Die approbierte Originalversion dieser Dissertation ist an der Hauptbibliothek der Technischen Universität Wien aufgestellt (http://www.ub.tuwien.ac.at).

The approved original version of this thesis is available at the main library of the Vienna University of Technology (http://www.ub.tuwien.ac.at/englweb/).

Doctoral Thesis

Process based flood forecasting: Estimation methods and uncertainties

submitted in satisfaction of the requirements for the degree of
Doctor of Science in Civil Engineering
of the Vienna University of Technology, Faculty of Civil Engineering

---

Dissertation

Prozessorientierte Hochwasservorhersage: Berechnungsmethoden und Unsicherheiten

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaft
eingereicht an der Technischen Universität Wien Fakultät für Bauingenieurwesen
von

Dipl.-Ing. Jürgen Komma
Matrikelnummer 9426081
Austinstrasse 3c/6, 3107 St.Pölten

Gutachter: Univ.Prof. Dipl.-Ing. Dr. techn. Günter Blöschl
Institut für Wasserbau und Ingenieurhydrologie, TU Wien
Karlsplatz 13, 1040 Wien

Gutachter: Univ.Prof. Dipl.-Ing. Dr. techn. Wolfgang Wagner
Institut für Photogrammetrie und Fernerkundung, TU Wien
Gußhausstr.25-29, 1040 Wien

St. Pölten, Dezember 2011

Kurzfassung

Die Verfügbarkeit von genauen Hochwasservorhersagen, bis zu Prognosefristen von mehreren Tagen, wird in zunehmendem Maß auch für kleinere Einzugsgebiete gewünscht. Die Änderung des Anforderungsprofils macht die methodische Weiterentwicklung von Hochwasservorhersagesystemen zu einem wichtigen, interdisziplinären Betätigungsfeld in der Hydrologie, Meteorologie und Fernerkundung. Ziel dieser Arbeit ist die Entwicklung und Analyse von Methoden zur Beschreibung bzw. Reduktion von Unsicherheiten bei der Messung und Modellierung der abflussrelevanten hydrometeorologischen Prozesse. Die Analysen in dieser Arbeit basieren auf Simulationen mit einem flächendetaillierten hydrologischen Niederschlags-Abfluss-Modell zur Beschreibung von Schneeakkumulation und -schmelze, des Bodenfeuchtehaushaltes und des Abflusses am Hang und im Gerinne.

Der erste Teil der Arbeit beschäftigt sich mit der Quantifizierung der Größenordnung der Vorhersageunsicherheit durch die Verwendung von Ensembleprognosen, welche maßgebend durch die vorherrschenden meteorologischen und hydrologischen Randbedingungen bestimmt wird. Die Beurteilung der Prognosen erfolgt dabei durch die Analyse der Warncharakteristik, welche durch die Häufigkeit von Fehlwarnungen und zutreffenden Warnungen beschrieben wird. Die Ergebnisse dieser Arbeit zeigen, dass der abgeleitete Unsicherheitsbereich eine wertvolle Zusatzinformation für die Risikoabschätzung im Rahmen der Entscheidungsfindung beim Hochwassermanagement darstellt.

Der zweite Teil der Arbeit beschäftigt sich mit der Reduktion von hydrologischen Unsicherheiten durch die Einbeziehung von aktuell verfügbaren Abflussmessungen. Die Zielsetzung, den Prognosefehler so klein als möglich zu halten, kann durch die Verbindung der Modellergebnisse mit den Messdaten unter Abwägung der jeweils eingetragenen Unsicherheiten erreicht werden. Diese Vorgangsweise wird als Modellnachführung bezeichnet, und erfolgt in dieser Arbeit durch die Implementierung eines Ensemble-Kalman-Filters in die Modellstruktur. Zusätzlich erfolgt, abhängig von der jeweils vorherrschenden hydrometeorologischen Situation, eine Korrektur der Abflussvorhersagen durch ein Fehlermodell auf der Basis der zeitlichen Korrelationen. Durch das Fehlermodell können die Unsicherheiten während der ersten Stunden der

Prognosefrist deutlich reduziert werden. Die Wirkung des Ensemble-Kalman-Filters ist zwar geringer, erstreckt sich allerdings über die gesamte Vorhersagefrist.

Das Hauptaugenmerk im dritten Teil der Arbeit liegt auf der Beurteilung des Potentials der satellitenbasierten Messung von Bodenfeuchtemustern als zusätzliche Informationsquelle bei der Identifikation einer realistischen Modellstruktur und geeigneter Parameter. Der Problematik von geringen Eindringtiefen bei der satellitenbasierten Bodenfeuchtemessung, wird durch wird die Erweiterung des bestehenden hydrologischen Modells um eine dünne, oberflächennahe Bodenschichte Rechnung getragen. Der Vergleich von modellierter Bodenfeuchte und Bodenfeuchte aus der Fernerkundung zeigt eine gute Übereinstimmung beider Methoden hinsichtlich der räumlichen und zeitlichen Bodenfeuchtedynamik. Weiters zeigen die Auswertungen, dass die Maskierung der Satellitenbodenfeuchte für Zeitpunkte mit Schneebedeckung oder gefrorenem Boden zu einer deutlichen Verbesserung der Übereinstimmung beider Methoden während der Wintermonate führt.

Generell zeigen die Ergebnisse dieser Arbeit, dass die zusätzliche Verwendung von aktuellen Messdaten, in dieser Arbeit Abflussmessungen und Bodenfeuchtemuster aus der Fernerkundung, einen wertvollen Beitrag zur Reduktion von Unsicherheiten bei der hydrologischen Modellierung leistet. Damit werden operationelle Hochwasservorhersagen auf eine solide methodische Basis gestellt, und die Anwendbarkeit für außerordentliche hydrometeorologische Situation erhöht.

Abstract

Flood forecasting is becoming increasingly important for small catchments where the forecast uncertainties tend to be larger than in large catchments. In addition, also the increase of the forecast lead time is associated with larger uncertainties. These issues make the development of flood forecasting systems an important interdisciplinary task in hydrology, meteorology and remote sensing. In this study, the aim is the development and the analysis of methods to describe and reduce uncertainties in measurement and modelling of hydrometeorological processes. The analyses are based on simulations with a distributed hydrological rainfall-runoff-model which describes snow accumulation and melt, the changes in soil moisture and catchment and stream routing functions.

The quantification of the forecast uncertainty is in the focus of the first part of the study. The uncertainty is quantified using a set of equally probalbe forecasts (an ensemble) which are affected by the meteorological and hydrological boundary conditions. The assessment of the ensemble forecasts is based on the analysis of the frequency of false and correct alarms. The results indicate that the ensemble forecasts are a valuable and important source of information for flood forecasting. Even though the ensemble characteristics do not exactly match the forecast errors, they do provide information about the expected forecast errors.

In the second part of the study the hydrological uncertainties are reduced by using online available runoff measurements. To increase forecast accuracy, two real-time updating procedures are used in this study. The first procedure assimilates runoff data to update the catchment soil moisture state based on Ensemble Kalman filtering. The second procedure consists of an additive error model that updates runoff directly. This error model exploits the autocorrelation of the forecast error and involves an exponential decay of the correction. The error model clearly reduces the forecast uncertainties in the first hours of the forecast lead time. The impact of the Ensemble Kalman filter is smaller, but it affects the entire forecast lead time.

Remotly sensed soil moisture data are used in the third part of the study as additional source of information to identify a realistic model structure and parameters. To account for the shallow penetration depth of the remote sensing data the hydrological

model is extended by a skin soil layer which represents only the first centimetres of the landsurface. A comparison of simulated soil moisture and soil moisture derived from remote sensing data shows excellent consistency between the spatial patterns of soil moisture. Analyses indicate that the masking of remote sensing data with information about snow covered areas significantly improves the correlation between the simulated and remotely sensed soil moisture data.

The results of this study show that observed runoff data and remote sensing data are a valuable source of information to reduce uncertainties in hydrological modelling. They allow for a solid methodical basis of operational flood forecasts and guarantee the applicability of the flood forecasting system in extraordinary hydrometeorological situations.

Acknowledgment

The work on this PhD-thesis has been exhausting and trying, particularly in the last couple of weeks. Without the help and support and the encouragement from many people in my private and professional surrounding, I would still be working on it.

First, I want to thank Prof. Günter Blöschl for guiding me as a teacher and friend over the last couple of years. He has provided excellent input and ideas and kept me focused on the important things. His patience and support helped me overcome many crisis situations and finish this dissertation.

I also would like to thank Prof. Wolfgang Wagner for his input and ideas, particularly in the last part of the study.

For numerous constructive suggestions during the work on this PhD-thesis, my thanks also go to Prof. Dieter Gutknecht.

I also would like to thank my current and former colleagues at the Institute for Hydraulic and Water Resources Engineering at the Technical University in Vienna for their encouragement and their friendships and for providing a good working atmosphere. Special thanks for their friendship and support go to Christine and Thomas.

Financial support of the State Government of Lower Austria, the EVN Hydropower company, the EC (HYDRATE project) and the FFG (GMSM project) acknowledged.

For providing data and expert knowledge I would like to acknowledge the Federal Agency for Water Management (BAW-IKT), the Institute of Photogrammetry and Remote Sensing (IPF), the Austrian Hydrographic Service and the Central Institute for Meteorology and Geodynamics (ZAMG).

I would like to thank my parents and my brother for their support and encouragement.

At last but not least my beloved wife, Kati - thank you for supporting me with your love and understanding. I can’t thank you enough.

Contents

1 Introduction 1

2 Ensemble prediction of floods – catchment non-linearity and forecast probabilities 4

Abstract 4

2.1 Introduction 4

2.2 Data and methods 7

Study area and data 7

Hydrological model 10

Generating ensemble forecasts 11

2.3 Results and discussion 14

Model performance and deterministic forecasts 14

Ensemble forecasting and propagation of non linearity 16

Ensemble spread and forecast error 22

Relative Operating Characteristics 26

2.4 Conclusions 29

3 Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting 32

Abstract 32

3.1 Introduction 32

3.2 Data and methods 33

Study catchment and data 33

Hydrologic model 35

Ensemble Kalman Filter 36

Application of the Ensemble Kalman Filter to the Kamp catchment 43

Sensitivity to soil moisture 45
3.3 Results 47
Updating soil moisture in a simulation mode 47
Updating soil moisture in a forecast mode..... 49
Performance for large flood events 53
3.4 Discussion and Conclusions 58
4 A comparison of in-situ, ASCAT and model estimates of soil moisture in Austria 61
Abstract..... 61
4.1 Introduction 61
4.2 Field sites and data 64
4.3 Hydrologic model..... 69
4.4 Results 72
ASCAT Satellite soil moisture vs. model simulations 72
ASCAT Satellite soil moisture vs. in-situ data..... 79
Model simulations vs. in-situ data of soil moisture..... 82
4.5 Discussion and Conclusions 84
5 Summary and conclusions 88
Appendix A 94
Structure of the soil moisture model 94
References..... 96

List of Figures

2.1 Kamp catchment (622 km²) with telemetered rain gauges (circles) and stream gauge (triangle) shown. 7

2.2 Event precipitation and direct runoff depths for the largest events on record. Kamp at Zwettl, 622 km². 8

2.3 Average forecast errors ej of the five flood events of Table 2.2. 15

2.4 Ensemble forecasts on July 9, 2005 at 0h. Kamp at Zwettl, 622 km². 17

2.5 Ensemble forecasts on July 10, 2005 at 0h. Kamp at Zwettl, 622 km². 18

2.6 Ensemble forecasts on July 10, 2005 at 12h. Kamp at Zwettl, 622 km². 19

2.7 Mapping of the precipitation uncertainties to runoff uncertainties for forecast lead times of 24, 36 and 48 hours. 20

2.8 Mapping of the precipitation uncertainties to runoff uncertainties for forecast lead times of 48 hours in terms of the coefficient of variation of the ensemble spread of cumulative precipitation CVP and runoff CVQ. 21

2.9 Comparison of the distribution functions of the average ensemble spread 23

2.10 Range hit rates for lead times of 6, 12, 24 and 48 hours 25

2.11 Relative Operating Characteristic (ROC) curves 28

3.1 Kamp catchment (622 km²) 34

3.2 State variable and estimation variance for the recession from a linear reservoir 38

3.3 Schematics of the Ensemble Kalman Filter approach and the similarity approach 42

3.4 Scenario with low initial soil moisture 46

3.5 Scenario with high initial soil moisture 47

3.6 Simulations without updating and with updating 48

3.7 Simulations without updating and with updating 49

3.8 Effect of updating soil moisture in the forecast mode, 18 July 1997 50

3.9 Effect of updating soil moisture in the forecast mode, 15 August 2005 51

3.10 Spatial distributions of simulated and updated relative soil moisture 52

3.11 Comparison of the forecasted peak flows with and without updated initial conditions 55

3.12 Forecast errors for the six largest flood events 57

3.13 Nash-Sutcliffe model efficiency for the six largest flood events 57
4.1 Location of the study area and the field sites 65
4.2 TDR-Field measurement sites 67
4.3 Soil moisture measured at the Obersiebenbrunn site for the year 2008 68
4.4 Simulations for summer 2009 at the stream gauge Zwettl/Kamp (622 km²) 71
4.5 Spatial patterns of rainfall and soil moisture 72
4.6 Trade-off between accuracy and number of scenes available 76
4.7 Seasonality of simulated skin soil moisture and ASCAT SSM 77
4.8 Monthly coefficients R for the Kamp catchment without masking 78
4.9 Same as Figure 4.8 but with masking scenario 4 78
4.10 Seasonality of TDR measured soil moisture and ASCAT soil moisture 81
4.11 Seasonality of TDR measured soil moisture and simulated soil moisture 83
4.12 Temporal evolution of the simulated soil moisture and TDR measurements 84

List of Tables

2.1 Uncertainties and typical forecast lead times for a 1000 km² catchment. 5

2.2 Characteristics of the events for which flood forecasts are analysed in this paper. 9

3.1 Parameters of the Ensemble Kalman Filter. 45

3.2 Flood peaks, return periods and evaluation periods for the statistical error analysis of the six largest flood events. 54

4.1 Characteristics of the study area and the field sites. 66

4.2 Definition of criteria for the masking ASCAT surface soil moisture estimates due to snow and frozen ground. 73

4.3 Correlation coefficients between monthly mean ASCAT surface soil moisture and simulated skin layer soil moisture. 75

4.4 Bias (%) of monthly mean ASCAT surface soil moisture relative to simulated skin layer soil moisture. 75

4.5 Correlation coefficients between monthly mean ASCAT surface soil moisture and in-situ soil moisture. 80

4.6 Bias (%) of monthly mean ASCAT soil moisture relative to in-situ soil moisture at the three field sites. 80

4.7 Correlation coefficients and biases (%) between monthly mean simulated soil moisture (skin and main layer) and in-situ soil moisture. 82

1 Introduction

Every year severe flood events cause fatalities and enormous economic damage around the globe. The major flood event in 2002 in the Danube and Elbe catchments affected large parts in central Europe. As a consequence, the revision and adaption of flood management strategies has recently been an important issue. The flood framework directive of the European Union (European Union, 2007) engages the member states to develop flood risk management plans by December 2015. These plans include structural flood protection measures, such as retention basins and levee systems, as well as non-structural measures, such as flood warning systems. Flood warning is an important non-structural instrument to enhance preparedness immediately before a flood event (Blöschl, 2008).

Accurate flood forecasts rely on two important requirements. Firstly, the current hydrological situation at the time of the forecast has to be estimated appropriately. Secondly, mathematical description of the flood generation processes is necessary to predict future runoff. Recently, the trend in flood forecasting has been towards runoff predictions for longer lead times. For lead times longer than the response time of the catchment, meteorological forecasts are required.

Due to highly non-linear and variable hydro-meteorological processes the development of reliable flood forecasting systems is very challenging for the scientific community. The imperfect description of physical processes introduces a high amount of uncertainty to the system. Main sources of uncertainty can be attributed to the precipitation forecast (e.g., parameters, structure and initial conditions of the atmospheric model) and the hydrologic simulation (e.g., parameters and structure of the runoff model).

Quantifying the uncertainty of flood forecasts is becoming increasingly important for operational purposes. The uncertainties for longer lead times are dominated by the precipitation forecast uncertainty, due to propagation of small errors in the initial conditions of the atmospheric models (Buizza, 2003). The standard method of estimating this uncertainty is the generation of an ensemble (or set) of different forecasts of atmospheric processes that differ by their initial conditions (Taylor and Buizza, 2003) in

addition to the main (deterministic) forecast. The ensemble spread for different hydro-meteorological situations can be used as an indicator for the expected forecast error.

For shorter lead times erroneous initial conditions dominate the uncertainty of the flood forecasts. As it plays an important role in partitioning rainfall into runoff and infiltration, soil moisture is probably one of the key variables in flood forecasting models. The simulation of appropriate initial conditions, particularly soil moisture, relies on the accurate knowledge of the recent meteorological history.

A recent development in operational flood forecasting is to include data assimilation techniques for the reduction of uncertain initial soil moisture states caused by erroneous model inputs or inadequate parameters. Widely used updating methodologies are based on sequential assimilation of actual discharge measurements into hydrologic models. Specifically, Monte Carlo methods, e.g. Ensemble-Kalman-Filtering, are appealing because of their flexibility, ease of use and operational robustness (Madsen and Skotner, 2005).

However, appropriate model structure and parameters are a pre-requisite for the implementation of updating routines (Reefsgaard, 1997). This puts emphasis on the identification of appropriate model structure and parameters. As hydrological models are usually calibrated against runoff only, the benefit of additional information about internal model states, particularly soil moisture, may help to constrain uncertainties in model structure and parameters. Knowledge of the characteristics of the spatial and temporal dynamics of soil moisture, which are highly variable (Western et al., 2003), is of key importance. An alternative approach to estimate spatial patterns of soil moisture is through the use of satellite data. In a hydrologic context, the main advantage of using spaceborne sensors for soil moisture retrieval is that they provide an integral value over an area and are available at a global scale. The main limitation is the penetration depth, which is limited to the top few centimetres of the surface. Some assumptions hence need to be made on the vertical distribution of soil moisture in the soil profile to retrieve root zone soil moisture from surface soil moisture. Given the spatial scales and data limitations, more parsimonious approaches, like simple multi layer models, are more appealing than the more sophisticated ones. (Georgakakos and Baumer, 1996; Houser et al., 2000). Existing studies indicate that

satellite based surface soil moisture estimates do have the potential to improve simulated soil moisture states (Francois et al., 2003; Heathman et al., 2003)

These issues make the development of flood forecasting systems an important interdisciplinary task in hydrology, meteorology and remote sensing.

This study consists of three main parts. The quantification of the forecast uncertainty is the focus of the first part of the study (Chapter 2). The main aspects of investigation are (a) how the ensemble distribution of precipitation forecasts propagates in the catchment system, and (b) to interpret the flood forecast probabilities relative to the forecast errors. The second part of the study (Chapter 3) deals with data assimilation in an operational flood forecasting system. The benefit of Ensemble Kalman-Filter updating in forecasting large flood events is evaluated. Using actual runoff measurements, the soil moisture state of the catchment is updated which is then used as an initial condition for the forecasts. The analysis in Chapters 2 and 3 are based on a distributed rainfall-runoff model in the Kamp catchment in Austria that is part of an operational flood forecasting system. A comprehensive data set, including large floods, is available from this catchment. Remotely sensed soil moisture data is compared to simulated soil moisture and in-situ measurements in the third part of the study (Chapter 4). The main focus is to evaluate how realistic the spatial and temporal dynamics of soil moisture are estimated by the different methodologies. To account for the shallow penetration depth of the remote sensing data, an additional skin soil layer is introduced to the model structure, as used in Chapters 2 and 3.

2 Ensemble prediction of floods – catchment non-linearity and forecast probabilities

Abstract

Quantifying the uncertainty of flood forecasts by ensemble methods is becoming increasingly important for operational purposes. The aim of this paper is to examine how the ensemble distribution of precipitation forecasts propagates in the catchment system, and to interpret the flood forecast probabilities relative to the forecast errors. We use the 622 km² Kamp catchment in Austria as an example where a comprehensive data set, including a 500 yr and a 1000 yr flood, is available. A spatially-distributed continuous rainfall-runoff model is used along with ensemble and deterministic precipitation forecasts that combine rain gauge data, radar data and the forecast fields of the ALADIN and ECMWF numerical weather prediction models. The analyses indicate that, for long lead times, the variability of the precipitation ensemble is amplified as it propagates through the catchment system as a result of non-linear catchment response. In contrast, for lead times shorter than the catchment lag time (e.g. 12 hours and less), the variability of the precipitation ensemble is decreased as the forecasts are mainly controlled by observed upstream runoff and observed precipitation. Assuming that all ensemble members are equally likely, the statistical analyses for five flood events at the Kamp showed that the ensemble spread of the flood forecasts is always narrower than the distribution of the forecast errors. This is because the ensemble forecasts focus on the uncertainty in forecast precipitation as the dominant source of uncertainty, and other sources of uncertainty are not accounted for. However, a number of analyses, including Relative Operating Characteristic diagrams, indicate that the ensemble spread is a useful indicator to assess potential forecast errors for lead times larger than 12 hours.

2.1 Introduction

Quantifying the uncertainty of flood forecasts is becoming increasingly important for operational purposes. This is due to a number of reasons. First, the awareness of the

value of uncertainty bounds in flood management has increased. Indeed, it is the uncertainty bounds that will assist flood managers in the trade-off between alternative decisions as they provide information on the likelihood of making less than optimal decisions as a result of forecast errors. Second, flood forecasts are increasingly used for small catchments where the forecast uncertainties tend to be larger than in large catchments. Third, there is a tendency for making forecasts over longer lead times which are associated with larger uncertainties. The most accurate forecasts can be achieved by using observed runoff along with routing models but the forecast lead times are limited to the travel times in the streams. For, say, a 1000 km² catchment these are on the order of 2 hours (Figure 2.1). The values in Table 2.1 are based on simulation results and hydrograph analyses for various Austrian catchments. Runoff models that use observed precipitation allow to extend the lead times but at the cost of increased uncertainty. Precipitation forecasts allow to further extend the lead times but the uncertainties are still larger.

Table 2.1 Uncertainties and typical forecast lead times for a 1000 km² catchment.

	forecast lead time	forecast uncertainty
River routing using observed runoff	2 hours	small
Runoff model using observed precipitation	6 hours	medium
Runoff model using precipitation forecasts	48 hours	large

As the magnitude of the precipitation forecast uncertainty can be large it has been the topic of much recent research. Most of the uncertainty in precipitation forecasts stems from the propagation of small errors in the initial conditions of the atmospheric models (Buizza, 2003). The standard method of estimating this uncertainty is hence to generate an ensemble (or set) of different forecasts of atmospheric processes that differ by their initial conditions (Taylor and Buizza, 2003) in addition to the main (deterministic) forecast. Ensemble forecasts have been operationally issued by the US National Center for Environmental Predictions (NCEP) and the European Centre for Medium Range Weather Forecasts (ECMWF) for more than a decade. Each of the realisations (or members) of the ensemble is a possible trajectory of atmospheric processes over the lead time. By examining the distribution of the ensemble one then

gets a statistical measure of the forecast uncertainty. The value of making ensemble forecasts lies in the fact that the forecast error changes with time. For some meteorological situations, the likelihood of heavy precipitation will be nil while for others it may be large even though the deterministic forecast does not predict precipitation. More generally speaking, the more the ensemble spread deviates from its climatological mean, the more additional information is provided by the ensemble (Whitaker and Loughe, 1998). Often, the members of the ensemble are assumed to be equally likely and the ensemble spread is assumed to represent the distribution of the forecast errors. However, the statistical interpretation of the ensemble spread is not straightforward. For example, Schaake et al. (2004) analysed the statistical properties of NCEP ensemble precipitation forecasts from 1997-1999 and compared them with measured precipitation. He found that the ensemble forecasts were biased and that the ensemble spread was much smaller than the spread of the error distribution. He proposed methods for bias removal and adjusting the ensemble spread.

Calculating flood runoff from predicted precipitation will modulate the uncertainty of precipitation in two ways. First, additional sources of uncertainty will come in. These include uncertainties in estimating catchment precipitation and the spatial distribution of precipitation (Siccardi et al., 2005), uncertainties in the soil moisture state of the catchment, as well as uncertainties in the model structure and in the model parameters. Krzystofowicz (2001) presented a formal method of combining the hydrological uncertainties with those of the precipitation forecasts. Second, even when neglecting the hydrological uncertainties, the uncertainty in the flood forecasts will be different from those of precipitation because of the non-linearity of the catchment system. Small inaccuracies can amplify if the system shows strongly non-linear behaviour, for example, if threshold processes are present (Blöschl and Zehe, 2005). However, very little is known on exactly how the uncertainty of precipitation forecasts propagates in the catchment system. Part of the problem is that operational flood management is interested in large floods that tend to exhibit different characteristics from smaller floods, but they are – by definition – rare, so statistical analyses are notoriously limited by small sample sizes.

Given the current issue with ensemble forecasting methods, the aim of this paper is (a) to examine how the ensemble distribution of precipitation forecasts propagates in

the catchment system, and (b) to interpret the flood forecast probabilities relative to the forecast errors. We use the Kamp catchment in Austria as an example where an operational flood forecasting system has recently been implemented and a comprehensive data set, including two large floods, is available.

2.2 Data and methods

Study area and data

The Kamp catchment is located in northern Austria, approximately 120 km north-west of Vienna. At the Zwettl stream gauge the catchment size is 622 km² and elevations range from 500 to 1000 m a.s.l. (Figure 2.1).

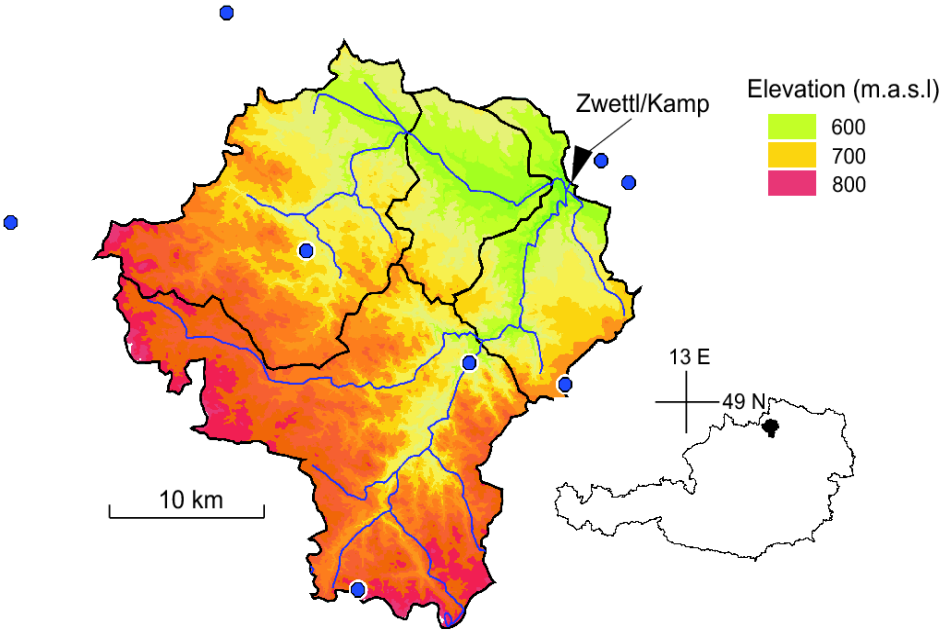

Figure 2.1 Kamp catchment (622 km²) with telemetered rain gauges (circles) and stream gauge (triangle) shown. Black lines represent the subcatchments, blue lines the river network. The figure in the lower left corner gives reference about the catchment location in Austria.

The higher parts of the catchment in the Southwest are hilly with deeply incised channels. Towards the catchment outlet in the Northeast the terrain is flatter and swampy areas exist along the streams. The geology of the catchment is mainly granite and gneiss. Weathering has produced sandy soils with a large storage capacity throughout the catchment. 50 % of the catchment is forested. Mean annual precipitation is about 900 mm of which about 300 mm become runoff (Parajka et. al., 2005a).

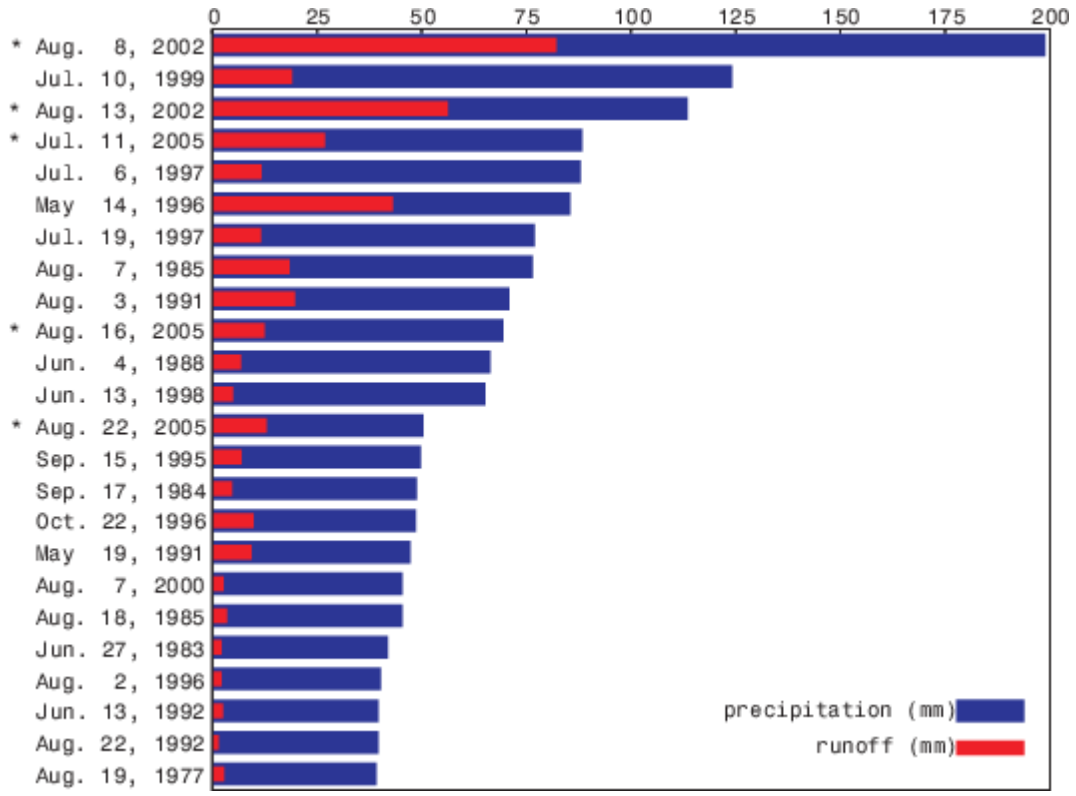

Figure 2.2 Event precipitation and direct runoff depths for the largest events on record. Kamp at Zwettl, 622 km². Events marked by asterisks are analysed in this paper.

To illustrate the nature of hydrologic response of the Kamp catchment the largest flood events on record and the associated rain events have been analysed using the telemetered rain gauges shown in Figure 2.1 and a number of additional rain gauges. Figure 2.2 shows the event precipitation of these events along with the direct runoff depths. The direct runoff depths were estimated by subtracting baseflow from the event hydrographs that was assumed constant during each event. The events have been ranked according to precipitation. There are two interesting findings. First, for the smallest events only around 10% of rainfall become runoff while the percentage can be much higher for the larger events. During the dry summer months large precipitation depths are necessary to exceed the soil capacity and produce any sizeable runoff as was the case for the extreme event of Aug. 8, 2002 (Aug. 2002a). Clearly, runoff generation is a non-linear process and as the magnitude of the event increases so does the proportion of runoff that is generated. Second, for the same precipitation depth, runoff can vary significantly. As a result of prior snow melt, antecedent soil moisture of the May 1996 event was high which produced a large proportion of runoff. On the other hand, the two July 1997 events had almost the same rainfall

as the May 1996 event but much less runoff. There was significant rainfall prior to the Aug. 13, 2002 (Aug. 2002b) event (namely the extreme Aug. 2002a event) which produced more than twice the runoff of the July 1999 event that had similar precipitation but very little antecedent rainfall. It is clear that soil moisture exerts a strong control on runoff response in the Kamp catchment.

Table 2.2 Characteristics of the events for which flood forecasts are analysed in this paper. Kamp at Zwettl, 622 km².

	Aug. 2002a	Aug. 2002b	July 2005	Aug. 2005a	Aug. 2005b
Precipitation (mm)	212	114	88	70	50
Direct runoff depth (mm)	82	56	27	13	13
Runoff coefficient (-)	0.39	0.49	0.30	0.18	0.26
Initial moisture state	Dry	Very wet	Wet	Dry	Wet
Peak discharge (m³/s)	459	367	95	68	65
Return period (yrs)	~ 1000	~ 500	5	3	3
Beginning of event	Aug. 6 0h	Aug. 11 0h	Jul. 5 0h	Aug. 14 0h	Aug. 20 0h
End of event	Aug. 10 21h	Aug. 15 21h	Jul. 15 0h	Aug. 19 21h	Aug. 26 21h
Beginning of rising limb	Aug. 6 12h	Aug. 11 12h	Jul. 10 12h	Aug. 16 0h	Aug. 21 12h
End of rising limb	Aug. 8 6h	Aug. 13 18h	Jul. 11 6h	Aug. 17 21h	Aug. 22 12h
Time to peak (hrs)	36	48	30	18	24

In this paper, the analyses of the ensemble forecasts are based on five flood events for which complete data sets of precipitation forecasts were available. These are marked by asterisks in Figure 2.2. Details of these events are given in Table 2.2. The initial moisture state was assessed by examining antecedent rainfall. Both August 2002 events were indeed extraordinary. More details of these events are given in Gutknecht et al. (2002).

