

DISS. ETH NO. 16832

**Uncertainty Analysis in Distributed Hydrological Modelling
Using a Bayesian Framework**

A dissertation submitted to
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZÜRICH

for the degree of
Doctor of Sciences

presented by

Jing Yang
Master of Engineering, Wuhan University
born 22 May 1977
citizen of China

accepted on the recommendation of

Prof. Dr. Peter Reichert, examiner
Prof. Dr. Alexander J.B. Zehnder, co-examiner
Dr. Karim C. Abbaspour, co-examiner
Dr. Hong Yang, co-examiner
Dr. Jasper A. Vrugt, co-examiner

Zürich, 2007

Seite Leer /
Blank leaf

Contents

Summary	5
Zusammenfassung	9
1 Introduction.....	13
1.1 Background and Motivation	13
1.2 Goals and Research Questions	15
1.3 Contents and Structure of the thesis	16
1.4 Reference	17
2 Hydrological Modelling of the Chaohe Basin in China: Statistical Model Formulation and Bayesian Inference.....	19
2.1 Introduction.....	20
2.2 Methods	23
2.3 Study Site and Data Compilation.....	30
2.4 Application	33
2.5 Results and Discussion.....	36
2.6 Summary and Conclusions.....	44
2.7 Acknowledgement.....	46
2.8 References	47
3 Bayesian Uncertainty Analysis in Distributed Hydrologic Modelling: A Case Study in the Thur River Basin (Switzerland).....	53
3.1 Introduction.....	54
3.2 Bayesian inference for a continuous-time autoregressive error model.....	56
3.3 Study Area and SWAT Model.....	61
3.4 Results and Discussion.....	66
3.5 Summary and Conclusion	74
3.6 Acknowledgments	75
3.7 References	75
4 Comparing Uncertainty Analysis Techniques for a SWAT Application to the Chaohe Basin in China.....	81
4.1 Introduction.....	82
4.2 Methodology, Selected Techniques, and Criteria for Comparison.....	84
4.3 Study Site, SWAT Watershed Simulation Program, and Model Application.....	93
4.4 Results and Discussion.....	97
4.5 Conclusions	116

4.6	<i>Acknowledgement</i>	118
4.7	<i>References</i>	118
5	Conclusions and Outlook	123
5.1	<i>Summary</i>	123
5.2	<i>Conclusions</i>	124
5.3	<i>Outlook</i>	124
5.4	<i>References</i>	125
Appendix A:	Interfacing SWAT with Systems Analysis Tools: A Generic Platform	127
Appendix B:	Explanation of some terms	138
	Curriculum Vitae	139
	Acknowledgements	141

Summary

Distributed hydrological modelling is useful to improve our understanding of the mechanics of natural processes in a watershed and their interaction with human activity. It can support the estimation of water availability, and the assessment of the impacts of climate and land use change or other activities within a watershed. This can support decision-making about measures to improve flood protection, water quality, aquatic ecosystems, and potential for recreational activities in the watershed. As all hydrologic modelling results are subject to uncertainty due to measurement errors in input and response and error in model structure, the reliability of modelling results must be assessed by estimating their uncertainty. In the last two decades, many uncertainty analysis techniques were developed and applied in the field of hydrology. Most of the uncertainty analysis (UA) methodologies focus only on parameter uncertainty and other sources of uncertainties are not or only partially represented, and the advantage and disadvantage of different UA techniques are not comparatively investigated.

The **aim** of this study is to develop a UA methodology which describes the effect of both parameter uncertainty and other sources of uncertainty and combines prior knowledge about parameter values with empirical evidence from the catchment to reduce prediction uncertainty. The applicability and effectiveness of this UA technique is tested by applying it to two case studies with different climatic conditions, and further by a comparison with other UA methodologies.

The two case study areas are the Chaohe basin and the Thur river basin. The Chaohe basin, with a drainage area of 5300 km², lies in the north of China and is a very important water source to Beijing city's water supply. Its climate is characterized as temperate continental and semi-arid. The Thur river basin, with a drainage area of 1700 km², is located in the north-eastern Switzerland, and has a pre-alpine/alpine climate.

Hydrology of the two watersheds was modelled using the program Soil and Water Assessment Tool (SWAT; Arnold et al., 1998). SWAT implements a semi-physically based and distributed hydrological model. This model accounts for the major processes influencing water transport in the watershed, such as soil water movement, surface water movement, groundwater movement, evapotranspiration, channel routing, etc. SWAT has been widely applied in the USA, Europe, Africa and Asia, and there are over 160 peer-reviewed published articles using this program (Gassman et al., 2005). As the distributed parameters of SWAT

Summary

are separated in distributed files, an interface which automatically manages the change of distributed parameter is highly desirable for the uncertainty analysis on SWAT. This interface, named iSWAT, was first developed to interface SWAT and systems analysis tool and facilitate our uncertainty analysis (Yang et al., 2005; or Appendix A).

In order to fulfil the aim of this study, the work is divided into three research tasks:

First a UA methodology is developed and applied to the Chaohe basin. This method must overcome difficulties of calibration of hydrologic models due to measurement errors in input and response, errors in model structure, and the large number of non-identifiable parameters of distributed models. The difficulties even increase in arid regions with high seasonal variation of precipitation, where the modelled residuals often exhibit high heteroscedasticity and autocorrelation. Extending earlier work in the field, we developed a procedure to overcome (i) the problem of non-identifiability of distributed parameters by introducing aggregate parameters and using Bayesian inference, (ii) the problem of heteroscedasticity of errors by combining a Box-Cox transformation of results and data with seasonally dependent error variances, (iii) the problems of autocorrelated errors, missing data and outlier omission with a continuous-time autoregressive error model, and (iv) the problem of the seasonal variation of error correlations with seasonally dependent characteristic correlation times. The posterior distribution of the parameters of the hydrologic model and the error model is calculated using a Markov Chain Monte Carlo (MCMC) technique. Our methodology was tested with the calibration of the hydrologic sub-model of the Soil and Water Assessment Tool (SWAT) in the Chaohe Basin in North China. The result demonstrated the good performance of this approach to uncertainty analysis, particularly with respect to fulfilment of statistical assumptions of the error model. A comparison with an independent error model clearly showed the superiority of our approach.

In the **second** step, the developed continuous-time autoregressive error model is further extended and tested with an application of SWAT to the Thur river basin in Switzerland, which has completely different climatic conditions compared to the Chaohe basin. This application corroborates the applicability of the approach, but also demonstrates the necessity of accounting for the heavy tails in the distributions of residuals and innovations. This is done by replacing the normal distribution of the innovations by a Student t distribution, the degrees of freedom of which is adapted to best represent the shape of the empirical distribution of the innovations. We conclude that with this extension the continuous-time autoregressive error model is applicable and flexible for hydrologic modelling under different climatic conditions.

The major remaining conceptual disadvantage is that this class of approaches does not lead to a separate identification of model input and model structural errors. The major practical disadvantage is the high computational demand characteristic for all MCMC techniques.

In a **third** step the developed technique is compared with other uncertainty analysis techniques widely used in hydrology to identify differences and similarities of these approaches. We compared 5 uncertainty analysis procedures: Generalized Likelihood Uncertainty Estimation (GLUE), Parameter Solution (ParaSol), Sequential Uncertainty Fitting algorithm (SUFI-2), and Bayesian-based continuous-time autoregressive model based on Markov Chain Monte Carlo (MCMC) and Importance Sampling (IS). For the comparison we used the SWAT model of the Chaohe Basin in China. As all of these techniques in fact are classes of techniques, we had to make choices of priors, likelihood functions and goal functions. We chose these according to their typical uses in applications of hydrological models. An analysis of the differences in the results of the selected techniques showed that many of the differences are consequences of not only the choice of the goal function but also the techniques. As far as the prediction uncertainty is concerned, except ParaSol and simple IS, all techniques lead to similar results. However, different techniques result in different posterior distributions of the parameters, best parameter sets, and performances of their corresponding simulation results. ParaSol leads to narrow parameter ranges because it only considers parameter uncertainty and uses an incorrect error model, while simple importance sampling failed due to its inefficient search strategy. From the point view of the authors, due to its superior theoretical foundation, Bayesian-based approaches are most recommendable. However, construction of the likelihood function and testing of the statistical assumption must require critical attention. Our continuous-time autoregressive error model contributes to this effort.

General conclusions:

It can be concluded that the developed continuous-time autoregressive error model is applicable and efficient for uncertainty analysis in distributed hydrological modelling. It accounts for the effects of parameter uncertainty, uncertainty in the input and response, and uncertainty in model structure on model predictions. The examination of the residuals and innovations between the observation and simulations shows that the assumption of independent t-distributions (or normal distributions) is adequate to describe the distribution of the innovations of the autoregressive error model. A comparison with the applications of other uncertainty analysis techniques in hydrology shows that the primary advantage of our

Summary

approach is not the difference in derived prediction uncertainty, but the testable fulfilment of the statistical assumptions of the error model. This improves the confidence in the uncertainty estimates.

The major conceptual disadvantage of the approach is the missing separation of error sources that contribute to total prediction uncertainty. It is an interesting research field to search for error models that would add this element. The major practical disadvantage is the high computational demand characteristic for all Markov Chain Monte Carlo techniques.

Zusammenfassung

Hydrologische Modellierung mit räumlich verteilten Modellen ist nützlich um unser Verständnis für die Mechanismen der natürlichen Prozesse in einem Einzugsgebiet und deren Wechselwirkung mit anthropogenen Einflüssen zu verbessern. Sie kann zur Schätzung der Wasserverfügbarkeit, der Beurteilung der Konsequenzen von Klima- und Landnutzungsänderungen, und zu den Effekten anderer Veränderungen im Einzugsgebiet beitragen. Damit können Entscheidungen über Massnahmen zur Verbesserung des Hochwasserschutzes, der Wasserqualität, der aquatischen Oekosysteme, und der Freizeitnutzung in einem Einzugsgebiet unterstützt werden. Da die Resultate aller hydrologischen Berechnungen aufgrund von Messfehlern in Eingangs- und Ausgangsgrössen, sowie von Modellstrukturfehlern unsicher sind, muss die Zuverlässigkeit der Modellresultate durch eine Unsicherheitsschätzung beurteilt werden. In den letzten zwei Jahrzehnten sind viele Methoden der Unsicherheitsschätzung in der Hydrologie entwickelt worden. Die meisten dieser Methoden fokussieren auf die Unsicherheit der Modellparameter und deren Auswirkungen und vernachlässigen andere Quellen der Unsicherheit. Zudem gibt es kaum vergleichende Studien über Vor- und Nachteile der verschiedenen Methoden.

Das **Ziel** dieser Untersuchung ist es, eine Methode der Unsicherheitsanalyse von hydrologischen Modellen zu entwickeln, die sowohl den Effekt von Parameterunsicherheit als auch den von Unsicherheit anderer Ursache berücksichtigt, zusätzlich Vorwissen über Parameterwerte mit empirischer Evidenz aus dem Einzugsgebiet kombiniert und dadurch die Unsicherheit der Prognosen reduziert. Die Anwendbarkeit und Effektivität dieser Methode wird in zwei Fallstudien und durch Vergleich mit anderen Methoden getestet.

Die beiden Fallstudien betreffen die Einzugsgebiete der Flüsse Chaohe und Thur. Das Chaohe-Einzugsgebiet liegt in Nordchina und hat eine Fläche von 5300 km². Es steht unter dem Einfluss von kontinentalem, semi-aridem Klima und hat eine sehr grosse Bedeutung für die Trinkwasserversorgung von Peking. Das Thur-Einzugsgebiet liegt in der Nordostschweiz und hat eine Fläche von 1700 km². Das Klima ist voralpin/alpin.

Die Hydrologie der beiden Einzugsgebiete wurde mit Hilfe des Programms "Soil and Water Assessment Tool" (SWAT, Arnold et al., 1998) modelliert. In SWAT ist ein auf physikalischen Grundlagen basierendes verteiltes hydrologisches Modell implementiert, das die wichtigsten den Wassertransport beeinflussenden Prozesse, wie Wassertransport im Boden, Oberflächenabfluss, Grundwassertransport, Verdunstung, Wassertransport in

Zusammenfassung

Fliessgewässern, usw. beschreibt. SWAT wurde schon sehr oft auf Einzugsgebiete in den Vereinigten Staaten, Europa, Afrika und Asien angewandt. Das hat zu mehr als 160 begutachteten wissenschaftlichen Publikationen geführt, die SWAT anwenden (Grassman et al., 2005).

Um die Ziele dieser Arbeit zu erreichen, wurde sie in drei Forschungsarbeiten unterteilt:

Als **erster Schritt** wurde eine Methode für die Analyse der Unsicherheit von hydrologischen Modellen entwickelt und auf das Chaohe-Einzugsgebiet angewandt. Diese Methode musste mit den Schwierigkeiten der Kalibrierung hydrologischer Modelle durch Messfehler in Eingangs- und Ausgangsgrössen, durch Fehler in der Modellstruktur und durch die grosse Zahl nicht-identifizierbarer Parameter umgehen. Diese Schwierigkeiten wurden in dieser Fallstudie noch durch das aride Klima mit der grossen saisonalen Veränderung des Niederschlags erhöht, das zu einer Vergrösserung der Veränderungen der Varianz und Autokorrelation der Residuen führt. Als Erweiterung früher publizierter Methoden wurde ein Verfahren entwickelt, das (i) die Schwierigkeiten der nicht-Identifizierbarkeit der verteilten Parameter durch die Einführung aggregierter Parameter und die Verwendung von Bayesscher Inferenz, (ii) das Problem der Veränderungen der Varianz der Fehler durch die Kombination einer Box-Cox Transformation mit saisonal variierenden Fehlervarianzen, (iii) die Probleme der Autokorrelation der Fehler, fehlender Daten und Ausreisserelimination mit einem zeitlich kontinuierlichen autoregressiven Fehlermodell und (iv) das Problem der saisonal abhängigen Stärke der Autokorrelation mit saisonal variierenden charakteristischen Korrelationszeiten überwindet. Die a posteriori Verteilung der Parameter des hydrologischen Modells und des Fehlermodells wird mit einem Markovketten Monte Carlo Verfahren berechnet. Die Methodik wurde durch die Kalibration des in SWAT implementierten hydrologischen Modells im Chaohe-Einzugsgebiet in Nordchina getestet.

Als **zweiter Schritt** wurde das Fehlermodell erweitert und auf das Thur-Einzugsgebiet in der Schweiz angewandt. Dieses hat ein Klima, das sehr stark vom Klima im Chaohe-Einzugsgebiet abweicht. Diese Anwendung bestätigte die Anwendbarkeit der Methodik, zeigte aber auch die Notwendigkeit auf, die Abweichung der Verteilungsform der Residuen von einer Normalverteilung zu berücksichtigen. Dies wurde durch die Einführung einer t-Verteilung implementiert, deren Anzahl Freiheitsgrade ein Anpassen der Form an die empirische Verteilung der Residuen erlaubt. Damit ergibt jetzt das Verfahren gute Resultate für Einzugsgebiete mit sehr unterschiedlichen klimatischen Einflüssen. Der hauptsächliche verbleibende konzeptionelle Nachteil ist, dass diese Klasse von Fehlermodellen nicht zu einer

separaten Identifikation von Input- und Modellstrukturfehlern führt. Der hauptsächlich praktische Nachteil ist der hohe Rechenaufwand, der charakteristisch ist für alle Markovketten-Verfahren.

In einem **dritten Schritt** wurde das entwickelte Verfahren mit andern in der Hydrologie häufig verwendeten Unsicherheitsanalyseverfahren verglichen. Dies diente der Identifikation von Unterschieden und Ähnlichkeiten der verschiedenen Ansätze. Es wurden die folgenden Verfahren verglichen: Generalized Likelihood Uncertainty Estimation (GLUE), Parameter Solution (ParaSol), Sequential Uncertainty Fitting algorithm (SUFI-2), und Bayessche Inferenz mit dem neuen zeitlich kontinuierlichen autoregressiven Fehlermodell basierend auf der numerischen Approximation mittels Markovketten Monte Carlo (MCMC) und mittels Importance Sampling (IS). Für den Vergleich wurde das in SWAT implementierte Modell für das Chaohc-Einzugsgebiet in China gewählt. Da alle diese Techniken eigentlich Klassen von Techniken sind, mussten noch a priori Verteilungen und Likelihood-Funktionen oder Zielfunktionen gewählt werden. Diese wurden gemäss typischen Anwendungen in der Hydrologie gewählt. Danach wurde analysiert, inwieweit die Unterschiede in den Resultaten der verschiedenen Verfahren die Konsequenzen der Zielfunktionen oder der Verfahren an sich sind. Was die Unsicherheit der Modellprognosen betrifft, ergaben alle Verfahren mit Ausnahme von ParaSol und einfachem IS ähnliche Resultate. Demgegenüber waren die Resultate für die a posteriori Verteilungen der Parameter verschieden. ParaSol führt wegen der ausschliesslichen Berücksichtigung der Parameterunsicherheit und wegen eines unkorrekten Fehlermodells zu sehr kleinen Unsicherheitsbereichen. Primitives Importance Sampling ausgehend von der a priori Verteilung führt wegen der zu grossen Ineffizienz zu keinen brauchbaren Resultaten. Aus unserer Sicht sind die Bayesschen Methoden wegen ihrer besseren theoretischen Begründung am empfehlenswertesten. Bei der Anwendung dieser Methoden muss aber der Konstruktion der Likelihood Funktion und dem Testen der statistischen Annahmen entscheidende Beachtung geschenkt werden. Unser neuer Ansatz trägt zu einer solchen Entwicklung bei.

Allgemeine Schlussfolgerungen

Das in dieser Arbeit entwickelte zeitlich kontinuierliche autoregressive Fehlermodell scheint sich sehr gut für die Anwendung auf hydrologische Probleme zu eignen. Es berücksichtigt die Effekte von Parameterunsicherheit, Unsicherheit in Eingangs- und Ausgangsgrössen und Unsicherheit in der Modellstruktur auf Modellprognosen. Ein Vergleich der empirischen und der angenommenen Verteilungen der Inkremente des autoregressiven

Zusammenfassung

Modells zeigt die Angemessenheit der Annahmen von unabhängigen t-Verteilungen (oder Normalverteilungen). Ein Vergleich mit den Resultaten anderer Unsicherheitsanalysetechniken zeigt, dass der Hauptvorteil unseres Ansatzes die testbare Erfüllung der statistischen Annahmen des Fehlermodells ist. Das erhöht das Vertrauen in die Resultate der Unsicherheitsschätzung.

Der grösste verbleibende konzeptionelle Nachteil unsers Verfahrens ist die fehlende Aufspaltung der Fehlerquellen, die zur Gesamtunsicherheit beitragen. Dies ist ein interessantes Forschungsgebiet für weitere Untersuchungen. Der grösste verbleibende praktische Nachteil ist der hohe Rechenaufwand, der für alle Markovketten Monte Carlo Verfahren typisch ist.

Reference

- Arnold, J.G., Srinivasan R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment - Part 1: Model development. *Journal of the American Water Resources Association* 34(1), 73-89.
- Gassman, P.W., Reyes, M.r, Green, C.H., Arnold, J.G, 2005. SWAT Peer-reviewed literature: A review. In: Srinivasan, R., Jacobs, J., Day, D., Abbaspour, K. (Eds.), 2005 3rd International SWAT Conference Proceedings, July 11-15, Zurich, Switzerland.
- Yang, J., Abbaspour, K.C., Reichert, P., 2005, Interfacing SWAT with Systems Analysis Tools: A Generic Platform. In: Srinivasan, R., Jacobs, J., Day, D., Abbaspour, K. (Eds.), 2005 3rd International SWAT Conference Proceedings, July 11-15, Zurich, Switzerland, pp. 169-178

1 Introduction

1.1 Background and Motivation

Distributed hydrological models are widely used in many applications such as estimating water availability and assessing the impacts of climate change and land use change within the study watershed. They assist decision-making in water management and contribute in research to understanding the mechanisms of the natural process and the interaction with human activities. However, all modelling results are subject to uncertainty due to the measurement errors in input and response and error in model structure. The assessment of the reliability of the modelling results is very much dependent on the way the uncertainties are described.

A. Distributed hydrological modelling

In hydrological models the watershed can be characterized differently depending on the modelling purpose. Hydrologic models can be classified as lumped or distributed models based on the description of the processes of the system geometry, model input, governing laws, initial and boundary conditions, and model output. Different from lumped models, distributed models take an explicit account of spatial variability of processes, input, boundary conditions, and output (Singh, 1995). Examples of implementations of such models include SHE (Abbott et al., 1986a and 1986 b) and SWMM 9 (Metcalf and Eddy, Inc., et al., 1971). Based on the physical basis of the described processes, the distributed model can be further classified as conceptual distributed model, physically based distributed model, and semi-physically based / semi-conceptual distributed model. However, no matter how spatially explicit and how physically based the distributed model is, some parameters should be estimated indirectly through calibration. For example, the soil hydraulic conductivity can be correctly measured for a particular location, but such a measurement is often invalid as a representative average over the model grid cell, let alone the entire domain (Kavestski et al., 2002).

Recently, distributed models were coupled with Geographic Information System (GIS) or Remote sensing (RS) to make use of the topographic data, land use data, remote sensing data, etc for more precise and reasonable prediction, and integrated with biological and ecological sub-models to model water related issues. Examples of such models include the Soil and

Chapter 1

Water Assessment Tool (SWAT) (Arnold et al., 1998), AGNPS (Young et al., 1989), and HSPF (Bicknell et al., 2000).

On the other hand, users of distributed models face the challenging task of calibration and uncertainty analysis. The difficulty is due to the interaction of different processes and parameters, the nonlinear and non-monotone characteristics of relationships parameterized in the model, and the large number of non-identifiable distributed parameters, while only a relatively small number of observations are available.

The program SWAT (Arnold et al., 1998) was chosen in this study. SWAT is a semi-physically based and distributed watershed model. It describes the climatic and topographic heterogeneity through sub-basins based on DEM and climatic stations. It describes the heterogeneities in land use, soil, management practices through HRUs (Hydrologic Response Units), which is the unique combination of land use, soil, and management practises for each HRU. The SWAT program has been widely applied in the USA, Europe, Africa and Asia, and there are over 160 peer-reviewed published articles using this program (Gassman et al., 2005).

In this paper, the SWAT is applied to the Chaohe basin and the Thur river basin. The Chaohe basin, in North China, is characterized by a temperate continental and semi-arid climate, and the Thur river basin, in north-east of Switzerland, is characterized by a pre-alpine/alpine climate.

B. Uncertainty analysis in hydrological modelling

The uncertainties in hydrologic modelling are normally classified as input uncertainty, model parameter uncertainty, model structural uncertainty and uncertainty in the measurement of response which is used for model calibration. Input uncertainty is often related to imprecise measurement of model input or initial condition and spatial aggregation of model input, such as DEM data, land use data, rainfall, temperature and initial groundwater level, etc. Model parameter uncertainty is caused by the indirect/dependent measurement, imprecise measurement or conceptualization process of the model parameters. Model structural uncertainty may arise from the simplification of the reality, or in erroneous conceptualization of the processes. Uncertainty in the measurement response often refers to imprecision of the measured response (e.g. observed streamflow and groundwater level). In some literatures, the input uncertainty and uncertainty in the measurement response are called data uncertainty (Gupta, Beven and Wagener, 2005).

Because of the existence of those uncertainties, it has been accepted by most hydrologists that the process of calibration cannot lead to a single optimal parameter set but one has to find a probability distribution of parameters that represents the knowledge of parameter values. This is called the principle of “equifinality” by Beven (2001), but it fits more generally into any Bayesian approach of statistical inference (Gelman et al., 1995).

In the last two decades, many uncertainty analysis methodologies have been developed/introduced and applied in hydrological modelling. These methodologies include Generalized Likelihood Uncertainty Estimation (GLUE; Beven and Binley, 1992), Parameter Solution (ParaSol; Van Griensven et al., 2006), Markov Chain Monte Carlo (MCMC) and Importance Sampling (IS) within the Bayesian Approach, Sequential Uncertainty Fitting algorithm (SUFI-2; Abbaspour et al., 2006), etc. Most of these methodologies and/or their applications only focus on parameter uncertainty. To better understand the contribution of different uncertainty sources to the prediction uncertainty, there is a need to investigate the uncertainties in other sources in addition to parameter uncertainty. It is also useful to investigate the strengths and weaknesses of different methodologies so as to provide an overview on how to select a suitable UA methodology in (distributed) hydrologic modelling.

1.2 Goals and Research Questions

Goals:

The primary goal of this project is to develop a UA methodology that accounts for different uncertainty sources in hydrologic modelling and the statistical assumptions of which are testable and not violated. The second goal is to compare this technique with other UA methodologies by studying their strengths and weaknesses, and providing guidance for UA methodology selection.

In order to illustrate the usefulness and applicability/flexibility of our methodology, two study sites with different climatic conditions were selected. One of these sites is the Chaohe basin in North China, which is characterized by a temperate continental and semi-arid climate, the other is the Thur river basin, in north-eastern Switzerland, which is characterized by a pre-alpine/alpine climate. Detailed data records such as rainfall, temperature, DEM, land use and river discharge are available.

To achieve the above goals, the work is divided into three major research tasks:

Chapter 1

1. Development of the continuous-time autoregressive error model and its application to the Chaohe basin;
2. Extension of the developed methodology and its application to the Thur river basin;
3. Comparison of our technique with other UA methodologies.

Research Questions:

The research objective is addressed by answering three major questions:

1. How can the uncertainty in model structure and in the measurement of input and response be described within the Bayesian framework in hydrological modelling?
2. Is the developed methodology applicable for other watersheds as well?
3. What are the advantages and disadvantages of the developed methodology in comparison with other UA methodologies that are also used in hydrological modelling?

1.3 Contents and Structure of the thesis

This thesis is structured in 4 main sections as described below:

Section 2: Hydrological Modelling of the Chaohe Basin in China: Statistical Model Formulation and Bayesian Inference. In this section, the importance of uncertainty analysis is addressed and a brief literature review on uncertainty analysis is given. In addition, the problems and difficulties in the separation of uncertainty sources are discussed. In the methodology part, a continuous-time autoregressive error model within the Bayesian framework is developed. This is the key issue of this paper. The developed methodology is then applied to a SWAT model for the Chaohe basin, China, which has an obvious seasonal rainfall/flow variation.

Section 3: Bayesian Uncertainty Analysis in Distributed Hydrologic Modelling: A Case Study in the Thur River Basin (Switzerland). Following section 2, the statistical assumption of the continuous-time autoregressive error model is strengthened and further generalized in this section. The generalized methodology is then applied to the application of the SWAT model to the Thur river basin, Switzerland, which has different climatic conditions compared to the Chaohe basin.

Section 4: Comparing different uncertainty analysis techniques in a SWAT application to the Chaohe Basin in China. This section is concerned about the relative advantages and disadvantages of different UA methodologies: GLUE, ParaSol, SUFI-2, and

the continuous-time autoregressive model based on two different numerical implementations. The different methodologies are introduced and applied to the application of the SWAT model to the Chaohe basin with the same prior setup. The derived posterior parameter distributions, the quality of the best fit, prediction uncertainty, the efficiency of the techniques and the conceptual basis of the techniques are compared.

Section 5: Conclusions and outlook. The results from the above sections are analyzed, and the effectiveness of the developed methodology is discussed. An outlook is given for future research on the separation of uncertainty source in hydrologic modelling.

1.4 Reference

- Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., Srinivasan, R., 2007. Spatially-distributed modelling of hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of Hydrology*, 333:413-430.
- Abbott, M.B., Bathurst, J.C., Cunge, J.A., Oconnell, P.E. and Rasmussen, J., 1986a. An Introduction to the European Hydrological System - Systeme Hydrologique Europeen, She .1. History and Philosophy of a Physically-Based, Distributed Modeling System. *Journal of Hydrology*, 87(1-2): 45-59.
- Abbott, M.B., Bathurst, J.C., Cunge, J.A., Oconnell, P.E. and Rasmussen, J., 1986b. An Introduction to the European Hydrological System - Systeme Hydrologique Europeen, She .2. Structure of a Physically-Based, Distributed Modeling System. *Journal of Hydrology*, 87(1-2): 61-77.
- Arnold, J.G., Srinivasan R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment - Part 1: Model development. *Journal of the American Water Resources Association* 34(1), 73-89.
- Beven, K. and Binley, A., 1992. The Future of Distributed Models - Model Calibration and Uncertainty Prediction. *Hydrological Processes*, 6(3): 279-298.
- Beven, K. 2001. How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences* 5(1), 1-12.
- Bicknell, B.R, Imhoff, J., Kittle, J., Jobes, T., Donigian, A.S., 2000. *Hydrological Simulation Program – Fortran User’s Manual*, Release 12, U.S EPA.
- Gassman, P.W., Reyes, M.r, Green, C.H., Arnold, J.G, 2005. SWAT Peer-reviewed literature: A review. In *Proceeding of the 2005 3rd international SWAT conference*, Zurich, Switzerland.

Chapter 1

- Gelman, S., Carlin, J. B., Sten, H. S., Rubin, D. B., 1995. Bayesian Data Analysis, Chapman and Hall, New York, USA.
- Gupta, H.V., Beven, K.J., Wagener, T., 2005. Model Calibration and Uncertainty Estimation. In: Anderson, M.G., (Eds.), Encyclopedia of Hydrological Sciences, John Wiley & Sons, Ltd. pp2015-2031.
- Metcalf and Eddy, Inc., University of Florida and Water Resources Engineers, Inc., 1971. Storm water management model, Vol. 1, final report. Water Pollution Control Research Series 11024 DOC 07/71, U.S. Environmental Protection Agency, Washington, D.C, USA.
- Kavetski, D, Franks, S.W., Kuczera, G., 2003. Confronting Input Uncertainty in Environmental Modelling. In: Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N. and Turcotte, R. (Eds.), Calibration of Watershed Models. American Geophysical Union, Washington, DC., USA, pp 49-68.
- Singh, V.P., 1995. Chapter 1: Watershed modelling. In: Singh, V.P. (Eds.), Computer Models of watershed Hydrology, Water Resources Publications, Colorado, USA, pp1-22.
- van Griensven, A. and Meixner, T., 2006. Methods to quantify and identify the sources of uncertainty for river basin water quality models. Water Science and Technology, 53(1): 51-59.
- Young, R.A., Onstad, C.A., Bosch, D.D., Anderson, W.P. 1989. AGNPS: A nonpoint-source pollution model for evaluating agricultural watersheds. Journal of Soil and Water Conservation 44(2), 168-173.

2 Hydrological Modelling of the Chaohe Basin in China: Statistical Model Formulation and Bayesian Inference

Jing Yang, Peter Reichert, Karim C. Abbaspour, Hong Yang

(Accepted by Journal of Hydrology)

Abstract

Calibration of hydrologic models is very difficult because of measurement errors in input and response, errors in model structure, and the large number of non-identifiable parameters of distributed models. The difficulties even increase in arid regions with high seasonal variation of precipitation, where the modelled residuals often exhibit high heteroscedasticity and autocorrelation. On the other hand, support of water management by hydrologic models is important in arid regions, particularly if there is increasing water demand due to urbanization. The use and assessment of model results for this purpose requires a careful calibration and uncertainty analysis. Extending earlier work in this field, we developed a procedure to overcome (i) the problem of non-identifiability of distributed parameters by introducing aggregate parameters and using Bayesian inference, (ii) the problem of heteroscedasticity of errors by combining a Box-Cox transformation of results and data with seasonally dependent error variances, (iii) the problems of autocorrelated errors, missing data and outlier omission with a continuous-time autoregressive error model, and (iv) the problem of the seasonal variation of error correlations with seasonally dependent characteristic correlation times. The technique was tested with the calibration of the hydrologic sub-model of the Soil and Water Assessment Tool (SWAT) in the Chaohe Basin in North China. The results demonstrated the good performance of this approach to uncertainty analysis, particularly with respect to the fulfilment of statistical assumptions of the error model. A comparison with an independent error model and with error models that only considered a subset of the suggested techniques clearly showed the superiority of the approach based on all the features (i) to (iv) mentioned above.

Keywords: Watershed model calibration; Uncertainty analysis; Bayesian inference; Continuous-time autoregressive error model; MCMC; SWAT; UNCSIM; Aggregate parameters.

2.1 Introduction

With continuous urbanization and economic development, water scarcity and deterioration of water quality have become increasingly severe in many river basins in the world, especially in arid regions, such as North China. Tackling these problems with effective water management strategies is crucial for sustaining the economic development and meeting the water demand of a growing population. Hydrologic models can assist decision-makers in dealing with these problems by providing systematic and consistent information on water availability, water quality, and impacts of human activities, particularly land use change, on the hydrologic systems. However, the confidence in model predictions relies on their uncertainties. These are difficult to estimate. As hydrologic models need site-specific calibration, uncertainty estimation must be based on the results of the calibration and validation processes (Yapo et al, 1996; Duan et al., 2003).

Parameter uncertainty in hydrological modelling has gained a lot of interest over the past two decades. It has been accepted by most hydrologists that the process of calibration cannot lead to a single “optimal” parameter set but one has to find a probability distribution of parameters that represents the knowledge about parameter values. This is called the principle of “equifinality” by Beven (2001) and Beven and Freer (2001), but it fits more generally into any Bayesian approach of statistical inference (Gelman et al., 1995). Many techniques have been proposed to quantify parameter uncertainty of hydrologic models. Early approaches with quite sophisticated error models were based on first-order approximations of the model equations for Bayesian inference (e.g. Kuczera 1983). Due to the difficulty of quantifying the errors of linearization (Vrugt and Bouten, 2002) and the increasing availability of computational power, these approaches have been replaced by Monte Carlo based numerical approximations to the posterior that account for model nonlinearity. Monte Carlo approaches can be divided into global random (mostly uniform) importance sampling approaches including some generalizations (Beven and Binley 1992; Lamb, 1999; Beven and Freer 2001), regional or iterative importance sampling and similar approaches (Abbaspour et al., 1997, 2004, 2007), and Markov Chain Monte Carlo techniques (Kuczera and Parent, 1998; Bates and Campbell, 2001; Vrugt et al., 2003). The techniques based on global scanning of the parameter space have the conceptual advantage of being, in principle, able to deal with arbitrary shapes of posterior distributions. In particular, this includes multi-modal distributions. However, as shown by Kuczera and Parent (1998) and others, they are very inefficient and can even lead to misleading results unless a very large sample of parameters is

drawn. This is increasingly difficult if the parameter space has a high dimension. The efficiency of importance sampling can be improved by iteratively adapting the sampling distribution and by using efficient sampling techniques (Reichert et al., 2002). Such a “regionalization” should be based on a global search and not on a small sample from global scanning, as this can propagate the above-mentioned problem of a misleading parameter selection due to a too small sample to the local search. Nevertheless, iterative adaptation of the sampling distribution to approximate the posterior in importance sampling remains difficult, particularly in high dimensional parameter spaces. For these reasons, the Markov Chain Monte Carlo approach seems to be the most promising general approach. In order to avoid problems of finding the maximum of the posterior and long burn-in phases, Markov Chains should be started in the neighbourhood of the maximum of the posterior probability density calculated with a global search algorithm (e.g. Duan et al., 1992, 1993, 1994). This is the numerical approach we follow also in this paper. Still the choice of an adequate error model to construct the model likelihood function remains a challenge.

Conventional watershed models consist of a deterministic description of rainfall, runoff, evaporation, storage and transport processes. Due to the representation of internal storage processes by the model, measurement errors of input and errors in model structure lead to sequentially dependent errors in model results. These errors, together with the measurement errors of response, can be accounted for by an overall additive autoregressive error model (e.g. Kuczera 1983, Bates and Campbell, 2001) or by considering the error sources separately and propagating them through the model. The methodology of including input uncertainty in Bayesian inference is well known (Zellner, 1971). However, it has rarely been applied in hydrological modelling (Kavetski et al., 2003), probably because its application to (rainfall) time series introduces a large number of additional parameters and interferes with errors in model structure. Errors in model structure have been addressed by making the deterministic hydrologic model stochastic and combining parameter estimation with the estimation of model state variables (e.g. Vrugt et al., 2005). As the first approach involves many additional parameters to be estimated and the second is hard to implement in an existing large hydrologic simulation program, we will base our analysis on an overall additive autoregressive error model that accounts for the joint effect of measurement errors of input and response and errors in model structure. We extend previous approaches with discrete-time autoregressive error models by introducing a continuous-time autoregressive error model.

Since the development of the Stanford Watershed Model (Crawford and Linsley, 1966), there has been a proliferation of watershed models and corresponding simulation programs. Currently, such simulation programs are coupled with GIS and are being integrated with biological and ecological sub-models. Such simulation programs include AGNPS (Young et al., 1989), SWAT (Arnold et al., 1998) and HSPF (Bicknell et al., 2000). Coupling with GIS makes it easier to represent the watersheds in more and more detail. This increases the number of model parameters, decreases their identifiability, and makes calibration and uncertainty analysis even more difficult. To limit this increase in model complexity, we define aggregate global or regional parameters to modify distributed parameters. With this concept, distributed parameters are changed by additively or multiplicatively modifying their initial, spatially varying values, or by defining values that depend on potentially important influence factors, such as soil or land use categories or sub-basin index.

The Chaohe Basin in North China is selected as a case study for model calibration and uncertainty analysis. The severe water scarcity and growing population makes efficient water management an important issue in North China. The Chaohe Basin is a large part of the catchment of the Miyun reservoir, an important drinking water source for the city of Beijing. Previous hydrological modelling studies in North China had the objectives of simulating river discharge and water quality and assessing the impact of land use and climate change. Some commonly used models include the Xinanjiang model (Zhao, 1992; Zhao and Liu, 1995), the Distributed Time Variant Gain Model (Wang et al., 2002), the TOPKAPI model (TOPographic Kinematic APproximation and Integration) (Liu, 2004), and also SWAT. The application of SWAT in China includes the application in the Heihe Basin (Huang and Zhang, 2004; Wang et al., 2003), the Suomo Basin (Chen and Chen, 2004), the Luohe Watershed (Zhang et al., 2003a, 2003b), the Yuzhou Reservoir Basin (Zhang et al., 2004) and the Luxi Watershed (Hu et al., 2003). However, to the authors' knowledge, none of the above applications includes an uncertainty analysis, and hence, these studies are of limited use for water management as their reliability cannot be quantified.

The goal of the present study is to calibrate the SWAT program (Arnold et al., 1998) for the Chaohe Basin in North China and to perform a state of the art uncertainty analysis for this model application. The calibration of the model duplicates to some degree earlier efforts (see above). However, it is hoped that the calibration and uncertainty estimation techniques and tools developed and used for this case study will stimulate similar development in future

studies. This would improve future support of watershed management by hydrologic models in general and particularly in North China.

This paper is structured as follows: Section 2.2 outlines the techniques used in this paper. This section starts with an introduction of the Soil and Water Assessment Tool (SWAT), a description of the aggregate parameters used for model calibration and a brief description of the software developed for interfacing SWAT (Yang et al., 2005) with our systems analysis tool UNCSIM (Reichert, 2005). We then focus on the construction of the likelihood function for heteroscedastic and autocorrelated errors by a detailed description of the continuous-time autoregressive error model which we will use for optimal flexibility in representing the deviation of Box-Cox transformed measurements from (transformed) SWAT output. This section ends with a brief outline of the (standard) techniques of numerical Bayesian inference by Markov Chain Monte Carlo as we will use them in our application. Section 2.3 contains a description of the study site and of data acquisition and compilation. In section 2.4, we describe our application of the model to the Chaohe Basin, and in section 2.5, we present and discuss the results. Finally the main results are summarized and conclusions are drawn in section 2.6.

