TU WIEN bl.

Doctoral Thesis

Real time flood forecasting - a regional analysis

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

Operationelle Hochwasservorhersage - eine regionale Analyse

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. Thomas Nester

Matrikelnummer 9725738

Embelgasse 36/10, 1050 Wien

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

Gutachter: Prof. Ján Szolgay, PhD.
Department of Land and Water Resources Management,
Slovak University of Technology in Bratislava
Radlinského 11, blok C, 12. posch, 813 68 Bratislava 15

Wien, November 2011

How high's the water, Mama?

Five feet high and risin'

How high's the water, Papa?

It's five feet high and risin'

Johnny Cash, *Five Feet High And Rising*

Kurzfassung

Hochwasserereignisse, wie sie im Jahr 2002 im Donaueinzugsgebiet aufgetreten sind, haben das Interesse der Öffentlichkeit an möglichst frühzeitigen Hochwasserwarnungen und Vorhersagen steigen lassen. In den folgenden Jahren wurden deshalb Vorhersagesysteme für die meisten Flüsse in Österreich eingerichtet. Ziel dieser Arbeit ist die Evaluierung des Hochwasservorhersagesystems für die österreichischen Donauzubringer, das von der TU Wien entwickelt wurde. Bei der Entwicklung von operationellen Vorhersagesystemen ist es notwendig, Routinen zu verwenden, die auch mit einem Minimum an online verfügbaren Daten plausible und genaue Ergebnisse liefern. Das in dieser Arbeit verwendete Modell beschreibt die Schneeakkumulation und -schmelze, den Bodenfeuchtehaushalt und den Abfluss am Hang und im Gerinne.

In Kapitel 2 wird der Einfluss von klimatologischen und gebietsspezifischen Eigenschaften auf die Güte der Modellergebnisse untersucht. Für die Interpretation der Vorhersagen ist ein Verständnis der Funktionsweise eines hydrologischen Modells in unterschiedlichen hydrologischen Situationen wichtig. Die Ergebnisse zeigen, dass in feuchten Gebieten bessere Modellergebnisse erzielt werden können als in trockenen Gebieten, und dass simulierte Hochwasserscheitel in feuchten Gebieten kleinere Fehler aufweisen als in trockenen Gebieten. Die Güte der Modellergebnisse wird am stärksten von der Größe des Einzugsgebietes und dem Anteil von Regen am gesamten Niederschlag beeinflusst.

In Kapitel 3 liegt der Fokus auf dem Schneemodul des hydrologischen Modells. Zur zeitlichen und räumlichen Validierung der simulierten Schneeprozesse wurden Satellitendaten verwendet. Ein Vergleich von Modellergebnissen und den verwendeten MODIS-Satellitendaten zeigt, dass sie nützliche Zusatzinformation darstellen, wenn die Wolkenbedeckung weniger als 80% beträgt. Die Schneebedeckung in den Voralpen und in bewaldeten Gebieten wird vom Modell etwas unterschätzt, in alpinen Bereichen und nicht bewaldeten Gebieten stimmen Modell und Satellitendaten gut überein.

Hochwasservorhersagen sind grundsätzlich mit Fehlern behaftet, die auf Unsicherheiten der meteorologischen Prognosen und des hydrologischen Modells zurückzuführen sind. Um diese Unsicherheiten zu quantifizieren, wurden in Kapitel 4 mehrere gleich wahrscheinliche Niederschlagsvorhersagen, sogenannte Ensembles, als Eingangsdaten in das Modell verwendet. Die Ergebnisse zeigen, dass die Modellunsicherheit dominant ist, wenn die Prognosefristen kurz sind und sich die prognostizierten möglichen Niederschläge wenig voneinander unterscheiden. Bei längeren Prognosefristen wird jedoch die Unsicherheit des Niederschlags dominant. Die Auswertungen zeigen auch, dass die Vorhersageensembles in allen Fällen ein guter Indikator für die Prognoseunsicherheiten sind.

Mit dem an der TU Wien entwickelten Modell ist es möglich, nicht nur den wahrscheinlichsten Wert des zukünftigen Hochwassers in einem Gebiet vorherzusagen, sondern auch Aussagen über die zu erwartende Streubreite zu treffen. Im Rahmen des Hochwasserrisikomanagements, wie durch die EU Hochwasserrichtlinie vorgegeben, ist beides eine wertvolle Information.

Abstract

Recent flood events, such as the 2002 floods in the Danube catchment, have raised the public awareness for the need for flood warnings and forecasts. Following these floods, operational flood forecasting systems were developed for most rivers in Austria. The aim of this study is the evaluation of the flood forecasting system for the Austrian Danube tributaries, which was designed at the Vienna University of Technology. When developing operational flood forecasting systems, it is important to use routines that can be used with a limited amoung of real time data. The model used in the study describes snowaccumulation and -melt, the soil moisture accounting and and catchment and stream routing functions.

For the interpretation of runoff forecasts it is vital to understand how the model works in different hydrological situations. In chapter 2, the model performance is evaluated as a function of climatological and catchment characteristics. The results indicate that the that the model performance increases with increasing wetness of the catchment, whereas the peak errors tend to decrease with increasing wetness. However, the results suggest that the catchment size and the ratio of rain to total precipitation are the most important controls on the performance of the runoff model.

The evaluation of the snow routine of the hydrological model is in the focus of chapter 3. To validate the simulated snow processes on a temporal and spatial scale satellite data were used. A comparison of the model results and the MODIS-satellite data shows that the satellite data are a useful additional information if the cloud coverage is less than 80%. Results indicate that the model tends to underestimate snow cover in prealpine areas and forested areas while it performs better in alpine catchments and open land.

Flood forecasts are generally associated with errors which can be attributed to uncertainties in the meteorological forecasts and the hydrologic simulations. To quantifiy the uncertainties, a set of equally probable precipitation forecast (an ensemble) is used as input into the hydrological model in chapter 4. The results indicate that the hydrologic simulation uncertainty dominates for short lead times and for narrow ensemble spreads. For longer lead times the uncertainty from the precipitation forecasts dominate. The results also indicate that the ensembles are a good indicator for the forecast uncertainty.

The flood forecasting system designed at the Vienna University of Technology allows not only to forecast the most likely flood peak in a catchment, but also to make conclusions about the expected spread of the forecasts. In the context of integrated flood management, as demanded by the the flood directive of the European union, this is a valuable information.

Acknowledgement

Writing a PhD thesis is like climbing a mountain, a very high mountain. Step after step, for many hours, until you reach the peak. Sometimes the steps are rather easy, but it can also be one step forward and two or three steps back as well.

But finally – I am at the top of the mountain – but I know that it would have been impossible for me alone to get to the top. I would like to thank all the people who helped me to get to the top of the mountain.

Above all, I would like to thank Prof. Günter Blöschl for his support, help and motivation – not only to work on this thesis, but also to finalize it. He helped me to focus on the important things throughout all stages of my work and provided excellent input and ideas.

I am also grateful to Prof. Dieter Gutknecht, who gave me the opportunity to work on the flood forecasting project many years ago when I had just received my diploma and had no idea what would expect me. I learned a lot while working on the project.

The good advice, support and friendship of Robert Kirnbauer have been invaluable on both an academic and a personal level, for which I am extremely grateful.

My thanks also go to Prof. Ján Szolgay from the Slovak University of Technology in Bratislava for reviewing this thesis.

I would also like to thank all my current and former colleagues at the Institute for Hydraulic and Water Resources Engineering at the Technical University in Vienna – not only for the help you offered me and the pleasure of working together with you, but also for the friendships which have developed and for the serious and not so serious talks. Thank you, Alberto, Christian, Christine, Duro, Gerti, Herbert, Horst, Ines, Jon, Jose-Luis, Julia, Julia, Judit, Jürgen, Magdalena, Maria, Martine, Michael, Paul, Peter, Peter, Ralf, Richard, Tim, and last but not least, Ulli.

I would like to thank the Hydrographic Services of Lower Austria and Upper Austria for providing the discharge data and the Central Institute for Meteorology and Geodynamics (ZAMG) in Vienna for providing the meteorological data. The study was performed as part of developing an operational flood foresting system for the Austrian Danube and tributaries.

Finally, I would also like to thank my family - my parents and my brothers and sisters for their support and encouragement. And of course, thank you Meri for your love and support throughout the years, and Alexander for bringing so much joy into our lives.

Contents

List of Figures

List of Tables

1 Introduction

Floods are among the most frequent and costly natural disasters in terms of human hardship and economic loss. In Austria, flood damages were estimated to be in the order of €3 billion and €550 million for the major flood events of August 2002 and August 2005. After these flood events, the public awareness has increased that flood warnings and forecasts can substantially reduce the damage to property and life.

The purpose of the flood framework directive of the European Union (European Union, 2007) *"is to establish a framework for the assessment and management of flood risks, aiming at the reduction of the adverse consequences for human health, the environment, cultural heritage and economic activity associated with floods in the Community."* Flood risk management plans have to be completed and published by the member states by December 2015. Flood forecasting is an important component of the flood risk management plan.

However, flood forecasting is a challenge as physical processes that control runoff generation and flow routing in catchments are complex and highly variable. Moore et al. (2005) summarized the challenges in flood forecasting as *"Imperfect estimates of rainfall and river flow are used with mathematical models of river systems that aim, in an approximate way, to represent the physical processes affecting water movement."* Several assumptions have to be met to estimate future runoff (Gutknecht, 1988): First, the current hydrological situation (the "initial condition") at the time of forecast has to be known with adequate accuracy. Secondly, mathematical models are required to estimate future runoff.

The knowledge of the current hydrological situation means that the meteorological and hydrological histories prior to the current situation have to be known as adequately as possible. This is because the runoff estimates highly depend on the state of, for example, the soil moisture and the snow conditions. The sensitivity of runoff simulations on the initial conditions has been shown by a number of authors (e.g., Zehe and Blöschl, 2004; Komma et al., 2008). The predictability of a flood event does not depend on the initial conditions alone; it can also depend on the character and the seasonality of the event. A wide range of process types that can cause floods including long-rain floods, short-rain floods, flash floods, rain-on-snow floods, and snowmelt was identified by Merz and Blöschl (2003). The same authors analysed process indicators such as the timing of the floods, the storm duration, the rainfall depth and the snowmelt, the catchment state, the runoff response dynamics and the spatial coherence. A rainfall-runoff model used for flood forecasting is ideally able to forecast all different runoff situations. However, models are always a simplification and approximation of reality meaning that there is also uncertainty and errors associated with the models. Ewen et al. (2006) classified the errors in rainfall-runoff modelling into three groups: (1) model structure

errors which are associated with the model equations, (2) parameter errors, associated with parameter values used in the equations and (3) run time errors, associated with rainfall and other forcing data. This means that the rainfall-runoff models have imperfect structures and use imperfect data, which the forecaster has to deal with.

If the initial conditions are well-known, future runoff can be estimated using (1) runoff routing models with observations of runoff, (2) precipitation measurements and (3) precipitation forecasts in a rainfall-runoff model. An approach to combine the different sources of information was adopted in the forecasting system for the Danube tributaries. Using real time observations of runoff in a routing model yields the most accurate runoff forecasts despite the uncertainty in the runoff measurements. However, when using runoff routing models the lead time is limited to the range of the travel time in the stream. Using a rainfall-runoff model with measurements of precipitation and temperature increases the lead time, but the accuracy of the forecasts decreases as the meteorological observations are associated with uncertainty as well. The accuracy of the forecast further decreases when using precipitation and temperature forecasts in the rainfall-runoff model (e.g., Blöschl, 2008; Komma et al., 2009).

The study consists of three main parts. In the first part (Chapter 2) the performance of the model with regard to flood simulation is analysed using different statistical measures such as the Nash-Sutcliffe model efficiency and peak errors. The understanding of the model performance is important for both practical and theoretical perspectives. The dependence of the model performance on the quality and quantity of the data used for the calibration is quite obvious. However, it is also interesting to see how catchment characteristics such as mean annual runoff and precipitation or the catchment area can affect the model performance. An analysis of the snow model is carried out in the second part (Chapter 3) of the study. Snow model results are compared to snow cover data derived from the MODIS satellites on a temporal and spatial scale. The third part (Chapter 4) consists of an analysis of ensemble runoff forecasts, which are used to quantify the uncertainty of forecasts. Sources of uncertainty include the precipitation forecast uncertainty and the hydrologic simulation uncertainty, which are quantified by means of a spread-skill analysis.

The analyses are based upon the flood forecasting system for the Danube tributaries in Austria. They draw on data from 57 catchments covering different hydrological regimes in Austria and Bavaria. Observed meteorological and hydrological data from 2003-2009 were to calibrate and validate the rainfall-runoff model; meteorological forecasts were available for 2006-2009 and remote sensing data for 2003-2009.

2 Climate and catchment controls on the performance of regional flood simulations

Abstract

Flood runoff is simulated for 57 catchments in Austria and Southern Germany. Catchment sizes range from 70 to 25600 km^2, elevations from 200 to 3800 m and mean annual precipitation from 700 to 2000 mm. A semi-distributed conceptual water balance model on an hourly time step is used to examine how model performance (both calibration and validation) is related to the hydroclimatic characteristics of the catchments. Model performance of runoff is measured in terms of four indices, the Nash-Sutcliffe model efficiency, the volume error, the percent absolute peak errors and the error in the timing of the peaks. The simulation results indicate that the model performance in terms of the Nash-Sutcliffe model efficiency has a tendency to increase with mean annual precipitation, mean annual runoff, the long-term ratio of rainfall and total precipitation and catchment size. Peak errors have a tendency to decrease with mean annual precipitation and mean annual runoff as well as with catchment size. Catchment size is the most important control on the model performance but also the ratio rain/precipitation is an important factor. The hydrograph shapes tend to improve with the spatial scale and magnitude of the precipitation events. Calibration and validation results are consistent in terms of these controls on model performance.

2.1 Introduction

Understanding the performance of hydrological models is important for a number of reasons. From a practical perspective, it is essential to know how well streamflow and flood forecasts will perform. For short lead times, streamflow and flood forecasts are mainly limited by the hydrological model (Blöschl et al., 2008) as the short term forecasts are highly dependant on the observed precipitation. From a more theoretical perspective it is also of interest to understand what the limits of hydrological predictability (Blöschl and Zehe, 2005) are which may give guidance on selecting model complexity (Sivapalan, 2003; Jakeman and Hornberger, 1993). It is clear that model performance depends on the type and amount of data available as well as the model type (Grayson et al., 2002) but there is also evidence in the literature that the model performance depends on climate and catchment characteristics although these relationships are less apparent.

Many studies have been performed on the catchment scale (e.g., Robinson et al., 1995; Ogden and Dawdy, 2003; Vivoni et al., 2007; Merz et al., 2009) and a number of modelling studies found catchment size to be a major control on the performance of a model. The study of Hellebrand and van den Bos (2008) performed on 18 catchments in Germany ranging between 8 and 4000 km^2 showed that model performance was higher in larger catchments. Similarly, the results of Das et al. (2008) indicated that model performance is higher for larger subcatchments as random errors are likely to be cancelled out on larger scales. Oudin et al. (2008) obtained higher model efficiencies for the larger, ground-water dominated and highland catchments than the smaller catchments in the South of France and explained the differences by averaging and storage effects. Merz et al. (2009) found that the long-term water balance could be modelled more reliably with increasing catchment scale and the scatter of the model performance between catchments decreased as well. They attributed both effects to the larger number of climate stations in any one catchment. There is a line of argument suggesting that part of the hydrological variability averages out as one increases the catchment scale but there will also be additional variability that needs to be captured as the catchments become large (Sivapalan, 2003; Skøien and Blöschl, 2006).