Hydrological model

The model used in this paper is a spatially-distributed continuous rainfall-runoff model (Reszler et al., 2006). The model runs on a 15 minute time step and consists of a snow routine, a soil moisture routine and a flow routing routine. The snow routine represents snow accumulation and melt by the degree-day concept. The soil moisture routine represents runoff generation and changes in the soil moisture state of the catchment and involves three parameters: the maximum soil moisture storage *FC*, a parameter representing the soil moisture state above which evaporation is at its potential rate, termed the limit for potential evaporation *LP*, and a parameter in the non-linear function relating runoff generation to the soil moisture state, termed the non-linearity parameter *ß*. The details of the soil moisture routine are given in Appendix A. Runoff routing on the hillslopes is represented by an upper and two lower soil reservoirs. Excess rainfall enters the upper zone reservoir and leaves this reservoir through three paths, outflow from the reservoir based on a fast storage coefficient K_1; percolation to the lower zone with a percolation rate C_P; and, if a threshold of the storage state LS_{UZ} is exceeded, through an additional outlet based on a very fast storage coefficient K_0. Water leaves the lower zones based on the slow storage coefficients K_2 and K_3. Bypass flow dQ_{by} is accounted for by recharging the lower zone reservoir directly by a fraction of the excess rainfall. K_1 and K_2 as well as C_P have been related to the soil moisture state in a linear way. The outflow from the reservoirs represents the total runoff Q_g on the hillslope scale. These processes are represented on a 1 km x 1 km grid. The model structure and the model parameters have been identified by a five step procedure using field data, comprehensive hydrographic data as well as qualitative evidence during floods (Reszler et al., 2006).

Runoff routing in the stream network is represented by cascades of linear reservoirs with parameters *n* (number of reservoirs) and *k* (storage coefficient) that are a function of runoff. Decreasing travel times with increasing flood levels are represented by linearly decreasing *k* with runoff over a certain range but as the flood water exceeds bank full discharge, *k* is decreased to represent flood attenuation on the flood plains. The parameters have been found by calibration against observed hydrographs and results of hydro-dynamic simulation models. In the context of this study it is important that the model represents the catchment non-linearities well. Comprehensive tests

have shown that this is indeed the case (Blöschl et al., 2006). Of particular value have been the extreme flood events in August 2002 as they allowed to test the model over a wide range of event magnitudes, from small to extreme, along with the smaller events on record (Figure 2.2).

To increase forecast accuracy, two real-time updating procedures are used in the Kamp flood forecasting system. The first procedure assimilates runoff data to update the catchment soil moisture state based on Ensemble Kalman filtering (EnKF) (Evensen, 1994). The strength of the EnKF is that it can accommodate model non-linearity. The model variance represents the errors in the precipitation and evaporation inputs that control the soil moisture state of the catchments and was set to a constant value of 0.005 (mm/15min)² based on sensitivity analyses. The model update is performed for every timestep and the updating is uniform within each gauged catchment. The observation variance represents the discharge measurement errors and is assumed to increase with runoff. The observation variance of runoff was set to $\xi \cdot Q_i^2$ where ξ = 0.0025 was obtained from sensitivity analyses and Q_i denotes the observed runoff at timestep *i*. The soil moisture state of the catchment estimated by the EnKF is used as the initial condition of all forecast runs. The second procedure consists of an additive error model (termed MOS or model output statistics) that updates runoff directly. This error model exploits the autocorrelation of the forecast error and involves an exponential decay of the correction. The autocorrelation lag was found from error analyses of events as 4 hours.

Generating ensemble forecasts

At each time step, precipitation observed at the telemetered rain gauges (Figure 2.1) over the past 15 minutes is interpolated on the 1 km grid using climatologically scaled radar information (Haiden et al., 2007). The climatological scaling is derived from a comparison of monthly totals of the radar and raingauge data at the station locations and varies with location and season. The scaled radar field is linearly combined with the field derived by station interpolation, the weights of this combination depending on the climatological scaling factor. In regions where this factor is large (i.e., the visibility by the radar network is low), most of the weight is with the station interpolation. Where the factor is close to unity, the scaled radar field dominates the final estimate.

The final precipitation analysis reproduces the observed values at the raingauge locations.

Additionally, at each time step, deterministic precipitation forecasts are made at 15 min temporal resolution over a lead time of 48 hrs. The forecasts consist of two components. The first component is an observation-based extrapolation or nowcast of the interpolated precipitation field using motion vectors. They are determined from consecutive analyses by searching for the spatial shift which gives the best match (lowest root mean square difference) of precipitation patterns (Steinheimer and Haiden, 2007). The second component is a weighted mean of the forecast fields of the ALADIN (Wang et al., 2006) and ECMWF numerical weather prediction (NWP) models. The weighting function to estimate the optimised precipitation forecast P_{opt} can be written as

$$P_{opt} = w_{ALA} \cdot P_{ALA} + w_{ECM} \cdot P_{ECM} \tag{2.1}$$

were w_{ALA} and w_{ECM} are the weights for the ALADIN and the ECMWF precipitation forecasts P_{ALA} and P_{ECM}, respectively. The weights w_{ALA} and w_{ECM} have been derived from several years of comparisons of ALADIN and ECMWF forecasts with observed precipitation by minimising

$$e^* = e_M + 0.5 \cdot e_B \tag{2.2}$$

where e_M is the mean absolute error and e_B is the absolute value of the bias. This optimisation has been performed separately for moderate (>5mm/24 hrs) and heavy (>10mm/24 hrs) precipitation events. The sum of the optimized weights w_{ALA} and w_{ECM} can differ from unity to account for biases. The mean error of the combined precipitation forecasts in that period was 20-25% smaller than that of the individual ALADIN and ECMWF forecasts (Haiden et al., 2007) and the biases were negligible.

Another weighting function is used for a smooth transition between the two components (nowcast and NWP forecast) (Golding, 1998). Analyses of the forecast performance indicated that, in most cases, over the first 2-6 hours of the forecast the nowcast had smaller errors than the NWP forecast combination. A weighting function was hence chosen that gives full weight to the nowcast during the first 2 hours, decreases linearly to zero at 6 hours, and remains at zero for larger lead times.

It should be noted that beyond the nowcasting range, the 15 min temporal resolution of the precipitation forecast does not reflect the actual information content of the meteorological models. ALADIN provides output for every hour, and ECMWF provides 6-hourly totals, both of which are linearly interpolated to a uniform 15 min resolution. Similarly, the spatial grid scale of ALADIN (9.6 km) and ECMWF (~25 km) is much larger than the 1 km grid of the hydrological model. Although the scales of the meteorological models and the hydrological model do not match, sensitivity analyses indicated that the first order effect of precipitation uncertainty on runoff is related to (average) catchment precipitation, while the uncertainty resulting from a lack of knowledge of the spatial detail of precipitation is a second order effect.

In order to quantify the uncertainty of the precipitation forecasts, ensemble forecasts are generated. They are constructed, in a similar way as the deterministic forecasts. The ECMWF model provides, at each run, a set of 50 ensemble forecasts in addition to the main (deterministic) run. The ALADIN model currently does not produce ensemble forecasts operationally, so a set of 25 pseudo-ensembles is generated by spatially shifting the ALADIN forecast in both the x and y directions by a random space lag of up to 40 km. This spatial lag has been introduced to account for some of the small-scale uncertainty in the position of the precipitation forecasts. Each of the ECMWF members is then randomly combined with one of the ALADIN pseudo-ensemble members, and with the nowcast. No uncertainty has been assigned to the nowcasts. This means that, up to 2 hours lead time, all ensemble members are identical (zero spread) and the spread increases at longer lead times. We use the pseudo-ensembles of the ALADIN model in the construction of the ensembles because they provide small-scale variance and spread not present in the ECMWF forecasts.

In the case of air temperature, station data are interpolated. The forecasts are based on a combination of the station data with the ALADIN forecasts. No temperature ensembles are computed as their effect on the flood forecasting uncertainty is deemed to be small.

The interpolated precipitation and air temperature fields are used to estimate the state variables of the runoff model such as soil moisture, reservoir storages and snow water equivalent at each time step allowing for EnKF updating. These state

variables are used as the initial conditions for the flood forecasts. All members of the ensemble forecasts use the same initial conditions as the deterministic forecast without any perturbation.

The deterministic forecast fields (both precipitation and temperature) are used as an input to the runoff model to compute deterministic flood forecasts. The 50 members of the ensemble forecasts of precipitation along with the deterministic temperature forecasts are used to compute the ensemble flood forecasts, i.e., 50 realisations of runoff over a lead time of 48 hrs. These are analysed in this paper. Each member of the ensemble forecasts is updated by the additive error model in the same way which means that the error model does not affect the ensemble spread.

2.3 Results and discussion

Model performance and deterministic forecasts

To get an appreciation of the performance of the components of the flood forecasting system, the forecast errors e_j were examined for the five flood events of Table 2.2:

$$e_j = \frac{1}{i_2 - i_1} \sum_{i=i_1}^{i_2} \frac{\left| \hat{Q}_{ij} - Q_i \right|}{Q_i} \tag{2.3}$$

where e_j is the mean absolute normalized error in percent for lead time *j*, $\hat{Q}_{ij}$ is runoff at time step *i* that is forecast with a lead time of *j*, Q_i is the observed runoff at time step *i*, and i_1 and i_2 are the beginning and the end of the analysis interval, respectively. The error analyses were performed separately for the entire flood events (i.e. between the beginning and the end of the flood event as in Table 2.2) and the rising limbs only (as in Table 2.2). The rising limb of a flood hydrograph is the period that is of most interest for the users of real-time flood prediction system. For a given lead time *j*, the errors of the five events were averaged and are shown in Figure 2.3. Four cases were considered.

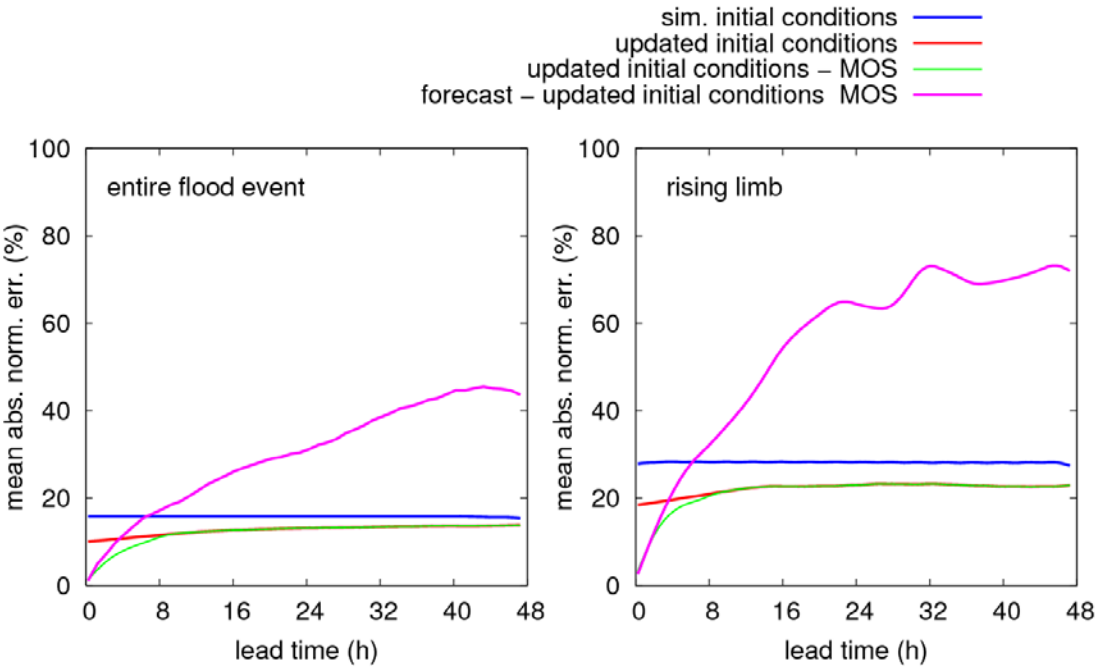

Figure 2.3 Average forecast errors e_j of the five flood events of Table 2.2. Entire flood event (left), rising limb only (right). Kamp at Zwettl, 622 km².

In the first case (blue lines in Figure 2.3) we assumed that future precipitation and temperature were known and used their observed interpolated fields as inputs to the runoff model without any updating. Figure 2.3 indicates that, for this case, the errors do not depend on the forecast lead time. This would be expected as this is the simulation mode. The model errors for the entire flood events are about 15% (Figure 2.3 – left) while they are about 30% if the rising limb alone is analysed (Figure 2.3 – right). The better model accuracy for the entire flood events results from including the time periods with no rainfall, i.e. the recession. In these periods the errors are small as no uncertainties about the amount and the spatial distribution of the input rainfall fields are propagated through the rainfall-runoff model. In contrast, the rising limbs are more difficult to simulate.

In the second case we ran the model in a similar way as in case 1, however, allowed for the EnKF updating to estimate the initial conditions of the forecasts (termed updated initial conditions, red lines in Figure 2.3). Updating the initial conditions reduces the errors for both analysis periods. During the rising limbs the updating reduces the errors from about 30% to about 20%. For the entire events the updating reduces the errors from about 15% to about 12%. There is a slight dependence of the error on the

lead time with smaller errors for short lead times. This dependence is related to the memory of the hydrological system which is taken advantage of by the updating.

The third case was as case 2 but, in addition, allowed for updating by the additive error model (termed updated initial conditions - MOS, green lines in Figure 2.3). The benefit of the additive error model is limited to the first eight hours of the forecasts which is the interval over which the errors are correlated. For larger lead times the additive error model has no effect on the forecasts, so the errors are identical with those of case 2. The fourth case was as case 3 but used forecast precipitation and temperatures rather than the observations (termed forecast - updated initial conditions MOS, purple lines Figure 2.3). The fourth case represents the operational real time configuration, where both updating procedures are used along with the deterministic precipitation and temperature forecasts. In this case, the forecast performance shows a clear dependence on the forecast lead time. In the first 8 hours the forecast, errors are less than 30% (rising limb alone) and about 18% (entire flood events). For lead times of 48 hours the errors are 75% (rising limb alone) and about 50% (entire flood events). This is much larger than the errors of case 3 where observed precipitation has been used as an input. This clearly demonstrates that the main error source for lead times larger than the travel times is the uncertainty in the precipitation forecasts. The difference between the errors of the two analysis periods is particularly large in case 4, as would be expected, as the precipitation forecasts will be most significant in the rising limb where the rainfall occurs.

Ensemble forecasting and propagation of non linearity

Depending on the soil moisture state, a change in precipitation input can be amplified (wet conditions) or dampened (dry conditions) by the catchment system. In the Kamp catchment the occurrence of big floods is associated with wet catchment conditions or very large rainfall depths that wet up the catchment during the event. In the case of the flood events examined here, one would hence expect that the precipitation forecast errors will be amplified as they are propagated through the hydrological model. To illustrate the propagation characteristics ensemble forecasts for the flood event of July 2005 are shown in Figure 2.4, Figure **2.5** and Figure **2.6**. The plotted time window ranges from July 9, 2005 0h to July 12, 2005 0h in all three figures. The thick red lines represent observed runoff, the black lines represent the deterministic

forecasts and the thin blue lines represent the 50 ensemble members. The light blue shading represents the confidence interval between the 10% and 90% quantiles of the ensemble forecasts.

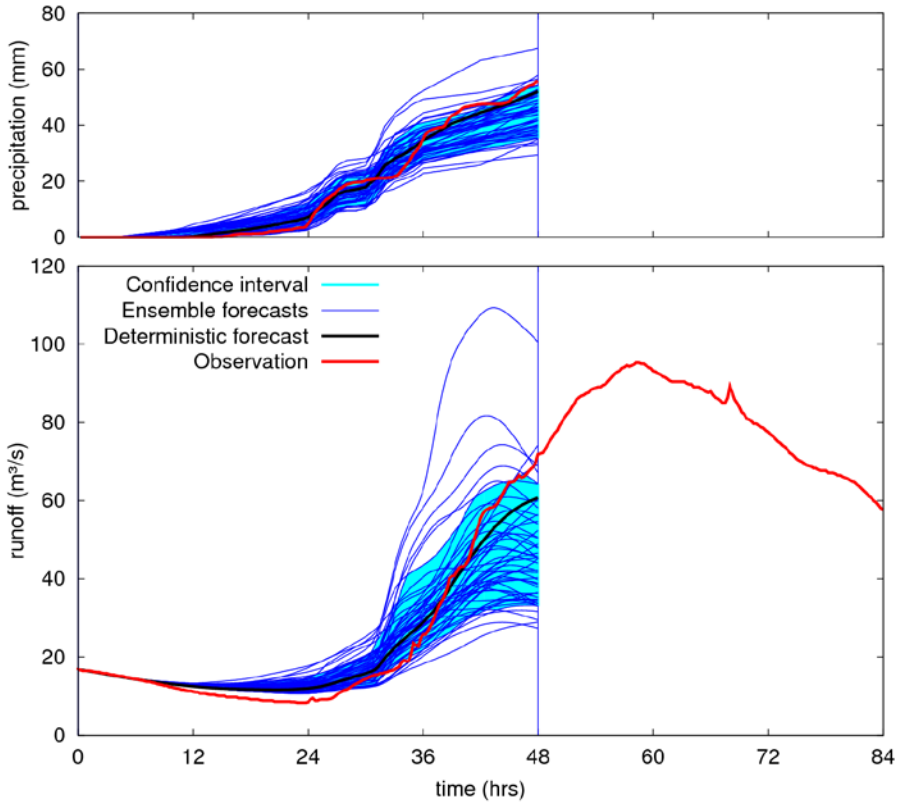

Figure 2.4 Ensemble forecasts (top: cumulative catchment precipitation, bottom: runoff) on July 9, 2005 at 0h (time 0 in the figure). Kamp at Zwettl, 622 km².

Figure 2.4 shows the forecasts on July 9, 2005 0h. In the first 24 hours of the forecast lead time only 5 mm of precipitation have been observed. The deterministic precipitation forecast is very accurate but most of the ensemble members predict more than that. In the second 24 hours of the forecast lead time about 50 mm of precipitation have been observed. Again, the deterministic forecast is accurate. However, most of the ensemble forecasts underestimate precipitation slightly. The deterministic flood forecast matches the observed hydrograph closely. During the first 12 hours of the forecast lead time, the ensemble members are very similar to each other. This is because the forecasts are controlled by observed upstream runoff and observed precipitation through the routing and runoff model components, respectively. In both model components no uncertainty was introduced, i.e. the same data and parameters were used for all members of the ensemble. For lead times of 30 hours and

more, some of the ensemble members indicate a sudden increase in discharge with a maximum flood peak of 100 m³/s. As compared to precipitation, the ensemble spread for these lead times is much larger. This is where the uncertainty in forecast precipitation becomes important.

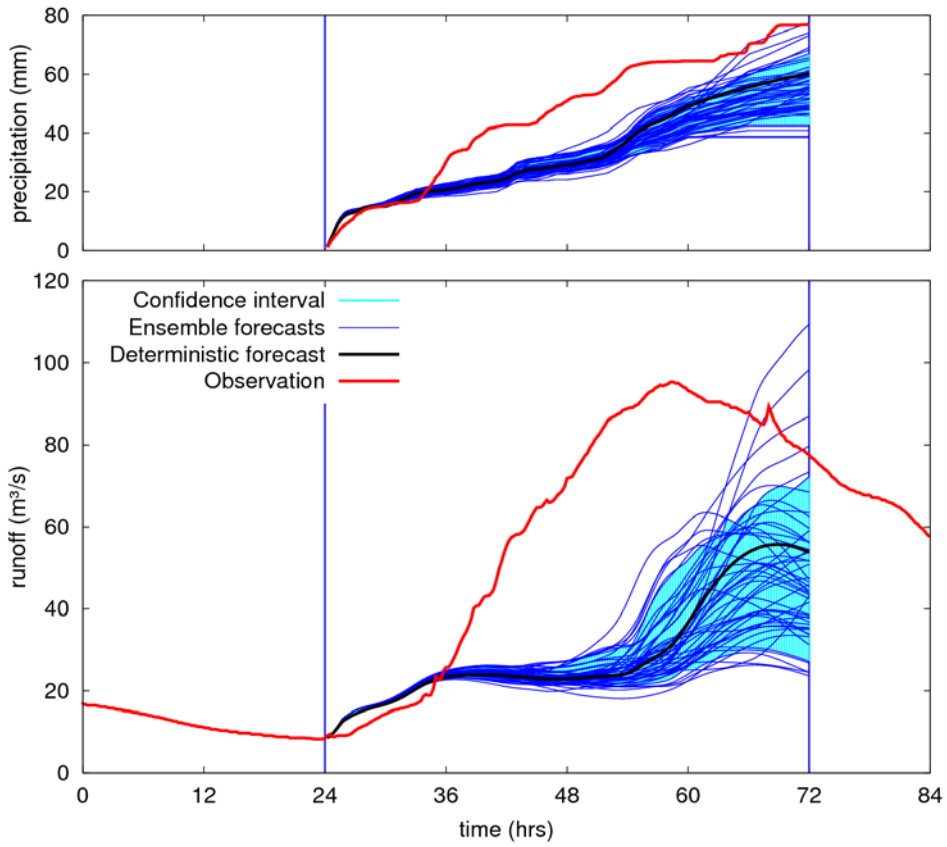

Figure 2.5 Ensemble forecasts (top: cumulative catchment precipitation, bottom: runoff) on July 10, 2005 at 0h (time 24 in the figure). Kamp at Zwettl, 622 km².

The results of the forecast run on July 10, 2005 0h are plotted in Figure 2.5. The total observed precipitation during the forecast lead time is about 75 mm while the deterministic forecast predicts about 60 mm. This relatively moderate underestimation of 15% translates into a larger underestimation of runoff with an estimated peak of 57 m³/s as compared to an observed peak of 95 m³/s, i.e., an underestimation of 40%, and the rising limb is almost completely missed. The main reason is the missing precipitation block at time 36 hours. Most ensemble members underestimate precipitation and the ensemble spread is very small up to a lead time of 24 hours. Similarly, the spread of the runoff forecasts is small during the first 24 hours but in the last 24 hours of the forecasts the spread increases significantly. Clearly, this increase is re-

lated to the non-linearity of the catchment system that translates the somewhat larger spread of precipitation at the end of the forecast period in a much larger spread in runoff. While the observed hydrograph is never within the confidence interval, at the end of the forecast some of the ensemble members do indicate the possibility of a flood on the order of 100 m³/s. Twelve hours later, the forecasts are much more accurate (Figure 2.6). The deterministic precipitation forecast estimates the observed precipitation very well over the entire lead time. The runoff forecasts are very good for the first 12 hrs but do underestimate runoff for larger lead times. In this case, the underestimation is not a result of precipitation errors but is related to the initial conditions of the hydrologic model that are somewhat too dry.

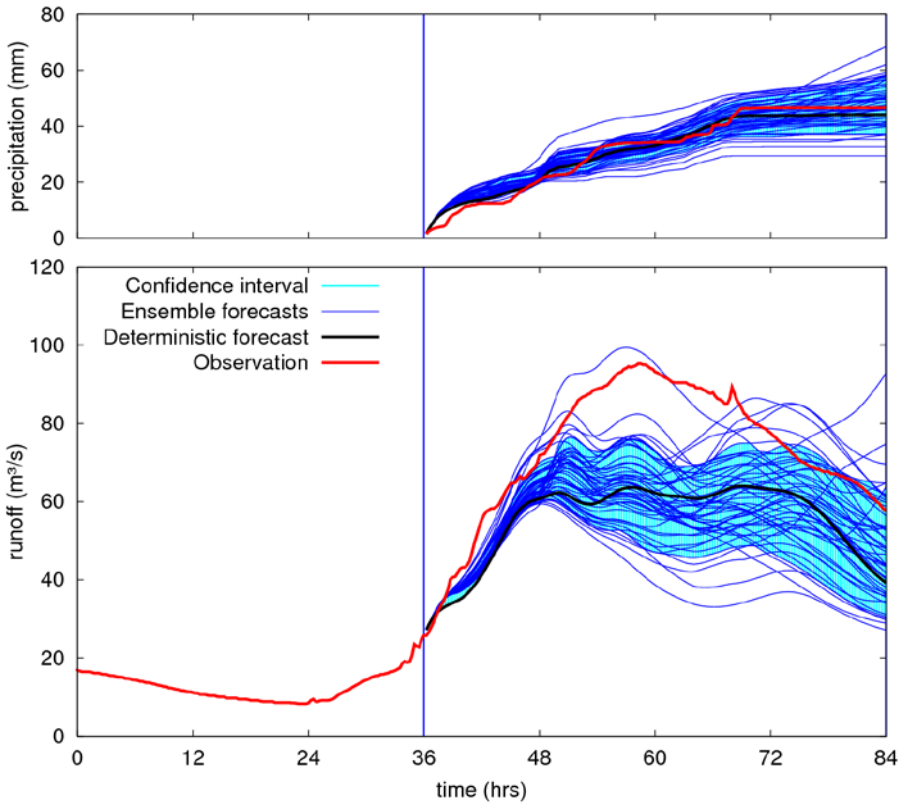

Figure 2.6 Ensemble forecasts (top: cumulative catchment precipitation, bottom: runoff) on July 10, 2005 at 12h (time 36 in the figure). Kamp at Zwettl, 622 km².

To illustrate how the spread of the precipitation ensemble, representing the uncertainties in the precipitation forecasts, is propagated through the hydrologic model we analysed the probability distributions of the precipitation ensembles (model input) and the runoff ensemble (model output). In both cases it was assumed that all ensemble members are equally likely. As an example, Figure 2.7 shows the probability distribu-

tions of the forecast on July 10, 2005 at 0h as in Figure 2.5. For a lead time of 24 hours the distributions of the precipitation and runoff ensemble are similar. They exhibit a narrow spread and are symmetric. For a lead time of 36 hours the two distributions are somewhat different. While the precipitation ensemble spread remains small, the runoff ensemble spread is larger and skewed to the right. This effect is even stronger for a lead time of 48 hours and the largest 20% of the runoff ensemble members have increased their spread dramatically. The total range of the precipitation uncertainty at 48 hours is 40 mm or 70% of the median precipitation while the total range of the runoff uncertainty at 48 hours is 90 m³/s or 200% of the median runoff. Small errors in rainfall may translate into larger errors in runoff. The example of Figure 2.7 has been extended to all the 232 forecasts examined in this paper. For each of these forecasts, the coefficient of variation of the ensemble members of precipitation and runoff has been calculated. The results for a lead time of 48 hours are shown in Figure 2.8, stratified by the mean cumulative precipitation.

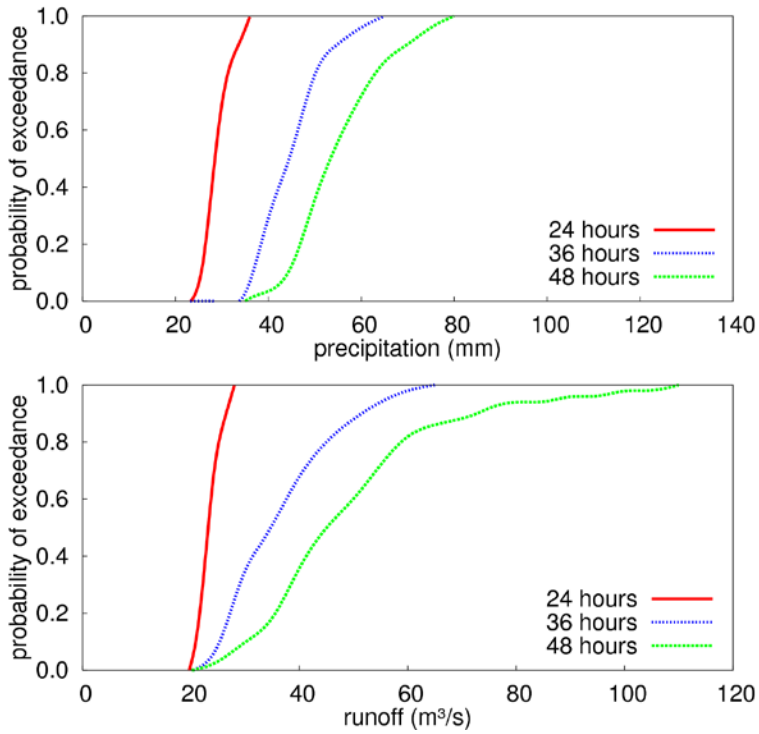

Figure 2.7 Mapping of the precipitation uncertainties to runoff uncertainties for forecast lead times of 24, 36 and 48 hours. cdf is the cumulative distribution functions assuming all ensemble members are equally likely. Forecast on July 10, 2005 at 0h as in Figure 2.5. Kamp at Zwettl, 622 km².

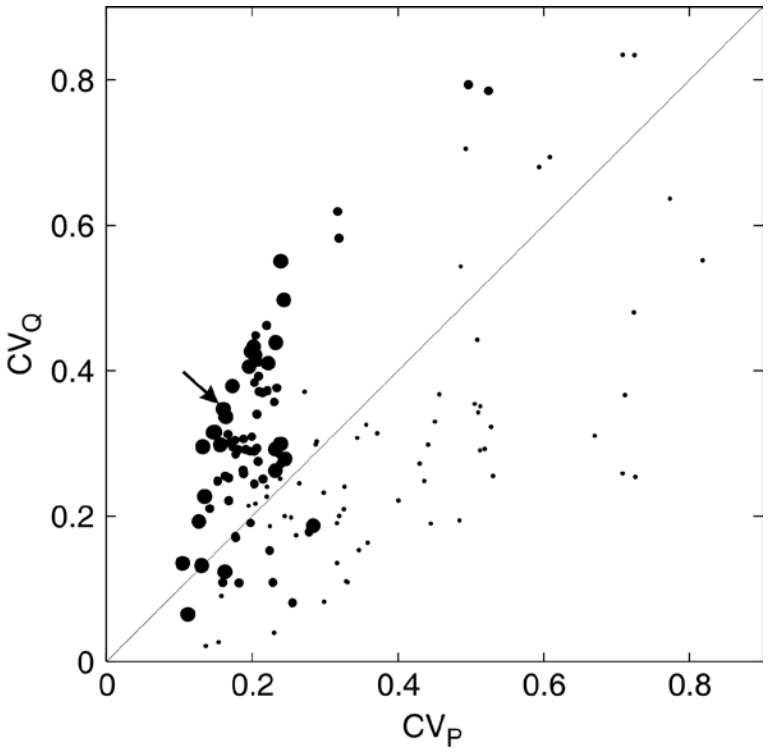

Figure 2.8 Mapping of the precipitation uncertainties to runoff uncertainties for forecast lead times of 48 hours in terms of the coefficient of variation of the ensemble spread of cumulative precipitation CV_P and runoff CV_Q. Five flood events of Table 2.2, Kamp at Zwettl, 622 km². Arrow indicates 48hr forecasts of Figure 2.7. Largest dots relate to forecasts with mean cumulative precipitation $\overline{P}$ > 50mm, and the medium and smallest dots relate to forecasts with 50 > $\overline{P}$ > 30mm and 30 > $\overline{P}$ > 10mm, respectively. Clearly, this kind of mapping of precipitation uncertainties to runoff uncertainties for large forecast lead times is related to the non-linear nature of catchment response. Non-linearity in runoff response has been observed at all space time scales. Often, the non-linearity is more pronounced in dry climates than in wet ones (Chiew et al., 2006). With 300 mm of annual runoff the Kamp is a rather dry catchment in an Austrian context.

The largest dots in Figure 2.8 relate to forecasts with mean cumulative precipitation $\overline{P}$ > 50mm, and the medium and smallest dots relate to forecasts with 50 > $\overline{P}$ > 30mm and 30 > $\overline{P}$ > 10mm, respectively. For small precipitation depths, the uncertainty in precipitation may or may not matter for runoff. Indeed, if precipitation is very small, runoff will be controlled by groundwater response, so any uncertainty in precipitation will not appear in the runoff forecasts. In contrast, for the instances when the forecasted precipitation was large (largest dots in Figure 2.8), the coefficients of variations may more than double when moving from precipitation to runoff. In the ex-

ample of Figure 2.7, the coefficient of variation increases from 0.16 to 0.34 when moving from precipitation to runoff (arrow in Figure 2.8).

