be physically more correct than the uniform rate, causes a physically unreasonable, upslope redistribution of water.

The steady-state assumption causes all simulated groundwater levels in a catchment to always rise and fall in parallel. This did not agree with examples of measured data from three catchments in Sweden and studies found in literature. As a consequence of the steady-state assumption, topography has only little effect on the simulated groundwater dynamics and runoff. Groundwater levels and runoff from the saturated zone are neither delayed nor dampened. The upslope sub-catchment at a specific location is represented only by the value of a , *i.e.*, topography within this sub-basin is of no importance for the groundwater level at this location. If the slope and the upslope area are equal for two locations, the simulated groundwater levels will always be exactly the same, independent of any difference in their upslope topography.

The steady-state assumption combined with the spatially-uniform-recharge assumption implies that the simulated contribution of groundwater to discharge per unit area is spatially uniform over the basin at any time. Results from field experiments indicate that the situation can be very different in reality (Hinton *et al.*, 1993; Sidle *et al.*, 1995).

The conclusion in paper V was that the fundamental assumptions underpinning the TOPMODEL approach obstruct a correct simulation of the spatial and temporal dynamics of the groundwater table.

Multi-criteria calibration to runoff and groundwater levels (paper VI)

Calibration against more than one output variable of a model makes the simulations of internal processes more reliable. The HBV model, together with a genetic algorithm for optimisation, was applied in two catchments with different geology

where, in addition to observed runoff, groundwater-level time series were available for calibration. In the first catchment (Lilla Tivsjön) it was possible to calibrate the model according to both runoff and groundwater levels. The respective goodness-of-fit values were slightly lower compared to the values when calibrating with respect to only one criterion. Calibrating against only runoff resulted in poor simulations of groundwater levels and considering only groundwater levels during calibration led to poor runoff simulations. The effect of the multi-criteria calibration on the parameter-value identifiability was studied using a number of calibration trials. The parameter uncertainty was reduced significantly, compared to the calibration to only runoff, when groundwater levels were included (Fig. 9).

In the second catchment (Tärnsjö) the drop of goodness when calibrating simultaneously to both runoff and groundwater levels was larger than in the Lilla Tivsjön catchment. This difficulty to simulate both runoff and groundwater levels with the same parameter set was related to the special geological situation with an esker running through part of the catchment. The decrease of the respective goodness-of-fit values was of the same order as in the Lilla Tivsjön catchment for a modified model structure, which was assumed to agree better with the real situation.

To summarise, the multi-criteria calibration both helped to reduce the parameter uncertainty and motivated the use of a more adequate model structure.

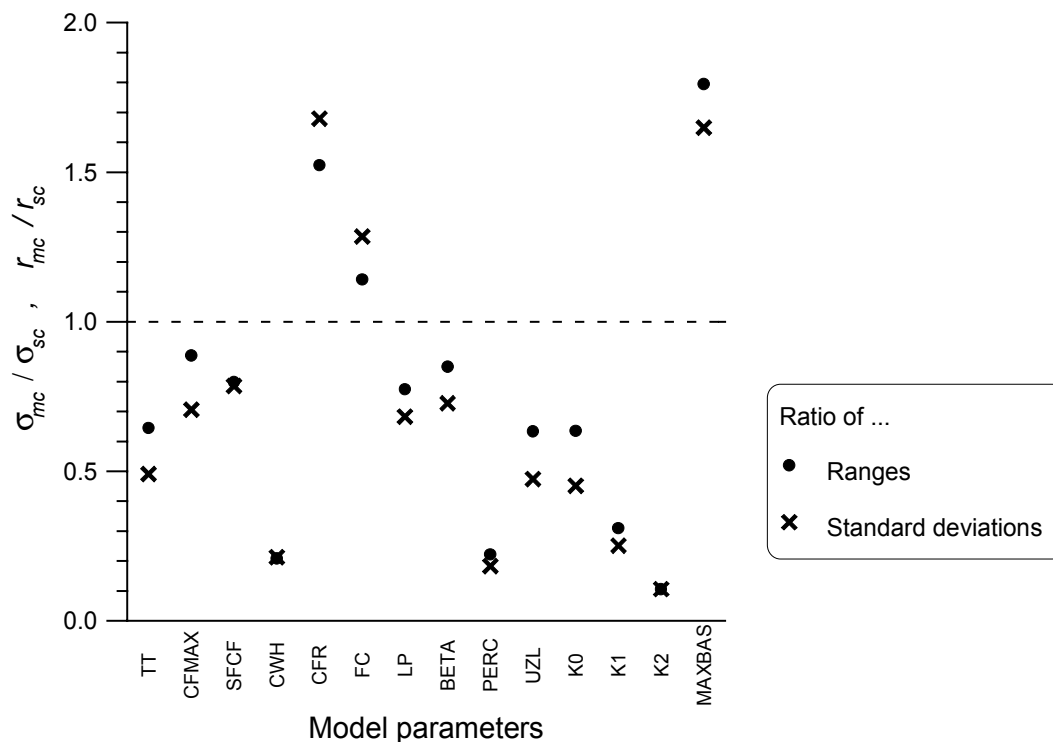


Figure 9. Comparison of the variations of calibrated parameter values obtained by single- and multi-criteria calibration of the HBV model in the Lilla Tivsjön catchment. Both ranges and standard deviations were calculated from the 20 best of 25 calibration trials. The ratios were computed by dividing those values for standard deviation and range from calibrations against both runoff and groundwater levels, σ_{mc} and r_{mc} , by the values from calibration against only runoff, σ_{sc} and r_{sc} . (from paper VI)

Validation based on regionalisation

Relationships between model parameters and catchment characteristics were used to discuss the validity of the HBV model (Seibert, 1999). The model was applied to 11 catchments in the NOPEX region and the calibrated parameter values were related to the catchment characteristics forest and lake percentages, and catchment area. Different relationships between model parameters and catchment characteristics could be detected (Fig. 10). Relationships between forest percentage and snow parameters supported the physical soundness of the model as they had been expected from results of experimental studies found in literature. On the other hand, relationships between lake percentage and soil parameters called the physical soundness of the model into question as they could not be explained by the physical processes in the soil but by the dominating effect of lakes to runoff variations.