Climate is another important control on model performance. Generally, it is well accepted that catchments in dry climates seem to be more difficult to simulate than catchments in humid climates (e.g., Xiong and Guo, 2004). Braun and Renner (1992) reported the catchments in the Swiss lowlands to be more difficult to simulate than the alpine and high-alpine catchments. The lowland catchments had less mean annual precipitation (1000 to 1230 mm/yr) than the Alpine catchments (1490 and 2400 mm/yr). Similarly, the results of Lidén and Harlin (2000) show that the model performance decreased with increasing catchment dryness for four catchments evaluated in Tanzania, Zimbabwe, Bolivia and Turkey. This is consistent with the French results of Oudin et al. (2008). One reason for the lower model performance in arid climates may be the flashier and smaller scale rainfall patterns (Yatheendradas et al., 2008). Goodrich et al. (1997), however, related the differences in model performance to the more non-linear character of the rainfall-runoff relationship in arid than in wetter regimes. Xiong and Guo (2004) state that "*it is nearly impossible to establish a clear relationship between the humidity/aridity of catchments and the model performance.*" Clearly, in arid regions, runoff tends to become an ephemeral process with threshold character (and hence non-linear), while for wet climates the rainfall-runoff relationship is more linear. In this context, Mimikou et al. (1992) have shown in their study that the model efficiency is increasing with basin humidity in five semi-arid to humid catchments in Greece. In more general terms, predictability tends to increase as the system states move away from the threshold states (Zehe et al., 2007), which is the case for increasingly wet climates. This is also true of wetter years, as compared to drier years, as noted by Gupta et al. (1999). However, catchments denoted as dry in this paper would not be considered as a dry catchment in most climate regions around the world but as catchments with moderate rainfall rates. We used the term for clarity in the context of the Alpine region where most catchments, in fact, have a mean annual runoff of 600 mm/yr or more. We define a dry catchment as catchment with a mean annual runoff of around 250 mm/yr.

Snow melt is another climate related control. Snow dominated runoff regimes tend to be easier to model than rainfall dominated regimes for a number of reasons. First, snow data can be used in model setup and calibration which gives additional information on this part of the hydrological model (Blöschl and Kirnbauer, 1991, 1992; Parajka et al., 2007). Second, and equally important, the snow dominated runoff regime tends to have a clear annual devolution with winter low flows and spring snow melt which is more predictable. Merz et al. (2009) found that the model performance for a model run on a daily time step significantly increased with the ratio of snowfall and total precipitation which they attributed to the stronger seasonality of the runoff regime.

While these studies found interesting controls on the performance of hydrological models, most of them conducted the simulations on a daily time step and for a relatively small number of catchments. As Micovic and Quick (2009) noted, the model performance may strongly depend on the temporal resolution of the model. It is hence of interest to examine the controls for a higher temporal resolution where routing effects become more important and the flood peaks are more accurately represented, and to extend the analyses to a larger number of catchments than is usually done.

The aim of this paper is to analyse the controls on the performance of a hydrological model with an hourly time resolution that includes channel routing processes. Specifically, we examine whether the model performance can be related to climatic and hydrological characteristics of the catchments. We do not only focus on the Nash-Sutcliffe model efficiency, but also on peak error measures. The rainfall runoff model used is a conceptual hydrological model (Blöschl et al., 2008) which is applied in a semi-distributed mode to 57 Danube tributaries in Austria and Germany over a period of seven years.

The organisation of the paper is as follows: Following the description of the study region and data used in Section 2.2, a short description of the model is given in Section 2.3. Section 2.4 presents the results found in the simulation runs, and in Section 2.5 the results are discussed. Appendix A gives an overview over the statistical measures used to evaluate the model and finally, appendix B gives a description of the model used.

2.2 Study region and data

The study region is hydrologically diverse covering large parts of Austria and some parts of Bavaria (Figure 2.1). The West of the region is Alpine with elevations of up to 3800 m a.s.l. while the North and East consist of prealpine terrain and lowlands with elevations between 200 and 800 m a.s.l.

Figure 2.2 (top) shows the mean annual precipitation in the study region ranging from 600 mm/yr in the East to almost 2000 mm/yr in the West. Figure 2.2 (bottom) shows the mean annual runoff depths calculated from the discharge data used in the study. The Alpine catchments generally show much higher runoff depths ranging from around 100 mm/yr in
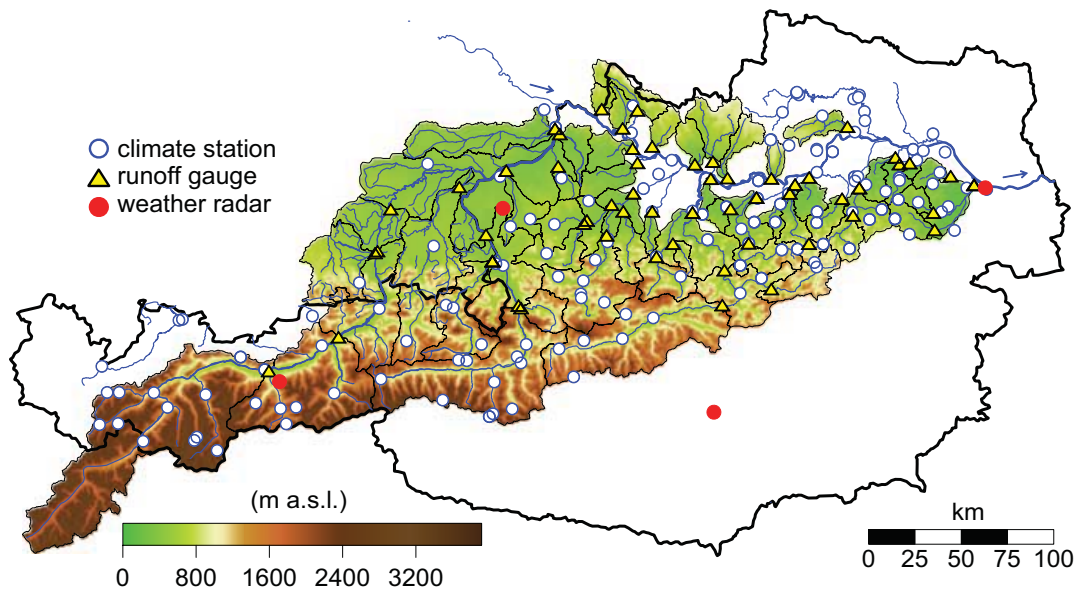
Figure 2.1: Topography of Austria and parts of Southern Germany. Elevations within the model region are shown as darker colours. The stream gauges used in the study are indicated by triangles, the precipitation gauges by circles. Weather radar stations are indicated by red circles.

the East to almost 1600 mm/yr in the West. Both figures indicate that the alpine regions are much wetter than the lowlands.

The model region consists of 57 gauged catchments with sizes ranging from 70 km^2 to 25600 km^2 and a total size of 43800 km^2. The median size of the catchments is around 400 km^2. The small catchments are mostly nested catchments. Land use is mainly agricultural in the lowlands, forested in the medium elevation ranges and alpine vegetation, rocks and glaciers in the alpine catchments.

The study was carried out with hourly data from the years 2002 to 2009. Model input data are hourly values of precipitation, air temperature and potential evapotranspiration. The data for 2002 were used as a warm-up period for the model, 2003 to 2006 was the calibration period and 2007 to 2009 was the validation period. Meteorological input data were spatially interpolated by the Central Institute for Meteorology and Geodynamics (ZAMG) in Vienna using the algorithm implemented in the INCA system (Haiden and Pistotnik, 2009; Haiden et al., 2010). The INCA system is used operationally for forecasting in Austria, but can also be used with historical data. The system operates on a horizontal resolution of 1 km and has a vertical resolution of 100-200 m. It combines surface station data, remote sensing data (radar, satellite), forecast fields of the numerical weather prediction model ALADIN, and high-resolution topographic data (Haiden et al., 2010). Currently, 408 online
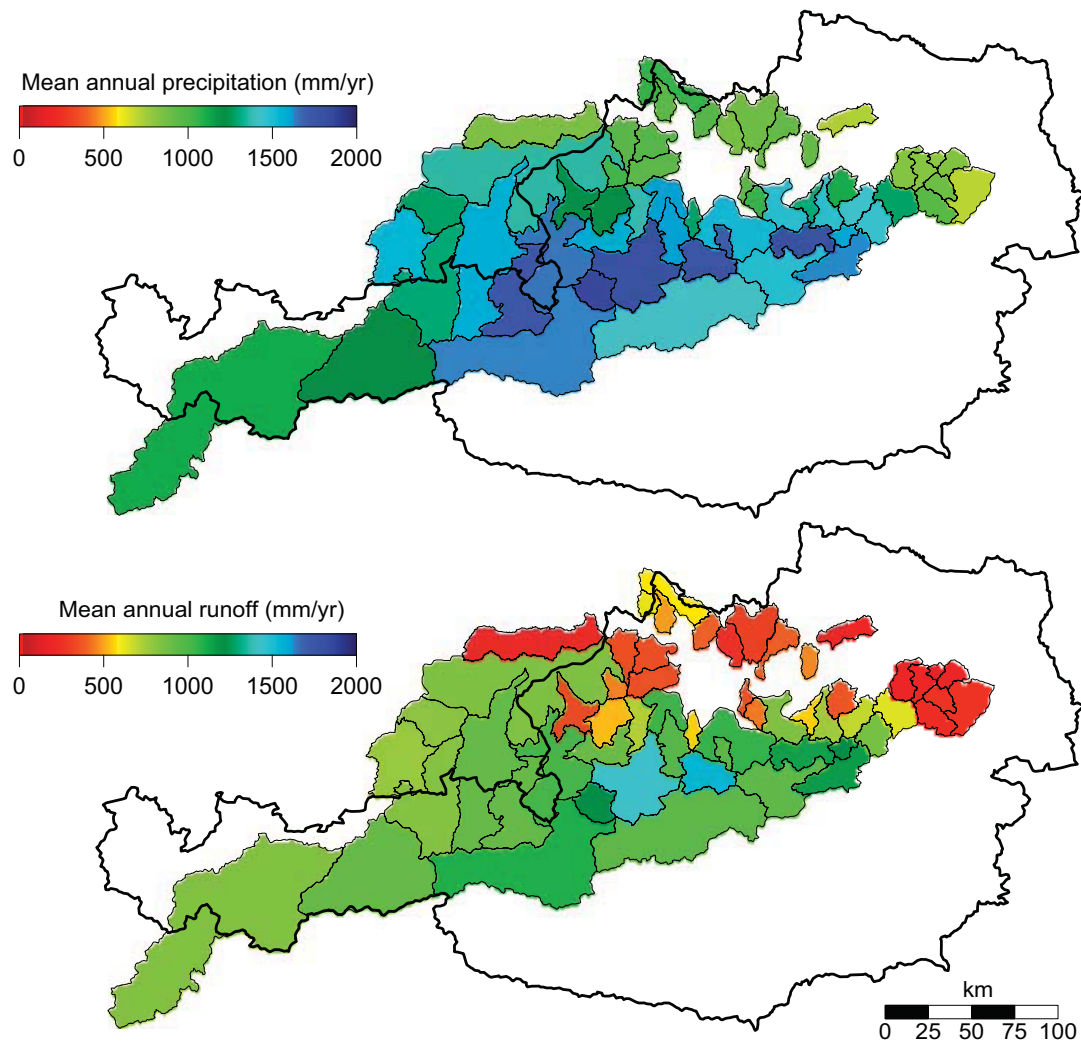
Figure 2.2: Top - mean annual precipitation calculated from the precipitation data used in the study for the years 2003-2009, bottom - mean annual runoff depths calculated from discharge data for the years 2003-2009

available climate stations are implemented in INCA; 169 of which lie within the model region, which equals to one climate station every 258 km^2. In small catchments, on average 0.35 stations per 100 km^2 are available whereas in large catchments on average 0.45 stations are available per 100 km^2. 70% of the stations are below 1000 m a.s.l., 24% are between 1000 and 2000 m a.s.l. and the remaining 6% are above 2000 m a.s.l. with the highest station at 3100 m a.s.l. Station data are interpolated to a 1x1 km grid using distance weighting, producing a gridded precipitation field that reproduces the observed precipitation values at the station locations. This grid is then superimposed with data from four weather radar stations

in Austria, combining the accuracy of the point measurements and the spatial structure of the radar field. This approach has two advantages: (1) the radar can detect precipitating cells that do not hit a station and (2) station interpolation can provide a precipitation analysis in areas not accessible to the radar beam (Haiden et al., 2010). However, there are two error sources on the precipitation side: First, the gauge deficit which is about 5% in summer and 10-20% in winter and second, the interpolation error which depends on the precipitation type. For convective storms, errors can be large even though radar is used for detection. For synoptic events, the errors are around 5-10% or less (Viglione et al., 2010a,b).

The spatial distribution of potential evapotranspiration was estimated from hourly temperature and daily potential sunshine duration by a modified Blaney-Criddle equation (DVWK, 1996). This method has been shown to give plausible results in Austria (Parajka et al., 2003). The gridded weather data fields were superimposed on the subcatchment boundaries to estimate hourly catchment average values. For air temperature and potential evapotranspiration, elevation was additionally accounted for by dividing all catchments into 500 m elevation zones. To calibrate and verify the model, hourly discharge data from 57 stream gauges were used. The data were checked for errors and in cases where a plausible correction could be made they were corrected. Otherwise they were marked as missing data.

2.3 Model

Model structure

Figure 2.3 shows the spatial layout of the model. A total of 57 sub-catchments and 58 routing modules are accounted for in the model. Each of the catchments is further divided into 500 m elevation zones to account for differences in air temperature and potential evapotranspiration. The stream gauges used in the model are shown as red points and the confluences with the main stream of the Danube are shown as yellow points. Table 2.1 gives details about catchments named in the paper such as area, mean annual precipitation and runoff.

Table 2.1: Stream gauges used in the study. MAP and MAR (2002-2009) is mean annual precipitation and runoff, respectively. * denotes catchments we consider to be dry.

Number	Gauge/catchment	area (km^2)	MAP (mm/a)	MAR (mm/a)
1	Rosenheim/Mangfall	1100	1520	770
2	Schärding/Inn	25600	1040	830
3	Haging/Antiesen*	160	1030	440
4	Fraham/Innbach*	360	940	340
5	Wels/Traun	3400	1550	1040
6	Steyr/Enns	5900	1500	1060
7	Molln/Steyrling	130	1700	955
8	Greimpersdorf/Ybbs	1100	1100	840
9	Opponitz/Ybbs	510	1140	1140
10	Krems/Imbach*	300	720	210
11	Cholerakapelle/Schwechat	180	890	250
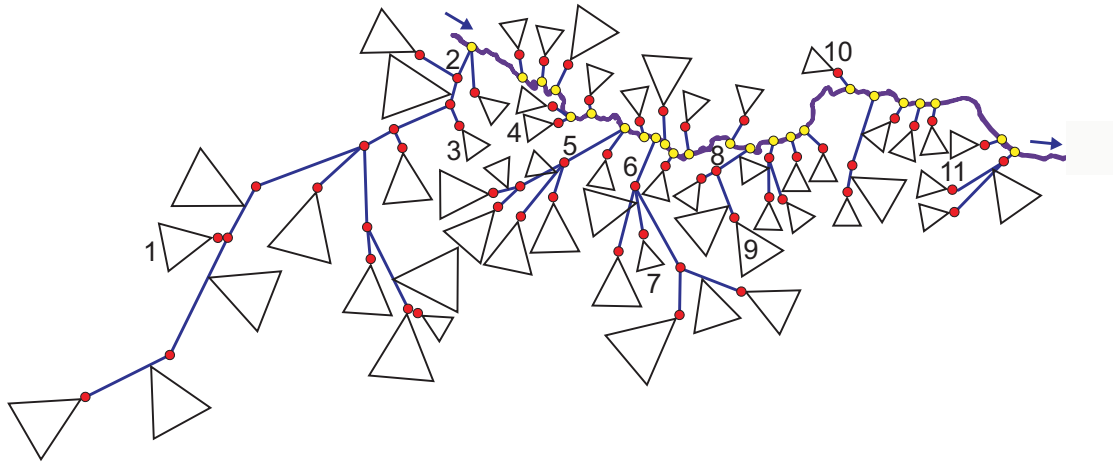
Figure 2.3: Model layout. Numbers refer to the gauges and catchments in Table 2.1. Triangles represent catchments, lines represent routing modules. The size of the triangles indicates the size of the sub-catchments: the smallest triangles stand for catchments with less than 400 km^2, and the largest for areas larger than 1890 km^2.