Ensemble spread and forecast error

Ideally, the ensemble spread should be an estimator of the distribution of the forecast errors. However, in the present study not all sources of uncertainty have been represented in the ensembles. Rain gauge measurement errors, small scale precipitation variability between the raingauges, uncertainty in the routing and runoff models as well as uncertainties in initial soil moisture have not been represented in the ensembles. Also, it is unclear whether the ECMWF ensemble forecasts and the ALADIN pseudo-ensembles are equally probable forecasts in the study region. The obvious method of examining to what degree the ensemble spread actually captures the distribution of the forecast errors is a comparison of the two, based on an analysis of observed flood events. However, the forecast error distribution changes with time. Typically, the forecast errors are large during the rising limbs of floods and small during the recession and low flow periods. Most importantly, one is interested in the forecast errors of large flood events but large events are always rare, so statistical analyses are limited by small sample sizes. As a simplification we assumed here that the forecast errors of all 232 time steps of the five flood events of Table 2.2 can be combined into a single distribution function for each forecast lead time. It should be noted, however, that not all of the 232 forecasts are completely independent from each other. We calculated the forecast errors as the difference between the deterministic forecast and the observed runoff (positive error for overestimation) from which we derived the distribution function. In a similar vein, we calculated the deviations between the ensemble forecasts and the deterministic forecast (positive deviation if ensemble forecast is larger than the deterministic forecast), and calculated the distribution function for the same time steps as in the case of the errors, assuming that all ensemble members are equally probable.

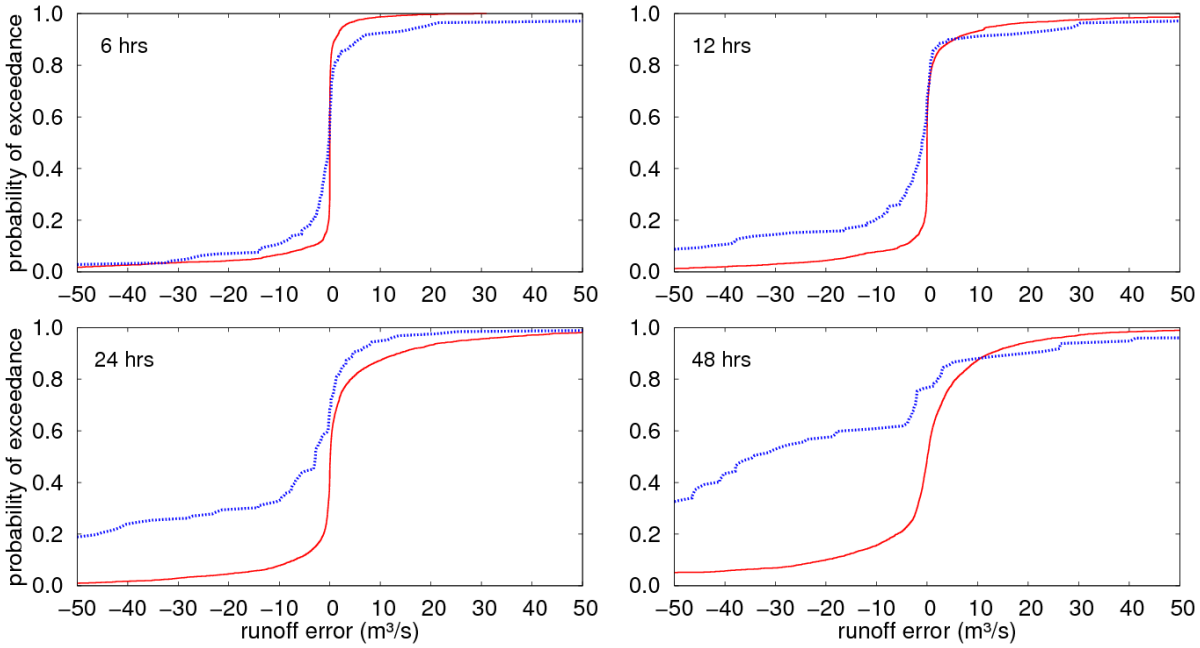

Figure 2.9 Comparison of the distribution functions of the average ensemble spread around the deterministic forecast (solid red lines) and forecast errors of the deterministic forecasts (dotted blue lines) for lead times of 6, 12, 24 and 48 hours for the five flood events of Table 2.2. Kamp at Zwettl, 622 km².

Figure 2.9 shows the results of this comparison. The solid red lines represent the average ensemble spread around the deterministic forecasts and the dotted blue lines show the error distributions of the deterministic forecasts. For a forecast lead time of 6 hours, about 90% of the ensemble members hardly differ from the deterministic discharge forecast (upper left panel). The remaining members show a slight tendency of being smaller than the deterministic forecast. The forecast errors for this lead time exhibit a slightly wider distribution than that of the ensemble members. For a forecast lead time of 12 hours, the spread of the ensemble members increases as would be expected but the forecast errors increase even more. In particular, there are a number of time steps where runoff was significantly underestimated (i.e. negative errors). This tendency continues as one moves to 24 and 48 hour lead times. For 48 hours, the ensemble spread is larger than that of the other lead times and so are the forecast errors. In about half the time steps, the deterministic forecasts underestimate runoff by more than 30 m³/s while less than 10% of the ensemble forecasts indicate deviations of less than -30 m³/s.

The median forecast errors of the 6, 12 and 24 hour lead times are close to zero but the large negative errors are more frequent than the large positive errors, i.e., there exists a negative skew. This is even more the case for a lead time of 48 hours. This means that the deterministic forecasts underestimate runoff more often than they overestimate runoff in the case of the five flood events. This effect can be potentially related to the tendency of the deterministic precipitation forecast to underestimate extremely high amounts of precipitation during the five flood events. Although a bias correction is used in preparing the forecasts it is based on the analysis of moderate (>5mm/24hrs, >10mm/24hrs) precipitation events. What is of most interest in a flood forecasting context are the very large precipitation events but such extreme events are rare, so sample size is very small. It is likely that the forecast errors and biases of the extreme events will differ from those of the moderate events as one would assume that the error characteristics are heteroscedastic. However, accounting for such biases in practice is very difficult.

The ensemble forecasts are almost symmetric although the 48 hr lead times do indicate a slight negative skew. Also, the ensemble spread is always narrower than the distribution of the forecast errors. This would be expected as not all error sources have been represented in the ensembles. However, the ensemble spread increases with lead time in a similar way as the forecast errors. This means that the ensemble spread does provide an indicator to assess potential forecast errors over a range of lead times. Also, one would expect that the most significant changes in the forecast errors as a function of time are captured in the ensembles as they are related to precipitation.

As another possibility of assessing the ability of the ensemble flood forecasts to capture the forecast errors we analysed what we term "range hit rates". A range hit is counted when the observed discharge value is within the range of a certain number of discharge ensemble members. How many of the 50 ensemble members are used to define the upper and lower range is described by the quantile. For the entire forecast ensemble the quantile is 100%. A quantile of 60% means that the highest 20% and the lowest 20% of the ensemble forecast values are not taken into account, i.e., a range hit is counted if the observed runoff is within the range covered by the remaining 60% of the ensemble. A quantile of 0% relates here to the deterministic fore-

cast alone, i.e., a range hit is counted if the observed runoff is identical with the deterministic forecast within the numerical accuracy of 2 digits used here. With this definition, the range hit rate indicates in how many cases, relative to the total number of forecasts, the observed discharge value lies within the range of the ensemble quantiles. The range hit rates were calculated for the same forecasts as used for Figure 2.9.

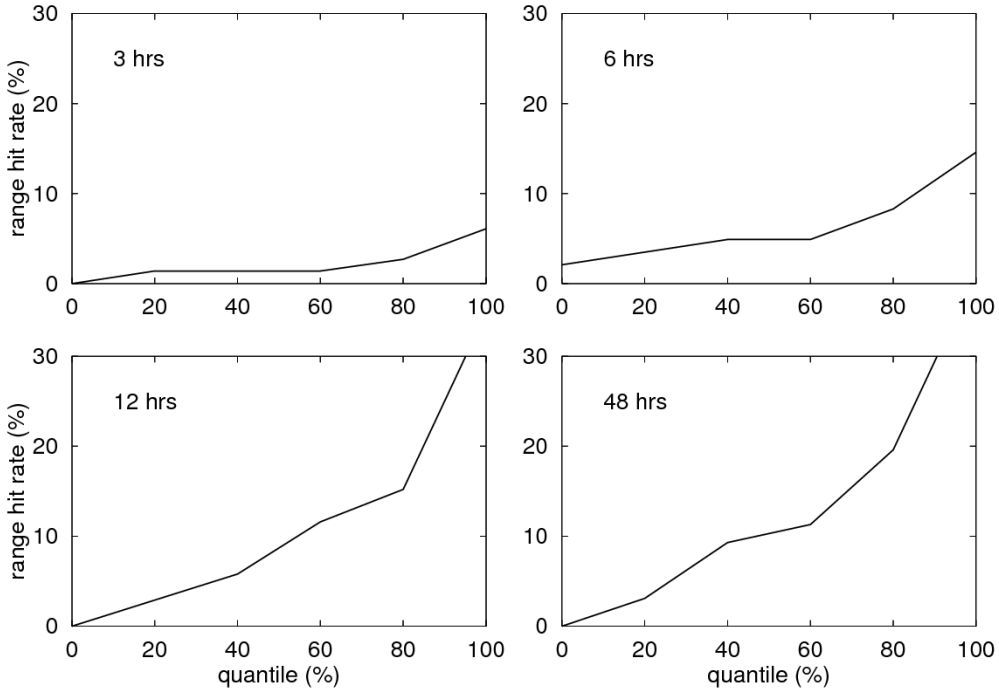

Figure 2.10 Range hit rates for lead times of 6, 12, 24 and 48 hours for the five flood events of Table 2.2. Kamp at Zwettl, 622 km². The range hit rate indicates in how many cases, relative to the total number of forecasts, the observed discharge value lies within the rnage of the ensemble quantiles.

Figure 2.10 shows the results of this analysis for different forecast lead times. For all lead times, the range hit rate increases with the quantiles. Clearly, the wider the uncertainty range the easier it is to capture the observed runoff. For quantiles larger than 60% the range hit rate increases more strongly which is related to the influence of the hydrologic non-linearity combined with the growing deviation from the ensemble mean for the peripheral ensemble members. For a lead time of 3 hours, the range hit rates are always very small. This is because most of the forecast error is due to the routing model and the discharge measurements and both error sources have not been included in the ensemble forecasts. However, as the lead time increases, the

precipitation forecast error becomes more important and the range hit rates increase. The range hit rates are still much smaller than the quantiles. In fact, if the ensemble forecasts captured all the forecast errors one would expect the range hit rates to lie on the 1:1 line. Clearly, this is not the case as the ensemble forecasts focus on the dominant source of uncertainty, i.e., uncertainty in forecast precipitation. The range hit rates are similar for the 12 hour and the 48 hour lead times (as well as for 24 and 36 hours not shown here) which suggests, again, that the ensemble spread does provide an indicator to assess potential forecast errors over a range of lead times, provided the lead times are 12 hours or larger.

Relative Operating Characteristics

Flood management decisions are often based on discharge thresholds, i.e., if a threshold will be exceeded some kind of alarm is triggered. Depending on the context, the warning may result in an alert as is usually the case in early warning, or it may result in flood mitigation action for shorter lead times. When ensemble forecasts are available, any of the ensemble quantiles could be used to trigger an alarm. There is a tradeoff between the ensemble quantile that is used and the usefulness of the alarm. Ensemble members at the upper end (100% quantile) will more likely trigger an alarm but there will also be more false alarms. The opposite is true of the ensemble members at the lower end (0% quantiles). In meteorology, the method of ROC (Relative Operating Characteristic) - diagrams based on threshold analysis are a common method for assessing this tradeoff and hence the performance of probabilistic forecasting systems (Mason and Graham, 1999, Buizza et al., 1999). In analogy, ROC-diagrams are used here to illustrate the alert characteristic of the ensemble flood forecasts for predefined discharge thresholds.

As a first step, hit rates and false alarm rates are defined. The hit rate *HR* is the ratio of the number of correct alarms *H* and the total number of observed events defined by the sum of correct alarms *H* and missed alarms *M*, i.e.

$$HR = \frac{H}{H + M} \tag{2.4}$$

where a correct alarm is counted if both the observed and forecast hydrograph exceed the threshold within the forecast lead time, and an observed event is counted if

the observed hydrograph exceeds the threshold within the forecast lead time. In analogy, the false alarm rate *FAR* is the ratio of the number of false alarms *F* and the total number of no-events defined by the sum of false alarms *F* and correct rejections *C* (neither the observed hydrograph nor the forecast exceeds the threshold), i.e.

$$FAR = \frac{F}{F + C} \tag{2.5}$$

where a false alarm is counted if the forecast hydrograph exceeds the threshold within the forecast lead time but the observed hydrograph does not, and a no-event is counted if the observed hydrograph does not exceed the threshold within the forecast lead time. *H+M+F+C* is 232 in this paper as this is the total number of time steps for which forecasts have been analysed. The hit rate and the false alarm rate of the deterministic forecasts can be plotted as a single point on a hit/false alarm rate graph. The same procedure is then repeated for each quantile of the ensemble forecasts separately which gives a set of points in the hit/false alarm rate graph known as relative operating characteristic (ROC). A perfect forecasting system gives a hit rate of 100% and a false alarm rate of 0%, i.e. the point plots in the top left corner of the ROC diagram. Systems with no skill result in a ROC curve on the 1:1 line.

The relative operation characteristics were calculated for the same forecasts as used for Figure 2.9 with a forecast lead time of 48 hours and are shown in Figure 2.11. Four thresholds were selected. The discharge thresholds of 50 and 100 m³/s are relevant values for flood warning at the Kamp (see Table 2.2), the smaller thresholds of 10 and 30 m³/s were examined for illustrative purposes. The dots represent the ensemble forecasts at intervals of 5%, the crosses represent the deterministic forecasts. A hit rate of 100% is reached for a threshold of 10 m³/s and the 100% quantile of the ensemble forecast (i.e. the largest of the ensemble members). This perfect hit rate is associated with a false alarm rate of more than 80%. With decreasing ensemble quantiles the false alarm rate decreases to less than 20% while the hit rate is greater than 50% for all ensemble quantiles. For a threshold of 30 m³/s, the hit rate ranges from about 80% (100% quantile) to 15%, while the false alarm rate ranges from 20% to nil. The false alarm rates are even smaller for the 50 and 100 m³/s thresholds with similar hit rates. For the 100 m³/s threshold the forecasts of the 100% quantile produce only 10% false alarms with a hit rate of about 80%.

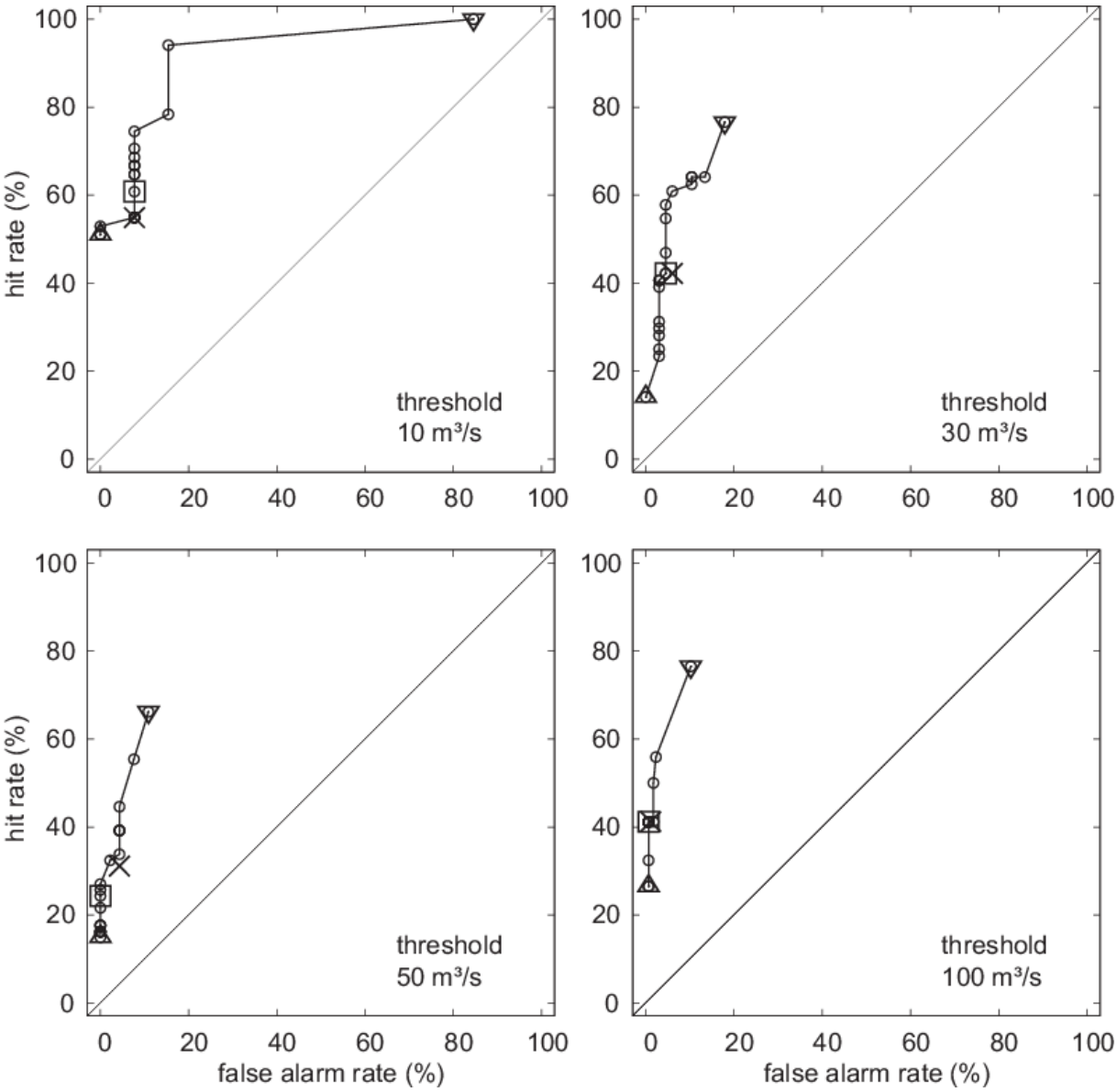

Figure 2.11 Relative Operating Characteristic (ROC) curves for the probabilistic (dots) and deterministic (crosses) 48 hour flood forecasts for discharge thresholds of 10, 30, 50 and 100 m³/s for the five flood events of Table 2. Kamp at Zwettl, 622 km². 5% quantile (upward pointing triangle), 50% quantile (square) and 100% quantile (downward pointing triangle).

For all discharge thresholds, the hit rates of the 100% quantiles of the ensemble forecasts are larger than those of the deterministic forecasts as would be expected. This is the main reason of using ensemble forecasts in flood management. In general, the ROC curves show that the deterministic and probabilistic forecasts at the Kamp produce only few false alarms and the percentage of false alarms decreases with the magnitude of the discharge threshold. The hit rate is limited to about 80% for the 50 and 100 m³/s thresholds. The tendency of very small false alarm rates and

maximum hit rates below 100% in the ROC curves indicate that the flood forecasts tend to underestimate the observed discharges, particularly for the 50 and 100 m³/s thresholds. This tendency is consistent with results of meteorological analyses, which have shown that meteorological model forecasts tend to overestimate small precipitation values and underestimate large precipitation values (Buizza, 1999). The analysis in this paper is based on five big flood events with heavy precipitation, so some underestimation of precipitation would be expected. The tendency towards underestimating precipitation during the floods is amplified through the non-linearity of the hydrologic response at the Kamp. Therefore, the flood quantile of choice for flood alarm purposes would be a high ensemble quantile, for example the 90% quantile. There is another argument for using a large quantile which are the relative costs of false alarms and missed alarms. If false alarms are inexpensive it may pay to choose higher flood quantiles as if false alarms were as expensive as missed alarms. However, decisions on alarms are often made based on maximising the credibility of the forecasts rather than cost arguments.

The area under the ROC curve is sometimes used as a measure for the forecast skill (Stanski et al., 1989). The area under the curve decreases from 1 for a perfect prediction system to 0.5 for a prediction system with no skill. Fitting a cubic spline to the ROC curves in Figure 2.11 gives areas of 0.90, 0.85, 0.85 and 0.90 for the 10, 30, 50 and 100 m³/s thresholds, respectively. As compared to precipitation forecasts in the literature, this is a favourable skill. For example, Buizza et al. (1999) found skills in the range of 0.70 and 0.83, depending on precipitation thresholds and for a maximum forecast lead time of 3 days.

2.4 Conclusions

The real-time flood forecasting system of the Kamp catchment in Austria has been operational since January 2006. It is used in this paper to examine how the ensemble distribution of precipitation forecasts propagates in the catchment system, and to interpret the flood forecast probabilities relative to the forecast errors. The model was tested on five large flood events including a 500 yr flood and a 1000 yr flood.

The analyses indicated that, for long lead times (e.g. 48 hours), the variability of the precipitation ensemble is amplified as it propagates through the catchment system.

For the example examined, the total range of the precipitation uncertainty is 70% of the median precipitation while the total range of the runoff uncertainty is 200% of the median runoff. Small errors in rainfall may translate into larger errors in runoff. An analysis of the coefficients of variation of the ensemble members of precipitation and runoff suggests that, for small precipitation depths, the uncertainty in precipitation may or may not matter for runoff. In contrast, for the instances when future precipitation is large, the coefficients of variations may more than double when moving from precipitation to runoff. Also, the ensemble distribution of precipitation is symmetric while that of the flood forecasts is skewed to the right. Clearly, this kind of mapping of precipitation uncertainties to runoff uncertainties for large forecast lead times is related to the non-linear nature of catchment response. In contrast, for short lead times (e.g. 12 hours and less), the variability of the precipitation ensemble is decreased as it propagates through the catchment system. This is because the forecasts are mainly controlled by observed upstream runoff and observed precipitation through the routing and runoff model components, as the forecasting system is operated in a real-time mode. The ensemble forecasts focus on the dominant source of uncertainty, i.e., uncertainty in forecast precipitation. For lead times of 12 hours and less the ensemble spread is very narrow as other sources uncertainty such as rain gauge measurement errors, small scale precipitation variability between the raingauges, uncertainty in the routing and runoff models as well as uncertainties in the initial soil moisture have not been represented in the ensembles. More generally speaking, it can be expected that the lead time where the uncertainty of the precipitation forecasts starts to amplify will depend on the catchment response characteristics, such as travel times in the river reaches and runoff concentration. In small and flashy catchments this will be a short lead time while for large catchments it will be longer.

The paper also examined the ability of the probabilistic forecasts to capture the distribution of the flood forecast errors. Assuming that all ensemble members are equally likely, the statistical analyses of the ensemble forecasts for five flood events at the Kamp showed that the ensemble spread is always narrower than the distribution of the forecast errors. This would be expected as not all error sources have been represented in the ensembles. Although two updating procedures based on observed runoff have been used to improve the flood forecasts over the simulation mode, there always remains a certain amount of hydrologic uncertainty in the forecasting system.

It is also likely, that the precipitation ensembles do not fully represent the precipitation forecast errors (Schaake et al., 2004). However, the ensemble spread increases with lead time in a similar way as the forecast errors. This means that the ensemble spread does provide an indicator to assess potential forecast errors over a range of lead times. Also, one would expect that the most significant changes in the forecast errors as a function of time are captured in the ensembles as they are related to precipitation. A "range hit rate" was defined as the number of cases, relative to the total number of forecasts, in which the observed discharge value lies within the range of the ensemble quantiles. Analyses of the range hit rates indicate that they are small for short lead times but increase with lead time. The range hit rates are similar for lead time of 12 hours and more which suggests, again, that the ensemble spread does provide an indicator to assess potential forecast errors over a range of lead times, provided the lead times are 12 hours or larger. Finally, the forecast skill of the 48 hour ensemble forecasts was tested by ROC (Relative Operating Characteristic) diagrams based on threshold analyses. For all discharge thresholds, the hit rates of the 100% quantiles of the ensemble forecasts are larger than those of the deterministic forecasts. This is the main reason of using ensemble forecasts in flood management. For the largest discharge threshold examined here (100 m^3/s) the forecasts of the 100% quantile produce only 10% false alarms with a hit rate of about 80%. The flood quantile of choice for flood alarm purposes would be a high ensemble quantile, for example the 90% quantile.

Even though the ensemble characteristics do not exactly match the forecast errors, they do provide information about the expected forecast errors. The comparisons indicated that, for lead times larger than 12 hours in the case of the 622 km^2 Kamp catchment, the ensemble spread is a useful indicator to the forecast errors. While additional error sources could be included in estimating the flood ensembles it may not be necessary for operational flood forecasting purposes as the uncertainty in forecast precipitation is the dominant source of flood forecast uncertainty for lead times of more than 12 hours in catchments such as the Kamp.

3 Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting

Abstract

The aim of this paper is to examine the benefits of updating model soil moisture in forecasting large floods. The soil moisture of a distributed rainfall runoff model is updated by the Ensemble Kalman Filter based on observed runoff in a real-time mode, and is then used as an initial condition for the flood forecasts. The case study is set in the 622 km² Kamp catchment, Austria. The results indicate that the Ensemble Kalman Filter indeed improves the forecasts substantially. The mean absolute normalised error of the peak flows of six large floods decreases from 25 to 12% (3 hour lead time), and from 25 to 19% (48 hour lead time). The Nash-Sutcliffe efficiency of forecasting runoff for these flood events increases from 0.79 to 0.92 (3 hour lead time), and from 0.79 to 0.88 (48 hour lead time). The flood forecasting system has been in operational use since early 2006.

3.1 Introduction

Updating methods in real-time flood forecasting have enjoyed wide popularity in the late 1970s and early 1980s with the increasing use of telemetry in the control of water resource systems (Wood, 1980). While numerous national flood forecasting systems have indeed implemented updating procedures (e.g. Gutknecht, 1991), scientific interest soon ebbed off. The reasons may well be as O'Connell and Clarke (1981, pp. 202-203) noted: "*The above discussion suggests that there are still considerable unsolved estimation problems in real-time forecasting, but it is not clear to what extent their solution would result in improved forecasts. It may be more beneficial to seek a better representation of the spatial variation in rainfall and its effect on streamflow response, and in improving the structure of real-time forecasting models than to expend effort in solving estimation problems. Information on where efforts will be best rewarded can only be obtained by feedback from case studies*." Indeed, distributed modelling and use of radar rainfall have been key topics in hydrologic re-

search in the 1990s (e.g. Grayson and Blöschl, 2000). In the mean time, updating methods have been developed along a separate avenue where the interest resided in how to best use soil moisture satellite data in hydrological models to improve climate forecasts (McLaughlin, 1995). In this context, updating is usually referred to as data assimilation. Methods have been gleaned from oceanography and atmospheric sciences (Reichle et al., 2002) rather than from control theory as had been the case in the earlier flood forecasting research. The availability of new methods has kindled renewed interest in the updating problem of flood forecasting. Specifically, Monte Carlo methods are appealing because of their flexibility, ease of use and operational robustness (Madsen and Skotner, 2005). The Ensemble Kalman Filter (Evensen, 1994) extends the traditional Kalman Filter (Kalman, 1960) concept by Monte Carlo techniques and is able to deal with non-linear model dynamics in a natural way without linearised model equations. Moradkhani et al. (2005) found that the Ensemble Kalman Filter improved runoff forecasts of a conceptual hydrologic model when using on-line measured runoff. Weerts and Serafy (2006) compared the performance of three methods of updating a conceptual runoff model - Ensemble Kalman Filtering, Particle Filtering and Residual Resampling. They suggested that the Ensemble Kalman Filter technique was the most efficient method in case of a small number of realisations, and was generally more robust than the other methods. The Ensemble Kalman Filter is hence an obvious choice for updating flood forecasts.

The aim of this paper is to examine the benefit of Ensemble Kalman Filter updating in forecasting large floods. Using observed runoff, the soil moisture state of the catchment is updated which is then used as an initial condition for the forecasts. The analysis is based on a distributed rainfall-runoff model in the Kamp catchment in Austria that is part of a flood forecasting system that has been in operational use since early 2006.

3.2 Data and methods

Study catchment and data

The Kamp catchment is located in northern Austria, approximately 120 km north-west of Vienna. At the Zwettl stream gauge the catchment size is 622 km² and elevations range from 500 to 1000 m a.s.l. The higher parts of the catchment in the Southwest

are hilly with deeply incised channels. Towards the catchment outlet in the Northeast the terrain is flatter and swampy areas exist along the streams. Typical flow travel times in the river system range from 2 to 4 hours. The geology of the catchment is mainly granite and gneiss. Weathering has produced sandy soils with a large storage capacity throughout the catchment. A catchment fraction of 50 % is forested. Mean annual precipitation is about 900 mm of which about 300 mm becomes runoff (Parajka et al., 2005c). During flood events, only a small proportion of rainfall contributes to runoff. Typically, the event runoff coefficients are 10% or less (Merz and Blöschl, 2005). As rainfall increases in magnitude, the runoff response characteristics change fundamentally because of the soil moisture changes in the catchment and the runoff coefficients can easily exceed 50%. The catchment is hence highly non-linear in its rainfall-runoff response. Representing catchment soil moisture well is hence of utmost importance for producing accurate flood forecasts.

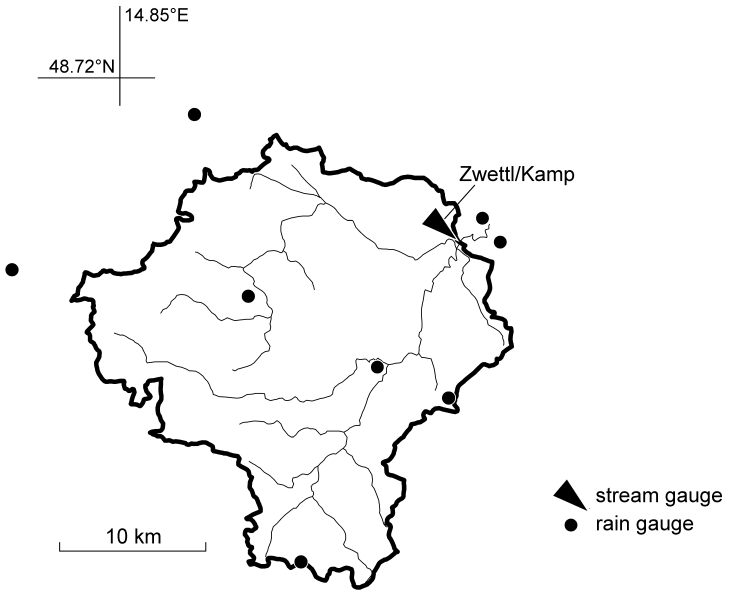

Figure 3.1 Kamp catchment (622 km²) with telemetered rain gauges and stream gauge shown. Thick line represents the catchment boundary, thin lines the river network.

For the development of the distributed model, data from a total of 16 rain gauges were used. Out of these, 10 rain gauges recorded at a time interval of 15 minutes, the others were daily gauges. Eight of the recording rain gauges are telemetered (Figure 3.1) and are used for the operational forecasting. At each time step, the rain gauge data are spatially interpolated to a 1 km grid, supported by climatologically scaled radar information. While the operational system uses rainfall forecasts, all

analyses in this paper are based on the assumption that future rainfall were known from the rain gauge data to focus on the value of the updating procedure in reducing forecasting errors.

Hydrologic model

The model used in this paper is a spatially-distributed continuous rainfall-runoff model (Reszler et al., 2006 and Blöschl et al., 2008). The model runs on a 15 minute time step and consists of a snow routine, a soil moisture routine and a flow routing routine. The snow routine represents snow accumulation and melt by the degree-day concept. The soil moisture routine represents runoff generation and changes in the soil moisture state of the catchment and involves three parameters: the maximum soil moisture storage L_s, a parameter representing the soil moisture state above which evaporation is at its potential rate, termed the limit for potential evaporation L_P, and a parameter in the non-linear function relating runoff generation to the soil moisture state, termed the non-linearity parameter *ß*. The details of the soil moisture routine are given in Appendix A. Runoff routing on the hillslopes is represented by an upper and two lower soil reservoirs. Excess rainfall Q_p enters the upper zone reservoir and leaves this reservoir through three paths, outflow from the reservoir based on a fast storage coefficient k_1; percolation to the lower zones with a percolation rate c_P; and, if a threshold of the storage state L_1 is exceeded, through an additional outlet based on a very fast storage coefficient k_0. Water leaves the lower zones based on the slow storage coefficients k_2 and k_3. Bypass flow Q_{by} is accounted for by recharging the lower zone reservoir (k_2) directly by a fraction of the excess rainfall. k_1 and k_2 as well as c_P have been related to the soil moisture state in a linear way. The outflow from the reservoirs represents the total runoff Q_t on the hillslope scale. These processes are represented on a 1 km x 1 km grid. The model states for each grid element are the snow water equivalent, soil moisture S_s of the top soil layer, the storage of the soil reservoirs S_1, S_2, S_3 associated with the storage coefficients k_1, k_2, k_3, with $k_1 < k_2 < k_3$. The model parameters for each grid element were identified based on the 'dominant processes concept' of Grayson and Blöschl (2000) which suggests that, at different locations and different points in time, a small number of processes will dominate over the rest. Land use, soil type, landscape morphology (e.g. the degree

of incision of streams) and information on soil moisture and water logging based on field surveys were used. Discussions with locals provided information on flow pathways during past floods. Runoff simulations, stratified by time scale and hydrological situations, were then compared with runoff data, and the simulated subsurface dynamics were compared with piezometric head data. The various pieces of information were finally combined in an iterative way to construct a coherent picture of the functioning of the catchment system, on the basis of which plausible parameters for each grid element were chosen. The model was extensively tested against independent runoff data both at the seasonal and event scales. Data from 1993-2003 were used for model identification and parameter calibration. Data from 2004-2006 were used for model verification.