2.2 Methods

2.2.1 Deterministic Hydrological Model

We use the Soil and Water Assessment Tool (SWAT; Arnold et al., 1998; <http://www.brc.tamus.edu/swat>) as the simulation software that implements the deterministic hydrologic model to describe the hydrologic processes in the catchment. A major reason for this choice was that, in addition to hydrology, SWAT provides modules for the simulation of sediment, nutrients and pesticides in the watershed. This gives us the opportunity to extend the present work to water quality in future studies. SWAT implements a partially physically based and semi-distributed model that operates on a daily time step. In SWAT, a watershed is divided into a number of sub-basins based on a given DEM (Digital Elevation Model). Within each sub-basin, soil and land use maps are overlaid to create a number of hydrologic response units (HRUs), which are the basic working units. SWAT simulates the land phase of the hydrologic cycle for each HRU. The land phase controls the amount of water, sediment, nutrients and pesticides delivered to the main channel in each sub-basin. The resulting loads are then routed through the channel network of the watershed to the basin outlet.

Chapter 2

Water storage in each HRU in SWAT is represented by four storage volumes: snow, soil profile, shallow aquifer and deep aquifer. The water mass balance then considers precipitation, interception, runoff, infiltration, evapotranspiration, percolation, “revap” (water flux from the shallow aquifer to the soil by evaporation, diffusion and condensation), lateral movement and, finally, routing in the rivers. Surface runoff from daily rainfall is calculated using a modified SCS curve number method (Soil Conservation Service, 1972), which estimates the amount of runoff based on local land use, hydrologic soil group, and antecedent soil moisture. SWAT can estimate potential evapotranspiration using Penman-Monteith (Monteith, 1965; Allen, 1986; Allen et al., 1989), Priestley-Taylor (Priestley and Taylor, 1972), or Hargreaves (Hargreaves et al., 1985) methods based on data availability. The Hargreaves method is used in this study to estimate potential evapotranspiration. For actual evapotranspiration, SWAT first evaporates any rainfall intercepted by the plant canopy, and then calculates transpiration, sublimation and soil evaporation based on potential evapotranspiration and water availability. SWAT’s root zone water processes include evapotranspiration, percolation into deep soil, and lateral movement, while the shallow runoff, lateral flow and aquifer contribute to the stream flow. SWAT provides two water routing methods, the variable storage (Williams, 1969) and Muskingum (Cunge, 1969; Chow et al., 1988) methods. In this study we use the Muskingum routing method.

2.2.2 Aggregate Parameters

Calibration of a distributed hydrologic model using discharge data from a small number of river sites always leads to non-identifiable parameters due to strong overparameterization. One way of dealing with overparameterized models is to combine prior knowledge about parameter values with data using Bayesian inference. Due to the very large number of model parameters, this would be computationally very demanding. In order to use the information about spatial variation or about the dependence on important influence factors, but to keep the number of parameters small, an alternative approach is used in this study. For distributed parameters either a value or a multiplicative or additive modification term to the prior parameter values can be used instead of using the parameter values in all HRUs. Such a modification term can be chosen to have a global value, or different values for different categories of important influence factors such as soil type, land use, etc.

To do this, an interface program iSWAT (Yang et al., 2005; <http://www.uncsim.eawag.ch/interfaces/swat>) was implemented that allows its users to encode information into an extended parameter name on how to apply a parameter value conditionally on important influence

factors and location and hence aggregate distributed parameters. The name of the SWAT aggregate parameters uses the following format:

$$x_ \langle \text{parname} \rangle . \langle \text{ext} \rangle _ \langle \text{hydrogrp} \rangle _ \langle \text{soltext} \rangle _ \langle \text{landuse} \rangle _ \langle \text{subbsn} \rangle \quad (2.1)$$

Where x represents the type of change to be applied to the parameter (v: value; a: absolute change; r: relative change), <parname> is the SWAT parameter name; <ext> represents the extension of the SWAT input file which contains the parameter value; <hydrogrp>, <soltext>, <landuse>, and <subbsn> represent the dependent factors, referring to soil hydrologic group, the type of soil texture, the land use category, and sub-basin number/crop index/fertilizer index, respectively. For these factors, single values or groups of values can be specified, or they can be omitted to indicate that the change is applied independently of the factor. For example, v__CN2.mgt = 75 will cause a global replacement of CN2 values (v=value) in the management files by 75, and a__CN2.mgt_____AGRR_1,5 = 5 will increase the CN2 values by 5 (a=absolute change) in subbasin 1 and 5 in which the landuse types are “AGRR” independent of hydrologic group and soil texture (corresponding codes omitted in the extended parameter name).

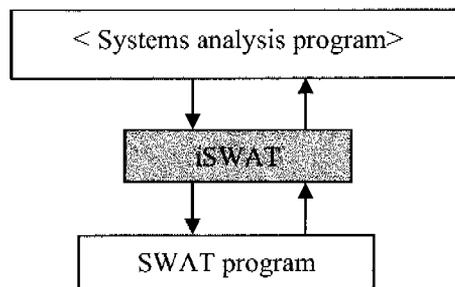


Figure 2.1: A schematic flowchart shows the linkage of the systems analysis program and SWAT model.

The interface program iSWAT reads the parameter names and values from a file written by a systems analysis tool and modifies the SWAT input files accordingly. After execution of SWAT it compiles the results to a format that can easily be interpreted by the systems analysis tool. This makes it possible to couple SWAT with any systems analysis tool that supports the simple file-based information exchange format described by Reichert (2006). Figure 2.1 illustrates the interaction between iSWAT, SWAT and the systems analysis tool. More details are given in Yang et al. (2005).

2.2.3 Likelihood Functions

The deterministic hydraulic simulation model can be written in the form of the function

$$\mathbf{y}^M(\boldsymbol{\theta}) = (y_{t_0}^M(\boldsymbol{\theta}), y_{t_1}^M(\boldsymbol{\theta}), \dots, y_{t_n}^M(\boldsymbol{\theta})) \quad (2.2)$$

where $y_{t_i}^M(\boldsymbol{\theta})$ represents the model output at time t_i for model parameter values $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ (in our case mostly aggregate parameters as described in section 2.2.2), and M indexes the model.

As mentioned in the introduction, measurement errors of input and response and errors in model structure lead to deviations of simulation results from measurements. These are modelled as an additive random process to Box-Cox transformed model results (Box and Cox, 1964; 1982). The parameters of the Box-Cox transformation give us degrees of freedom to improve the degree of fulfillment of simple distributional assumptions of the errors. After adding the random error to the transformed results, we need a transformation back to the original scale for comparison with data. This leads to the following model formulation as random variables at all observation time points:

$$Y_{t_i}^M(\boldsymbol{\theta}) = g^{-1}(g(y_{t_i}^M(\boldsymbol{\theta})) + E_{t_i}) \quad (2.3)$$

In this equation, g and g^{-1} are forward and backward Box-Cox transformations

$$g(y) = \begin{cases} \frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & \lambda_1 \neq 0 \\ \ln(y + \lambda_2) & \lambda_1 = 0 \end{cases}, \quad g^{-1}(z) = \begin{cases} (\lambda_1 z + 1)^{1/\lambda_1} - \lambda_2 & \lambda_1 \neq 0 \\ \exp(z) - \lambda_2 & \lambda_1 = 0 \end{cases}, \quad \frac{dg}{dy} = (y + \lambda_2)^{\lambda_1 - 1} \quad (2.4)$$

λ_1 and λ_2 are Box-Cox transformation parameters ($y + \lambda_2$ must be larger than zero for all values of y ; $\lambda_1 = \lambda_2 = 1$ leads to the identity transformation), E_{t_i} is the random variable quantifying the total effects of measurement errors of input and response and errors in model structure on model results, and $Y_{t_i}^M$ is the random variable describing model response at time t_i .

The simplest assumption for the error term, E_{t_i} , is that it consists of independent, normally distributed random variables with mean zero and standard deviation σ . In this case, the probability density of E_{t_i} is given by

$$f_{E_{t_i}}^*(\varepsilon_{t_i}) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{\varepsilon_{t_i}^2}{\sigma^2}\right) \quad (2.5)$$

However, due to the memory effect of storage processes, even independent input and model structure errors will lead to correlated response errors. For this reason, we use an autoregressive model to formulate the error term E_{t_i} . In the past, this has usually been done with discrete-time autoregressive error models (e.g. Kuczera, 1983; Bates and Campbell, 2001). Our approach is similar, but we use a continuous-time autoregressive error model (e.g. Brockwell and Davis, 1996; Brockwell, 2001) because this seems to be a more reasonable representation of continuous-time processes in the catchment and because this significantly facilitates dealing with missing data and outliers. Because of the adequateness and simplicity of the mean-reverting Ornstein-Uhlenbeck process, we use it to describe this error term (e.g. Kloeden and Platen, 1992; the same process was used for describing continuous, time-dependent model parameters in Tomassini et al. 2007). The conditional probability densities of the individual errors are then given by

$$f_{E_{t_0}}(\varepsilon_{t_0}) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{\varepsilon_{t_0}^2}{\sigma^2}\right) \quad (2.6)$$

$$f_{E_{t_i}|E_{t_{i-1}}}(\varepsilon_{t_i}|\varepsilon_{t_{i-1}}) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma \sqrt{1 - \exp\left(-2 \frac{t_i - t_{i-1}}{\tau}\right)}} \exp\left(-\frac{1}{2} \frac{\left(\varepsilon_{t_i} - \varepsilon_{t_{i-1}} \exp\left(-\frac{t_i - t_{i-1}}{\tau}\right)\right)^2}{\sigma^2 \left(1 - \exp\left(-2 \frac{t_i - t_{i-1}}{\tau}\right)\right)}\right)$$

where σ is the asymptotic standard deviation of the errors and τ the characteristic correlation time. The assumption here is that the random disturbances, sometimes called innovations (Chatfield, 2003),

$$I_{t_i} = E_{t_i} - E_{t_{i-1}} \exp\left(-\frac{t_i - t_{i-1}}{\tau}\right) \quad (2.7)$$

rather than the individual errors, E_{t_i} , are independent and normally distributed. Keeping the asymptotic standard deviation of the errors E_{t_i} at σ , the innovations must have standard deviations of

$$\sigma_{I_{t_i}} = \sigma \sqrt{1 - \exp\left(-2 \frac{t_i - t_{i-1}}{\tau}\right)} \quad (2.8)$$

They reach σ if the time difference between two observations is large compared to the characteristic correlation time, τ , and they are significantly smaller if subsequent observations are within that time or even closer. Note that the formulation of this likelihood function is

similar to the approach suggested by Duan et al. (1988) for use with unequally spaced data. However, equation (2.8) formulates the essential difference: when decreasing temporal distance of measurement points in our error model not only the correlation increases, but also the standard deviation of the error term decreases. This guarantees its applicability on a continuous time scale.

Combining the deterministic hydrologic model (2.2) with the Box-Cox transformation (2.3, 2.4) and the independent error model (2.5), we end up with the following likelihood function:

$$f_{\mathbf{y}^M|\boldsymbol{\theta}}^*(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=0}^n \left[\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{[g(y_{t_i}) - g(y_{t_i}^M(\boldsymbol{\theta}))]^2}{\sigma^2}\right) \cdot \left| \frac{dg}{dy} \right|_{y=y_{t_i}} \right] \quad (2.9)$$

Note that when keeping the transformation parameters λ_1 and λ_2 constant, maximum likelihood parameter estimation results in minimizing the sum of weighted squares of the deviations of transformed model results from transformed data. In the special case of the identity transformation this reduces to the minimization of the sum of squares of model results from measured data what is equivalent to maximizing the Nash-Sutcliffe coefficient (Nash and Sutcliffe, 1970).

Similarly, combining the deterministic hydrologic model (2.2) with the Box-Cox transformation (2.3, 2.4) and the Ornstein-Uhlenbeck continuous-time autoregressive error model (2.6) leads to the likelihood function

$$f_{\mathbf{y}^M|\boldsymbol{\theta}}(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{[g(y_{t_0}) - g(y_{t_0}^M(\boldsymbol{\theta}))]^2}{\sigma^2}\right) \cdot \left| \frac{dg}{dy} \right|_{y=y_{t_0}} \cdot \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma \sqrt{1 - \exp\left(-2 \frac{t_i - t_{i-1}}{\tau}\right)}} \exp\left(-\frac{1}{2} \frac{\left[\begin{array}{c} g(y_{t_i}) - g(y_{t_i}^M(\boldsymbol{\theta})) \\ - [g(y_{t_{i-1}}) - g(y_{t_{i-1}}^M(\boldsymbol{\theta}))] \exp\left(-\frac{t_i - t_{i-1}}{\tau}\right) \end{array} \right]^2}{\sigma^2 \left(1 - \exp\left(-2 \frac{t_i - t_{i-1}}{\tau}\right)\right)}\right) \cdot \left| \frac{dg}{dy} \right|_{y=y_{t_i}} \right] \quad (2.10)$$

In the following, we will call the likelihood function (2.9) independent error model and we will use it for comparative purposes only. We call the likelihood function (2.10) autoregressive error model and will use it for the actual application of the model.

In order to test the statistical model assumptions of the likelihood function (2.9), we will check the standardized residuals of transformed data and model results

$$r_{t_i}(\boldsymbol{\theta}, \mathbf{y}^{\text{obs}}) = \frac{g(y_{t_i}^{\text{obs}}) - g(y_{t_i}^M(\boldsymbol{\theta}))}{\sigma} \quad (2.11)$$

for independent normality (with mean zero and standard deviation unity) and for the likelihood function (10) the standardized observed innovations of the transformed data and model results

$$i_{t_i}(\boldsymbol{\theta}, \mathbf{y}^{\text{obs}}) = \frac{g(y_{t_i}^{\text{obs}}) - g(y_{t_i}^M(\boldsymbol{\theta})) - (g(y_{t_{i-1}}^{\text{obs}}) - g(y_{t_{i-1}}^M(\boldsymbol{\theta}))) \exp\left(-\frac{t_i - t_{i-1}}{\tau}\right)}{\sigma \sqrt{1 - \exp\left(-2\frac{t_i - t_{i-1}}{\tau}\right)}} \quad (2.12)$$

for fulfillment of the same statistical assumption. In these equations, \mathbf{y}^{obs} are the observations corresponding to the model outputs. To check for heteroscedasticity and correlation of standardized residuals (Equation 2.11) and standardized observed innovations (Equation 2.12), we plotted their time series, autocorrelation functions and cumulative periodograms.

2.2.4 Bayesian Inference and Numerical Implementation

We will derive a posterior probability density function of the parameters, $f_{\boldsymbol{\theta}_{\text{post}}|\mathbf{Y}}(\boldsymbol{\theta}|\mathbf{y}^{\text{obs}})$, from the prior density, $f_{\boldsymbol{\theta}_{\text{pri}}}(\boldsymbol{\theta})$, and data, \mathbf{y}^{obs} , according to Bayes' theorem

$$f_{\boldsymbol{\theta}_{\text{post}}|\mathbf{Y}}(\boldsymbol{\theta}|\mathbf{y}^{\text{obs}}) = \frac{f_{\mathbf{Y}^M|\boldsymbol{\theta}}(\mathbf{y}^{\text{obs}}|\boldsymbol{\theta}) \cdot f_{\boldsymbol{\theta}_{\text{pri}}}(\boldsymbol{\theta})}{\int f_{\mathbf{Y}^M|\boldsymbol{\theta}'}(\mathbf{y}^{\text{obs}}|\boldsymbol{\theta}') f_{\boldsymbol{\theta}_{\text{pri}}}(\boldsymbol{\theta}') d\boldsymbol{\theta}'} \quad (2.13)$$

where the model likelihood function, $f_{\mathbf{Y}^M|\boldsymbol{\theta}}(\mathbf{y}^{\text{obs}}|\boldsymbol{\theta})$, is either (for comparative purposes) given by equation (2.9) or by equation (2.10). A numerical sample of the posterior distribution is derived by applying the Metropolis-Hastings Markov Chain Monte Carlo algorithm (Gelman et al., 1995). In order to avoid long burn-in periods (or even lack of convergence to the distribution) the chain is started at a numerical approximation to the maximum of the posterior density calculated with the aid of the shuffled complex global optimization algorithm (Duan et al., 1992, 1993, 1994). Markov Chains were run until 20 000 model runs were reached after the convergence criterion of Heidelberger and Welch was fulfilled (Cowles and Carlin, 1996; Best et al., 1995). The likelihood function, optimization

algorithm and Markov Chain algorithm were implemented in an updated version of the UNCSIM program (Reichert, 2005).

2.3 Study Site and Data Compilation

2.3.1 Site Description

The Chaohe Basin is situated in North China with a drainage area of 5,300 km² above the Xiahui station (see Figure 2.2). The characteristic of the climate is temperate continental and semi-arid. From 1980 to 1990 the average daily maximum temperature was 6.2°C, the average daily minimum temperature in this period was 0.9°C, and the yearly rainfall varied between 350 to 690 mm. The elevation varies from 200 m at the basin outlet to 2,400 m at the highest point in the catchment. The topography is characterized by high mountain ranges, steep slopes and deep valleys. Water flows fast in the river and the average channel slope is 1.87%. Average daily flow at the catchment outlet is 9.3 m³ s⁻¹ and varies irregularly from around 800 m³ s⁻¹ during the flood season to lower than 1 m³ s⁻¹ in the dry season at the Xiahui station. The ratio of runoff at the Xiahui station to the rainfall in this basin decreased from 0.24 in 1980 to 0.09 in 1990. It is believed that the decline is mainly due to the intensified human activities, including increasing water use and build up of more water retention structures.

The Chaohe River is one of two tributaries flowing into the Miyun reservoir, which is an important drinking water reservoir for Beijing city and provides nearly half of the city's water supply (Jia and Cheng, 2002). As a major drinking water source of Beijing city, both water quality and water quantity are important concerns in this river basin. However, due to the decrease in incoming water and increase in soil loss and pollution from the upper stream, the water level of the Miyun Reservoir has been declining continuously and the water quality has been deteriorating. The reservoir is severely affected by dissolved pollutants and pollutants attached to suspended particles and sedimentation (see e.g. Wang et al, 2001; <http://www.china.com.cn/chinese/zhuanti/qyjjfz/1169096.htm>). Although the Chinese government has taken measures to improve water quantity and quality in the inflows to the reservoir, such as implementing reforms on water prices and land conservation programs, the problems have not decreased and are even exacerbating. One of the reasons for the ineffective control of the problems has been the lack of quantitative understanding of the hydrologic system and of how it is affected by human activities in the river basin. Given this

background, a hydrologic and water quality model for this basin is useful for providing more reliable information to improve water resource management.

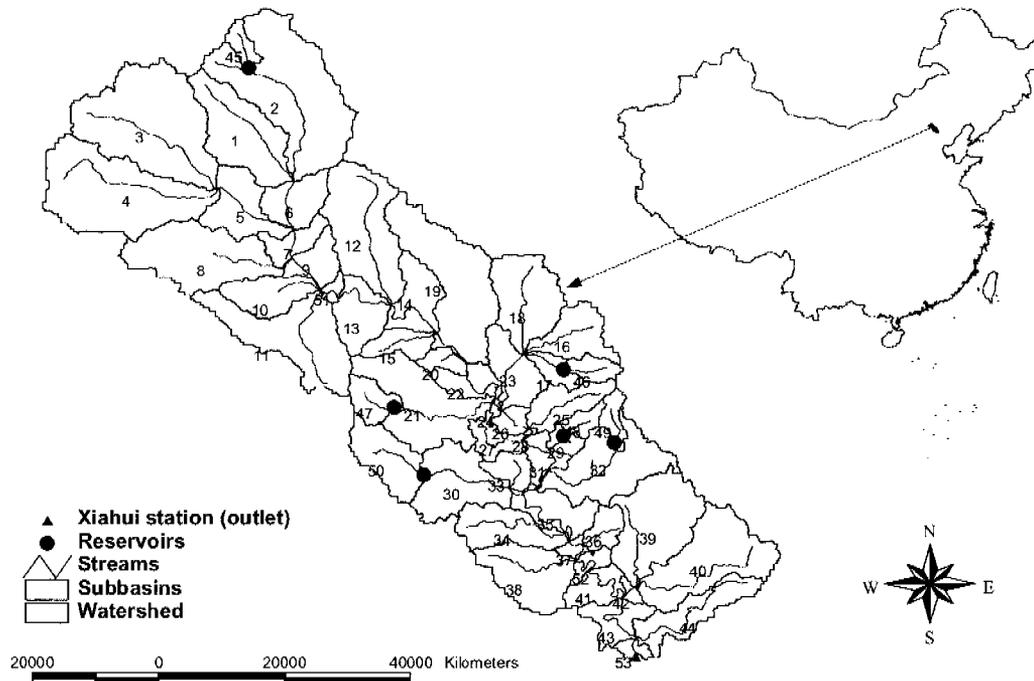


Figure 2.2: Location of the Chaohe Basin in North China. Solid circles represent reservoirs, irregular polygons stand for subbasins in this project, and the triangle is the outlet (Xiahui station).

2.3.2 Data Compilation

In the following, a short description is given of the data gathered for the Chaohe Basin and its processing for the application of the model:

(i) A digital elevation map (DEM) at a scale of 1:1,000,000 was obtained from the “China Data Centre” of the University of Michigan (<http://chinadatacenter.org/newcdc/>).

(ii) A soil map at a scale of 1:1,000,000 was provided by the Institute of Soil Science, Chinese Academy of Sciences, Nanjing (Shi et al, 2004), (http://issas.ac.cn/english/soil_database.htm). The soil data is aggregated into 35 soil profiles in the Chaohe Basin. The original soil data only contains the percentages of the texture components, bulk density and organic carbon content. The soil erodibility factor, USLE_K, was estimated by an equation proposed by Williams (1995), saturated hydraulic conductivity, SOL_K, and available water storage capacity, SOL_AWC, were initially estimated using

pedotransfer functions proposed by Schaap et al.(1996 and 2001) and later calibrated for the region.

(iii) A land use map at a scale of 1:1,000,000 was provided by the Institute of Geography Chinese Academy of Sciences, Beijing. The dominant land uses in the Chaohe Basin are forest (49.5%), grassland (27.3%), and agricultural land (21.3%). Agricultural land is mainly distributed on both sides of the main channel and tributaries including the river flood plains. The main crops planted in this area are corn and wheat.

(iv) Daily precipitation data contains 15 stations over a period of 6 years (1985-1990) (Hydrologic Yearbook, Ministry of Water Resources, China), and daily maximum and minimum temperatures for 2 stations (Fengning and Luanping) over a period of 40 years (1960-2000) (China Meteorological Administration, www.cma.gov.cn).

(v) Six reservoirs were built in the Chaohe Basin during the 1970s or earlier. However, the total catchment area of these reservoirs contributes to only 5% of the watershed area and the total storage capacity of the reservoirs is 2.4% of the yearly water discharge passing through the Xiaohui station. The properties of those reservoirs are listed in Table 2.1.

(vi) Daily discharge data used is at the basin outlet Xiaohui (5340 km²) for 6 years (1985-1990) (Hydrologic Yearbook, Ministry of Water Resources, China).

Table 2.1: Properties of the reservoirs in the Chaohe Basin

Reservoir name	County	Longitude (degree)	Latitude (degree)	Watershed Area (ha)	Storage capacity (10 ⁴ m ³)
Mujiang	Fengning	116.60	41.52	2541	84.2
Shanshengmiao	Fengning	116.77	41.02	3600	10.4
Hongqi	Fengning	117.10	40.95	5360	52.4
lingying	Fengning	117.10	41.05	3050	144.0
Caoyingzi	Luanping	117.18	40.93	2400	127.0
Longtanmiao	Luanping	116.80	40.92	10300	286.0

According to the natural river network, the topography of the basin, and the distribution of rainfall stations, the basin was divided into 53 sub-basins. To get a reasonable resolution of soil properties and land use and management practices, these sub-basins were divided into a total of 262 HRUs (see section 2.2.1 for an explanation of HRUs). The watershed parameterization and the model input were obtained using the ArcView interface to SWAT (AVSWAT; Di Luzio et al., 2002), which provides a graphical support for the disaggregation scheme and allows the construction of the model input from digital maps. The initial values of distributed parameters (hereafter referred as “initial (parameter) estimates”) are either

directly obtained from the database (e.g. parameters concerning the soil property, crop property, rainfall, etc) or estimated by AVSWAT based on input maps and database (e.g., curve number, manning roughness coefficients).

2.4 Application

2.4.1 Choice and Prior Distribution of Parameters

An application of SWAT with initial parameter estimates extremely over-predicted the observed flow. The reason may be the incorrectness of the initial estimates of soil parameters and the fact that the initial estimates of SWAT's land use parameters can not be directly applied to the Chaohe Basin. Based on a literature review and preliminary sensitivity analyses, 10 SWAT aggregate parameters related to river flow were selected for calibration (Table 2.2).

The prior distribution of the aggregate parameters was assumed to be the combination of independent marginal distributions for the parameters. For the SWAT aggregate parameters, uniform priors within reasonable ranges were assumed. These ranges were selected based on recommendations given in the SWAT user manual (Neitsch et al., 2001). For the parameters σ and τ , characterizing the statistical part of the likelihood functions (2.9) and (2.10), densities proportional to $1/\sigma$ and $1/\tau$ were chosen, which is equivalent to assuming that the logarithms of these parameters are uniformly distributed. The prior for the parameter λ_1 of the Box-Cox transformation was chosen to be uniform in the interval $[0,1]$, and λ_2 was kept fixed at a value of zero. Table 2.2 gives an overview of the parameters used for calibration and their marginal prior distributions.

In the Chaohe Basin, wet and dry seasons can be clearly distinguished. Hydrology during the wet season is driven by highly variable precipitation, whereas during the dry season a slowly decreasing base flow dominates the hydrograph pattern. This can have consequences for the model error. For this reason, we inferred different values of the parameters σ and τ of the statistical error model for dry season (October to May) and for the wet season (July and August) assuming a linear transition from one value to the other in June and September.

Table 2.2: Model parameters used for inference and their prior distributions

Aggregate Parameters ^{*1}	Name and meaning of underlying SWAT parameter	Range of Initial parameter underlying SWAT parameter ^{*2}	Prior distribution of aggregate parameter ^{*3}
a_CN2.mgt	CN2: curve number	72-92	U[-30, 5]
v_ESCO.hru	ESCO: soil evaporation compensation factor	0.95	U[0,1]
v_EPCO.hru	EPCO: plant uptake compensation factor	1.00	U[0,1]
r_SOL_K.sol	SOL_K: soil hydraulic conductivity (mm/hr)	1.6-328.27	U[-0.8, 0.8]
a_SOL_AWC.sol	SOL_AWC: soil available water capacity (mm H ₂ O/mm soil)	0-0.13	U[0,0.15]
v_ALPHA_BF.gw	ALPHA_BF: base flow alpha factor (1/day)	0.048	U[0,1]
r_SLSUBBSN.hru ^{*3}	SLSUBBSN: average slope length (m)	9.461-91.463	U[-0.6, 0.6]
a_CH_K2.rte	CH_K2: effective hydraulic conductivity in main channel alluvium (mm/hr)	0	U[0,150]
a_OV_N.hru	OV_N: overland manning roughness	0.06-0.15	U[0,0.2]
v_GW_DELAY.gw	GW_DELAY: groundwater delay time (days)	31	U[0,300]
λ_1	parameter of Box-Cox transformation		U[0,1]
σ	σ_{dry} σ_{wet}	standard deviation during dry season in Equation (2.9) and (2.10) standard deviation during wet season in Equation (2.9) and (2.10)	Inv ^{*4} Inv ^{*4}
τ	τ_{dry} τ_{wet}	characteristic correlation time of autoregressive process during dry season (days) in Equation (2.10) characteristic correlation time of autoregressive process during wet season (days) in Equation (2.10)	Inv ^{*4} Inv ^{*4}

^{*1} Aggregate parameters are constructed based on Eq. (2.1). “a_”, “v_” and “r_” means an absolute increase, a replacement, and a relative change to the initial parameter value respectively.

^{*2} This range is based on the initial parameter estimates of the project.

^{*3} Prior distributions of aggregate parameters are based on Neitsch, et al (2001); U[a,b] in this column means the distribution of this parameter/aggregate parameters is uniform over the interval [a,b].

^{*4} “Inv” means that the probability density at the value x is proportional to $1/x$ (the logarithm of the parameter is uniform distributed)

2.4.2 Approach

As we cannot specify reasonable initial values for all storage volumes considered in the model, SWAT was operated for a “warm-up period” of 5 years without comparison of model results with data. We found that such a “warm-up period” was sufficient to minimize the effects of the initial state of SWAT variables on river flow. Furthermore, to test the calibrated model parameters, the model was calibrated and tested based on the observed discharges at the watershed outlet (Xiahui station, Figure 2.2) using a split sample procedure. The data from the years 1985-1988 with omission of a single outlier in 1985 was used for calibration, and the data from 1989-1990 was used to test the model.

To analyse and demonstrate the effect of the Box-Cox transformation, of the seasonal dependence of the parameters of the error model, and of the autoregressive error model, we compare the results of four different calibrations:

1. Application of the independent and normally distributed error model (2.9) with λ_1 and λ_2 set equal to unity. This is used to get a reference to traditional hydrological modelling.
2. Application of the independent error model (2.9) with Box-Cox transformation (λ_1 estimated and $\lambda_2 = 0$), but without seasonally dependent parameters of the independent error model.
3. Application of the independent error model (2.9) with Box-Cox transformation (λ_1 estimated and $\lambda_2 = 0$) and with seasonally dependent parameters (σ_{dry} , and σ_{wet}).
4. Application of the autoregressive error model (2.10) with Box-Cox transformation (λ_1 estimated and $\lambda_2 = 0$) and with seasonally dependent parameters (σ_{dry} , σ_{wet} , τ_{dry} and τ_{wet}).

For each of these calibrations, heteroscedasticity and autocorrelation of standardized residuals (Equation (2.11) for the independent error model) or of standardized observed innovations (Equation (2.12) for the autoregressive error model) were checked.

To quantify prediction uncertainty we plot the 95% posterior uncertainty bands together with the simulation corresponding to the parameters at the maximum posterior density and the observed data points. In addition, we calculate the standard deviation of the model results considering all sources of uncertainty (parameter uncertainty and input, model structure and

output uncertainty described by the autoregressive model) and the standard deviation of the results of the deterministic model due to parameter uncertainty only. Finally, we calculate the fraction of data points contained in the 95% prediction uncertainty band.

2.5 Results and Discussion

Figures 2.3, 2.4 and 2.5 show regression diagnostics for the simulations during the calibration period for all four calibrations described in section 2.4.2. Figure 2.3 shows time series of the standardized residuals (for calibration 1; according to equation 2.9 without transformation), standardized residuals of transformed model results and data (for calibrations 2 and 3; according to equation 2.9) and standardized observed innovations of the autoregressive error model (for calibration 4; according to equation 2.10) at the maxima of the posterior densities during the calibration period. The standardized residuals corresponding to calibrations 1 and 2 exhibit strong heteroscedasticity and high autocorrelation. By combining the Box-Cox transformation with the seasonally dependent standard deviation (σ_{dry} , and σ_{wet}), the heteroscedasticity of the residuals could be considerably decreased in calibration 3. However, all three calibrations show a strong autocorrelation of residuals, particularly during the dry season. Calibration 4, based on the autoregressive error model with Box-Cox transformation and seasonally dependent parameters of the error model, obviously decreases the degree of these problems considerably. This is quantified in Figures 2.4 and 2.5 which show the autocorrelation functions and cumulative periodograms of residuals or observed innovations for all four calibrations. Figure 2.4 clearly shows the high autocorrelation of the residuals for the calibrations 1, 2 and 3. This is in contradiction to the independence assumption of the error models. On the other hand, autocorrelation of the observed innovations of the autoregressive error model are very small (calibration 4). Figure 2.5 demonstrates that the white-noise assumption of the standardized residuals of the transformed output is clearly violated in calibrations 1, 2 and 3, whereas it can be accepted for the standardized observed innovations of the autoregressive error model.

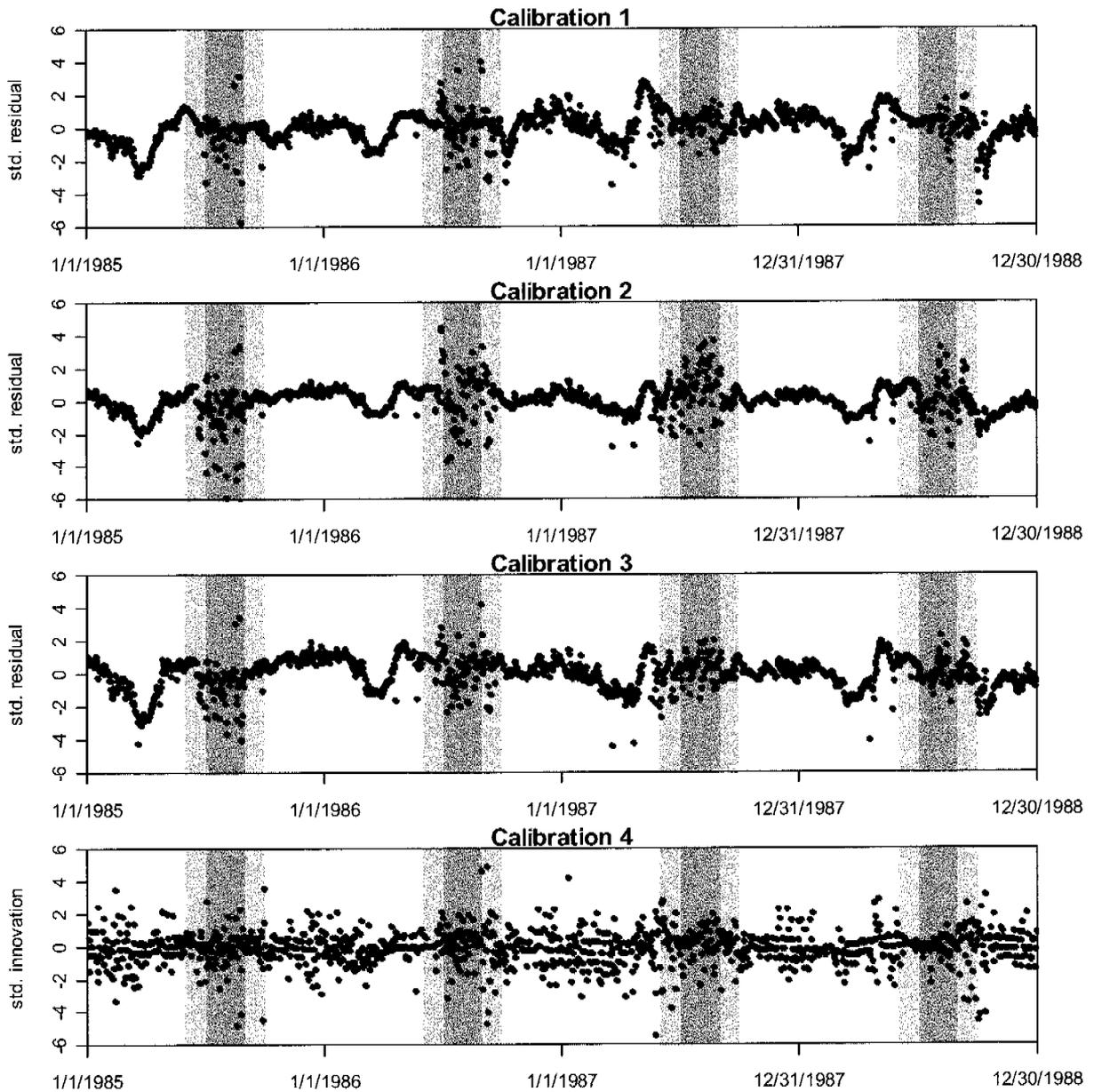


Figure 2.3: Time series of standardized residuals (calibration 1), standardized residuals of transformed model results and data (calibrations 2 and 3) and standardized observed innovations of the continuous-time autoregressive model (calibration 4) at the maxima of the posterior distributions. The dark shaded areas indicate the wet seasons, light shaded areas indicate transition periods, and white areas indicate dry seasons.

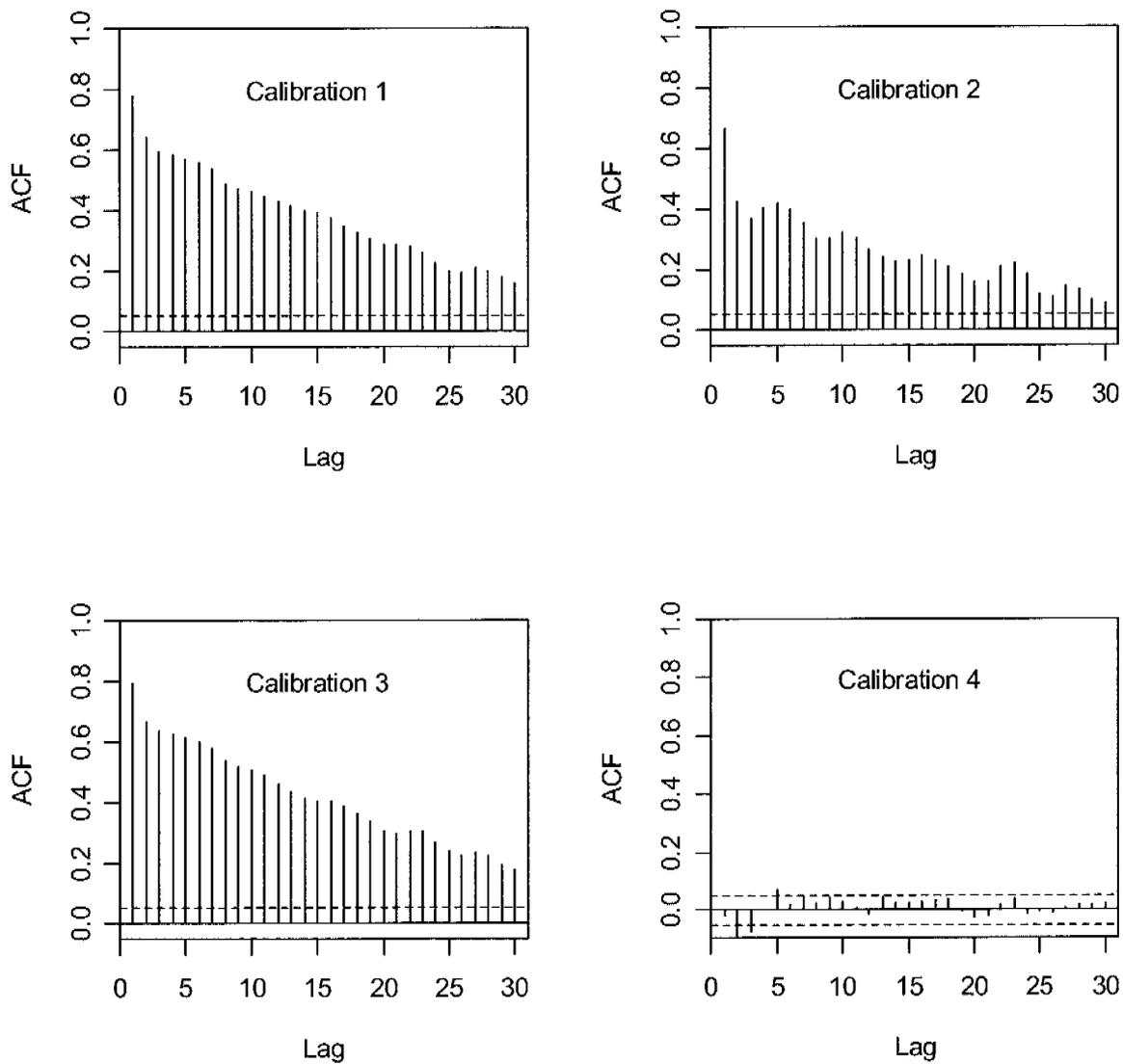


Figure 2.4: Autocorrelation functions of standardized residuals (calibration 1), standardized residuals of transformed model results and data (calibrations 2 and 3) and standardized observed innovations of the continuous-time autoregressive model (Calibration 4) at the maxima of the posterior distributions.