The rainfall runoff model used in this study is a conceptual hydrological model (Blöschl et al., 2008) which is applied in a semi-distributed mode. The structure is similar to that of the HBV model (Bergström, 1976) but several modifications were made including an additional ground water storage, a bypass flow (Blöschl et al., 2008; Komma et al., 2008) and a modified routing routine (Szolgay, 2004). Fig. 2.4 shows the model scheme for one 500 m elevation zone of a catchment.

For each elevation zone, snow processes, soil moisture processes and hill slope scale routing are simulated on an hourly time step. In the snow routine, snow accumulation and melt are represented by a simple degree-day concept, involving the degree-day factor D (mm.°C^{-1}.day^{-1}) and melt temperature T_m (°C). Catch deficit of the precipitation gauges is corrected by a snow correction factor, C_S (-). Precipitation is considered to fall as rain if the air temperature T_a (°C) is above a threshold temperature T_r (°C), as snow if T_a (°C) is below a threshold temperature T_s (°C), and as a mix if T_a (°C) is between T_r (°C) and T_s (°C). Runoff generation and changes in soil moisture storage are represented in the soil moisture routine with three parameters: the maximum soil moisture storage L_S (mm), a parameter representing the soil moisture state above which evaporation is at its potential rate, termed the limit for potential evaporation L_P (mm), and a parameter in the nonlinear function relating runoff generation to the soil moisture state, termed the nonlinearity parameter β (-). Runoff routing in the elevation zones is represented by three reservoirs: the upper and lower zones and a groundwater reservoir. 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 (h); percolation to the lower zone with a constant percolation rate c_p (mm/day); and if a threshold of the storage state L_1 (mm) is exceeded, through an additional outlet controlled by a very fast storage coefficient k_0 (h). Water leaves the lower zone based on a

slow storage coefficient k_2 (h). k_3 (h) controls the outflow from the groundwater storage. Additionally, a bypass flow Q_{by} (mm) is introduced to account for precipitation that bypasses the soil matrix and directly contributes to the storage in the lower soil levels (Blöschl et al., 2008). Outflow from all reservoirs is then routed by a transfer function which consists of a linear storage cascade with the parameters N (-; number of reservoirs) and K (h; time parameter of each reservoir).

Model calibration

(Madsen et al., 2002) have compared different methods of automated and manual calibration to find that the different methods put emphasis on different aspect of the hydrograph, but none of the methods were superior with respect to the performance measures. However, we believe that manual calibration based on hydrological reasoning will yield model parameters that are more suitable for the extrapolation of extreme conditions so the manual calibration was used here as modelling floods was the main interest in this study. This is supported by several studies. E.g., Franchini and Pacciani (1991) have stated that "*it is apparent that between an automatic calibration procedure and a procedure based on successive rational attempts, the latter is preferable as it is the only one which makes it possible to use prior knowledge of the nature of the watershed.*" Ivanov et al. (2004) have stated that "*manual streamflow-based calibration is a stepwise approach that includes analysis of a number of variables considered at different spatial and temporal scales.*" We have applied a similar method affecting the spatial scale as in Ivanov et al. (2004), where each nested basin was calibrated first. Parameter values for the nested catchments were then considered to be fixed and the remaining parts of the catchment were calibrated. The calibration process followed a number of steps (Blöschl et al., 2008): First, an approximation of the annual water balance was sought to be achieved. This was done by setting initial parameters for the snow routine, for the maximum soil moisture storage and for the slow runoff components. In a second step, the initial model parameters were adjusted in order to reproduce seasonal patterns correctly. Threshold temperatures were adjusted, as well as parameters influencing the slow and, if necessary, the groundwater runoff component. The third step included the parameterization of the fast runoff components and the parameters of the linear storage cascade by looking at single flood events as well as a fine tuning of the parameters of the snow and soil moisture routines. The timing of the rising limbs and the peak discharge was sought to be estimated as correctly as possible, as well as the magnitude of the peak discharge. After each model run, we visualised the model simulations and evaluated the results using statistical measures (measures used are given in Appendix A). It showed that the manual calibration had two main advantages. First, the structure of events in different hydrological situations could be captured better by using manual calibration. Second, the timing of the rising limbs of the flood waves could be simulated well. Additionally, the approach has been found to be efficient as looking at a lot of different flood events helped to gain a deep insight into the runoff processes throughout the catchments.
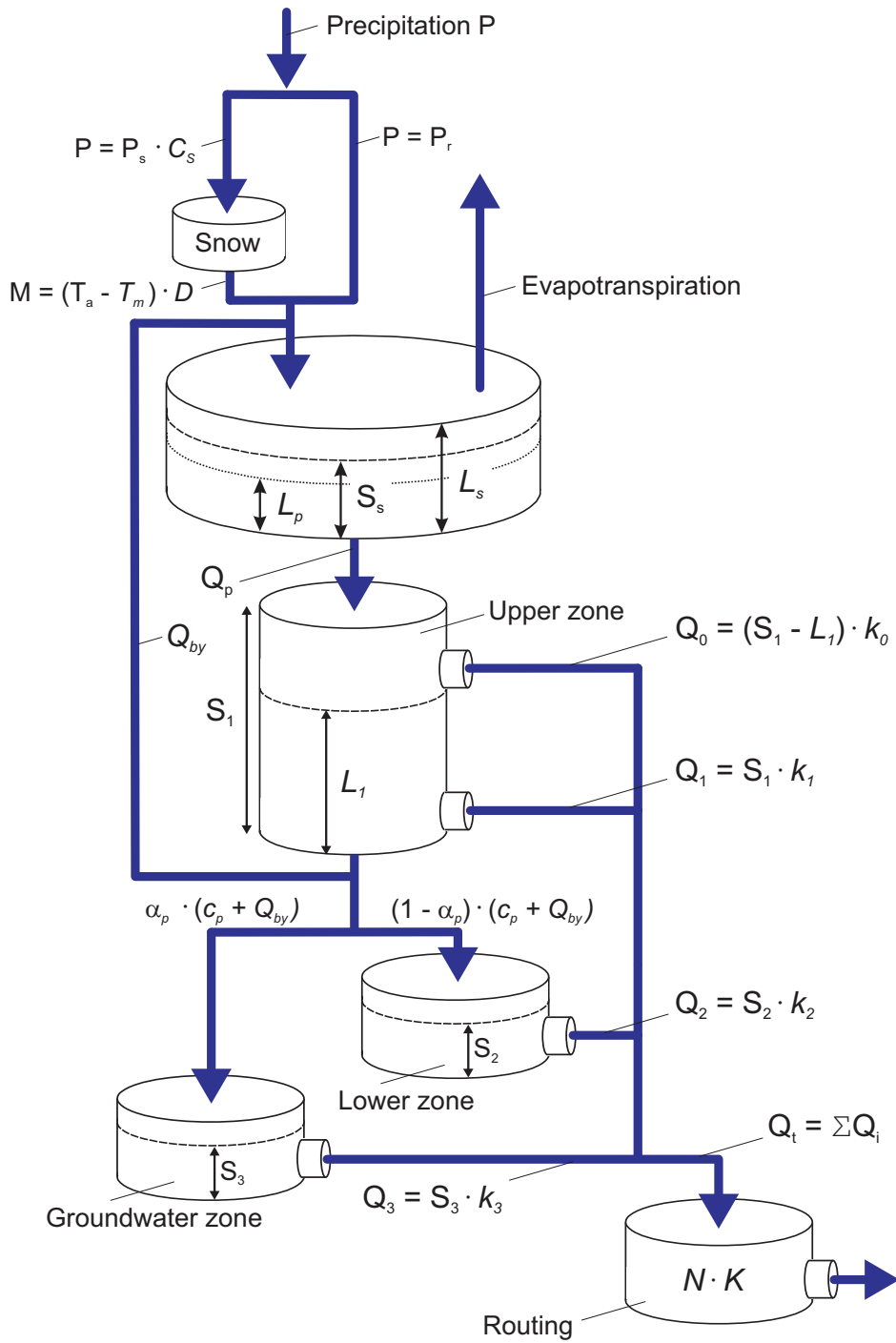
Figure 2.4: Model scheme. Parameters see Appendix B.

In the calibration process, the snow correction factor C_S (-) has been set to a value of 1, as an elevation-based parameterization for precipitation is implemented in the INCA system (Haiden, 2009). However, in one elevation zone ranging from 3250 to 3750 m a.s.l. the model results have shown that a snow correction factor of 1 results in a constantly increasing SWE value for this elevation zone. Therefore, a C_S value of 0.8 has been used in this elevation zone to yield better and more realistic results. Initial model parameters for the calibration of the snow routine were taken from the literature and were adjusted in several model runs. Initial values were taken from, e.g., Seibert (1999) who has used threshold temperatures ranging from -1.5 to 2.5°C and a degree day factor ranging from 1 to 10 mm.°C^{-1}.day^{-1} for his Monte Carlo based calibration in Sweden. A temperature range from -2.0 to 4.0°C is reported in Braun (1985) where a mix of rain and snow can occur in lowland and lower-alpine catchments in Switzerland. Kienzle (2008) proposed threshold temperatures ranging from $T_s = $-4°C to $T_r = 8$°C for Canada. Merz et al. (2009) have not calibrated the threshold temperatures T_s and T_r but have set them to constant values of 0°C and 2°C, respectively, prior to calibration, in Austrian catchments. We have estimated parameters for T_s in the range of -1.8 to -0.4°C and for T_r in the range of 0.8 to 1.6°C in our model area. The upper threshold temperature is well in the range of other studies, and also the lower threshold temperature is in the range of the studies in which the lower threshold temperature has been calibrated. Merz and Blöschl (2004) have estimated the remaining parameters of the snow routine using an automatic algorithm; the upper and lower bounds for the degree day factor were set to 0 and 5 mm.°C^{-1}.day^{-1}, for the snow correction factor the bounds were set to 1.0 and 1.5, and the range of the melt temperature was set to -1.0 and 3.0°C. During rain-on-snow events large melt rates are likely to occur in northern Austria (Sui and Koehler, 2001). This enhanced melting is represented in the model by increasing D (mm.°C^{-1}.day^{-1}) by a factor of 2 if rain falls on an existing snow pack. We used values for D in the order of 1.3 to 2.3 mm.°C^{-1}.day^{-1}, which is in the range of existing studies, even when considering the doubling of the factor in the case of rain on snow.

Table 2.2: Hydrologic model parameters

Model parameter	Description	min in region	max in region
D	Degree day factor (mm.°C^{-1}.day^{-1})	1.3	2.3
T_s	Threshold temperature (°C)	-1.8	-0.4
T_r	Threshold temperature (°C)	0.3	1.6
T_m	Melt temperature (°C)	0.1	0.9
C_S	Snow correction factor (-)	0.8	1.0
L_S	Max. soil moisture (mm)	70	725
L_P	Limit for pot. evaporation (mm)	9.5	360
β	nonlinearity parameter (-)	1.3	4.7
k_0	storage coefficient (h)	0.5	200
k_1	storage coefficient (h)	10	550
k_2	storage coefficient (h)	75	2500
k_3	storage coefficient (h)	100	2500
c_p	constant percolation (mm.day^{-1})	2.2	24

Initial values for the model parameters for change in soil moisture and runoff generation were taken from previous works in the study region (e.g., Merz and Blöschl, 2004; Parajka et al., 2007; Merz et al., 2009). They have used an automated calibration routine; values for the maximum soil moisture L_S are in the range of 0 to 600 mm and the nonlinearity parameter β is varied in the range of 0 to 20 (-); values for storage parameters are in the range of 0 to 2 days for k_0, 2 to 30 days for k_1 and 30 to 250 days for k_2 and the percolation rate c_p is varied between 0 and 8 mm.day^{-1}. Parameters controlling the soil moisture were chosen depending on catchment characteristics (land use, geology) which were analysed prior to the calibration. E.g., in catchments dominated by open land medium values for the maximum soil moisture L_S were used, in catchments dominated by forests, larger values of L_S were used as it was assumed that the storage capacity is higher in forested areas. In alpine catchments small values of L_S were chosen as the storage capacity was assumed to be smaller due to rocks and shallow soil. The storage coefficients were chosen depending on the shape of the catchments. E.g., the fast runoff component of a catchment which is stretched was assumed to be slower (and hence the storage coefficients larger) compared to a more compact catchment which was assumed to react quicker (and hence the storage coefficients being smaller). Table 2.2 gives an overview over the calibrated minimum and maximum parameters values in the region.

For each catchment, the model performance was evaluated by several statistical measures, including (1) the Nash and Sutcliffe (1970) coefficient of efficiency $nsme$, (2) the volume error VE, (3) peak discharge errors pde, (4) mean absolute peak discharge errors $mapde$ and (5) mean absolute peak time errors $mapte$. To identify the dependence of the statistical measures on various catchment attributes, we used Spearman's rank correlation coefficient r_s, a non-parametric measure of statistical dependence between two variables. Partial correlation was used to describe the relationship between two variables whilst taking away the effects of another variable on this relationship. Definitions of the metrics can be found in Appendix A.

2.4 Results

Examples of different runoff regimes

In order to provide a first insight into the runoff model performance for different hydrological regimes we present two example catchments. We denote them as climatologically wet and dry catchments, respectively, but as already mentioned in the introduction, it is realised that the latter would not be considered as a dry catchment in most climate regimes around the world. The wet catchment (Opponitz/Ybbs; gauge number 9 in Fig. 2.3) has a mean annual precipitation of around 1800 mm/yr and mean annual runoff of around 1100 mm/yr. The catchment that is denoted as dry here (Cholerakapelle/Schwechat; gauge number 11 in Fig. 2.3) has a mean annual precipitation of around 890 mm/yr and mean annual runoff of around 250 mm/yr.
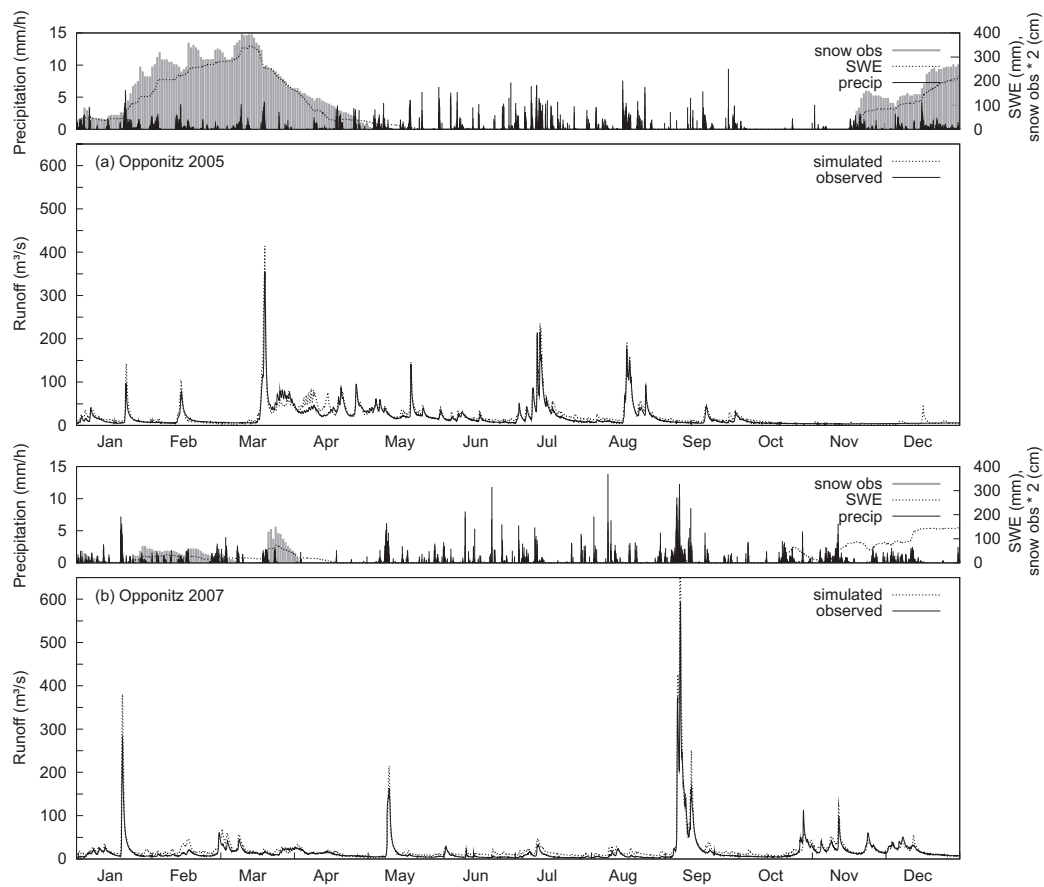
Figure 2.5: Calibration (top) and validation (bottom) results for a climatologically wet catchment: Gauge Opponitz/Ybbs. Mean annual runoff depth is 1100 mm/yr, catchment size is 510 km^2. Gauge number 9 in Figure 2.3. SWE is the snow water equivalent. Snow depth measurement for a single representative station (approx. mean catchment elevation) in the catchment is shown (data only available until September 2007).