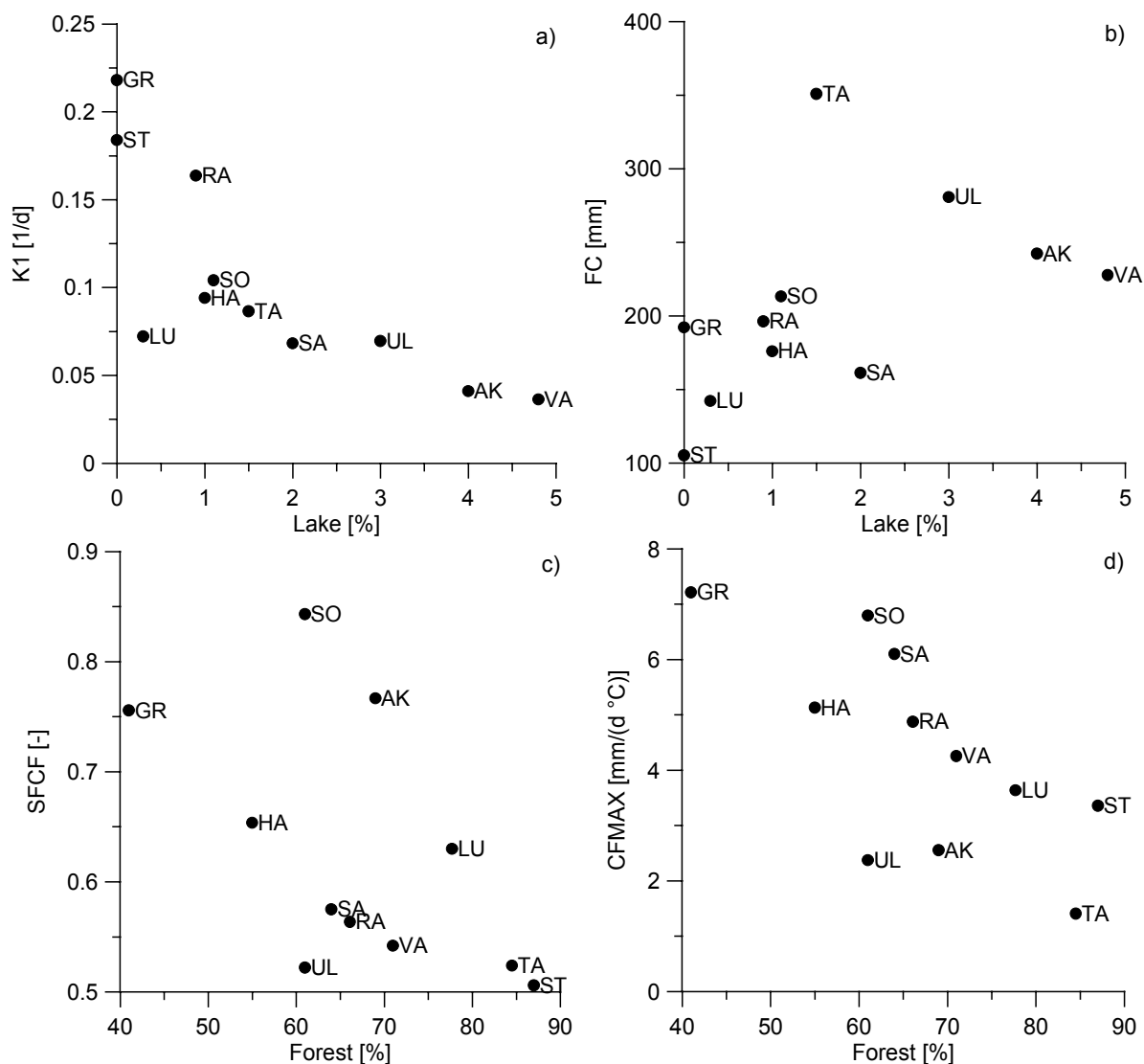


Figure 10. Relationships between catchment characteristics and model parameter values of the HBV model, calibrated with a combination of three goodness-of-fit measures (efficiency, log-efficiency and volume error) against observed runoff for the period 1981-1990. The abbreviations refer to the eleven different catchments, which are located within the NOPEX region. K1 is a recession coefficient (upper box of response function), FC is the maximal storage in soil box, SFCF is the correction factor for snowfall, and CFMAX is the degree-day factor. (from Seibert, 1999)

Discussion

Parsimony and complexity

A central question in runoff modelling is to find a middle course between parsimonious and complex alternatives in model development. Unique parameter values could not be derived for the HBV model (paper I, Uhlenbrook *et al.*, 1999). This agrees with the commonly accepted fact that a small number of parameters are sufficient to simulate runoff and that only few parameters can be identified from the information contained in the precipitation-runoff relationship (*e.g.*, Beven, 1989; Jakeman and Hornberger, 1993; Gaume *et al.*, 1998). In a parsimonious model with only few parameters, each parameter represents a conglomerate of catchment properties and can, thus, not be determined from measurable physical properties. A direct relationship between measurable physical properties and model parameters can – if at all - only be expected if an ample number of various parameters is used.

Parsimonious models may allow identifying unique parameter values, but extrapolation beyond the conditions used for calibration may be less reliable for such models than for complex models. This dilemma has been formulated aptly by Kuczera and Mroczkowski (1998, p. 1482): “A simple model cannot be relied upon to make meaningful extrapolative predictions, whereas a complex model may have the potential but because of information constraints may be unable to realize it.”

Additional information may help improving the identifiability of parameter values as demonstrated in paper VI for the use of groundwater level observations. On the other hand, incorporation of additional variables used for calibration and validation often requires extending the model and the number of parameters may increase faster than the amount of additional information.

The testability of models increases with increasing model complexity. It is difficult to test very parsimonious models, *e.g.*, monthly models with 3-5 parameters, against other variables than runoff since measurable quantities have no clear counterparts in the model. Lumped models can only be compared to averaged values whereas (semi-)distributed models allow a more direct testing against measurements. This is also one reason why it was easier to find inaccuracies in simulations of TOPMODEL than in those of the HBV model. Contrary to TOPMODEL, the HBV model does not claim the prediction of spatially variable groundwater levels but simulates some catchment-wide mean groundwater level, which is obviously a crude representation of the real processes. This difference can be seen from two points of view: (1) With TOPMODEL it is at least attempted to simulate groundwater levels in a more realistic way; (2) given the poor results of the distributed simulations of TOPMODEL it would be more honest to abstain from predicting local groundwater levels.

Limitations of conceptual models

There are two main sources of errors when using a model to simulate runoff from an ungauged catchment. The model structure is incorrect to some degree, *i.e.*, the fit would not be perfect even with optimisation. The parameter values, which are estimated from regionalisation equations (or measurements), differ from those that would give the optimal fit. When transferring series of specific runoff directly without using any model, the variations in both catchment characteristics and climatic variables are ignored. The use of a model will provide more accurate results if it is easy to estimate parameter values and if climatic variables differ. The mountainous catchments in the Black Forest (Seibert *et al.*, 1999) were similar but both precipitation and temperature varied significantly. The catchments within and close to the NOPEX region differed more from each other in their characteristics and less in the climatic variables. Here the use of the model was less superior (Seibert, 1999).

Both TOPMODEL and the HBV model performed well in terms of simulated runoff after calibration to runoff. The models were also able to simulate groundwater levels with reasonable accuracy if groundwater levels were included in the calibration (paper II and VI). On the other hand, groundwater levels and other variables were simulated with poor results when they had not been used for calibration (papers II, IV, VI). This demonstrated that an acceptable fit of simulated and observed catchment runoff does not ensure internal consistency. Simulations of variables, which have not been looked at during calibration, should in general be interpreted with great care.

The multi-criteria validation of TOPMODEL indicated that its concept of runoff generation was not appropriate for the Brugga catchment and it might be argued that it was wrong to apply the model in this catchment. Prior to the application and testing of TOPMODEL several prerequisites for its use were supposed to be met in the study area, which is similar to other regions where model applications have been considered as successful (*e.g.*, the Vosges, France, Ambroise *et al.*, 1996).

Conceptual runoff models are crude simplifications and it was not surprising that the validation of TOPMODEL against hydrological common sense revealed unrealistic assumptions. Two points are important to note: (1) The name TOP(ography-based)MODEL (Beven *et al.*, 1984) can be deceptive. Much of the information given by DEMs cannot be utilised by the TOPMODEL approach. The influence of topography to the temporal variations in the response of groundwater levels is neglected because of the steady-state assumption, and as a result of the assumption of a spatially uniform recharge it is not possible to use an atmospheric forcing which depends on topography. (2) The limitation of assuming a spatially uniform recharge has been ignored in many TOPMODEL applications. The use of local recharge as a function of groundwater depth is misleading and causes implicitly an unrealistic redistribution of water. Moreover, much of the point in using TOPMODEL to simulate patterns of evaporation (Famiglietti and Wood, 1994a; Quinn and Beven, 1993) is missed, because the interdependence of hydrological

fluxes in vertical and horizontal directions is not captured. As a result of the spatially-uniform-recharge assumption, simulated spatial patterns of evaporation are dependent on spatial patterns of groundwater levels but not *vice versa*.

If one is interested in simulating only runoff, there is no problem to do this on large scales. It is often easier to obtain good fits for large basin since different errors cancel each other (Bergström and Graham, 1998). For the same reason, a large-scale model may be a poor representation of reality. It works even if it does so for the wrong reasons. Observable quantities such as soil moisture and their counterparts in the model are not equivalent. The simulations of internal variables must be considered as extremely uncertain. This has also to be taken into account when large-scale, conceptual runoff models are used within GCMs (*e.g.*, Dümenil and Todini, 1992): simulated quantities are uncertain and the validity of the model and its parameter values is questionable for changed climatic conditions.

Problems of validation

Validation against spatially distributed measurements of hydrological variables is important for relevant model testing. Because of the availability of suitable data, this type of validation is mainly restricted to small experimental studies (*e.g.*, paper **II**). Other, *e.g.* remotely sensed, information can be used to test spatially distributed simulations for larger areas in some cases. The use of mires to test the TOPMODEL index (paper **III**) illustrates such a test, but it also exemplifies that 'indirect' tests are associated with problems. The definition of mires in commercially available maps is not clear. Furthermore, a fundamental objection could be raised against the use of mires as field data. The development of many mires is connected to downslope damming resulting in flat ground surfaces. Once the development of a mire has been initiated, local TOPMODEL index values may change because the mire modifies its ground surface by biological activity.

Additional data, which may be used to test a model, are often measured on scales much smaller than the modelling scale. Therefore, up- or downscaling (Blöschl and Sivapalan, 1995) is required before simulations can be compared with these data. This may cause new sources of errors and uncertainties.