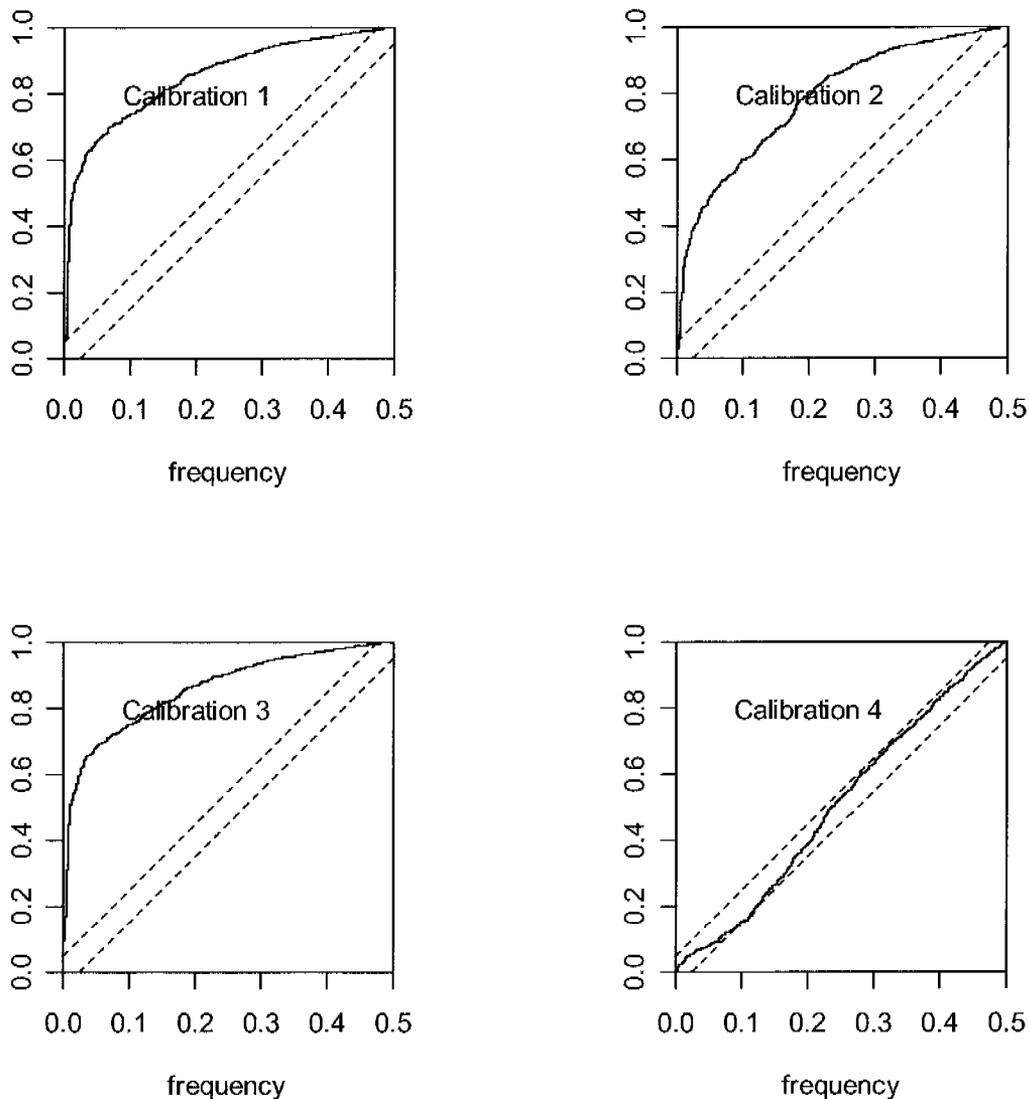


Figure 2.5: Cumulative periodograms with 95% limit lines for standardized residuals (calibration 1), standardized residuals of transformed model results and data (calibrations 2 and 3) and standardized observed innovations of the continuous-time autoregressive model (calibration 4) at the maxima of the posterior distributions.

Figure 2.6 shows histograms approximating the marginals of the posterior parameter distribution. The decrease in CN2 (negative value of $a_{CN2.mgt}$; see section 2.2.2) and the increase in SOL_AWC (positive value of $a_{SOL_AWC.sol}$; see section 2.2.2) reflects the overestimation of flow in the default simulation. The high increase in CH_K2 (positive value of $a_{CH_K2.rte}$) indicates that there is a strong interaction between the river channel and

Chapter 2

groundwater. σ_{wet} is much larger than σ_{dry} as a consequence of larger fluctuations of measurements around simulation results during the wet season. The reason for this difference is the driving force of hydrologic response: it is driven by highly variable and sometimes intensive rainfall during the wet season and by groundwater feed during the dry season. The characteristic correlation time during the dry season, τ_{dry} , is nearly 10 times longer than during the wet season, τ_{wet} . This is caused by the long extension of the dependence of model results on the “initial” value at the beginning of the dry season. Because of the high temporal dynamics of the input during the wet season, dependence of errors is much weaker. The marginal distributions of the parameters are quite narrow compared to their prior distributions (see Table 2.2). This shows that only a small part of the output errors can be mapped to parameter uncertainty. Our simulation with this narrow parameter distribution still does not show a serious violation of the statistical assumptions, as the error model adds sufficient uncertainty to the model results to “explain” the deviations from the simulations. This demonstration of the compatibility of the statistical model with the data is satisfying and gives us confidence into the uncertainty estimates of model predictions. However, the posterior distribution of the model parameters may be multi-modal and similarly good predictions may be possible within other local maxima of the posterior. This makes it difficult to interpret the posterior marginals as realistic uncertainty estimates of the parameters (despite the realistic uncertainty estimates of the predictions dominated by the error model).

In order to compare the traditional hydrological calibration method (calibration 1 with likelihood function 2.9) with the autoregressive error model approach, optimal SWAT aggregate parameters for calibration 1 are marked by an asterisk (“*”) in Figure 2.6. In addition, the quality of discharge calibration is compared for these two approaches in Table 2.3 using different performance measures. As can be seen from Figure 2.6, there are significant differences in aggregate parameter values between these two approaches. This is caused by different choices of the likelihood function. Although maximizing the likelihood function of the traditional method is equivalent to maximizing the *Nash-Sutcliffe* coefficient, in Table 2.3, the values of the *Nash-Sutcliffe* coefficient and of R^2 are only slightly smaller for the autoregressive error model than the traditional method. On the other hand, obviously, the log posterior densities for both calibration and validation period are smaller for calibration 1 than those of the autoregressive error model. The comparison of numerical criteria cannot be used to assess these two approaches. The

essential difference is that the distributional assumptions of the independent error model are strongly violated. This is not the case for the autoregressive error model.

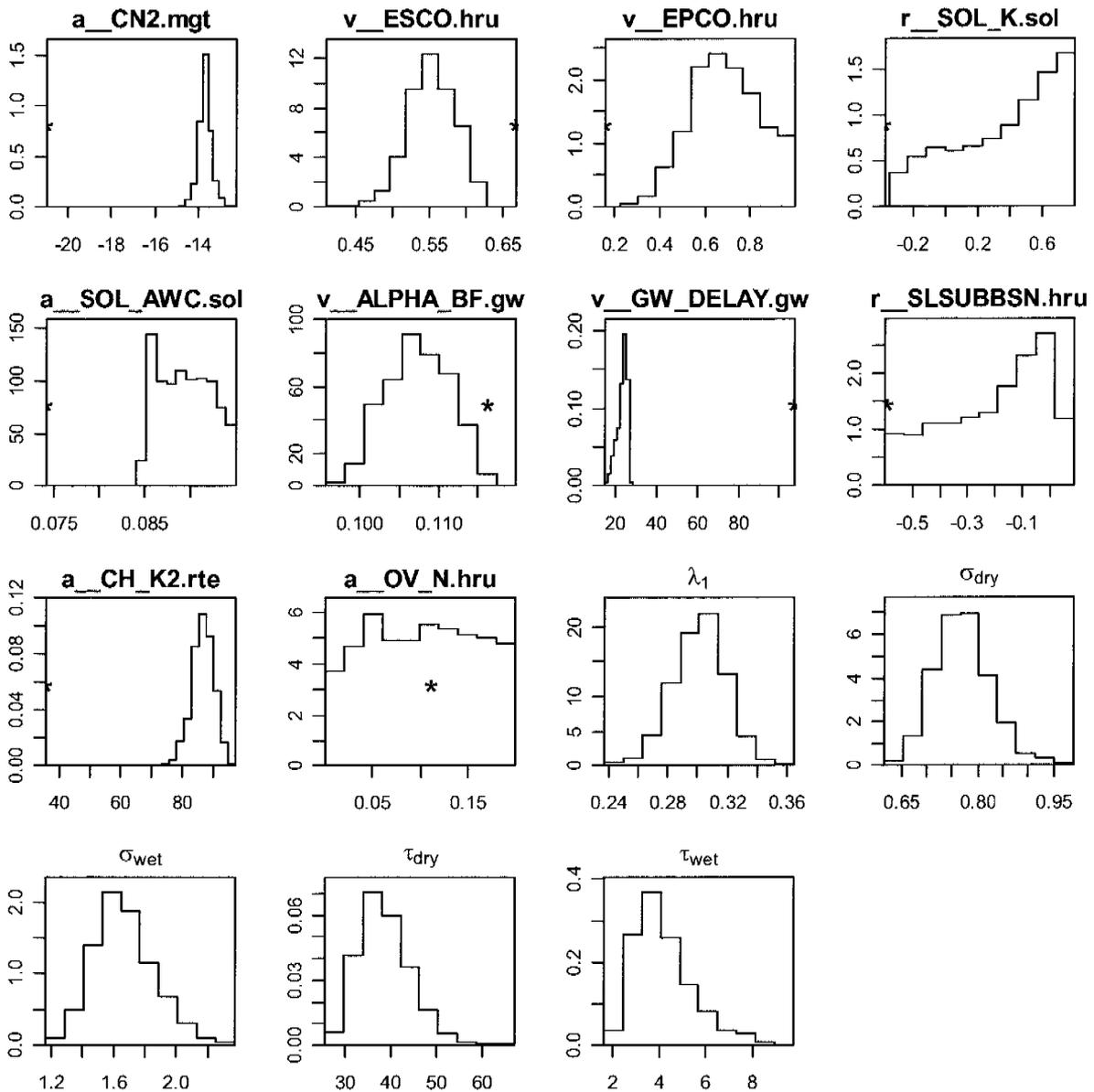


Figure 2.6: Histograms approximating the marginals of the posterior parameter distribution. Asterisks (“*”) indicate the optimized aggregate parameter values by traditional method (calibration 1).

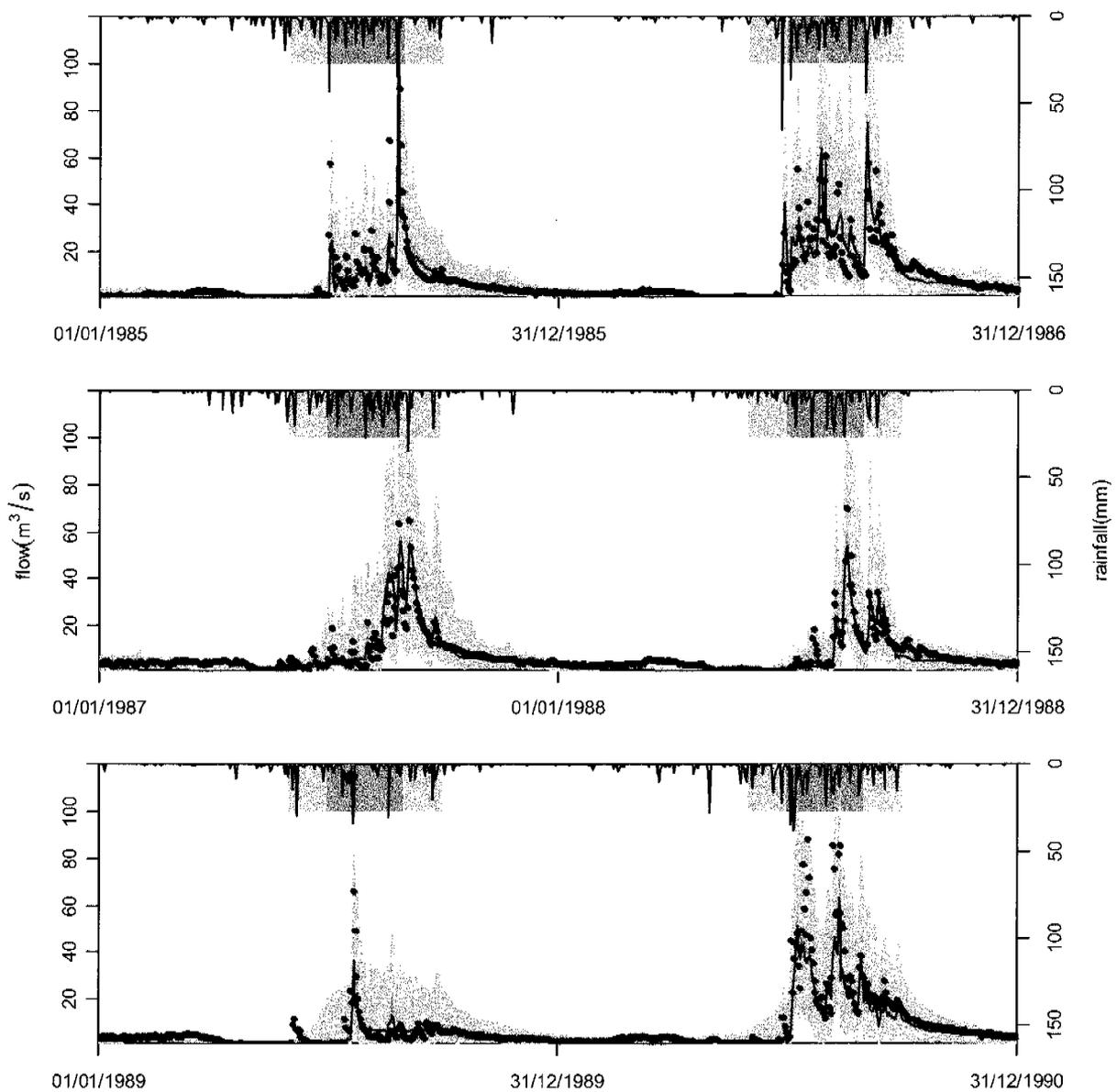


Figure 2.7: 95% prediction uncertainty bands associated with parameter uncertainty (dark shaded area), and with parameter uncertainty and continuous-time autoregressive model (light shaded area) during the calibration period (top and middle) and validation period (bottom). The dots correspond to the observed discharge at Xiahui station and the line stands for the simulated discharge at the maximum of the posterior distribution. The light and dark shaded areas on the top of each plot indicate the transition periods and wet seasons, and the line on the top represents rainfall series.

Table 2.3: Performance comparison between traditional method and autoregressive error model

		Nash-Sutcliffe	R ²	Log posterior density
Traditional method	Calibration period	0.82	0.80	-4872
	validation period	0.81	0.80	-3507
Autoregressive error model	Calibration period	0.77	0.78	-3078
	validation period	0.73	0.81	-3217

Figure 2.7 shows the best model prediction (at the maximum of the posterior density) and 95% prediction uncertainty bands associated with parameter uncertainty only (dark shaded area) and with total uncertainty (parameter uncertainty and uncertainty described by the continuous-time autoregressive error model; light shaded area) both during the calibration and validation periods. As can be seen, although the prediction uncertainty from parameter uncertainty (dark shaded area) is very narrow (it covers 10% of the measured data points during the calibration period), the 95% uncertainty bands representing total uncertainty brackets most of the observations (85%). This indicates that our proposed approach can mimic the prediction uncertainty. Despite the severe violation of statistical assumptions demonstrated in Figures 2.2, 2.3 and 2.4 for the independent error model (simulation 3), the prediction uncertainty estimates are quite similar to those of the autoregressive error model. This is caused by the dominance of the error of the additive error model over the error caused by uncertain model parameters. There is no reason that the independent error model reproduces the standard deviation of the residuals less adequately than the autoregressive error model. The inadequate description of the correlation of residuals makes individual realizations of model predictions unrealistic (particularly during the dry season), but the 95% prediction uncertainty bands are not seriously affected.

Uncertainty of predicted river discharge, quantified by its standard deviation, is mainly dependent on the predicted discharge. This allows us to approximately summarize prediction uncertainty as a function of predicted discharge. Figure 2.8 shows how the standard deviation of the predicted discharge increases with the predicted discharge. The figure distinguishes the dry and wet seasons. Prediction uncertainty is significantly larger during the wet season. As discussed earlier, parameter uncertainty contributes only to a small part to total uncertainty. The approximate relationship shown in Figure 2.8 is very precise for the total uncertainty, whereas there is more scatter around the relationship for the parameter uncertainty only.

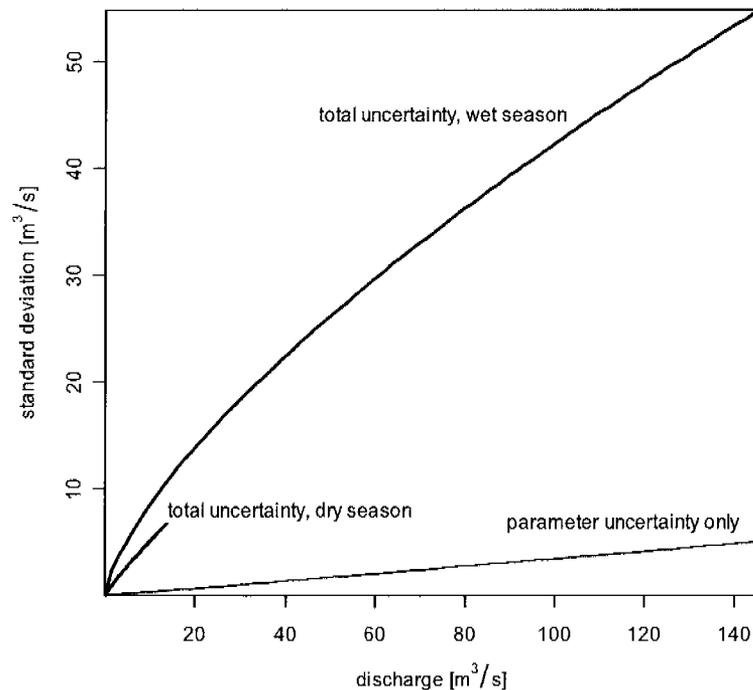


Figure 2.8: Approximate standard deviation of model predictions as a function of predicted discharge. “Total uncertainty” refers to uncertainty due to parameter uncertainty and the autoregressive error model. “Parameter uncertainty only” refers to deterministic model results based on uncertain (posterior) model parameter without consideration of the autoregressive error model that accounts for input, model structure and measurement error.

2.6 Summary and Conclusions

While calibrating the hydrologic model for the Chaohe Basin, we encountered the following problems:

- The distributed parameters are (obviously) not identifiable from the data of a single watershed outlet station.
- Multiple local maxima of the posterior make it difficult to find the “true” maximum and get a reasonable uncertainty estimate of model parameters.

- Measurement errors of input and response and model structure deficits lead to errors which could not be made homoscedastic by a Box-Cox transformation of data and model output or by the use of different error variances for the dry and wet seasons.
- As input and model structure errors are propagated through a model with memory effect due to water storage, residuals are substantially correlated even for the best model fit.
- The structure of the residuals was significantly different during the wet (larger variance, less autocorrelation) and the dry season (smaller variance, long range of higher autocorrelation). This is caused by the highly variable input during the wet season and by the sensitivity of the model results to the “initial” condition at the beginning of the dry season and the storage release parameters of the model.

These problems could be overcome by applying the following techniques:

- Non-identifiability of distributed parameters was overcome (i) by the use of aggregate parameters that use the spatial structure of distributed parameters based on prior information and (ii) by applying a Bayesian inference technique that does not rely on parameter identifiability. Technically, this was implemented with the aid of an interface program that has a very high flexibility in modifying parameters on SWAT input files (Reichert, 2006; Yang et al., 2005).
- The effect of measurement errors of input and response and errors in model structure were described by a continuous-time autoregressive error model. This model was applied as follows:
 - It was used to describe an additive error between Box-Cox transformed model results and data.
 - It was used with different error variances during wet and dry seasons. Together with the Box-Cox transformation, this led to a reasonably good homoscedasticity of the standardized residuals between the transformed model result and data. (The combination of these two measures was necessary to achieve approximate homoscedasticity and normality.)
 - The characteristic correlation time was chosen differently for wet and dry seasons. A significantly longer correlation time during the dry season than during the wet

season led to very low autocorrelation in the standardized observed innovations of the autoregressive error model.

- Numerically, Bayesian inference was done by a Metropolis-Hastings Markov Chain algorithm that was started at a numerical approximation to the maximum of the posterior density calculated by an implementation of the shuffled complex global optimization algorithm (Gelman et al., 1995; Duan et al., 1992, 1993, 1994; Reichert, 2005).
- The advantage of the autoregressive error model over the traditional calibration method is shown by comparing the results of both approaches.

In contrast to a discrete-time autoregressive error model, the continuous-time autoregressive model seems conceptually more satisfying as a description of the effects of input and model structural errors that are of a continuous nature. The residual diagnostics demonstrated that the model application is consistent with the underlying statistical assumptions. However, despite the obvious violation of the statistical assumptions by the independent error model, both error models led to similar prediction uncertainty estimates.

This study also demonstrates that prediction uncertainty in hydrological modelling can hardly be described by parameter uncertainty only. Our technique provides a statistical description of the effect of input, model structure, and output uncertainty on the model results. More research is needed, however, to separate these error sources and to get a description that better addresses their cause in addition to their effect.

We hope that the development of this technique and its provision in a generally applicable system analysis program (Reichert, 2005) will stimulate the application of consistent uncertainty analyses in hydrological modelling. This provision of prediction uncertainty estimates on a routinely basis could increase the awareness of decision makers about scientific uncertainty and improve the model-based support of their decisions.

2.7 Acknowledgement

We would like to thank Prof Xia, J., Prof. Jia, S., Dr. Wang, G., and Dr. Dong, W. of Institute of Geographical Science and Natural Resources Research, CAS, and Dr. Liu, C. of Beijing Normal University for helping us with data collection, and Mr. Li, X. and Mr. Zhang X. from the Water Authority of the Fenning County for guiding our field trip.

2.8 References

- Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., Srinivasan, R., 2007. Spatially-distributed modelling of hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of Hydrology*, 333:413-430.
- Abbaspour, K.C., Johnson, A., van Genuchten, M.T., 2004. Estimation of uncertain flow and transport parameters by a sequential uncertainty fitting procedure: SUFI-2. *Vadose Zone Journal* 3(4), 1340-1352.
- Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., Srinivasan, R., 2007. Spatially-distributed modelling of hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of Hydrology*, 333:413-430.
- Allen, R.G., 1986. A Penman for all seasons. *Journal of Irrigation and Drainage Engineering-ASCE* 112(4), 348-368.
- Allen, R.G., Jensen, M.E., Wright, J.L., Burman, R.D., 1989. Operational Estimates of Reference Evapotranspiration. *Agronomy Journal* 81(4), 650-662.
- Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment - Part 1: Model development. *Journal of the American Water Resources Association* 34(1), 73-89.
- Bates, B.V., Campbell, E.P. 2001. A Markov Chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research* 37(4), 937-947.
- Best, N.G., Cowles, M.K., Vines, S.K., 1995. Convergence Diagnosis and Output Analysis software for Gibbs Sampler output: Version 0.3. Cambridge: Medical Research Council Biostatistics Unit.
- Beven, K., 2001. How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences* 5(1), 1-12.
- Beven, K., Binley, A., 1992. The Future of Distributed Models - Model Calibration and Uncertainty Prediction. *Hydrological Processes* 6(3), 279-298.
- Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* 249(1-4), 11-29.

Chapter 2

- Bicknell, B.R., Imhoff, J., Kittle, J., Jobes, T., Donigian, A.S., 2000. Hydrological Simulation Program – Fortran User's Manual, Release 12, U.S EPA.
- Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. *Journal of the Royal Statistical Society Series B*, 211-252.
- Box, G.E.P., Cox, D.R., 1982. An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association* 77(377), 209-210.
- Brockwell, P.J., Davis, R.A., 1996. *Introduction to Time Series and Forecasting*, Springer, New York.
- Brockwell, P.J., 2001. Continuous-time ARMA processes. In: Shanbhag, D.N. & Rao, C.R., (Eds.), *Stochastic Processes: Theory and Methods*, Vol. 19 of *Handbook of Statistics*, Elsevier, Amsterdam, pp. 249–276.
- Chatfield, C., 2003. *The Analysis of Time Series: An introduction (Sixth Edition)*. Chapman & hall/CRC pub, P40.
- Chen, J., Chen X., 2004. Water Balance of the SWAT Model and its Application in the Suomo Basin. *Acta Scientiarum Naturalium-Universitatis Pekinensis* 40(2) (in Chinese).
- Chow V.T., Maidment, D.R., Mays, L.W., 1988. *Applied Hydrology*. McGraw-Hill, Inc., New York, NY.
- Cowles, M.K., Carlin, B.P., 1996. Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91(434), 883-904.
- Crawford, N.H., Linsley, R.S., 1966. Digital simulation in hydrology: The Stanford Watershed Model IV. Technical Report No. 39, Department of Civil Engineering, Stanford University, Palo Alto, California.
- Cunge, J.A., 1969. On the subject of a flood propagation method (Muskingum method). *J. Hydraulics Research* 7(2), 205-230.
- Di Luzio, M., Srinivasan, R., Arnold, J.G., Neitsch, S.L., 2002. ArcView Interface for SWAT2000, Blackland Research&Extension Center, Texas Agricultural Experiment Station and Grassland, Soil and Water Research Laboratory, USDA Agricultural Research Service.
- Duan, Q.Y., Gupta, V.K., Sorooshian, S., 1993. Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization. *Journal of Optimization Theory and Applications* 76(3), 501-521.

- Duan, Q.Y., Sorooshian, S., Ibbitt, R.P., 1988. A maximum likelihood criterion for use with data collected at unequal time intervals. *Water Resources Research* 24(7): 1163-1173.
- Duan, Q.Y., Sorooshian, S., Gupta, V.K., 1992. Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models. *Water Resources Research* 28(4), 1015-1031.
- Duan, Q.Y., Sorooshian, S., Gupta, V.K., 1994. Optimal Use of the SCE-UA Global Optimization Method for Calibrating Watershed Models. *Journal of Hydrology* 158(3-4), 265-284.
- Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N., Turcotte, R., eds., 2003. *Calibration of Watershed Models*, American Geophysical Union, Washington, DC, USA.
- Gelman, S., Carlin, J.B., Stren, H.S., Rubin, D.B., 1995. *Bayesian Data Analysis*, Chapman and Hall, New York, USA.
- Hargreaves, G.H., Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture* 1, 96-99.
- Huang, Q., Zhang W., 2004. Improvement and Application of GIS-based distributed SWAT Hydrological Modeling on High Altitude, Cold, Semi-arid Catchment of Heihe River Basin, China. *Journal of Nanjing Forestry University (Natural Sciences Edition)* 28(2), 22-26 (in Chinese).
- Hu, Y., Cheng, S., Jia, H., 2003. Hydrologic Simulation in NPS Models: Case of Application of SWAT in Luxi Watershed. *Research of Environmental Sciences* 16(5), 29-32 (in Chinese).
- Jia, H., Cheng, S., 2002. Spatial and dynamic simulation for Miyun Reservoir waters in Beijing. *Water Science and Technology* 46(11-12), 473-479.
- Kavetski, D, Franks, S.W., Kuczera, G., 2003. Confronting Input Uncertainty in Environmental Modelling. In: Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N. and Turcotte, R. (Eds.), *Calibration of Watershed Models*. American Geophysical Union, Washington, DC., USA, pp 49-68.
- Kloeden, P.E., Platen, E., 1992. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, Germany.
- Kuczera, G., 1983. Improved parameter inference in catchment models - 1. Evaluating parameter uncertainty. *Water Resources Research* 19(5), 1151-1162.
- Kuczera, G., Parent, E. 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology* 211(1-4), 69-85.

Chapter 2

- Lamb, R., 1999. Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation, *Water Resources Research* 35(10), 3103-3114.
- Liu, Z.Y., 2004. Application of GIS-based distributed hydrological model to flood forecasting. *Shuli Xuebao* 5, 70-75 (in Chinese).
- Monteith, J.L., 1965. Evaporation and the environment. In: *The state and movement of water I living organisms*, XIXth Symposium. Soc. For Exp. Biol., Swansea, Cambridge University Press, 205-234.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles, *Journal of Hydrology*, 10 (3), 282–290.
- Neitsch, S.L., Arnold, J.G, Kiniry, J.R., Williams, J.R., 2001. *Soil and Water Assessment Tool User's Manual*, version 2000. Grassland, Soil and Water Research Laboratory, Agricultural Research Service, Blackland Research Center, Texas Agricultural Experiment Station.
- Priestley, C.H.B., Taylor, .R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Weather. Rev.* 100, 81-92.
- Reichert, P., 2005. UNCSIM - A computer programme for statistical inference and sensitivity, identifiability, and uncertainty analysis, In: Teixeira, J.M.F. and Carvalho-Brito, A.E. (Eds.), *Proceedings of the 2005 European Simulation and Modelling Conference (ESM 2005)*, Oct. 24-26, Porto, Portugal, EUROSIS-ETI, pp. 51-55.
- Reichert, P., 2006. A standard interface between simulation programs and systems analysis software. *Water Science and Technology* 53(1), 267-275.
- Reichert, P., Schervish, M., Small, M.J., 2002. An efficient sampling technique for Bayesian inference with computationally demanding models. *Technometrics* 44(4), 318-327.
- Schaap, M.G., Bouten, W., 1996. Modeling water retention curves of sandy soils using neural networks. *Water Resources Research* 32(10), 3033-3040.
- Schaap, M.G., Leij, F.J., van Genuchten, M.T., 2001. ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology* 251(3-4), 163-176.
- Shi, X.Z., Yu, D.S., Warner, E.D., Pan, X.Z., Petersen, G.W., Gong, Z.G, Weindorf, D.C., 2004. Soil Database of 1:1,000,000 Digital Soil Survey and Reference System of the Chinese Genetic Soil Classification System. *Soil Survey Horizons* 45, 129–136.

- Soil Conservation Service, 1972. Section 4: Hydrology. In: National Engineering Handbook. SCS.
- Tomassini, L., Reichert, P., Künsch, H.-R., Buser, C., Borsuk, M.E., 2007. A smoothing algorithm for estimating stochastic, continuous-time model parameters and an application to a simple climate model. Submitted.
- Vrugt, J. A., Bouten, W., 2002. Validity of first-order approximations to describe parameter uncertainty in soil hydrologic models. *Soil Science Society of America Journal* 66(6), 1740-1751.
- Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modelling: Combining the strengths of global optimization and data assimilation, *Water Resources Research* 41, art. no. W01017
- Vrugt, J. A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* 39(8), art. no.-1201.
- Wang, G.S, Xia, J., Tan, G., Lv, A.F., 2002. A Research on Distributed Time Variant Gain Model: a Case Study on Chaohe River Basin. *Progress in Geography* 21(6), 573-582 (in Chinese).
- Wang, X., Li, T., Xu, Q., He, W., 2001. Study of the distribution of non-point source pollution in the watershed of the Miyun Reservoir, Beijing, China. *Water Science and Technology*, 44(7), 35-40.
- Wang, Z.G., Liu, C.M, Huang, Y.B., 2003. The Theory of SWAT Model and its Application in Heihe Basin. *Progress in Geography* 22(1), 79-86 (in Chinese).
- Williams, J.R., 1969. Flood routing with variable travel time or variable storage coefficients. *Transactions of the ASAE* 12(1), 100-103.
- Williams, J.R., 1995. The EPIC model. In Vijay P. Singh (Eds.), *Computer Models of Watershed Hydrology*. Water Resources Publication, US, pp909-1000.
- Yang, J., Abbaspour, K.C., Reichert, P., 2005, Interfacing SWAT with Systems Analysis Tools: A Generic Platform. In: Srinivasan, R., Jacobs, J., Day, D., Abbaspour, K. (Eds.), 2005 3rd International SWAT Conference Proceedings, July 11-15, Zurich, Switzerland, pp. 169-178.
- Yapo, P. O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data. *Journal of Hydrology* 181(1-4), 23-48.

Chapter 2

- Young, R.A., Onstad, C.A., Bosch, D.D., Anderson, W.P., 1989. AGNPS: A nonpoint-source pollution model for evaluating agricultural watersheds. *Journal of Soil and Water Conservation* 44(2), 168-173.
- Zellner, A., 1971. *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, Inc.
- Zhang, X.S., Hao, F.H., Yang, Z.F., Cheng, H.G., Li, D.H., 2003a. Runoff and Sediment Yield Modeling in Meso-scale Watershed Based on SWAT Model. *Research of Soil and Water Conservation* 10(4), 38-42 (in Chinese).
- Zhang, X.S., Hao, F.H., Cheng, H.G., Li, D.F., 2003b. Application of SWAT model in the upstream watershed of the Luohe River. *Chinese Geographical Science* 13(4), 334-339.
- Zhang, L.N., Li, X.B., Wang, Z.H., Li, X., 2004. Study and Application of SWAT model in the Yuzhou Reservoir Basin. *Hydrology* 22(3), 4-8 (in Chinese).
- Zhao, R.J., 1992. The Xinanjiang Model Applied in China. *Journal of Hydrology* 35(1-4), 371-381.
- Zhao, R.J., Liu, X.R., 1995. The Xinanjiang Model. In: Singh, V.P. (Eds.), *Computer Models of Watershed Hydrology*. Water Resources Publication, US, pp215-232.

3 Bayesian Uncertainty Analysis in Distributed Hydrologic Modelling: A Case Study in the Thur River Basin (Switzerland)

Jing Yang, Peter Reichert, Karim C. Abbaspour

(Submitted to Water Resources Research)

Abstract

Calibration and uncertainty analysis in hydrologic modelling are affected by measurement errors in inputs and response and errors in model structure. Recently, extending similar approaches in discrete-time, a continuous-time autoregressive error model was proposed for statistical inference and uncertainty analysis in hydrologic modelling. The major advantages over discrete-time formulation are the use of a continuous-time error model for describing continuous processes, the possibility of accounting for seasonal variations of parameters in the error model, the easier treatment of missing data or omitted outliers, and the opportunity for continuous-time predictions. The model was developed for the Chaohe Basin in China and had some features specific for this semi-arid climatic region (in particular the seasonal variation of parameters in the error model in response to seasonal variation in precipitation). This paper tests and extends this approach with an application to the Thur river basin in Switzerland, which is subject to completely different climatic conditions. This application corroborates the general applicability of the approach, but also demonstrates the necessity of accounting for the heavy tails in the distributions of residuals and innovations. This is done by replacing the normal distribution of the innovations by a Student t distribution, the degrees of freedom of which is adapted to best represent the shape of the empirical distribution of the innovations. We conclude that with this extension the continuous-time autoregressive error model is applicable and flexible for hydrologic modelling under different climatic conditions. The major remaining conceptual disadvantage is that this class of approaches does not lead to a separate identification of model input and model structural errors. The major practical disadvantage is the high computational demand characteristic for all MCMC techniques.

Keywords

Uncertainty Analysis; Hydrologic Modeling; MCMC; Continuous-time Autoregressive Error Model; Bayesian Inference

3.1 Introduction

Due to measurement errors in input and response and errors in model structure, predictions of hydrologic models are inevitably affected by uncertainty. Hydrologic models play an important role in supporting environmental decisions, e.g. by assessing water availability, exploring vulnerability to environmental change, or predicting the effect of management measures in the watershed. Therefore, to be able to support environmental decisions under consideration of prediction uncertainty, careful analysis and quantification of uncertainty are crucial in hydrologic modelling.

A significant number of techniques have been developed to estimate parameters and assess prediction uncertainty in hydrologic modelling. These include: first-order approximation [Carrera and Neuman, 1986; Kool and Parker, 1988; Vrugt and Bouten, 2002], Bayesian inference based on importance sampling (IS) [e.g., Kuczera and Parent, 1998] or Markov Chain Monte Carlo (MCMC) [e.g., Vrugt et al., 2003, 2004; Kuczera and Parent, 1998], Generalized Likelihood Uncertainty Estimation (GLUE) [Beven and Binley, 1992], Sequential Uncertainty Fitting SUFI-2 [Abbaspour et al., 2004, 2007], Parameter Solution (ParaSol) [Van Griensven and Meixner, 2006], and Sources of Uncertainty Global Assessment using Split Samples (SUNGLASSESS) [Van Griensven and Meixner, 2006]. With respect to model results and their uncertainty bands, many applications of these techniques give similar results [Yang et al., 2007a]. However, there are differences in the statistical foundations of these techniques. Some of these techniques, such as GLUE [e.g. Beven and Binley, 1992] or SUFI-2 [Abbaspour et al., 2004, 2007], apply a philosophy that is not based on a statistical foundation [see e.g. Beven, 2006 for an explanation]. On the other hand, applications of techniques that are based on a statistical foundation often use statistical assumptions, such as independent errors, which are obviously violated [e.g., Vrugt et al., 2003]. The violation of the statistical assumptions, particularly of homoscedasticity and independence of errors, is clearly and visually demonstrated by Vrugt et al. [2005]. Under such strong violations of the statistical assumptions, the derived parameter and prediction uncertainties are unreliable. As this is not a problem of the statistical inference procedure but of the formulation of the likelihood function, we think that the key to solving this problem is to improve the formulation of the likelihood function, rather than the development of new inference techniques with a poor conceptual foundation. The focus of such an improvement

must be on the inclusion of input and model structure uncertainty in addition to parameter and output errors.