Figure 2.5 shows the model results for Opponitz/Ybbs. The area is 510 km^2, elevations range from 500 to 1800 m a.s.l., and 85% of the catchment are covered by forest. The geology is mainly limestone. The top panels (Figure 2.5a) show the year 2005 (calibration period). Simulated snow water equivalent (SWE) is plotted as a dotted line; hourly precipitation is shown as impulses. Overall, the seasonal pattern of runoff is simulated well with an $nsme$ of 0.83, a VE of 0.16 and a $mapde$ of 25.7. The largest observed runoff in 2005 was induced by snow melt and reached around 350 m^3/s in March; in summer the largest runoff was around 220 m^3/s in July and August. The devolution of snow is simulated well; however, in Figures 2.5 and 2.6 we compare model SWE and observed snow depths. The lower panels provide results for the year 2007 which is part of the validation period. There is less snow than in 2005 and the maximum discharge is higher. Again, the seasonal devolution of snow is simulated well at the beginning of the year; towards the end of the year no snow data are available. Three major events were recorded, all of which are somewhat overestimated.

Several storms were recorded over the summer but there is almost no runoff response in the catchment. The timing of the rising limbs is good. $nsme$ for this year is 0.80, VE is 0.07 and $mapde$ is 36.2.

Figure 2.6 provides the results of a drier catchment in the forelands of the Alps (Cholera-kapelle/Schwechat). The area is 180 km^2, elevations range from 280 to 830 m a.s.l. and 77% of the catchment are forested and 21% are meadows. The geology includes Molasse, Flysch and limestone. This means that the catchment is quite different from the catchment in Figure 2.5 in terms of climate and geology. The baseflow is at a low level and short storm events typically cause a quick rise of the runoff which makes the simulations difficult. Also, the rainfall events are shorter and of higher intensity than in Opponitz and there is less snow. The short rainfall events cause the runoff to drop to the level of the baseflow within a short period of time. The simulation results for the calibration period are reasonable, but some of the peaks are overestimated or underestimated. $nsme$ is 0.79, VE is 0.14 and $mapde$ is 34.0. The performance of the model in the validation period is lower in terms of the $nsme$ (0.50), better in terms of the VE (-0.01), but similar in terms of $mapde$ (35.0). The observed hydrograph does not show a lot of variability, and several short small-scale storms do not have a lot of influence on the runoff. At the beginning of September a large scale precipitation event caused fast response. The discharge increased to 110 m^3/s, but the model estimated 165 m^3/s. The comparison of the two catchments (Fig. 2.5 and 2.6) suggests that the flashier runoff pattern in the drier catchment is more difficult to model. Small precipitation events can lead to unexpected runoff response, and rain-on-snow events occur in these prealpine areas (Merz and Blöschl, 2003, 2008). Alpine catchments with a larger elevation range such as Opponitz have a more distinct annual cycle related to snow melt in spring.

Effect of catchment scale on the model performance

To assess the model performance more comprehensively, the model error measures based on hourly data of all catchments have been plotted in Fig. 2.7 against catchment area. Additionally, we have calculated the Spearman's rank correlation coefficients r_s between model error measures and catchment attributes for the entire period and for winter and summer seasons (Table 2.3). Overall, the model performs well. The median Nash-Sutcliffe model efficiencies ($nsme$) for the calibration and validation periods are 0.69 and 0.67, respectively. Median $nsme$ in the summer months (June-November) are 0.69 (calibration) and 0.71 (validation), in the winter months (December-May) the values are 0.64 (calibration) and 0.56 (validation). The distribution of the $nsme$ in the validation period is similar to that in the calibration period. In the validation period, 80% of the $nsme$ values are larger than 0.5 (90% in the calibration period), 41% are larger than 0.7 (46% in the calibration period) and 7% are larger than 0.8 (9% in the calibration period). Catchments with high $nsme$ are mostly large catchments. For the calibration period, 5 out of 57 catchments have an $nsme$ below 0.50. Out of these, four catchments have an area less than 400 km^2. This indicates that there is a trend of increasing model performance with catchment scale. This is confirmed by the Spearman's rank correlation coefficient r_s which is 0.43 for catchment area and $nsme$
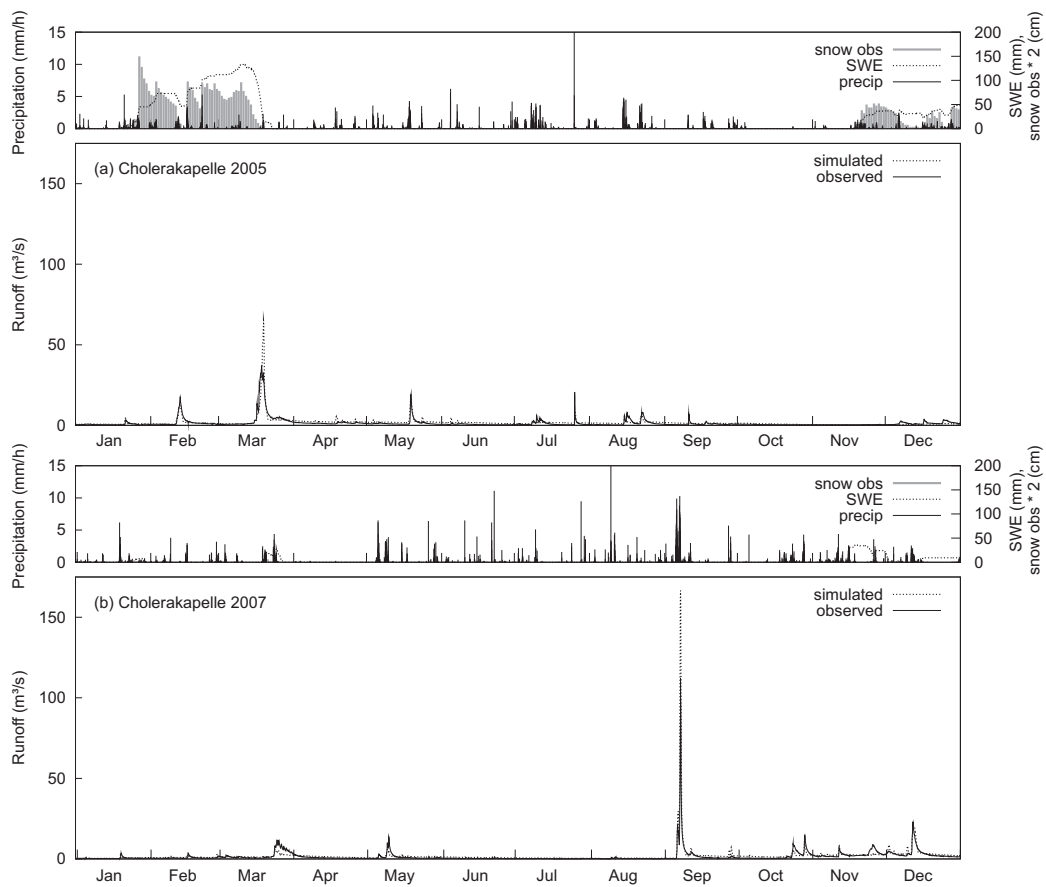
Figure 2.6: Calibration (top) and validation (bottom) results for a dry catchment: Gauge Cholerakapelle/Schwechat. Mean annual runoff depth is 250 mm/yr, catchment size is 181 km². Gauge is labelled as number 11 in Figure 2.3. Snow depth measurement for a single representative station (approx. mean catchment elevation) in the catchment is shown (data only available until September 2007).

for the entire period (Table 2.3). The correlation for summer and winter periods is slightly smaller with 0.37 and 0.38, respectively. Volume errors are in the range of -0.20 to 0.40 for the small and medium sized catchments and in the range of -0.10 to 0.20 for the larger catchments with no trend of increasing model performance and catchment area. There is however a small correlation between the absolute volume error and catchment scale as the errors tend to decrease with catchment scale, especially in the summer period. The mean absolute peak discharge error ($mapde$) is clearly decreasing with catchment scale, which is confirmed with a Spearman's rank correlation coefficient of -0.60 for the entire period. Again, the correlation is much stronger in the summer period. The trend for the percent absolute peak time error ($mapte$) to decrease with catchment scale is not as distinct as for mapde. Interestingly, the correlation for mapte is stronger in the winter periods. This may be due to the fact that flash floods mainly occur in the summer months.

Table 2.3: Correlations of Nash-Sutcliffe model efficiencies ($nsme$), volume error (VE), mean absolute peak discharge error ($mapde$), mean absolute peak time error ($mapte$) and absolute volume error ($|VE|$) to mean annual runoff MAR (mm.yr^{-1}), mean annual precipitation MAP (mm.yr^{-1}), ratio rain/precipitation and catchment area (km^2) for different seasons. Winter is December to May (runoff influenced by snow and/or snow melt), summer is June to November. Correlation coefficients that are significant at the 95% level are printed in bold.

		nsme			VE			mapde			mapte			|VE|		
		calib	valid	total	calib	valid	total	calib	valid	total	calib	valid	total	calib	valid	total
TOTAL	MAR	**0.39**	**0.46**	**0.38**	0.17	-0.03	0.12	**-0.34**	-0.24	**-0.42**	**-0.37**	-0.08	**-0.35**	-0.23	-0.10	-0.19
	MAP	**0.32**	**0.46**	**0.31**	0.18	0.02	0.14	**-0.33**	-0.25	**-0.40**	**-0.42**	-0.13	**-0.38**	-0.09	-0.05	-0.08
	rain/precip	**-0.32**	**-0.34**	**-0.45**	0.06	0.22	0.16	**0.44**	**0.57**	**0.55**	**0.36**	**0.36**	**0.43**	0.19	**0.48**	**0.28**
	area	**0.40**	**0.44**	**0.43**	-0.04	-0.06	-0.06	**-0.57**	**-0.31**	**-0.60**	-0.22	**-0.01**	**-0.27**	-0.08	-0.12	-0.24
SUMMER	MAR	**0.57**	**0.47**	**0.55**	-0.02	-0.23	-0.07	**-0.33**	-0.11	**-0.43**	-0.14	0.08	-0.17	**-0.36**	-0.21	**-0.34**
	MAP	**0.62**	**0.45**	**0.57**	-0.03	0.18	-0.06	**-0.33**	-0.12	**-0.41**	**-0.24**	0.04	-0.23	-0.23	-0.15	**-0.28**
	rain/precip	**-0.49**	-0.18	**-0.46**	0.16	**0.53**	0.23	**0.42**	**0.33**	**0.52**	0.20	0.23	**0.30**	**0.33**	**0.59**	**0.41**
	area	**0.39**	**0.36**	**0.37**	-0.11	-0.10	-0.06	**-0.52**	-0.19	**-0.58**	-0.19	-0.02	-0.22	-0.21	-0.09	**-0.27**
WINTER	MAR	**0.22**	**0.30**	0.25	0.24	0.00	0.17	-0.15	-0.05	-0.15	**-0.38**	-0.17	**-0.35**	-0.03	-0.16	-0.22
	MAP	0.10	**0.30**	0.13	**0.30**	0.06	0.22	-0.19	-0.10	-0.20	**-0.42**	-0.17	**-0.34**	0.07	-0.22	-0.11
	rain/precip	-0.01	-0.09	-0.18	0.00	0.13	0.10	**0.31**	**0.5**	**0.46**	**0.39**	**0.34**	**0.43**	-0.01	**0.45**	0.19
	area	**0.27**	**0.35**	**0.38**	-0.04	-0.17	-0.08	**-0.40**	-0.05	**-0.37**	**-0.35**	-0.04	**-0.30**	0.03	-0.09	-0.16
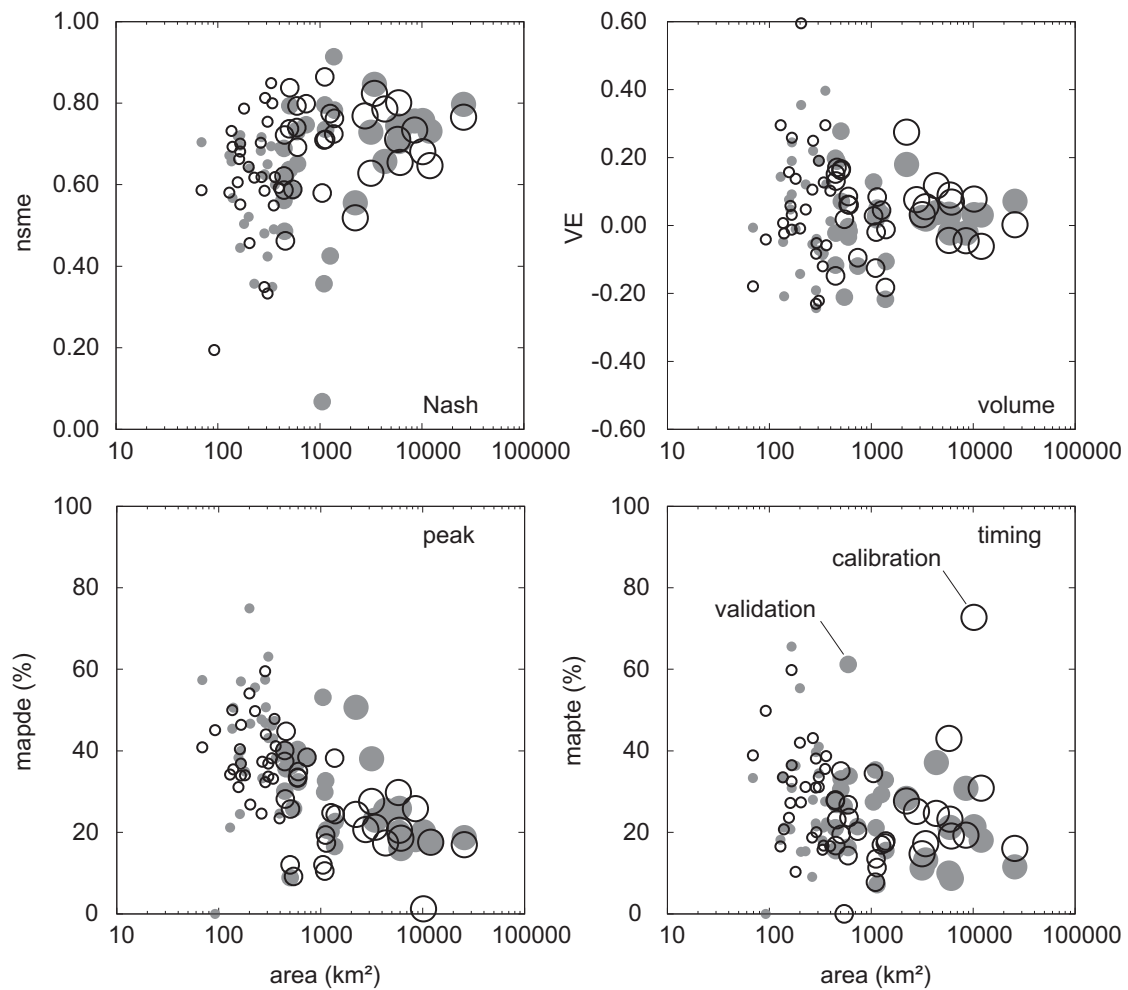
Figure 2.7: Nash-Sutcliffe model efficiency ($nsme$), volume error (VE), mean absolute peak discharge error ($mapde$) and mean absolute peak time error ($mapte$) plotted versus catchment area. The marker sizes indicate the catchment size. The open circles relate to the calibration period, the full circles to the validation period.