Runoff routing in the stream network is represented by cascades of linear reservoirs with parameters *n* (number of reservoirs) and *k* (storage coefficient) that are a function of runoff. Decreasing travel times with increasing flood levels are represented by linearly decreasing *k* with runoff over a certain range but as the flood water exceeds bank full runoff, *k* is decreased to represent flood attenuation on the flood plains. The model parameters for each reach have been found by calibration against observed hydrographs and results of hydro-dynamic simulation models. The effect of stream routing on the runoff hydrograph is relatively small as compared to runoff generation within the catchment, so most of the effort was devoted to obtaining a realistic representation of catchment processes. All model equations have been implemented in state-space notation to facilitate use of the Ensemble Kalman Filter.

Ensemble Kalman Filter

The idea of the Kalman Filter is to provide an estimate of a state vector based on model information and measurement information, balancing out the errors of the two. It is a sequential algorithm for minimising the state error variance. The Kalman filter is optimal for linear systems. In this paper, runoff has been chosen as the state vector. For consistency with the usual notation (e.g. Madsen et al., 2003) it is denoted by *x* here. The measurement error is attributed to the error in runoff measurements, the model error to the error in precipitation and evaporation input. In the Ensemble Kalman Filter, the model $\Phi(.)$ is now applied to each of the *M* members of the ensemble to estimate the runoff:

$$x_{m,i}^{f} = \Phi(x_{m,i-1}^{a}, u_i + \varepsilon_{m,i}), \qquad m = 1,2,...,M \tag{3.6}$$

where $x_{m,i}$ is the runoff of ensemble member *m* at time step *i*, $x_{m,i-1}$ is the runoff at the previous time step, superscript *f* stands for forecast, superscript *a* stands for analysed, u_i is the model input (precipitation, evaporation) and $\varepsilon_{m,i}$ is the model error which is randomly drawn from a normal distribution with zero mean and model error covariance $\mathbf{V}_i$. As an *a priori* forecast, the mean value of the ensemble forecasts is adopted:

$$x_i^f = \overline{x_i^f} = \frac{1}{M}\sum_{m=1}^{M} x_{m,i}^f \tag{3.7}$$

The error covariance matrix $\mathbf{P}_i^f$ of the forecast is estimated from the ensemble forecasts as:

$$\mathbf{P}_i^f = \mathbf{S}_i^f (\mathbf{S}_i^f)^T \tag{3.8}$$

with

$$s_{m,i}^f = \frac{1}{\sqrt{M-1}}(x_{m,i}^f - \overline{x_i^f}) \tag{3.9}$$

where $s_{m,i}^f$ is the *m* - th column of $\mathbf{S}_i^f$. In a next step, the measurements z_i of runoff are contaminated by a measurement error $\eta_{m,i}$ to generate an ensemble of *M* possible measurements:

$$z_{m,i} = z_i + \eta_{m,i}, \qquad m = 1,2,...,M \tag{3.10}$$

where $\eta_{m,i}$ is randomly drawn from a normal distribution with zero mean and covariance $\mathbf{W}_i$. Each ensemble member $x_{m,i}^f$ is then updated according to

$$x_{m,i}^a = x_{m,i}^f + \mathbf{K}_i(z_{m,i} - \mathbf{C}_i x_{m,i}^f) \tag{3.11}$$

where $\mathbf{K}_i$ is the Kalman gain:

$$\mathbf{K}_i = \mathbf{P}_i^f \mathbf{C}_i^T [\mathbf{C}_i \mathbf{P}_i^f \mathbf{C}_i^T + \mathbf{W}_i]^{-1} \tag{3.12}$$

and $\mathbf{C}_i$ is a matrix that relates the measurements and the state vector. Based on the updated ensemble members, the updated a posteriori estimates of the state vector x_i^a and the error covariance matrix $\mathbf{P}_i^a$ are calculated analogously to Equations (3.2) and (3.3).

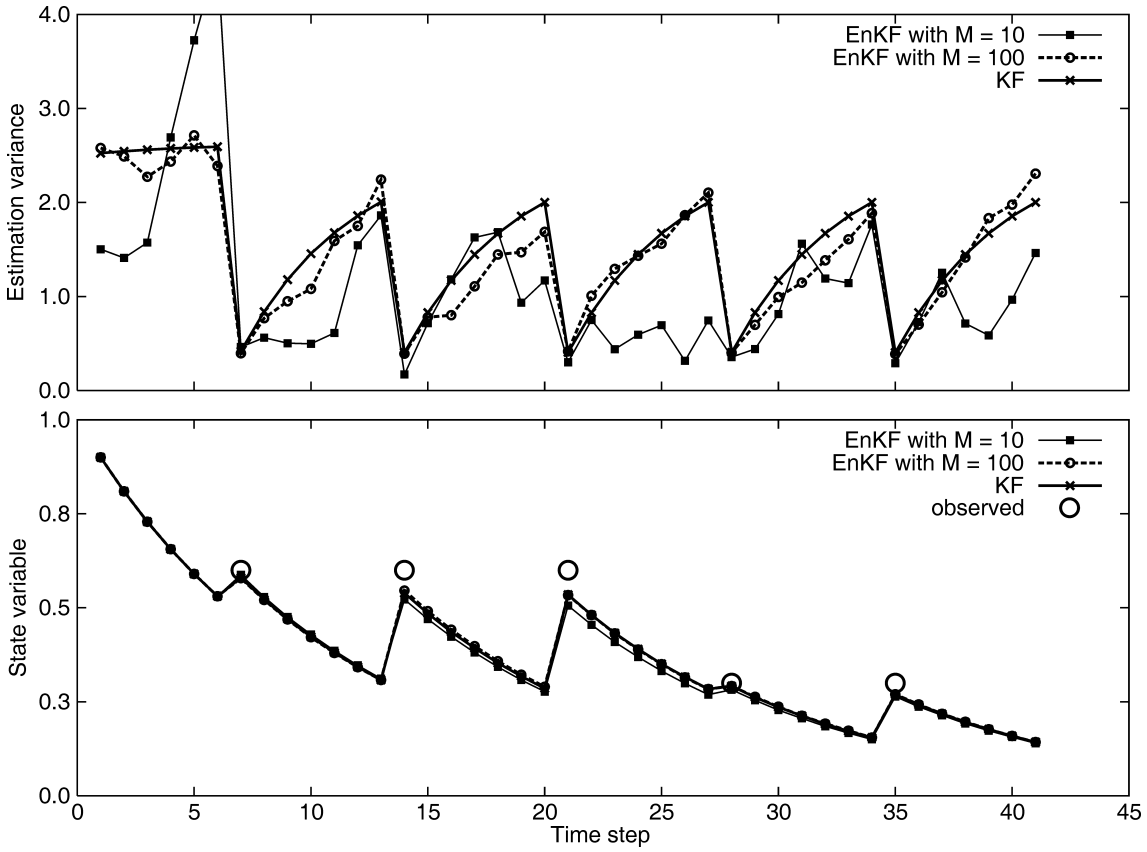

Figure 3.2 State variable (i.e. runoff) and estimation variance for the recession from a linear reservoir estimated by the Kalman Filter (KF) and the Ensemble-Kalman-Filter (EnKF) making use of runoff data at intervals of 7 time steps. M is the ensemble size (i.e. the number of realisations).

To illustrate the dynamics of the Ensemble Kalman Filter for a simple case, Figure 3.2 shows a comparison of updated outflows from a linear reservoir using the original Kalman Filter (KF) and the Ensemble Kalman Filter (EnKF) with ensemble sizes of *M* = 10 and 100. The model equation is $x_i = \kappa \cdot x_{i-1}$ with the recession parameter chosen as $\kappa = 0.9$, the measurement error variance *W* and the model error variance *V* both chosen as 0.5 (m³/s)² and the initial flow chosen as x_0 =1 m³/s. The example can be interpreted as the recession of a flood hydrograph. Again for illustrative purposes, it was assumed that runoff measurements are available every seventh time step. As

the measurements become available, the estimation variance of the Kalman Filter decreases to about 0.4 and increases as the system loses the memory of that information. The degree to which the Ensemble Kalman Filter matches the pattern of the estimation variance depends on the size *M* of the ensemble. While for *M* = 10 the patterns is not represented well, for *M* = 100 the match is much closer. Of course, in the limit of $M \to \infty$, the results of the Ensemble Kalman Filter should approach those of the Kalman Filter for this simple linear case. The estimated state variable (i.e. runoff) is adjusted as the runoff measurements become available (lower panel of Figure 3.2). In contrast to the estimation variance, the estimated state variable is represented well for an ensemble size as small as 10. This is hardly surprising in the light of the efficiency of the method pointed out by Weerts and Serafy (2006), but nevertheless satisfying for the purposes of flood forecasting. While the flood forecasting model is non-linear, so the efficiency of the estimates will be different, the simple comparison does point to an order of magnitude of the ensemble size needed of *M* = 10, if the main interest lies in representing the state variable (i.e. runoff) well.

There are a number of possibilities for implementing the Ensemble Kalman Filter with a flood forecasting model that are related to formulating the errors and the states. One can separately represent different sources of the model error by different error terms. The advantage of doing this is that the physical basis of individual error sources remains clear. For example, one can separately represent errors in precipitation estimation, evaporation, as well as errors in model structure and model parameters. While the separate representation of many error sources is conceptually appealing it may be difficult in a practical application to specify the error distribution for each of the sources in a reliable way. If the model error assumptions are inappropriate, the updating may degrade the model performance as compared to the case without updating, as illustrated by Crow and Van Loon (2006) for the case of assimilating remotely sensed surface soil moisture. Also, some of the errors are likely correlated and, if the approach of separately representing component errors were adopted one would also have to account for these interrelationship. In this paper we have hence chosen to represent the errors in precipitation and evaporation input as the model error in an aggregate way, both for simplicity and parsimony.

In terms of formulating the states one possibility is a dual-state scheme where both runoff and soil moisture are treated as states with runoff observed but both states estimated. However, dual-state schemes may, potentially, give rise to identifiability issues. For example, Crow and Van Loon (2006) found that it was difficult to estimate two states (surface soil moisture and root zone soil moisture) from remote sensing data of surface soil moisture alone. As a remedy they recommended dual assimilation of both runoff observations and surface soil moisture observations (from remotely sensed data) that may allow more robust estimates of the two states. Dual assimilation of runoff observations and surface soil moisture observations would also be a possibility here but Parajka et al. (2005b) demonstrated that very little can be gained in terms of runoff prediction capabilities when assimilating remotely sensed soil moisture in Austria. While in this paper the suitability of a dual-state scheme has not been tested, a single-state scheme was hence considered a robust choice. The main idea of the approach chosen here is hence that the state vector is the runoff at the catchment outlet which is estimated according to Equation (3.6). The updated a posteriori runoff x_i^a represents the best estimate of the current runoff considering uncertainties of the model results and the runoff measurements. Therefore the updated runoff x_i^a provides a logical basis for the real-time flood forecast at the current time step *i*. However, the catchment soil moisture S_s, and the storage of the soil reservoirs S_1, S_2, S_3 of each grid element associated with x_i^a are unknown as they are propagated forward in time according to the non-linear model equations while x_i^a is estimated directly from Equation (3.6). To run the model in a forecast mode from the updated initial conditions, soil moisture and the storage of the soil reservoirs are required. They are also required for the forward propagation of the estimation covariance derived from the runoff ensemble. A simple similarity approach is hence adopted here to find soil moisture and the storage of the soil reservoirs of each pixel that is consistent with the a posteriori runoff x_i^a. For each ensemble member *m*, a set of *N* additional realisations is generated by forward propagation of the hydrologic model Φ which is the runoff model as presented in Blöschl et al. (2008):

$$x_{n,m,i}^f = \Phi(x_{n,m,i-1}^f, u_i + \varepsilon_{n,i}), \qquad n = 1,2,...,N \tag{3.13}$$

adding random errors $\varepsilon_{n,i}$ of precipitation and evaporation that are spatially uniform. These realisations are termed auxiliary realisations while the ensemble of *m* = 1, *M* contains the main realisations. The auxiliary realisations start from a time step where soil moisture and the storage of the soil reservoirs are known. At time step *j* the auxiliary realisations $x^f_{n,m,i}$ differ because of the random errors. One of the auxiliary realisations $x^f_{n,m,i}$ is closest to the a posteriori runoff $x^a_{m,i}$. This realisation $x^f_{n,m,i}$ is assumed to be consistent with $x^a_{m,i}$, i.e.

$$\left| x^a_{m,i} - x^f_{n,m,i} \right| \rightarrow \min \tag{3.14}$$

which gives the soil moisture and the storage of the soil reservoirs for all grid elements at time i for each realisation m. As the initial conditions of the real-time forecasts the realisation m is selected that is closest to the mean value of all realisations in terms of runoff, i.e. $\left| \overline{x^a_i} - x^a_{m,i} \right| \rightarrow \min$ with

$$\overline{x^a_i} = \frac{1}{M} \sum_{m=1}^{M} x^a_{m,i} \tag{3.15}$$

A schematic overview of the real-time model update with Ensemble Kalman Filter and similarity approach is given in Figure 3.3. The current time step is labelled *i* and the time increment is 1. Tests with the procedure suggested that it is useful to start the realisations at *u* time intervals before time step *i* for numerical reasons. Figure 3.3a shows three ensemble members of the Ensemble Kalman Filter with their respective values $x^f_{1,i}$, $x^f_{2,i}$, and $x^f_{3,i}$ at time step *i*. They approximate the probability density function (pdf) of the a priori estimates (dashed dotted line in Figure 3.3a). The perturbed observations $z_{1,i}$, $z_{2,i}$ and $z_{3,i}$ that approximate the pdf of the observation errors (dotted line in Figure 3.3a) are combined with the $x^f_{1,i}$, $x^f_{2,i}$, $x^f_{3,i}$ by Equation (3.6) to obtain the $x^a_{1,i}$, $x^a_{2,i}$, $x^a_{3,i}$ which approximate the pdf of the a posteriori estimates (solid line in Figure 3.3a). In the schematic of Figure 3.3, the Kalman gain has been chosen as K_i = 0.6. To obtain the soil moisture and the storage of the soil reservoirs of each pixel, auxiliary realisations are started at time step *i-u* .

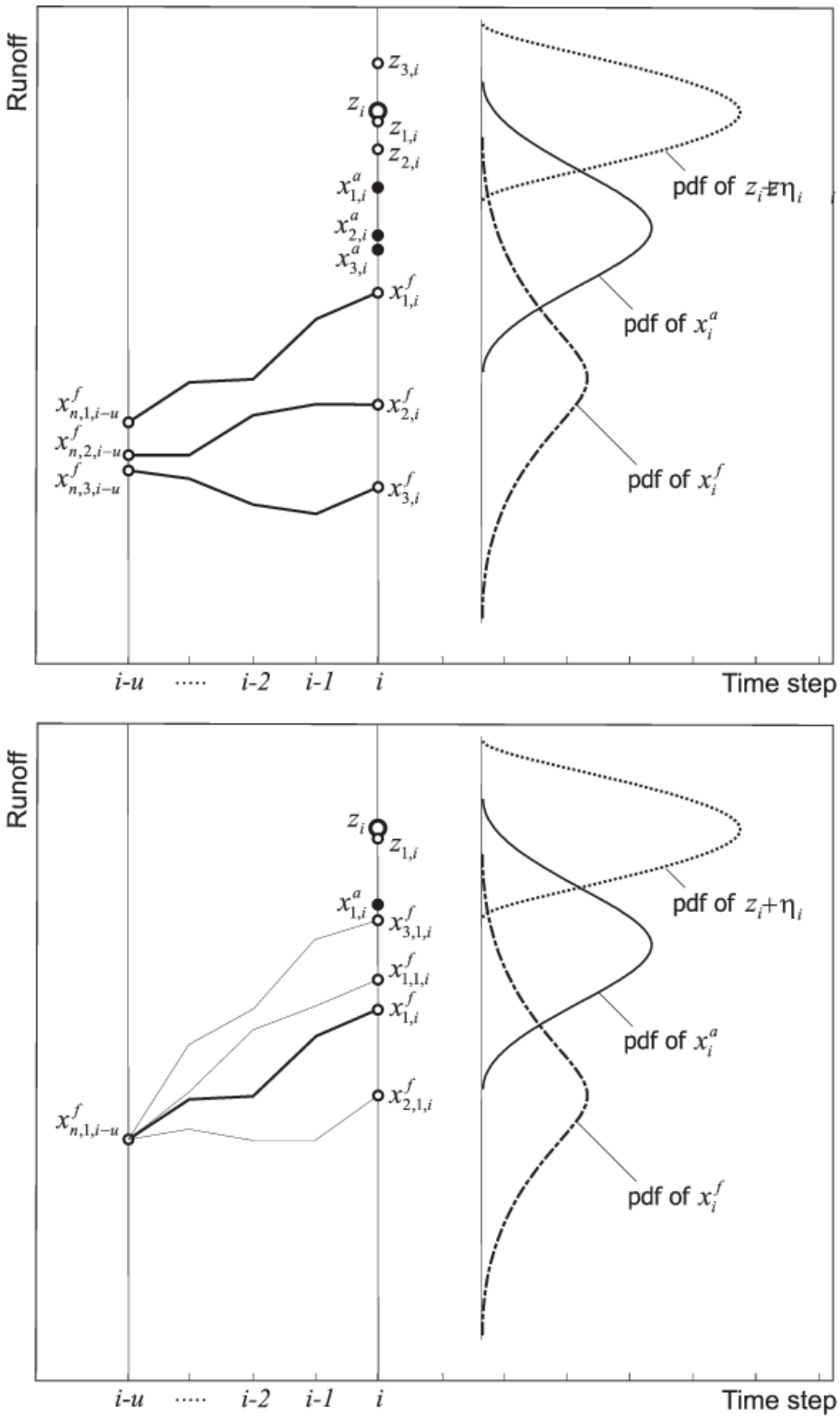

Figure 3.3 (a) Schematic of the Ensemble Kalman Filter approach. (b) Schematic of the similarity approach. For symbols see text.

In Figure 3.3b, N = 3 auxiliary realisations are shown for the main realisation m = 1 which produce $x^f_{1,1,i}$, $x^f_{2,1,i}$ and $x^f_{3,1,i}$. In the schematic, the auxiliary realisation n = 3 is the one that is closest to the a posteriori estimate of realisation m = 1 as $\left|x^a_{1,i} - x^f_{3,1,i}\right|$ is small. The soil moisture and the storage of the soil reservoirs of each pixel associated with the auxiliary realisation n = 3 is hence used to represent the a posteriori estimate of realisation m = 1. For the example in Figure 3.3 the initial conditions for the a posteriori estimate $x^a_{2,i}$ are used for the forecasts according to Equation (3.10).

Application of the Ensemble Kalman Filter to the Kamp catchment

The soil moisture and the storage of the soil reservoirs of the grid elements of the hydrologic model at the beginning of a flood event are clearly important for reliable flood forecasts. If the initial system state deviates from the optimal state, the flood forecasts will also be less than perfect. An overestimation of soil moisture at the beginning of a flood event would be expected to lead to an overestimation of the observed flood peak and, in a similar way, an underestimation of soil moisture would cause an underestimation of flood peaks. Biases in the soil moisture may be the result of small biases in the input, i.e. precipitation and evapotranspiration, that may accumulate over weeks and months. It is these biases the updating procedure of this papers aims to correct. While updating methods commonly used in real-time flood forecasting (e.g. Gutknecht, 1991) update runoff generation during events, the procedure presented here updates the evolution of soil moisture between events by attributing the model uncertainty to rainfall and evapotranspiration inputs. This means, it is the slow component of soil moisture change in the catchment that is adjusted. The model error variance must hence be set to reflect the slow processes. As there is a single stream gauge, the covariances simplify to scalar variances.

In order to find suitable parameters for the Ensemble Kalman Filter we performed extensive test calculations with different sets of parameters and different error models (white and red noise) for time periods including floods and low flow conditions. Based on the results of these calculations, the model variances for the ensemble of main and auxiliary realisations are set to V_i = 0.005 (mm/15min)². As the variance of the sum of independent random variables scales with the number of aggregation steps, this value is equivalent to an error standard deviation of 1.8 mm/week (with a

time step of 0.25 hours). This is the order of magnitude one would expect for the uncertainty of the precipitation measurements and estimation of evaporation, although it is difficult to separate the individual effects. It is clear that this magnitude relates to the small biases over a relatively long time period rather than to precipitation errors during a flood event which could be much larger. The small model variance updates the system states between the flood events to improve the initial conditions for the forecasts of future flood events.

The accuracy of runoff measurements tends to decrease with increasing runoff. Typically, the error standard deviation is set to a fixed percentage of runoff. The measurement error variance of runoff was hence formulated as $W_i = \xi \cdot z_i^2$. Again based on test simulations, ξ was set to ξ = 0.0025. Runoff measurement errors depend on the sampling method and on the local stream geometry but, typically, the error standard deviations are on the order of 5% of the runoff (Herschy, 2002). This means that the measurement error variance used here is the order of magnitude one would expect for the uncertainty of the runoff measurements.

One of the advantages of the Ensemble Kalman Filter is its flexibility with regards to the statistical characteristics of the model and measurement errors. During the parameterisation of the update procedure test calculations with red noise (i.e. temporally correlated) model errors were carried out. The test calculations indicated that for the red noise case, larger ensemble sizes *M* are needed than in the white noise case to get similar results. White noise error terms without a correlation in time were hence used for the model and measurement errors. As there is a single stream gauge, $\mathbf{C}_i = C_i = 1$.

Test simulations were performed to determine a suitable ensemble size. For a given time period the Ensemble Kalman Filter was run with a large ensemble size and the ensemble size was gradually reduced, similar to Figure 3.2. These comparisons suggested that an ensemble size of *M* = 10 gives very similar estimates of runoff to the case of large ensemble sizes with less than 1% difference. *M* = 10 was hence adopted in this case study. *N* was set to *N* = 10 in a similar comparison. The update interval was set to *u* = 12. With a time step of Δt = 0.25 hours the update interval Δt_u is hence 3 hours. This lag is needed because of the non-linearity – any additional

rainfall will not immediately produce a response at the catchment outlet as there is some time lag within the catchment. Table 3.1 summarises the parameters of the Ensemble Kalman Filter used in this paper.

Table 3.1 Parameters of the Ensemble Kalman Filter. z_i is measured runoff.

Parameter	Symbol	Unit	value
Time step	Δt	hrs	0.25
Updating time step	Δt_u	hrs	3
Measurement error variance of runoff	W_i	$(m^3/s)^2$	$\xi \cdot z_i^2$, ξ =0.0025
Model error variance	V_i	$(mm/15min)^2$	0.005
Ensemble size	M		10
Auxiliary ensemble size	N		10

Sensitivity to soil moisture

The way the Ensemble Kalman Filter operates in the Kamp model is illustrated in two scenarios (Figure 3.4 and Figure 3.5). To emulate the situation in real-time flood forecasting, the update of the system states is only performed during the low flow period before the first rise of the hydrograph on October 20, which would be the past in a forecast situation. From October 20 (which would be the future), the simulation is performed without any updating and observed precipitation from rain gauges is used as a model input for clarity. In Figure 3.4, the initial soil moisture at the beginning of the calculation period was set to 0%, i.e. the top soil was assumed to be perfectly dry. Shown in the graphs is the mean value of the relative soil moisture within the catchment

$$\overline{S_s/L_s} = \frac{1}{n_p} \cdot \sum_{k=1}^{n_p} \frac{S_{s_k}}{L_{s_k}} \tag{3.16}$$

where S_{s_k} and L_{s_k} are the simulated soil moisture and their limit at grid element k. n_p is the number of grid elements with n_p = 622. In a simulation mode without updating (dashed lines) the model consistently underestimates soil moisture and hence runoff.

When updating is allowed (thin solid lines) the model adjusts the perceived errors in precipitation and evaporation and hence increases soil moisture more quickly than in the simulation case. The updating of the model input affects the entire hydrologic system including the storage of the soil reservoirs not shown. The updated runoff hence reaches the level of the observed hydrograph after a short time period. In case of the simulation without updating, the flood peak on October 23 is clearly underestimated while the forecast that uses updated initial conditions (thin solid lines) is much closer to the observed hydrograph. As noted above, it is antecedent soil moisture that is aimed to be improved on by the updating procedure.

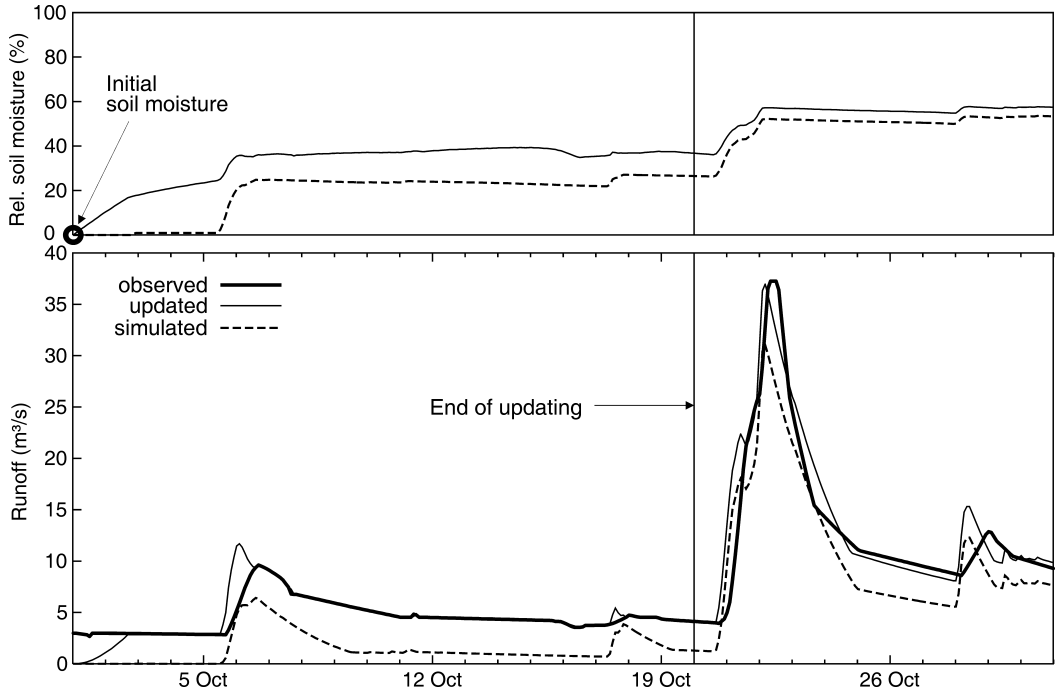

Figure 3.4 Scenario with initial soil moisture set to an arbitrary low value to illustrate the effect of updating by the Ensemble Kalman Filter on mean relative soil moisture and runoff. October 1996, stream gauge Zwettl/Kamp.

A similar scenario, but with very wet initial conditions is shown in Figure 3.5. The effect of the updating is similar in that it adjusts the soil moisture to a reasonable value. Without updating the flood peak is vastly overestimated as a consequence of the overestimated soil moisture at the beginning of the flood event. A comparison of Figure 3.4 and Figure 3.5 indicates that, in both cases, updated soil moisture converges to a value that is consistent with runoff. On Oct. 20 (i.e. the hypothetical time of the

forecast) soil moisture in both Figure 3.4 and Figure 3.5 was 38%, while without updating, it was 23% and 79%, respectively.

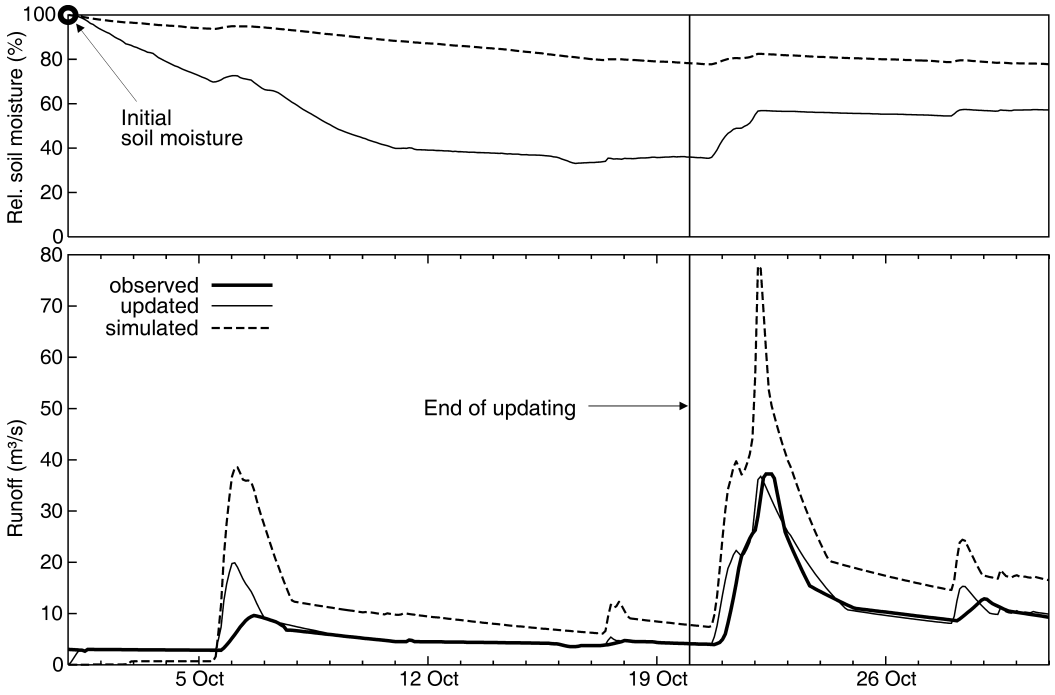

Figure 3.5 Scenario with initial soil moisture set to an arbitrary large value to illustrate the effect of updating by the Ensemble Kalman Filter on mean relative soil moisture and runoff. October 1996, stream gauge Zwettl/Kamp.

The scenarios illustrate that accurate estimates of antecedent soil moisture are indeed of utmost importance for producing accurate forecasts. Inadequate initial moisture can be corrected and suitable moisture conditions can be estimated by updating the model input during the dry period before the flood event on October 23.

3.3 Results

Updating soil moisture in a simulation mode

Figure 3.6 shows the results of simulation runs with and without model update from May to September 2005. The calculation results with model update are simply the analysed state estimates x_i^a. During this period, the simulation without updating performs very well. Both the shape and the peaks of the simulated flood hydrographs are close to the observations. The cumulative errors (lower part of Figure 3.6) are

very small. The cumulative error never exceeds $7*10^6$ m³ within this period which is small as compared to the total flow volume of $140*10^6$ m³. This is because of the favourable model performance. There is a slight improvement in the May event and the August events, but overall there is hardly any difference between the simulations with and without updating. This example is the ideal case for real-time flood forecasting, where the model performs well in the simulation mode, so one would also expect the model to work well in the forecasts.

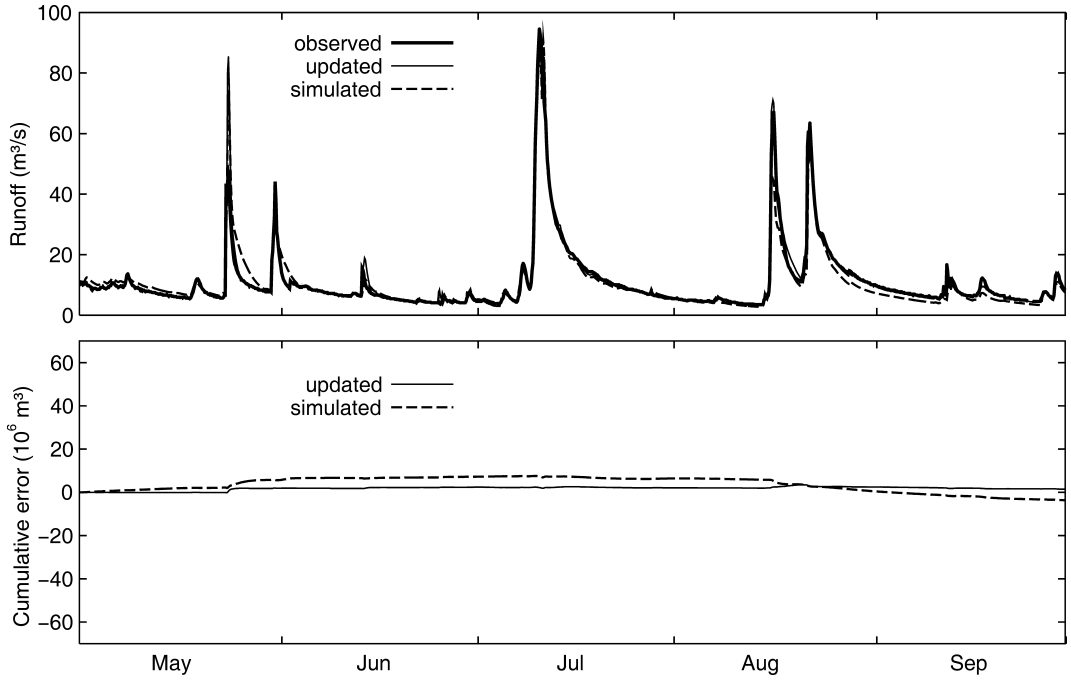

Figure 3.6 Simulations without updating (dashed lines) and updating (thin solid lines) of runoff (top) and cumulative errors (bottom) at Zwettl/Kamp from May to September 2005. Example of excellent model performance where the benefits of updating are small.