Soil moisture is a central variable in runoff modelling as it influences both the amount of runoff caused by a rainfall or snowmelt event and the reduction of the potential evaporation. In most conceptual models there is some representation of soil moisture, but validation against field data is difficult because of at least two problems: soil moisture is very variable in space and volumetric water content is seldom directly comparable with its counterpart in a model. Currently available routine measurement techniques for soil moisture measure only small volumes of soil. Areal values have to be estimated from these point measurements. Recent studies indicate that it might be possible to obtain reliable areal estimates of soil moisture from a limited number of point measurements, if the locations of these are chosen thoughtfully (Grayson and Western, 1998).

Environmental tracers allow data to be obtained at the same scale as runoff is measured. The interpretation of the measurements by tracer-based hydrograph separation may be affected by additional uncertainties (Rodhe, 1987; Bazemore *et al.*, 1994; Genereux, 1998; Rice and Hornberger, 1998). These uncertainties can be related to assumptions made for the hydrograph separation (*e.g.*, that each source of runoff has a uniform concentration of the tracer), sampling errors and too similar concentrations of the tracers for different runoff components. Furthermore, runoff components of a model and those determined by hydrograph separation may not be directly equivalent (paper IV, Holko and Lepistö, 1997).

Physically-based models

The idea of physically-based, fully distributed models has been proposed already by Freeze and Harlan (1969) and has been realised for operational use, for instance, in the SHE model (Abbott *et al.*, 1986a,b). It might be assumed to be a better approximation of reality than conceptual models. Yet there has been much debate on whether physically-based, distributed catchment modelling is feasible (Beven, 1987; 1989; 1996a,b,c; Bathurst and O'Connell, 1992; Jensen and Mantonoglou, 1992; Grayson *et al.*, 1992, 1994; Jakeman and Hornberger, 1993; Smith *et al.*, 1994; Barnes, 1995; Refsgaard *et al.*, 1996; Bronstert, 1999).

At least for snowmelt modelling, one might expect physically-based modelling to be straightforward simply using the equations of energy and mass balance. However, it is commonly accepted that simple degree-day methods are equally, or even more, suitable for simulations at the catchment scale. There are mainly three reasons for this: (1) The processes in a melting snowpack are more complex than expected at first glance, *e.g.*, different layers are needed to represent percolation of meltwater through the snowpack; (2) at the catchment scale neither parameter values nor climatic input data can be estimated with acceptable accuracy; (3) the computational burden of detailed snowmelt models is high even for the computers available nowadays.

There has been much research on detailed snowmelt models (Tuteja and Cunnane, 1997). Looking at these studies, the list of processes, which have to be included to achieve a 'complete' snowmelt model, seems overwhelming. One might, for instance, include sub-surface melting caused by penetration of solar radiation into the snow cover (Koh and Jordan, 1995). Tuteja and Cunnane (1997) presented a model for the transport of mass and energy into the snowpack. They demonstrate the model to represent the key processes on the point scale, but it seems impossible to apply the model on larger scales mainly because of the amount of data required for running and calibrating the model. It should be noted that, although the model is very detailed, it still contains empirical equations with non-measurable parameters (*e.g.*, for the compaction of the snowpack).

When using a degree-day method to compute melt rates it is assumed that the air temperature is always proportional to the energy available for melt. This can not be expected if snowmelt occurs during different seasons or for locations with different

exposure within a catchment. Therefore, it might be argued that energy-balance methods should be included into catchment models. For conceptual models, the approach of improving degree-day methods with a melt factor varying with season (Braun and Renner, 1992) or depending on topography (Hottelet *et al.*, 1994; Cazorzi and Dalla Fontana, 1996) seems to be a convenient compromise between incorporating physical reality, data availability and computational burden.

To conclude, physically-based representations do not seem to be feasible for the problem of snowmelt modelling at the catchment scale. Simplified energy-balance methods may be used, but these can hardly be classified as more 'correct' than modifications of the simple degree-day methods. The determination of the spatial distribution of input variables and parameter values is problematic and internal processes may be represented very poorly, even in a model based on the energy balance, when the snowpack is represented by a single layer (Blöschl and Kirnbauer, 1991).

One advantage claimed for physically-based, distributed models is that their parameters have a direct physical meaning, *e.g.* the hydraulic conductivity, and are, thus, measurable (Bathurst and O'Connell, 1992). It must be noted that these models work with spatial resolutions much larger than the measurement scale of most parameters (Beven, 1989). Natural heterogeneity causes large subgrid variability for many parameters, and effective values, which are not directly measurable, have to be used. The use of effective parameter values limits the physical basis of the underlying equations. Furthermore, in many cases it is impossible to reproduce the effects of spatial variability with a single effective parameter (Binley *et al.*, 1989).

Measurement of most parameters is restricted in practice to a few points in a catchment. Soil parameters can only be estimated from measurements at a limited number of locations. For modelling the interception of rain (*e.g.*, Lankreijer *et al.*, 1999) and of snow (*e.g.*, Lundberg *et al.*, 1998), parameter values can not be measured directly even at the plot scale and detailed measurements and analyses are needed to derive them. Accurate estimation of the spatial distribution of parameter values seems impossible in such cases.

Even complex, physically-based models such as the SHE model require calibration in most cases (*e.g.*, Refsgaard and Knudsen, 1996; Jayatilaka *et al.*, 1998).

Mroczkowski *et al.* (1997) define a model as conceptual if at least one of its parameters has to be calibrated. It may not be helpful to use such an extreme definition, which applies to almost all models, but their point of view illustrate that the difference between conceptual and physically-based models is of degree rather than of kind.

Future directions

Attempts to improve conceptual runoff models often result in frustrating conclusions. Although a model has been modified towards a better description of the real processes, the quality of runoff simulations does not increase significantly. For the

HBV model, such an experience has been reported for tests of different formulations of evaporation and snowmelt (Andersson, 1992), the insertion of evaporation depending with altitude (Evremar, 1994) and the use of an explicit interception routine (Lindström *et al.*, 1997). Similar results were obtained by Uhlenbrook *et al.* (1999) who tested variants of the HBV model with different numbers of elevation and land use zones and various runoff-generation conceptualisations. Good results in terms of runoff simulations could be obtained with different and even unrealistic concepts. The difference in numerical measures such as R_{eff} was usually small even in cases where the model performance increased after modifications (Bergström and Lindström, 1992; Bergström *et al.*, 1997; Lindström *et al.*, 1997; Uhlenbrook *et al.*, 1999). These unsatisfactory results can partly be attributed to the way of model evaluation. The model efficiency, which has been used to assess model goodness in most of the mentioned studies, is not sensitive to improvements of runoff during low flow conditions and, more important, the improvements may vanish in the simulated runoff, although they may be significant if looking on internal variables.

Progress in remote sensing may provide new ways to parameterise and to validate models (Rango, 1994; Burke *et al.*, 1997; Franks *et al.*, 1998). Land-use classifications for modelling are often derived from satellite data (*e.g.*, Kite, 1991). The spatial distribution of some variables can be estimated (or this will probably be possible in the near future) with remote-sensing techniques (*e.g.*, snow cover and snow-water equivalent, extension of surface-saturated areas, soil moisture in areas with no or only little vegetation). Remotely-sensed quantities may be used as proxies for other variables. Mauser and Schädlich (1998), for instance, validate an evaporation model against satellite-measured patterns of surface temperature. Combining DEMs and remotely-sensed data may further increase the total information extractable from the two types of data (Florinsky, 1998). Difficulties in the use of remotely-sensed data are the limited availability, the costs and the large uncertainties. Furthermore, even with an optimistic point of view, important variables such as soil moisture in forested areas or groundwater levels are not assumed to be accessible by remote-sensing techniques in a foreseeable future.

The increasing computer power may be utilised in different ways. It may allow refining the resolution of distributed models or to include additional process representations and to enlarge the model complexity, *i.e.*, executing more calculations per model run. On the other hand, it may be used to address the model uncertainty by Monte-Carlo procedures (*e.g.*, the GLUE methodology, Beven and Binley, 1992), *i.e.*, performing more model runs. Both ways are reasonable, but a quantification of the prediction uncertainties is of central importance, especially in practical applications, given the large uncertainties associated with the use of runoff models. For the HBV model, Langsrud *et al.* (1998a,b) recently proposed methods for quantification of the uncertainty in runoff forecasts in operational use. Increasing model complexity should mainly aim at improving model testability. It is of limited value to extend a model with routines that cannot be tested against available data.

Failings in modelling studies

Model validation, as it has been defined in this thesis, is an undertaking that can not, and should not, be carried out by a single researcher or research group, but requires a scientific dialogue. Improper model applications and ambiguously presented results sometimes impede this dialogue. Examples of pitfalls of hydrological modelling are discussed in the following section. The aim is to help avoiding these pitfalls and to reduce confusion in hydrological modelling.