Input and model structure uncertainty can be addressed by explicitly including these uncertainty sources into the formulation of the likelihood function, or by formulating an error model that jointly accounts for the effects of all uncertainty sources. There has been recent progress in this research field. Kavetski et al. [2006] explicitly takes into account input and output uncertainty in the formulation of the likelihood function. However, this approach does not consider the errors in model structure. Vrugt et al. [2005] presents a Simultaneous Optimization and Data Assimilation (SODA) procedure to separate parameter uncertainty from input and model structural uncertainty. The main characteristic of SODA is to make the deterministic hydrologic model stochastic and combine parameter with state estimation. The difficulty of this approach is that it involves state estimate (which is equivalent to the estimation of many additional parameters) in addition to parameter estimation. This increases the computational burden and requires modifications to existing simulation programs. A simpler approach to address input and model structural errors is by adding a “bias” or “model inadequacy” term to model output that provides a statistical description of the effect of model deficiencies on model output. This approach has recently gained attention in the literature [Kennedy and O’Hagan, 2001; Bayarri et al., 2002; Bayarri et al., 2007] in the context of interpolation (emulation) of the output of complex computer models. This approach is a more general formulation of the use of autoregressive error models to account for the effect of all error sources on the output of time-series models, which has been applied frequently in hydrological modelling [see e.g. Kuczera, 1983; Bates and Campbell, 2001]. Yang et al. [2007b] further developed this discrete-time overall additive autoregressive error model into a continuous-time additive autoregressive model and successfully applied it in the Chaohe Basin in China with the hydrological simulation program implemented in the Soil and Water Assessment Tool (SWAT) [Arnold et al., 1998]. This approach is an extension of the approach proposed by Duan et al. [1988] for unequally spaced data. In contrast to discrete-time autoregressive error models, the continuous-time autoregressive model seems more satisfying because it can better describe the effects of input and model structural error that are of a continuous time nature, it makes it easier to describe seasonal dependence of error model properties, it eliminates the problems associated with missing data or omitted outliers, and it offers the opportunity for continuous-time predictions [Yang et al.,

2007b]. This paper further tests this procedure by applying it to the Thur river basin in Switzerland. This is important to corroborate the universal applicability of the procedure under different climatic conditions and to gain experience with typical values of the parameters of the error model. In addition, we will extend the continuous-time additive autoregressive error model by relaxing the assumption of normally distributed innovations to t-distributed innovations to account for the heavy tails of the distributions of innovations observed in the application to the Thur river basin.

The remainder of this paper is organized as follows. In Section 3.2, the continuous-time autoregressive error model introduced by Yang et al. [2007b] is described and extended. Section 3.3 will briefly describe the Thur river basin and the distributed hydrologic model implemented in the Soil and Water Assessment Tool [Arnold et al., 1998]. The results of the analysis are then discussed and compared to those for the Chaohe Basin [Yang et al., 2007b] in Section 3.4. Finally, a summary with conclusions is provided in Section 3.5.

3.2 Bayesian inference for a continuous-time autoregressive error model

3.2.1 Bayesian Inference

A deterministic hydrologic model can be written in the form of a function

$$\mathbf{y}^M(\boldsymbol{\theta}) = (y_{t_0}^M(\boldsymbol{\theta}), y_{t_1}^M(\boldsymbol{\theta}), \dots, y_{t_n}^M(\boldsymbol{\theta})) \quad (3.1)$$

where $y_{t_i}^M(\boldsymbol{\theta})$ represents the model output at time t_i as a function of the model parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$, and M indexes the model.

According to Bayes' theorem, the probability density of the posterior parameter distribution, $f_{\boldsymbol{\theta}|Y}(\boldsymbol{\theta}|\mathbf{y}^{\text{obs}})$, is derived from the prior density, $f_{\boldsymbol{\theta}_{\text{pri}}}(\boldsymbol{\theta})$, the likelihood function of the model, $f_{Y^M|\boldsymbol{\theta}}(\mathbf{y}^{\text{obs}}|\boldsymbol{\theta})$, and data, \mathbf{y}^{obs} , according to

$$f_{\boldsymbol{\theta}|Y}(\boldsymbol{\theta}|\mathbf{y}^{\text{obs}}) = \frac{f_{Y^M|\boldsymbol{\theta}}(\mathbf{y}^{\text{obs}}|\boldsymbol{\theta}) \cdot f_{\boldsymbol{\theta}_{\text{pri}}}(\boldsymbol{\theta})}{\int f_{Y^M|\boldsymbol{\theta}'}(\mathbf{y}^{\text{obs}}|\boldsymbol{\theta}') f_{\boldsymbol{\theta}_{\text{pri}}}(\boldsymbol{\theta}') d\boldsymbol{\theta}'} \quad (3.2)$$

Numerically, there are two generic Monte Carlo approaches to approximate the posterior parameter distribution (Eq. 3.2), i.e., Markov Chain Monte Carlo (MCMC) and Importance Sampling (IS) [Gelman et al., 1995].

In hydrology, the likelihood function is often constructed by assuming the residuals between the observations, \mathbf{y}^{obs} , and model results, $\mathbf{y}^M(\boldsymbol{\theta})$, are identically, independently and normally distributed:

$$f_{\mathbf{y}^M|\boldsymbol{\theta}}(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=0}^n \left[\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{[y_{t_i} - y_{t_i}^M(\boldsymbol{\theta})]^2}{\sigma^2}\right) \right] \quad (3.3)$$

However, due to measurement errors in the model inputs and response and errors in model structure [Yang et al., 2007b], this assumption is usually not satisfied and residuals are often heteroscedastic and autocorrelated. Therefore, in order to correctly apply Bayesian inference, the likelihood function must either address these errors explicitly or contain an autocorrelated component of residuals to describe their effect on model output.

3.2.2 The additive continuous-time autoregressive error model

As an extension of the discrete-time autoregressive error models introduced earlier [e.g. Kuczera, 1983; Bates and Campbell, 2001], an additive continuous-time autoregressive error model was introduced by Yang et al. [2007b]. This model can account for heteroscedasticity and autocorrelation of residuals and it can easily handle missing data or omitted outliers. Briefly, the likelihood function is constructed as follows:

For an autocorrelated random time series E_{t_i} representing the effect of input, model structure and output errors we assume the probability density

$$f_{E_{t_0}}(\varepsilon_{t_0}) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{\varepsilon_{t_0}^2}{\sigma^2}\right) \quad (3.4)$$

$$f_{E_{t_i}|E_{t_{i-1}}}(\varepsilon_{t_i}|\varepsilon_{t_{i-1}}) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma \sqrt{1 - \exp\left(-2\frac{t_i - t_{i-1}}{\tau}\right)}} \exp\left(-\frac{1}{2} \frac{\left(\varepsilon_{t_i} - \varepsilon_{t_{i-1}} \exp\left(-\frac{t_i - t_{i-1}}{\tau}\right)\right)^2}{\sigma^2 \left(1 - \exp\left(-2\frac{t_i - t_{i-1}}{\tau}\right)\right)}\right)$$

Chapter 3

where σ is the asymptotic standard deviation of the errors and τ the characteristic correlation time. The assumption here is that the random disturbances, sometimes called innovations [Chatfield, 2003],

$$I_{t_i} = E_{t_i} - E_{t_{i-1}} \exp\left(-\frac{t_i - t_{i-1}}{\tau}\right) \quad (3.5)$$

rather than E_{t_i} , are independent and normally distributed. Keeping the asymptotic standard deviation of the errors E_{t_i} at σ , the innovations must have standard deviations of

$$\sigma_{I_{t_i}} = \sigma \sqrt{1 - \exp\left(-2\frac{t_i - t_{i-1}}{\tau}\right)} \quad (3.6)$$

They reach σ if the time difference between two observations is large compared to the characteristic correlation time, τ , and they are significantly smaller if succeeding observations are within that time or even closer. This error model is the analytical solution of an Ornstein-Uhlenbeck stochastic process [e.g., Kloeden and Platen, 1992]. The same process was used to describe continuous, time-dependent model parameters in Tomassini et al. [2007]. Note that the formulation of the continuous-time error model (3.4) is similar to the approach suggested by Duan et al. [1988] for use with unequally spaced data. However, there is an essential difference between the two approaches: a decreasing temporal distance of measurement points in our error model leads not only to an increase of the correlation, but also to a decrease in the standard deviation of the error term. This guarantees continuity of the process realizations.

Combining the deterministic hydrologic model (3.1) with the Box-Cox transformation [Box and Cox, 1964, 1982] and the error model (3.4), the likelihood function of the continuous-time autoregressive model can be written as:

$$f_{Y^M|\Theta}(\mathbf{y}|\Theta) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{[g(y_{t_0}) - g(y_{t_0}^M(\Theta))]^2}{\sigma^2}\right) \cdot \left| \frac{dg}{dy} \right|_{y=y_{t_0}} \left| \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma \sqrt{1 - \exp\left(-2 \frac{t_i - t_{i-1}}{\tau}\right)}} \exp\left(-\frac{1}{2} \frac{\left[\begin{array}{l} g(y_{t_i}) - g(y_{t_i}^M(\Theta)) \\ - [g(y_{t_{i-1}}) - g(y_{t_{i-1}}^M(\Theta))] \exp\left(-\frac{t_i - t_{i-1}}{\tau}\right) \end{array} \right]^2}{\sigma^2 \left(1 - \exp\left(-2 \frac{t_i - t_{i-1}}{\tau}\right)\right)}\right)} \cdot \left| \frac{dg}{dy} \right|_{y=y_{t_i}} \right| \quad (3.7)$$

where the function g represents the Box-Cox transformation with parameters λ_1 and λ_2 :

$$g(y) = \begin{cases} \frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & \lambda_1 \neq 0 \\ \ln(y + \lambda_2) & \lambda_1 = 0 \end{cases}, \quad g^{-1}(z) = \begin{cases} (\lambda_1 z + 1)^{1/\lambda_1} - \lambda_2 & \lambda_1 \neq 0 \\ \exp(z) - \lambda_2 & \lambda_1 = 0 \end{cases}, \quad \frac{dg}{dy} = (y + \lambda_2)^{\lambda_1 - 1} \quad (3.8)$$

In order to test the statistical assumptions of the likelihood function (3.7), a test should be made for the empirical distribution of the standardized observed innovations of the transformed observations $g(y_{t_i}^{\text{obs}})$ and the transformed model results $g(y_{t_i}^M(\Theta))$:

$$i_{t_i}(\Theta, \mathbf{y}^{\text{obs}}) = \frac{g(y_{t_i}^{\text{obs}}) - g(y_{t_i}^M(\Theta)) - (g(y_{t_{i-1}}^{\text{obs}}) - g(y_{t_{i-1}}^M(\Theta))) \exp\left(-\frac{t_i - t_{i-1}}{\tau}\right)}{\sigma \sqrt{1 - \exp\left(-2 \frac{t_i - t_{i-1}}{\tau}\right)}} \quad (3.9)$$

The suggested tests [c.g. Kuczera, 1983; Bates and Campbell, 2001; Yang et al., 2007b] include plots of time series of innovations, of the autocorrelation function of innovations, of the cumulative periodogram, and of a normal quantile-quantile plot of the innovations.

3.2.3 Error model Extension

To be able to account for heavy tails of the innovations, we extend the assumption of normally distributed, independent innovations in Eq. (3.4) to independent t-distributions with the same standard deviations, i.e.

$$f_{\varepsilon_{t_0}}(\varepsilon_{t_0}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi(\nu-2)}} \frac{1}{\sigma} \left(1 + \frac{\varepsilon_{t_0}^2}{(\nu-2)\sigma^2}\right)^{-\frac{\nu+1}{2}}$$

$$f_{\varepsilon_{t_i}|\varepsilon_{t_{i-1}}}(\varepsilon_{t_i}|\varepsilon_{t_{i-1}}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi(\nu-2)}} \frac{1}{\sigma\sqrt{1-\exp\left(-2\frac{t_i-t_{i-1}}{\tau}\right)}} \left(1 + \frac{\left(\varepsilon_{t_i} - \varepsilon_{t_{i-1}} \exp\left(-\frac{t_i-t_{i-1}}{\tau}\right)\right)^2}{(\nu-2)\sigma^2\left(1-\exp\left(-2\frac{t_i-t_{i-1}}{\tau}\right)\right)}\right)^{-\frac{\nu+1}{2}}$$

for $\nu > 2$ (3.10)

where Γ denotes the gamma function and ν the degrees of freedom of the t-distribution (note that the degrees of freedom of the t distribution must be larger than 2 in order to guarantee the existence of the standard deviation).

Accordingly, the likelihood function is adapted to:

$$f_{y^M|\theta}(y|\theta, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi(\nu-2)}} \frac{1}{\sigma} \left(1 + \frac{[g(y_{t_0}) - g(y_{t_0}^M(\theta))]^2}{(\nu-2)\sigma^2}\right)^{-\frac{\nu+1}{2}} \cdot \left|\frac{dg}{dy}\right|_{y=y_{t_0}}$$

$$\cdot \prod_{i=1}^n \left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi(\nu-2)}} \frac{1}{\sigma\sqrt{1-\exp\left(-2\frac{t_i-t_{i-1}}{\tau}\right)}} \cdot \left(1 + \frac{[g(y_{t_i}) - g(y_{t_i}^M(\theta)) - [g(y_{t_{i-1}}) - g(y_{t_{i-1}}^M(\theta))] \exp\left(-\frac{t_i-t_{i-1}}{\tau}\right)]^2}{(\nu-2)\sigma^2\left(1-\exp\left(-2\frac{t_i-t_{i-1}}{\tau}\right)\right)}\right)^{-\frac{\nu+1}{2}} \cdot \left|\frac{dg}{dy}\right|_{y=y_{t_i}} \right]$$

(3.11)

The statistical tests to be used to assess the hypotheses of the error model are the plot of time series of innovations, autocorrelation functions of innovations, and t-distribution quantile-quantile plot of innovations.

As the degrees of freedom (ν) approaches infinity, the t distribution will approximate the normal distribution. Therefore, the additional flexibility of the error model provided by the degrees of freedom of the t distribution, extends our ability to approximate the observed

distribution of innovations, while keeping the normal distribution as a limiting case. Lowering the degrees of freedom (ν) of the t distribution leads to heavier tails, as it is often observed in hydrologic modelling. Increasing the number of degrees of freedom leads us back to the previous assumption of normally distributed innovations.

3.2.4 Uncertainty analysis procedure

Parameters to be estimated within the Bayesian framework with the autoregressive error model (Eq. 3.11) include the parameters θ of the hydrologic model, the parameters λ_1 and λ_2 of the Box-Cox transformation, the characteristic correlation time τ , the standard deviation σ , and the degrees of freedom ν of the error model. Except ν which characterizes the shape of the t distribution of the innovations, all of these parameters should be estimated jointly. This was done by applying a Markov Chain Monte Carlo (MCMC) technique to approximate the posterior distribution of these parameters. In order to avoid long burn-in periods (or even lack of convergence to the distribution) of the Markov chain, the chain was started at a numerical approximation to the maximum of the posterior distribution calculated with the aid of the shuffled complex global optimization (SCE-UA) algorithm [Duan et al., 1992]. Markov chains were run until 20,000 model runs were reached with fulfillment of the convergence criterion by the Heidelberger and Welch [Cowles and Carlin, 1996; Best et al., 1995].

The implementation of the modified likelihood function as well as the numerical realization of Bayesian inference was done in UNCSIM [Reichert, 2005].

3.3 Study Area and SWAT Model

3.3.1. Description of the study area

The Thur river basin, with a drainage area of 1,700 km², is situated in north-eastern Switzerland near the border to Germany (Figure 3.1). Mean elevation of the watershed is about 769 meters above sea level and mean slope is around 7.5°. The climate of the watershed is the pre-alpine/alpine climate, which is characterized by moderate winters in hilly dissected terrain area, cold winters in mountainous areas and summer seasons with relatively large annual temperature variations. Mean monthly temperature ranges from about 10 °C to 25 °C in the summer and from -15 °C to 7 °C during the winter. The average precipitation is 1,460 mm year⁻¹ with high precipitation (about 2,200-2,500 mm year⁻¹) in the mountain area and about 1,000 mm year⁻¹ in

Chapter 3

the lower (sub-mountain) part of the watershed, and most of precipitation falls during the summer months. The mean actual evapotranspiration is about 565 mm year⁻¹, and runoff 895 mm year⁻¹. The climate data used in this study are from seventeen precipitation, eight air temperature, five solar radiation, five relative humidity, and five wind speed gages (see Figure 3.1) over 20 years (1980-2000), which were obtained from the Swiss Federal Office of Meteorology and Climatology (<http://www.meteoschweiz.ch/web/en/weather.html>). The daily discharge is available at the basin outlet (Andelfingen station) from 1991- 2000 from the Swiss Federal River Survey Program (NADUF; <http://www.naduf.ch>).

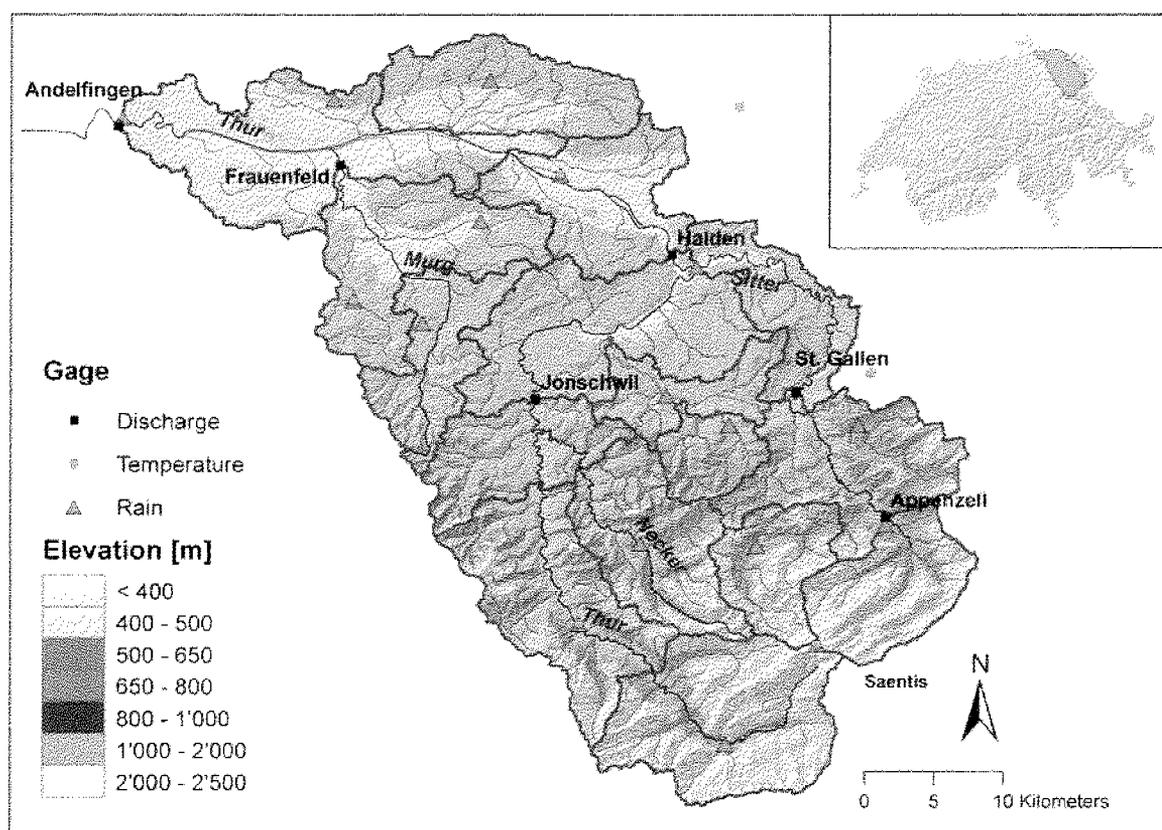


Figure 3.1: The Thur river basin with SWAT-delineated sub-basins, DEM map, river network, and meteorological stations.

The dominant land use (around 60%) in this area is agriculture, most of which are meadows for feeding cows, alpine pastures, and arable lands. Close to 30% of the total area is covered by forests, about 3% are orchards. The rest of the area is barren land, surface waters, and urban areas. Hogs and cattle are the main livestock raised in the study area.

Most of the Thur river basin is underlain by conglomerates, marl incrustations and sandstone with medium to low storage capacity and rather high permeability. Groundwater is mainly found in areas with till deposits [Gurtz et al., 1999].

3.3.2 Description of SWAT and iSWAT

Soil and Water Assessment Tool (SWAT) [Arnold et al., 1998; <http://www.brc.tamus.edu/swat>] implements a semi-distributed and semi-physically based watershed model. SWAT describes the climatic and topographic heterogeneity through subbasins based on a digital elevation map and climatic stations, while it describes the heterogeneities in land use, soil, and management practice through HRUs (Hydrologic Response Units) which consist of unique combinations of land use, soil type, and management practice within the subbasin.

At the HRU level, SWAT accounts for rainfall, interception, evapotranspiration, percolation, sediment yield, nutrient cycles, crop growth and management practice. Then, runoff, sediment yield and nutrient loads are aggregated to the subbasin level by taking the weighted average based on the areas of the HRUs. Water flow, sediment yield, and nutrient loading obtained at the subbasin level are then routed through the river system under consideration of in-stream transformation, deposition and re-mobilization processes. Channel routing is simulated using either the variable storage technique [Williams, 1969] or the Muskingum method [Cunge, 1969; Chow et al., 1988]. More detailed descriptions of the model can be found in Arnold et al. [1998] and in SWAT manuals (available at <http://www.brc.tamus.edu/swat>).

iSWAT is an interface between SWAT and an arbitrary system analysis tool that supports a simple, file-based interface [Reichert, 2006]. *iSWAT* was developed to facilitate the application of systems analysis techniques to hydrologic modelling based on using SWAT [Yang et al., 2006]. In *iSWAT*, SWAT parameters can be aggregated based on important influential factors, such as land use, soil texture, soil hydrologic group or subbasin as follows:

$$x_ \langle \text{parname} \rangle . \langle \text{ext} \rangle _ _ \langle \text{hydrogrp} \rangle _ _ \langle \text{soltext} \rangle _ _ \langle \text{landuse} \rangle _ _ \langle \text{subbsn} \rangle \quad (3.12)$$

where x represents the type of change to be applied to the parameter (v: value; a: absolute change; or r: relative change), $\langle \text{parname} \rangle$ is the SWAT parameter name; $\langle \text{ext} \rangle$ represents the extension of the SWAT input file which contains the parameter, $\langle \text{hydrogrp} \rangle$ is the identifier for the hydrologic group, $\langle \text{soltext} \rangle$ is the soil texture, $\langle \text{landuse} \rangle$ is the landuse, and $\langle \text{subbsn} \rangle$ is the subbasin number, or the crop index, or the fertilizer index. For example, $v_ \text{CN2.mgt} = 69$, will

Chapter 3

cause a global replacement of CN2 value in the management files by 69, and $r_CN2.mgt_{23,25} = 0.3$, will cause a replacement of the CN2 value in the management files associated with subbasins 23 and 25 by a value equal to their current CN2 values multiplied by 1.3, etc.

3.3.3 Choice of parameters and priors

After setting-up the project, a manual calibration and then an automatic calibration were done on some parameters of the Thur SWAT project. All the simulations in this paper are based on the calibrated project for all parameters not included in the analysis (i.e. not explicitly mentioned). To distinguish these simulations from the following new simulations, they are referred to as “previous simulations” in the following text.

The choice of parameters is based on the LH-OAT (Latin-Hypercube-One-factor-At-a-Time) method [Van Griensven et al., 2006]. LH-OAT is a global screening sensitivity analysis technique and its characteristic is that it combines the Latin Hypercube sampling [McKay et al., 1979] and OAT (One-factor-At-a-Time) method by taking the Latin-Hypercube samples as initial points for the OAT method. Based on LH-OAT, 20 aggregate SWAT parameters related to river flow were selected for calibration (Table 3.1).

Together with the parameters λ_1 , λ_2 , σ and τ of the autoregressive error model in Eq. (3.7) or (3.11), there are 24 parameters. The prior distributions of all these parameters are assumed to be independent. For the 20 aggregate SWAT parameters, uniform priors with reasonable ranges were assumed (see the last column in Table 3.1). And transformation parameters λ_1 and λ_2 are assumed to be uniformly distributed. For the parameters σ and τ , densities proportional to $1/\sigma$ and $1/\tau$ were chosen. Table 3.1 gives an overview of the parameters used for calibration and their prior distribution.

Bayesian Uncertainty Analysis in Distributed Hydrologic Modelling

Table 3.1: Selected parameters for inference and their initial values and prior distributions

Aggregate Parameter	Name and meaning of underlying SWAT parameter	Initial par. range of underlying SWAT parameter	Prior dist. of aggregate parameter
v TIMP.bsn ^{*1}	Snow pack temperature lag factor	0.307	U[0.01,1] ^{*3}
v SFTMP.bsn	Snowfall temperature	-1	U[-5,5]
v SMTMP.bsn	Snowmelt base temperature	2.585	U[-5,5]
v SMFMX.bsn	Melt factor for snow on June 21	4.473	U[0,10]
v SMFMN.bsn	Melt factor for snow on Dec 21	0.923	U[0,10]
v_MSK_CO1.bsn	Muskingum coefficient to control impact of the storage time constant for normal flow	0	U[0,10]
v_MSK_CO2.bsn	Muskingum coefficient to control impact of the storage time constant for low flow	0.2	U[0,10]
v_MSK_X.bsn	A weighting factor that controls the relative importance of inflow and outflow in determining the storage in a reach in Muskingum method	0.1	U[0,0.3]
v_CH_K1.sub	Effect hydraulic conductivity in tributary channel alluvium (mm/hr)	0.5	U[0,150]
r_CN2.mgt ^{*2}	CN2: curve number	47-73	U[-0.35,0.35]
r_CH_N2.rte	Manning roughness for main channel	0.052/0.3	U[-0.5,0.5]
v_CH_K2.rte	Effective hydraulic conductivity in main channel alluvium (mm/hr)	6.325	U[0,150]
v_ALPHA_BF.gw	Base flow alpha factor (1/day)	0.0625	U[0,1]
v_GWQMN.gw	Threshold depth of water in the shallow aquifer required for return flow to occur (mm H ₂ O)	0	U[0,5000]
v_GW_REVAP.gw	Groundwater "revap" coefficient	0.02	U[0.02,0.2]
v_GW_DELAY.gw	Groundwater delay time (days)	43.338	U[0,300]
v_CANMX.hru	Maximum canopy storage	5.275	U[0,10]
v_ESCO.hru	Soil evaporation compensation factor	0.154	U[0,1]
r_SOL_AWC.sol	Soil avail. water capacity (mm H ₂ O/mm soil)	0-0.28	U[-0.5,0.5]
r_SOL_K.sol	Soil hydraulic conductivity (mm/hr)	0.01-279.71	U[-0.8,0.8]
λ_1 ^{*4}	Transformation factor in Equation (3.7) or (3.11)		U[0,1]
λ_2 ^{*4}	Transformation factor in Equation (3.7) or (3.11)		U[0,50]
σ	Standard deviation in Equation (3.7) or (3.11)		Inv ^{*3}
τ	Characteristic correlation time of autoregressive process(days)		Inv

^{*1} "v_" in "v_TIMP.bsn" means "replace TIMP with a given value"

^{*2} "r_" in "r_CN2.mgt" means "a relative change (of the default value) of CN2" and hence r_CN2.mgt is dimensionless.

^{*3} U[x,y] represents the uniform distribution over the interval [x,y] for the given aggregate parameter; Inv denotes the probability distribution with probability density at the value x proportional to 1/x

^{*4} λ_1 and λ_2 are fixed to 0 and excluded in the final MCMC as they are very close to 0.

Except for the analysis with likelihood function (3.11), 2 additional analyses with likelihood functions (3.3) and (3.7) were also carried out as a comparison to the analysis based on the likelihood function (3.11). It is worth noting that the likelihood function (3.3) is widely used in hydrology (and many other fields) and the likelihood function (3.7) was used in Yang et al.

[2007b]. Hereafter, simulations based on analyses with likelihood functions (3.11), (3.3) and (3.7) are referred to as simulation 1, simulation 2 and simulation 3, respectively.

Obviously the initial values of storage volumes (e.g. soil water content) will influence the river flow. As we cannot specify reasonable initial values for all storage volumes considered in the model, SWAT is operated for a “warm-up period” of 6 years (1985-1990) without comparison of model results with observed data. We found that such a “warm-up period” was sufficient to minimize the effects of the initial state of SWAT variables on river flow. Furthermore, in order to verify the calibrated model parameters, the model was calibrated and tested based on the observed discharges at the basin outlet (Andelfingen station, Figure 3.1) using a split sample procedure. The data from the years 1991-1995 was used for calibration, and the data from 1996-2000 was used to test the model.

3.4 Results and Discussion

3.4.1. Results for the Thur river basin

To determine the optimum value of the degrees of freedom, ν , of the t-distribution in the likelihood function given by Eq. (3.11), we compared regression diagnostics for analyses performed with different values of ν . The comparisons were done for simulation results at the maximum of the posterior density obtained with the aid of the global optimization algorithm SCE-UA [Duan et al., 1992]. The comparisons showed that the simulation with $\nu = 8$ led to the smallest deviations of the residuals from the theoretical assumptions made by the model. These results of regression diagnostics with $\nu = 8$ are illustrated in Figure 3.2. The top panel in Figure 3.2 shows the time series of observed (circles) and simulated (line) flows. For this simulation, R^2 equals 0.80 and the Nash-Sutcliffe coefficient equals 0.77. The middle panel in Figure 3.2 shows the time series of the innovations. There seems to be no serious violation of the assumptions of independence and of distribution shape. This is further corroborated by the autocorrelation function (bottom left panel) and the t-distribution quantile-quantile plot (bottom right panel). The autocorrelations are very small except for the first order coefficient. The quantile-quantile plot in the bottom right panel demonstrates that the empirical quantiles of the innovations are in good agreement with the theoretical t-distribution quantiles.

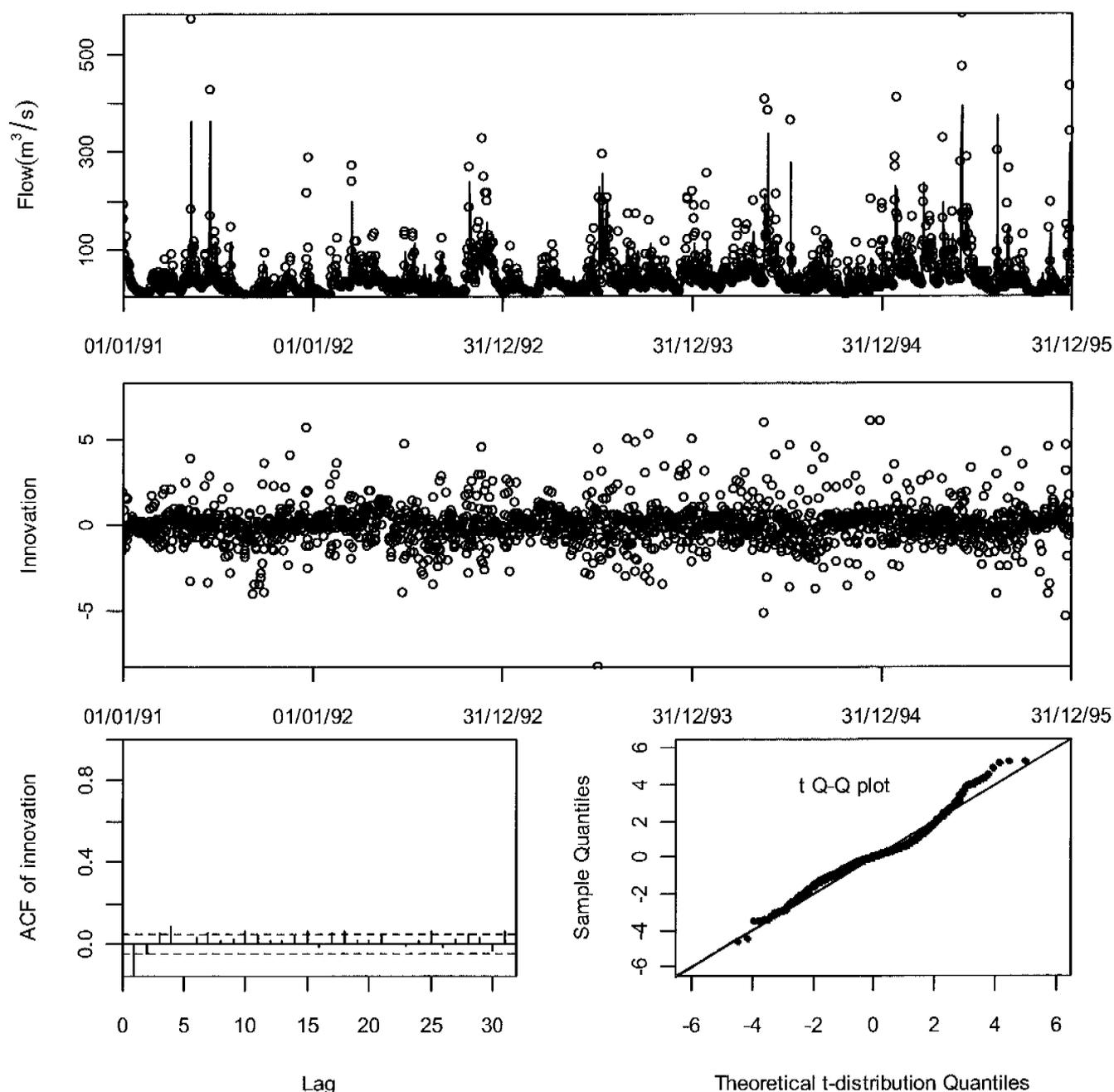


Figure 3.2: Statistics diagnostics for simulation 1 with likelihood function (3.11) with degrees of freedom 8. From top to bottom: time series of the observed (circles) and simulated (line) flows, time series of the normalized innovations, and the autocorrelation function and t -distribution quantile-quantile plot of normalized innovations.

For comparative purposes, Figures 3.3 and 3.4 show the corresponding results and diagnostics for analyses of simulations 2 and 3, respectively, and Table 3.2 lists the performances of 3 simulations at the maxima of posterior densities. The top panels in Figures 3.2, 3.3 and 3.4 give

the impression that all three simulations led to similarly good agreement with data, although simulation 2 captured several peaks better than simulations 1 and 3 (e.g., flow at 1991-8-22). This led to the highest R^2 and the Nash-Sutcliffe coefficient calculated with the simulated flow and observed flow (Table 3.2). This is because the Box-Cox transformation with $\lambda_1=0$ (the optimized λ_1 's in simulations 1 and 3 are very close to 0) puts less weight on the good approximation of high peaks to account for the lower measurement accuracy. However, the significant heteroscedasticity in the residuals of simulation 2 violates the statistical assumptions and makes its uncertainty estimates unreliable (middle panel of Figure 3.3). There are also slightly higher autocorrelation coefficients (bottom left panel of Figure 3.3), and the assumption of normally distributed residuals is severely violated especially in the tails of the distribution (bottom right panel of Figure 3.3). Also for simulation 3, the distribution of the innovations is far from normal (especially in the tails) although better than that of simulation 2 (bottom right panel of Figure 3.4). In conclusion, simulation 1 is the only one that does not significantly violate its statistical assumptions. In Table 3.2, simulation 2 obtained the highest values R^2 and Nash-Sutcliffe coefficient calculated with the simulated flow and observed flow. This demonstrates that unweighted least squares regression is an efficient technique to find a good fit solution. However, as mentioned above, this technique cannot be used to get reliable uncertainty estimates of model parameters and results.

Table 3.2: Performance of 3 simulations at the maxima of the posterior distribution

Simulation	Test data	Nash-Sutcliffe	R^2	Log posterior density
Simulation 1 with likelihood function (3.11)	Calibration period	0.77	0.80	-6510
	validation period	0.79	0.82	-6586
Simulation 2 with likelihood function (3.3)	Calibration period	0.85	0.85	-8615
	validation period	0.86	0.86	-8597
Simulation 3 with likelihood function (3.7)	Calibration period	0.77	0.80	-6668
	validation period	0.79	0.83	-6742

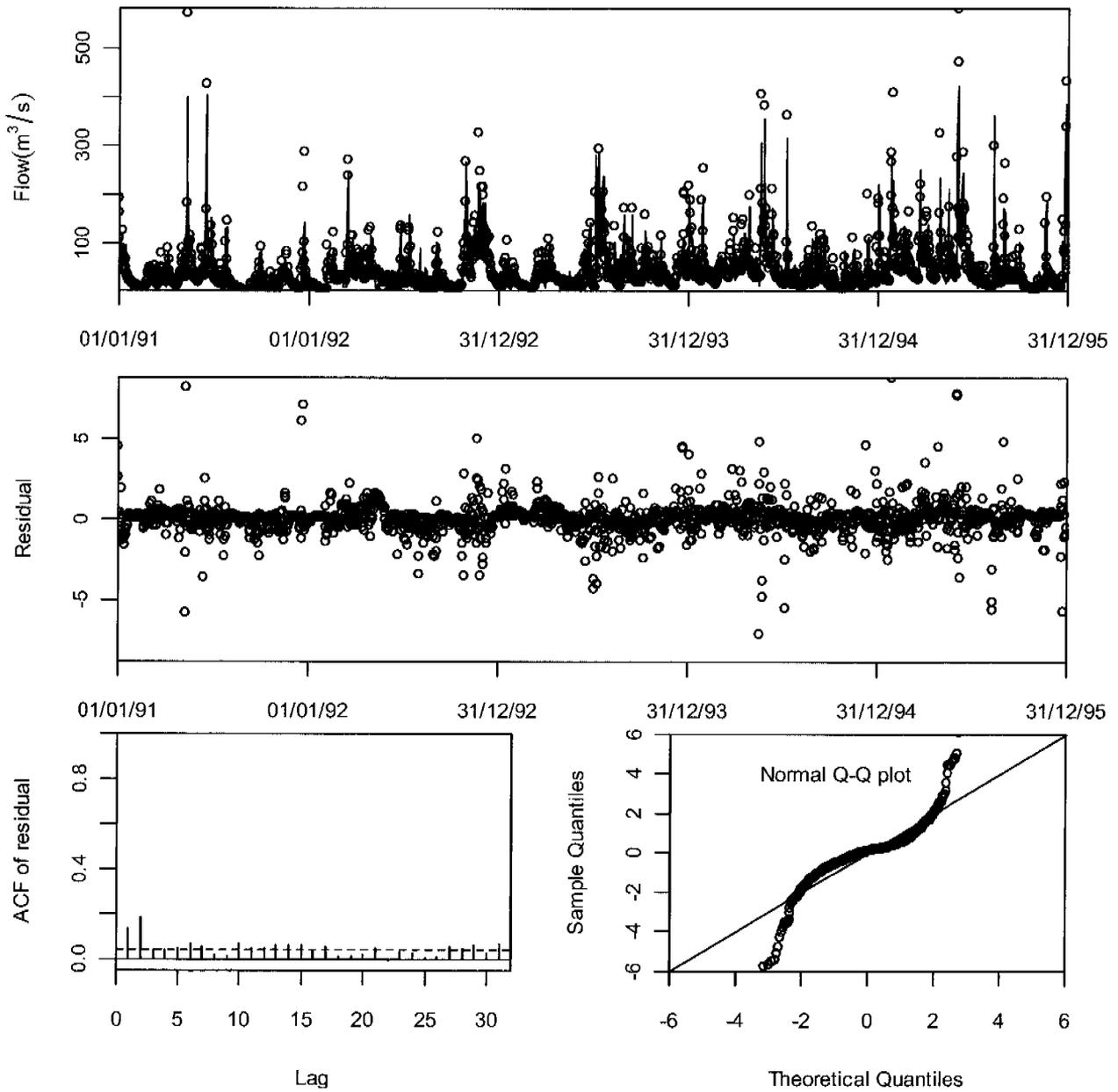


Figure 3.3: Statistics diagnostics for simulation 2 with likelihood function (3.3). From top to bottom: time series of the observed and simulated flows, time series of the normalized residuals, and the autocorrelation function and the normal quantile-quantile plot of normalized residuals.