An alternative example is shown in Figure 3.7 for the period from November to April 2006. Until the end of December the simulated hydrograph is slightly lower than the data. This is most likely due to uncertain precipitation and evaporation inputs during this relatively dry period. From January until the end of March the differences between simulation and observation increases which is reflected in a progressive increase in the negative cumulative errors. During this period the likely reason for this underestimation are the uncertainties in simulating snow accumulation and snow melt. The effect of these biases is the underestimation of the soil moisture at the beginning of the flood event in April 2006. As a result, the entire flood event in April is substantially underestimated. In contrast, the simulation with updating performs much

better during the low flow period until the end of March. The antecedent soil moisture at the beginning of the flood event in April is larger than for the simulation case without updating and the flood event is represented much more accurately. For this example, the advantage of the updating during the low flow period is obvious.

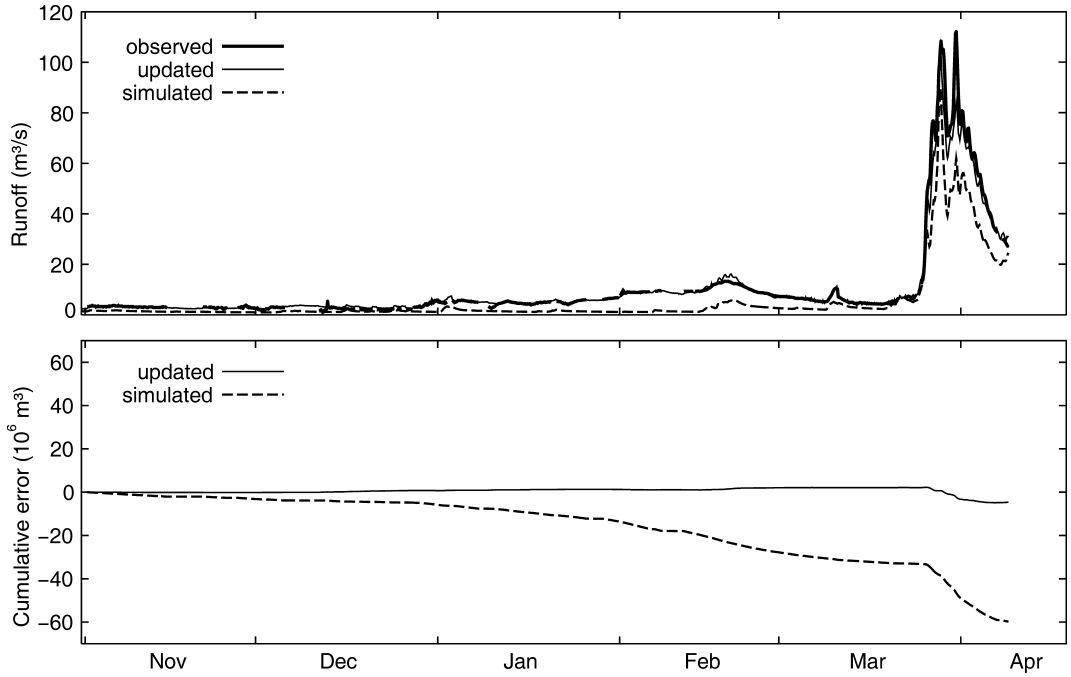

Figure 3.*7* Simulations without updating (dashed lines) and updating (thin solid lines) of runoff (top) and cumulative errors (bottom) at Zwettl/Kamp from November 2005 to April 2006. Example of poor model performance where the benefits of updating are significant.

Updating soil moisture in a forecast mode

The examples in Figure 3.6 and Figure 3.*7* were illustrative of the merits of updating, depending on the performance of the simulation per se. In a forecast situation, however, the updating is for the past only. The forecast starts with the updated initial conditions but, of course, with no additional updating of the forecast as future runoff data are not available. This situation is illustrated in Figure 3.8 and Figure 3.9. Up to the time the forecast is made (vertical lines in Figure 3.8 and Figure 3.9), the updating is as in Figure 3.6 and Figure 3.*7* but beyond that point in time no more updating is allowed although future precipitation is assumed to be known. The difference between the updating and no updating (simulation) cases in Figure 3.8 and Figure 3.9

for the points in time later than the forecast time is hence only related to the difference in the initial conditions at the forecast time.

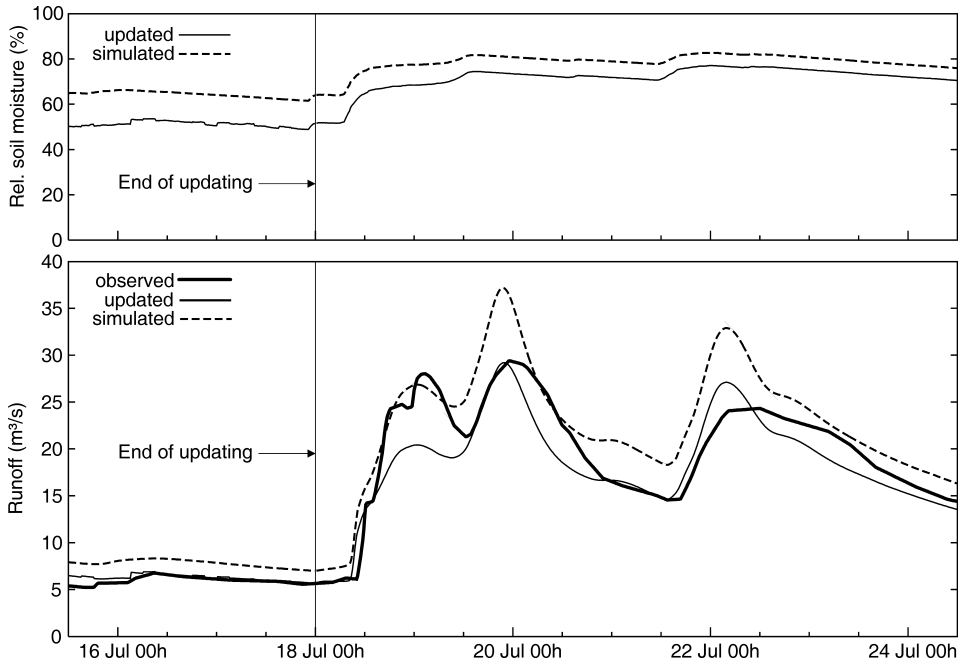

Figure 3.8 Effect of updating soil moisture in the forecast mode. The forecast was started from simulated and updated initial conditions on 18 July 1997 at 0 h. Future precipitation is assumed to be known but no updating is performed beyond the forecast time (vertical line). Zwettl/Bahnbrücke (622 km²).

The upper panel of Figure 3.8 shows simulated and updated mean relative soil moisture $\overline{S_s/L_s}$, the lower panel shows the associated hydrographs. During July 16 and 17 before the start of the event, runoff is overestimated in the simulation (no updating) case because soil moisture and the storage of the soil reservoirs are overestimated as a result of biases accumulated over the previous months. The updating brings soil moisture and the storage of the soil reservoirs as well as runoff down, so that runoff is very similar to the data. At the time the forecast is made, relative soil moisture is 62% and 51% in the simulation and updating cases, respectively. These are the initial conditions for the forecasts along with the storage of the soil reservoirs S_1, S_2, S_3 not shown. The forecast based on the simulated initial conditions overestimates the observed hydrograph during most of the forecast lead time (19-22 July). The forecast based on the updated initial conditions does underestimate the first peak but performs substantially better for the remaining forecast lead time.

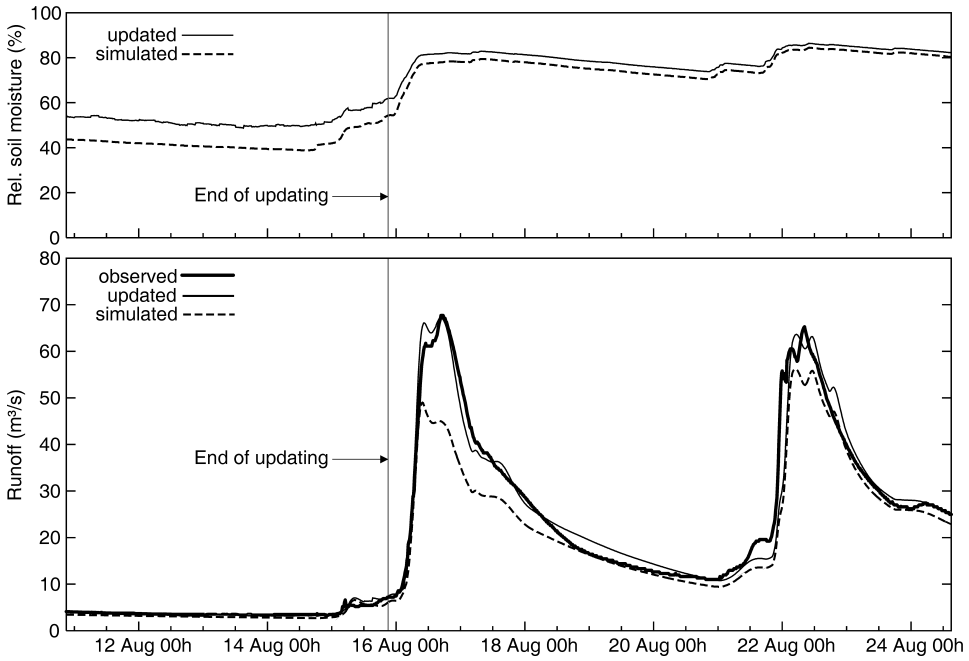

Figure 3.9 Effect of updating soil moisture in the forecast mode. The forecast was started from simulated and updated initial conditions on 15 August 2005 at 21 h. Future precipitation is assumed to be known but no updating is performed beyond the forecast time (vertical line). Zwettl/Bahnbrücke (622 km²).

Figure 3.8 is an example where soil moisture (without updating) is overestimated prior to the event which is quite apparent in the overestimation of runoff. Figure 3.9 shows the converse example where soil moisture (without updating) is underestimated prior to the event but this is not so obvious in the hydrograph. In fact, the simulated initial runoff is only slightly lower than the measurement but the flood peak of the following event is clearly underestimated by the simulation. In this example, the updated initial soil moisture improves the forecast accuracy very substantially which is due to the updating of soil moisture during the dry period before the flood event. It is interesting that the non linearity of the rainfall-runoff model amplifies the small differences in runoff prior to the event. This means that small differences between simulated and observed hydrographs can have a great effect on the runoff forecast. Conversely, these small differences can be exploited to improve the forecasts. It is also interesting that the difference in soil moisture of the updated and simulated forecast runs decreases during the forecast period. This is due to the formulation of the soil moisture accounting scheme (Equation A.1) which is a stable dynamic system where small perturbations in the initial conditions vanish over time. For the second event,

hence, the difference between the two runoff forecasts (with and without updating) is much smaller than for the first event in Figure 3.9.

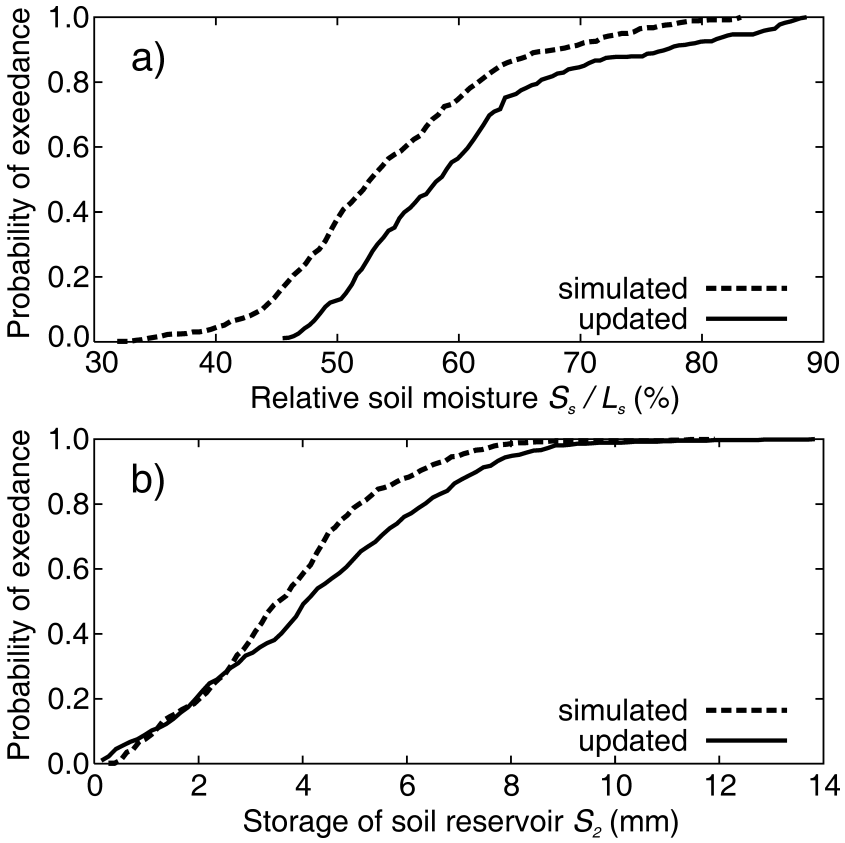

Figure 3.10 Spatial distributions of simulated and updated relative soil moisture S_s/L_s and soil storage S_2 on 15 August 2005 at 21 h at Zwettl/Bahnbrücke (622 km²) used as initial conditions for the forecasts in Figure 3.9.

The previous figures have illustrated the temporal evolution of mean relative soil moisture. The model used is a distributed model where the model parameters are non-uniform in space and the inputs also differ spatially. The soil moisture is hence variable within the catchment. It is of interest to see how this spatial distribution changes with the updating. Figure 3.10a shows a comparison of the spatial distribution of relative soil moisture within the catchment at the start of the forecast run on 15 August 2005 at 21 h (vertical line in Figure 3.9). In this example, the updating increases mean relative soil moisture from 0.54 to 0.60 (Figure 3.9) which is also apparent in Figure 3.10a. It is mainly the mean that increases while the shape of the distribution does not change much. This spatial distribution indicates that most of the runoff stems from a relatively small portion of the catchment with above soil moisture (Equation A.1) and this spatial distribution is maintained in the updating. Indeed, the

assumptions involve spatially uniform random errors $\varepsilon_{n,i}$ of precipitation and evaporation. Figure 3.10b shows the corresponding spatial distribution of the storage of the soil reservoir S_2. This soil reservoir has a storage parameter k_2 that ranges between 6 and 17 days within the catchment, so represents an intermediate component in terms of the timing of runoff response. It is interesting that it is mainly the wetter parts of the catchment where the updating increases the soil storage, while the relatively dry parts remain almost unaffected. The wetter parts (larger S_2) are those that are hydrologically more active, and are also those that are more affected by the updating as one would expect.

Performance for large flood events

Most of the time, updating soil moisture leads to an improvement of the forecast accuracy. In particular, during low flow and average flow conditions the forecasts are very close to the data. However, the main interest in this paper is on flood forecasting, and in particular on the forecasting of large floods. The six largest flood events on record at the Kamp have hence been examined in more detail (Table 3.2). Some of these events are indeed extraordinary events. Flood records at the Kamp have been available since, 1977, and flood marks and archive information from the early 19th century. Based on this information, the largest flood on record (first event in August 2002) was assessed to be on the order of a 1000 year flood (Blöschl and Zehe, 2005). Some of the other floods are also large (second event in August 2002, about 500 years; March 2006 about ten years return period). The data set is hence particularly well suited to address the science question of whether the updating prior to events will actually improve the forecasts of large floods.

As in the previous analyses, two cases were examined, with and without updating soil moisture. In a first step the ability of the updating procedure to improve on the forecast of the flood peaks is examined. To this end, the forecasts are analysed that have been made 3 hours before each flood peak occurred. For example, for the first event in August 2002, the flood peak occurred on August 8 at 0 hrs, so the forecast made on August 7, 21 hrs is analysed. Future precipitation was assumed to be known as in all the previous analyses, but no updating beyond the forecast time was allowed.

Table 3.2 Flood peaks, return periods and evaluation periods for the statistical error analysis of the six largest flood events on record at Zwettl/Bahnbrücke (622 km²).

	Aug 2002a	Aug 2002b	July 2005	Aug 2005a	Aug 2005b	Mar 2006
Observed flood peak (m³/s)	459	367	95	68	65	112
Return period of peak (yrs)	~ 1000	~ 500	5	3	3	10
Peak time	8 Aug, 0h	13 Aug, 13h	11 Jul, 10h	16 Aug, 17h	22 Aug, 8h	31 Mar, 23h
Beginning of entire event	6 Aug, 0h	11 Aug, 0h	5 Jul, 0h	14 Aug, 0h	20 Aug, 0h	25 Mar, 0h
End of entire event	10 Aug, 21h	15 Aug, 21h	15 Jul, 0h	19 Aug, 21h	26 Aug, 21h	5 Apr, 12h
Beginning of rising limb	6 Aug, 12h	11 Aug, 12h	10 Jul, 12h	16 Aug, 0h	21 Aug, 12h	26 Mar, 6h
End of rising limb	8 Aug, 6h	13 Aug, 18h	11 Jul, 6h	17 Aug, 21h	22 Aug, 12h	2 Apr, 3h

The results of the comparison are shown in Figure 3.11. For five out of the six flood events, the flood peaks are indeed improved. For example, the peak flow of the largest event was observed as 459 m³/s while the forecast without and with updating soil moisture gives 508 and 470 m³/s, respectively. The improvement of updating is larger for those events that are not represented so well in the simulation case. For the smallest event, the peak flow is slightly deteriorated (65 m³/s observed and 56 and 53 m³/s, respectively, without and with updating). The mean normalised absolute error of the peaks

$$e = \frac{1}{p}\sum_{k=1}^{p}\frac{\left|\hat{Q}_k - Q_k\right|}{Q_k} \tag{3.17}$$

was evaluated where Q_k are the observed flood peaks and $\hat{Q}_k$ are the flood peak forecasts and p = 6. For the six peaks in Figure 3.11 the mean normalised absolute error of the peaks is 25% without updating and decreases to 12% with updating. This is for a lead time of 3 hours. For a lead time of 48 hours the mean normalised abso-

lute error of the peaks is 25% without updating and decreases to 19% with updating. It is clear, that overall, there are significant merits of the updating in terms of forecasting peak flows.

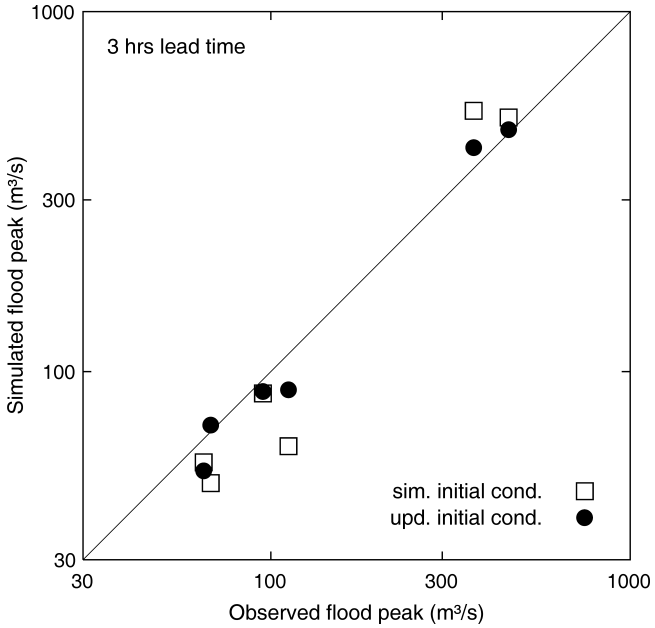

Figure 3.11 Comparison of the forecasted peak flows with and without updated initial conditions for the six largest flood events on record as of Table 2. Both forecast runs (updated and simulated) were started three hours before the observed flood peaks (forecast lead time of three hours) based on observed precipitation inputs.

In a second step, the forecast accuracy of the two cases is analysed for the entire events rather than the peaks only. Two error measures are used, the mean normalised absolute error e_j (Equation 3.13) and the Nash-Sutcliffe efficiency E_j (Equation 3.14):

$$e_j = \frac{1}{i_2 - i_1} \sum_{i=i_1}^{i_2} \frac{\left| \hat{Q}_{ij} - Q_i \right|}{Q_i} \tag{3.18}$$

$$E_j = 1 - \frac{\sum_{i=i_1}^{i_2} (Q_i - \hat{Q}_{ij})^2}{\sum_{i=i_1}^{i_2} (\overline{Q} - Q_i)^2} \tag{3.19}$$

where *j* is the forecast lead time, $\hat{Q}_{ij}$ is runoff at time step *i* that is forecasted with a lead time of *j*, Q_i is the observed runoff at time step *i*, and i_1 and i_2 are the beginning

and the end of the analysis interval, respectively (Table 3.2). In this analysis, the forecasts were made at three hour intervals and different lead times of up to 48 hours were analysed. We analysed two evaluation periods (i_1 to i_2); entire flood events, and the rising limbs only (see Table 3.2). The forecast errors for the entire events and the risings limbs are shown in Figure 3.12 a) and b) respectively. In all instances, the updating of soil moisture reduces the forecast errors. For a lead time of 3 hours, for example, the errors decrease from 20 to 12% in the case of the entire events, and 33 to 15% in the case of analysing the rising limbs only. For the case of simulated initial conditions, the forecast errors do not change with lead time as would be expected, as this is a simulation case where the forecast time does not come into play. In contrast, for the case of updated initial conditions, the errors are smallest for the short lead times, which again would be intuitively expected. At the time of the forecast, observed runoff at the time of the forecast captures some of the hydrological process dynamics that continue over the following hours. As the memory fades away with time, the improvement in forecast accuracy is largest for the short lead times. It is interesting that even after a forecast lead time of 48 hours the updated initial conditions improve the forecasts substantially. Quite clearly, it is not only the fast runoff components that contribute to a given forecast accuracy.

The errors for the rising limbs (Figure 3.12a) are generally larger than those for the entire flood events (Figure 3.12b). This is because the forecast errors during the rising flood limbs are larger than those during the falling limbs due to rainfall uncertainty. During the falling limb rainfall is zero or very small, so rainfall uncertainty is small too. It is also possible, that the fast components of runoff are more uncertain than the slow components. Generally speaking, the rising limbs are more difficult to predict than the falling limbs but it are the former that are of most interest to flood management. Figure 3.13 shows the results of the Nash-Sutcliffe efficiency. This error measure involves squared errors (Equation 3.14), so the large deviations from the data are weighted more strongly than in the case of the mean absolute error.

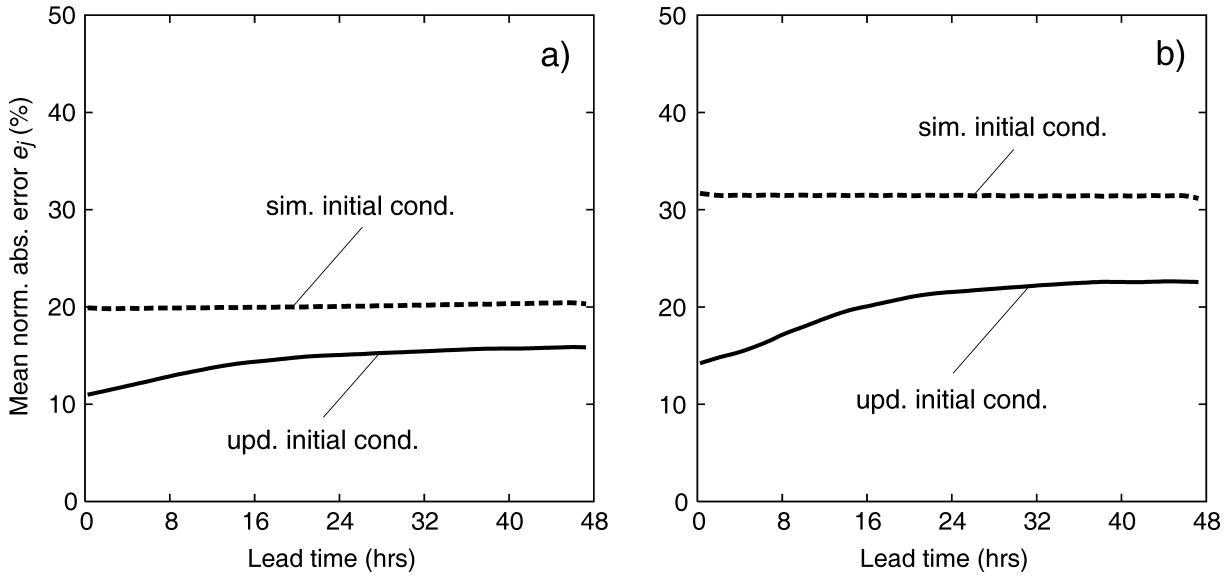

Figure 3.12 Forecast errors (Equation 3.13) for the six largest flood events on record (Table 3.2) assuming future precipitation is known. Dashed lines relate to the forecasts with simulated soil moisture (no-updating), solid lines to the forecasts with updated soil moisture. (a) Entire flood events; (b) rising limbs only.

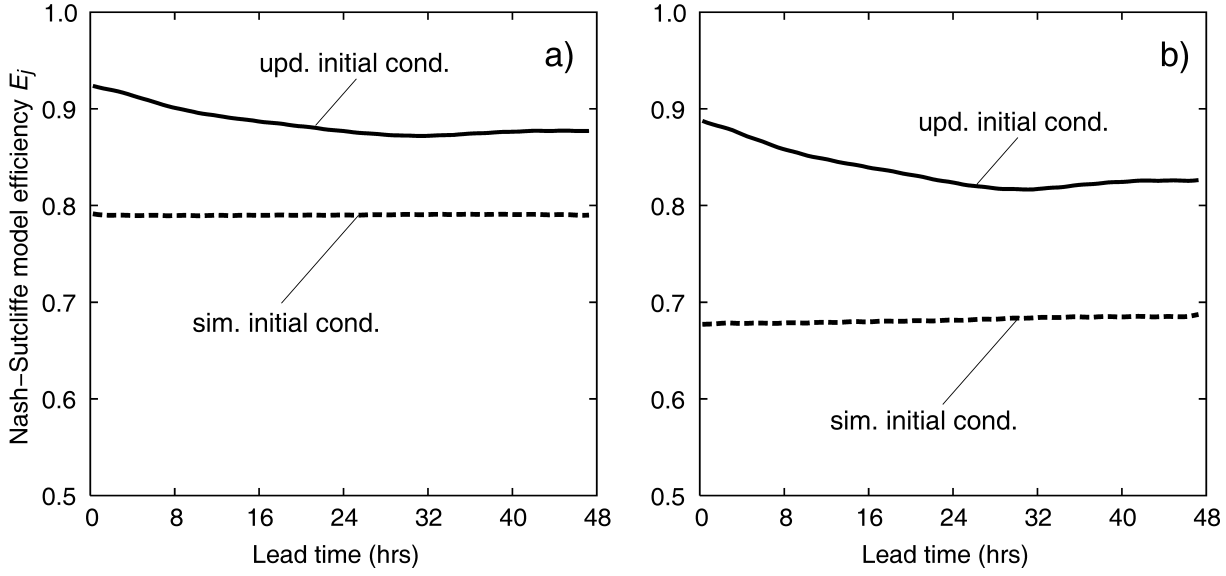

Figure 3.13 Nash-Sutcliffe model efficiency of the forecasts (Equation 3.14) for the six largest flood events on record (Table 3.2) assuming future precipitation is known. Dashed lines relate to the forecasts with simulated soil moisture (no-updating), solid lines to the forecasts with updated soil moisture. (a) Entire flood events; (b) rising limbs only.

The error pattern of the Nash-Sutcliffe efficiency is similar to that of the mean absolute error. In both evaluation periods, rising limbs and entire flood events, the model efficiency is improved by the updating. For the entire flood events, the Nash-Sutcliffe

efficiency at a forecast lead time of 48 hours increases from 0.79 to 0.88 by the updating, and it increases from 0.79 to 0.92 at a lead time of 3 hours. For the rising limbs, the Nash-Sutcliffe efficiency at a forecast lead time of 48 hours increases from 0.68 to 0.82 by the updating, and it increases from 0.68 to 0.88 at a lead time of 3 hours.

3.4 Discussion and Conclusions

Renewed interest in updating methods in hydrology has come from the availability of Monte Carlo methods because of their flexibility, ease of use and operational robustness. The Ensemble Kalman Filter extends the traditional Kalman Filter concept by Monte Carlo techniques and is able to deal with non-linear model dynamics in a natural way. The aim of this paper is to examine the benefit of Ensemble Kalman Filter updating in forecasting large floods. The soil moisture of a distributed runoff model is updated based on observed runoff. The updated soil moisture is then used as an initial condition for the forecasts. The ensemble size was set to M = 10 with N = 10 auxiliary realisations. Hardly any improvement in forecast accuracy was obtained when increasing the ensemble size in test simulations. A typical ensemble size used in the updating of hydrological models is 50 (Moradkhani et al., 2005). The main interest of this paper was, however, not in the particular formulation of the Ensemble Kalman Filter but in the degree it will actually improve the forecasts for a real world case. The value of updating is obvious because of the long time scales associated with the hydrological processes during low and average flows. For large flood flows, the difficulty with updating runoff during an event is that phase errors usually cannot be handled well. There can be overshooting of the forecasts if phase errors are interpreted as volume errors. The procedure examined here mainly updates the slow runoff components, i.e., soil moisture between events which is then used as an initial condition for the flood forecasts. Sensitivity analyses and comparisons of individual events suggest that the concept of updating the slow component is plausible and robust. It is interesting that the non linearity of the rainfall-runoff model amplifies the small differences in runoff prior to the event. This means that small differences between simulated and observed hydrographs can have a great effect on the runoff forecast (Zehe and Blöschl, 2004). Conversely, these small differences can be exploited to improve the forecasts. The updating mainly changes the mean value of the catchment soil

moisture, while the spatial structure of the moisture distribution is preserved during the update. Therefore, increasing catchment soil moisture leads to an increasing fraction of runoff contributing areas within the catchment. Analyses of six large flood events at the Kamp indicate that the updating indeed reduces forecast errors substantially during the flood events. It is considered a strength of this case study that data on a number of large floods (including two extreme floods) were available which is not usually the case in practical applications. This is important as one of the main motivations of implementing flood forecasting systems is to improve on the forecasting of extreme events where the damage potential is largest (Apel et al., 2006).

Nash-Sutcliffe efficiencies of runoff models without updating reported in the literature are, typically, on the order of 0.7 to 0.9 (e.g. Parajka et al, 2005a). The efficiencies without updating found in this paper are at the lower end of this range (Figure 3.13). It should be noted that low flow and average flow conditions can usually be simulated much more accurately than flood flows. For comparison, the Nash-Sutcliffe forecast efficiency at the Kamp was evaluated for entire years (as opposed to events) following an analogous procedure. The efficiencies without updating were always larger than 0.85 and increased to more than 0.98 if updating of soil moisture was allowed. Clearly, the updating is most efficient for low and medium flows, but from a practical perspective the flood flows are usually of much more interest. However, these tend to be more difficult to predict and errors are usually much larger. For example, a model comparison of Reed et al. (2004, their Figure 18b) gave mean normalised absolute errors of peak flows in a typical range of 20 to 50%, depending on the model and the catchment analysed. Based on the results of this study, one would expect that such errors could be substantially reduced if soil moisture were updated. In the present paper, the peak flow errors for 3 hour forecasts were reduced from 25% to 12% by the updating procedure, and from 25% to 19% for 48 hours forecasts. It should be noted that the forecast lead time of 48 hours is much larger than typical flow travel time in the streams within the catchment which are less than 2 hours. It is hence the water in the landscape rather than that in the stream that needs to be adjusted in this case study.

Remotely sensed soil moisture is sometimes used for updating the soil moisture of hydrological models. The significant increase in forecast accuracy found here sug-

gests that use of runoff data to infer catchment soil moisture may be an efficient alterative to remote sensing data. In fact, in the study area examined here it appears that updating soil moisture through observed runoff is a better choice than to directly use remotely sensed soil moisture data for updating (Parajka et al., 2005b).

The model parameters and structure were chosen very carefully in this case study. The model identification procedure went substantially beyond the calibration to runoff. Piezometric head data, and information from local surveys and other sources (such as snow data, Parajka and Blöschl, 2006) were used and combined by hydrological reasoning. This means that the model can be expected to represent the hydrological processes in the Kamp catchment reasonably well. We believe it is important to very carefully adjust the model to the local conditions (going beyond calibration to runoff) for the updating procedure to work efficiently. The events in 2005 and 2006 (Table 3.2) were not used for calibration but retained for model validation. In the current procedure, the main error source is attributed to the inputs (rainfall, evaporation) and their effect on soil moisture, so model parameters are not updated. A plausible model structure and carefully adjusted model parameters are hence the basis for a good performance of the updating routine. This is important as it then avoids the "flogging a dead horse" syndrome, i.e. attempting to update models that do not represent the processes well. Also, the availability of input data (16 rain gauges for model development, 8 telemetered rain gauges in a 622 km² catchment) along with radar data in this study is probably more than what one usually encounters in operational applications. With these caveats, it is suggested that updating procedures such as the one proposed in this paper can indeed substantially improve the forecasting of large floods at the catchment scale examined here.