For a more detailed analysis we have chosen the event in June 2009 which was the largest event in the validation period, not only in terms of runoff but also in precipitation. Statistical analyses have shown that in the affected area 226% of the normal precipitation in June has been recorded (BMLFUW, 2009). Specific discharges are shown for ease of comparison. Results are shown in Figure 2.8 for nine of the 57 catchments for which (1) observed runoff data were available and (2) which were affected most by the storm. Precipitation ranged from 70 to 167 mm in 5 days (Table 2.4). The flood return periods in the region ranged from 2 to > 50 years with the largest values at the Ybbs and Traun. The model results suggest that the response of the catchments to the large scale precipitation in this case can be simulated well. The event was mainly produced by large scale precipitation caused

by a low pressure area in Italy. Throughout the catchments along the Northern rim of the Alps, precipitation started on June 22, raining nearly continuously until June 24. Embedded in this large scale precipitation were local convective heavy rain storms (BMLFUW, 2009; Haiden, 2009). Within 12 hours, the water levels at the gauges started to rise. Intensities increased for a few hours and it continued to rain for another day at low intensities. The main peaks in all nine catchments are simulated well, also the timing of the rising limbs. However, the first increase of this event is underestimated at the gauges Wels, Molln, Steyr and Greimpersdorf. This is likely due to local convective heavy rain storms (BMLFUW, 2009; Haiden, 2009) close to the outlet of the catchments which have not been captured by the rain gauges and the radar. The undercatch has two effects: First, areal mean precipitation is underestimated. Second, the response times of the catchments in this event are shorter than if uniform precipitation occurred which is not captured well.

Table 2.4: Catchment precipitation totals for the event June 22-26, 2009 shown in Figure 2.8.

Gauge	Number in Figure 2.3	Precipitation (mm)
Rosenheim/Mangfall	1	85
Schärding/Inn	2	70
Haging/Antiesen	3	104
Fraham/Innbach	4	98
Wels/Traun	5	130
Molln/Steyrling	7	163
Steyr/Enns	6	109
Greimpersdorf/Ybbs	8	167
Imbach/Krems	10	110

As the model has been designed for flood forecasting, particular interest resides in the performance of simulating flood peaks. For every gauge included in the model flood peaks were identified. Observed and simulated peak discharges were compared as well as the time from the beginning of the rising limb to the peak.

In the left panels of Figure 2.9, simulated peaks are plotted against the observed peaks. On this double logarithmic scale there is good correspondence. However, there is a tendency in both calibration and validation periods for peak discharges less than 100 m^3/s to be underestimated. On the other hand, peak discharges larger than 100 m^3/s are fairly well estimated with some overestimation for the largest peaks. In the right panels of Figure 2.9 the cumulative distribution functions of the peak discharge errors (pde) are plotted. For the calibration period, the cdfs of the small and medium sized catchments indicate that 75% of the peaks are underestimated and 25% are overestimated while in the large catchments around 60% of the peaks are underestimated. The peak discharge errors are in the range of -0.9 to 1.5. In the validation period, 70% of the peaks in small catchments are underestimated which is similar to the calibration period; 50% of the peaks in medium sized catchments and 40% of the peaks in large catchments are underestimated. This means that, when moving from calibration to validation, there is a shift from underestimation to overestimation with increasing catchment size. This is mainly due to a slight overestimation during the low flow periods. As the system states prior to the events are of great importance for the performance
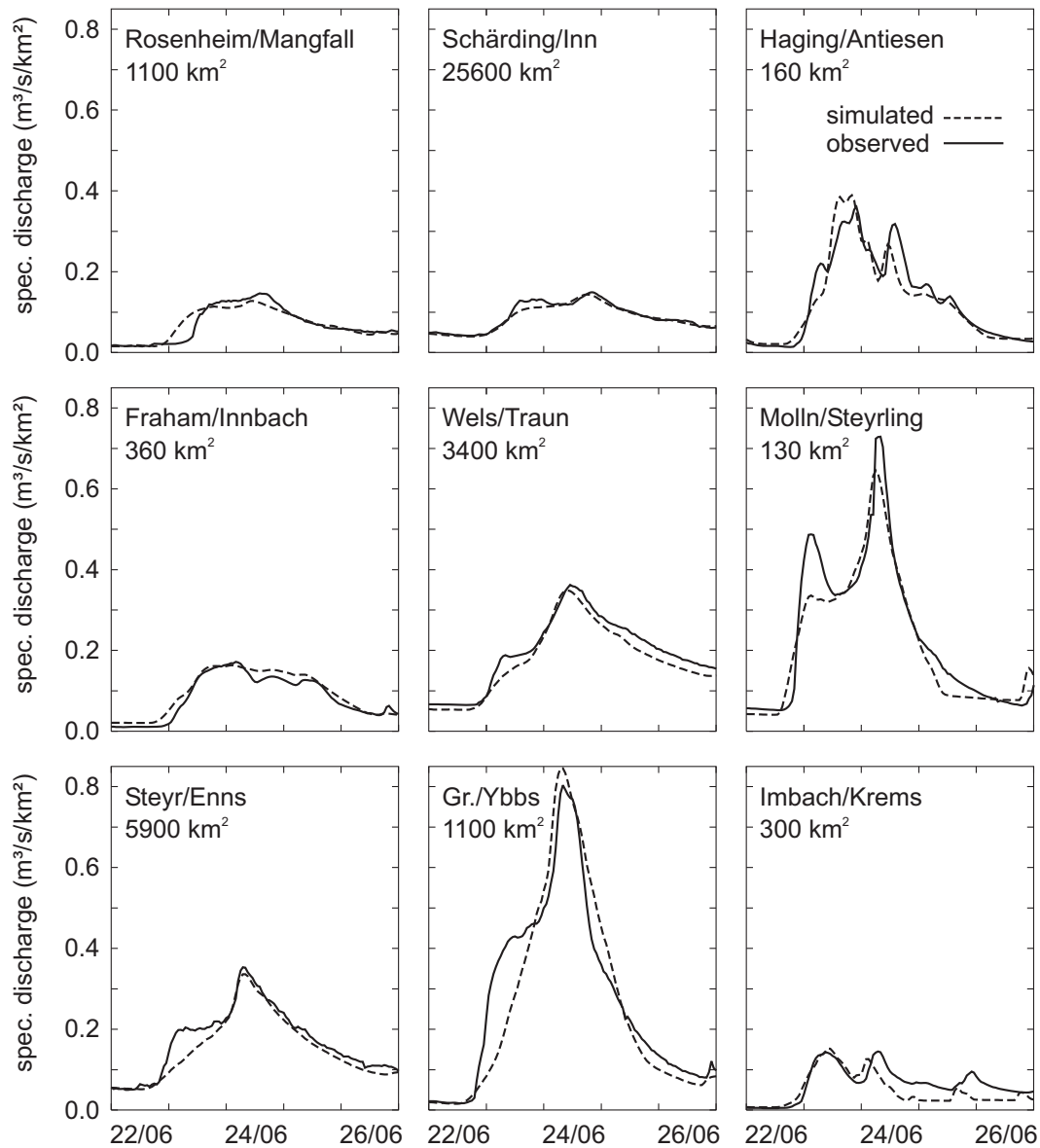
Figure 2.8: Event hydrographs for the June 2009 event during the validation period. Gr./Ybbs stands for Greimpersdorf/Ybbs.

of the model during the event, an overestimated baseflow tends to lead to an overestimation of the peak.

Table 2.5 gives a summary of error statistics with and without seasonal distinction. The $nsme$ values have a tendency to increase with increasing catchment scale whereas $mapde$ clearly decreases with increasing catchment scale, both for the calibration and validation periods. These trends can be observed in the seasonal statistics as well. Interestingly, in winter the $nsme$ of the small catchments is better for the calibration period compared to the
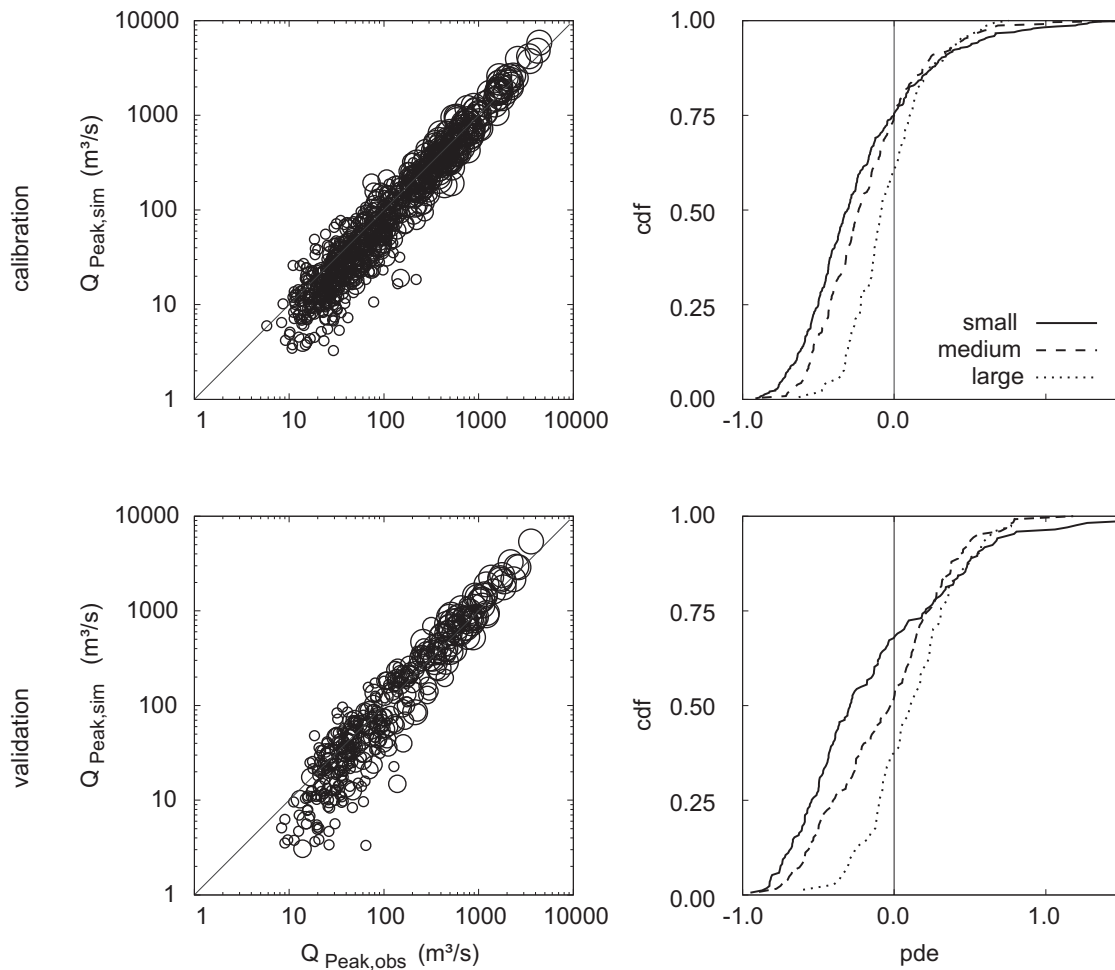
Figure 2.9: Scatter plots of $Q_{obs,peak}$ vs. $Q_{sim,peak}$ (left side) and cdf of peak discharge errors (pde) (right side). Size of markers represents catchment size. Calibration (top), validation (bottom). The line types in the right panels relate to three groups of catchment sizes (< 400 km^2, 400 to 1890 km^2, and > 1890 km^2).

value in summer and mapde is better in winter for small and medium catchments. This can be attributed to the small flashy events in summer, which are simulated with less accuracy than the large events. There is no real trend for the volume error. The magnitude of the peak errors seems to be large. However, they are perfectly in the range of existing studies. Senarath et al. (2000) have given values for the average absolute peak discharge error in the range from 32 to 66%; Reed et al. (2004) have shown that the percent absolute peak error ($pape$) is typically on the order of 30 to 50% for calibrated models with much larger errors for the smallest basins. Similarly, Reed et al. (2007) report percent absolute peak errors from 24 to 88% with increasing errors with decreasing catchment size. Modarres (2009) gives values of 11 to 41% for the medium absolute percentage error of peak discharges.

Table 2.5: Error statistics of runoff: Nash-Sutcliffe model efficiency ($nsme$), volume error (VE), mean absolute peak discharge errors ($mapde$) and mean absolute peak timing error ($mapte$). A is catchment area, N the number of catchments.

Average

A (km²)	N	$nsme$ calib	$nsme$ valid	$nsme$ total	VE calib	VE valid	VE total	$mapde$ calib	$mapde$ valid	$mapde$ total	$mapte$ calib	$mapte$ valid	$mapte$ total
Small catchments <400	27	0.62	0.60	0.62	0.06	0.04	0.05	38	43	42	30	27	30
Medium catchments 400-1890	18	0.71	0.63	0.67	0.03	0.01	0.01	27	31	31	20	27	24
Large catchments >1890	13	0.71	0.73	0.70	0.05	0.03	0.04	20	25	23	28	19	22
All catchments 70-25600	57	0.66	0.63	0.65	0.05	0.03	0.03	31	36	34	26	25	26

Median

A (km²)	N	$nsme$ calib	$nsme$ valid	$nsme$ total	VE calib	VE valid	VE total	$mapde$ calib	$mapde$ valid	$mapde$ total	$mapte$ calib	$mapte$ valid	$mapte$ total
Small catchments <400	27	0.62	0.64	0.64	0.04	0.01	0.03	37	46	42	31	22	30
Medium catchments 400-1890	18	0.72	0.67	0.69	0.05	0.01	0.03	27	33	34	19	27	23
Large catchments >1890	13	0.72	0.74	0.73	0.06	0.03	0.04	21	20	21	24	18	25
All catchments 70-25600	57	0.69	0.68	0.67	0.05	0.02	0.03	33	37	35	24	22	25

Climate effects on model performance

The results shown so far have indicated that the model performance depends on both the catchment size and the wetness of the catchments. To provide further insights into these findings, the model performance indices were plotted in Figure 2.10 against the mean annual precipitation (MAP). While there is a lot of scatter in the diagrams, they also indicate interesting patterns. The lowest $nsme$ only occur for the driest catchments which are also among the smallest catchments. The correlation coefficient between $nsme$ and MAP is in the range of 0.31 for the entire period 2003-2009, with a larger correlation in summer (0.57) and a lower correlation in winter (0.13) (see Table 2.3). The VE ranges between -0.25 and 0.25 with a few outliers. Correlation coefficients between VE and MAP are small. The range of VE is somewhat smaller for the catchments with higher mean annual precipitation, suggesting that the model performance in these catchments is slightly better. The generally rather large volume errors are due to the fact that the model calibration was guided by an attempt to simulate the peaks well, as the main purpose of the study was flood forecasting. Also, hydrologically realistic parameters were given preference over minimizing biases with a view of representing extreme events well. There is quite a clear tendency for the peak errors in both terms of maximum discharge ($mapde$) and time to peak ($mapte$) to decrease with mean annual precipitation which is consistent with the other error measures. Interestingly, correlation coefficients for mapde and mapte are quite different for the winter and summer periods. For $mapte$, the correlation coefficients are higher in winter, indicating that the timing of the peaks can be simulated better in winter. This is no surprise as the flashier events which are more difficult to simulate are more likely to occur in summer. For $mapde$, the correlation coefficients are higher in summer, indicating that the peaks are simulated somewhat better in summer. For comparison, the performance measures have been plotted against the mean annual runoff (MAR) in Figure 2.11. The patterns are similar to those in Figure 2.10, and also the correlation coefficients are similar.