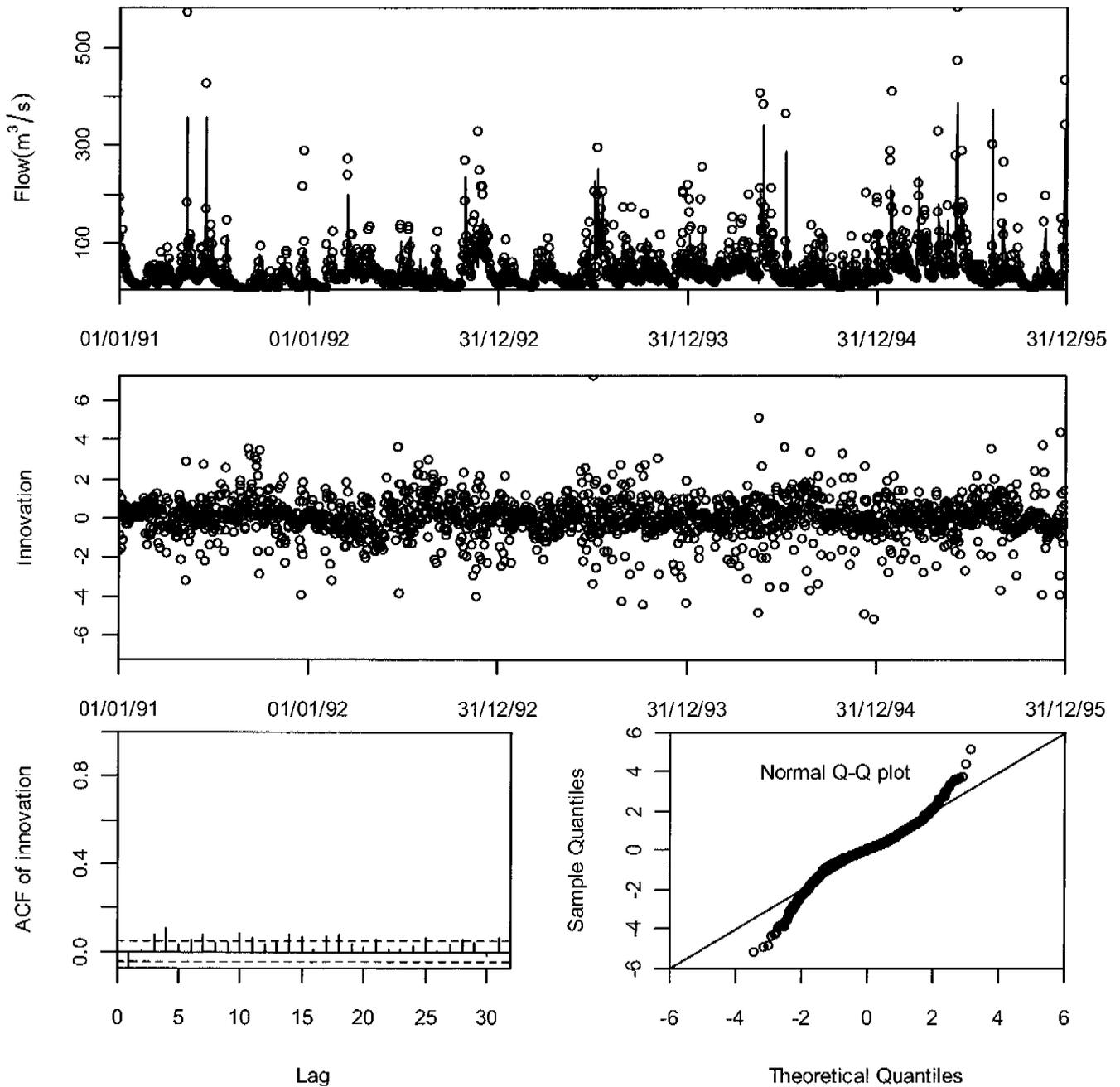


Figure 3.4: Statistics diagnostics for simulation 3 with likelihood function (3.7). From top to bottom: time series of observed and simulated flows, time series of innovations, and autocorrelation function and normal quantile-quantile plot of innovations.

For simulation 1, a Markov chain was started from the approximation to the maximum of the posterior density obtained above to get an approximation to the posterior distribution. The preliminary Markov chain led to the conclusion that both λ_1 and λ_2 are very close to 0. To decrease the complexity of the MCMC process, we fixed λ_1 and λ_2 to 0 and excluded them from

further MCMC processes. After a burn-in period of 40,000 model runs, 20,000 model runs were used to obtain the posterior parameter distribution and prediction uncertainty.

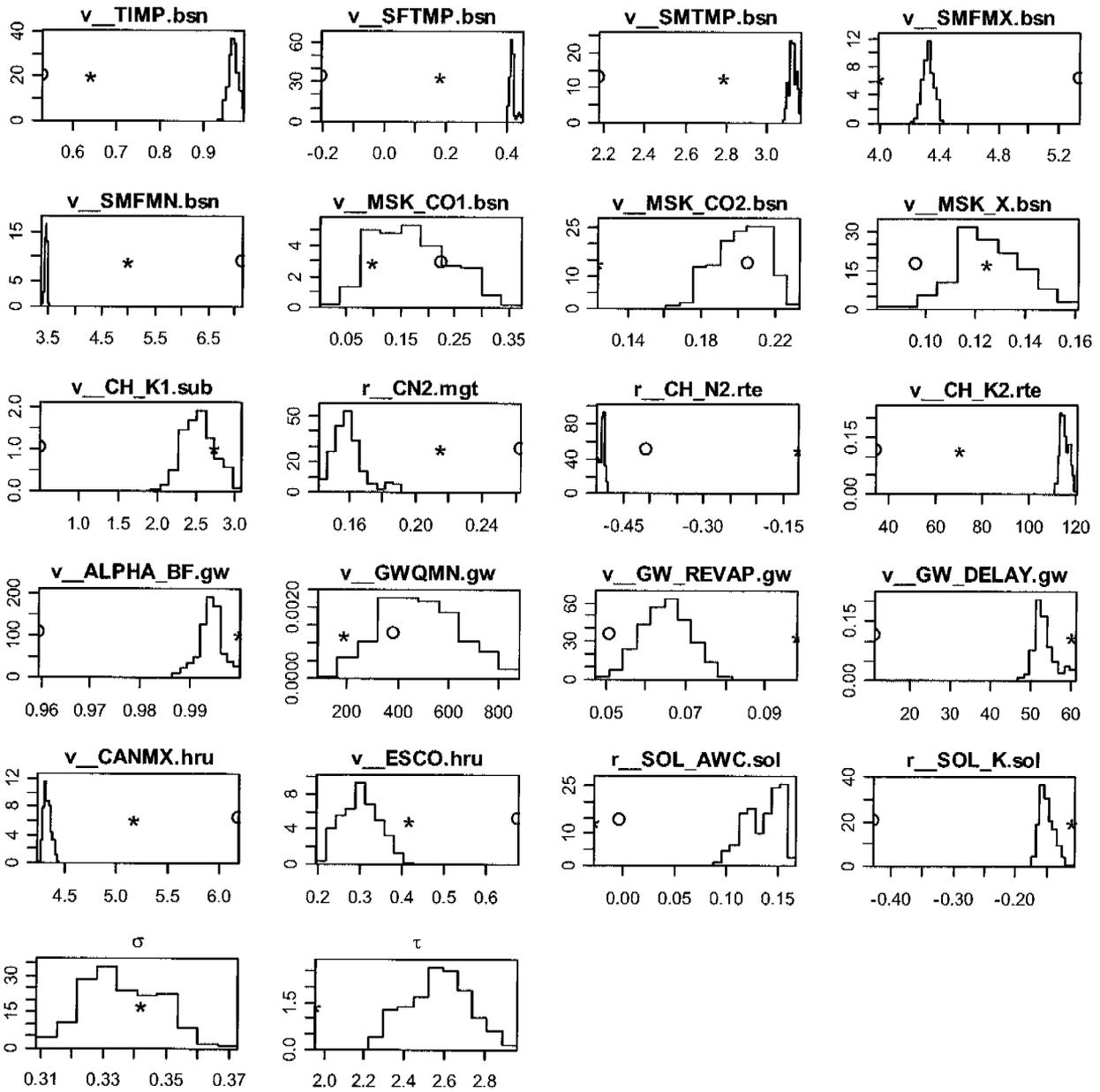


Figure 3.5: Histograms approximating the marginals of the posterior parameter distribution for simulation 1 and optimized parameters for simulation 2 (circles) and simulation 3 (asterisks).

Chapter 3

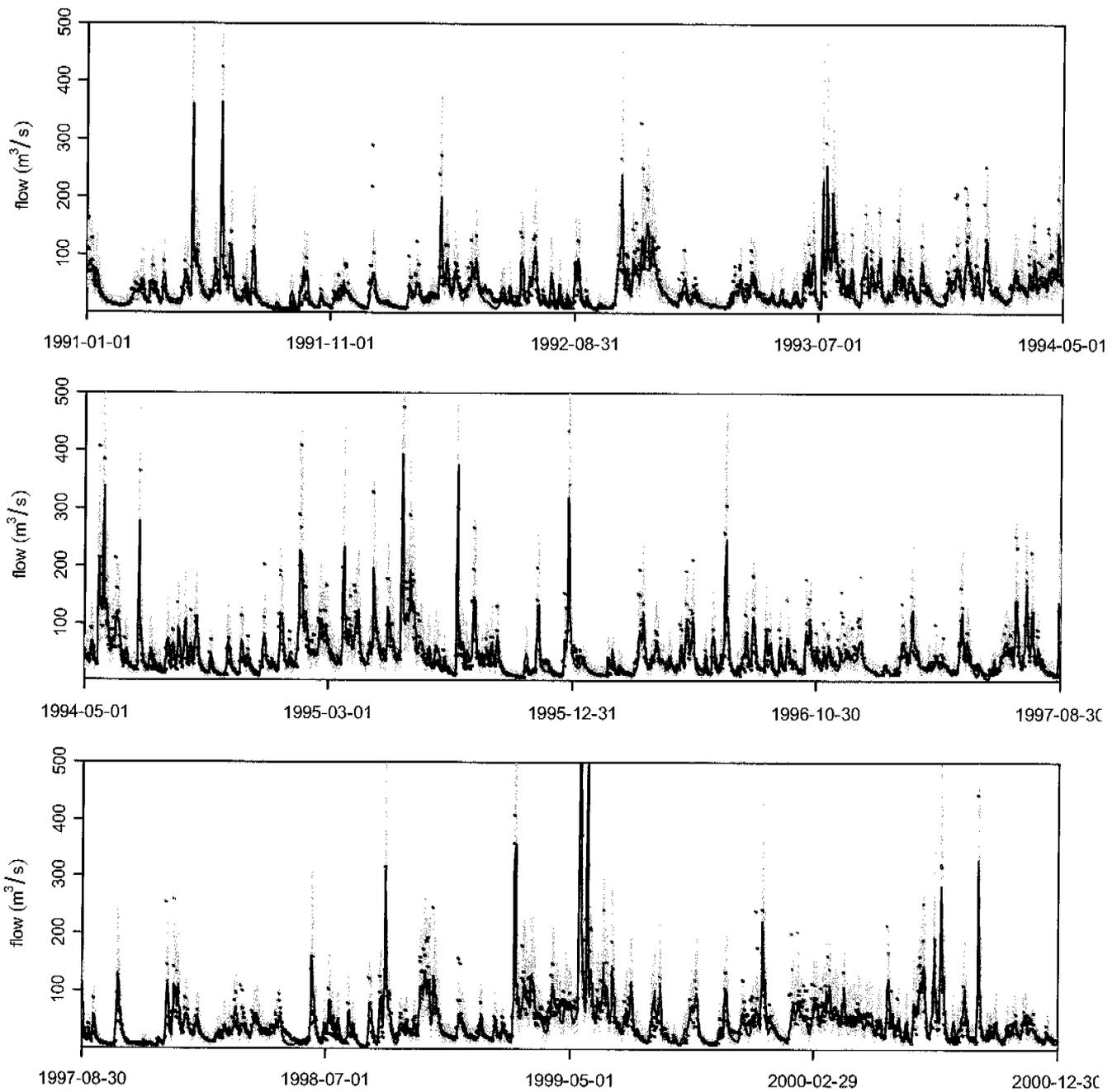


Figure 3.6: 95% prediction uncertainty bands associated with parameter uncertainty (dark shaded area), and with parameter uncertainty and continuous-time autoregressive model (light shaded area) for simulation 1 for both calibration period (1991-1995) and test period (1996-2000). The dots correspond to the observed flow series at the basin outlet and the line stands for the simulated discharge at the maximum of the posterior distribution.

Figure 3.5 shows the marginal distributions of the posterior parameter distribution. The increase in CN2 (positive value of $r_CN2.mgt$) reflects higher surface runoff than in the previous simulation, while an increase in ESCO (value around 0.32 instead of 0.154) indicates smaller evapotranspiration than in the previous simulation. The changes in the temperature related parameters (TIMP, SFTMP, SMTMP, SMFMX and SMFMN) demonstrate that temperature factors have a significant influence on river flow. The marginal posteriors of some parameters are at the boundary of the prior interval. This can be an indication for very poor identifiability due to strong correlations in the posterior. The large increase in CH_K2 reflects a stronger interaction between channel and groundwater. The characteristic correlation time is around 2~3 days. This indicates that there is no long-term correlation in the residuals. The parameter values corresponding to the maximum posterior density for simulations 2 and 3 are also plotted in Figure 3.5 as circles and asterisks, respectively. As we can see, due to different objective functions, optimum parameter values vary a lot.

Figure 3.6 shows the 95% prediction uncertainty bands associated with parameter uncertainty (dark shaded area), and with parameter uncertainty and continuous-time autoregressive model uncertainty (light shaded area) for both calibration period and validation period. As can be seen, although the prediction uncertainty from parameter uncertainty (dark shaded area) is very narrow (it only covers 7.2% of the observations), the 95% uncertainty bands from parameter uncertainty and autoregressive model brackets most of the observations, which indicates that our proposed approach can mimic the prediction uncertainty (if covers 92.3% of the observations). The dots correspond to the observed discharge at the basin outlet and the line represents the simulated discharge at the maximum of the posterior distribution.

3.4.2 Comparison with the results of the Chaohe Basin

Compared to the application of the continuous-time autoregressive model in the Chaohe Basin in China by Yang et al. [2007b], we can find some differences and similarities:

- 1) SWAT parameters. In the Chaohe Basin, river discharge is only sensitive to runoff generation (e.g., CN2, SOL_AWC and ESCO) during the wet season and the snow accumulation and melting processes are negligible. In the Thur river basin, flow is not only due to runoff generation (e.g., CN2, SOL_AWC and ESCO), but snow accumulation and melting processes are relevant (e.g., TIMP, SFTMP, SMTMP, SMFMX and SMFMN).

2) Standard deviation (σ) and characteristic correlation time (τ) of the error model. In the Chaohe Basin, these 2 parameters have a strong seasonal dependence, i.e., high σ and low τ during the wet season, and low σ and very high τ during the dry season. In the Thur river basin τ is relatively small. This can be explained by the climate difference of these 2 basins. In the temperate continental and semi-arid climate in the Chaohe Basin with over 80% rainfall in July and August, the flow during the dry weather season is strongly dependent on the water stored during the wet season. This leads to the very high value of the correlation time during the dry season. In the pre-alpine/alpine climate in the Thur river basin, river discharge is much more strongly dependent on rain events distributed throughout the year.

3) Prediction uncertainty. No matter how the continuous-time autoregressive error model is applied, the characteristic of the prediction is the same: narrow prediction uncertainty band from parameter uncertainty and substantially wider prediction uncertainty band from the continuous-time autoregressive error model. This difference between the 2 uncertainty bands indicates a high fraction of uncertainty due to input and model structure. The uncertainty due to parameters of the deterministic models may be underestimated by this procedure.

4) Convergence of MCMC. The Markov Chain for the simulation of the Thur river basin converged slower than that of the Chaohe Basin. The reason might be that the number of parameters in the simulation of the Thur river basin is large and the shape of the posterior is more complicated than that of the Chaohe Basin (possibly multi-modal with many local maxima).

3.5 Summary and Conclusion

The continuous-time autoregressive error model developed by Yang et al. [2007b] for hydrologic modelling was tested for a watershed with completely different characteristics than the one in Yang et al. [2007b]. This application required an extension of the distributional shape of the innovations from a Normal distribution to a Student t distribution to account for heavier tails of the innovations. The extended model was successfully applied (empirical results are not in disagreement with distributional assumptions made by the model) to an implementation of the hydrologic model of the Soil and Water Assessment Tool (SWAT) [Arnold et al., 1998] for the Thur river basin in Switzerland. The results for the Thur river basin are compared to those for the Chaohe Basin in China described in the previous paper [Yang et al., 2007b].

These analyses led to the following conclusions:

1) Our case studies indicate that the extended continuous-time autoregressive model is generally applicable as an error model for hydrologic simulations under significantly different climatic conditions (case studies for semi-arid climate in North China and pre-alpine/alpine climate in Switzerland). This was confirmed by statistical tests of the distributional assumptions of the model.

2) 2 case studies indicate that the parameters of the hydrologic model as well as the parameters of the error model need careful site-specific priors and calibration. Particularly, the degrees of freedom of the t distribution proved to be an effective parameter to adjust the distributional shape of the innovations (to account for heavy tails), and the standard deviation and characteristic correlation time of the error model required a seasonal variation for the semi-arid climate in North China that was not required under the pre-alpine/alpine climate in Switzerland. The reason for this is that river discharge during very long dry weather periods is dependent on precipitation during the rainy season before, whereas the dominant influence on river discharge during wet periods is rain event over a much shorter preceding period.

While our approach leads to a satisfactory mechanistic and statistical description of runoff, it does not separate input and model structural uncertainty. The resolution of this should continue to be a future effort in hydrological systems analysis.

3.6 Acknowledgments

We thank the Swiss Federal Office of Meteorology and Climatology (<http://www.meteoschweiz.ch/web/en/weather.html>) and the Swiss Federal River Survey Program (NADUF; <http://www.naduf.ch>) for the provision of climatological and river discharge data.

3.7 References

Abbaspour, K. C., C. A. Johnson, and M. T. van Genuchten (2004), Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure, *Vadose Zone Journal*, 3(4), 1340-1352.

Chapter 3

- Abbaspour, K. C., J. Yang, I. Maximov, R. Siber, K. Bogner, J. Mieleitner, Z. Jobrist, and S. Raghavan (2007), Spatially-distributed modelling of hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT, *Journal of Hydrology*, 333:413-430.
- Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams (1998), Large area hydrologic modeling and assessment - Part 1: Model development, *Journal of the American Water Resources Association*, 34(1), 73-89.
- Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resources Research*, 37(4), 937-947.
- Bayarri, M.J., J.O. Berger, D. Higdon, M.C. Kennedy, A. Kottas, R. Paulo, J. Sacks, J.A. Cafeo, J. Cavendish, C.H. Lin, J. Tu (2002), A framework for validation of computer models, Technical Report Number 128, National Institute of Statistical Sciences, Research Triangle Park, NC, USA (<http://www.niss.org>).
- Bayarri, M.J., J.O. Berger, R. Paulo, J. Sacks, J.A. Cafeo, J. Cavendish, C.H. Lin, J. Tu (2007), A framework for validation of computer models. In press.
- Beven, K. (1992), Future of Distributed Modeling - Special Issue of Hydrological Processes, *Hydrological Processes*, 6(3), 253-254.
- Beven, K., A. Binley (1992), The Future of Distributed Models - Model Calibration and Uncertainty Prediction, *Hydrological Processes* 6(3), 279-298.
- Beven, K. (2006), A manifesto for the equifinality thesis, *Journal of Hydrology*, 320, 18-36.
- Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *Journal of the Royal Statistical Society Series B*, 211-252.
- Box, G. E. P., and D. R. Cox (1982), An analysis of transformations revisited, rebutted, *Journal of the American Statistical Association*, 77(377), 209-210.
- Carrera, J., and S. P. Neuman (1986), Estimation of Aquifer Parameters under Transient and Steady-State Conditions .1. Maximum-Likelihood Method Incorporating Prior Information, *Water Resources Research*, 22(2), 199-210.
- Chatfield, C. (2003), *The Analysis of Time Series: An introduction (Sixth Edition)*, Chapman & hall/CRC pub, P40.
- Chow V. T., D. R. Maidment, and L. W. Mays (1988), *Applied Hydrology*, McGraw-Hill, Inc., New York, NY.

- Cunge, J. A. (1969), On the subject of a flood propagation method (Muskingum method), *J. Hydraulics Research*, 7(2), 205-230.
- Duan, Q. Y., S. Sorooshian, and R.P. Ibbitt (1988), A maximum likelihood criterion for use with data collected at unequal time intervals. *Water Resources Research* 24(7), 1163-1173.
- Duan, Q. Y., S. Sorooshian, and V. Gupta (1992), Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models, *Water Resources Research*, 28(4), 1015-1031.
- Gelman, S., J. B. Carlin, H. S. Stren, and D. B. Rubin (1995), *Bayesian Data Analysis*, Chapman and Hall, New York, USA.
- Gurtz, J., A. Baltensweiler, and H. Lang (1999), Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins, *Hydrological Processes*, 13(17), 2751-2768.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368.
- Kennedy, M.C. and A. O'Hagan (2001), Bayesian calibration of computer models (with discussion), *Journal of the Royal Statistical Society B* 63, 425-464.
- Kloeden, P. E., and E. Platen (1992), *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, Germany.
- Kool, J. B., and J. C. Parker (1988), Analysis of the Inverse Problem for Transient Unsaturated Flow, *Water Resources Research*, 24(6), 817-830.
- Kuczera, G. (1983), Improved Parameter Inference in Catchment Models .1. Evaluating Parameter Uncertainty, *Water Resources Research*, 19(5), 1151-1162.
- Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *Journal of Hydrology*, 211(1-4), 69-85.
- McKay, M. D., R. J. Beckman, and W. J. Conover (1979), Comparison of 3 Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics*, 21(2), 239-245.
- Reichert, P. (2005), UNCSIM - A computer programme for statistical inference and sensitivity, identifiability, and uncertainty analysis, in *Proceedings of the 2005 European Simulation and*

Chapter 3

Modelling Conference (ESM 2005), edited by J. M. F. Teixeira and A. E. Carvalho-Brito, Oct. 24-26, Porto, Portugal, EUROSIS-ETI, pp. 51-55.

- Reichert, P. (2006), A standard interface between simulation programs and systems analysis software, *Water Science and Technology*, 53(1), 267-275.
- Tomassini, L., P. Reichert, H. R. Künsch, C. Buser, and M. E. Borsuk (2007), A smoothing algorithm for estimating stochastic, continuous-time model parameters and an application to a simple climate model. Submitted.
- van Griensven, A., and T. Meixner (2006), Methods to quantify and identify the sources of uncertainty for river basin water quality models, *Water Science and Technology*, 53(1), 51-59.
- van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, A. Diluzio, and R. Srinivasan (2006), A global sensitivity analysis tool for the parameters of multi-variable catchment models, *Journal of Hydrology*, 324(1-4), 10-23.
- Vrugt, J. A., W. Bouten (2002), Validity of first-order approximations to describe parameter uncertainty in soil hydrologic models. *Soil Science Society of America Journal* 66(6), 1740-1751.
- Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian, A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39(8), 1201, doi:10.1029/2002WR001642, 2003.
- Vrugt, J. A., G. Schoups, J. W. Hopmans, C. Young, W. W. Wallender, T. Harter, and W. Bouten (2004), Inverse modeling of large-scale spatially distributed vadose zone properties using global optimization, *Water Resour. Res.*, 40, W06503, doi:10.1029/2003WR002706.
- Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, 41, W01017, doi:10.1029/2004WR003059.
- Williams, J. R. (1969), Flood routing with variable travel time or variable storage coefficients, *Transactions of the Asae*, 12(1), 100-103.
- Yang, J., Abbaspour, K.C., Reichert, P., 2005, Interfacing SWAT with Systems Analysis Tools: A Generic Platform. In: Srinivasan, R., Jacobs, J., Day, D., Abbaspour, K. (Eds.), 2005 3rd International SWAT Conference Proceedings, July 11-15, Zurich, Switzerland, pp. 169-178

Yang, J., K. C. Abbaspour, P. Reichert, H. Yang, and J. Xia (2007a), Comparing Uncertainty Analysis Techniques for a SWAT Application to the Chaohe Basin in China, Submitted to Journal of Hydrology.

Yang, J., P. Reichert, K. C. Abbaspour, and H. Yang (2007b), Hydrological Modelling of the Chaohe Basin in China: Statistical Model Formulation and Bayesian Inference, Accepted by Journal of Hydrology.

Seite Leer /
Blank leaf

4 Comparing Uncertainty Analysis Techniques for a SWAT Application to the Chaohe Basin in China

Jing Yang, K.C. Abbaspour, Peter Reichert, Hong Yang, Jun Xia

(Submitted to Journal of Hydrology)

Abstract

Distributed watershed models are increasingly being used to support decisions about alternative management strategies in the areas of landuse change, climate change, water allocation, and pollution control. For this reason it is important that these models pass through a careful calibration and uncertainty analysis. To fulfil this demand, in recent years, scientists have come up with various uncertainty analysis techniques for watershed models. To determine the differences and similarities of these techniques we compared 5 uncertainty analysis procedures: Generalized Likelihood Uncertainty Estimation (GLUE), Parameter Solution (ParaSol), Sequential Uncertainty Fitting algorithm (SUFI-2), and a continuous-time autoregressive error model applied in a Bayesian framework and implemented with Markov Chain Monte Carlo (MCMC) and Importance Sampling (IS) techniques. For the comparison, we used the program Soil and Water Assessment Tool (SWAT) and applied it to the Chaohe Basin in China. As the uncertainty analysis techniques are different in their philosophies and leave the user free to make subjective choices, a direct comparison between the techniques is difficult. In this study, we applied each technique according to its typical use in hydrology and compared the posterior parameter distributions, performances of their best solutions, prediction uncertainty, conceptual bases, efficiency, and difficulty of implementation. The comparison results for these categories are listed and the advantages and disadvantages are analyzed. The final choice of the uncertainty analysis technique is left to the reader. From the point of view of the authors, Bayesian-based approaches are most recommendable because of their conceptual basis, but construction and test of the likelihood function requires critical attention.

Keywords

Uncertainty analysis; watershed modeling; Bayesian inference, Markov Chain Monte Carlo (MCMC); Importance Sampling (IS); SUFI-2; GLUE; ParaSol.

4.1 Introduction

Simulation programs implementing models of watershed hydrology and river water quality are important tools for watershed management for both operational and research purposes. In recent years many such simulation programs were developed such as AGNPS (Agricultural Non Point Source) (Young et al., 1989), SWAT (Soil and Water Assessment Tool) (Arnold et al., 1998) and HSPF (Hydrologic Simulation Program – Fortran) (Bicknell et al., 2000). Areas of application of watershed models include: integrated watershed management (e.g., Zacharias et al., 2005), peak flow forecasting (e.g. Jorgeson and Julien, 2005), test of the effectiveness of measures for the reduction of non-point source pollutants (e.g., Bekele and Nicklow, 2005; Santhi et al., 2001), soil loss prediction (e.g. Cochrane and Flanagan 2005), assessment of the effect of landuse change (e.g. Hundecha and Bardossy, 2004, Claessens et al., 2006; Cotler and Ortega-Larrocea, 2006), analysis of causes of nutrient loss (e.g. Abbaspour et al., 2007; Adeuya et al., 2005), and climate change impact assessment (e.g. Claessens et al., 2006; Huang et al., 2005; Pednekar et al. 2005) among many others. This large number of various, and often very specific, applications led to the development of a multitude of watershed models starting in the early 1960s (see Todini, 1988 for a historical review).

As distributed watershed models are increasingly being used to support decisions about alternative management strategies, it is important that these models should pass through a careful calibration and uncertainty analysis. Calibration of watershed models, however, is a challenging task because conceptual model uncertainty is quite large. Sources of model uncertainty include input uncertainty, model structural uncertainty, parameter uncertainty, and uncertainty in the calibration data. Sources of model structural uncertainty include processes not accounted for in the model, unknown activities in the watershed, and model inaccuracy due to over-simplification of the processes that are considered in the model. Some examples of this type of uncertainty are: effects of wetlands and reservoirs on hydrology and chemical transport; interaction between surface and groundwater; occurrence of landslides, and large constructions (e.g., roads, dams, tunnels, bridges) that could produce large amounts of sediment during short time periods affecting water quantity and quality; unknown wastewater discharges into the streams from factories and water treatment plants; imprecisely known application of fertilizers and pesticides, unknown irrigation activities and water diversions, and other activities in the river basin. The

input uncertainty is often related to imprecise or spatially interpolated measurements of model input or initial conditions, such as elevation data, landuse data, rainfall, temperature and initial groundwater levels. Other uncertainties in distributed models may also arise due to the large number of unknown parameters and the errors in the data used for parameter calibration.

To account for modelling uncertainties, in the last two decades, many uncertainty-analysis-techniques have been developed and applied to various catchments. However, only rarely more than one technique has been applied in the same case study in the literature. The objective of this paper is to fill this gap. We apply 5 different calibration and uncertainty analysis techniques to the same catchment to compare their performances. These include Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992), Parameter Solution (ParaSol) (Van Griensven and Meixner, 2006), Sequential Uncertainty Fitting (SUFI-2) (Abbaspour et al., 2004; 2007), Bayesian inference based on Markov Chain Monte Carlo (MCMC) (e.g., Kuczera and Parent, 1998; Vrugt et al., 2003; Yang et al., 2007), and Bayesian inference based on Importance Sampling (IS) (e.g., Kuczera and Parent, 1998). For the comparison, we used the program Soil and Water Assessment Tool (SWAT) applied to the Chaohe Basin in China. As the uncertainty analysis techniques are different in their philosophies and formulations, a literal comparison is impossible. Hence, despite the subjective nature of such an assessment, we formulated a typical application for each technique according to its typical use in hydrology and compare their posterior parameter distribution, performances of their best solutions, prediction uncertainty, conceptual basis, difficulty and efficiency of implementation.

The remainder of this paper is structured as follows. In section 4.2, we introduce the methodology used for the comparison, give a brief overview of the selected techniques, and then list the criteria for the assessment. In section 4.3, we give an overview of the study site, the SWAT hydrological model, and our model application (priors and selection and aggregation of parameters). In section 4.4 the results are presented and discussed. Section 4.5 contains the conclusions.

4.2 Methodology, Selected Techniques, and Criteria for Comparison

4.2.1 General Methodology

There are various difficulties in comparing uncertainty analysis techniques in hydrological modelling. The following list addresses the most important concerns and how we handled them.

- Most techniques are different in their philosophies and subjective choices have to be made in their formulation with respect to prior parameter distributions, likelihood functions and/or goal functions. We addressed this problem by choosing priors and goal functions for each technique as they would typically be used in hydrological applications. This leads necessarily to different goal functions for different techniques. When discussing the results, we will analyze whether a problem is caused by the conceptual formulation of a particular technique or by the choice of the goal function.
- Different underlying concepts and goal functions from different techniques make the comparison difficult. The values of the goal functions from all techniques will be calculated for the best solution for each technique to allow for a fair comparison. In addition we use measures of efficiency and an assessment of the conceptual basis as criteria for the comparison.
- Different techniques obviously lead to different results for different criteria. We will outline the results in all criteria so that the reader can draw his/her own conclusions. Our own conclusions depend to some degree on a subjective judgment. As an example, not all readers may agree with our preference for the conceptual basis of Bayesian inference.
- The results of the comparison inherently depend on the application. We try to separate the results of specific application from generic results in the discussion.

4.2.2 Selected Techniques

4.2.2.1. GLUE

GLUE is an uncertainty analysis technique inspired by Importance Sampling and Regional Sensitivity Analysis (GSA; Hornberger and Spear, 1981). In GLUE, parameter uncertainty accounts for all sources of uncertainty, i.e., input uncertainty, structural uncertainty, parameter uncertainty and response uncertainty, because “the likelihood measure value is associated with a

parameter set and reflects all these sources of error and any effects of the covariation of parameter values on model performance implicitly” (Beven and Freer, 2001). Also, from a practical point of view, “disaggregation of the error into its source components is difficult, particularly in cases common to hydrology where the model is nonlinear and different sources of error may interact to produce the measured deviation” (Gupta & Beven, 2005). In GLUE, Parameter uncertainty is described as a set of discrete “behavioral” parameter sets with corresponding “likelihood weights”.

A GLUE analysis consists of the following three steps:

1) After the definition of the “generalized likelihood measure” $L(\theta)$, a large number of parameter sets are randomly sampled from the prior distribution and each parameter set is assessed as either “behavioral” or “non-behavioral” through a comparison of the “likelihood measure” with a given threshold value.

2) Each behavioral parameter is given a “likelihood weight” according to:

$$w_i = \frac{L(\theta_i)}{\sum_{k=1}^N L(\theta_k)} \quad (4.1)$$

where N is the number of behavioral parameter sets.

3) Finally, the prediction uncertainty is described as prediction quantile from the cumulative distribution realized from the weighted behavioral parameter sets.

In the literature, the most frequently used likelihood measure for GLUE is the *Nash-Sutcliffe* coefficient (NS) (e.g. Beven and Freer, 2001; Freer et al., 1996), which is also used in this paper:

$$NS = 1 - \frac{\sum_{t_i=1}^n (y_{t_i}^M(\boldsymbol{\theta}) - y_{t_i})^2}{\sum_{t_i=1}^n (y_{t_i} - \bar{y})^2} \quad (4.2)$$

where n is the number of the observed data points, and y_{t_i} and $y_{t_i}^M(\boldsymbol{\theta})$ represents the observation and model simulation with parameters $\boldsymbol{\theta}$ at time t_i , respectively, and \bar{y} is the average value of the observations.

4.2.2.2. *ParaSol and modified ParaSol*

ParaSol is based on a modification to the global optimization algorithm SCE-UA (Duan et al., 1992). The idea is to use the simulations performed during optimization to derive prediction uncertainty because “the simulations gathered by SCE-UA are very valuable as the algorithm samples over the entire parameter space with a focus on solutions near the optimum/optima” (Van Griensven and Meixner, 2006). Hence, ParaSol only accounts for parameter uncertainty.

The procedure of ParaSol is as follows:

1) After optimization applying the modified SCE-UA (the randomness of the algorithm SCE-UA is increased to improve the coverage of the parameter space), the simulations performed are divided into ‘good’ simulations and ‘not good’ simulations by a threshold value of the goal function as in GLUE. This leads to ‘good’ parameter sets and ‘not good’ parameter sets.

2) Prediction uncertainty is constructed by equally weighting all ‘good’ simulations.

The Objective function used in ParaSol is the sum of the squares of the residuals (SSQ):

$$SSQ = \sum_{t_i=1}^n (y_{t_i}^M(\boldsymbol{\theta}) - y_{t_i})^2 \quad (4.3)$$

The relationship between NS and SSQ is

$$NS = 1 - \frac{1}{\sum_{t_i=1}^n (y_{t_i} - \bar{y})^2} \cdot SSQ \quad (4.4)$$

where $\sum_{t_i=1}^n (y_{t_i} - \bar{y})^2$ is a fixed value for given observations. To improve the comparability with

GLUE, all objective function values of ParaSol were converted to NS .

As the choice of the threshold of the objective function in ParaSol is based on the χ^2 -statistics it mainly accounts for parameter uncertainty under the assumption of independent measurements. For the purpose of comparison with GLUE, as an alternative, we choose the same threshold as used by GLUE and we call this method “modified ParaSol”.

4.2.2.3. SUFI-2 procedure

In SUFI-2, the uncertainty of input parameters is described by a multivariate uniform distribution in a parameter hypercube, while the output uncertainty is quantified by the 95% prediction uncertainty band (95PPU) calculated at the 2.5% and 97.5% levels of the cumulative distribution function of the output variables (Abbaspour et al., 2007). Latin hypercube sampling is used to draw independent parameter sets (Abbaspour et al., 2007). Similar to GLUE, SUFI-2 represents uncertainties of all sources through parameter uncertainty in the hydrological model.

The procedure of SUFI-2 is as follows:

1) In the first step, the goal function $g(\boldsymbol{\theta})$ and meaningful parameter ranges $[\boldsymbol{\theta}_{\text{abs min}}, \boldsymbol{\theta}_{\text{abs max}}]$ are defined.

2) Then a Latin Hypercube sampling is carried out in the hypercube $[\boldsymbol{\theta}_{\text{min}}, \boldsymbol{\theta}_{\text{max}}]$ (initially set to $[\boldsymbol{\theta}_{\text{abs min}}, \boldsymbol{\theta}_{\text{abs max}}]$), the corresponding goal functions are evaluated, and the sensitivity matrix \mathbf{J} and the parameter covariance matrix \mathbf{C} are calculated according to:

$$J_{ij} = \frac{\Delta g_i}{\Delta \theta_j} \quad i = 1, \dots, C_2^m, \quad j = 1, \dots, n, \quad (4.5)$$

$$\mathbf{C} = s_g^2 (\mathbf{J}^T \mathbf{J})^{-1} \quad (4.6)$$

where s_g^2 is the variance of the objective function values resulting from the m model runs.

3) The 95% confidence interval of a parameter θ_j is computed as follows:

$$\theta_{j,lower} = \theta_j^* - t_{v,0.025} \sqrt{C_{jj}} \quad , \quad \theta_{j,upper} = \theta_j^* + t_{v,0.025} \sqrt{C_{jj}} \quad (4.7)$$

where θ_j^* is the parameter θ_j for the best solutions (i.e., parameters which produce the optimal goal function), and v is the degrees of freedom ($m - n$).

4) The 95PPU is calculated. And then two indices, i.e., p -factor (the percent of observation bracketed by 95PPU) and r -factor, are calculated:

$$r\text{-factor} = \frac{\frac{1}{n} \sum_{t_i=1}^n (y_{t_i,97.5\%}^M - y_{t_i,2.5\%}^M)}{\sigma_{obs}} \quad (4.8)$$

where $y_{t_i,97.5\%}^M$ and $y_{t_i,2.5\%}^M$ represent the upper part and lower part of the 95PPU, and σ_{obs} stands for the standard deviation of the measured data.

The goodness of calibration and prediction uncertainty is judged on the basis of the closeness of *p-factor* to 100% (i.e., all observations bracketed by the prediction uncertainty) and *r-factor* to 1 (i.e., achievement of rather small uncertainty band). As all uncertainties in the conceptual model and inputs are reflected in the measurements (e.g., discharge), bracketing most of the measured data in the prediction 95PPU ensures that all uncertainties are depicted by the parameter uncertainties. Hence, if the two factors have satisfactory values, then the parameter range $[\theta_{min}, \theta_{max}]$ is the posterior parameter distribution. Otherwise, $[\theta_{min}, \theta_{max}]$ is updated according to:

$$\theta_{j,min,new} = \theta_{j,lower} - \max\left(\frac{\theta_{j,lower} - \theta_{j,min}}{2}, \frac{\theta_{j,max} - \theta_{j,upper}}{2}\right) \quad (4.9)$$

$$\theta_{j,max,new} = \theta_{j,upper} + \max\left(\frac{\theta_{j,lower} - \theta_{j,min}}{2}, \frac{\theta_{j,max} - \theta_{j,upper}}{2}\right)$$

and another iteration needs to be performed.