4 A comparison of in-situ, ASCAT and model estimates of soil moisture in Austria

Abstract

Soil moisture is a key variable in both the land energy balance and the generation of runoff from catchments. Unfortunately, soil moisture is very difficult to measure at the scales of hydrologic interest. The aim of this paper is to examine the spatial and temporal dynamics of soil moisture estimates from model simulations, in-situ measurements and remote sensing. Ground based measurements in a soil depth of 10 cm at three field sites, remotely sensed surface soil moisture estimates from the Advanced Scatterometer (ASCAT), on board the Meteorological Operational (METOP) satellite series and simulated soil moisture from a spatially distributed hydrological model based on a dual layer soil moisture accounting scheme were used in this study. The evaluation is based on a three year period from 2007 to 2010. The study area at the Kamp river (1550 km²) and the in-situ field sites are located in the north-eastern part of Austria The comparison of the different soil moisture products gives correlations *R* of 0.68, 0.38 (0.61) and 0.51 (0.8) for ASCAT SSM and modelled skin layer, in-situ and ASCAT SSM (SWI) and in-situ and modelled skin (main) layer for period from April to September. An important issue in such cold regions is how to deal with periods of snow and frozen ground as microwaves can only detect liquid water. In this paper the benefit of a number of dynamic masking criteria is examined. The comparison of simulated and TDR measurements to ASCAT soil moisture estimates during the winter period indicates a significant improvement of the consistency between the data sets is indicated. The greatest improvement, with correlations from October to March rising from -0.12 to 0.71, was observed for the coldest region, the Kamp catchment.

4.1 Introduction

Soil moisture is a key variable in both the land energy balance and the generation of runoff from catchments. Unfortunately, soil moisture is very difficult to measure at the

scales of hydrologic interest. Ground based (or in-situ) measurements of soil moisture are usually performed for individual soil profiles at a number of locations within a catchment (Western et al., 2002). With appropriate sensors and measurement setups the accuracies can be very good and the temporal resolution can be on the order of minutes and less. However, it is difficult to cover large areas by the sensors due to logistic constraints, and the spatial support or foot print of one measurement is usually only a few centimetres (Grayson and Blöschl, 2000). This makes it very difficult to estimate meaningful spatial averages over catchments. While methods for estimating catchment average soil moisture from limited point measurements have been developed (Grayson and Western, 1998), identification of representative points in a particular catchment remains a challenge.

An alternative to ground based soil moisture measurements is the use of remote sensing methods. Spaceborne microwave sensors have been widely used for soil moisture retrieval because of their large sensitivity of the microwave response to water content in the soil surface layer (Ulaby et al., 1982; Wagner et al., 2007). The main advantage of spaceborne sensors is that they provide an integral value over an area rather than point values and most of the data are available at a global scale. However, the scale problem in the case of spaceborne sensors are relatively large footprints (or pixel sizes) relative to the hydrological processes of interest, relatively low repeat cycles (typically one or a few days) relative to the soil moisture dynamics and, perhaps most importantly, limited penetration depths of the microwave signal into the ground of a few centimetres or less. When comparing spaceborne microwave data with in-situ measurements to assess their relative accuracy these scale incompatibilities need to be accounted for in some way. To account for the scale differences in the footprints, often, relative soil moisture values (scaled by the minimum and maximum over a long record) are compared with the scaled in-situ data. To account for the low repeat cycles, comparisons use the exact time of the satellite overpass. To account for limited penetration depths Wagner et al. (1999) proposed a soil water index, *SWI*, which represents satellite based soil moisture time series for a pixel filtered by an exponential kernel with a set time constant T found from calibration. Similar scale inconsistencies apply when comparing space borne data with hydrological models but limited penetration depth are usually dealt with by using a skin layer soil moisture model (e.g., Georgakakos and Baumer, 1996). There is also an

issue with the interpretation of the microwave response for snow covered areas and frozen soils (Bartalis et al., 2008), as the microwave response can only detect liquid water while it is insensitive to water in the solid phase.

Because of the scale incompatibility, comparisons of ground based (in-situ), space-borne and modelled soil moisture usually give larger discrepancies than would result from the sensor characteristics (and modelling uncertainties alone). Also, because of the effect of snow and ice one would expect that the consistency of the different methods depends on the climate of the catchment under study. However, comparisons are very useful to understand the uncertainties involved in the measuring or modelling soil moisture by the three methods.

Anguela et al. (2008) compared the three methods in a catchment near Paris, France. In situ measurements were collected at 6 locations at depths of 5 cm to 155 cm, ERS-Scatterometer surface soil moisture and SWI were used, and the soil moisture was simulated by the SIM model that involves a skin layer and parameters from a global soil data base. The comparison for the skin layer gave coefficients of determination (r^2) of 0.47 (in-situ and simulation), 0.44 (simulation and ERS) and 0.53 (in-situ and ERS surface soil moisture) and the correlations increased to 0.90, 0.78 and 0.66, respectively, when the root zone was compared instead of the skin layer. Albergel et al. (2010) compared the same model with ASCAT surface soil moisture and in-situ measurements at depths of 5 cm at 11 field sites in south-western France during 2007 and 2008. The comparison gave correlations R of 0.7 (in-situ and simulation), 0.65 (simulation and ASCAT) and 0.59 (in-situ and ASCAT surface soil moisture). The correlations increased in winter which was related to more consistent rainstorms in winter. Because of the Mediterranean climate, minimum snow and soil freezing effects were observed. A similar study in a warm climate (the Tiber catchment in Italy) was performed by Brocca et al. (2010) with correlations around 0.67 (in-situ and ASCAT surface soil moisture), around 0.93 (in-situ and ASCAT SWI) and above 0.74 (simulation and ASCAT surface soil moisture). When snow and freezing processes are important one would expect lower correlations. For a catchment in Luxembourg, Matgen et al. (2011) post-processed the ASCAT data to obtain the same distribution function as those of the in situ-data at 16 sites. With this distribution

matching they obtained a mean correlation *R* 0.82 but with the unprocessed ASCAT surface soil moisture product the mean correlation was 0.50.

While a number of comparisons of in-situ, space-borne and modelled soil moisture in warm climates have been published in the literature, much less is known about comparisons in climates where snow and frozen ground play a significant role in winter. Also, masking methods for dealing with snow and frozen ground have not been fully evaluated for such comparisons. The aim of this paper therefore is to compare space-borne (ASCAT) soil moisture with in-situ and modelled soil moisture in Austria where winters can be cold, and to test the value of masking criteria.

4.2 Field sites and data

Four sites are used in this study, which are all located in northern Austria (Figure 4.1). The Kamp catchment has a total catchment area of 1550 km² and is located in the north-west of Austria. The geology of the catchment is mainly granite and gneiss. Weathering has produced sandy soils with a large storage capacity throughout the catchment. 40% of the catchment is forested and the rest is mainly cropland and grassland. For the Kamp catchment ASCAT soil moisture was compared with model simulations. The model was driven by rainfall interpolated from eight rain gauges were combined with radar information as well as air temperature from 8 stations. Runoff data were collected at a number of locations to test the hydrological model.

For the remaining three sites ASCAT soil moisture was compared with model simulations and TDR soil moisture measurements.

The Petzenkirchen site is located in Lower Austria south of the Danube (Figure 4.1). The soil in the upper horizon (0-20 cm) is sandy loam and the land use is grassland (Figure 4.2). The TDR sensors (TRASE-System, SOIL MOISTURE EQUIPMENT, California) were horizontally installed from a pit at different soil depths from 10 cm to 180 cm. The data were recorded at 3 hour intervals by a data logger at the site. Rainfall and air temperature data used in the modelling were recorded at the site with a temporal resolution of 10 minutes.

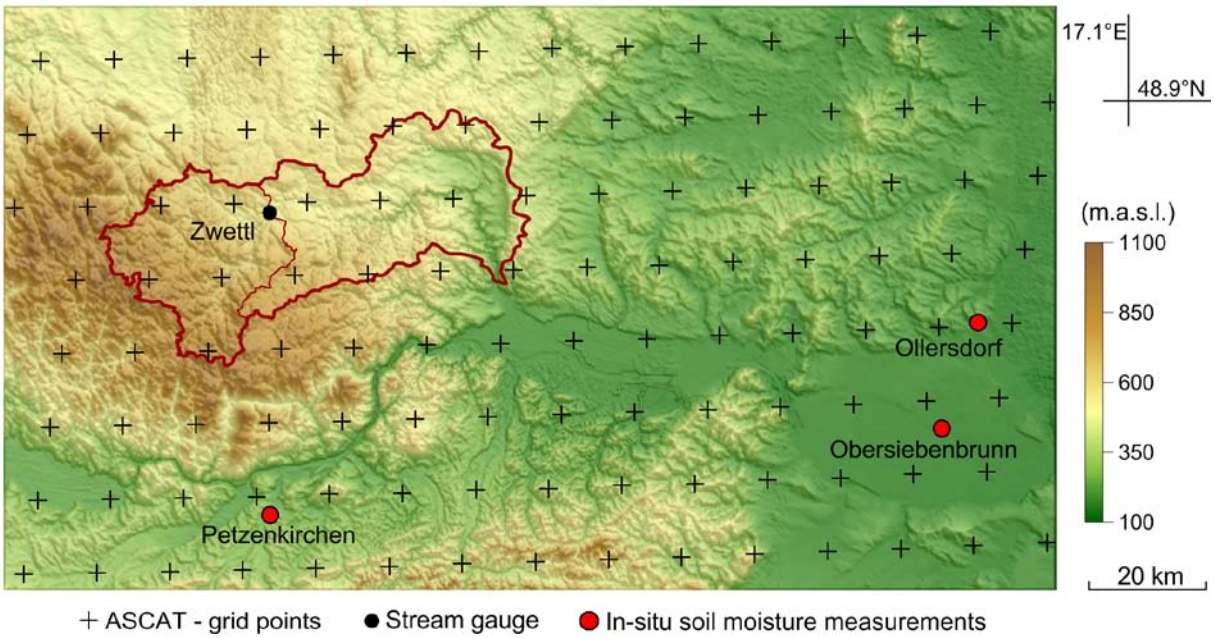

Figure 4.1 Location of the study area and the field sites (red circles) equipped with TDR sensors at different soil depths. The thick red line represents the catchment boundaries of the Kamp catchment (1550 km²). The Zwettl subcatchment (gauged; 622 km²) is indicated by a thin red line. The crosses represent the grid points of the ASCAT data.

The Obersiebenbrunn site is located in Lower Austria 25 km east of Vienna (Figure 4.1). The soil in the upper horizon (0-30 cm) is loamy sand and the land use is grassland (Figure 4.2). The TDR sensors (TRASE-System, SOIL MOISTURE EQUIPMENT, USA) were horizontally installed from a pit at different soil depths from 10 cm to 160 cm. The data were recorded at hourly intervals by a data logger at the site. Rainfall and air temperature data used in the modelling were recorded at the site with a temporal resolution of 10 minutes. An example of the data is shown in Figure 4.3. The decreasing dynamics of the soil moisture with increasing depth is indicating of the movement of the wetting front through the soil profile. Since the focus of this study is on the comparison with ASCAT data on the records at a depth of 10 cm are used here. Rainfall and air temperature data used in the modelling were recorded at a weather station close to the site.

The Ollersdorf site is located close to Obersiebenbrunn in a reforestation area (Figure 4.1). The soil in the upper horizon (0-20 cm) is loamy silt (Figure 4.2). The TDR sensors (TRASE-System, TDR-IAPAS Devices, Poland) were horizontally installed from a pit at different soil depths from 10 cm to 240 cm. The data were recorded at hourly intervals by a data logger at the site. Rainfall and air temperature

data used in the modelling were recorded at the site with a temporal resolution of 10 minutes.

Table 4.1 Characteristics of the field site. Meteorological data relate to the period 2002-2011.

	Kamp catchment	Petzenkirchen	Obersieben-brunn	Ollersdorf
Elevation (m a.s.l.)	300 -1000	258	151	154
Soils (general)	Orthic Podzol Dystric cambisol	Calcaric Fluvisol Orthic Luvisol	Calcic Chernozern	Calcic Chernozern
Soils (upper layer)	sandy	sandy loam	loamy sand	loamy silt
Landuse	Forest, crop-land, grassland	Grassland	Grassland	Reforestation area
Mean annual precipitation (mm)	900 / 600	1050	560	560
Average winter temperature (Dec-Feb) (°C)	-2.1 / -0.7	-0.3	0.3	0.3
Average duration of snow cover period (days)	66 / 37	48	27	27

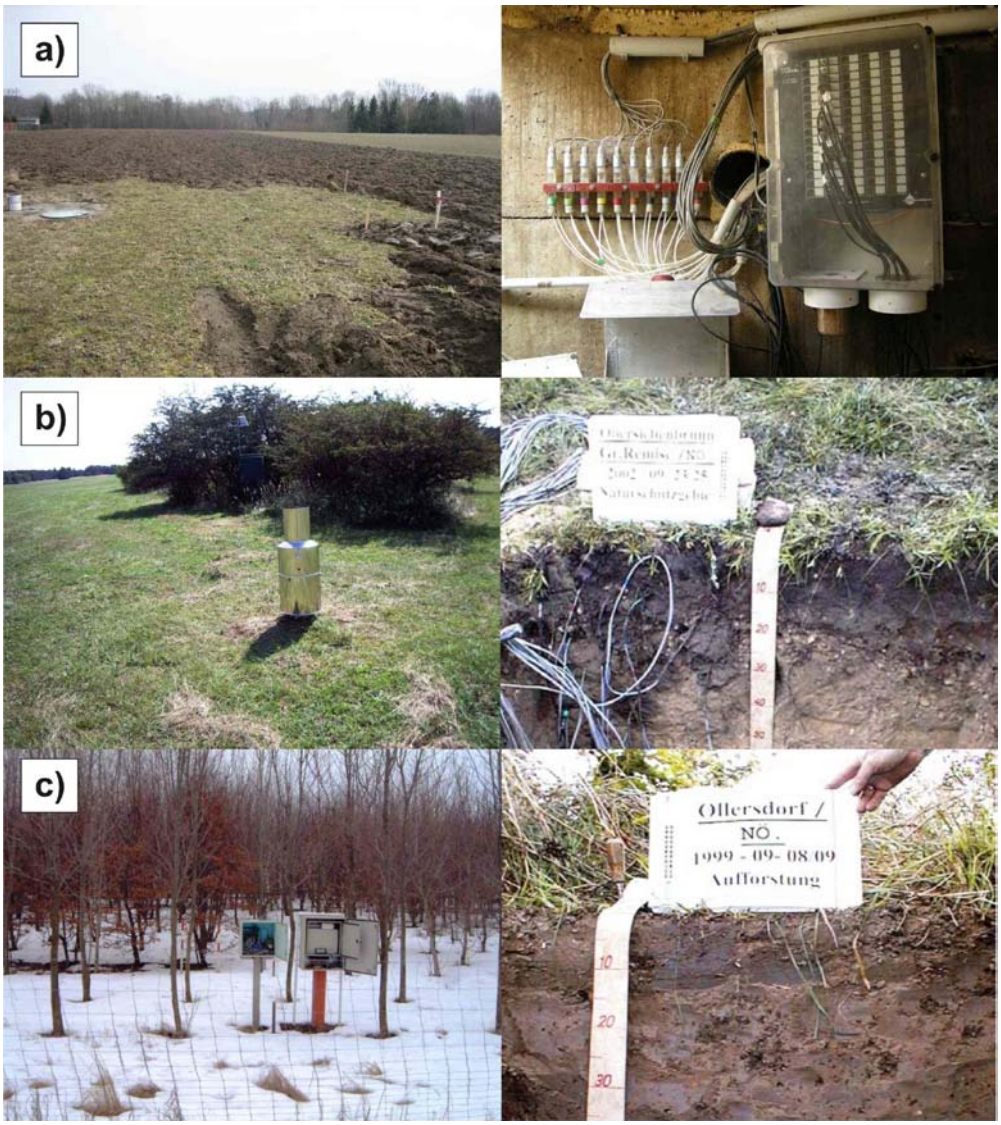

Figure 4.2 TDR-Field measurement sites in Petzenkirchen (a), Obersiebenbrunn (b) and Ollersdorf (c). Left panels: Landscape at the field site. Right panels: Sensor electronics (a) and soil profiles (b, c). Photos: © BAW/IKT

For all four sites, space-borne soil moisture data from the Advanced Scatterometer (ASCAT), on board the Meteorological Operational (METOP) satellite series were used (Bartalis et al., 2007). The period of the data is from January 2007 to January 2010. On average, 1.0 measurements per day were available. The ASCAT sampling nodes (12.5 km x 12.5 km) over the study region are indicated by crosses in Figure 4.1. Soil moisture was estimated from the ASCAT data by a detection method (Wagner et al., 1999) that rescales the instantaneous backscatter with the lowest and highest backscatter coefficients observed in the entire period. Under the assumption that within this period each pixel had been completely dry (lowest backscatter coeffi-

cient) and wet (highest backscatter coefficient), the soil moisture so obtained, SSM is equivalent to the degree of saturation in relative units (ranging between 0 and 100%). The penetration depth of ASCAT is about 0.5 cm to 2 cm, depending on soil characteristics and soil moisture. The ASCAT data therefore represent the top 2 cm of the soil. In order to make the dynamics of the ASCAT data more consistent with the root zone soil moisture an attempt is made to estimate root zone soil moisture from the surface values. The main idea is that surface soil moisture tends to fluctuate much more rapidly than root zone soil moisture. Rather than solving the flow equations the dampening effect is represented by a linear, exponential filter in the time domain. The filtered values are termed the Soil Water Index, SWI. The time parameter of the filter was set to 10 days based on a qualitative comparison to the temporal dynamics of in-situ soil moisture measurements in a depth of 10 cm. Parajka et al. (2006) give a discussion of the filter parameter.

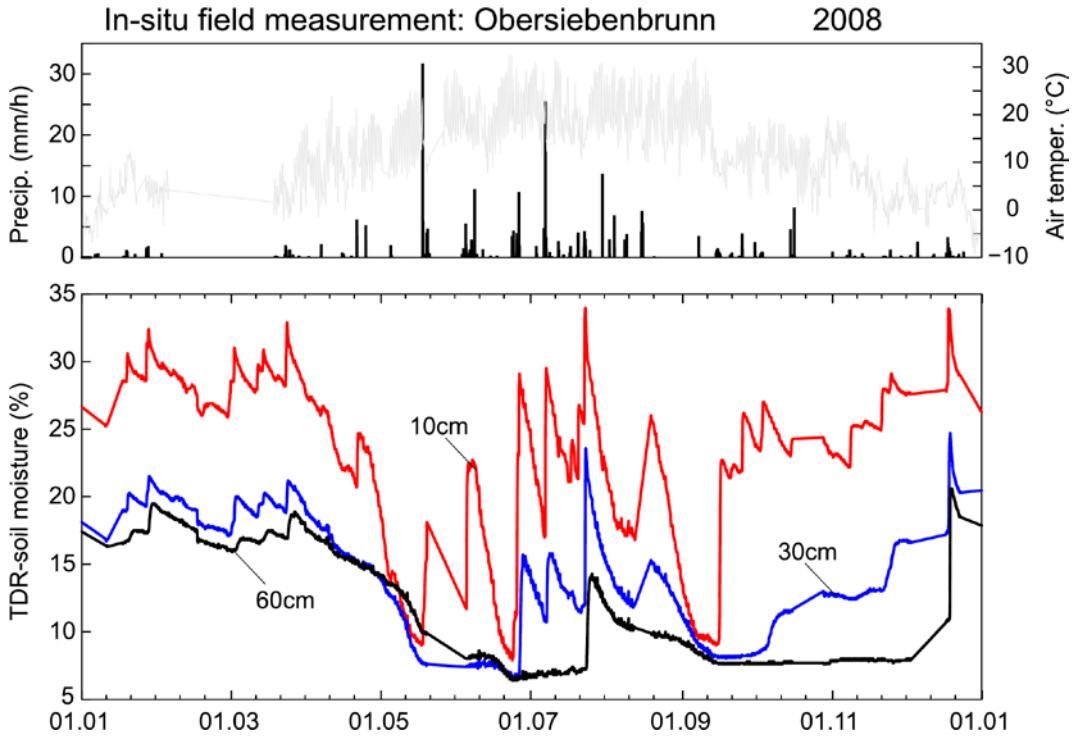

Figure 4.3 Soil moisture measured at different depths at the Obersiebenbrunn site for the year 2008. Top: precipitation intensities (black impulses) and air temperature (gray line). Bottom Soil moisture at 10, 30 and 60 cm depths.

4.3 Hydrologic model

The model used in this paper is a spatially-distributed continuous rainfall-runoff model based on a dual layer soil moisture accounting scheme (Parajka et al., 2009). The model runs on an hourly time step and combines a snow routine, a dual layer soil moisture routine and a flow routing routine. The snow routine represents snow accumulation and melt by a degree-day concept. The soil moisture routine represents runoff generation and changes in the soil moisture states, and the interaction between the skin soil layer and the main soil layer. The main soil layer is represented by three parameters: the maximum soil moisture storage of the main layer L_S, a parameter representing the soil moisture state above which evaporation is at its potential rate, termed the limit for potential evaporation L_P, a parameter in the non-linear function relating runoff generation to the soil moisture state, termed the non-linearity parameter $ß$. The skin layer is represented by the maximum soil moisture storage of the skin soil reservoir L_{Skin}, a parameter subdividing the evaporation from the main layer and the skin soil layer termed φ, and the flux gradient α_s that describes the exchange between the skin and the main soil layer. Runoff routing on the hillslopes and the channels is also represented by reservoirs.

For the Kamp catchment these processes modelled on a 1 km x 1 km grid. The model parameters of the main layer and the routing were identified by a five step procedure using field data, comprehensive hydrographic data as well as qualitative evidence during floods (Blöschl et al., 2009). The parameters of the skin layer were selected on the basis of literature values. Following Houser et al. (2000) and Ottle and Vidal-Madjar (1994) the storage capacity of the skin soil layer is only a small fraction of the main storage of the order of 5-10%. With a typical storage capacity of the main layer of 70 mm a storage capacity of the skin layer L_{skin} of 5 mm was chosen. This is also consistent with the penetration depth of the microwaves (Wagner et al., 1999). During the summer month evaporation rates up to 4 mm per day are possible in the study area. From field experience it is known that, typically, the soil surface turns from saturated to dry conditions within less then two days. This is consistent with an evaporation subdivision parameter of φ= 15% indicating that 15% of the soil water directly evaporated from the root zone layer. To keep the numbers of parameters small φ was assumed to be uniform within the catchment although some

dependence on vegetation would be expected. The vertical exchange flux between the skin and the main soil layer is limited by a maximum flux gradient of α_m =10 mm per day based on the maximal rates for capillary rise and infiltration used by Houser et al. (2000) and Ottle and Vidal-Madjar (1994) respectively. As the internal moisture flux is driven by the moisture difference in the two soil layers, the maximum flux rate is not very likely to occur.

For the Petzenkirchen, Obersiebenbrunn and Ollersdorf sites the same skin layer parameters as in the Kamp were used. The parameter of the main layers were also transferred from the Kamp from model units that were similar to the three sites in terms of soils and vegetation. This parameter transfers were made in an attempt to make the comparison representative of the regional case were no local soil characteristics are available but data from regional sources need to be used such as in the study of Anguela et al. (2008). This makes the comparison more generally applicable for large areas were soil characteristics need to be inferred from regionally available data such as soil texture. No routing was used for the three sites.

Figure 4.4 shows simulations of the Kamp during summer 2009 which is part of the validation period of the model. The runoff simulations closely match the observations in terms of the temporal structure and the magnitude of the flood peak during the event in June. However, the peak in early August is missed which is due to a local, convective precipitation event which was not captured in the rainfall data, even though radar data were used in to spatially distribute the raingauge data. The temporal evolution of the simulated soil moisture highlights the much faster dynamics of the skin layer. While the remotely sensed ASCAT surface soil moisture is not an exact match to the simulations that dynamics are surprisingly similar to those modelled. Estimates show a similar temporal behaviour like the simulated skin soil moisture. Both simulations and ASCAT indicate saturation of the skin layer after heavy rainfall events and a dry down during the following few days between 40 and 60%.

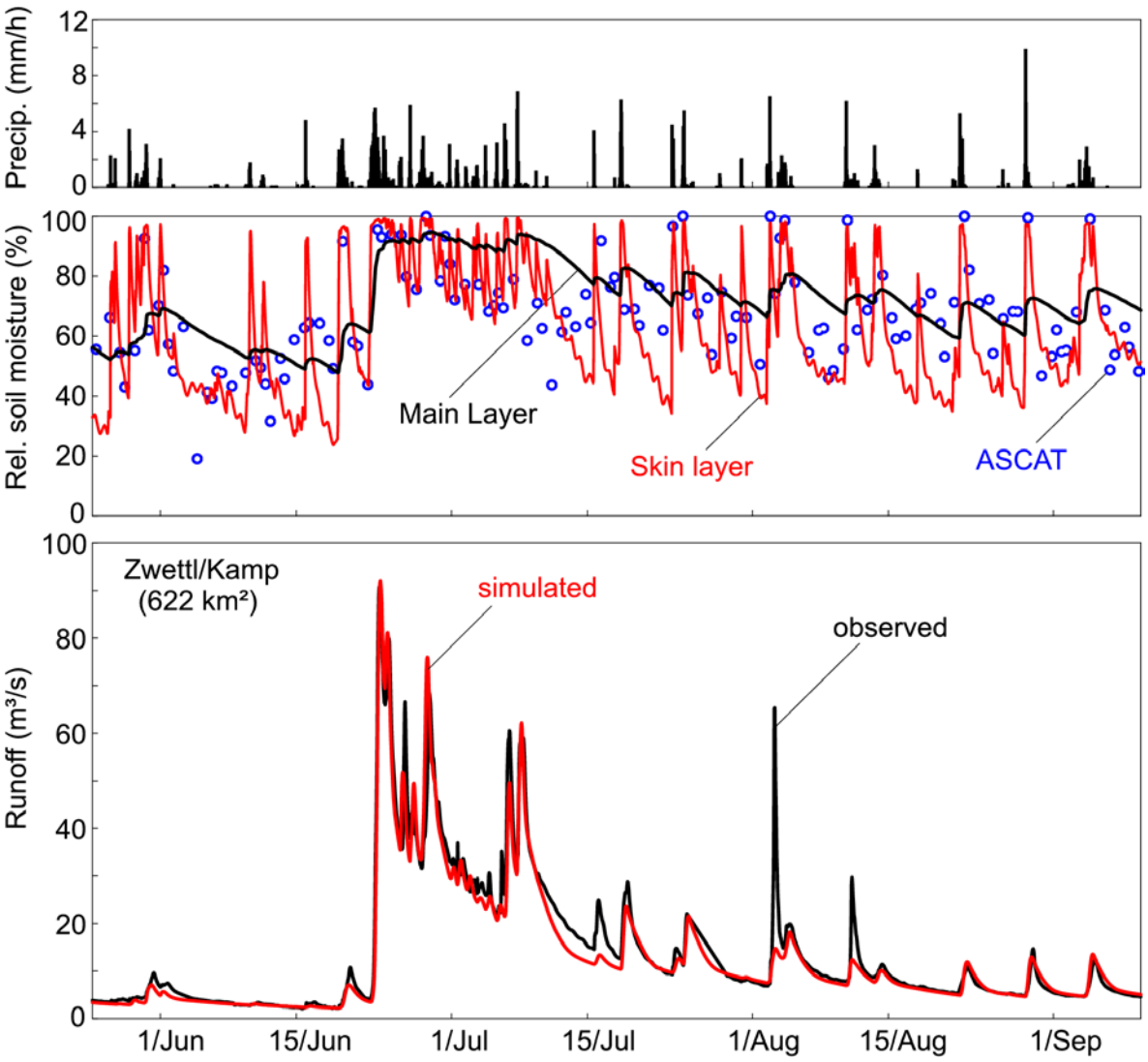

Figure 4.4 Simulations for summer 2009 at the stream gauge Zwettl/Kamp (622 km²). Top panel: catchment average precipitation. Centre panel: simulated mean catchment soil moisture of the skin (red line) and main layer (black line) and mean catchment ASCAT soil moisture (circles). Lower panel: simulated (red line) and observed (black line) runoff hydrographs.

Figure 4.5 shows the spatial soil moisture patterns of ASCAT and the simulations for the same flood event on 23 June 2009. On June 22 in the morning (left panels in Fig. 5) there was hardly any antecedent precipitation. The simulations of skin soil moisture and the remotely sensed (ASCAT) surface soil moisture give very similar patterns. Soil moisture varies from 25% in the East to 55% in the West. The East has lower catchment elevations and therefore higher evaporation. Also, the soils tend to be coarser than in the west. 34 hours later, on June 23 in the evening most of the event precipitation had already fallen. There is a rainfall gradient from west to east. Again, simulations and ASCAT soil moisture give vary similar patterns. While on

June 22 the patterns are controlled by evaporation, on June 23 the patterns are controlled by the spatial rainfall distribution. Also the effect of the coarse soils in the west is apparent in the simulations as lower soil moisture which is also reflected, albeit with lower resolution, in the ASCAT data.

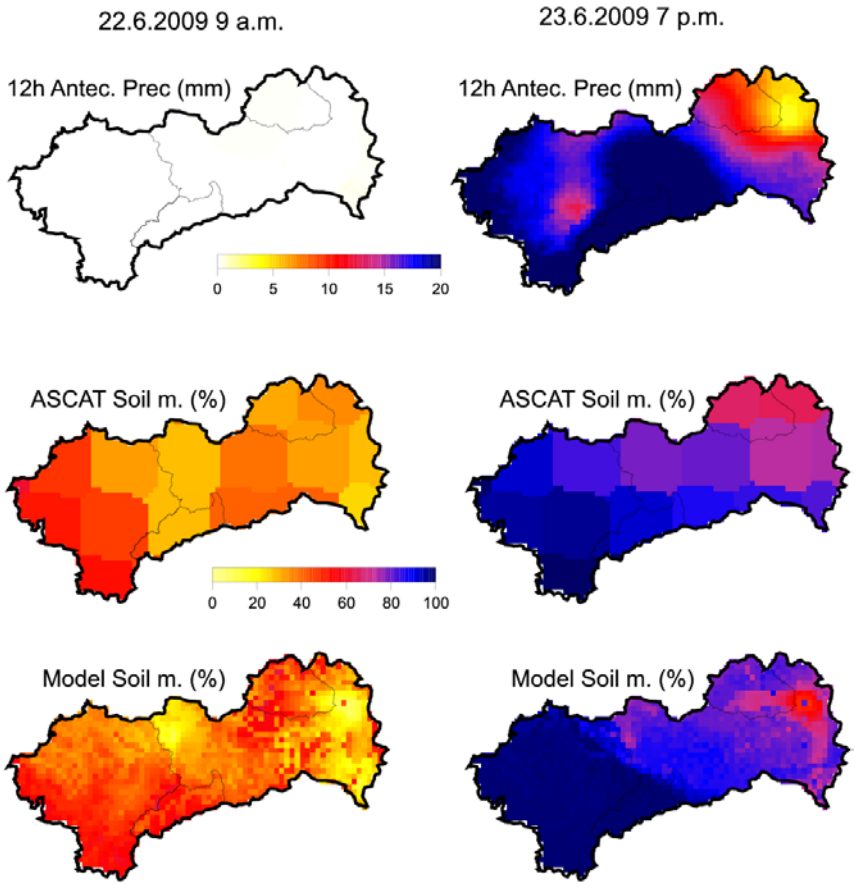

Figure 4.5 Top: Antecedent rainfall during the past 12 hours, Centre: remotely sensed ASCAT surface soil moisture, Bottom: Simulated skin soil moisture in the Kamp catchment (1550 km²). Left: 22 June 2009 immediately before a 5 year flood. Right: 23 June 2009 during the flood.

4.4 Results

ASCAT Satellite soil moisture vs. model simulations

An important issue of space-borne soil moisture estimation is how to deal with periods of snow and frozen ground as microwaves can only detect liquid water while for hydrological purposes one is interested in the total water stored in the soil. The standard approach is to mask the satellite scenes that are affected by snow and frozen

ground based on various criteria. Usually a static masking method is chose that depends on the location and the day of the year (Bartalis, 2007) but this is unable to account for specific hydrometeorological situations that can be critically important for snow and frozen ground is present. In this paper a number of dynamic masking criteria are examined (Table 4.2). In a first masking scenario, the ASCAT pixels are masked if 50% of the pixel area is covered by snow. To identify snow cover MODIS satellite data are used. MODIS data from Aqua and Terra are combined to minimise the shading effects due to clouds based on the method of Parajka and Blöschl (2008). In a second masking scenario, ASCAT pixels are additionally masked if the average snow water equivalent in the pixel exceeds 2 mm. Estimates of snow water equivalent are obtained from the hydrological simulations. In a third masking scenario, ASCAT pixels are additionally masked if the average air temperature over the past 12 hours exceeded 0 °C to reflect the effects of frozen ground. The fourth masking scenario masks the ASCAT pixels if the air temperatures exceed 2°C in order to assess the effect of changing the threshold temperature. For comparison scenario 0 relates to the ASCAT data without masking.

Table 4.2 Definition of criteria for the masking ASCAT surface soil moisture estimates due to snow and frozen ground. A pixel is masked if at least one of the criteria applies.