Presentation of models and model results

Researchers, and modellers in particular, are human beings with all their weaknesses. They often prefer to build new models and to modify them, than to test the capabilities and limitations of existing models. They strive for ‘good’ results, while negative results are assumed to be of less value. These attitudes certainly are wrong as pointed out by Bergström (1991), but they can still be recognised in the research on hydrological modelling.

Modellers are often much better in claiming what models are capable of, than in telling about what can not be done with the models. TOPMODEL is a good example of such biased information, where limitations are often not stated clearly enough or even not at all. There seems to be a need for such statements since TOPMODEL’s capabilities are often overrated in the literature. White and Running (1994) coupled TOPMODEL with an ecological model and call TOPMODEL a “complex process model” (p. 701), while it is in fact a simple conceptual model. Hinton *et al.* (1993) describe results of measurements which provide examples where the TOPMODEL assumptions are not fulfilled. Nevertheless, they concluded that modelling “... the effect of such spatial differences in hydrological processes [...] would require distributed models such as TOPMODEL ...” (p. 246).

Various methods can be used to present misleading graphs comparing observations and simulation results: plotting cumulative values (*e.g.*, Viney and Sivapalan, 1996), using log-scale (*e.g.*, Stagnitti *et al.*, 1992) or squeezing long series into a graph. Each of these ‘techniques’ may be justified in special situations, but in general they should be avoided.

Model efficiency

Nash and Sutcliffe (1970) defined the efficiency of a model (Tab. 1) “analogous to the coefficient of determination” (p. 288) and used the notation R^2 . The model efficiency has become one of the most widely used goodness-of-fit measures in hydrological modelling. It is dimensionless and provides a quick impression about the goodness-of-fit. On the other hand, the notation R^2 has caused confusion.

Different terms are used for the (model) efficiency, *e.g.*, coefficient of efficiency, Nash-Sutcliffe criteria, Nash criteria, determination coefficient (Franchini *et al.*, 1996) or explained variance. Various notations are used: R^2 (Nash and Sutcliffe, 1970), CE (Kuczera, 1983), E (Beven *et al.*, 1984), F^2 (Kite, 1991), E_m (Moore, 1993), EQ (Ambroise *et al.*, 1995), DC (Franchini *et al.*, 1996), r^2 (Freer *et al.*,

1996), NS (Gupta *et al.*, 1998), NSC (Mohseni and Stefan, 1998), E^2 (Krysanova *et al.*, 1998), R_{eff} (this thesis). All notations including a power two are mathematically improper because the model efficiency can take negative values but the model goodness is hardly a complex number.

The choice of R^2 as notation for the model efficiency by Nash and Sutcliffe (1970) causes confusion with the coefficient of determination not only for students but also for scientists. Pietroniro *et al.* (1996), for instance, wrote in the abstract that the calibration "... yielded an r^2 value of 0.86", which most readers will understand as value of the coefficient of determination unless they find the line some pages later where r^2 is called Nash-Sutcliffe criteria. Another example is found in the paper of Ye *et al.* (1998), who state erroneously that the "... Nash-Sutcliffe coefficient R^2 [is] termed the Coefficient of Determination in more standard statistical contexts" (p. 67).

The true coefficient of determination (r^2 , Tab. 1) is unsuitable to evaluate model performance; good agreements between simulations and observations give rise to high values of r^2 , but high values do not ensure a good agreement. The coefficient of determination has been used to assess the goodness of fit in some studies (*e.g.*, Wigmosta, 1994; Flügel, 1995; Suyanto *et al.*, 1995). It is conceivable that some of these authors computed the model efficiency and just called it coefficient of determination. Flügel (1995), for instance, applied the PRMS runoff model (Leavesley *et al.*, 1983) and gives the model goodness as 'correlation coefficient'. On the other hand, the PRMS software package provides the model efficiency (even if called coefficient of determination by Leavesley *et al.*, 1983). Here the reader is left in the dark about which criterion has been used and whether it has been calculated correctly.

A generally accepted goodness-of-fit measure is definitely needed in hydrological modelling. Figures intended to show how well simulations agree with observations often only provide limited information because long runoff series are squeezed in and lines for observed and simulated runoff are not easily distinguishable. Not all authors provide numerical information, but only state that the model was in 'good agreement' with the observations (*e.g.*, Buchtele *et al.*, 1996; Bergström and Graham, 1998, Gellens and Roulin, 1998).

Even if a measure of goodness is given, it does not always provide the relevant information. Panagoulia and Dimou (1997) studied the sensitivity of flood events to climate change using a conceptual model with a daily time step. The only measures of goodness of the calibration found in the paper are relative differences between observations and simulations based on monthly sums. This information is insufficient to assess the validity of the model for the purpose to study flood events.

The study of Mohseni and Stefan (1998) is an example of another type of misleading use of goodness-of-fit criteria. They tested a monthly streamflow model and computed a 'mean monthly efficiency' from the long-term mean values of both simulated and observed runoff for each month of the year. They emphasise

the high values (0.87 to 0.99) of this model evaluation of long-term seasonal runoff variations, but do scarcely discuss the much lower efficiency when computed directly from the time series (0.41 to 0.81), *i.e.*, the fact that the errors simply cancel each other when averaging over several years.

Insufficient model testing

The most widespread type of model validation is the simulation and comparison with observed runoff for a period that has not been used for calibration (split-sample test). Examples where the result of such a test is called ‘not successful’ are seldom found in literature. This may be because this kind of validation is a simple task (Kirchner *et al.*, 1996) or because of a tendency not to publish negative results. An exception is Piotroniro *et al.* (1996) where very poor validation results (negative model efficiency) are frankly reported and discussed. In the study of Mohseni and Stefan (1998), the efficiency of a monthly streamflow model dropped from 0.68 for the calibration period to 0.41 for the validation period. They partly explain this drop by some measurement errors and state that “the results show that the model is valid” (p. 1294). In such a case one might ask how poor model results are required to invalidate a model.

When a model shall be used to predict effects of a climate change, tests should be performed to assert the reliability of a model for nonstationary conditions, *e.g.*, a (proxy-basin) differential split-sample test (Xu, 1999a, b), but such tests are not commonly performed. Viney and Sivapalan (1996) or Panagoulia and Dimou (1997), for instance, predict the response of runoff to climatic changes without any kind of validation of their calibrated models.

Many modelling studies suffer from the lack of adequate data. When tested only against runoff at the catchment outlet, *i.e.*, ‘lumped’ data that integrate over the catchment, distributed models can seldom be demonstrated to be superior to lumped or semi-distributed models. Whenever a model is aimed at simulating more than just runoff its capability to do so should be demonstrated. Much too often powerful tests are not performed ‘because there was no data available’ (*e.g.*, Stagnitti *et al.*, 1992; Yao *et al.*, 1996; Krysanova *et al.*, 1998). The development of complex models based on limited data is likely to give misleading results (Pilgrim, 1986).

Another problem is the drawing of unfounded conclusions from testing a model. For example, Ponce *et al.* (1996) conclude that “the close agreement between analytical and numerical results underscores the utility of Muskingum-Cunge routing as a viable and accurate method for routine applications in flood hydrology.” (p. 235). What they showed is that the simpler Muskingum-Cunge routing agrees for a number of numerical tests with an analytical model and, thus, may be used instead. They did not compare any of the two models with observed data.

Conclusions

- Parameter uncertainty is a significant source of uncertainty in model predictions. Predictions should be given as ranges, which can be computed using Monte-Carlo-based methods, rather than as single values.
- Model predictions are of limited practical use without clear information about their reliability and accuracy.
- A conceptual runoff model that has only been calibrated against runoff does not provide reliable simulations of internal variables such as groundwater levels.
- Multi-criteria calibration can help to reduce parameter uncertainty.
- Powerful validation is essential for further development of a model for two reasons: identification of weak parts and evaluation of improvements.

Finally, the answer to the question posed in the title is that conceptual runoff models are rather fiction than a representation of reality. Models may provide good and useful fiction. Similar to literary fiction, the primary value of models may be their use as an intellectual tool, which helps to understand and reflect on reality. By this, models support experts to make estimates about the future, but models alone can not provide these estimates.