SUFI-2 allows several choices of the objective function (for instance *NS* coefficient). In the literature, the weighted root mean square error (*RMSE*) (Abbaspour et al., 2004) and the weighted sum of squares *SSQ* (Abbaspour et al., 2007) were used. In this study we chose the *NS* coefficient for the sake of comparison with other techniques.

4.2.2.4. Bayesian Inference

According to Bayes' theorem, the probability density of the posterior parameter distribution $f_{\theta_{post}|Y}(\theta|y_{meas})$ is derived from the prior density $f_{\theta_{pri}}(\theta)$ and data y_{meas} as:

$$f_{\theta_{post}|Y}(\theta|y_{meas}) = \frac{f_{Y^M|\theta}(y_{meas}|\theta) \cdot f_{\theta_{pri}}(\theta)}{\int f_{Y^M|\theta'}(y_{meas}|\theta') f_{\theta_{pri}}(\theta') d\theta'} \quad (4.10)$$

where $f_{Y^M|\theta}(y_{meas}|\theta)$ is the likelihood function of the model, i.e. the probability density for model results for given parameters with the measurements substituted for the model results.

Posterior prediction uncertainty is usually represented by quantiles of the posterior distribution. The crucial point in applying this technique is the formulation of the likelihood function. If the statistical assumptions for formulating the likelihood function are violated, the results of Bayesian inference are unreliable. Unfortunately, when formulating likelihood functions in hydrological applications, it is often assumed that the residuals between measurements and simulations are independently and identically (usually normally) distributed (iid). However this assumption is often violated. To avoid this problem in our case study, we constructed the likelihood function by combining a Box-Cox transformation (Box and Cox, 1964; 1982) with a continuous-time autoregressive error model (Brockwell and Davis, 1996; Brockwell, 2001) as follows:

$$f_{Y^M|\theta}(y|\theta) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{[g(y_{t_0}) - g(y_{t_0}^M(\theta))]^2}{\sigma^2}\right) \cdot \left| \frac{dg}{dy} \right|_{y=y_{t_0}} \cdot \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma \sqrt{1 - \exp\left(-2 \frac{t_i - t_{i-1}}{\tau}\right)}} \exp\left(-\frac{1}{2} \frac{\left[\begin{array}{c} g(y_{t_i}) - g(y_{t_i}^M(\theta)) \\ - [g(y_{t_{i-1}}) - g(y_{t_{i-1}}^M(\theta))] \exp\left(-\frac{t_i - t_{i-1}}{\tau}\right) \end{array} \right]^2}{\sigma^2 \left(1 - \exp\left(-2 \frac{t_i - t_{i-1}}{\tau}\right)\right)}\right) \cdot \left| \frac{dg}{dy} \right|_{y=y_{t_i}} \right] \quad (4.11)$$

where σ is the asymptotic standard deviation of the errors, τ is the characteristic correlation time, θ is the vector of model parameters, y_{t_i} and $y_{t_i}^M(\theta)$ are the observation and model simulation, respectively, at time t_i , and $g(\cdot)$ represents the Box-Cox transformation (Box and Cox, 1964; 1982):

$$g(y) = \begin{cases} \frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & \lambda_1 \neq 0 \\ \ln(y + \lambda_2) & \lambda_1 = 0 \end{cases}, \quad g^{-1}(z) = \begin{cases} (\lambda_1 z + 1)^{1/\lambda_1} - \lambda_2 & \lambda_1 \neq 0 \\ \exp(z) - \lambda_2 & \lambda_1 = 0 \end{cases}, \quad \frac{dg}{dy} = (y + \lambda_2)^{\lambda_1 - 1} \quad (4.12)$$

This model extends earlier works with discrete-time autoregressive error models in hydrological applications (e.g. Kuczera 1983, Duan et al. 1988, Bates and Campbell, 2001). More details are given by Yang et al. (2007).

Chapter 4

Two generic Monte Carlo approaches to sample from a posterior distribution are Markov Chain Monte Carlo and Importance Sampling (Gelman et al. 1995; Kuzera and Parent, 1998). Both techniques are used as implemented in the systems analysis tool UNCSIM (Reichert 2005; <http://www.uncsim.eawag.ch>).

Markov Chain Monte Carlo (MCMC)

MCMC generates samples from a random walk which adapts to the posterior distribution (Kuczera and Parent, 1998). The simplest technique from this class is the Metropolis-Hasting algorithm (Gelman et al. 1995), which is applied in this study. A sequence (Markov Chain) of parameter sets representing the posterior distribution is constructed as follows:

- 1) An initial starting point in the parameter space is chosen.
- 2) A candidate for the next point is proposed by adding a random realization from a symmetrical jump distribution, f_{jump} , to the coordinates of the previous point of the sequence:

$$\theta_{k+1}^* = \theta_k + rand(f_{jump}) \quad (4.13)$$

- 3) The acceptance of the candidate points depends on the ratio r :

$$r = \frac{f_{\theta_{post}|y}(\theta_{k+1}^* | y_{meas})}{f_{\theta_{post}|y}(\theta_k | y_{meas})} \quad (4.14)$$

If $r \geq 1$, then the candidate point is accepted as a new point, else it is accepted with probability r . If the candidate point is rejected, the previous point is used as the next point of the sequence.

In order to avoid long burn-in periods (or even lack of convergence to the posterior distribution) the chain is started at a numerical approximation to the maximum of the posterior distribution calculated with the aid of the shuffled complex global optimization algorithm (Duan et al., 1992).

Importance Sampling (IS)

Importance Sampling is a well established technique for randomly sampling from a probability distribution (Gelman et al. 1995; Kuzera and Parent, 1998). The idea is to draw randomly from a

sampling distribution f_{sample} and calculate weights for the sampling points to make the weighted sample a sample from the posterior distribution. The procedure consists of the following steps:

- 1) Choose a sampling distribution and draw a random sample from this sampling distribution.
- 2) For each parameter set, θ_i , of the sample, calculate a weight according to

$$w_i = \frac{f_{\theta_{post}|Y}(\theta_i | \mathbf{y}_{meas}) / f_{sample}(\theta_i)}{\sum_{k=1}^N f_{\theta_{post}|Y}(\theta_k | \mathbf{y}_{meas}) / f_{sample}(\theta_k)} \quad (4.15)$$

- 3) Use the weighted sample to derive properties of the posterior distribution, for example, by calculating the expected value of a function h according to:

$$E_{f_{post}}(h) \approx \sum_{k=1}^N w_k h(\theta_k) \quad (4.16)$$

The efficiency of this procedure depends strongly on how close the sampling distribution is to the posterior distribution, and hence, the choice of the sampling distribution is crucial (Tanner, 1992; Gelman et al., 1995). Three practical choices for the sampling distribution are: sampling from the prior distribution (often uniform sampling over a hypercube referred to in the following as primitive IS or naive IS), the approximation with over-dispersed multi-normal distribution (e.g. Kuczera, 1998), and the method of iteratively adapting the sampling distribution and using efficient sampling techniques (Reichert et al., 2002). Each of the above methods has some disadvantages. Primitive IS is very inefficient if the posterior is significantly different from the prior, particularly for high dimensional parameter spaces. It is also worth nothing that primitive IS is a special case of GLUE, in which no generalizations are made to the likelihood function and all parameter sets are accepted as behavioral (although some will get a very small weight). For the method with over-dispersed multi-normal distribution, it is difficult to determine a prior for the dispersion coefficients (Kuczera, 1998). The method of iteratively adapting the sampling distribution becomes more and more difficult to implement as the dimensionality of the parameter space increases (Reichert et al., 2002). This is because larger samples are required to get sufficient information on the shape of the posterior and it becomes more and more difficult to find a reasonable parameterized sampling distribution to approximate the posterior. In this study,

only the Primitive IS is implemented, as this also allows us to study the behaviour of GLUE with different likelihood measures.

4.2.3 Criteria for the Comparison

We use the following five categories to compare the performances of the uncertainty analysis techniques:

1. Parameter estimates and parameter uncertainty (values, uncertainty ranges, correlation coefficients).
2. Performance of the simulation at the mode of the posterior distribution was evaluated for all criteria.
3. The model prediction uncertainty

Three indices are used to compare the derived 95% probability band (95PPU). Those indices are the width of 95PPU (i.e., *r-factor* as used in SUFI-2), percentage of the measurements bracketed by this band (i.e. *p-factor* in SUFI-2), and the Continuous Rank Probability Score (CRPS). CRPS is widely used in weather forecast as a measure of the closeness of the predicted and occurred cumulative distributions and sharpness of the predicted PDF (e.g., Hersbach, 2000). For a time series, the CRPS at time t can be defined as:

$$CRPS_t = \int_{-\infty}^{\infty} (F_t(y) - H(y - y_t))^2 dy \quad (4.17)$$

where $F_t(y)$ stands for the predicted CDF at time t , H is the Heaviside function (returning zero for negative and unity for non-negative arguments), and y_t is the observed at time t .

In practice the CRPS is averaged over a time series:

$$CRPS = \sum_t w_t \cdot CRPS_t \quad (4.18)$$

where w_t is the weight for corresponding $CRPS_t$ at time t and we take equal weights in our study.

4. The conceptual basis of the technique (theoretical basis, testability and fulfillment of statistical assumptions, capability of exploring the parameter space, coverage of regions with high goal function values).
5. Difficulty of implementation and efficiency of the technique (programming effort and number of simulations required to get reasonable results).

4.3 Study Site, SWAT Watershed Simulation Program, and Model Application

4.3.1. The Chaohe Watershed and Data

The Chaohe watershed is situated in North China with a drainage area of 5,300 km² upstream of the Xiahui station (see Figure 2.1 in Chapter 2). The climate is temperate continental, semi-humid and semi-arid. From 1980 to 1990 the average daily maximum temperature was 6.2 °C, the average daily minimum temperature 0.9 °C, and the yearly rainfall varied between 350 to 690 mm. The elevation varies from 200 m at the basin outlet to 2,400 meters at the highest point in the catchment. The topography is characterized by high mountain ranges, steep slopes and deep valleys. The average channel slope is 1.87% which leads to fast water flow in the river. Average daily flow at the catchment outlet (Xiahui station) is 9.3 m³ s⁻¹ and varies irregularly from around 798 m³ s⁻¹ during the flood season to lower than 1 m³ s⁻¹ in the dry season. The runoff coefficient (the ratio of runoff to precipitation) at the Xiahui station to the rainfall in this basin decreased from 0.24 in 1980 to 0.09 in 1990. It is believed that the decline is mainly due to the intensified human activities, including increasing water use and building of more (small scale) water retention structures.

4.3.2. The Watershed Simulation Program

The Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998) is a continuous-time, spatially distributed simulator of water, sediment, nutrients and pesticides transport at a catchment scale. It runs on a daily time step. In SWAT, a watershed is divided into a number of sub-basins based on a given DEM (Digital Elevation Model) map. Within each sub-basin, soil and landuse maps are overlaid to create a number of unique hydrologic response units (HURUs). SWAT simulates surface and subsurface processes, accounting for snow fall and snow melt, vadose zone processes (i.e. infiltration, evaporation, plant uptake, lateral flows, and percolation

into aquifer). Runoff volume is calculated using the Curve Number method (USDA Soil conservation Service, 1972). Sediment yield from each sub-basin is generated using the Modified Universal Soil Loss Equation (MUSLE) (Williams, 1995). The model updates the C factor of the MUSLE on a daily basis using information from the crop growth module. The routing phase controls the movement of water using the variable storage method or the Muskingum method (Cunge, 1969; Chow, 1988).

4.3.3. Model Application

Parameterization of spatially-distributed hydrologic models can potentially lead to a large number of parameters. To effectively limit the number of parameters, we developed an aggregating scheme based on hydrologic group (A, B, C, or D), soil texture, landuse, sub-basin, and the spatial distribution of default values. This scheme was implemented in an interface, iSWAT, that allows systems analysis programs to access SWAT parameters that are distributed over many input files (Yang et al., 2005; <http://www.uncsim.eawag.ch/interfaces/swat>). The names of aggregate parameters specified in the interface iSWAT have the following format:

$$x_ \langle \text{parname} \rangle . \langle \text{ext} \rangle _ _ \langle \text{hydrogrp} \rangle _ _ \langle \text{soltext} \rangle _ _ \langle \text{landuse} \rangle _ _ \langle \text{subbsn} \rangle \quad (4.19)$$

where x represents the type of change to be applied to the parameter (v : replacement; a : absolute change; or r : relative change), $\langle \text{parname} \rangle$ is the SWAT parameter name; $\langle \text{ext} \rangle$ represents the extension of the SWAT input file which contains the parameter; $\langle \text{hydrogrp} \rangle$ is the identifier for the hydrologic group; $\langle \text{soltext} \rangle$ is the soil texture; $\langle \text{landuse} \rangle$ is the landuse; and $\langle \text{subbsn} \rangle$ is the subbasin number, the crop index, or the fertilizer index. The interface exchanges parameter values with the systems analysis tool based on a simple text file-based interface (Reichert, 2006).

Following our previous work (Yang et al., 2007), 10 aggregate SWAT parameters related to discharge at the watershed outlet were selected. These parameters, listed in Table 4.1, represent single global values, global multipliers, or global additive terms to the distributed default values of the corresponding SWAT parameters (compare parameter names in Table 4.1 with the explanations of Expression 19). The likelihood function for the Bayesian approach requires the additional parameters σ and τ characterizing the standard deviation and characteristic correlation time of the continuous-time autoregressive error model (see Eq. 4.11). These parameters were considered to be dependent on the seasons, i.e., σ_{dry} and τ_{dry} were used for dry season (October to

May), and σ_{wet} and τ_{wet} were used for wet season (July to August), and we assumed a linear transition from one value to the other in June and September (Yang et al. 2007).

The priors of all the parameters above were assumed to be independent of each other. For the SWAT parameters, uniform priors within reasonable ranges were assumed for all the techniques. For the parameters σ and τ , densities proportional to $1/\sigma$ and $1/\tau$ were chosen, which is equivalent to assuming that the logarithms of these parameters are uniformly distributed. Table 4.1 gives an overview of the parameters used for calibration and their prior distributions.

As we cannot specify reasonable initial values for all storage volumes considered in the model, SWAT is operated for a “warm-up” period of 5 years without comparison of model results with data. We found that such a “warm-up period” was sufficient to minimize the effects of the initial state of SWAT variables on river flow. Furthermore, in order to verify the calibrated model parameters, the model was calibrated and tested based on the observed discharges at the watershed outlet (Xiahui station) using a split sample procedure. The data from the years 1985-1988 with omission of a single outlier in 1985 was used for calibration, and the data from 1989-1990 was used to test the model. This strategy was applied for all the techniques.

Table 4.1: Selected parameters for uncertainty analysis and their prior distributions. The last 4 parameters are only used for MCMC and IS

Aggregate Parameters ^{*1}	Name and meaning of underlying SWAT parameter	Range of Initial parameter underlying SWAT parameter ^{*2}	Prior distribution of aggregate parameter ^{*3}
a_CN2.mgt	CN2: curve number	72-92	U[-30, 5]
v_ESCO.hru	ESCO: soil evaporation compensation factor	0.95	U[0,1]
v_EPCO.hru	EPCO: plant uptake compensation factor	1.00	U[0,1]
r_SOL_K.sol	SOL_K : soil hydraulic conductivity (mm/hr)	1.6-328.27	U[-0.8, 0.8]
a_SOL_AWC.sol	SOL_AWC: soil available water capacity (mm H ₂ O/mm soil)	0-0.13	U[0,0.15]
v_ALPHA_BF.gw	ALPHA_BF: base flow alpha factor (1/day)	0.048	U[0,1]
r_SLSUBBSN.hru ^{*3}	SLSUBBSN: average slope length (m)	9.461-91.463	U[-0.6, 0.6]
a_CH_K2.rte	CH_K2: effective hydraulic conductivity in main channel alluvium (mm/hr)	0	U[0,150]
a_OV_N.hru	OV_N: overland Manning roughness	0.06-0.15	U[0,0.2]
v_GW_DELAY.gw	GW_DELAY: groundwater delay time (days)	31	U[0,300]
λ_1	parameter of Box-Cox transformation		U[0,1]
σ			Inv ^{*4}
σ_{dry}	standard deviation during dry season in Equation (4.11)		Inv ^{*4}
σ_{wet}	standard deviation during wet season in Equation (4.11)		Inv ^{*4}
τ			Inv ^{*4}
τ_{dry}	characteristic correlation time of autoregressive process during dry season (days) in Equation (4.11)		Inv ^{*4}
τ_{wet}	characteristic correlation time of autoregressive process during wet season (days) in Equation (4.11)		Inv ^{*4}

^{*1} Aggregate parameters are constructed based on Eq (4.19). "a __", "v __" and "r __" means an absolute increase, a replacement, and a relative change to the initial parameter value respectively.

^{*2} This range is based on the initial parameter estimates of the project.

^{*3} Prior distributions of aggregate parameters are based on Neitsch, et al (2001); U[a,b] in this column means the distribution of this parameter/aggregate parameters is uniform over the interval [a,b]

^{*4} "Inv" means that the probability density at the value x is proportional to 1/x (the logarithm of the parameter is uniform distributed)

4.4 Results and Discussion

We start with a description of the results for each technique and then compare and discuss the results according to the categories of criteria given in section 4.2.3.

4.4.1. Results of GLUE implementation with likelihood measure *NS*

GLUE is convenient and easy to implement and widely used in hydrology (e.g. Freer et al., 1996; Cameron et al., 2000a and 2000b; Blazkova et al., 2002). The drawback of this approach is its prohibitive computational burden imposed by its random sampling strategy (Hossain et al., 2004).

In this study, the threshold value of GLUE application is chosen to be 0.70, i.e. the simulations with *NS* values larger than 0.70 are behavioral otherwise non-behavioral. Four GLUE simulations were performed with sample sizes of 1000, 5000, 10000, and 20000. For each simulation, the dot plot, cumulative posterior distribution and 95PPU are analyzed. The comparison shows that there are some differences in the results between 1000, 5000 and 10000 while there is no significant difference between 10000 and 20000. The following analysis of results and comparison are based on a sample size of 10000. The dot plot shown in Figure 4.1 demonstrates that for each parameter solutions with similarly good values of the *NS* coefficient can be found within the complete prior range. The posteriors of most aggregate parameters follow closely the uniform prior distribution. Table 4.2 shows the mean, standard deviation and correlation matrix of the posterior parameter distribution. The correlations between most parameters are small except between `a__CN2.mgt` and `a__SOL_AWC.sol`, `v__ESCO.hru` and `a__SOL_AWC.sol`, and `r__SOL_K.sol` and `r__SLSUBBSN.hru`, with values of 0.44, 0.56 and 0.67, respectively. The third column in Table 4.2 shows the standard deviations of the parameters. Figure 4.2 shows the 95PPU of the model results for both calibration and validation periods. Most of the observations are bracketed by the 95PPUs (79% during the calibration period and 69% during the validation period, see *p-factor* in Table 4.5).

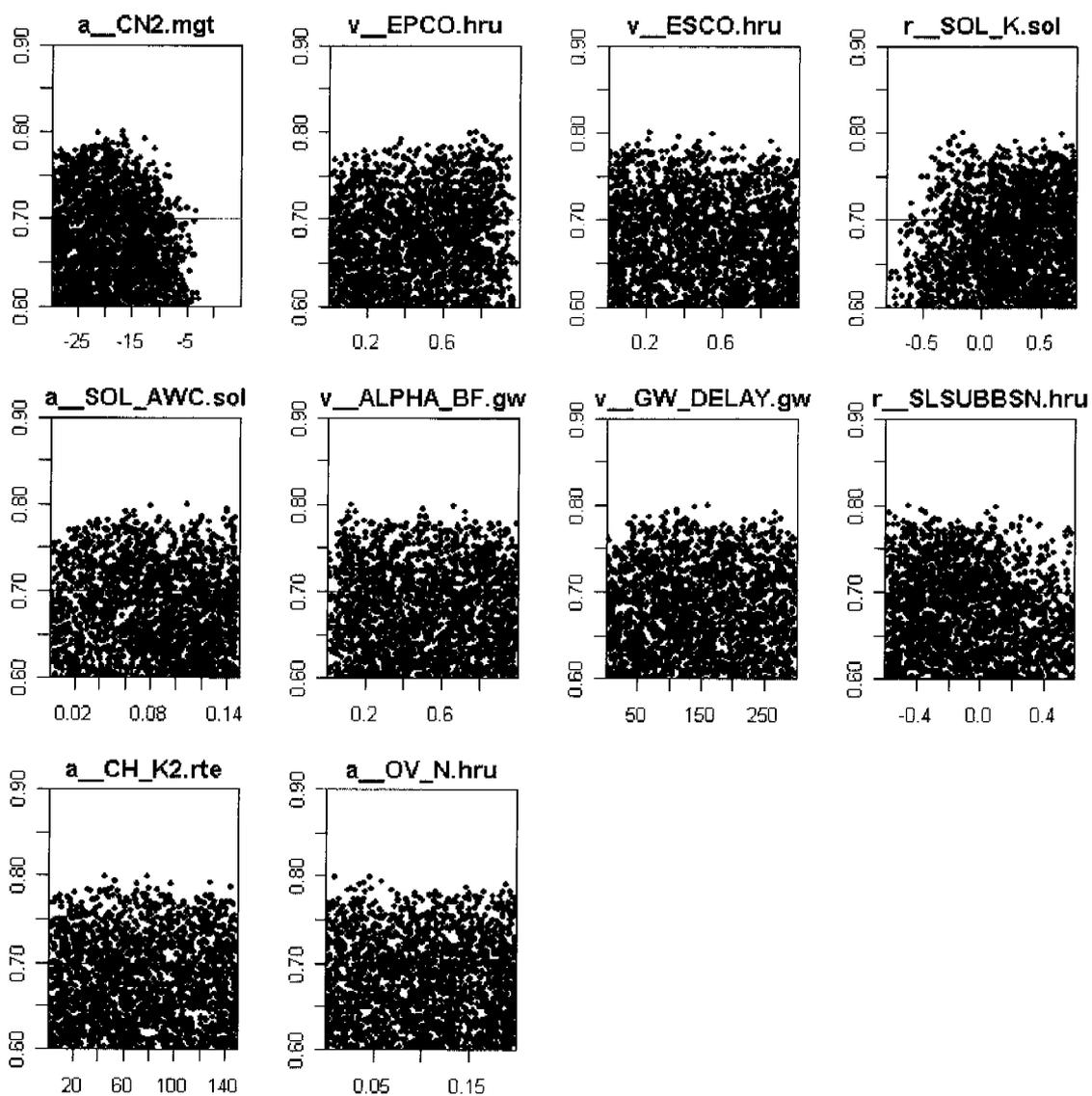


Figure 4.1: Dotty plot of NS coefficient against each aggregate SWAT parameter conditioning with GLUE based on 10,000 samples with threshold 0.70 (red line), above which the parameter sets are behavioral.

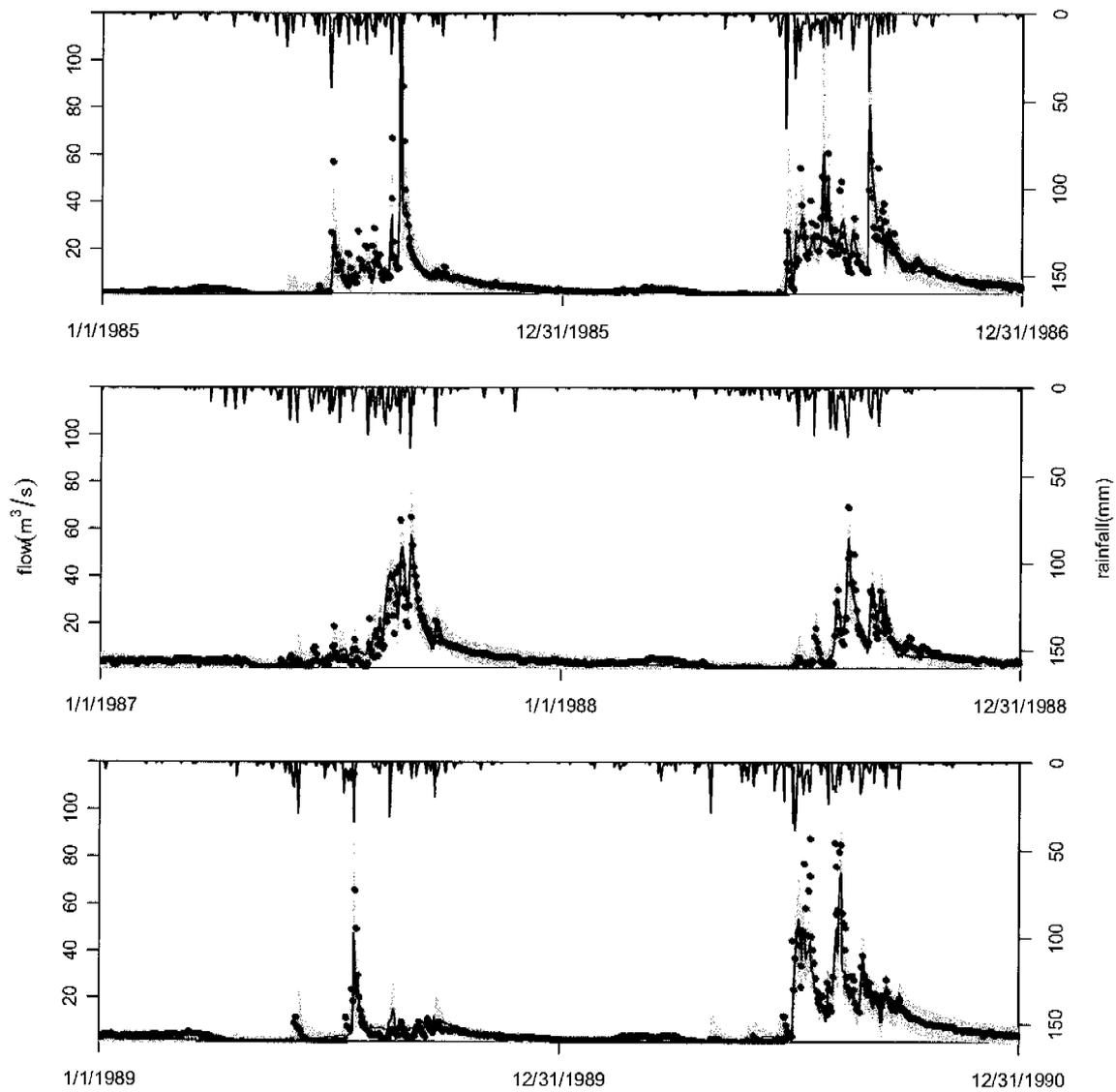


Figure 4.2: 95PPU (shaded area) derived by GLUE. The dots correspond to the observed discharge at the basin outlet, while the solid line represents the best simulation obtained by GLUE.

Table 4.2: Mean, standard deviation (SD) and correlation matrix of the posterior distribution resulting from application of the GLUE technique

parameter	mean	stdev											
a__CN2.mgt	-20.27	5.56	1										
v__EPCO.hru	0.47	0.29	-0.02	1									
v__ESCO.hru	0.51	0.25	0.04	0.18	1								
r__SOL_K.sol	0.30	0.33	-0.04	0.03	-0.10	1							
a__SOL_AWC.sol	0.08	0.04	0.44	-0.09	0.56	-0.09	1						
v__ALPHA_BF.gw	0.51	0.28	-0.19	0.03	0.02	-0.07	-0.06	1					
v__GW_DELAY.gw	149.46	81.96	0.03	0.01	0.15	0.06	-0.04	-0.14	1				
r__SLSUBBSN.hru	-0.13	0.28	0.09	-0.03	0.24	0.67	0.19	0.07	0.01	1			
a__CH_K2.rte	74.99	42.42	0.16	-0.02	-0.12	0.00	-0.08	-0.01	0.04	-0.13	1		
a__OV_N.hru	0.10	0.06	-0.02	0.03	0.01	0.05	-0.02	-0.03	0.03	0.02	0.08	1	

4.4.2. Results of ParaSol and modified ParaSol implementations with objective function SSQ

Implementation of ParaSol is relatively easy and the computation depends only on the convergence of the optimization process (SCE-UA algorithm). Once the optimization is done, ParaSol will determine the behavioral and non-behavioral parameter sets and produce prediction uncertainty.

The application of ParaSol resulted in 851 behavioral parameter sets out of a total of 7550 samples (the threshold value based on the χ^2 -statistics is equivalent to NS 0.819). Figure 4.3 shows the dot plot of the NS coefficient against each parameter. Clearly, the parameter samples are very dense around the maximum. This is confirmed by very steep cumulative distribution functions (not shown) and small standard deviations of the estimated model parameters (third column in Table 4.3). ParaSol based on the SCE-UA is very efficient in detecting the area with high goal-function values in the response surface. The threshold line (blue line) in Figure 4.3 separates the parameters sets into behavioral parameter sets (above the blue line) and non-behavioral parameter sets (below the blue line). However, as can be seen, both the number and area of the behavioral parameter sets are extremely small, and the corresponding parameter ranges are very narrow. This also leads to a very narrow 95PPU for model predictions shown in Figure 4.4 (dark gray area). ParaSol failed to derive the prediction uncertainty (only 18% of measurements were bracketed by 95PPU during the calibration period) though the best simulation matches the observation quite well with NS equals 0.82 during the calibration period. This is because ParaSol doesn't consider the error in the model structure, measured input and measured

response, which results in underestimation of the prediction uncertainty. The developer of ParaSol solved this problem by reducing the threshold to include the correct number of data points (technique “SUNGLASSES”). SUNGLASSES is not applied here because it needs to take into account the observations during the validation period, which will complicate the comparison.

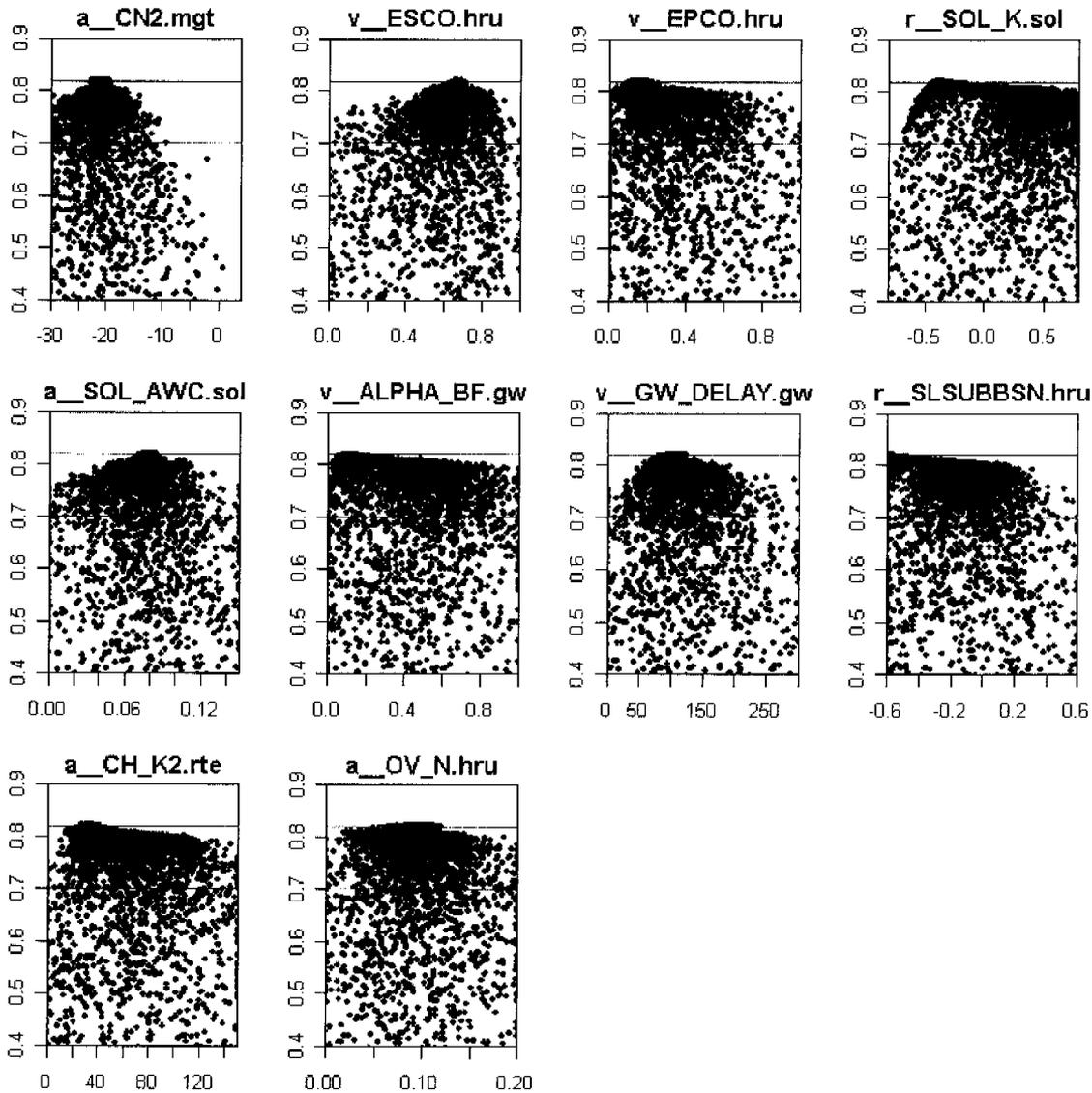


Figure 4.3: Dotty plot of NS coefficient against aggregate SWAT parameters conditioning with ParaSol. The blue line is the threshold determined by ParaSol, and red line is the threshold with value 0.70 for modified ParaSol.

As to the modified ParaSol with threshold value 0.70, Figure 4.3 shows the behavioral and non-behavioral parameter sets separated by threshold line with value 0.70 (red line), and light

Chapter 4

grey area in Figure 4.4 describes the 95PPU. There are 60% of measurements bracketed by 95PPU during calibration period and 52% during validation period (see *p-factor* in Table 4.5).

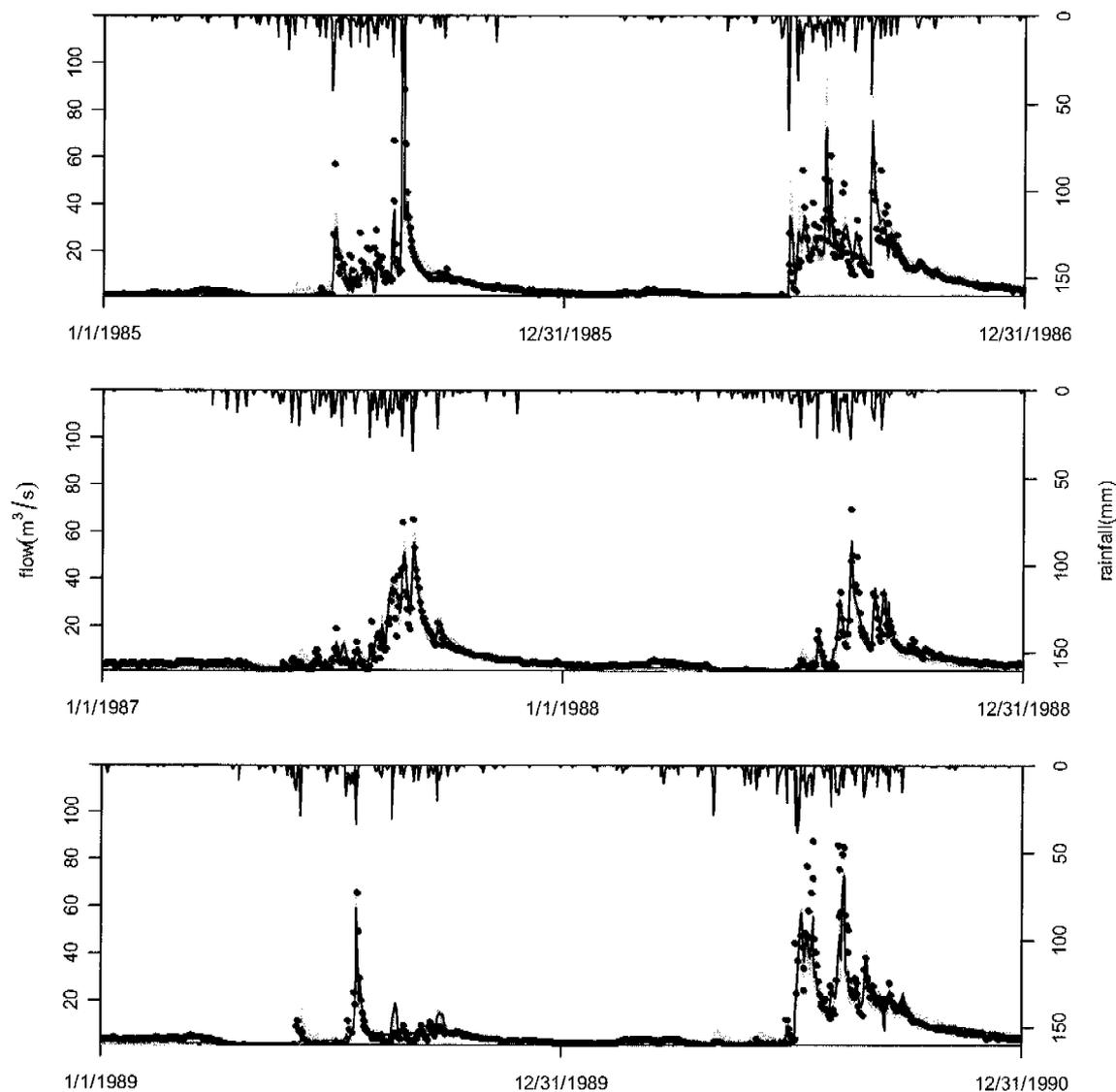


Figure 4.4: 95PPUs derived by ParaSol (dark gray area) and modified ParaSol (light gray area). The dots correspond to the observed discharge at the basin outlet, while the solid line represents the best simulation obtained by ParaSol.

Table 4.3: Mean, standard deviation (SD) and correlation matrix of the posterior distribution resulting from application of the ParaSol technique

Aggragate parameters	mean	stdev											
a__CN2.mgt	-21.08	1.81	1										
v__ESCO.hru	0.65	0.07	0.18	1									
v__EPCO.hru	0.22	0.13	-0.06	0.04	1								
r__SOL_K.sol	0.00	0.38	-0.03	-0.14	0.50	1							
a__SOL_AWC.sol	0.08	0.01	0.42	0.54	-0.18	-0.20	1						
v__ALPHA_BF.gw	0.29	0.21	-0.15	-0.08	0.57	0.84	-0.14	1					
v__GW_DELAY.gw	106.62	24.91	-0.08	-0.02	0.35	0.36	-0.27	0.38	1				
r__SLSUBBSN.hru	-0.35	0.24	0.03	-0.07	0.48	0.96	-0.11	0.85	0.33	1			
a__CH_K2.rte	49.58	23.41	-0.07	-0.19	0.54	0.76	-0.36	0.73	0.45	0.72	1		
a__OV_N.hru	0.09	0.02	-0.09	-0.11	0.30	0.31	-0.18	0.26	0.16	0.28	0.36	1	

4.4.3. Result of SUFI-2 implementation with objective function NS

SUFI-2 is also convenient to use. The drawback of this approach is that it is semi-automated and requires the interaction of the modeler for checking a set of suggested posterior parameters, hence, requiring a good knowledge of the parameters and their effects on the output. This may add an additional error, i.e., “modeler’s uncertainty” to the list of other uncertainties.