Masking scenario	0	1	2	3	4
MODIS snow cover	-	> 50%	> 50%	> 50%	> 50%
Simulated snow water equivalent	-	-	> 2mm	> 2mm	> 2mm
Air temperature (12h average)	-	-	-	< 0 °C	< 2 °C

Each of the masking scenarios are applied to the ASCAT data of the Kamp catchment and the three field sites. Pearson correlation between the remaining ASCAT pixels and the simulated skin are calculated separately for the summer and winter months. The biases of ASCAT relative to the simulations are also analysed. For the Kamp catchment the correlations and biases have been calculated from catchment average soil moisture. The correlations (Table 4.3) in the summer months range from 0.64 (Ollersdorf) to 0.71 (Petzenkirchen) which is a similar order of magnitude as other studies (e.g., Albergel et al., 2010). As would be expected the correlations do

not change with the scenario as the masking only applies do cold conditions. The correlations in the winter months strongly depend on the masking method. Without masking the correlations range from -0.12 (Kamp) to 0.43 (Obersiebenbrunn). These differences are closely related to the different climates of the two locations. The mean annual winter temperatures of the Kamp are -1.3 °C while they are +0.3 °C at Obersiebenbrunn (Table 4.1). Also there are major differences between the two sites in terms of snow cover durations. Masking scenario 1 based on snow cover significantly increases the winter correlations for Kamp, but it does not change the correlations for the other sites much. Because Kamp is the locations with the longest snow durations one would expect the effect of this masking scenario to be largest. If simulated snow water equivalent is also considered (masking scenario 2) the correlations increase and range from 0.41 (Kamp) to =0.56 (Petzenkirchen and Ollersdorf). A further substantial increase of correlations is achieved for all study sites, if air temperature is used to mask frozen soils (masking scenarios 3 and 4). For the Kamp selection of a threshold air temperature from 2°C instead of 0°C increases the winter correlations from 0.49 to 0.71. Clearly, there are long periods during winter with temperatures around freezing where the ground is still frozen, so the increase in the threshold temperature substantially improves the correlations.

A similar effect of the masking scenarios can be observed for the biases, i.e. the differences between ASCAT soil moisture and the simulated skin soil moisture. While the biases are always small for the summer, substantial biases occur during winter if no masking is applied. The biases range from 11% (Obersiebenbrunn, Ollersdorf) to 34% (Kamp). Clearly, the differences in the climates have a major affect on the ASCAT soil moisture interpretation as the ASCAT readings appear too low due to snow and frozen ground. As the various masking variants are applied, the biases decrease to between 1% (Ollersdorf) and 3% (Kamp). The exception is Petzenkirchen where significant biases remain for masking scenario 4. Maybe this is related to the seasonal tendency of very low surface soil moisture estimates during the winter season at this specific ASCAT grid element.

Table 4.3 Correlation coefficients between monthly mean ASCAT surface soil moisture and simulated skin layer soil moisture at the four field sites for the five masking sce-narios of Table 4.1. First number relates to summer (April to September), second number relates to winter (October to March).

	Masking 0	Masking 1	Masking 2	Masking 3	Masking 4
Kamp catchment	0.68 / -0.12	0.68 / 0.33	0.68 / 0.41	0.68 / 0.49	0.68 / 0.71
Petzenkirchen	0.71 / 0.28	0.71 / 0.31	0.71 / 0.56	0.71 / 0.62	0.71 / 0.69
Obersiebenbrunn	0.68 / 0.43	0.68 / 0.45	0.68 / 0.55	0.68 / 0.69	0.68 / 0.72
Ollersdorf	0.64 / 0.38	0.64 / 0.40	0.64 / 0.56	0.64 / 0.69	0.64 / 0.72

Table 4.4 Bias (%) of monthly mean ASCAT surface soil moisture relative to simulated skin layer soil moisture at the four field sites and for the five masking scenarios of Table 4.1. First number relates to summer (April to September), second number relates to winter (October to March).

	Masking 0	Masking 1	Masking 2	Masking 3	Masking 4
Kamp catchment	-1 / 34	-1 / 13	-1 / 9	-1 / 6	-1 / 3
Petzenkirchen	9 / 28	9 / 26	9 / 20	9 / 17	9 / 14
Obersiebenbrunn	-2 / 11	-2 / 10	-2 / 7	-2 / 3	-2 / 2
Ollersdorf	1 / 11	1 / 10	1 / 6	1 / 3	1 / 1

As would be expected the highest correlations and smallest biases are obtained if the remotely sensed soil moisture data set is masked based on snow cover, simulated snow water equivalents and the mean air temperature with a threshold of 2°C (masking scenario 4). However, this is also the scenario that removes the largest number of pixels. To understand the tradeoff between the strictness of the masking method and the accuracy of the ASCAT soil moisture the number of ASCAT Level 2 images was analyses for the four sites. Without masking there are a total of about 450 ASCAT images in the Kamp and at Petzenkirchen in the four winters which is equivalent to 0.83 ASCAT images per day. For the Kamp, occasionally, only part of the catchment was covered, so the number of scenes was calculated as the catchment aver-

age. Figure 4.6 shows that, as the masking get stricter, the correlations of ASCAT soil moisture and simulations increase but the number of scenes decreases. For the strictest masking (scenario 4) the number of scenes ranges between 95 (Kamp) and 280 (Ollersdof) which is equivalent to 0.18 and 0.52 ASCAT images per day, respectively. If the ASCAT data are to be assimilated into a hydrological model one would probably use the strictest masking, particularly in the Kamp. At the other three sites, however, there seems to be little extra benefit of using a 2°C threshold (scenario 4) over the 0°C threshold (scenario 3).

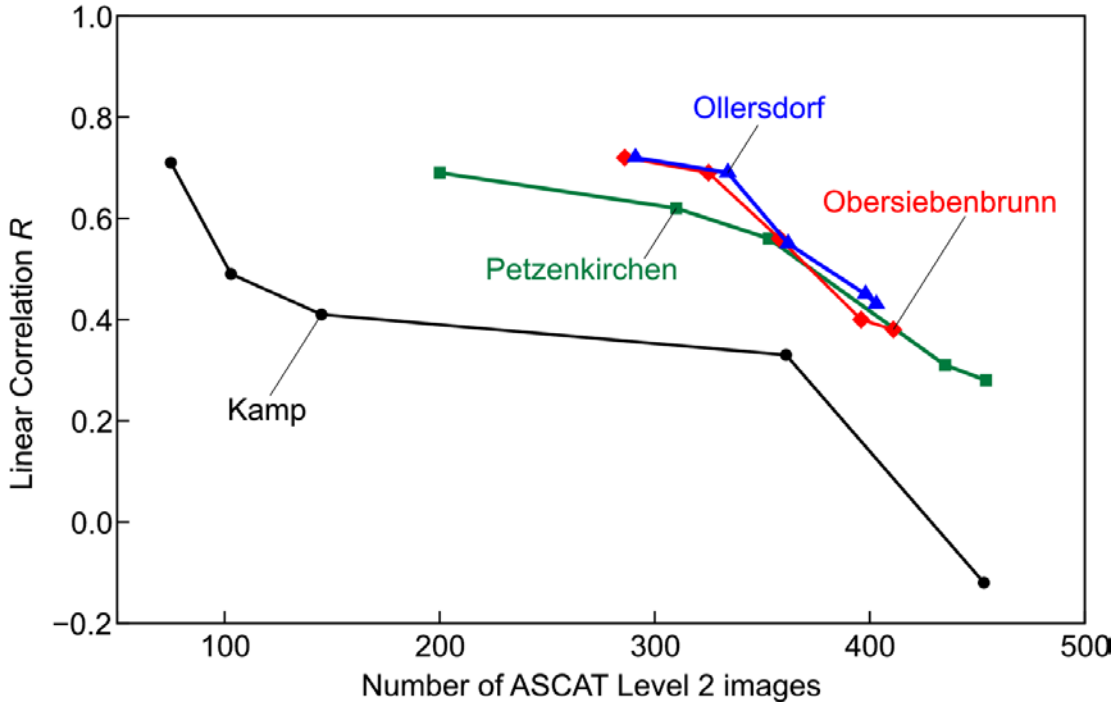

Figure 4.6 Trade-off between accuracy of winter (October to March) surface soil moisture of ASCAT and number of scenes available. Each point relates to one filter scenario (Table 2). As the masking get stricter, the correlations of ASCAT soil moisture and simulations increase but the number of scenes decreases

To illustrate the seasonal fluctuations of the correlations and biases Figure 4.7 shows the monthly values for Obersiebenbrunn as an example. The inconsistency of the correlations and biases mainly occurs in December, January and February which are the months with frequent frost and snow. For the Kamp (not shown here), the period of inconsistency is longer and ranges from November do March (with particularly low values from January to March) as would be expected from the longer snow cover period at the Kamp.

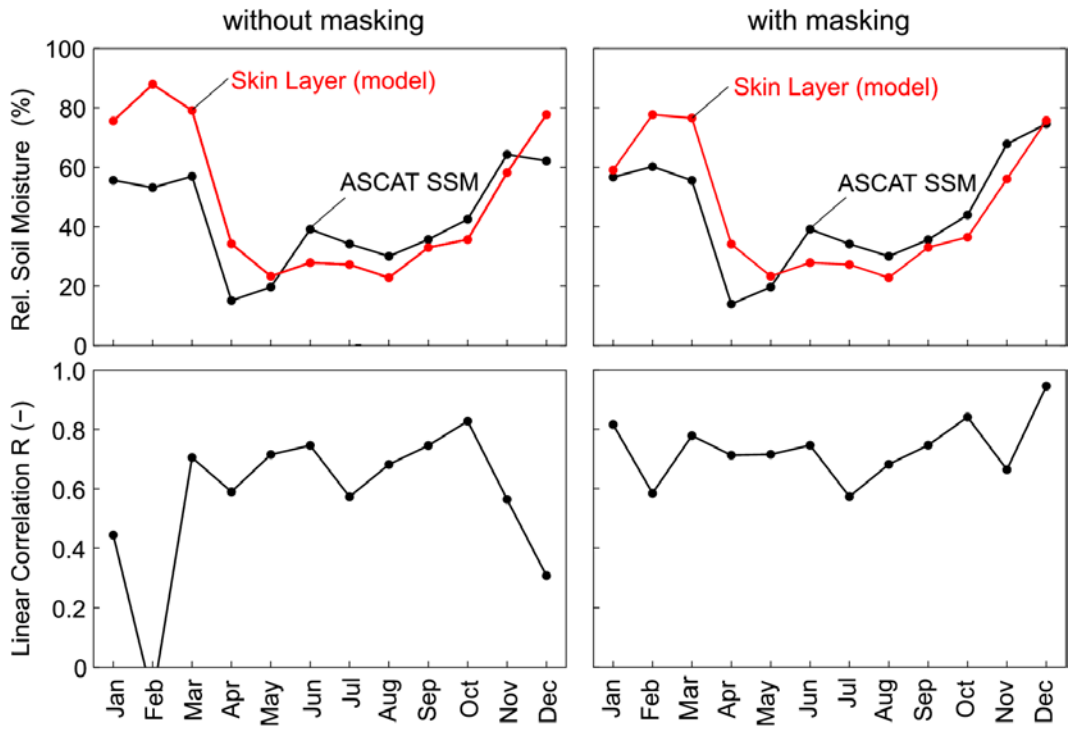

Figure 4.7 Seasonality of simulated skin soil moisture (red line) and ASCAT surface soil moisture (black line, upper panels) and correlation coefficients (lower panels) at Obersiebenbrunn. Left: without masking. Right: with masking (scenario 4, as of table 2)

For the case of the Kamp, spatially distributed soil moisture simulations are available, so it is of interest to examine the spatial patterns of the correlations and biases. For the case without masking these are shown in Figure 4.8. During the summer months (April to October) the correlations coefficients are around 0.7. In contrast from January to March parts of the catchment have negative correlations. This is particularly the case for the western part of the Kamp catchment which has higher elevations and therefore more snow and more extended periods of frozen soil can be expected. In the eastern (lower) parts of the catchment the correlations remain positive. The better correlations in the lower part of the catchment are also apparent in the correlations evaluated over the entire years annual (bottom right panel). If the masking is applied (Figure 4.9) the correlations improve in all winter months. Interestingly, the correlations in the eastern parts of catchment are particularly high during the winter months (eg. in February, November), if not masked. However, the effect on the number of scenes is also apparent in February and December as, for some pixels, there were fewer than 3 scenes, so no correlation coefficients were calculated (white areas in Figure 4.9).

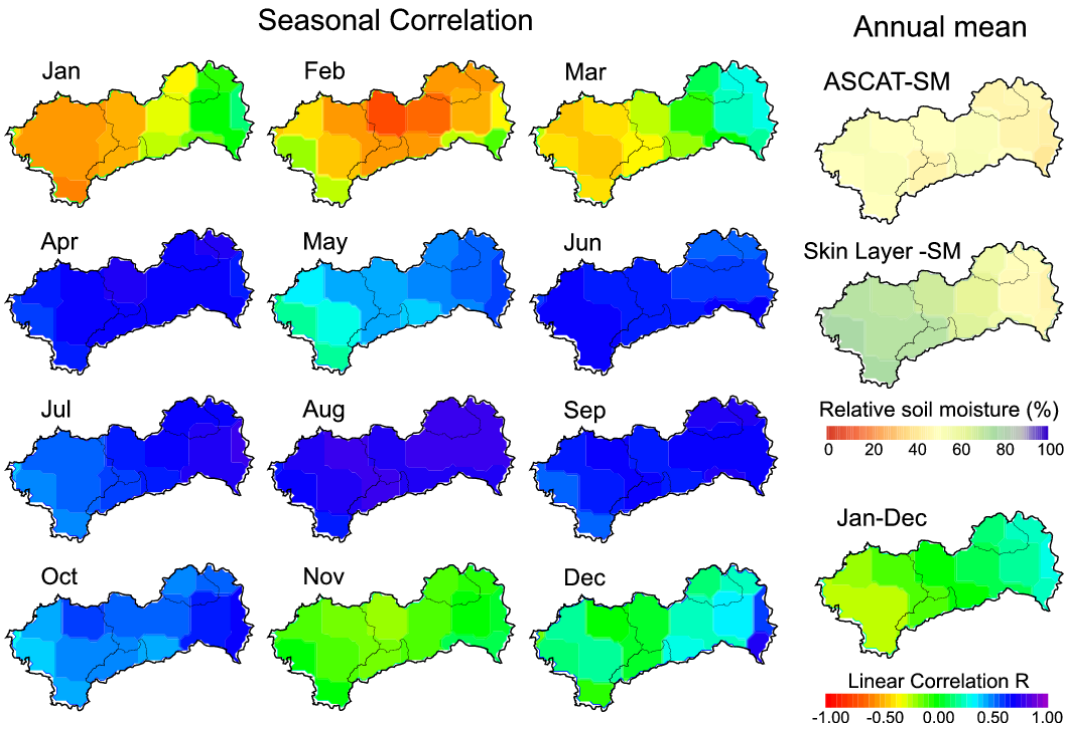

Figure 4.8 Monthly coefficients R for the Kamp catchment for the period 2007 to 2010 without masking (12 panels). The top right panel shows the mean annual ASCAT surface soil moisture, the panel below the simulated skin soil moisture. The bottom right panel shows the correlations over the entire period

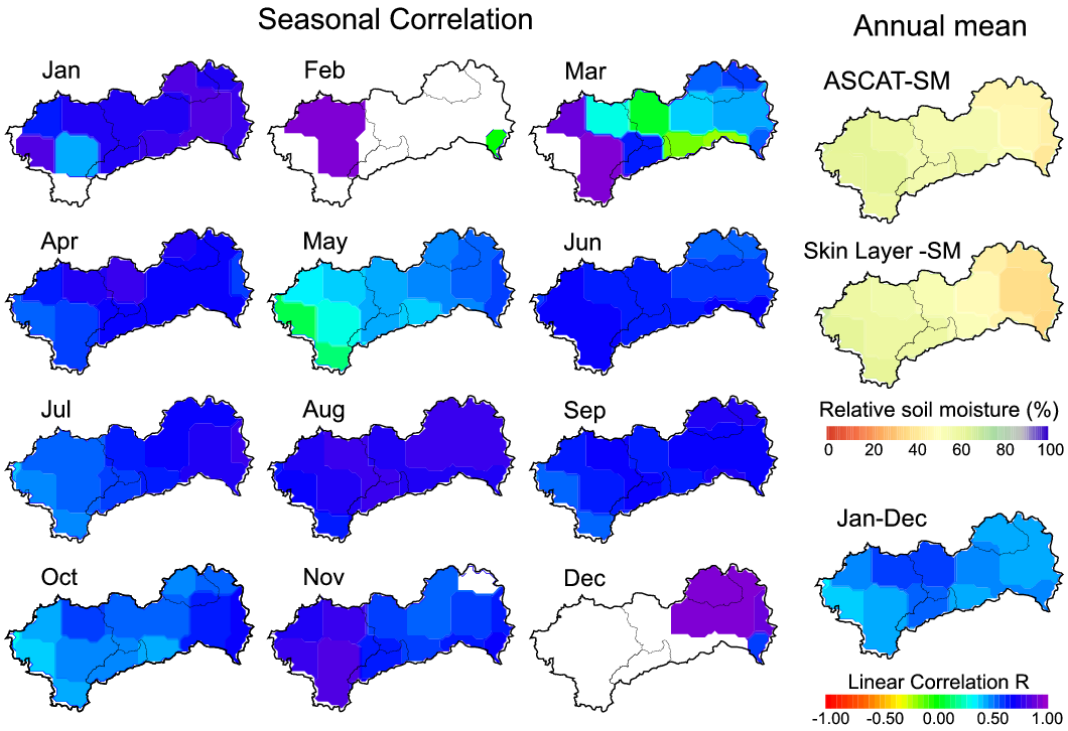

Figure 4.9 Same as Figure 4.8 but with masking scenario 4 (Table 2).

ASCAT Satellite soil moisture vs. in-situ data

In a comparison of space-borne and in-situ soil moisture measurements one would expect particularly large differences resulting from scale inconsistencies due to vastly support volumes (footprints). To account for limited penetration depths, in addition to ASCAT surface moisture SSM the soil water index SWI derived from the ASCAT data is compared to the in-situ measurements of the TDR-sensors in a soil depth of 10 cm. However, a scale inconsistency of the spatial footprint of about 10 orders of magnitude (10^8 m² for ASCAT as opposed to 10^{-2} m² for the TDR) remains.

During the summer months the correlations of ASCAT surface soil moisture with the TDR at 10 cm depth ranges around 0.4 for all stations (Table 4.5). During the winter months the correlations range from 0.22 to 0.44. As different masking methods are applied there is a slight increase in the correlations to a range of 0.27 (Ollersdorf) to 0.51 (Petzenkirchen). It is interesting that the winter correlations are higher than the summer correlations which is likely due to the spatially and temporally more consistent rainfall regime in winter than in summer. However, it is clear that correlations compares soil moisture at very different scales, both horizontally and vertically. The SWI assists in accounting for a deeper penetration depth, so should be more comparable with the TDR measurements. Indeed, the correlations during the summer months range between 0.52 and 0.67 which is almost twice the correlations coefficient of surface soil moisture. In winter the correlations are lower and here the masking makes a difference. Without masking the correlations range between 0.33 and 0.51 while with masking (masking scenario 4) the correlations range from 0.45 to 0.61. However, the spatial scale inconsistency remains.

The biases (Table 4.6) in summer range from –8 to – 15 (i.e. ASCAT tends to give lower values then the TDR). Maybe this is due to fact that the in-situ measurements indicate wetter soil conditions than the ASCAT surface soil moisture during hot and dry periods in summer, when the surface is drier than the deeper soil layers. The biases winter range between -14 and -19 without masking and are slightly reduced for if masking is applied (range of -8 to -18 for masking scenario 4). There is very little difference between ASCAT SWI and ASCAT surface soil moisture since the SWI is a linear aggregation so should not affect biases. The differences that do occur are probably due to smaller effect of extreme dry soil moisture estimates after applying

the filter. For all of the three field sites the masking scenario 4 performs best but, again, there is a trade-off between the ASCAT accuracy and the number of images.

Table 4.5 Correlation coefficients between monthly mean ASCAT surface soil moisture and in-situ soil moisture at the three field sites and for the five masking scenarios of Table 4.1. First number relates to summer (April to September), second number relates to winter (October to March). SSM relates to surface soil moisture from ASCAT, SWI is the soil water index obtained by filtering SSM with a time parameter of T = 10 days

	Masking 0	Masking 1	Masking 2	Masking 3	Masking 4
Petzenkirchen - SSM	0.44 / 0.38	0.44 / 0.39	0.44 / 0.42	0.44 / 0.49	0.44 / 0.51
Obersiebenbrunn - SSM	0.39 / 0.44	0.39 / 0.45	0.39 / 0.48	0.39 / 0.55	0.39 / 0.58
Ollersdorf - SSM	0.31 / 0.22	0.31 / 0.21	0.31 / 0.24	0.31 / 0.26	0.31 / 0.27
Petzenkirchen - SWI	0.62 / 0.43	0.62 / 0.46	0.62 / 0.51	0.62 / 0.56	0.62 / 0.61
Obersiebenbrunn - SWI	0.67 / 0.51	0.67 / 0.52	0.67 / 0.57	0.67 / 0.59	0.67 / 0.61
Ollersdorf - SWI	0.52 / 0.33	0.52 / 0.33	0.52 / 0.36	0.52 / 0.42	0.52 / 0.45

Table 4.6 Bias (%) of monthly mean ASCAT surface soil moisture (SSM) and ASCAT Soil Water Index (SWI) relative to in-situ soil moisture at the three field sites and for the five masking scenarios of Table 4.1. First number relates to summer (April to September), second number relates to winter (October to March). SSM relates to surface soil moisture from ASCAT, SWI is the soil water index obtained by filtering SSM with a time parameter of T = 10 days

	Masking 0	Masking 1	Masking 2	Masking 3	Masking 4
Petzenkirchen - SSM	-9 / -19	-9 / -19	-9 / -17	-9 / -15	-9 / -13
Obersiebenbrunn - SSM	-12 / -18	-12 / -16	-12 / -14	-12 / -13	-12 / -13
Ollersdorf - SSM	-15 / -19	-15 / -19	-15 / -19	-15 / -19	-15 / -18
Petzenkirchen - SWI	-8 / -14	-8 / -13	-8 / -12	-8 / -10	-8 / -9
Obersiebenbrunn - SWI	-7 / -13	-7 / -12	-7 / -12	-7 / -11	-7 / -9
Ollersdorf - SWI	-13 / -16	-13 / -14	-13 / -13	-13 / -12	-13 / -12

To illustrate the seasonal fluctuations of the correlations and biases Figure 4.10 shows the monthly values for Obersiebenbrunn as an example. The seasonal patterns of soil moisture with drier conditions during the summer from May to September and wet conditions during winter are remarkably similar (top panels). The high ASCAT soil moisture estimates for the winter months, even if no masking is applied, maybe related to the minor importance of snow cover and frozen soils at Obersiebenbrunn. The masking slightly increases the ASCAT data in January. For the correlation coefficients between ASCAT surface soil moisture and the in-situ data, no seasonal patterns (lower panels). The correlations of the SWI based on ASCAT and the in-situ data are higher, particularly in summer. The masking produces a moderate increase of the correlations in winter but, again, the scale inconsistencies seem to dominate in all months.

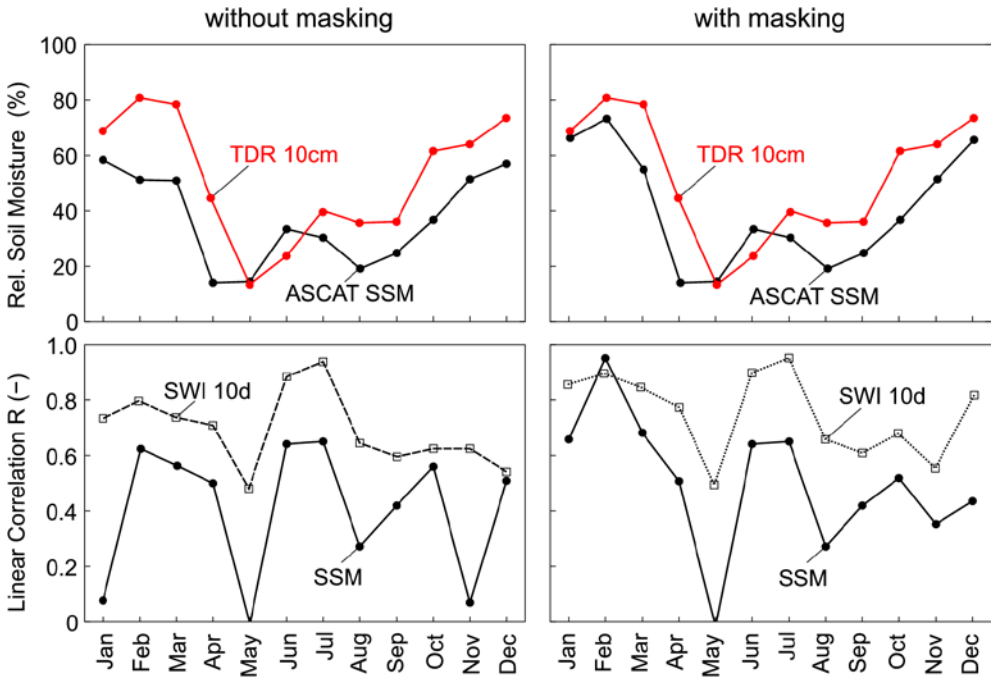

Figure 4.10 Seasonality of TDR measured soil moisture (red line) and ASCAT surface soil moisture SSM (black line, upper panels) and correlation coefficients (lower panels) between TDR and SSM (solid line) and between TDR and SWI (dashed line) at Obersiebenbrunn. Left: without masking. Right: with masking (scenario 4, as of Table 4.2). TDR readings represent relative saturation.

Model simulations vs. in-situ data of soil moisture

To complete the inter-comparison of hydrologic modelling, remote sensing and in-situ measurements, this section compares the simulation results with the TDR measurements. The correlations between the soil moisture simulated in the skin layer and the TDR measurements gives correlations ranging from 0.39 to 0.65 (summer) and ranging from 0.65 to 0.68 (winter). While one would the skin layer not expect to closely represent the TDR at 10 cm depth it is interesting that the correlations are large in winter at all three sites. This is because of differences in the dynamics in summer and winter. In summer there are frequent short rainstorms producing a highly dynamics soil moisture response while in winter there are fewer and longer storms. For the more dampened dynamics the scale mismatch is less important, so the correlations are higher. The correlations for the main layer – which is more consistent with the TDR measurements – range from 0.71 to 0.86 (summer) and ranging from 0.69 to 0.79 (winter). Clearly, if the scales are consistent one can expect much more consistent soil moisture dynamics of the model and the TDR measurements.

At all sites negative biases indicate the tendency of the model to underestimate the in-situ measurements of soil moisture. When using the main layer the biases are smaller than when using the skin layer, once again pointing to the role of the scale mismatch.

Table 4.7 Correlation coefficients and biases (%) between monthly mean simulated soil moisture (skin and main layer) and in-situ soil moisture at the three field sites. First number relates to summer (April to September), second number relates to winter (October to March).

	Correlation	Bias
Petzenkirchen - Skin Layer	0.65 / 0.68	-9 / -4
Obersiebenbrunn - Skin Layer	0.50 / 0.65	-8 / -1
Ollersdorf - Skin Layer	0.39 / 0.65	-18 / -13
Petzenkirchen - Main Layer	0.82 / 0.69	-6 / -3
Obersiebenbrunn - Main Layer	0.86 / 0.79	-2 / 4
Ollersdorf - Main Layer	0.71 / 0.72	-13 / -11

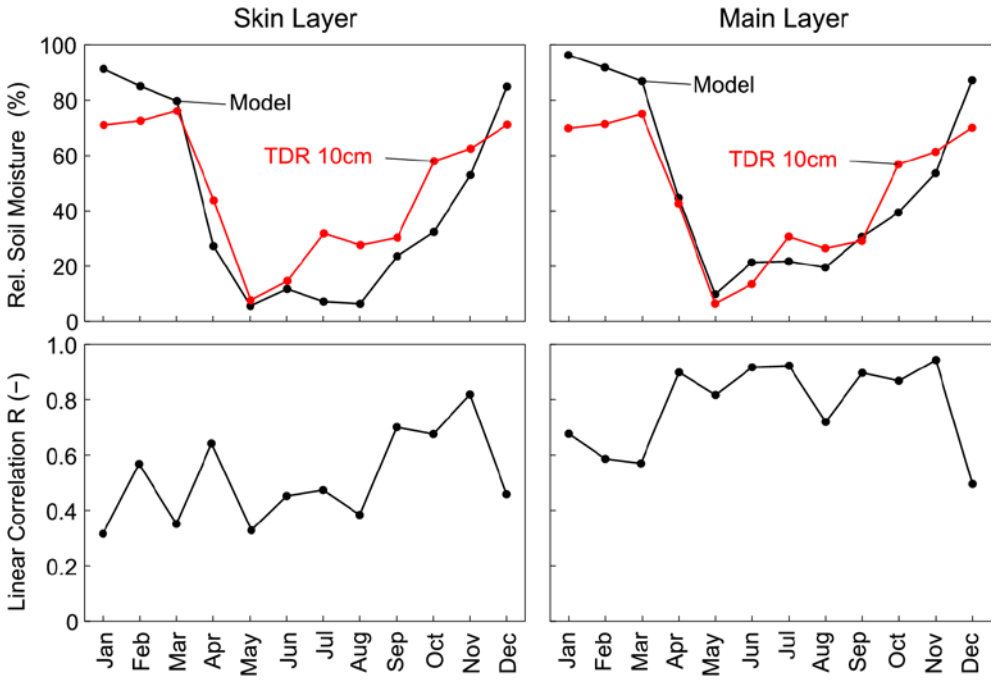

Figure 4.11 Seasonality of TDR measured soil moisture (red line) and simulated soil moisture (black line, upper panels) and correlation coefficients (lower panels) at Obersiebenbrunn. Left: comparison with the skin layer simulations. Right: Comparison with the main layer simulations. TDR readings represent relative saturation.

The monthly correlations and biases are shown in more detail in Figure 4.11 for the Obersiebenbrunn site. During December, January and February the model gives near saturation conditions while the TDR gives somewhat smaller relative soil moistures of around 80%. From July to September the skin model underestimates soil moisture relative to the TDR while the simulated soil moisture of the main layer is very similar to the TDR. The correlations of skin layer soil moisture and TDR measurements vary between 0.3 and 0.8 with the largest values from September to November. As indicated above this is a period with fewer convective storms than in summer. The comparison of the simulated main layer soil moisture and the TDR-soil moisture measurements show a very good agreement, particularly from April to November. During this period the correlation coefficients are around 0.9 or larger, with the exception of August with a correlation of 0.7. Somewhat lower correlations during the winter period from December to March are consistently large soil moisture values with less temporal variability so any differences are more apparent. This is illustrated in Figure 4.12 for the year 2009. While the fluctuations of the simulated skin layer soil moisture are much faster than those indicated by the TDR-measurement, the tempo-

ral dynamic of the simulated main layer soil moisture shows excellent agreement with the in-situ measurement in terms of the slope of the fluctuations. The different temporal dynamics are very pronounced during the warm period May to September. The high temporal variability is induced by intensive rainfalls and high evaporation rates during the summer. The agreement in terms of the dynamics indicates that the parameters of the soil moisture model indeed describe snow melt processes and the drying of soils in spring in a realistic way. The other two field sites show similar patterns with a similar consistency of the model and the TDR measurements.

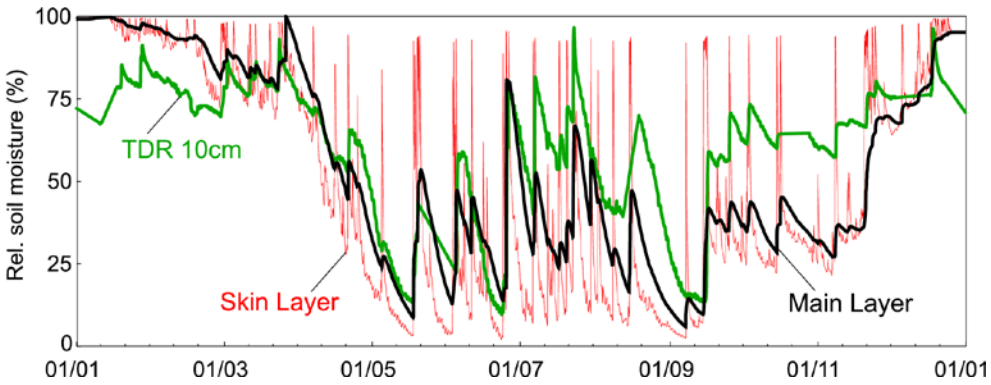

Figure 4.12 Temporal evolution of the simulated skin (red line) and main layer (black line) soil moisture and TDR measurements at 10 cm depth (green line) at Obersiebenbrunn for the year 2009.

4.5 Discussion and Conclusions

Soil moisture plays an important role in hydrologic forecasting models. As soil moisutre is difficult to measure across the spatial scales, combining available informaiton from different sources seems to be an obvious strategy. The inter-comparison of remotely sensed soil moisture products with in situ measurements and simulated soil moisture to assess the potential of satellite based surface soil moisture estimates to reduce uncertaintys in hydrologic applications has recently been an important issue in the hydrologic and remote sensing sciences.