References

- Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E. and Rasmussen, J., 1986a. An introduction to the European Hydrological System - Système Hydrologique Européen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system. *Journal of Hydrology* 87: 45-59
- Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E. and Rasmussen, J., 1986b. An introduction to the European Hydrological System - Système Hydrologique Européen, "SHE", 2: Structure of a physically-based, distributed modelling system. *Journal of Hydrology* 87: 61-77
- Allen, R.E. (ed.), 1990. *The Concise Oxford Dictionary of Current English*, 8th ed.. Oxford University Press, New York, United States, 1454 pp.
- Ambroise, B., Perrin, J.L. and Reutenauer, D., 1995. Multicriterion validation of a semidistributed conceptual model of the water cycle in the Fecht Catchment (Vosges Massif, France). *Water Resources Research* 31(6): 1467-1481
- Ambroise, B., Freer, J. and Beven, K.J., 1996. Application of a generalized TOPMODEL to the small Ringelbach catchment, Vosges, France. *Water Resources Research* 32(7): 2147-2159
- Andersson, L., 1992. Improvements of runoff models - what way to go? *Nordic Hydrology* 23: 315-332
- Andersson, I., Bishop, K.H., Borg, G.Chr., Giesler, R., Hultberg, H., Huse, M., Moldan, F., Nyberg, L., Nygaard, P. H. and Nyström, U., 1998. The covered catchment site: a description of the physiography, climate and vegetation of three small coniferous forest catchments at Gårdsjön, South-west Sweden. In: H. Hultberg and R. Skeffington (eds.), *Experimental Reversal of Acid Rain Effects: The Gårdsjön Roof Project*, Wiley and Sons, London, Great Britain, pp. 25-70
- Arheimer, B. and Brandt, M., 1998. Modelling nitrogen transport and retention in the catchments of southern Sweden. *Ambio* 27(6): 471-480
- Bair, E.S., 1994. Model (in)validation - a view from the courtroom (editorial). *Ground Water* 32(4): 530-531
- Band, L.E., 1993. Effect of land surface representation on forest water and carbon budgets. *Journal of Hydrology* 150: 749-772
- Band, L.E., Patterson, P., Nemani, R. and Running, S.W., 1993. Forest ecosystem processes at the watershed scale: incorporating hillslope hydrology. *Agricultural and Forest Meteorology* 63: 93-126
- Barnes, C.J., 1995. Commentary: The art of catchment modeling: what is a good model? *Environment International* 21(5): 747-751

- Bathurst, J.C. and Cooley, K.R., 1996. Use of the SHE hydrological modelling system to investigate basin response to snowmelt at Reynolds Creek, Idaho. *Journal of Hydrology* 175: 181-211
- Bathurst, J.C. and O'Connell, P.E., 1992. The future of distributed modelling: the Système Hydrologique Européen. *Hydrological Processes* 6: 265-277
- Bazemore, D.E., Eshleman, K.N. and Hollenbeck, K.J., 1994. The role of soil water in stormflow generation in a forested headwater catchment: synthesis of natural tracer and hydrometric evidence. *Journal of Hydrology* 162: 47-75
- Beck, M.B., Ravetz, J.R., Mulkey, L.A. and Barnwell, T.O., 1997. On the problem of model validation for predictive exposure assessments. *Stochastic Hydrology and Hydraulics* 11: 229-254
- Bergqvist, E., 1971. *Nåsten and Marsta - two small drainage basins in central Sweden, 1. Purpose of the investigation and description of the basins*. UNGI Report 5, Uppsala Univ., Dept. Phys. Geogr. (In Swedish with English summary and figure captions.), 49 pp.
- Bergström, S., 1976. *Development and application of a conceptual runoff model for Scandinavian catchments*. SMHI, RHO No. 7, Norrköping, 134 pp.
- Bergström, S., 1990. *Parametervärden för HBV-modellen i Sverige, Erfarenheter från modelkalibreringar under perioden 1975-1989* (in Swedish). SMHI, Nr. 28, Norrköping, 35 pp.
- Bergström, S., 1991. Principles and confidence in hydrological modelling. *Nordic Hydrology* 22: 123-136
- Bergström, S., 1992. *The HBV model - its structure and applications*. SMHI, RH No. 4, Norrköping, 35 pp.
- Bergström, S., 1995. The HBV model. In: V.P. Singh (ed.) *Computer models of watershed hydrology*. Water Resources Publications, Highlands Ranch, Colorado, U.S.A., 443-476
- Bergström, S. and Graham, L.P., 1998. On the scale problem in hydrological modelling. *Journal of Hydrology* 211: 253-265
- Bergström, S. and Lindström, G., 1992. Recharge and discharge areas in hydrological modelling - a new model approach. *Vannet i Norden* 3: 5-12
- Bergström, S. and Sandberg, G., 1983. Simulation of groundwater response by conceptual models - three case studies. *Nordic Hydrology* 14: 85-92
- Bergström, S., Carlsson, B., Sandberg, G. and Maxe, L., 1985. Integrated modelling of runoff, alkalinity, and pH on a daily basis. *Nordic Hydrology* 16: 89-104
- Bergström, S., Harlin, J. and Lindström, G., 1992. Spillway design floods in Sweden: I. New guidelines. *Hydrological Sciences Journal* 37: 505-519
- Bergström, S., Carlsson, B., Grahn, G. and Johansson, B., 1997. A more consistent approach to catchment response in the HBV model. *Vannet i Norden* 4: 6-13
- Beven, K.J., 1987. Towards a new paradigm in hydrology. In: *Water for the future: Hydrology in Perspective (Proceedings of a symposium held at Rome, April 1987)* (ed. by J.C. Rodda and N.C. Matalas), IAHS Publication 164: 393-403
- Beven, K.J., 1989. Changing ideas in hydrology - the case of physically-based models. *Journal of Hydrology* 105: 157-172
- Beven, K.J., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources* 16: 41-51
- Beven, K.J., 1996a. A discussion of distributed hydrological modelling (chapter 13A). In: M.B. Abbott and J.C. Refsgaard (eds.), *Distributed hydrological modelling*, Water Science and Technology Library, Vol.22, Kluwer Academic Publishers, Dordrecht, The Netherlands, 321 pp.,: 255-278
- Beven, K.J., 1996b. Response to comments on 'A discussion of distributed hydrological modelling' by J C Refsgaard et al. (chapter 13C). In: M.B. Abbott and J.C. Refsgaard (eds.), *Distributed hydrological modelling*, Water Science and Technology Library, Vol.22, Kluwer Academic Publishers, Dordrecht, The Netherlands, 321 pp.,: 289-295
- Beven, K.J., 1996c. The limits of splitting: Hydrology. *The Science of the Total Environment* 183: 89-97
- Beven, K.J. and Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* 6: 279-298
- Beven, K.J. and Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal* 24: 43-69
- Beven, K.J., Kirkby, M.J., Schofield, N. and Tagg, A.F., 1984. Testing a physically based flood forecasting model (TOPMODEL) for three U.K. catchments. *Journal of Hydrology* 69: 119-143
- Beven, K.J., Lamb, R., Quinn, P., Romanowicz, R. and Freer, J. 1995. TOPMODEL. In: V.P. Singh (ed.), *Computer Models of Watershed Hydrology*, Water Resources Publications, Highlands Ranch, Colorado, pp. 627-668
- Binley, A., Beven, K.J. and Elgy, J., 1989. A physically based model of heterogeneous hillslopes, 2. Effective hydraulic conductivities. *Water Resources Research* 25(6): 1219-1226