As the smallest catchments also tend to be among the drier catchments and the larger catchments tend to have more snow, we have calculated the partial correlation coefficient based on the Spearman's rank correlation coefficient rs to separate the two effects. Correlation coefficients and partial correlation coefficients variables are summarized in Table 2.6. Taking away the effects of the climate related variable (MAP, MAR, rain/precip) decreases the correlation between $nsme$ and area on the order of 20 to 36% whereas taking away the effects of the area causes a decrease on the order of 30 to 55% of the correlation between $nsme$ and the climate variable. This indicates that the catchment area has a strong impact on the correlation coefficients. Similar results are obtained for the remaining correlations as well. The correlation between mapde and area decreases on the order of 15 to 25% when taking away the effects of the climate related variable. It shows that the mean annual precipitation MAP has the least influence on the correlations, whereas the ratio rain/precip has the largest impact on the correlations. On the contrary, taking away the effects of the climate related variables on the correlations between |VE| and area and mapte and area has a larger influence on the correlations; however, for |VE|, area and MAP there is no difference
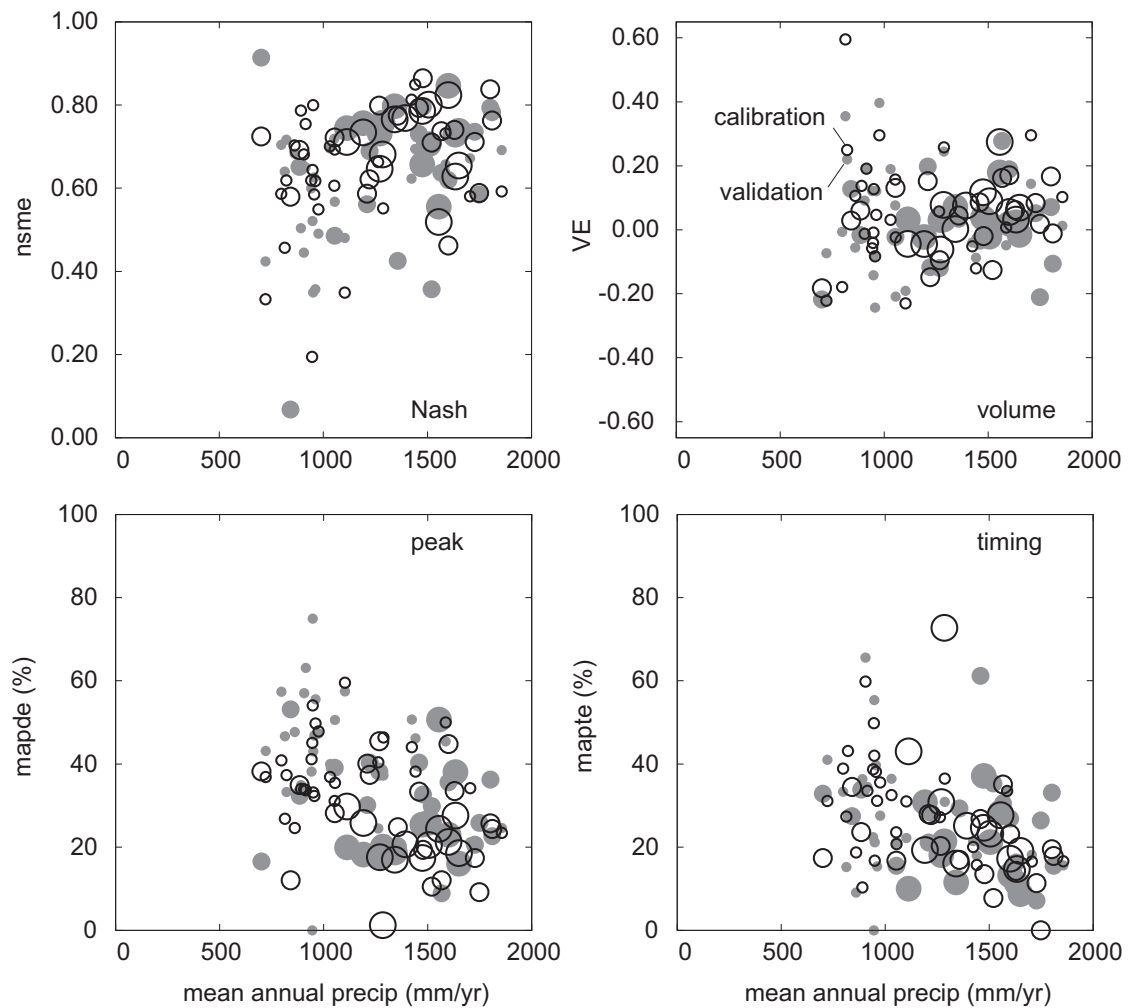
Figure 2.10: Nash-Sutcliffe model efficiency ($nsme$), volume error (VE), mean absolute peak discharge error ($mapde$) and mean absolute peak time error ($mapte$) plotted versus mean annual precipitation. The marker sizes indicate the catchment size. The open circles relate to the calibration period, the full circles to the validation period.

between the correlations. The influence of the ratio rain/precip has the largest influence on the correlations between timing error and area.

Figure 2.12 shows the effects of another climate related variable, the ratio of long term liquid precipitation (rainfall) and total precipitation. The larger the ratio, the more precipitation falls as rain, so the catchments are dominated by rainfall runoff processes rather than by snowmelt. Figure 2.12 also indicates that the snow dominated catchments tend to be larger than the rainfall dominated catchments. This indicates that catchment area is a stronger control on model performance than are the snow processes per se. However, the $nsme$ values tend to decrease with increasing long-term ratio of rainfall and precipitation from 55
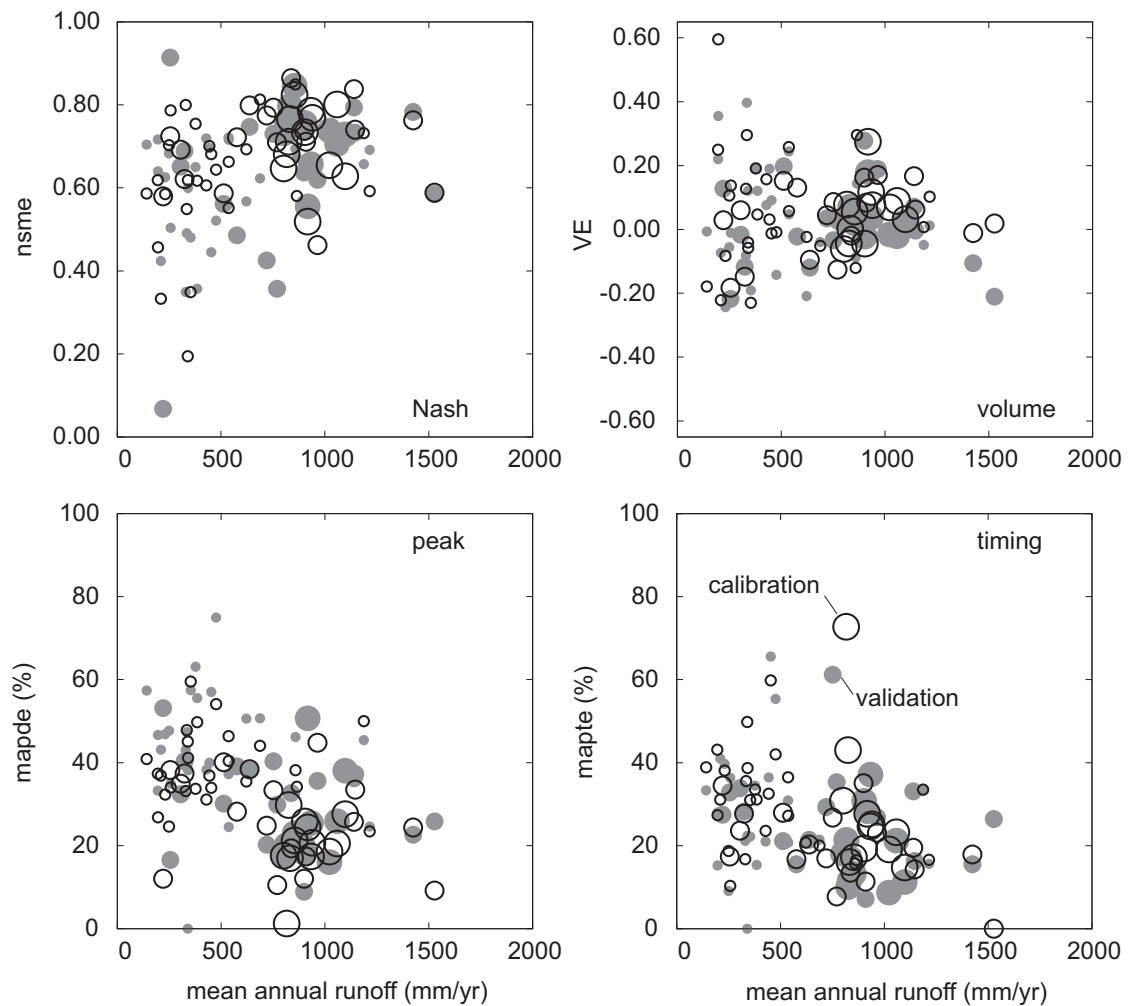
Figure 2.11: Nash-Sutcliffe model efficiency ($nsme$), volume error (VE), mean absolute peak discharge error ($mapde$) and mean absolute peak time error ($mapte$) plotted versus mean annual runoff. The marker sizes indicate the catchment size. The open circles relate to the calibration period, the full circles to the validation period.

to 90%, with a peak in model performance for a ratio of 75 to 80%. The Spearman's rank correlation between $nsme$ and the ratio rain/precip is -0.45 for the entire period, with better correlation in the summer. However, when taking away the effects of the catchment area on the correlation as shown in Table 2.6, the correlation decreases to -0.31. In catchments where 55% of the precipitation falls as rain the $nsme$ is on the order of 0.70. These catchments are the large mountainous catchments in the West of the study region. The model performs best in catchments where 75 to 80% of the total precipitation is recorded as liquid rain with $nsme$ values on the order of 0.70 to 0.85. These catchments comprise mainly Alpine catchments with a mean annual precipitation of more than 1500 mm/yr. obviously, there is a much larger variability in $nsme$ values in rain dominated catchments. In catchments

where more than 85% of the precipitation is rain (15% snow) $nsme$ is ranging from 0.20 to 0.70 in the calibration period and from 0.06 to 0.70 in the validation period. These catchments are mostly small catchments situated in the Eastern part of the model region and along the Northern rim of the Alps where elevation does not change much. Hence, the large variability in model error measures could be related to the use of the 500 m elevation zones which might be too coarse to simulate runoff in these catchments appropriately. The results for the volume error are similar with a smaller variability of the error measures in snow dominated catchments compared to rainfall dominated catchments. The correlation between the ratio rain/precip and the absolute volume error is significantly larger in summer with a larger variability in the smaller, rain dominated catchments. The peak discharge errors mapde show a consistent trend with larger errors in rainfall dominated catchments. This is confirmed by Spearman's r_s of 0.55 with similar values for both summer and winter periods. Peak timing errors mapte show a similar behaviour. All this can be attributed to the fact that flash floods mainly occur in small catchments in summer. Table 2.7 summarizes the model error measures as a function of the ration rain/precip. Values are given for separated summer and winter periods and for the entire period, respectively. It shows that in rain dominated catchments the error measures for the winter period are somewhat better; however, in the validation period $nsme$ and $mapde$ obtain better values in the summer. In rain and snow dominated catchments the summer statistics are somewhat better in both calibration and validation periods; in the snow dominated catchments the winter statistics are better in terms of $nsme$ and $mapde$.

Table 2.6: Partial correlation coefficient based on Spearman's rank correlation coefficient r_s from Table 2.3.

| X	Y	Z	r_{XY}	r_{XZ}	r_{YZ}	$r_{XY,Z}$	$r_{XZ,Y}$
$nsme$	area	MAP	0.43	0.31	0.45	0.34	0.14
$nsme$	area	MAR	0.43	0.38	0.50	0.33	0.22
$nsme$	area	rain/precip	0.43	-0.45	-0.49	0.27	-0.31
|VE|	area	MAP	-0.24	-0.08	0.45	-0.23	0.03
|VE|	area	MAR	-0.24	-0.19	0.50	-0.23	-0.08
|VE|	area	rain/precip	-0.24	0.28	-0.49	-0.12	0.19
$mapde$	area	MAP	-0.60	-0.4	0.45	-0.51	-0.16
$mapde$	area	MAR	-0.60	-0.42	0.50	-0.49	-0.13
$mapde$	area	rain/precip	-0.60	0.55	-0.49	-0.45	0.35
$mapte$	area	MAP	-0.27	-0.38	0.45	-0.12	-0.31
$mapte$	area	MAR	-0.27	-0.34	0.50	-0.12	-0.25
$mapte$	area	rain/precip	-0.27	0.43	-0.49	-0.08	0.38

Table 2.7: Error statistics of runoff: Nash-Sutcliffe model efficiency ($nsme$), volume error (VE), mean absolute peak discharge errors ($mapde$) and mean absolute peak timing error ($mapte$). R/P is the long-term ratio of rainfall and total precipitation, N the number of catchments.