The comparison of the spatial surface soil moisture patterns from remote sensing (ASCAT SSM) and hydrologic modelling (Skin layer SM) indicated very good consistency for the Kamp catchment (1550 km²) in northern Austria. It is difficult to relate these results directly to the findings of recent studies. Most studies mainly focus on evaluations of different soil moisture estimates at the plot scale (in-situ measurement

networks) or for lumped hydrologic models (Albergel et al., 2010; Brocca et al., 2010; Matgen et al., 2011). For illustration purposes Anguela et al. (2008) plot soil moisture patterns from the SIM model and ERS-Scatterometer for a catchment in France for three different days (winter, summer and autumn). However, no specific statements about the accordance of the spatial patterns are given. The results at the Kamp catchment are very promising, as both data soil moisture estimates predict wet soils in the higher western part of the catchment (figure 4.9). The realistic representation of soil moisture patterns is from utmost importance for the simulation of appropriate initial conditions for flood forecasts, as they play an important role in the highly non-linear process of runoff generation. As this study focuses on the assessment of the spatial and temporal dynamics of the surface soil layer, no comparison of the spatial patterns of the derived soil water indices and simulated main layer soil moisture were analysed. However, the positve impact of surface soil moisture products on the prediction of root zone soil moisture is support by the results of several studies (Ottlé and Vidal-Madjar, 1994; Georgakakos and Baumer, 1996; Heathman et al., 2003, Draper et al., 2011).

The evaluation results based on the comparison at the Kamp catchment and three field site locations (plot scale) showed the highest correlations for the comparison of TDR measurements and simulated main layer soil moisture. We have included the simulated main layer soil moisture, because the lower temporal dynamics in a soil depth of 10 cm show a much better consistency with the main layer of the model, representing the root zone (Figure 4.12). The correlations are much better compared to those obtained when using the skin layer soil moisture. This finding is not very surprising and generally supported by the results obtained by Anguela (2010) and Brocca et al. (2010). However, at the Austrian field sites the seasonality of the mismatch in temporal dynamics between simulated skin layer soil moisture and TDR measurements seems to be more pronounced. In the light of previous studies the correlation coefficients and biases for the Austrian field sites between skin and main layer soil moisture and in-situ soil moisture measurements in a soil depth of 10 cm are slightly lower than obtained at in-situ field sites in France and Italy. The reason therefore maybe found in the fact that at the Austrian field sites the model parameters for the parsimnious soil moisture accounting schemes were transferred from

comparable hydrologic response units at the Kamp catchment without any further calibration.

Similar to the results obtained by Aguela et al. (2008), Albergel et al. (2009) and Brocca et al. (2010), the correlations and biases between in-situ measurements in a soil depth of 10 cm and the remotely sensed soil moisture estimates are much better if the ASCAT SWI, instead of the ASCAT SSM product, is used for the comparison at the three field sites in Austria. The correlation coefficients are in the order of magnitude of the correlations obtained at the French in-situ measurement network. The correlations in the Tiber region are higher, particularly for the SWI comparison. A reason therefore maybe can be found in the extensive calibration of the time parmeter T for the Italian field sites. However, the resutls at the Austrian field sites indicate slightly higher correlations for the summer periods from April to September. No distinct seasonality has been found for the Italian and French studies.

Looking at the comparison of ASCAT surface soil moisture and simulated surface soil moisture the results at the Kamp catchment and the three field sites in Austria the correlations and biases are quite good and show consistency with the results found in a previous studies in France (Albergel et al., 2010). The results from the Tiber catchment in central Italy indicate slightly higher correlations (Brocca et al., 2010). While in Austria very distinct seasonality effects are observed, the Italian and French analysis do not indicate seasonal differences of the correlation coefficients. The seasonal differnces are obviously related to the climatic conditions at the different locations, as the greatest seasonal differnces are obtained for the study area with the coldest mean winter temperature. According to Parajka and Blöschl (2008), the remotely sensed measurements are invalid for situations affected by snow cover or frozen soils during the winter season in wide parts of Austria. However, the good results during summer are very promising, as the high temporal soil moisture dynamics during this period, dominated by convective rainfall events and high evaporation rates, seem to be well captured by the ASCAT surface soil moisture product. The ASCAT soil moisture product is potentially valuable to reduce uncertainties of initial soil moisture conditions for flood forecasting, particularly because they also provide valueable information about the spatial distribution of soil moisture.

The results for the winter periods raise the question, how to deal with the influence of snow cover and frozen soils on the remotely sensed soil moisture estimates, as microwaves can only detect liquid water. Usually a static masking method is chosen that depends on the location and the day of the year (Bartalis, 2007), but this is unable to account for specific hydrometeorological situations that can be critically important. Parajka and Blöschl (2008) analysed the effect of different masking criteria on the number of valid scatterometer images. Based on the results from the comparison of ASCAT surface soil moisture to simulated skin layer soil moisture and to ground based measurements a clear improvement in terms of biases and Pearson correlation coefficients for the cases with masking is evident. The correlations in the winter months strongly depend on the masking method. Without masking the correlations range from -0.12 (Kamp) to 0.43 (Obersiebenbrunn). This is closely related to the different climates of the two locations. The mean annual winter temperatures of the Kamp are -1.3°C while they are +0.3°C at Obersiebenbrunn (Table 4.1). Clearly, the differences in the climates have a major affect on the ASCAT soil moisture interpretation as the ASCAT readings appear too low due to snow and frozen ground. As would be expected the highest correlations and smallest biases are obtained if the remotely sensed soil moisture data set is masked based on snow cover, simulated snow water equivalents and the mean air temperature with a threshold of 2°C. However, this is also the scenario that removes the largest number of pixels. To understand the tradeoff between the strictness of the masking method and the accuracy of the ASCAT soil moisture the number of ASCAT Level 2 images was analyses for the four sites. As the masking gets stricter, the correlations of ASCAT soil moisture and simulations increase but the number of scenes decreases. If the ASCAT data are to be assimilated into a hydrological model one would probably use the strictest masking, particularly in the Kamp.

The results of this study are very promising in terms of the capability of ASCAT surface soil moisture estimates to adequately describe the large scale organisation of spatial soil moisture patterns, as well as the temporal dynamics in the skin soil layer. Hence, ASCAT surface soil moisture estimates are potentially valuable to identify appropriate model structure and parameters, as well as to decrease uncertainties of initial soil moisture conditions as a basis for accurate flood forecasts.

5 Summary and conclusions

The aim of this study is the development and the analysis of methods to quantify and reduce uncertainties in modelling and measuring hydrometeorological processes envolved in operational flood predictions. There are large forecasting uncertainties related to the meteorological model inputs, model structure and model parameters. This study discusses different methodolgies to deal with the predictive uncertainties. The different strategies are to quantive the the forecast uncertainties depending on the actual hydrometeorological situation, to reduce the uncertainty of simulated initial conditions by the use of actual runoff measurements and to validate the spatial and temporal dynamics of the simulated soil moisture, based on the comparison to remotely sensed soil moisture data. The main part of the study is carried out at the Kamp catchment, with a total catchment size of 1550 km², in the north eastern part of Austria. A spatially distributed conceptual hydrologic model, including a snow routine, a soil moisture accounting scheme and runoff routing at the hillslopes and in the river reaches, is used in all parts of this study. The model runs on a 15 minute or hourly timestep and the spatial resolution is 1km x 1km and is part of an operational flood forecasting system at the Kamp. On model input combined measurements from rain gauges and information from weather radar is used to simulate appropriate initial conditions. Quantitative precipitation forecast up to lead times of 48 hours are the basis for the flood forecasts. The model parameters and structure were chosen very carefully in this case study. The model identification procedure went substantially beyond the calibration to runoff. This means that the model can be expected to represent the hydrological processes in the Kamp catchment reasonably well.

The quantification of the forecast uncertainty is the focus of the first part of the study (Chapter 2). The main aspects of investigation are (a) how the ensemble distribution of precipitation forecasts propagates in the catchment system, and (b) to interpret the flood forecast probabilities relative to the forecast errors. The analyses indicated that, for long lead times (e.g. 48 hours), the variability of the precipitation ensemble is amplified as it propagates through the catchment system. Small errors in rainfall may translate into larger errors in runoff. The ensemble distribution of precipitation is symmetric while that of the flood forecasts is skewed to the right. Clearly, this kind of

mapping of precipitation uncertainties to runoff uncertainties for large forecast lead times is related to the non-linear nature of catchment response. In contrast, for short lead times (e.g. 12 hours and less), the variability of the precipitation ensemble is decreased as it propagates through the catchment system. This is because the forecasts are mainly controlled by observed upstream runoff and observed precipitation through the routing and runoff model components, as the forecasting system is operated in a real-time mode. The ensemble forecasts focus on the dominant source of uncertainty, i.e., uncertainty in forecast precipitation. It can be expected that the lead time where the uncertainty of the precipitation forecasts starts to amplify will depend on the catchment response characteristics, such as travel times in the river reaches and runoff concentration. In small and flashy catchments this will be a short lead time while for large catchments it will be longer.

The paper also examined the ability of the probabilistic forecasts to capture the distribution of the flood forecast errors. The statistical analyses of the ensemble forecasts for five flood events at the Kamp showed that the ensemble spread is always narrower than the distribution of the forecast errors. This would be expected as not all error sources have been represented in the ensembles. However, the ensemble spread increases with lead time in a similar way as the forecast errors. This means that the ensemble spread does provide an indicator to assess potential forecast errors over a range of lead times. A "range hit rate" was defined as the number of cases, relative to the total number of forecasts, in which the observed discharge value lies within the range of the ensemble quantiles. Analyses of the range hit rates indicate that they are small for short lead times but increase with lead time. The range hit rates are similar for lead time of 12 hours and more which suggests, again, that the ensemble spread does provide an indicator to assess potential forecast errors over a range of lead times, provided the lead times are 12 hours or larger. Even though the ensemble characteristics do not exactly match the forecast errors, they do provide information about the expected forecast errors. The comparisons indicated that, for lead times larger than 12 hours in the case of the 622 km² Kamp catchment, the ensemble spread is a useful indicator to the forecast errors. While additional error sources could be included in estimating the flood ensembles it may not be necessary for operational flood forecasting purposes as the uncertainty in forecast precipitation

is the dominant source of flood forecast uncertainty for lead times of more than 12 hours in catchments such as the Kamp.

The second part of the study (Chapter 3) deals with data assimilation in an operational flood forecasting system. The benefit of Ensemble Kalman-Filter updating in forecasting large flood events is evaluated. The aim of this paper is to examine the benefit of Ensemble Kalman Filter updating in forecasting large floods. The soil moisture of a distributed runoff model is updated based on observed runoff. The updated soil moisture is then used as an initial condition for the forecasts. The value of updating is obvious because of the long time scales associated with the hydrological processes during low and average flows. For large flood flows, the difficulty with updating runoff during an event is that phase errors usually cannot be handled well. There can be overshooting of the forecasts if phase errors are interpreted as volume errors. The procedure examined here mainly updates the slow runoff components, i.e., soil moisture between events which is then used as an initial condition for the flood forecasts. Analyses of six large flood events at the Kamp indicate that the updating indeed reduces forecast errors substantially during the flood events. It is considered a strength of this case study that data on a number of large floods (including two extreme floods) were available which is not usually the case in practical applications. This is important as one of the main motivations of implementing flood forecasting systems is to improve on the forecasting of extreme events where the damage potential is largest (Apel et al., 2006).

It should be noted that low flow and average flow conditions can usually be simulated much more accurately than flood flows. Clearly, the updating is most efficient for low and medium flows, but from a practical perspective the flood flows are usually of much more interest. However, these tend to be more difficult to predict and errors are usually much larger. In the present paper, the peak flow errors for 3 hour forecasts were reduced from 25% to 12% by the updating procedure, and from 25% to 19% for 48 hours forecasts. It should be noted that the forecast lead time of 48 hours is much larger than typical flow travel time in the streams within the catchment which are less than 2 hours. It is hence the water in the landscape rather than that in the stream that needs to be adjusted in this case study.

We believe it is important to very carefully adjust the model to the local conditions (going beyond calibration to runoff) for the updating procedure to work efficiently. In the current procedure, the main error source is attributed to the inputs and their effect on soil moisture, so model parameters are not updated. A plausible model structure and carefully adjusted model parameters are hence the basis for a good performance of the updating routine. Also, the availability of input data (16 rain gauges for model development, 8 telemetered rain gauges in a 622 km² catchment) along with radar data in this study is probably more than what one usually encounters in operational applications. With these caveats, it is suggested that updating procedures such as the one proposed in this paper can indeed substantially improve the forecasting of large floods at the catchment scale examined here.

Remotly sensed soil moisture data are used in the third part of the study as additional source of information to identify a realistic model structure and parameters. As soil moisutre is difficult to measure across the spatial scales, combining available informaiton from different sources seems to be an obvious strategy. The inter-comparison of remotely sensed soil moisture products with in situ measurements and simulated soil moisture to assess the potential of satellite based surface soil moisture estimates to reduce uncertaintys in hydrologic applications has recently been an important issue in the hydrologic and remote sensing sciences.

The comparison of the spatial surface soil moisture patterns from remote sensing (ASCAT SSM) and hydrologic modelling (Skin layer SM) indicated very good consistency for the Kamp catchment (1550 km²) in northern Austria. The realistic representation of soil moisture patterns is from utmost importance for the simulation of appropriate initial conditions for flood forecasts, as they play an important role in the highly non-linear process of runoff generation. However, the positve impact of surface soil moisture products on the prediction of root zone soil moisture is support by the results of several studies (Ottlé and Vidal-Madjar, 1994; Georgakakos and Baumer, 1996; Heathman et al., 2003, Draper et al., 2011).

The evaluation results based on the comparison at the Kamp catchment and three field site locations (plot scale) showed the highest correlations for the comparison of TDR measurements and simulated main layer soil moisture. Similar to the results obtained by Aguela et al. (2008), Albergel et al. (2009) and Brocca et al. (2010), the

correlations and biases between in-situ measurements in a soil depth of 10 cm and the remotely sensed soil moisture estimates are much better if the ASCAT SWI, instead of the ASCAT SSM product, is used for the comparison at the three field sites in Austria. The correlation coefficients are in the order of magnitude of the correlations obtained at the French in-situ measurement network. However, the resutls at the Austrian field sites indicate slightly higher correlations for the summer periods from April to September. This can be related to the lower accuracy of the ASCAT surface soil moisture estimates for the winter period.

Looking at the comparison of ASCAT surface soil moisture and simulated surface soil moisture the results at the Kamp catchment and the three field sites in Austria the correlations and biases are quite good and show consistency with the results found in a previous studies in France (Albergel et al., 2010). The seasonal differnces are obviously related to the climatic conditions at the different locations, as the greatest seasonal differnces are obtained for the study area with the coldest mean winter temperature. The results for the winter periods raise the question, how to deal with the influence of snow cover and frozen soils on the remotely sensed soil moisture estimates, as microwaves can only detect liquid water. Usually a static masking method is chosen that depends on the location and the day of the year (Bartalis, 2007), but this is unable to account for specific hydrometeorological situations that can be critically important. Parajka and Blöschl (2008) analysed the effect of different masking criteria on the number of valid scatterometer images. Based on the results from the comparison of ASCAT surface soil moisture to simulated skin layer soil moisture and to ground based measurements a clear improvement in terms of biases and Pearson correlation coefficients for the cases with masking is evident. The correlations in the winter months strongly depend on the masking method. Without masking the correlations range from –0.12 (Kamp) to 0.43 (Obersiebenbrunn). This is closely related to the different climates of the two locations. The mean annual winter temperatures of the Kamp are -1.3°C while they are +0.3°C at Obersiebenbrunn (Table 4.1). Clearly, the differences in the climates have a major affect on the ASCAT soil moisture interpretation as the ASCAT readings appear too low due to snow and frozen ground. As would be expected the highest correlations and smallest biases are obtained if the remotely sensed soil moisture data set is masked based on snow cover, simulated snow water equivalents and the mean air temperature with a

threshold of 2°C. However, this is also the scenario that removes the largest number of pixels. To understand the tradeoff between the strictness of the masking method and the accuracy of the ASCAT soil moisture the number of ASCAT Level 2 images was analyses for the four sites. As the masking gets stricter, the correlations of ASCAT soil moisture and simulations increase but the number of scenes decreases. If the ASCAT data are to be assimilated into a hydrological model one would probably use the strictest masking, particularly in the Kamp.

The results of this study are very promising in terms of the capability of ASCAT surface soil moisture estimates to adequately describe the large scale organisation of spatial soil moisture patterns, as well as the temporal dynamics in the skin soil layer. Hence, ASCAT surface soil moisture estimates are potentially valuable to identify appropriate model structure and parameters, as well as to decrease uncertainties of initial soil moisture conditions as a basis for accurate flood forecasts.

Appendix A

Structure of the soil moisture model

A conceptual soil moisture accounting scheme is used at the model grid scale. The sum of rain and melt, $P_r + M$, is split into a component dS that increases soil moisture of a top layer, S_s, and a component Q_p that contributes to runoff. The components are split as a function of S_s:

$$Q_p = \left(\frac{S_s}{L_s}\right)^{\beta} \cdot (P_r + M) \tag{A.1}$$

L_s is the maximum soil moisture storage. β controls the characteristics of runoff generation and is termed the non-linearity parameter. If the top soil layer is saturated, i.e., $S_s = L_s$, all rainfall and snowmelt contributes to runoff and dS is 0. If the top soil layer is not saturated, i.e., $S_s < L_s$, rainfall and snowmelt contribute to runoff as well as to increasing S_s through $dS > 0$:

$$\begin{aligned} dS &= P_r + M - Q_p - Q_{by} && \text{if} \quad P_r + M - Q_p - Q_{by} > 0 \\ dS &= 0 && \text{otherwise} \end{aligned} \tag{A.2}$$

where, additionally, bypass flow Q_{by} is accounted for. Analysis of the runoff data at the Kamp indicated that flow that bypasses the soil matrix and directly contributes to the storage of the lower soil zone is important for intermediate soil moisture states S_s. For $\xi_1 \cdot L_s < S_s < \xi_2 \cdot L_s$ (with ξ_1=0.4, ξ_2=0.9) bypass flow was assumed to occur as

$$\begin{aligned} Q_{by} &= \alpha_{by} \cdot (P_r + M) && \text{if} \quad \alpha_{by} \cdot (P_r + M) < L_{by} \\ Q_{by} &= L_{by} && \text{otherwise} \end{aligned} \tag{A.3}$$

while no by pass flow was assumed to occur for dry and very wet soils. Changes in the soil moisture of the top soil layer S_s from time step *i* – 1 to *i* are accounted for by

$$S_{s,i} = S_{s,i-1} + (dS - E_A) \cdot \Delta t \tag{A.4}$$

The only process that decreases S_s is evaporation E_A which is calculated from potential evaporation, E_P, by a piecewise linear function of the soil moisture of the top layer:

$$E_A = E_P \cdot \frac{S_s}{L_P} \qquad \text{if} \quad S_s < L_p$$
$$E_A = E_P \qquad \text{otherwise} \tag{A.5}$$

where L_p is a parameter termed the limit for potential evaporation. Potential evaporation was estimated by the modified Blaney-Criddle method (DVWK, 1996) as a function of air temperature. This representation of potential evaporation was compared to other methods in Parajka et al. (2003) suggesting that it gives plausible results in Austria.

References

Albergel, C., Rüdiger, C., Carrer, D., Calvet, J.-C., Fritz, N., Naeimi, V., Bartalis, Z., Hasenauer, S. (2010). An evaluation of ASCAT surface soil moisture products with in-situ observations in Southwestern France, Hydrol. Earth Syst. Sci., 13, 115-124.

Albergel, C., Calvet, J.-C., de Rosnay, T., Balsamo, G., Wagner, W., Hasenauer, S., Naeimi, V., Martin, E., Bazile, E., Bouyssel, F. and Mahfouf, J.-F. (2010). Cross-evaluation of modelled and remotely sensed surface soil moisture with in situ data in southwestern France, Hydrol. Earth Syst. Sci., 14, 2177–2191.

Apel, H., Thieken, A.H. , Merz, B. and Blöschl, G. (2006). A probabilistic modelling system for assessing flood risks. Nat. Hazards, 38, pp. 79-100.

Blöschl, G. and Zehe, E. (2005). On hydrological predictability. INVITED commentary. Hydrol. Process., 19 (19), pp. 3923-3929.

Blöschl, G., Reszler, Ch. and Komma, J. (2006). Hochwasservorhersage Kamp - Hydrologie (Kamp flood forecasting system - hydrology). Final Report to the Lower Austria State Government and EVN Energy Supply, Feb. 2006. Institute of Hydraulic and Water Resources Engineering, Vienna University of Technology.

Blöschl, G., Reszler, C. and Komma, J. (2008). A spatially distributed flash flood forecasting model. Environ. Modell. Softw., 23 (4), pp. 464-478.

Brocca, L., Melone, F., Moramarco, T., Wagner, W., and Hasenauer, S. (2010). ASCAT soil wetness index validation through in situ and modeled soil moisture data in central Italy, Remote Sens. Environ. 114, 2745–2755.

Buizza, R., Hollingsworth, A., Lalaurette, F. and Ghelli, A. (1999). Probabilistic Predictions of Precipitation Using the ECMWF Ensemble Prediction System. Weather Forecast., 14, pp. 168-189.

Buizza, R. (2003). Weather Prediction: Ensemble Prediction. Encyclopaedia of Atmospheric Sciences, Academic Press, London, pp. 2546-2557.

Chiew, F.H.S., Peel, M.C., McMahon, T.A. and Siriwardena, L.W. (2006). Precipitation elasticity of streamflow in catchments across the world. Climate Variability

and Change-Hydrological Impacts. Proceedings of the Fifth FRIEND World Conference held at Havana, Cuba, November 2006, IAHS Publ. 308, pp. 256-262.

Crow, W.T. and Van Loon, E. (2006). Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture. J. Hydrometeorol. 7 (3),pp. 421-432.

Das N.N., Mohanty, B.P., Cosh, M.H., and Jackson, T.J. (2008). Modeling and assimilation of root zone soil moisture using remote sensing observations in Walnut Gulch Watershed during SMEX04. Remote Sens. Environ. 112, 415-429.

DVWK (1996). Ermittlung der Verdunstung von Land- und Wasserflächen, DVWK-Merkblätter, Heft 238, Bonn.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res., 99 (C5), pp. 10,143-10,162.

Francois C., Quesney A. and Ottle, C. (2003). Sequential Assimilation of ERS-1 SAR Data into a Coupled Land Surface-Hydrological Model Using an Extended Kalman Filter. American Meteorological Society.

Georgakakos, K.P. and Baumer, O.W. (1996). Measurement and utilization of on-site soil moisture data. J. Hydrol. 184, 131-152.

Golding, B.W. (1998). Nimrod: A system for generating automated very short range forecasts. Meteorol. Appl., 5, pp. 1-16.

Grayson, R. and Blöschl G. (2000). Spatial modelling of catchment dynamics. Chapter 3 in R. Grayson and G. Blöschl (Eds.).Spatial Patterns in Catchment Hydrology: Observations and Modelling. Cambridge University Press, Cambridge, pp. 51-81.

Grayson, R., Blöschl, G., Western A. and McMahon, T. (2002). Advances in the use of observed spatial patterns of catchment hydrological response. Adv. Water Resour. 25, pp. 1313-1334.

Gutknecht, D. (1991). On the development of „applicable“ models for flood forecasting. In: Hydrology for the Water Management of Large River Basins (ed. by

F.H.M. van de Ven, D. Gutknecht, D.P. Loucks & K. A. Salewicz), (Proc. Vienna Symp., August 1991), pp. 337 - 345. IAHS Publ. no. 201

Gutknecht, D., Reszler, Ch. and Blöschl, G. (2002). Das Katastrophenhochwasser vom 7. August 2002 am Kamp - eine erste Einschätzung (The August 7, 2002 - flood of the Kamp - a first assessment), Elektrotechnik und Informationstechnik, 119 (12), pp. 411-413.

Haiden, T., Kann, A., Stadlbacher, K., Steinheimer, M. and Wittmann, C. (2007). Integrated Nowcasting through Comprehensive Analysis (INCA).- System overview. ZAMG report, 49p. http://www.zamg.ac.at/fix/INCA_system.doc accessed March 26, 2007.

Hall, D.K., Riggs, G.A., Salomonson, V.V., DiGirolamo, N.E., and Bayr, K. J. (2002). MODIS snow-cover products, Remote Sens. Environ., Volume 83, Issues 1-2, Pages 181-194.

Heathman, G.C., Starks, P.J., Ahuja, L.R., Jackson, T.J. (2003). Assimilation of surface soil moisture to estimate profile soil water content. J. Hydrol. 279, 1-17.

Herschy, R.W. (2002). The uncertainty in a current meter measurement. Flow Meas. Instrum. 13, pp. 281–284.

Houser P., Goodrich, D. and Syed, K. (2000).Runoff, precipitation, and soil moisture at Walnut Gulch. Chapter 6 in R. Grayson and G. Blöschl (Eds.).Spatial Patterns in Catchment Hydrology: Observations and Modelling. Cambridge University Press, Cambridge, pp. 125-157.

Jackson, T.C., Schmugge, T.J., Nicks, A.D., Coleman, G.A. and Engman E.T. (1981). Soil moisture updating and microwave remote sensing for hydrological simulations, Hydrol. Sci. Bull. 26 (3).203-319.

Kalman, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems. J. Basic Eng.-T ASME, 82 (D), pp. 35-45.

Kann, A. and Haiden, T. (2005). The August 2002 flood in Austria: sensitivity of precipitation forecast skill to area size and duration. Meteorol. Z., 14, pp. 369-377.

Krzysztofowicz, R. (2001). Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model. J. Hydrol., 249, pp. 69-85.

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

Madsen, H., Rosbjerg, D., Damgard, J. and Hansen, F.S. (2003). Data assimilation in the MIKE 11 Flood Forecasting system using Kalman Filtering. In: Water resources systems, Blöschl, G. (ed.), IAHS publication, no 281, pp. 75-81.

Mason, S.J. and Graham, N.E. (1999). Conditional Probabilities, Relative Operating Characteristics, and Relative Operating Levels. Weather Forecast., 14(5), pp. 713-725.

Matgen, P., Heitz, S., Hasenauer, S., Hissler, C., Brocca, L., Hoffmann, L., Wagner W. and Savenije, H.H.G. (2011). On the potential of METOP ASCAT-derived soil wetness indices as a new aperture for hydrological monitoring and prediction: a field evaluation over Luxembourg, Hydrol. Process., Accepted Article, doi: 10.1002/hyp.8316

McLaughlin, D. (1994). Recent advances in hydrologic data assimilation, In U.S. National Report to the IUGG (1991-1994), Rev. Geophys., Supplement, pp. 977-984.

Merz, R. and Blöschl, G. (2005). Flood Frequency Regionalisation - spatial proximity vs. catchment attributes, J. Hydrol., 302 (1-4), pp. 283-306.

Moradkhani, H., Sorooshian, S., Gupta, H.V. and Houser, P.R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman Filter. Adv. Water Res., 28, pp. 135-147.

O'Connell, P.E. and Clarke, R.T. (1981). Adaptive hydrological forecasting - a review. Hydrological Sciences-Bulletin-des Sciences Hydrologiques, 26, pp. 179-205.

Ottlé, C., and Vidal-Madjar, D. (1994). Assimilation of Soil Moisture Inferred from Infrared Remote Sensing in a Hydrological Model Over the HAPEX-MOBILHY Region. J. Hydrol., 158, 241-264.

Parajka, J. and Blöschl, G. (2006). Validation of MODIS snow cover images over Austria. Hydrol. Earth Syst. Sc, 10, pp. 679-689.

Parajka J. and Blöschl, G. (2008). Spatio-temporal combination of MODIS images – potential for snow cover mapping. Water Resour. Res., VOL. 44, W03406, 13 PP.

Parajka, J., Merz, R. and Blöschl, G. (2003). Estimation of daily potential evapotranspiration for regional water balance modeling in Austria. In: 11th. International Poster Day and Institute of Hydrology Open Day "Transport of Water, Chemicals and Energy in the Soil - Crop Canopy - Atmosphere System", 20. November 2003, Bratislava, Slovakia. Published on CD-ROM, Slovak Academy of Sciences, ISBN 80 - 89139 -02- 7, pp. 299-306.

Parajka, J., Merz, R. and Blöschl, G. (2005a). A comparison of regionalisation methods for catchment model parameters. Hydrol. Earth Syst. Sc., 9, pp. 157-171.

Parajka, J., Merz, R. and Blöschl, G., (2005c). Regionale Wasserbilanzkomponenten für Österreich auf Tagesbasis (Regional water balance components in Austria on a daily basis). Österreichische Wasser- und Abfallwirtschaft, 57, H 3/4. pp. 43-56.

Parajka, J., Naeimi, F., Blöschl, G. and Komma, J. (2009). Matching ERS scatterometer based soil moisture patterns with simulations of a conceptual dual layer hydrologic model over Austria, Hydrol. Earth Syst. Sci., 13, 259-271, 2009.

Parajka, J., Naeimi, V., Blöschl, G., Wagner, W., Merz, R. and Scipal, K. (2005b). Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale. Hydrol. Earth Syst. Sc., 10, pp. 353-368.

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

Reichle, R., McLaughlin, D.B. and Entekhabi, D. (2002). Hydrologic Data Assimilation with the Ensemble Kalman Filter. Mon. Weather Rev., 130, 1, pp. 103-114.

Reszler, Ch., Komma, J., Blöschl, G. and Gutknecht, D. (2006). Ein Ansatz zur Identifikation flächendetaillierter Abflussmodelle für die Hochwasservorhersage (An approach to identifying spatially distributed runoff models for flood forecasting). Hydrol. Wasserbewirts., 50 (5), pp. 220-232.

Schaake, J., Perica, S., Mullusky, M., Demargne J., Welles, E. and Wu, L. (2004). Pre-processing of Atmospheric Forcing for Ensemble Streamflow Prediction. Proceedings of the the 84th AMS Annual Meeting held in Seattle, USA, January 2004, 5p. http://ams.confex.com/ams/pdfpapers/72172.pdf, accessed Dec. 18, 2006.

Siccardi F., Boni, G., Ferraris L. and Rudari, R. (2005). A hydro-meteorological approach for probabilistic flood forecast, J. Geophys. Res, 110, No. D5, 10.1029/2004JD005314.

Sinclair, S. and Pegram, G.G.S. (2010).A comparison of ASCAT and modelled soil moisture over South Africa, using TOPKAPI in land surface mode, Hydrol. Earth Syst. Sci., 14, 613–626.

Steinheimer, M. and Haiden, T. (2007). Improved nowcasting of precipitation based on convective analysis fields. Adv. Geosci., 10:125-131.

Taylor, J.W. and Buizza, R. (2003). Using weather ensemble predictions in electricity demand forecasting. Int. J. Forecasting, 19, pp. 57-70.

Ulaby F.T., R.K. Moore and A.K. Fung (1982)."Radar Remote Sensing and Surface Scattering and Emission Theory", Vol. II, Microwave Remote Sensing, Active and Passive, Addison-Wesley Publishing Company, Reading, MA, 1069 pages

Vischel, T., Pegram, G.G.S., Sinclair, S., Wagner, W. and Bartsch, A. (2008). Comparison of soil moisture fields estimated by catchment modelling and remote sensing: a case study in South Africa, Hydrol. Earth Syst. Sci., 12, 751–767, 2008

Wagner, W., Lemoine, G., and Rott, H. (1999). A method for estimating soil moisture from ERS scatterometer and soil data, Remote Sens. Environ., 70, 191-207.

Wagner, W., Lemoine, G., Borgeaud, M., and Rott, H.(1999). A Study of Vegetation Cover Effects on ERS Scatterometer Data, IEEE Trans. Geosci. Remote Sensing, Vol. 37, No. 2, pp. 938-948.

Wagner, W., Blöschl, G., Pampaloni, P., Calvet, J.-C., Bizzarri, B., Wigneron, J.-P. and Kerr, Y. (2007).Operational readiness of microwave remote sensing of soil moisture for hydrologic applications. Nordic Hydrology, 38 (1), pp. 1-20.

Walker J.P., Willgoose, G.R. and Kalma, J.D. (2001). One-Dimensional Soil Moisture Profile Retrieval by Assimilation of Near-Surface Measurements: A Simplified Soil Moisture Model and Field Application. J. Hydrometeorol., 2.

Wang, Y., Haiden, T. and Kann, A. (2006). The operational limited area modelling system at ZAMG: ALADIN-AUSTRIA, Österr. Beiträge zu Meteorologie und Geophysik, 37, 33p.

Weerts, A.H. and Serafy, Y.H., (2006). Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resour. Res., 42, doi:10.1029/2005WR004093.

Western, A.W. and Blöschl, G. (1999). On the spatial scaling of soil moisture. J. Hydrol., 217, pp. 203-224.

Western, A., Grayson, R. and Blöschl, G. (2002).Scaling of soil moisture: a hydrologic perspective. Ann. Rev. Earth and Planetary Sci., 30, pp. 149-180.

Whitaker, J.S. and Loughe, A.F. (1998). The relationship between ensemble spread and ensemble mean skill. Mon. Weather Rev., 126, pp. 3292–3302.

Wood, E.F. (1980). Real-Time Forecasting/Control of Water Resources Systems – selected papers from an IIASA Workshop. IIASA Proceedings Series, 8, pp. 37-46.

Zehe, E. and Blöschl, G. (2004). Predictability of hydrologic response at the plot and catchment scales – the role of initial conditions. Water Resour. Res., 40: W10202.