- Bishop, K.H. and Hultberg, H., 1995. Reversing acidification in a forest ecosystem: the Gårdsjön covered catchment experiment. *Ambio* 24(2): 85-91
- Blöschl, G., 1991. The influence of uncertainty in air temperature and albedo on snowmelt. *Nordic Hydrology* 22: 95-108
- Blöschl, G. and Kirnbauer, R., 1991. Point snowmelt models with different degrees of complexity - internal processes. *Journal of Hydrology* 129: 127-147
- Blöschl, G. and Sivapalan, M., 1995. Scale issues in hydrological modelling: a review. *Hydrological Processes* 9: 251-290
- Brandt, M., Bergström, S. and Gardelin, M., 1988. Modelling the effects of clearcutting on runoff - examples from Central Sweden. *Ambio* 17(5): 307-313
- Braun, L. and Renner, C.B., 1992. Application of a conceptual runoff model in different physiographic regions of Switzerland. *Hydrological Sciences Journal* 37: 217-231
- Bredehoeff, J.D. and Konikow, L.F., 1993. Ground-water models: validate or invalidate (editorial). *Ground Water* 31(2): 178-179
- Bronstert, A., 1999. Capabilities and limitations of physically based hydrological modelling on the hillslope scale. *Hydrological Processes* (in press)
- Brown, T.N. and Kulasiri, D., 1996. Validating models of complex, stochastic, biological systems. *Ecological Modelling* 86: 129-134
- Buchtele, J., Elias, V., Tesar, M. and Herrmann, A., 1996. Runoff components simulated by rainfall-runoff models. *Hydrological Sciences Journal* 41: 49-60
- Burke, E.J., Banks, A.C. and Gurney, R.J., 1997. Remote sensing of soil-vegetation-atmosphere transfer processes (progress report). *Progress in Physical Geography* 21(4): 549-572
- Calder, R.I., Hall, R.L., Bastable, H.G., Gunston, H.M., Shela, O., Chirwa, A. and Kafundu, R., 1995. The impact of land use change on water resources in sub-Saharan Africa: a modelling study of Lake Malawi. *Journal of Hydrology* 170: 123-135
- Caspary, H.J., 1990. An ecohydrological framework for water yield changes of forested catchments due to forest decline and soil acidification. *Water Resources Research* 26(6): 1121-1131
- Cazorzi, F. and Dalla Fontana, G., 1996. Snowmelt modelling by combining air temperature and a distributed radiation index. *Journal of Hydrology* 181: 169-187
- Cirno, C.P. and McDonnell, J.J., 1997. Linking the hydrologic and biochemical controls of nitrogen transport in near-stream zones of temperate-forested catchments: a review. *Journal of Hydrology* 199: 88-120
- de Groisbois, E., Hooper, R.P. and Christophersen, N., 1988. A multisignal automatic calibration methodology for hydrochemical models: a case study of the Birkenes model. *Water Resources Research* 24(8): 1299-1307
- Dümenil, L. and Todini, E., 1992. A rainfall-runoff scheme for use in the Hamburg climate model (chapter 9). In J.P. O'Kane (ed.) *Advances in Theoretical Hydrology, A Tribute to James Dooge*, European Geophysical Society Series on Hydrological Sciences, 1, Elsevier, Amsterdam, The Netherlands, pp. 129-157
- Dunn, S.M. and Mackay, R., 1995. Spatial variations in evapotranspiration and the influence of land use on catchment hydrology. *Journal of Hydrology* 171: 49-73
- Eeles, C.W.O. and Blackie, J.R., 1993. Land-use changes in the Balquhider catchments simulated by a daily streamflow model. *Journal of Hydrology* 145: 315-336
- Evans, R., 1996. Some soil factors influencing accelerated water erosion of arable land (progress report). *Progress in Physical Geography* 20(2): 205-215
- Evremar, Å., 1994. *Avdunstningens höjdberoende i svenska fjällområden bestämd ur vattenbalans och med modellering* (in Swedish with English abstract, English title: The altitude dependence of the evapotranspiration in Swedish mountains calculated from water balance and modelling). Thesis paper (examensarbete 20p), Uppsala University, Department of Earth Sciences, Hydrology, Uppsala, 40 pp.
- Ewen, J. and Parkin, G., 1996. Validation of catchment models for predicting land-use and climate change impacts. 1. Method. *Journal of Hydrology* 175: 583-594
- Famiglietti, J. S. and Wood, E.F., 1994a. Application of multiscale water and energy balance models on a tallgrass prairie. *Water Resources Research* 30(11): 3079-3093
- Famiglietti, J. S. and Wood, E.F., 1994b. Multiscale modeling of spatially variable water and energy balance processes. *Water Resources Research* 30(11): 3061-3078
- Florinsky, I.V., 1998. Combined analysis of digital terrain models and remotely sensed data in landscape investigations. *Progress in Physical Geography* 22(1): 33-60
- Flügel, W.-A., 1995. Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. *Hydrological Processes* 9: 423-436

- Franchini, M., Wendling, J., Obled, Ch. and Todini, E., 1996. Physical interpretation and sensitivity analysis of the TOPMODEL. *Journal of Hydrology* 175: 293-338
- Franks, S., Gineste, Ph., Beven, K.J. and Merot, Ph., 1998. On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process. *Water Resources Research* 34(4): 787-797
- Freer, J., Beven, K.J. and Ambrose, B., 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. *Water Resources Research* 32(7): 2161-2173
- Freeze, R.A. and Harlan, R.L., 1969. Blueprint for a physically-based, digitally-simulated hydrologic response model. *Journal of Hydrology* 9: 237-258
- Gaas, S.I., 1983. Decision-aiding models: validation, assessment, and related issues in policy analysis, *Operations Research* 31: 603-631
- Gabbard, D.S., Huang, C., Norton, L.D. and Steinhardt, G.C., 1998. Landscape position, surface hydraulic gradients and erosion processes. *Earth Surface Processes and Landforms* 23: 83-93
- Gaume, E., Villeneuve, J.-P. and Desbordes, M., 1998. Uncertainty assessment and analysis of the calibrated parameter values of an urban storm water quality model. *Journal of Hydrology* 210: 38-50
- Gellens, D. and Roulin, E., 1998. Streamflow response of Belgian catchments to IPCC climate change scenarios. *Journal of Hydrology* 210: 242-258
- Genereux, D.P., 1998. Quantifying uncertainty in tracer-based hydrograph separation. *Water Resources Research* 34(4): 915-919
- Gleick, P.H., 1987. Regional hydrologic consequences of increases in atmospheric CO₂ and other trace gases. *Climatic Change* 10: 137-161
- Grayson, R.B. and Western, A.W., 1998. Towards areal estimation of soil water content from point measurements: time and space stability of mean response. *Journal of Hydrology* 207: 68-82
- Grayson, R.B., Moore, I.D. and McMahon, T.A., 1992. Physically based hydrologic modeling, 2. Is the concept realistic? *Water Resources Research* 26(10): 2659-2666
- Grayson, R.B., Moore, I.D. and McMahon, T.A., 1994. Reply [to comment on "Physically based hydrologic modeling, 2, Is the concept realistic?" by R.B. Grayson, I.D. Moore, and T.A. McMahon]. *Water Resources Research* 30(3): 855-856
- Green, I.R.A. and Stephenson, D., 1986. Criteria for comparison of single event models. *Hydrological Sciences Journal* 31: 395-411
- Grimm, V., 1994. Mathematical models and understanding in ecology. *Ecological Modelling* 75/76: 641-651
- Güntner, A., Uhlenbrook, S., Seibert, J. and Leibundgut, Ch., 1999. Estimation of saturation excess overland flow areas - comparison of topographic index calculations with field mapping, In *Regionalization in Hydrology (Proc. Conf. at Braunschweig, March 1997)* (ed. by B. Dickkrüger, M.J. Kirkby and U. Schröder), *IAHS Publication* 254 (in press)
- Gupta, V.K., Sorooshian, S. and Yapo, P.O., 1998. Toward improved calibration of hydrological models: Multiple and noncommensurable measures of information. *Water Resources Research* 34(4): 751-763
- Halldin, S., Gottschalk, L., van de Griend, A.A., Gryning, S.-E., Heikinheimo, M., Högström, U., Jochum, A. and Lundin, L.C., 1998. NOPEX - a northern hemisphere climate processes land surface experiment. *Journal of Hydrology* 212-213: 172-187
- Hillman, G.R. and Verschuren, J.P., 1988. Simulation of the effects of forest cover, and its removal, on subsurface water. *Water Resources Research* 24(2): 305-314
- Hinton, M.J., Schiff, S.L. and English, M.C., 1993. Physical properties governing groundwater flow in a glacial till catchment. *Journal of Hydrology* 142: 229-249
- Holko, L. and Lepistö, A., 1997. Modelling the hydrological behaviour of a mountain catchment using TOPMODEL. *Journal of Hydrology* 196: 361-377
- Hornberger, G.M. and Boyer, E.W., 1995. Recent advances in watershed modelling. *Reviews of Geophysics* 33, supplement (U.S. National Report to International Union of Geodesy and Geophysics, 1991-1994): 949-957
- Hottelet, Ch., Blažková, Š. and Bličník, M., 1994. Application of the ETH Snow Model to three basins of different character in Central Europe. *Nordic Hydrology* 25: 113-128
- Iorgulescu, I. and Jordan, J.-P., 1994. Validation of TOPMODEL on a small Swiss catchment. *Journal of Hydrology* 159: 255-273
- Jakeman, A.J. and Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model? *Water Resources Research* 29(8): 2637-2649
- Jayatilaka, C.J., Storm, B. and Mudgway, L.B., 1998. Simulation of water flow on irrigation bay scale with MIKE-SHE. *Journal of Hydrology* 208: 108-130
- Jensen, K.H. and Mantoglou, A., 1992. Future of distributed modelling. *Hydrological Processes* 6: 255-264