Average

	R/P (-)	N	$nsme$ calib	$nsme$ valid	$nsme$ total	VE calib	VE valid	VE total	$mapde$ calib	$mapde$ valid	$mapde$ total	$mapte$ calib	$mapte$ valid	$mapte$ total
Rain dominated catchments	>0.85	17	0.60	0.56	0.57	0.00	0.03	0.00	38	42	42	26	20	23
Rain and snow dominated catchments	0.65-0.85	36	0.71	0.67	0.69	0.08	0.03	0.06	28	33	31	25	26	25
Snow dominated catchments	<0.65	4	0.67	0.73	0.69	-0.05	-0.02	-0.05	30	29	29	30	26	30
All catchments	0.55-0.92	57	0.66	0.63	0.65	0.05	0.03	0.03	31	36	34	26	25	26

Median

	R/P (-)	N	$nsme$ calib	$nsme$ valid	$nsme$ total	VE calib	VE valid	VE total	$mapde$ calib	$mapde$ valid	$mapde$ total	$mapte$ calib	$mapte$ valid	$mapte$ total
Rain dominated catchments	>0.85	17	0.62	0.56	0.61	0.03	0.04	0.01	40	43	40	23	20	25
Rain and snow dominated catchments	0.65-0.85	36	0.74	0.69	0.70	0.08	0.02	0.04	27	34	34	20	21	22
Snow dominated catchments	<0.65	4	0.67	0.74	0.69	-0.04	0.00	-0.04	29	29	28	31	22	29
All catchments	0.55-0.92	57	0.69	0.68	0.67	0.05	0.02	0.03	33	37	35	24	22	25
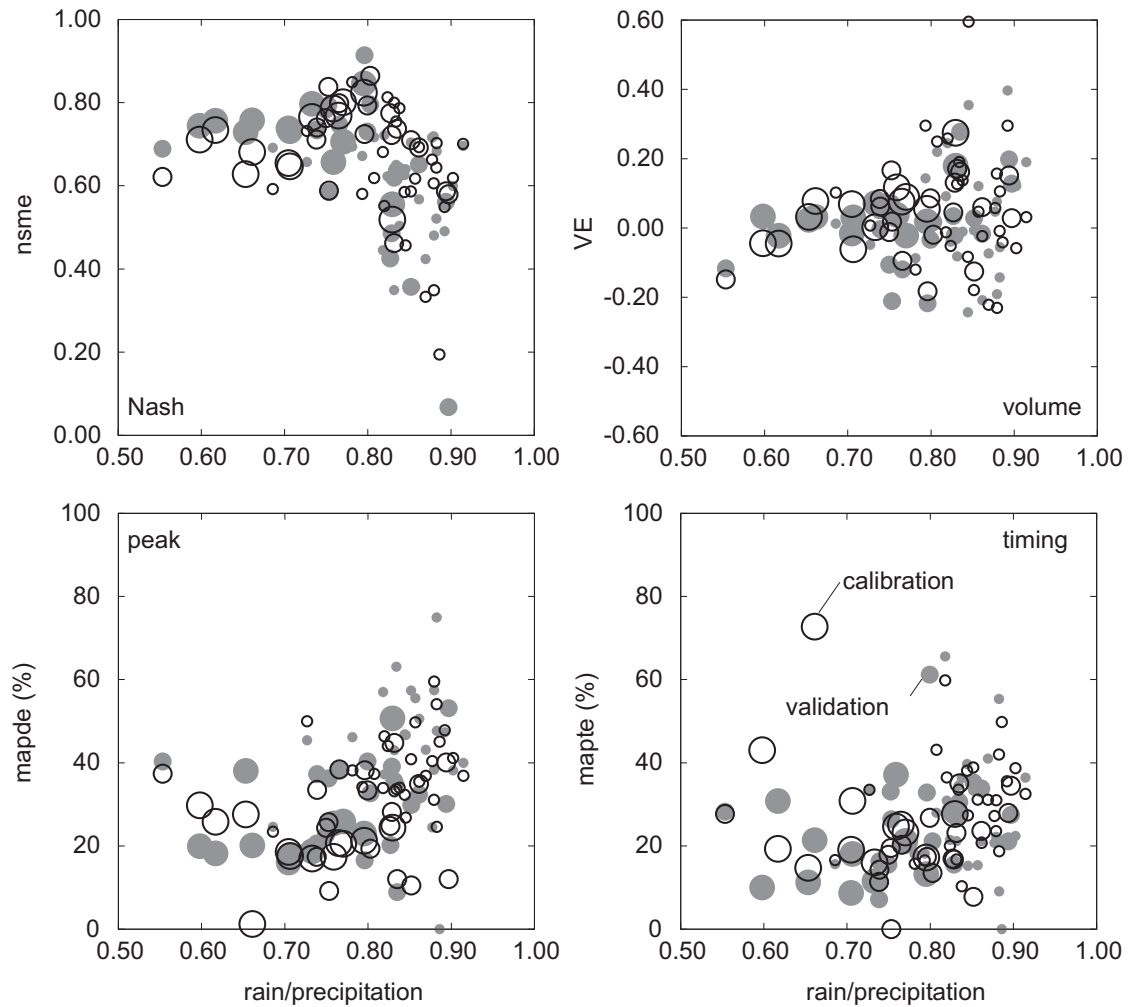
Figure 2.12: Nash-Sutcliffe model efficiency ($nsme$), volume error (VE), mean absolute peak discharge error ($mapde$) and mean absolute peak time error ($mapte$) plotted versus the long-term ratio of rainfall and total precipitation. The marker sizes indicate the catchment size. The open circles relate to the calibration period, the full circles to the validation period.

2.5 Discussion

When relating the results of this study to the literature it is important to note that we have used a simulation time step of 1 hour while most other studies on model performance have used a time step of 1 day. Das et al. (2008) note that the model performance increases with the aggregation time step which is consistent with averaging effects (Skøien et al., 2003). It is interesting to see what exactly the effect for the present case study is: The hourly model results were hence aggregated to daily time steps and nsme was revaluated. On average, the difference between nsme on the hourly and daily time step is on the order

of 0.05 with the largest difference in the smallest catchments (about 0.10), meaning that a $nsme$ of 0.70 (based on hourly data) would be equivalent to a $nsme$ of 0.80 (based on daily data). The larger differences in small catchments would be expected as these catchments tend to have flashier response and aggregating to a daily time step averages out some of this variability (and their errors) causing the $nsme$ to increase. Overall, the performance of the model found here is similar to the performance reported in other studies (e.g., Parajka et al., 2007; Das et al., 2008). The small catchments with areas less than 400 km^2 have median $nsme$ around 0.63 (0.68 on a daily time step) for the calibration and validation periods, and the largest catchments have median $nsme$ around 0.73 (0.76 on a daily time step). The median $nsme$ model performance of Merz et al. (2009) with a water balance model on a daily time step was 0.70 and 0.80 for groups of small and large catchments in Austria, respectively, which is similar to the efficiencies found here. Large peak errors are averaged out in daily $nsme$, but only 8% of the peak errors found here were larger than 70% and 25% of the peak errors were larger than 50%. The magnitude of the peak discharge errors pde is large for a few events with a maximum pde of about 160%. Similar results have been shown in Reed et al. (2004), who have obtained pde of 30 to 50% for half of the events analysed, and only 10% of the events had pde larger than 70%. The peak errors found here are larger in the small catchments with a tendency of underestimation whereas the absolute peak errors decrease with increasing catchment size. There are three main reasons for the increasing model performance with catchment scale: (1) the averaging effects as discussed by Sivapalan (2003) and Skøien and Blöschl (2006), (2) the decreasing variability in streamflow with increasing catchment scale as discussed by Reed et al. (2004), and (3) the increasing number of precipitation stations per catchment (0.35 stations per 100 km^2 in small catchments and 0.45 stations per 100 km^2 in large catchments) as discussed by Merz et al. (2009) which allow to better estimate catchment precipitation in the larger catchments.

The climatological wetness of the catchments also seems to be an important control on model performance. Wetness was evaluated in terms of mean annual precipitation (MAP) and mean annual runoff (MAR) in this study. The performance increased both with increasing MAP and MAR. For catchments with MAP of more than 1500 mm/yr, $nsme$ was around 0.70 while for the drier catchments with MAP around 1000 mm/yr the $nsme$ varied significantly and was on the order of 0.55. Lidén and Harlin (2000) presented similar values. The performance measure they used is $R_V^2 = nsme - 0.1 \cdot |nme|$. For the wet catchments they obtained R_V^2 = 0.80, and for the drier catchments their values were around 0.60. However, in the case of MAP and MAR the influence of the catchment area on the model errors is not negligible as MAP and MAR are correlated with the catchment area.

The analyses in this paper have shown that snow dominated catchments can be modelled somewhat better than rain dominated regimes. Catchments following a distinct annual hydrological cycle with snow accumulation and snow melt phases (and hence a lower ratio of long term liquid precipitation to total precipitation) tend to have a better model performance in terms of all the measures examined here. However, the snow dominated catchments in this study are also among the larger catchments and hence the influence of the catchment area again has to be considered (see Table 2.6). Merz et al. (2009) found $nsme$ around 0.78

for a ratio of liquid to total precipitation of 0.5, and around 0.60 for a ratio of 0.9, based on daily values. In this study, the corresponding numbers are 0.74 and 0.56 (medians in Table 2.7). Similar differences between snow and rainfall dominated regimes were found by Braun and Renner (1992) in Switzerland where the snow dominated high-alpine catchments had $nsme$ around 0.90, while rain dominated lowland catchments had $nsme$ from 0.66 to 0.80.

2.6 Conclusions and outlook

The simulation results indicate that the model performance in terms of all performance indices tends to increase with catchment size, mean annual precipitation, and mean annual runoff and the long-term ratio of snowfall and precipitation which is confirmed by the correlation coefficients. However, the latter are mainly due to the fact that there is a correlation between catchment size and the climatological indices, indicating that the catchment size is the most important control on model performance. The calibration and validation results are consistent in terms of these controls on model performance.

This study is based on observed meteorological data. Additional uncertainty will come in if rainfall forecasts are used (Blöschl et al., 2008). As the model presented in this study has been designed as a part of an operational forecasting system the total forecasting performance and its controls are also of interest. It is planned to examine these in more detail in the context of ensemble flood forecasting.

3 Evaluating the snow component of a flood forecasting model

Abstract

The objective of this study is to evaluate the snow routine of a semi-distributed conceptual water balance model calibrated to streamflow data alone. The model is used for operational flood forecasting in 57 catchments in Austria and Southern Germany with elevations ranging from 200 m a.s.l. to 3800 m a.s.l. We compared snow water equivalents (SWE) simulated by the hydrologic model with snow covered area (SCA) derived from a combined product of MODIS (version 5) Terra and Aqua satellite data for the period 2003-2009 using efficiency measures and a spatial analysis. In the comparison, thresholds for percent catchment snow cover and a cut-off water equivalent need to be chosen with care as they affect the snow model efficiency. Results indicate that the model has a tendency to underestimate snow cover in prealpine areas and forested areas while it performs better in alpine catchments and open land. The spatial analysis shows that for 88% of the analysed model area snow cover is modelled correctly on more than 80% of the days. The space borne snow cover data proved to be very useful for evaluating the snow model. We therefore suggest that the snow data will be similarly useful for data assimilation in real time flood forecasting.

3.1 Introduction

Recent flood events in Austria such as the 2002 flood in the Danube basin have raised the public awareness for the need for flood warnings to reduce damage to property and life. Following these floods, operational flood forecasting systems have been developed for most rivers in Austria. These include the Kamp river (Blöschl et al., 2008), the Inn River (Kirnbauer and Schönlaub, 2006), and the Mur River (Schatzl and Ruch, 2006).

The challenge of operational forecasting systems is the need for simple and robust, yet accurate, routines that can be used with a limited amount of real-time data. The forecasting is thus particularly difficult in mountainous regions because of the large spatial variability of hydrologic characteristics and the limited availability of ground based hydrologic data. As the prediction of streamflow depends on the accuracy of input data and the state variables of the model, it is important to estimate state variables such as soil moisture and snow water

equivalent well. Recent studies suggest that remote sensing data may be valuable for validating the snow component of hydrological models and assimilating them into forecast models. To evaluate snow models different remote sensing products have been used, especially in alpine regions where forests do not obstruct the detection of snow. Blöschl and Kirnbauer (1992) obtained snow cover patterns from aerial photographs and used them to validate a snow model (Blöschl and Kirnbauer, 1991). From the differences between simulated and observed patterns they evaluated the effects of radiation and wind transport on the snow distribution. Blöschl, G. (2002) used SPOT XS satellite data to reduce the biases in simulated snow water equivalent (SWE). Garen and Marks (2005) found good agreement between the temporal evolution of simulated and satellite snow covered area (SCA) for three snowmelt seasons in a basin in Idaho, but they also found that the satellite data underestimated SCA in forested areas. Koboltschnig et al. (2008) used LANDSAT TM and ASTER L1B data to compare simulated and snow covered area in a glaciated catchment in the Austrian Alps. Results showed that the model overestimated the observation by 1-9% in June and July and by 10-36% in August and September which they contributed to redistribution of snow by wind or avalanches not included in the model. Schöber et al. (2010) used LANDSAT images to evaluate SCA simulations in glaciated catchments in Tyrol and showed an average model underestimation of 17%. Studies in non-alpine areas include Wigmosta et al. (1994), who used NOAA-AVHRR SCA to validate simulated snow cover patterns in Montana. Roy et al. (2010) compared MODIS SCA to in-situ snow depth measurements and simulated SCA in a forested study region in Canada. They developed a direct-insertion approach defining an empirical threshold for SWE to compensate for discrepancies between modelled SWE and satellite derived SCA. Zappa (2008) assessed the performance of distributed snow cover simulations in Switzerland, adopting skill scores based on contingency tables for a quantitative evaluation of snow cover simulations and comparing NOAA-AVHRR snow cover data to the model results. He showed that the model captures the observed patterns with high accuracy and that the scores allow an objective quantification of such agreement. However, the results of Zappa (2008) reveal that the largest uncertainties are present in the regions of the transition zones between the valley plains and the upper part of the valley slopes.

MODIS snow cover data (SCA) are appealing for regional scale modelling and validation. The main advantage of MODIS imagery is the high temporal and spatial resolution and mapping accuracy. Comparisons of MODIS snow cover data with other satellite products and ground based snow depth measurements showed mapping accuracy between 90 and 95% in cloud free conditions, but varying with land cover, snow conditions and snow depth (see e.g., Klein and Barnett, 2003; Simic et al., 2004; Tekeli et al., 2005; Parajka and Blöschl, 2006; Hall and Riggs, 2007; Pu et al., 2007; Tong et al., 2009a; Parajka and Blöschl, 2012). The main limitation, however, is persistent cloud coverage, which can significantly limit MODIS application for snow cover mapping and its usefulness for assimilation into hydrologic models (e.g., Rodell and Houser, 2004; Parajka and Blöschl, 2008a; Şorman et al., 2009; Tong et al., 2009b). Different methods to reduce cloud coverage have been developed (see e.g., Parajka and Blöschl, 2008a; Gafurov and Bárdossy, 2009; Tong et al., 2009a; Hall et al., 2010; Parajka et al., 2010). Parajka and Blöschl (2006) show that MODIS classification errors are around 15% in the winter months and around 1% in summer; however, this is

related to the larger spatial extent of clouds in the winter months. On average, Parajka and Blöschl (2006) have estimated a spatial extent of clouds over Austria of 63% for the years 2000-2005. Recently, MODIS SCA data have been assimilated into hydrological models (see e.g., Andreadis and Lettenmaier, 2006; Roy et al., 2010; Thirel et al., 2011) and used for model calibration (see e.g., Déry et al., 2005; Tekeli et al., 2005; Udnæs et al., 2007; Şorman et al., 2009; Parajka and Blöschl, 2008a; Immerzeel et al., 2009), mostly indicating that using MODIS data improves the snow simulations more than it does the streamflow simulations.

In most of the basins in the present study springtime streamflow is highly influenced by the water stored in the snow pack. Especially for flood forecasting it is vital to estimate the available water storage as accurately as possible. As highlighted in the literature, using MODIS snow cover data show much promise for model evaluation. However, most studies evaluated the snow models over a short time period because prevailing cloud cover limited the available remote sensing data. Thus, the objective of the present study is (1) to investigate whether MODIS data with a large spatial extent of cloud coverage over a basin can be used for evaluating a snow model, (2) to examine the evaluation method in terms of the thresholds used, and (3) to analyse the temporal and spatial performance of the snow component of an operational flood forecasting model using observed SCA data derived from MODIS. We use a semi-distributed conceptual hydrological model in a simulation mode with historical data and MODIS data version 5 (Riggs et al., 2006) from 2003-2009 as an independent data set.

The paper is organised as follows. The data section 3.2 gives details on the study area and the ground and satellite data used in the paper. The methods section 3.3 gives a short description of the calibration of the model and the error measures used to evaluate the snow model performance. In the results section 3.4 a sensitivity analysis to evaluate threshold values on the snow model performance is carried out and the errors are analysed in regard to seasonality and elevation; the model performance is evaluated in space and time. The final section 3.5 discusses the results and concludes with remarks on potential future applications of remote sensing snow cover data in operational flood forecasting. In Appendix B, a description of the model is given.