For the SUFI-2 approach, we did 2 iterations with 1500 model runs in each iteration. In the second iteration, the 95PPU brackets 84% of the observations and *r-factor* equals 1.03 which is very close to a suggested value of 1. Posterior distributions in SUFI-2 are always independent and uniformly distributed, and expressed as narrowed parameter ranges (see the interval bracketed by parentheses in category 1 in Table 4.5). Figure 4.5 shows the dot plot conditioned on SUFI-2, and all these sampled parameter sets are taken as behavioral samples and contributing to the 95PPU. Obviously there are some parameter sets with low *NS* values (e.g., -1.5) in Figure 4.5. Figure 4.6 shows 95PPU for model results derived by SUFI-2 for the second iteration. As can be seen, most of the observations are bracket by the 95PPU (84% during calibration period and 82% during validation period), indicating SUFI-2 is capable of capturing the observations during both calibration and validation periods. The 95PPU is quite suitable to bracket the observations in 1985, 1988 and 1989, while it is somehow slightly overestimated in 1986, 1987 and 1990 especially in the recession part. This indicates there is a lot of uncertainty in the recession calculation of SWAT. However, as SUFI-2 is a sequential procedure, i.e., one more

Chapter 4

iteration can always be made leading to a smaller 95PPU at the expense of more observation points falling out of the prediction band.

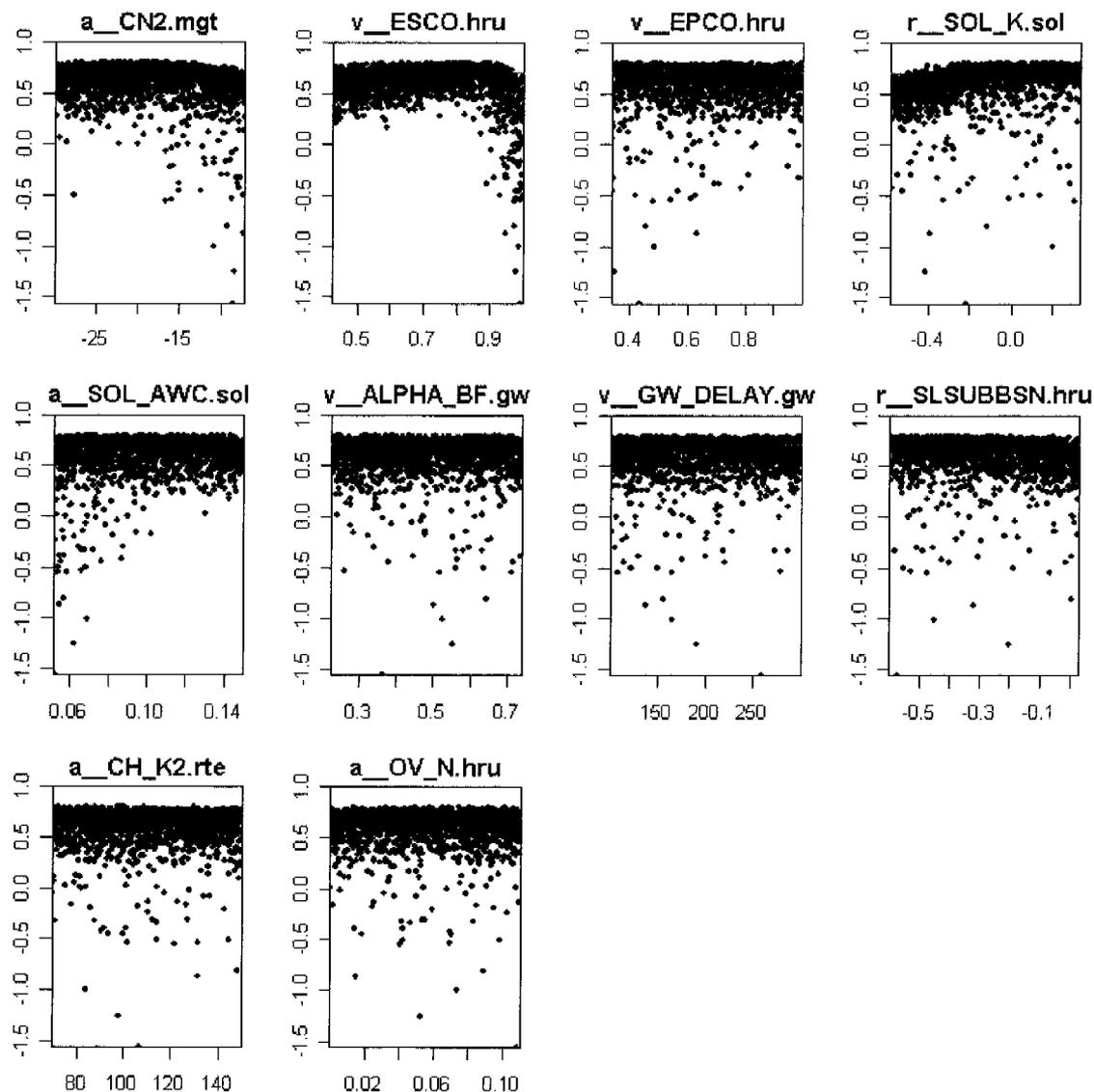


Figure 4.5: Dotty plot of NS coefficient against each aggregate SWAT parameter conditioning with SUFI-2. The red line represents the NS coefficient 0.70 used in the GLUE application.

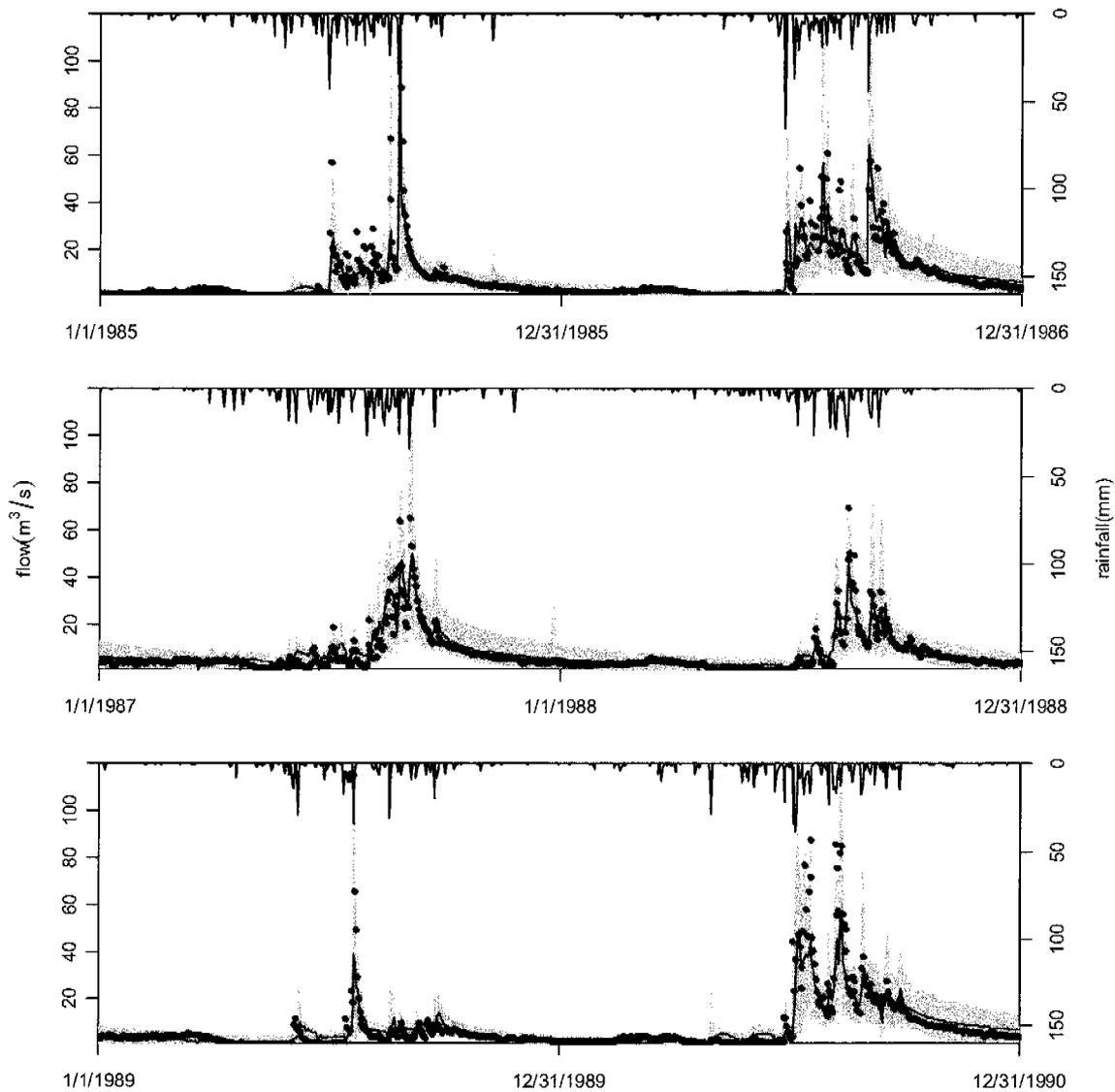


Figure 4.6: 95PPU (shaded area) derived by SUFI-2. The dots correspond to the observed discharge at the basin outlet, while the solid line represents the best simulation obtained by SUFI-2.

4.4.4. Result of MCMC implementation of Bayesian analysis with autoregressive error model

Implementation of Bayesian inference is not so easy especially for complex models because it requires formulating and testing of a likelihood function that characterizes the stochasticity of the observations. This usually requires several iterations of the inference procedure for different likelihood functions as statistical tests of residuals can only be performed after the analysis is completed. Once the constructed likelihood function is validated (i.e. the statistical assumptions

Chapter 4

for the likelihood function are validated), MCMC must be conducted and the resulting chain must be analyzed for the burn-in and stationary periods. Only points from the stationary period should be used for inference.

In this study, the Markov Chain was started at a numerical approximation to the maximum of the posterior distribution calculated with the aid of the SCE-UA (Duan et al., 1992) to keep the burn-in period short. The Markov Chain was run until 20,000 simulations were reached after the convergence of the chain to the stationary distribution monitored by the Heidelberger and Welch method (Heidelberger and Welch, 1983; Cowles and Carlin, 1996). The “CODA” package (Best et al., 1995) as implemented in the statistical software package R (<http://www.r-project.org>) was used to perform this test. As shown in Yang et al. (2007), the statistical assumptions of the likelihood function (Eq 4.11) were not significantly violated, so that we can be confident about the derived prediction uncertainties.

Figure 4.7 shows histograms which approximate the marginal posterior distributions of parameters conditioned with Bayesian MCMC. Except for the parameter `a__OV_N.hru` which has the approximate uniform distribution of its prior, all other parameters exhibit different posterior distributions than their priors in both parameter range and shape of the distributions. Table 4.4 lists the means, standard deviations, and correlation matrix of the posterior parameter distribution. As can be seen from Table 4.4, with the exception of the high correlation between the parameters `r__SOL_K.sol` and `r__SLSUBBSN.hru`, correlations between aggregate parameters are not very high. The high correlations between the parameters of the autoregressive error model (σ_{dry} , σ_{wet} , τ_{dry} , and τ_{wet}) indicate strong interactions among those parameters. Figure 4.8 shows the 95PPU of the model results arising from parameter uncertainty only (dark shaded area) and from total uncertainty (light shaded area) due to parameters, input, model structure and output represented by parameter uncertainty and the autoregressive error model. As can be seen, although the prediction uncertainty based on the parameter uncertainty alone in MCMC is quite narrow, that from parameter uncertainty and uncertainty sources represented by the autoregressive error model brackets over 80% of the observed points for both calibration and validation periods. This indicates that there is a large uncertainty in input, output and model structure in addition to parameter uncertainty. As can also be seen, there is a slight overestimation of prediction uncertainty during the wet season, and this suggests more attention should be paid to the wet season when constructing the likelihood function.

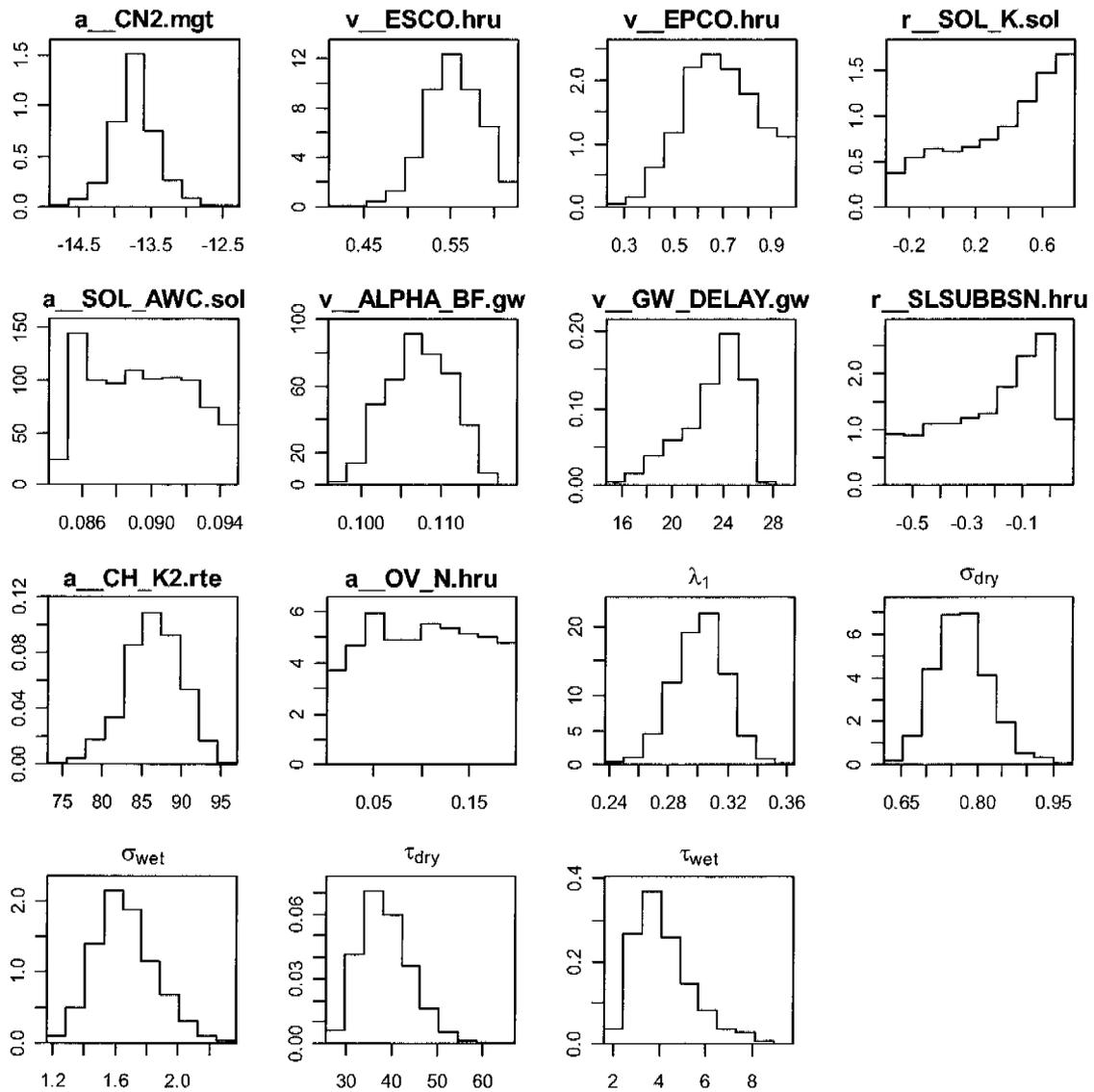


Figure 4.7: Histograms approximating the marginal posterior distributions of aggregate SWAT parameters conditioning with Bayesian MCMC.

Chapter 4

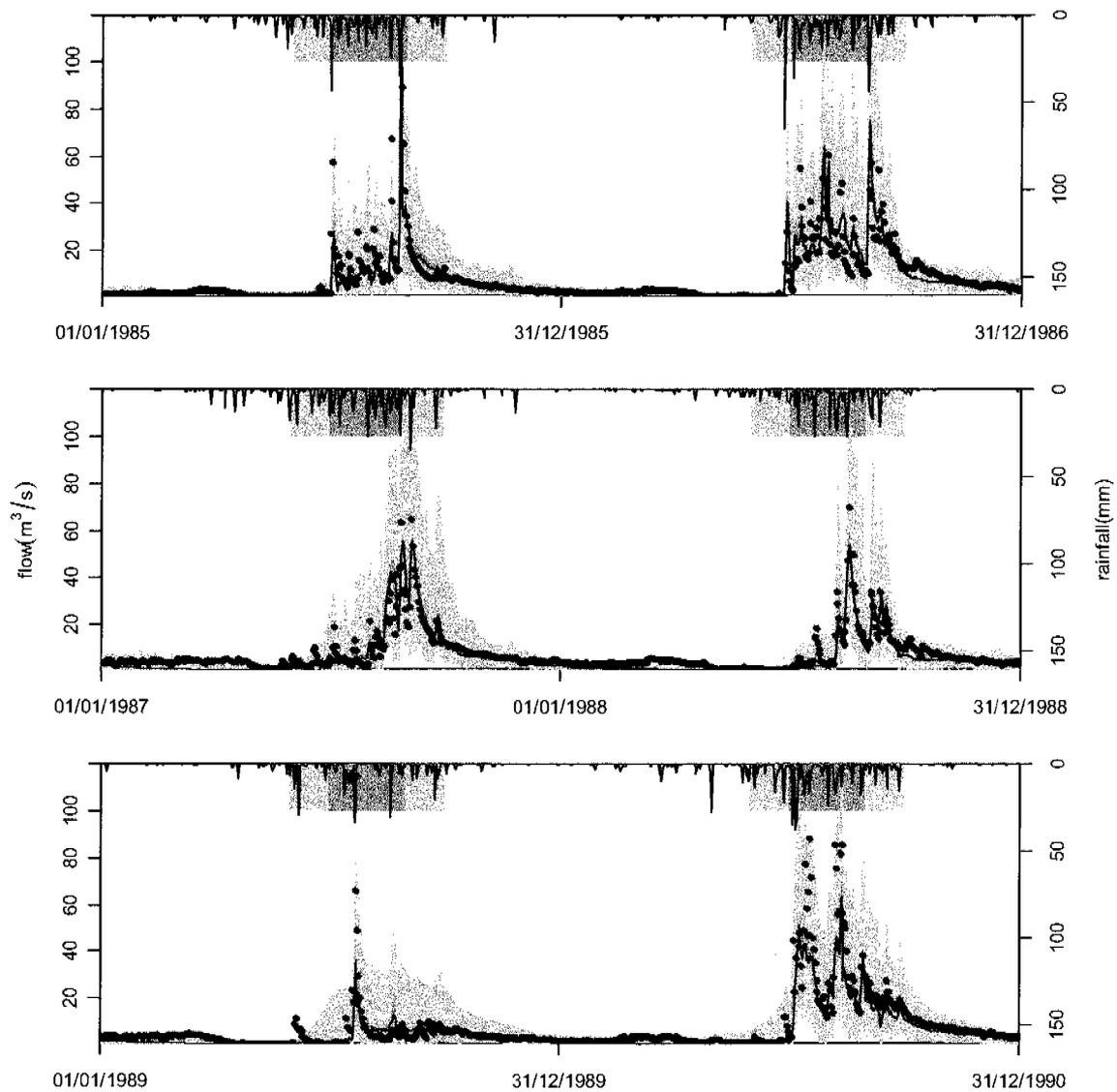


Figure 4.8: 95PPUs associated with parameter uncertainty (dark shaded area) and with total uncertainty (light shaded area) due to due to parameters, input, model structure and output represented by parameter uncertainty and the autoregressive error model during the calibration period (top and middle) and validation period (bottom). The dots correspond to the observed discharge at the basin outlet, while the line stands for the simulated discharge at the maximum of the posterior distribution.

Table 4.5: Comparison of criteria of the different inference and uncertainty analysis techniques (See section 4.2.5 for categories of criteria and text for the interpretation of the differences)

Cat	Criterion	GLUE [NS(Eq.4.2)] ^{*1}	ParaSol [SSQ(Eq.4.3)]	SUFI_2 [NS(Eq.4.2)]	Bayesian inference with cont. autoregr. error model [Log. unnorm. post. prob. den. (Eqs.4.10, 4.11, 4.12)] MCMC	Primitive IS
1	Best estimate and uncertainty range ^{*2} a_CN2.mgt v_ESCO.hru v_EPCO.hru r_SOL_K.sol a_SOL_AWC.sol v_ALPHA_BF.gw v_GW_DELAY.gw r_SLSUBSN.hru a_CH_K2.rte a_OV_N.hru λ σ_{dry} σ_{wet} τ_{dry} τ_{wet}	-16.78 (-29.58, -9.84) 0.76 (0.02, 0.97) 0.22 (0.04, 0.90) -0.16 (-0.36, 0.78) 0.11 (0.01, 0.15) 0.12 (0.06, 0.97) 159.58 (9.72,289.29) -0.45 (-0.56, 0.46) 78.19 (6.01,144.82) 0.05 (0.00, 0.20) - - - - - yes	-20.97 (-21.93, -20.08) 0.67 (0.65, 0.69) 0.16 (0.13, 0.20) -0.37 (-0.41, -0.34) 0.07 (0.08, 0.08) ^{*3} 0.12 (0.08, 0.13) 107.70 (91.23,115.20) -0.59 (-0.60, -0.58) 35.70 (27.72, 37.67) 0.11 (0.07, 0.10) - - - - - yes	-26.9 (-30.00, -7.23) 0.82 (0.43, 1.00) 1 (0.34, 1.00) -0.1 (-0.58, 0.34) 0.07 (0.05, 0.15) 0.51 (0.23, 0.74) 190.07 (100.24,300.00) -0.52 (-0.60, 0.03) 83.95 (69.42,150.00) 0.06 (0.00, 0.11) - - - - - no	-13.75 (-14.35, -13.04) 0.55 (0.49, 0.61) 0.62 (0.40, 0.98) 0.01 (-0.26, 0.78) 0.09 (0.09, 0.09) 0.10 (0.10, 0.11) 24.00 (17.42, 26.11) -0.41 (-0.57, 0.04) 90.18 (78.87, 93.26) 0.19 (0.01, 0.20) 0.31 (0.26, 0.34) 0.73 (0.67, 0.90) 1.38 (1.32, 2.08) 31.12 (29.63, 51.75) 2.48 (2.39, 7.48) yes	-11.57 0.16 0.73 -0.29 0.11 0.41 43.07 -0.39 83.85 0.17 0.34 1.11 6.05 93.47 27.64 yes
2	Parameter correlations NS for calibration NS for validation R^2 for calibration R^2 for validation LogPDF for calibration ^{*4} Log PDF for validation	0.80 0.78 0.80 0.84 -2124 -994	0.82 0.81 0.82 0.85 -2293 -1237	0.80 0.75 0.81 0.81 -2620 -1232	0.77 0.73 0.78 0.81 -1460 -815	0.60 0.51 0.73 0.80 -1662 -884
3	p-factor for calibration ^{*5} p-factor for validation r-factor for calibration ^{*6} r-factor for validation CRPS for calibration CRPS for validation	79% 69% 0.65 0.51 1.64 1.87	18% 20% 0.08 0.07 0.58 0.56	84% 82% 1.03 0.82 1.62 2.03	85%(10%) 85%(7%) 1.48(0.08) 1.16(0.06) 1.90 (0.54) 1.95 (0.57)	- - - - - -

(Table 4.5 continued)

Cat	Criterion	GLUE [NS(Eq.4.2)] ^{*1}	ParaSol [SSQ(Eq.4.3)]	SUFI_2 [NS(Eq.4.2)]	Bayesian inference with cont. autoregr. error model [Log. unnorm. post. prob. den. (Eqs.4.10, 4.11, 4.12)] MCMC	Primitive IS
4	Uncertainty described by parameter uncertainty Source of prediction uncertainty Theoretical basis Testability of stat. assum. Result of test	All sources of uncertainty Parameter uncertainty a. Normalization of generalized likelihood measure b. Primitive random sampling strategy No	Parameter uncertainty only Parameter uncertainty a. Least squares (probability theory) b. SCE-UA based sampling strategy Yes violated	All sources of uncertainty Parameter uncertainty a. Generalized objective function. b. Latin hypercube sampling; restriction of sampling intervals No	Parameter uncertainty only Parameter uncertainty + all other uncertainties described by the autoregressive error model a. Likelihood function (Probability theory) b. MCMC starting from optimal parameter set based on SCE-UA	Parameter uncertainty only Parameter uncertainty + all other uncertainties described by the autoregressive error model a. Likelihood function (Probability theory) b. Primitive random sampling strategy Yes
5	Difficulty of implement. Number of runs	very easy 10000	easy 7500	easy 1500 + 1500	more complicated 5000 - 20'000 + 20'000	more complicated 100'000

*1 The bracketed is the goal function used by the corresponding uncertainty analysis technique

*2 c (a, b) for each parameter means: c is the best parameter estimate, (a,b) is the 95% parameter uncertainty range except SUFI-2 (in SUFI-2, this interval denotes the posterior parameter distribution).

*3 In ParaSol for parameter a_SOL_AWC.sol, the optimal value and sampled range are 0.07416 and (0.07242, 0.085) respectively while 95% parameter uncertainty range is (0.075765, 0.084416). Therefore, the optimal value is outside of the 95% parameter uncertainty.

*4 the σ_{dry} , σ_{wet} , r_{dry} , and r_{wet} used to calculate the Calculate the logarithm of the posterior probability density function (PDF) are from the best of MCMC.

*5 p -factor means the percentage of observations covered by the 95PPU (See section 4.2.3)

*6 r -factor means relative width of 95PPU (See section 4.2.3, defined by Eq. 4.8)

4.4.5. Result of Primitive IS implementation of Bayesian analysis with autoregressive error model

The application with primitive IS is extremely inefficient. In this study, within 100,000 model runs only one parameter set got a weight significantly different from zero. This shows that IS based on the prior as a sampling distribution is too inefficient to be applied to such hydrological problems. An iterative narrowing of the sampling distribution that already starts with a good guess (e.g. close to the maximum of the posterior) would be required to make IS more efficient.

4.4.6. Comparison

Table 4.5 summarizes the results of the comparison in the categories of criteria introduced in section 4.2.3. We will exclude primitive IS from further discussion as obviously the numerical technique of primitive IS from the prior fails to give a reasonable approximation to the posterior at the sample sizes we can afford.

4.4.6.1 Parameter estimates and parameter uncertainty

Results of the marginal posterior parameter distributions are shown as dotted plots in Figures 4.1, 4.3, and 4.5 or marginal distributions in Figure 4.7. In addition, posterior means, standard deviations and correlation matrices of the techniques that provide these estimates are given in Tables 4.2, 4.3, and 4.4. Finally, best estimates and 95% parameter uncertainty ranges are summarized in Table 4.5 (category 1). In General, different techniques result in different posterior parameter distributions, which are represented by different 95% parameter uncertainty ranges, dotted plots and correlation matrices.

Category 1 in Table 4.5 shows the 95% uncertainty ranges of the marginals of all parameters resulting from GLUE, ParaSol and MCMC, and the posterior parameter intervals resulting from SUFI-2. As can be seen, GLUE provided the widest 95% parameter uncertainty ranges, followed by SUFI-2, MCMC and ParaSol. Most of the uncertainty intervals derived by GLUE contain the corresponding intervals from SUFI-2, MCMC and ParaSol. However, not all the parameter intervals derived by SUFI-2 contain the corresponding intervals of MCMC (for example, *a_OV_N.hru*). Some uncertainty intervals from SUFI-2 do not even overlap with those from MCMC (for example, *v_GW_DELAY.gw*). The marginals of GLUE are wider than those of SUFI-2; this may be because GLUE considers parameter correlations while SUFI-2 does not. The posterior shape in SUFI-2 is always a hypercube; therefore wide intervals would lead to too

many simulations with poor performance (poor values of the goal function). The marginals of MCMC are even narrower than those of SUFI-2 because the likelihood function (Eq. 4.11) applied considers input and model structural error separately while GLUE and SUFI-2 map those errors into parameter uncertainty. Therefore, in MCMC, parameter uncertainty contributes only partly to total prediction uncertainty. Different marginals from ParaSol and MCMC illustrate different response surfaces defined by different objective functions.

In principle the global sampling strategy of GLUE allows this technique to identify any shape of the posterior distribution including multimodal shapes. Unfortunately, the number of samples that can be run in practice is too small to realize this conceptual advantage. For example, a comparison between Figures 4.1 and 4.3 demonstrates that GLUE failed to cover the behavioral parameter sets of ParaSol (points above the blue line in Figure 4.3). In this sense, GLUE tends to flatten the true response surface by removing sharp peaks and valleys. This is also the problem of primitive IS (special case of GLUE) and SUFI-2. Primitive IS can only find several isolated points (e.g. its application in this study) because it uses a sampling distribution which is much wider than the posterior. In addition to the difficulty of locating multiple maxima, the hypercube shape of the posterior required for SUFI-2 does not allow this technique to describe multi-modal distributions. Although conceptually Bayesian inference can describe any posterior shape, the implementation of Markov Chain Monte Carlo will usually have a problem to jump from one mode to another in the multi-modal response surface. However, at least the global optimization preceding Markov Chain Monte Carlo helps to find the mode with maximum posterior density. Based on SCE-UA, ParaSol can also locate the best mode in the multi-modal response surface, and its capability to explore other modes is obviously questionable. This leads to the narrow 95% parameter uncertainty ranges listed in Table 4.5.

The parameter correlations in GLUE (Table 4.2) are smaller (the largest equal to 0.67 between $r_SOL_K.sol$ and $r_SLSUBBSN.hru$, and most are below 0.2) compared to those of ParaSol and MCMC (the strongest correlation between parameters of the hydrologic model is between $r_SLSUBBSN.hru$ and $r_SOL_K.sol$ (0.96 and 0.99) in both techniques). This indicates that the (behavioral) parameter sets with significant weight are quite uniformly dispersed over the parameter space. The weaker correlations in GLUE also indicate the phenomenon that the real response surface is flattened by GLUE. In SUFI-2, parameter correlations are neglected.

4.4.6.2 Performance of the simulation at the mode of the posterior distribution

The performances of the simulation at the mode of posterior distribution are listed in category 2 of Table 4.5. It is not astonishing that ParaSol (for *NS*) and MCMC (for log posterior) find the best fit of their respective goal functions because these techniques are based on global optimization algorithms (at least as a first step for MCMC). Such algorithms are much more efficient for finding the maximum of the goal function than random or Latin Hypercube searches. Despite the fact that *NS* is not the objective function of MCMC, the values of *NS* at the maximum of the posterior are not much smaller than those of the techniques which use *NS* as their goal function. The reader can compare other measures of performance in category 2 of Table 4.5.

4.4.6.3 Model prediction uncertainty

Category 3 in Table 4.5 lists the relative coverages of measurements (*p-factor*), the relative width (*r-factor*) and the CRPSs of the 95PPUs for model predictions for all techniques. For the reasons mentioned in section 4.6.1, ParaSol gave too narrow prediction uncertainty bands which are hardly distinguishable from its best prediction (i.e. the one with the best value of the goal function). GLUE and SUFI-2 led to similar *p-factors* but different *r-factors* during the calibration period, and both different *p-factors* and *r-factors* in the validation period. The reason for this may be that the uncertainty width (*r-factor*) of the 95PPU based on GLUE is determined not only by the threshold but also its capability of exploring the parameter space (instead of the multimodal shapes) while that of SUFI-2 is determined by the inclusion of some parameter set with poor goal function in the posterior hypercube. In MCMC, the *p-factors* are similar to those of GLUE and SUFI-2, however, the *r-factor* is a bit higher. This may be because of the overestimation of errors in the input and output and model structure. It is worth nothing that the coverage (*p-factor*) of GLUE and modified ParaSol can be increased at the expense of increasing *r-factor* by decreasing the threshold. This is not true for MCMC as the coverage does not depend on arbitrary threshold of the technique.

An examination on the dynamics of these 95PPUs in Figures 4.2, 4.4, 4.6 and 4.8 reveals: uncertainty analysis techniques based on *NS* show a better coverage in the recession part of the hydrographs than other parts (e.g., peak part), and there is also a clear yearly variation (overestimated in 1986, 1987 and 1990) for GLUE and SUFI-2, while MCMC has a better balance between years, but there seems to be a slight overestimation of prediction uncertainty

during the wet season. The reason is that in the application to the Chaohe Basin the autoregressive error model explicitly specifies the seasonally dependent values of the σ 's and τ 's which reflect the seasonal impacts of input uncertainty, model structural uncertainty and measured response uncertainty. In GLUE and SUFI-2 (at least when applied as in this case study), total uncertainty is expressed as parameter uncertainty, which leads to an equally weighted impact on wet season and dry season.

The CRPS values demonstrate the problem of this measure of combining quality of fit with prediction uncertainty into one common index (i.e. CRPS). The underestimation of prediction uncertainty combined with a good fit leads to the smallest values for ParaSol and MCMC with parameter uncertainty only whereas the values for the other techniques (with wider uncertainty bands) are larger and of a similar magnitude (See category 3 in Table 4.6). This indicates that a further decomposition of CRPS that accounts for different contributions to its value may be necessary in order to make CRPS a useful measure in the present context (see Hersbach, 2000).

4.4.6.4 Conceptual basis of the technique

The crucial criteria with respect of the conceptual basis of the techniques are summarized in category 4 of Table 4.5.

The first two criteria describe how different sources of uncertainty are dealt with. In GLUE and SUFI-2, all sources of uncertainty are mapped to (an enlarged) parameter uncertainty, which will result in wider parameter marginals than ParaSol, MCMC and IS. Parasol ignores other sources of uncertainty except parameter uncertainty. Finally the autoregressive error model maps the effect of input, output and model structure uncertainty to a continuous-time autoregressive error model. As this approach uses extra parameters rather than model parameters it does not enlarge parameter uncertainty.

The conceptual basis of ParaSol, MCMC and IS is probability theory. This has the advantage that the statistical assumptions must be clearly stated and are testable. The statistical assumptions underlying ParaSol (independent, normally distributed residuals) are clearly violated whereas there is no significant violation of the assumptions made by the autoregressive error model (see Yang et al. 2007). The conceptual bases of GLUE and SUFI-2 are different and their statistical bases are weak. GLUE and SUFI-2 allow the users to formulate different likelihood measures (or objective functions) which certainly include the likelihood function used for MCMC (example

Chapter 4

Eq. (4.11)). However, when using generalized rather than ordinary likelihood functions, GLUE and SUFI-2 lose the probabilistic interpretation of the results. In the last step of the GLUE application, weights are normalized and again interpreted as probabilities. This procedure lacks a consistent and testable statistical formulation. Also SUFI-2 lacks a rigorous probabilistic formulation. Parameter uncertainty formulated by a uniform distribution in a hypercube is propagated through the hydrologic model correctly, but the convergence criteria based on the values of *p-factor* and *r-factor* lack an assumption of the dependence structure of the errors.

4.4.6.5 Difficulty of implementation and efficiency

The final category of comparison criteria (category 5) in Table 4.5 is difficulty of implementation and efficiency.

Implementation of GLUE is straightforward and very easy. Due to the calculation of sensitivity measures and global optimization, implementation of SUFI-2 and ParaSol is somewhat more complicated but still quite easy. Due to the most complicated likelihood function and processing technique, the Bayesian techniques need more effort to be implemented.

Due to an efficient optimization procedure, ParaSol does not require extensive computations. Taking into account a relatively poor coverage of the parameter space, SUFI-2 is also not very expensive to run. Depending on the required coverage, GLUE can be run with smaller or bigger sample sizes. The computationally most expensive technique is Bayesian inference. This is certainly the major disadvantage of this technique.

4.5 Conclusions

After comparing the applications of different uncertainty analysis techniques to a distributed watershed model (SWAT) for the Chaohe watershed in North China, we come to the following conclusions:

1) *Application of GLUE based on the Nash-Sutcliffe coefficient.* This technique led to the widest marginal parameter uncertainty intervals of the model parameters (i.e. strong capability of exploring the parameter space), good prediction uncertainty (in the sense of coverage of measurements by the uncertainty bands), and problems of locating multimodal shapes of the posterior due to the inefficiency of global sampling. This technique tends to flatten the response surface defined by the likelihood measure NS. The wide parameter uncertainty ranges (or strong

capability of exploring the parameter space) are primarily caused by the use of the *Nash-Sutcliffe* coefficient as a generalized likelihood measure.

2) *Application of ParaSol based on the Nash-Sutcliffe coefficient.* ParaSol was able to find a good approximation to the global maximum of *NS*, however, it led to too narrow prediction uncertainty bands due to a violation of the statistical assumption of independently and normally distributed errors. Decreasing the threshold value in modified ParaSol increases its prediction uncertainty but the choice of the threshold value may be hard to justify.

3) *Application of SUFI-2 based on the Nash-Sutcliffe coefficient.* This technique could be run with the smallest number of model runs to achieve good prediction uncertainty (reasonable coverage of data points by the prediction uncertainty bands). This characteristic is very desirable for the uncertainty analysis on models which are computationally demanding. However, the choice of a small sample size obviously decreases exploration of the parameter space and this technique faces the same problems as GLUE encounters.

4) *Application of MCMC based on a continuous-time autoregressive error model.* Due to the global optimization performed before starting the Markov Chain, MCMC achieved a good approximation to the maximum of the posterior. The statistical assumptions of the error model are testable and in reasonable agreement with empirical evidence. The additional parameters of the error model give the user some freedom in the description of the effect of input and model structure error (such as seasonal dependence of the magnitude of these effects). The main disadvantages of this technique are the difficulty of constructing the likelihood function, the large number of simulations required to get a good approximation to the posterior, and the difficulty of covering multi-modal distributions caused by the numerical implementation of MCMC.

5) *Application of IS based on a continuous-time autoregressive error model.* The implementation of primitive importance sampling is much too inefficient to get a reasonable approximation to the posterior.

6) *About choosing the objective functions.* GLUE and SUFI-2 are very flexible by allowing for arbitrary likelihood measures / objective functions. On the other hand, GLUE and SUFI-2 lose their statistical basis when using this additional freedom. The real capability of exploring the parameter space is also seriously affected by the choice of the objective functions. In ParaSol, the objective function and the way to split the parameter set are statistically based. However, the

Chapter 4

underlying statistical assumptions are seriously violated. This makes the results unreliable. The likelihood function used for MCMC has a testable statistical basis and the test of our result did not indicate a severe violation of the assumptions. This makes the Bayesian inference which is based on this likelihood function conceptually the most satisfying technique.

Despite these big differences in concepts and performance, GLUE, SUFI-2 and MCMC led to similarly good prediction uncertainty bands. Our preference is for MCMC because Bayesian inference has a sound theoretical foundation and the statistical assumptions underlying the likelihood function based on the autoregressive error model is testable and did not indicate significant violations of the assumptions. However, further efforts are required to improve the formulation of likelihood functions used in hydrological applications. In particular, it would be interesting to formulate a likelihood function that not only describes the effect of input, model structure and output uncertainty (as our autoregressive error model does), but also resolves the different sources of uncertainty.