- Kirchner, J.W., Hooper, R.P., Kendall, C., Neal, C. and Leavesley, G., 1996. Testing and validating environmental models. *The Science of the Total Environment* 183: 33-47
- Kite, G.W., 1991. A watershed model using satellite data applied to a mountain basin in Canada. *Journal of Hydrology* 128: 157-169
- Kite, G.W., 1993. Application of a land class hydrological model to climatic change. *Water Resources Research* 29(7): 2377-2384
- Klemeš, V., 1986a. Operational testing of hydrological simulation models. *Hydrological Sciences Journal* 31: 13-24
- Klemeš, V., 1986b. Dilettantism in hydrology: transition or destiny. *Water Resources Research* 22(9): 177S-188S
- Klemke, E.D., Hollinger, R. and Kline, A.D. (eds.), 1988 (revised edition). *Introductory readings in the philosophy of science*. Prometheus Books, Buffalo, New York, 440 pp.
- Koh, G. and Jordan, R., 1995. Sub-surface melting in a seasonal snow cover. *Journal of Glaciology* 41(139): 474-482
- Konikow, L.F. and Bredehoeff, J.D., 1992. Ground-water models cannot be validated. *Advances in Water Resources* 15: 75-83
- Kryanova, V., Müller-Wohlfeil, D.-I. and Becker, A., 1998. Development and test of a spatially distributed hydrological/water quality model for mesoscale watersheds. *Ecological Modelling* 106: 261-289
- Kuczera, G., 1983. Improved parameter inference in catchment models, 1. Evaluating parameter uncertainty. *Water Resources Research* 19(5): 1151-1162
- Kuczera, G. and Mroczkowski, M., 1998. Assessment of hydrologic parameter uncertainty and the worth of data. *Water Resources Research* 34(6): 1481-1489
- Lamb, R., Beven, K.J. and Myrabø, S., 1997. Discharge and water table predictions using a generalized TOP-MODEL formulation, *Hydrological Processes* 11: 1145-1167
- Langsrud, Ø., Frigessi, A. and Høst, G., 1998a. *Pure model error of the HBV-model*. HYDRA-note 4, available from the Norwegian Water Resources and Energy Administration (NVE), Oslo, Norway, 28 pp.
- Langsrud, Ø., Høst, G., Follestad, T., Frigessi, A. and Hirst, D., 1998b. *Quantifying uncertainty in HBV runoff forecasts by stochastic simulations*. NR Note Nr. SAMBA/20/98, Norwegian Computing Center, P.O.Box 114 Blindern, N-0314 Oslo, Norway, 38 pp.
- Lankreijer, H., Lundberg, A., Grelle, A., Lindroth, A. and Seibert, J., 1999. Evaporation and storage of intercepted rain analysed by comparing two models applied to a boreal forest, *Agricultural and Forest Meteorology* (accepted for publication)
- Leavesley, G.H., Lichty, R.W., Troutman, B.M. and Saindon, L.G., 1983. *Precipitation-runoff modeling system – user's manual*: U.S. Geol. Surv. Water Resour. Invest. Rep. 83-4238, U.S. Geological Survey, Denver, 207 pp.
- Lindenlaub, M., Leibundgut, Ch., Mehlhorn, J. and Uhlenbrook, S., 1997. Interactions of hard rock aquifers and debris cover for runoff generation. In: *Hard Rock Hydrosystems (Proceedings of a symposium held during the Fifth IAHS Scientific Assembly at Rabat, Morocco, April–May 1997)* (ed. by T. Pointet), *IAHS Publication* 241: 63-74
- Lindström, G. and Rodhe, A., 1986. Modelling water exchange and transit times in till basins using O-18. *Nordic Hydrology* 17: 325-334
- Lindström, G., Johansson, B., Persson, M., Gardelin, M. and Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology* 201: 272-288
- Linsley, R.K., 1986. Flood estimates: how good are they?, *Water Resources Research* 22(9): 159S-164S
- Loaiciga, H.A., Valdes, J.B., Vogel, R., Garvey, J. and Schwarz, H., 1996. Global warming and the hydrologic cycle. *Journal of Hydrology* 174: 83-127
- Lundberg, A., Calder, I. and Harding, R., 1998. Evaporation of intercepted snow: measurement and modelling. *Journal of Hydrology* 206: 151-163
- Lørup, J.K., Refsgaard, J.C. and Mazvimavi, D., 1998. Assessing the effect of land use change on catchment runoff by combined use of statistical tests and hydrological modelling: case studies from Zimbabwe. *Journal of Hydrology* 205: 147-163
- Mauser, W. and Schädlich, S., 1998. Modelling the spatial distribution of evapotranspiration on different scales using remote sensing data. *Journal of Hydrology* 212-213: 250-267
- Mayer, D.G. and Butler, D.G., 1993. Statistical validation. *Ecological Modelling* 68: 21-32
- McCombie, C. and McKinley, I., 1993. Validation - another perspective (guest editorial). *Ground Water* 31(4): 530-531
- Mohseni, O. and Stefan, H.G., 1998. A monthly streamflow model. *Water Resources Research* 34(5): 1287-1298
- Moore, R.D., 1993. Application of a conceptual streamflow model in a glacierized drainage basin. *Journal of Hydrology* 150: 151-168

- Moore, I.D., Norton, T.W. and Williams, J.E., 1993. Modelling environmental heterogeneity in forested landscapes. *Journal of Hydrology* 150: 717-747
- Morton, A., 1993. Mathematical models: questions of thrustworthiness. *The British Journal for the Philosophy of Science* 44: 659-674
- Mroczkowski, M., Raper, G.P. and Kuczera, G., 1997. The quest for more powerful validation of conceptual catchment models. *Water Resources Research* 33(10): 2325-2335
- Nandakumar, N. and Mein, R.G., 1997. Uncertainty in rainfall-runoff model simulations and the implications for predicting the hydrological effects of land-use change. *Journal of Hydrology* 192: 211-232
- Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models, part 1 - a discussion of principles. *Journal of Hydrology* 10: 282-290
- Němec, J. and Schaake, J., 1982. Sensitivity of water resource systems to climate variation. *Hydrological Sciences Journal* 27: 327-343
- O'Connell, P.E. and Todini, E., 1996. Modelling of rainfall, flow and mass transport in hydrological systems: an overview. *Journal of Hydrology* 175: 3-16
- Oreskes, N., Shrader-Frechette, K. and Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263: 641-646
- Panagoulia, D. and Dimou, G., 1997. Sensitivity of flood events to global climate change. *Journal of Hydrology* 191: 208-222
- Parkin, G., O'Donnell, G., Ewen, J., Bathurst, J.C., O'Connell, P.E. and Lavabre, J., 1996. Validation of catchment models for predicting land-use and climate change impacts. 2. Case study for a Mediterranean catchment. *Journal of Hydrology* 175: 583-594
- Pilgrim, D.H., 1986. Bridging the gap between flood research and design practice. *Water Resources Research* 22(9): 165S-176S
- Piotroniro, A., Prowse, T., Hamlin, L., Kouwen, N. and Soulis, R., 1996. Application of a grouped response unit hydrological model to a northern wetland region. *Hydrological Processes* 10: 1245-1261
- Ponce, V.M., Lohani, A.K. and Scheyhing, C., 1996. Analytical verification of Muskingum-Cunge routing. *Journal of Hydrology* 174: 235-241
- Popper, K., 1934/1982 (7. Auflage). *Logik der Forschung*, Mohr, Tübingen, Germany, 468 pp.
- Popper, K., 1959/1968 (2nd ed.). *The Logic of Scientific Discovery*, Hutchinson & Co, London, Great Britain, 480 pp.
- Power, M., 1993. The predictive validation of ecological and environmental models. *Ecological Modelling* 68: 33-50
- Quinn, P.F. and Beven, K.J., 1993. Spatial and temporal predictions of soil moisture dynamics, runoff, variable source areas and evapotranspiration for Plynlimon, Mid-Wales. *Hydrological Processes* 7: 425-448
- Rango, A., 1994. Application of remote sensing methods to hydrology and water resources. *Hydrological Sciences Journal* 39: 309-320
- Refsgaard, J.C., 1996. Terminology, modelling protocol and classification of hydrological model codes (chapter 2). In: M.B. Abbott and J.C. Refsgaard (eds.), *Distributed hydrological modelling*, Water Science and Technology Library, Vol.22, Kluwer Academic Publishers, Dordrecht, The Netherlands, 321 pp., 17-39
- Refsgaard, J.C., 1997. Parameterisation, calibration and validation of distributed hydrological models. *Journal of Hydrology* 198: 69-97
- Refsgaard, J.C. and Knudsen, J., 1996. Operational validation and intercomparison of different types of hydrological models. *Water Resources Research* 32(7): 2189-2202
- Refsgaard, J.C., Storm, B. and Abbott, M.B., 1996. Comment on 'A discussion of distributed hydrological modelling' by K Beven (chapter 13B). In: M.B. Abbott and J.C. Refsgaard (eds.), *Distributed hydrological modelling*, Water Science and Technology Library, Vol.22, Kluwer Academic Publishers, Dordrecht, The Netherlands, 321 pp., 279-287
- Rice, K.C. and Hornberger, G.M., 1998. Comparison of hydrochemical tracers to estimate source contributions to peak flow in a small, forested, headwater catchment. *Water Resources Research* 34(7): 1755-1766
- Robson, A.C., Beven, K.J. and Neal, C., 1992. Towards identifying sources of subsurface flow: a comparison of components identified by a physically based runoff model and those determined by chemical mixing techniques. *Hydrological Processes* 6: 199-214
- Rodhe, A., 1987. *The origin of streamwater traced by oxygen-18*. Ph.D. Thesis, Uppsala University, UNGI Report Series A no. 41, Uppsala, Sweden, 260 pp., Appendix 73 pp.
- Rykiel, E.J., Jr., 1996. Testing ecological models: the meaning of validation. *Ecological Modelling* 90: 229-244
- Saelthun, N.R., 1996. *The 'Nordic' HBV Model. Description and documentation of the model version developed for the project Climate change and Energy Production*. Norwegian Water Resources and Energy Administration (Norges Vassdrags- og Energiverk, NVE), Oslo, Norway, 7, 26 pp.

- Seibert, P., 1994. *Hydrological characteristics of the NOPEX research area*, Thesis paper (examensarbete 20p), Uppsala University, Department of Earth Sciences, Hydrology, Uppsala, 51 pp.
- Seibert, J., 1999. Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and Forest Meteorology* (accepted for publication)
- Seibert, J., Uhlenbrook, S., Leibundgut, Ch., and Halldin, S.: 1999. Multiscale calibration and validation of a conceptual rainfall-runoff model. *Physics and Chemistry of the Earth* (in press)
- Servat, E. and Dezetter, A., 1991. Selection of calibration objective functions in the context of rainfall-runoff modelling in a sudanese savannah area. *Hydrological Sciences Journal* 36: 307-330
- Sidle, R.C., Tsuboyama, Y., Noguchi, S., Hosoda, I., Fujieda, M. and Shimizu, T., 1995. Seasonal hydrologic response at various spatial scales in a small forested catchment, Hitachi Ohta, Japan. *Journal of Hydrology* 168: 227-250
- Singh, V.P. (ed.), 1995. *Computer models of watershed hydrology*. Water Resources Publications, Highlands Ranch, Colorado, U.S.A., 1130 pp.
- Smith, R.E., Goodrich, D.R., Woolhiser, D.A. and Simanton, J.R., 1994. Comment on "Physically based hydrologic modeling, 2, Is the concept realistic?" by R.B. Grayson, I.D. Moore, and T.A. McMahon. *Water Resources Research* 30(3): 851-854
- Stagnitti, F., Parlange, J.-Y., Steenhuis, T.S. Parlange, M.B. and Rose, C.W., 1992. A mathematical model of hillslope and watershed discharge. *Water Resources Research* 28(8): 2111-2122
- Suyanto, A., O'Connell, P.E. and Metcalfe, A.V., 1995. The influence of storm characteristics and catchment conditions on extreme flood response: a case study based on the Brue river basin, U.K.. *Surveys in Geophysics* 16: 201-225
- Tsang, C.-F., 1991. The modeling process and model validation. *Ground Water* 29(6): 825-831
- Tuteja, N.K. and Cunnane, C., 1997. Modelling coupled transport of mass and energy into the snowpack - model development, validation and sensitivity analysis. *Journal of Hydrology* 195: 232-255
- Uhlenbrook, S., Seibert, J., Rodhe, A. and Leibundgut Ch., 1999. Prediction uncertainty of conceptual rainfall-runoff models caused by problems to identify model parameters and structure, *Hydrological Sciences Journal* (accepted for publication)
- van Oene, H. and Ågren, G.I., 1995. Complexity versus simplicity in modelling acid deposition effects on forest growth. *Ecological Bulletins* 40: 352-362
- Vandewiele, G.L., Xu, C.-Y. and Huybrechts, W., 1991. Regionalisation of physically-based water balance models in Belgium, application to ungauged catchments. *Water Resources Management* 5: 199-208
- Viney, N.R. and Sivapalan, M., 1996. The hydrological response of catchments to simulated changes in climate. *Ecological Modelling* 86: 189-193
- Waldenström, A., 1977. *Slutrapport över hydrologiska undersökningar i Kassjöåns representativa område*. HB Rapport nr 29, Swedish Meteorological Institute SMHI, Norrköping, Sweden. (In Swedish with English description of the area. English title: Final report on hydrological studies in the representative basin Kassjöån.), 59 pp.
- Węglarczyk, S., 1998. The interdependence and applicability of some statistical quality measures for hydrological models. *Journal of Hydrology* 206: 98-103
- White, J.D. and Running, S.W., 1994. Testing scale dependent assumptions in regional ecosystem simulations. *Journal of Vegetation Science* 5: 687-702
- Wigmosta, M.S., Vail, L.W. and Lettenmaier, D.P., 1994. A distributed hydrology-vegetation model for complex terrain. *Water Resources Research* 30(6): 1665-1679
- Xu, C.-Y., 1999a. Operational testing of a water balance model for predicting climate change impacts. *Agricultural and Forest Meteorology* (accepted for publication)
- Xu, C.-Y., 1999b. From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *Progress in Physical Geography* 23(1): 57-77
- Xu, C.-Y. and Vandewiele, G.L., 1995. Parsimonious monthly rainfall-runoff models for humid basins with different input requirements. *Advances in Water Resources* 18: 39-48
- Yao, H., Hashino, M. and Yoshida, H., 1996. Modeling energy and water cycle in a forested headwater basin. *Journal of Hydrology* 174: 221-234
- Ye, W., Jakeman, A.J. and Young, P.C., 1998. Identification of improved rainfall-runoff models for an ephemeral low-yielding Australian catchment. *Environmental Modelling & Software* 13: 59-74

Appendix: Terminology

The terminology used in hydrological modelling is sometimes used in different ways. Definitions of some important terms are given below, based on Singh (1995), Refsgaard (1996) and the general body of literature on modelling.

Calibration: the search for parameter values that provide the closest possible agreement between simulations and observations

Conceptual models: models that are built from a concept of the functioning of the studied real system. The routines are physically plausible, but they are not aimed to be exact descriptions

Distributed models: models in which spatial variations of all variables and parameters are considered

Lumped models: models in which the catchment is looked upon as a single unit

Model code: the implementation of a model on a computer

Model application: the use of a model for a specific catchment. This definition differs from Refsgaard (1996) who distinguished between modelling system and model (the latter being site-specific), but it is assumed to be more consistent with the general use of the terms model and model application.

Model: a simplified depiction of real systems. Models may be analogous (*e.g.*, scale models or electrical models) or mathematical. Only mathematical models are considered in this thesis.

Parameter: a constant used in the mathematical expressions of a model. Temporally, *e.g.*, seasonally, varying values may be defined for these 'constants'.

Physically-based model: model in which physical equations are used. These equations of mass and energy balances and flows have been shown to have direct physical relevance in small-scale experiments. Simple conceptual models are sometimes called physically-based (*e.g.*, Beven *et al.*, 1984; Vandewiele *et al.*, 1991; paper II). They may be physically reasonable, but the use of the term physically-based may be misleading.

Routine (or submodel): part of a larger model, *e.g.*, a part for simulation of snowmelt

Runoff model: a model that is intended to describe the transformation of water from precipitation to runoff and, depending on how detailed the model is, the variation of other variables such as groundwater levels (other terms found in literature are *rainfall-runoff model* or *catchment model*)

Semi-distributed model: model in which the spatial heterogeneity is taken into account through distribution functions using the concept of hydrological similarity (*e.g.*, elevation zones) or by divisions into different units, *e.g.*, HRUs (hydrological response units)

Variable: input (*e.g.*, precipitation) and output (*e.g.*, runoff, actual evaporation, spatial patterns of groundwater levels) of a model. *State variables* describe the state, *e.g.*, water level, in different compartments of a model