4.6 Acknowledgement

We would like to thank Dr. Ann van Griensven from UNESCO-IHE for providing the ParaSol program and suggestions, and Dr. Jasper Vrugt from the Los Alamos National Laboratory for constructive comments.

4.7 References

- Abbaspour, K.C., Johnson, C.A., van Genuchten, M.T., 2004. Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose Zone Journal*, 3(4): 1340-1352.
- Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., Srinivasan, R., 2007. Spatially-distributed modelling of hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of Hydrology*, 333:413-430.
- Adeuya, R.K., Lim, K.J., Engel, B.A., Thomas, M.A., 2005. Modeling the average annual nutrient losses of two watersheds in Indiana using GLEAMS-NAPRA. *Transactions of the Asae*, 48(5): 1739-1749.

- Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment - Part 1: Model development. *Journal of the American Water Resources Association*, 34(1): 73-89.
- Bates, B.C., Campbell, E.P., 2001. A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research*, 37(4): 937-947.
- Bekele, E.G., Nicklow, J.W., 2005. Multiobjective management of ecosystem services by integrative watershed modeling and evolutionary algorithms. *Water Resources Research*, 41(10).
- Best, N.G., Cowles, M.K., Vines, S.K., 1995. Convergence Diagnosis and Output Analysis software for Gibbs Sampler output: Version 0.3. Cambridge: Medical Research Council Biostatistics Unit.
- Beven, K., Binley, A., 1992. The Future of Distributed Models - Model Calibration and Uncertainty Prediction. *Hydrological Processes*, 6(3): 279-298.
- Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, 249(1-4): 11-29.
- Bicknell, B.R., Imhoff, J., Kittle, J., Jobs, T., Donigian, A.S., 2000. Hydrological Simulation Program – Fortran User's Manual, Release 12, U.S EPA.
- Blazkova, S., Beven, K., Tacheci, P., Kulasova, A., 2002. Testing the distributed water table predictions of TOPMODEL (allowing for uncertainty in model calibration): The death of TOPMODEL? *Water Resources Research*, 38(11).
- Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. *Journal of the Royal Statistical Society Series B*, 211-252.
- Box, G.E.P., Cox, D.R., 1982. An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association* 77(377), 209-210.
- Brockwell, P.J., Davis, R.A., 1996. *Introduction to Time Series and Forecasting*, Springer, New York.
- Brockwell, P. J., 2001. Continuous-time ARMA processes. In: Shanbhag, D.N., Rao, C.R. (Eds.), *Stochastic Processes: Theory and Methods*, Vol. 19 of *Handbook of Statistics*, Elsevier, Amsterdam, pp. 249–276.

Chapter 4

- Cameron, D., Beven, K., Naden, P., 2000a. Flood frequency estimation by continuous simulation under climate change (with uncertainty). *Hydrology and Earth System Sciences*, 4(3): 393-405.
- Cameron, D., Beven, K., Tawn, J, Naden, P., 2000b. Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation). *Hydrology and Earth System Sciences*, 4(1): 23-34.
- Chow V.T., Maidment, D.R., Mays, L.W., 1988. *Applied Hydrology*. McGraw-Hill, Inc., New York, NY.
- Claessens, L., Hopkinson, C., Rastetter, E., Vallino, J., 2006. Effect of historical changes in land use and climate on the water budget of an urbanizing watershed. *Water Resources Research*, 42(3).
- Cochrane, T.A., Flanagan, D.C., 2005. Effect of DEM resolutions in the runoff and soil loss predictions of the WEPP watershed model. *Transactions of the Asae*, 48(1): 109-120.
- Cotler, H., Ortega-Larrocea, M.P., 2006. Effects of land use on soil erosion in a tropical dry forest ecosystem, Chamela watershed, Mexico. *Catena*, 65(2): 107-117.
- Cowles, M. K., Carlin, B. P., 1996. Markov chain Monte Carlo convergence diagnostics: A comparative review, *Journal of the American Statistical Association*, 91(434), 883-904.
- Cunge, J.A., 1969. On the subject of a flood propagation method (Muskingum method). *J. Hydraulics Research*, 7(2): 205-230.
- Duan, Q.Y., Sorooshian, S., Ibbitt, R. P., 1988. A maximum likelihood criterion for use with data collected at unequal time intervals. *Water Resources Research* 24(7): 1163-1173.
- Duan, Q.Y., Sorooshian, S., Gupta, V., 1992. Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models. *Water Resources Research*, 28(4): 1015-1031.
- Freer, J., Beven, K., Ambroise, 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.
- Gelman, S., Carlin, J.B., Sten, H.S., Rubin, D.B., 1995. *Bayesian Data Analysis*, Chapman and Hall, New York, USA.
- Gupta, H.V., Beven, K.J., Wagener, T., 2005. Model Calibration and Uncertainty Estimation. In: Anderson, M.G. (Eds.), *Encyclopedia of Hydrological Sciences*, John Wiley & Sons, Ltd., pp. 2015-2031.

- Heidelberger, P., Welch, P.D., 1983. Simulation Run Length Control in the Presence of an Initial Transient, *Operations Research*, 31(6), 1109-1144.
- Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather and Forecasting*, 15(5), 559-570.
- Hornberger, G.M., Spear, R.C., 1981. An Approach to the Preliminary-Analysis of Environmental Systems. *Journal of Environmental Management*, 12(1): 7-18.
- Hossain, F., Anagnostou, E.N., Lee, K.H., 2004. A non-linear and stochastic response surface method for Bayesian estimation of uncertainty in soil moisture simulation from a land surface model. *Nonlinear Processes in Geophysics*, 11(4): 427-440.
- Huang, Y.F., Zou, Y., Huang, G.H., Maqsood, I., Chakma, A., 2005. Flood vulnerability to climate change through hydrological modeling - A case study of the swift current creek watershed in western Canada. *Water International*, 30(1): 31-39.
- Hundecha, Y., Bardossy, A., 2004. Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, 292(1-4): 281-295.
- Jorgeson, J., Julien, P., 2005. Peak flow forecasting with radar precipitation and the distributed model CASC2D. *Water International*, 30(1): 40-49.
- Kuczera, G., 1983. Improved Parameter Inference in Catchment Models .1. Evaluating Parameter Uncertainty. *Water Resources Research*, 19(5): 1151-1162.
- Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology*, 211(1-4): 69-85.
- Pednekar, A.M., Grant, S.B., Jeong, Y., Poon, Y., Oancea, C., 2005. Influence of climate change, tidal mixing, and watershed urbanization on historical water quality in Newport Bay, a saltwater wetland and tidal embayment in southern California. *Environmental Science & Technology*, 39(23): 9071-9082.
- Reichert, P., Schervish, M., Small, M.J., 2002. An efficient sampling technique for Bayesian inference with computationally demanding models, *Technometrics* 44(4), 318-327.
- Reichert, P., 2005. UNCSIM - A computer programme for statistical inference and sensitivity, identifiability, and uncertainty analysis. In: Teixeira, J.M.F., Carvalho-Brito, A.E. (Eds.), *Proceedings of the 2005 European Simulation and Modelling Conference (ESM 2005)*, Oct. 24-26, Porto, Portugal, EUROSIS-ETI, pp. 51-55.

Chapter 4

- Reichert, P., 2006. A standard interface between simulation programs and systems analysis software. *Water Science and Technology* 53(1), 267-275.
- Santhi, C., Arnold, J.G., Williams, J.R., Hauck, L.M., Dugas, W.A., 2001. Application of a watershed model to evaluate management effects on point and nonpoint source pollution. *Transactions of the Asae*, 44(6): 1559-1570.
- USDA Soil Conservation Service, 1972. Section 4: Hydrology. In: *National Engineering Handbook*. SCS.
- Todini, E., 1988. Rainfall-Runoff Modeling - Past, Present and Future. *Journal of Hydrology*, 100(1-3): 341-352.
- Van Griensven, A., Meixner, T., 2006. Methods to quantify and identify the sources of uncertainty for river basin water quality models. *Water Science and Technology*, 53(1): 51-59.
- Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, 39(8): art. no.-1201.
- Williams, J.R., 1995. Chapter 25: The EPIC model. In: V.P. Singh (eds.), *Computer models of Watershed hydrology*, pp. 909-1000.
- Young, R.A., Onstad, C.A., Bosch, D.D., Anderson, W.P., 1989. Agnps - a Nonpoint-Source Pollution Model for Evaluating Agricultural Watersheds. *Journal of Soil and Water Conservation*, 44(2): 168-173.
- Yang, J., Abbaspour, K.C., Reichert, P., 2005, Interfacing SWAT with Systems Analysis Tools: A Generic Platform. In: Srinivasan, R., Jacobs, J., Day, D., Abbaspour, K. (Eds.), 2005 3rd International SWAT Conference Proceedings, July 11-15, Zurich, Switzerland, pp. 169-178.
- Yang, J., Reichert, P., Abbaspour, K.C., Yang, H., 2007. Hydrological Modelling of the Chaohe Basin in China: Statistical Model Formulation and Bayesian Inference. Accepted by *Journal of Hydrology*.
- Zacharias, I., Dimitriou, E., Koussouris, T., 2005. Integrated water management scenarios for wetland protection: application in Trichonis Lake. *Environmental Modelling & Software*, 20(2): 177-185.

5 Conclusions and Outlook

In this section, firstly the main results from the previous sections are summarised, followed by the conclusions. In addition, a brief outlook on further research is provided.

5.1 Summary

In Section 2 we developed a continuous-time autoregressive error model within the Bayesian framework. The characteristic of this method is to construct the likelihood function in combination with the Box-Cox transformation and autoregressive error model in such a way that the observed innovations rather than the residuals are independently and normally distributed. The implementation is demonstrated by its application to the Chaohe Basin in North China. In the application, the statistical assumptions are tested and fulfilled. Statistical inference is numerically implemented by a global maximization of the posterior followed by Markov Chain Monte Carlo sampling.

In Section 3, the developed continuous-time autoregressive error model is extended by assuming that the innovations between the simulation and observation are realizations of independent t-distributions. As the degrees of freedom approaches infinity, the t-distribution approaches the normal distribution. Therefore, we can take the normal distribution as a special case of the t-distribution with sufficiently large degrees of freedom. With this additional parameter (i.e. degrees of freedom), it is possible to adapt the model to best match the shape of the empirical distribution of the innovations. This implementation is demonstrated by the application to the Thur river basin in Switzerland.

In Section 4, different UA approaches are compared. The comparison turns out to be difficult as these approaches differ in fundamental concepts, parameter prior choices, and goal function choice. Nevertheless, we conclude with the recommendation of the use of Bayesian inference with a carefully chosen likelihood function that accounts for all uncertainty sources, whenever it is possible. However, the computational demand involved may make this choice very difficult to apply.

5.2 Conclusions

Based on the work of the previous chapters, this study concludes:

1) The continuous-time autoregressive error model is applicable and efficient for the analysis of the effect of parameter uncertainty, uncertainties in the input, response and model structure in hydrologic modeling.

2) The assumption of independent t-distributions for describing the distribution of the innovations rather than the residuals makes the likelihood function have better statistical basis because the innovations are much less correlated and the shape of the t-distribution fits empirical evidence much better (particularly heavy tails). The continuous-time autoregressive error model can also address the problems of heteroscedasticity of the residuals by a combination of a Box-Cox transformation with seasonally dependent parameters of the error model.

3) Although different UA techniques are based on different fundamental concepts, prior parameter choices and goal functions, many of them lead to similar prediction uncertainties. Therefore, the choice of the UA techniques in hydrologic modelling depends on the preference of the modeller and the application at hand.

4) Because of the sound statistical basis and the testability of model assumptions, our preference is on the Bayesian approach with a careful choice of the likelihood function. The good reproduction of the empirical error distributions by our continuous-time autoregressive error model increases our confidence in predicted uncertainty bands. Computational limitations, however, may make this choice difficult for very time-consuming models.

5) The limitation of this method is that it doesn't separate input uncertainty from model structural uncertainty. And compared to other techniques, the major disadvantage is the high computational demand characteristic for all MCMC techniques.

5.3 Outlook

Building on this research and to the author's knowledge, the following research areas are potentially interesting and would be worth further research.

1 Further comparison of different UA techniques

A comparison of techniques based on a case study is obviously case and model dependent. More studies based on other watersheds and other hydrologic models should be done to investigate the advantage and disadvantage of different UA techniques.

2 Explicit consideration of input and model structural uncertainty

As mentioned in the conclusion part, the limitation of our approach is that it doesn't separate input uncertainty from model structural uncertainty. Recently, a method BATEA (Bayesian Total Error Analysis methodology; Kavetski et al., 2006a & 2006b) has been developed that considers input uncertainty within a Bayesian framework. BATEA uses additional (latent) variables to account for time-dependent input errors. Approaches to consider model structural errors by making the deterministic simulation model stochastic have also recently been published (Vrugt et al. 2005). A combination of such approaches and an extension to continuous-time formulation as done in the present study would be an interesting research field for the future.

3 Improvement of the numerical efficiency of Bayesian computation

The most limiting practical problem for applying Bayesian inference with large computational models is the relatively poor efficiency of the Markov Chain Monte Carlo approach. Optimization of the efficiency of this approach or development of alternative numerical approximations to the posterior would considerably extend the range of applicability of these techniques.

5.4 References

- Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resources Research*, 42, W03407, doi:10.1029/2005WR004368.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resources Research*, 42, W03408, doi:10.1029/2005WR004376.
- Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resources Research*, 41, W01017, doi:10.1029/2004WR003059.

Seite Leer /
Blank leaf

Appendix A:

Interfacing SWAT with Systems Analysis Tools: A Generic Platform

Jing Yang, Karim C. Abbaspour, and Peter Reichert

Abstract: Complex hydrologic watershed models need sophisticated techniques for statistical inference of parameters and uncertainty estimation of predictions. To perform such analyses, watershed simulation programs must be linked to systems analysis software. It is inefficient to do this by implementing a large set of systems analytical techniques directly into each simulation program. The more useful strategy is to implement a flexible interface independent of the simulation program that allows the user to link the simulation program to external systems analysis software. In this paper, the requirements for such an interface for distributed hydrological simulation programs are analyzed, and the implementation of such an interface for the Soil and Water Assessment Tool (SWAT) is described. The discussion of these requirements and the concepts of implementation are intended to stimulate similar development for other simulation programs; the implementation itself, which is freely available, facilitates the combination of systems analysis techniques with SWAT applications.

Keywords: Calibration; Interface; Systems Analysis Tools; SWAT; Watershed Models; Simulation Program

Software Availability

Name of product: iSWAT

Program language: C++

Software requirements: SWAT 2000/SWAT2005

Hardware requirements: PC with MS Windows or Linux (the interface could be compiled for other platforms easily)

Contact Address: Jing Yang, Eawag, Ueberlandstr 133, P.O.Box 611, 8600 Dübendorf, Switzerland, jing.yang@eawag.ch

Appendix

Availability: Source code and binaries for MS Windows and Linux can be freely downloaded from <http://www.uncsim.cawag.ch/interfaces/SWAT>

1 Introduction

Calibration of hydrologic models and uncertainty analyses of their predictions is a very active research field as it challenges current methodological know-how and is of highly practical relevance (Duan et al. 2002). Nevertheless, implementations of these hydrologic models are poorly supported by systems analysis programs. Improving this support is very important, as research requires the application of different and new techniques (e.g., uncertainty analysis) to a diversity of hydrologic simulation programs (e.g. Duan et al. 2002; Kavetski et al. 2006; Vrugt et al. 2006). To improve the support for such analyses, we need a generic interface between hydrologic simulation program and systems analysis software because this is much more flexible than implementing systems analysis techniques directly into each hydrologic simulation program. Recently, a simple version of such an interface has been proposed (Reichert, 2006). This paper describes how a interface is implemented for the Soil and Water Assessment Tool (SWAT; Arnold 1998). The implementation for SWAT includes 2 parts: one is to define a term “aggregate parameters” (which combines the parameter’s name and its influential factors, and this will lead to meaningful calibration and reduction of parameter number for calibration), the other is to communicate the “aggregate parameter” between the simulation program (SWAT executable program) and systems analysis programs. This paper is an updated version of an earlier conference contribution (Yang et al. 2004). Its intention is to stimulate the discussion on how to improve the application of systems analytical techniques for hydrologic models among hydrologists, to demonstrate the usefulness of the concept of flexibly defining “aggregate parameters” for distributed hydrologic models, and to provide a first introduction of the particular interface to future users.

This paper is structured as follows: In section 2 requirements for a generic interface between distributed watershed simulation program and systems analysis software are discussed. Section 3 contains a short summary of the features and structure of input and output files of SWAT. The interface is described in section 4. Section 5 gives a very brief overview of first applications of the interface. Finally, in section 6 conclusions are drawn.

2 Specific Interface Requirements for Distributed Watershed Models

Due to the high geographic resolution and the large number of processes required for a realistic description of watershed hydrology and the sparse data that is either local (soil properties, precipitation, etc) or integrative (e.g., river discharge), distributed hydrologic models are inevitably over-parameterized. In addition, most of parameters in distributed hydrologic models should be determined through so-called model calibration for some reason (for example, some parameters represent spatial averages of system properties) and at the same time are influenced by factors such as landuse type, soil texture, etc. To deal with these problems, it is useful to reduce the number of parameters for analysis by aggregating the distributed parameters in different ways. This will greatly increase the flexibility of the users' choice. Such aggregations should include applying relative or absolute changes to default parameter values (to keep the structure of spatial variation), to apply the same parameter values for the same soil and land use types, to regionally differentiate parameter values, etc.

3 The Soil and Water Assessment Tool (SWAT)

SWAT (Arnold et al., 1998) is a watershed simulation program that was originally developed by a research team in the US Department of Agriculture. SWAT solves water balances in hydrologic response units (HRUs) which are defined by unique land use – soil type combinations within sub-basins of the watershed. For each HRU the water balance is calculated considering precipitation, evapotranspiration, runoff, infiltration, interflow, and percolation into a shallow aquifer. River flow is routed downstream to the outlet of the watershed. The current version of SWAT (SWAT 2000) is linked to ArcView GIS (ESRI, <http://www.esri.com>) in order to facilitate handling of input and output. SWAT implements also a water quality submodel describing transport of sediment, and transport and transformation of nutrients and pesticides. Running SWAT is based on a three-step procedure:

1. In the first step, an ArcView GIS interface of SWAT (AVSWAT) is used to delineate sub-basins from digital elevation data, and then generate HRUs within each sub-basin by overlaying the soil and land use maps. As a final step, AVSWAT produces a large number of input text files. The content of these input text files and their corresponding spatial levels are summarized in Table 1.

Appendix

Table 1: Example of file types, file levels and corresponding parameter information.

File type (extension)	Spatial Level	Description
bsn	Basin level	Basin input file, containing parameters used for the whole basin, such as the snowmelt factor.
wwq	Basin level	Watershed water quality input file containing parameters used by the QUAL2E model applied in the main channels.
crp	Basin level	Land cover / plant growth database file containing plant growth parameters for all land covers simulated in the watershed.
pnd	Sub-basin level	Pond and wetland input file containing parameter information used to model the water, sediment, and nutrient balance for ponds and wetlands.
rte	Sub-basin level	Main channel routing input file containing parameters governing water and sediment movement in the main channel of the sub-basin.
sub	Sub-basin level	Sub-basin input file containing information related to features within the sub-basin, such as properties of tributary channels.
swq	Sub-basin level	Stream water quality input file containing parameters used to model pesticide and QUAL2E nutrient transformations in the main channel of the sub-basin.
wgn	Sub-basin level	Weather generator input file containing the statistical data needed to generate representative daily climate data for the sub-basin.
wus	Sub-basin level	Water use input file containing information for consumptive water use in the sub-basin.
chm	HRU level	Soil chemical input file containing information about initial nutrient and pesticide levels of the soil in the HRU.
gw	HRU level	Groundwater input file, containing information about the shallow and deep aquifer in the sub-basin.
hru	HRU level	HRU input file, containing information related to a diversity of features within the HRU, such as parameters affecting surface and subsurface water flow.
mgt	HRU level	Management input file, containing management scenarios simulated in the HRU.
sol	HRU level	Soil input file containing parameters about the physical characteristics of the soil in the HRU.

2. In the second step, the FORTRAN program “swat2000” reads these text input files, performs the simulation, and writes text output files.

3. ArcView-SWAT (AVSWAT) provides limited post processing capabilities and other programs must be used for output manipulation and display.

After the initial setup of a SWAT project, the text file-based project can be run and analyzed independent of the AVSWAT interface. This text file based project provides the easiest access for the implementation of an interface with systems analysis programs.

4 Interface Description

4.1 General Concept

The general concept of the interface is based on the suggestion by Reichert (2006). The systems analysis program first writes parameter names and corresponding values into the file “model.in”, then executes the simulation program (the part which is enclosed by the dashed box in Figure 1 in our case), and finally reads model result from the file “model.out” written by the simulation program (see Figure 1).

To allow for aggregate parameters (see below) and to make the interface independent of the SWAT code (as long as the input-file format is not changed), the interface is implemented as two executables (i.e., `sw_edit` and `sw_extract` as introduced below). The simulation program called by the system analysis program calls three executables:

1. First, the executable “`sw_edit`” is called. This executable first reads “aggregate parameters” names and values from the file “model.in” and default parameter values from a backup directory with SWAT input files, and then modifies SWAT input files accordingly.
2. Then “swat2000” is called to perform the simulation and write SWAT output files.
3. Finally, the interface program “`sw_extract`” is called. This executable extracts the values of selected output variables from the SWAT output files and writes them to the file “model.out”.

Appendix

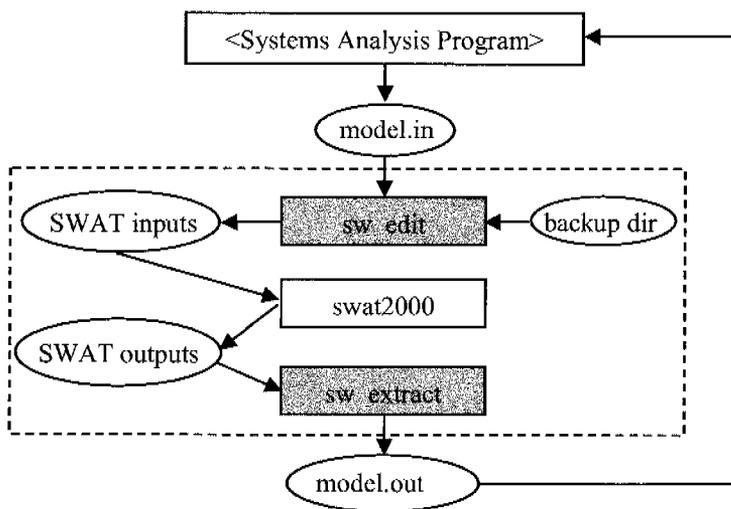


Figure 1: Interaction of the systems analysis program with the simulation program (enclosed by the dashed box) which consists of the interface programs “sw_edit” and “sw_extract”, and the SWAT simulation engine “swat2000”.

4.2 Aggregate Parameters

As mentioned in section 2, the interface should provide aggregate parameters in addition to the independent distributed SWAT parameters to the systems analysis procedures. In the interface iSWAT, parameter aggregation is implemented by encoding this information in a generalized parameter name. The structure of this name is as follows:

x_<parname>.<ext>__<hydrogrp>__<soltext>__<landuse>__<subbsn>

where

- x** Code to indicate the type of change to be applied to the parameter:
- v means the existing parameter value is to be replaced by the given value,
 - a means the given value is added to the default parameter value, and
 - r means the default parameter value is multiplied by (1+the given value);

<parname> SWAT parameter name;

<ext> SWAT file extension code for the file containing the parameter;

<hydrogrp> soil hydrological group ('A', 'B', 'C' or 'D') (optional);

<soltext> soil texture (optional);

- <landuse>** name of the land use category (optional);
- <subbsn>** number(s) of subbasin, crop index, fertilizer index and rainfall date (optional). The typical formats are “1”, “3-5” or “1,3-5,10-21,22” etc. The meaning depends on the extension code of the file containing the parameter (see <ext> above):
- sub: subbasin number(s),
 - crp: crop index,
 - frt: fertilizer index,
 - pcp: date(s) of the rainfall data.

The influential factors <hydrogrp>, <soltex>, <landuse>, and <subbsn> can be omitted, if the change applied to the (distributed) parameters is global. Any combination of these influential factors can be used to make distributed parameters dependent on important influential factors. The parameters can be kept regionally constant, modify a prior spatial pattern, or be changed globally. This gives the analyst a large freedom in selecting the complexity of distributed parameters. By using this flexibility, a calibration process can be started with a small number of aggregate parameters that only modify a given spatial pattern and with more complexity and regional resolution added as the learning process proceeds.

4.3 Implementation

There are a large number of SWAT model parameters distributed over a large number of input files (see section 3 and particularly Table 1). The number of files increases with the watershed disaggregation into sub-basins, and consequently the number of HRUs increases. To make the process of modifying input files efficient, this interface was implemented in C++ in two steps. In a first step, a library was built of classes corresponding to the input files listed in Table 1. The common functions for reading and writing parameter values from and to the text files are declared in a base class (CBaseParaFile) and implemented in the file-type specific classes. The functions for reading the output files of SWAT are implemented in a separate class (CMethodSet). This leads to the class hierarchy described in Table 2 and shown in Figure 2. In the second step, an application program called “sw_edit” is constructed to read the aggregate parameters and their values from “model.in”, and then parse the generalized parameter name, and change the SWAT input files by constructing the relevant class object and calling the member function of the object. Application “sw_extract” is also constructed to extract SWAT output.

Appendix

Table 2: Declared and implemented methods of the classes for modifying/extracting the SWAT input/output files

Class	Description	Declared methods not yet realized	Realized methods
CMethodSet	Collection of general purpose functions, of functions for collecting general SWAT project information, and of functions for extracting information from SWAT output.		<ul style="list-style-type: none"> - Procedure for different types of parameter modification some statistical functions. - Procedure for collecting general SWAT project information such as the number of subbasins. - Procedures for extracting output from .rch, bsb, and bsb files.
CBaseParaFile	Base class for swat input files.	<ul style="list-style-type: none"> - Procedure for reading a parameter value. - Procedure for changing a parameter value. 	<ul style="list-style-type: none"> - Procedure for retrieving the current file name, getting or checking the range of a given parameter.
CBasinBSNFile CBasinWWQFile CCropFile CHruCHMFile CHruGWFile CHruHRUFile CHruMGTFFile CHruSOLFile CSubPNDFile CSubRTEFile CSubSUBFile CSubSWQFile CSubWGNFile CSubWUSFile	Related to .bsn file. Related to .wwq file. Related to crop.dat file. Related to .chm file. Related to .gw file. Related to .hru file. Related to .mgt file. Related to .sol file. Related to .pnd file. Related to .rte file. Related to .sub file. Related to .swq file. Related to .wgn file. Related to .wus file.		<ul style="list-style-type: none"> - Procedure for reading a parameter value. - Procedure for changing a parameter value.

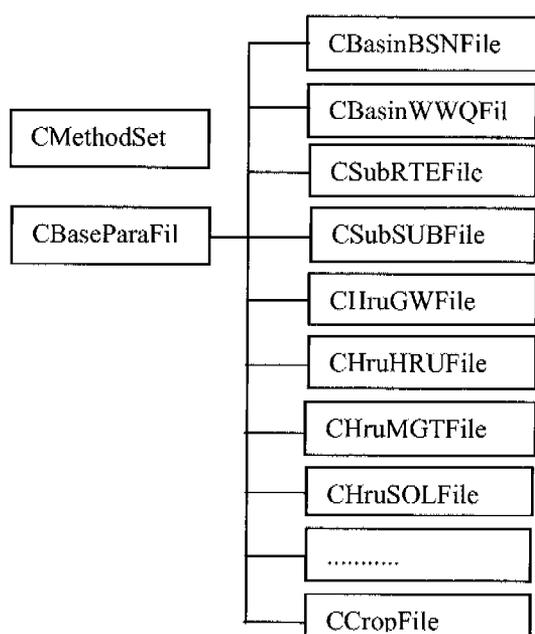


Figure 2: Class hierarchy for modifying or extracting the SWAT input/output files.

5 Applications

So far, the interface has been applied with the systems analysis tools SUFI (Abbaspour et al. 2004; 2007) and UNCSIM (Reichert 2005; <http://www.uncsim.eawag.ch>). These applications are described in Abbaspour et al. (2007) and Yang et al. (2007a, b, c). Without this interface, it would not have been possible to apply different systems analytical techniques in such a flexible way to SWAT applications.

6 Conclusions

Calibration of watershed models and uncertainty analysis of their prediction can significantly be improved by (i) a generic interface between the simulation program of the watershed model and systems analysis software, and (ii) a flexible way of aggregating distributed parameters to reflect important dependencies on soil properties and land use types and region (e.g. to reflect altitude or other regional influence factors) or by modifying default spatial distributions.

The interface we described to the hydrological simulation program “Soil and Water Assessment Tool” (SWAT) provides these two features. The generic interface is based on a simple text-file based data exchange between simulation program and systems analysis software.

Appendix

It uses “aggregate parameters” names to encode dependency information of the distributed parameters and modifies hundreds of SWAT files accordingly. The flexibility of this parameter aggregation scheme allows the researcher to start the analysis with a small number of highly aggregate parameters that only modify a given spatial pattern and adding more complexity and regional resolution later on as the analysis proceeds. This is not possible to achieve by the standard editors (provided by the operating system) or standard features of SWAT. The interface proved its practical value with several studies of SWAT model on calibrations and uncertainty analysis in different river basins.

We hope that the general concept of this interface will be adopted (and improved) for other hydrological simulation programs. This would significantly facilitate comparative studies of systems analytical techniques to different watershed models. In addition, the described iSWAT interface could stimulate comparative analyses of different systems analytical techniques to SWAT applications. The interface can be freely downloaded from <http://www.uncsim.eawag.ch/interfaces/SWAT>.

References

- Abbaspour, K.C., Johnson, A., van Genuchten, M.T, 2004. Estimation of uncertain flow and transport parameters by a sequential uncertainty fitting procedure: SUFI-2. *Vadose Zone Journal* 3(4), 1340-1352.
- Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., Srinivasan, R., 2007. Spatially-distributed modelling of hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of Hydrology*, 333:413-430.
- Arnold, J.G., Srinivasan R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment - Part 1: Model development. *Journal of the American Water Resources Association* 34(1), 73-89.
- Duan, Q., Gupta, H.V., Sorooshian, S., Rousseau, A.N. and Turcotte, R., eds. 2002. *Calibration of Watershed Models*. American Geophysical Union, Washington, DC, USA.
- Kavetski, D., Kuczera, G. and Franks, S.W., 2006. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, 42(3).

- Reichert, P., 2005. UNCSIM - A computer programme for statistical inference and sensitivity, identifiability, and uncertainty analysis, In: Teixeira, J.M.F. and Carvalho-Brito, A.E. (Eds.), Proceedings of the 2005 European Simulation and Modelling Conference (ESM 2005), Oct. 24-26, Porto, Portugal, EUROSIS-ETI, pp. 51-55.
- Reichert, P. 2006. A standard interface between simulation programs and systems analysis software. *Water Science and Technology* 53(1), 267-275.
- Vrugt, J.A., Robinson, B.A., Duan, Q. 2006. Treatment of Uncertainty Using Ensemble Methods: Comparison of Sequential Data Assimilation and Bayesian Model Averaging, submitted.
- Yang, J., Abbaspour, K.C. and Reichert, P. 2004. Interfacing Watershed Models with Systems Analysis Tools: Implementation for SWAT, Proceedings of the 6th International Symposium on Systems Analysis and Integrated Assessment, Watermatex 2004, Beijing, Nov. 3-5, 2004.
- Yang, J., Abbaspour, K.C., Reichert, P., 2007a. Comparing different uncertainty analysis techniques for a SWAT application to Chaohe Basin in China. Submitted to *Journal of Hydrology*.
- Yang, J., Reichert, P., Abbaspour, K.C., 2007b. Bayesian uncertainty analysis in distributed hydrologic modelling: A case study in the Thur river basin (Switzerland). Submitted to *Water Resources Research*
- Yang, J., Reichert, P., Abbaspour, K.C., Yang, H., 2007c. Hydrological Modelling of the Chaohe Basin in China: Statistical Model Formulation and Bayesian Inference. Accepted by *Journal of Hydrology*.

Appendix B: Explanation of some terms

1. Introduction to Autoregressive model and Innovation

A 1st-order Discrete-time Autoregressive Model (DAR(1)) can be formulated as follows:

$$y_t = \theta \cdot y_{t-1} + i_t$$

where y_t , and y_{t-1} are the values of the variable y at time t and $t-1$, θ is the autoregressive coefficient, and i_t is a random perturbation term, also called *innovation*, as this is the only new information that enters at time t , with respect to what it is already available from previous time.

A 1st-order Continuous-time Autoregressive Model (CAR(1)) can be formulated as follows:

$$y_{t_i} = \exp\left(-\frac{t_i - t_{i-1}}{\tau}\right) \cdot y_{t_{i-1}} + i_t$$

where τ is the characteristic correlation time, and the relationship between θ and τ is $\theta = \exp\left(-\frac{1}{\tau}\right)$. Compared to DAR(1), in CAR(1) the difference between y_t and y_{t-1} does not necessarily be 1.

2. Normal distribution and t distribution

Statistically, a random variable x of Normal distribution with mean μ and variance σ^2 ($N(\mu, \sigma^2)$) has the probability density function:

$$f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right)$$

And t distribution with degrees of freedom ν , mean μ and variance σ^2 :

$$f(x | \nu, \mu, \sigma^2) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi(\nu-2)}} \frac{1}{\sigma} \left(1 + \frac{(x - \mu)^2}{(\nu-2)\sigma^2}\right)^{-\frac{\nu+1}{2}}, \nu > 2$$

Curriculum Vitae

Jing Yang

born on 22 May 1977 in Hunan, China

Education:

05/2002 – 09/2006 PhD in Environmental Science at the Swiss Federal Institute of Aquatic Science and Technology (Eawag) and the Swiss Federal Institute of Technology (ETH) Zurich, Switzerland

Under the supervision of

Prof. Dr. Peter Reichert

Prof. Dr. Alexander J.B. Zehnder

Dr. Karim C. Abbaspour

Dr. Hong Yang

09/1998 – 07/2001 Master thesis in water resources, Wuhan University, China

09/1995 – 07/1999 Bachelor in water resources, Wuhan University of Hydropower Electric Engineering, China

09/1992 – 07/1995 High School, China

09/1989 – 07/1992 Middle School, China

09/1983 – 07/1989 Primary School, China

Employment Record:

07/2001 – 05/2002 SuperMap GIS Technologies, Inc. Beijing, China.

Publications:

Peer Reviewed Journals:

Yang, J., Reichert, P., and Abbaspour, K.C., Sensitivity analysis on a distributed hydrology model: A comparison of different techniques (In preparation)

Yang, J., Abbaspour, K.C., Reichert, P., and Yang, H., Comparing the applications of uncertainty analysis techniques for a distributed watershed model. (Submitted to Journal of Hydrology)

Yang, J., Reichert, P., and Abbaspour, K.C., Bayesian uncertainty analysis in distributed hydrologic modelling: A case study in the Thur river basin (Switzerland). (Submitted to Water Resources Research)

Yang, J., Reichert, P., Abbaspour, K.C., and Yang, H., Hydrological Modelling of the Chaohe Basin in China: Statistical Model Formulation and Bayesian Inference. (Accepted by Journal of Hydrology)

Abbaspour, K.C., **Yang, J.**, Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., Srinivasan, R., 2007. Spatially-distributed modelling of hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. Journal of Hydrology, 333:413-430.

Conference Proceedings:

van Griensven, A., Holvoet, K. and **Yang, J.**, 2006. Open sources model codes: critical review on the routing modules of the Soil and Water Assessment Tool (SWAT). The 7th international Conference on Hydroinformatics, September 4-8, 2006, Nice, France

Yang, J., Reichert, P., Abbaspour, K.C., 2006. A Bayesian methodology for estimating prediction uncertainty of a distributed watershed model. Geophysical Research Abstracts, Vol. 8, 08721, European Geosciences Union 2006.

Yang, J., Abbaspour, K. C., and Reichert, P., 2004. Interfacing watershed models with systems analysis tools: implementation for SWAT. In Proceedings of the 6th International Symposium on Systems Analysis and Integrated Assessment, Watermatex 2004. Beijing.

Acknowledgements

Until my dissertation goes here, I could sit down to thank all the people who have given and been giving kind support and assistance from my professional and private surrounding.

Firstly of all, I would like to thank my supervisors Peter Reichert, Alexander J.B. Zehnder, Karim C. Abbaspour, and Hong Yang for their engagement and professional guidance. And without them, I would not have had the chance to pursue my study in this beautiful country - Switzerland. Sascha is always very kind and friendly whenever I come to him. Hong has helped me a lot, such as improving my English and organizing my field trip, just to name a few. Karim has given me a lot of help, especially on the SWAT model and many valuable concepts in the modelling world. Peter has guided me in a lot of fields such as the Bayesian world, programming world, and system analysis. Although he is busy, he can always find his time for me, and I will remember for ever the following short conversation which benefits me a lot:

“Peter. Can I have your some minutes?”

“I will have a meeting with ... at 4:30. How long do you need? If it takes long we can schedule it tomorrow”

“Maybe I only need your 5 minutes”

“Five minutes? Ok”

.....

“I really have to go now. Let’s talk about this tomorrow. How about 5:00 tomorrow afternoon?”

I would also like to thank Ann Van Griensven, Raghavan Srinivasan, and Jeff Arnold for solving my questions on SWAT, and Jasper Vrugt for constructive comments on the papers.

I would like to express my thanks to Changming Liu, Jun Xia, and Shaofeng Jia for their kindly support on my China field trip.

Thanks must go to the doctoral students Gangsheng Wang and Wenhui Dong at the Institute of Geographic Sciences and Natural Resources Research, CAS, and Zhuo Yang from Beijing Water Authority for assistance in data collection; and Xinmin Li and Xiuyong Zhang from the Water Authority in Fenning County for guiding my field trip to the Chaohe basin.

I would like to thank Bettina Bluemling and Changshun Liu for the nice cooperation in our doctoral studies.

Thanks also should go to Jienan Yang, Baoping Qin, Ren Sun, Qiulian Wang, Guiling Yang, and Zhen Zhang from the Tianjin Environment Protection Bureau.

My special thanks go to Siamese. They have provided me a lot of help and convenience during my stay. I would never forget the happy time with them, e.g. climbing with Peter, barbecuing with Hong and Karim, joining all well-organized excursions by Karin, enjoying beer with Juergen, going to the cinema with Christian, talking about kindergarten with Markus, cooking with Lei and Junguo, discussing with Mats...

Many thanks also go to the staff working at Eawag for their kind helps.

I would also remember the cheering time with my Chinese friends at Empa/Eawag for ever. And I really enjoyed making dumplings with Wei, Lei, and Yingying.

Special thanks should be given to Johanna and Steffen for preparation for my defence, and Yuan, Juergen and Markus for checking the typos and format of my thesis, and Yingying for final printing.

Last but not least, I am very grateful to my family and my friends in China for their never ending support and encouragement.