3.2 Study region and data

The flood forecasting system for the Austrian Danube consists of three parts: (1) a meteorological, (2) a hydrological and (3) a hydraulic model part. The meteorological forecasts include deterministic and ensemble forecasts of precipitation and air temperature for 48 hours on an hourly time step; the hydrological model estimates deterministic and ensemble streamflow forecasts in the Danube tributaries; and the hydraulic model is run with the results from the hydrological model to estimate streamflow and water level for the Danube River. In this study we focus on the evaluation of the snow model. For this purpose we ran the hydrological model with observed meteorological data. The study region includes the tributaries to the Danube River which cover a large part of Austria and some parts of Bavaria. Hydrological

conditions are quite diverse. In the Alpine West elevations range up to 3800 m a.s.l. while the North and East consist of prealpine terrain and lowlands with elevations between 200 and 800 m a.s.l. (Figure 3.1). Land use is mainly agricultural in the lowlands, forests in the medium elevation ranges and alpine vegetation, rocks and glaciers in the alpine catchments. The alpine catchments are generally wetter with mean annual precipitation of almost 2000 mm/yr in the West compared to 600 mm/yr in the East.
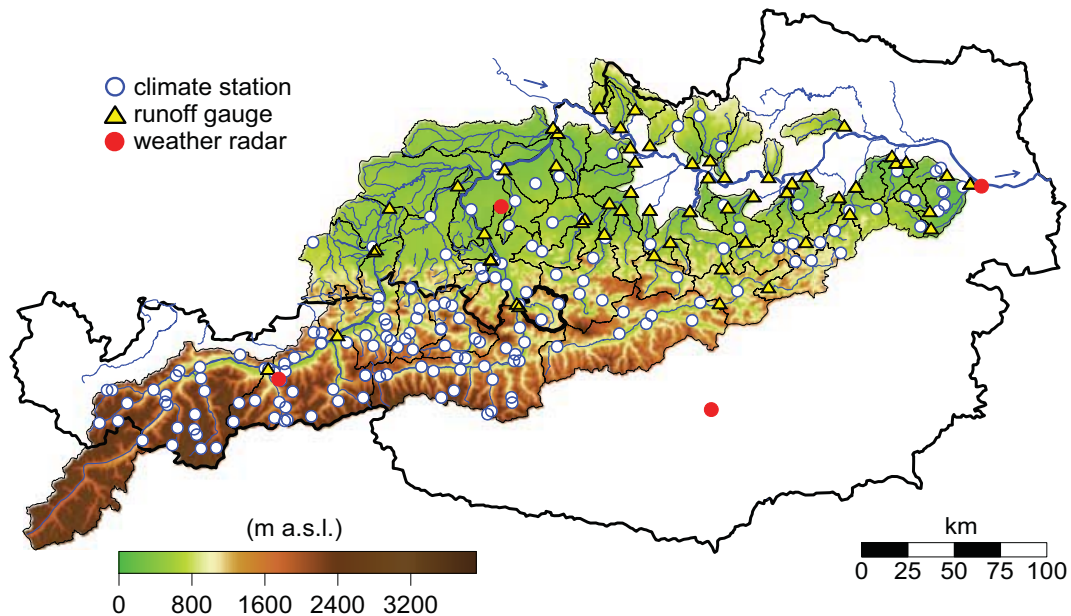
Figure 3.1: Topography of Austria and parts of Southern Germany. The stream gauges used in the study are indicated by triangles, precipitation gauges by white circles, weather radar stations by red circles. Thin black lines are catchment boundaries, the thick black line highlights the catchment Obergäu/Lammer used for detailed analyses.

The hydrologic data set used in this study includes hourly streamflow data of 57 gauged and telemetered catchments with sizes ranging from 70 km^2 to 25600 km^2 to calibrate and validate the model. The data set also includes hourly values of precipitation, air temperature and potential evapotranspiration. Precipitation and temperature measurements for the years 2003 to 2009 were spatially interpolated by the Central Institute for Meteorology and Geodynamics (ZAMG) in Vienna using the algorithm implemented in the INCA system (Steinheimer and Haiden, 2007; Haiden and Pistotnik, 2009). The INCA system was developed by the ZAMG mainly for meteorological forecasting, but it can also be used with historical data. INCA uses output from surface station observations, radar data and elevation data to generate gridded weather data. For the precipitation analysis, a combination of interpolated station data including elevation effects and spatially structured radar is used. The procedure of combining the different data is given in Haiden et al. (2010). The spatial distribution

of potential evapotranspiration was estimated from hourly temperature and daily potential sunshine duration by a modified Blaney-Criddle equation (DVWK, 1996). This method has been shown to give plausible results in Austria (Parajka et al., 2003). To calibrate and verify the model on streamflow data, the years 2003-2006 were used as calibration period; 2007-2009 were used as a validation period.

Figure 3.2: Top: cloud cover for the years 2003 to 2009 in the model region. Average cloud cover for the months October to May is 50%. Bottom: long term ratio of 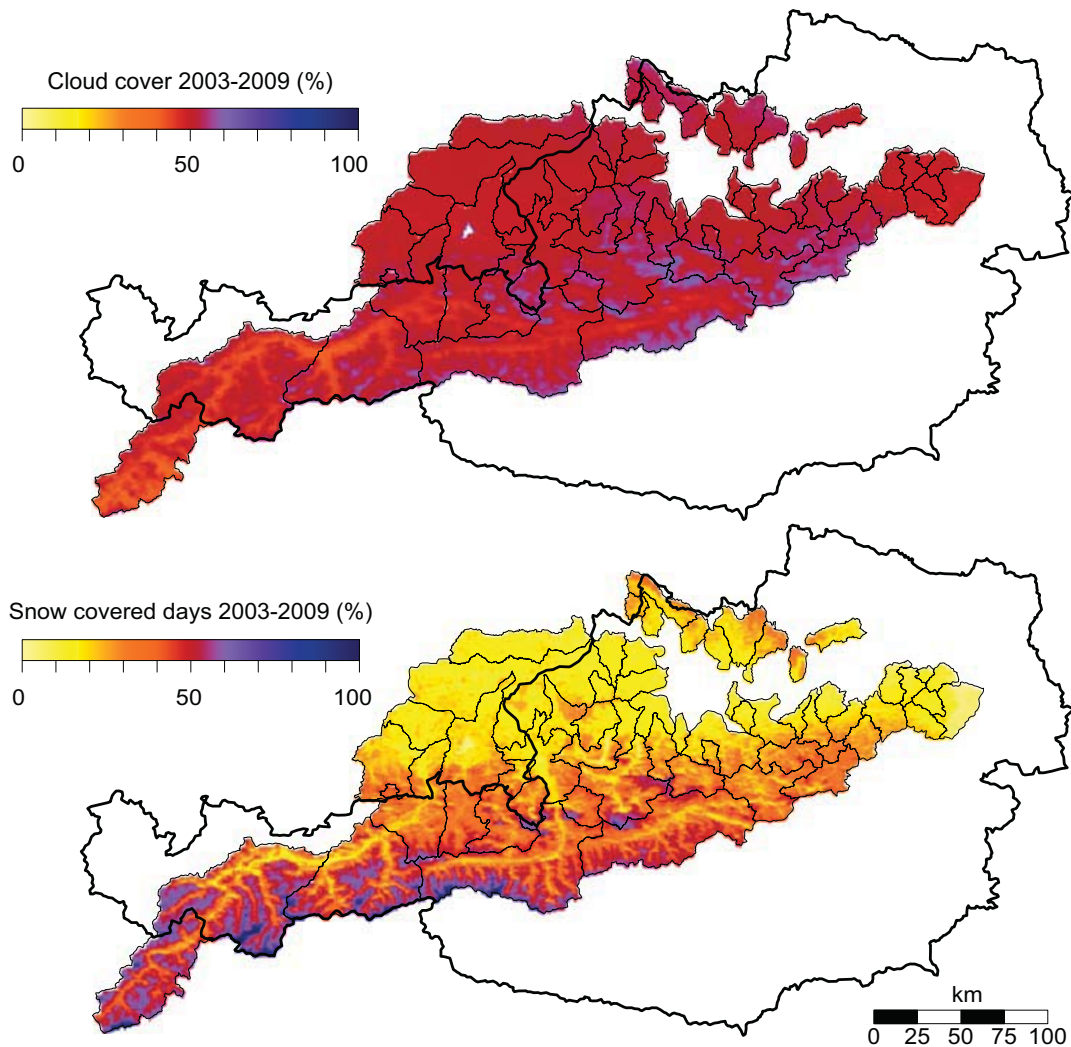snow covered days (S_{MODIS}) to snow covered plus snow free days (L_{MODIS}) for the years 2003 to 2009, according to MODIS. The thick black line indicates Austria; thin black lines indicate the model areas. White indicates no data and area outside the model region.

Once the hydrologic model was calibrated and in operational use, we were interested in how the model fits to the MODIS data which are an independent source of information as they were never used for calibration. The MODIS data used for evaluating the snow routine are based on daily observations acquired by the Terra and Aqua satellites of the NASA Earth Observation System. We used the approach of Parajka and Blöschl (2008a) to merge the original Version 5 Terra (MOD10A1) and Aqua (MYD10A1) MODIS products (Hall et al., 2006, 2007) in space and time on a pixel basis. The MODIS snow cover maps were reclassified from originally 16 pixel classes to three categories: snow, no snow (land) and clouds. The snow class was retained as snow. The snow-free land class was retained as no snow (land). The cloud, missing and erroneous data classes were combined into clouds. However, missing and erroneous data represent only a small portion of the total data. The remaining 11 original classes did not occur in the computations of this study. Based on the following assumptions, four pixels with the size of 500x500 m were aggregated to obtain snow cover maps with a pixel size of 1x1 km: (1) if all four pixels were marked as the same category, the category of the 1x1 km pixel remained the same, (2) if the number of pixels marked as no snow was greater than the number of snow pixels, the 1x1 km pixel was classified as no snow, and (3) if the number of pixels marked as snow was greater or equal than the number of pixels marked as no snow, the 1x1 km pixel was classified as snow.

The average cloud cover over the study region for the period 2003-2009 (total of 2555 days) is around 50% for the combined Aqua-Terra-MODIS data. Cloud coverage is around 40% in alpine valleys and around 60% over mountainous terrain. Figure 3.2 (top) shows the spatial distribution of the cloud coverage for the period 2003-2009 in the model region. Figure 3.2 (bottom) shows the long term ratio of snow covered days (SCD) for 2003-2009. For every pixel, the ratio of snow covered days (S_{MODIS}) to snow covered plus snow free days (L_{MODIS}) was determined; days with cloud cover were not considered. As expected, the ratio of SCD closely follows the elevation in the area: in the Alpine region the percentage of snow covered days is much higher than in the lowlands in the East and North of the Alps. Similarly, the Alpine valleys show a smaller percentage of snow covered days than the higher elevations. The mean snow cover in the model region is 30% for cloud free days for 2003-2009.

3.3 Methods

Model structure and calibration

The rainfall runoff model used in this study is a conceptual hydrological model which is applied in a semi-distributed mode. The model is operationally used for flood forecasting for the Austrian Danube tributaries; in this study, however, we use the model in a simulation mode with historical data. The structure is similar to that of the HBV model (Bergström, 1976). Detailed information about the model structure is given in Appendix B and in Blöschl et al. (2008). Parameters were estimated manually and separately for each of the 57 catchments for

the entire calibration period 2003-2006. First, nested catchments were calibrated. Parameter values of these basins were then considered to be constant and the remaining parts of the catchment were calibrated. The calibration process followed a number of steps (Blöschl et al., 2008). The first step was an approximation of the annual water balance. As snow is a major component of the water balance in the study area and can influence the soil moisture state throughout the year, initial parameters for the slow flow components, for the maximum soil moisture storage and the snow routine were set in this first step. Then, the initial model parameters were adjusted in order to reproduce seasonal patterns correctly. Threshold temperatures were adjusted, as well as parameters influencing the slow components. The last step was to parameterize the fast flow components and the parameters of the linear storage cascade by looking at single flood events as well as a fine tuning of the parameters of the snow and soil moisture routines. The main goal of the calibration was to estimate the timing of the rising limbs and the peak discharge as well as the magnitude of the peak discharge as correctly as possible. After each model run, we visualised the model simulations and evaluated the results using statistical measures for the entire calibration period.

Table 3.1: Hydrologic model parameters of the snow routine

Model parameter	Description	min in region	max in region
D	Degree day factor (mm.°C^{-1}.day^{-1})	1.3	2.3
T_s	Threshold temperature (°C)	-1.8	-0.4
T_r	Threshold temperature (°C)	0.8	1.6
T_m	Melt temperature (°C)	0.1	0.9
C_S	Snow correction factor (-)	0.8	1.0

For the calibration of the snow routine, the snow correction factor C_S was set to a value of 1, as an elevation-based correction of precipitation is part of the INCA system (Haiden, 2009). The choice of threshold temperatures was guided by Seibert (1999) who used values ranging from -1.5 to 2.5°C and a degree day factor ranging from 1 to 10 mm.°C^{-1}.day^{-1} for his Monte Carlo based calibration in Sweden. Braun (1985) used a temperature range from -2.0 to 4.0°C in lowland and lower-alpine catchments in Switzerland where a mix of rain and snow can occur, whereas Kienzle (2008) proposed a wider threshold temperatures range from $T_s = $ -4°C to $T_r = $ 8°C for Canada. Merz et al. (2009) preset values of 0°C and 2°C for the threshold temperatures T_s and T_r and produced accurate streamflow simulations at the daily time scale. We estimated parameters for T_s from -1.8 to -0.4°C and for T_r in the range of 0.8 to 1.6°C in our model area. The threshold temperatures are well in the range of other studies. The remaining parameters of the snow routine are in the range of the parameters in Merz and Blöschl (2004), who estimated the parameters with an automatic algorithm. The melt temperature T_m was set to values in the order of 0.1 to 0.9°C; the degree day factor D is in the range of 1.3 to 2.3 mm.°C^{-1}.day^{-1} and doubles during rain-on-snow events (Sui and Koehler, 2001). Table 3.1 gives an overview of the range of the calibrated snow routine model parameters; additional details on the calibration are given in Nester et al. (2011a).

Figure 3.3 shows the model results for the winter months (October to May) for the gauge Obergäu/Lammer (catchment highlighted in Figure 3.1). The gauge is shown as a rep-

resentative example for an Alpine catchment with elevations ranging from 470 m a.s.l. to 2400 m a.s.l. 10% of the catchment area are below 750 m a.s.l., 50% of the area are between 750 and 1250 m a.s.l., 30% are between 1250 and 1750 m a.s.l. and the remaining 10% of the catchment area are in elevations higher than 1750 m a.s.l. The top panels refer to the calibration period of the hydrological model and show the winter season 2004-2005. Starting at the end of October, most of the precipitation is accumulating as snow, and only a small amount of precipitation is directly contributing to runoff. For comparison of the snow model results we plotted modelled basin average SWE values and observed snow depths of the only station in the catchment where snow data were available (elevation 700 m a.s.l.). The peaks in the winter season are simulated well. At the beginning of March, the modelled snow water equivalent is close to 400 mm; snow melt starts at the middle of the month. The timing of the rise is simulated well but the peak is overestimated. Until mid April, the model is over-estimating the snow melt induced streamflow but the daily characteristics of the snow melt is reproduced well. Several storm events increase the runoff rapidly. The lower panels refer to the validation period of the model, showing the winter season 2006-2007. In this winter, less snow has been accumulated to a maximum snow water equivalent of around 100 mm. Several short storms directly contribute to the runoff or in a combination of melt and rain, e.g., in November and in January. Again, the snow melt starts in the middle of March but in this year the snow melt period is underestimated by the model. We used different statistical measures to evaluate the performance of the model including the Nash and Sutcliffe (1970) coefficient of efficiency ($nsme$):

$$nsme = 1 - \frac{\sum\limits_{i=1}^{n} (Q_{sim,i} - Q_{obs,i})^2}{\sum\limits_{i=1}^{n} \left(Q_{obs,i} - \overline{Q_{obs}}\right)^2} \tag{3.1}$$

where $Q_{obs,i}$ and $Q_{sim,i}$ are observed and simulated runoff at hour i, respectively, and $\overline{Q_{obs}}$ is the mean observed runoff over the calibration or validation period of n hours. $nsme$ values can range from ∞ to 1. A perfect match between simulation and observation implies $nsme = 1$; $nsme = 0$ indicates that the model predictions are as accurate as the mean of the observed data, and $nsme < 0$ occurs when the observed mean is a better predictor than the model. For the entire calibration period (summer and winter) the $nsme$ for the gauge Obergäu/Lammer is 0.60, for the validation period it is 0.69. For the periods shown in Figure 3.3, $nsme$ is 0.69 for the winter 2004-2005 and 0.65 for the winter 2006-2007. Details on the model performance are given in Nester et al. (2011a).

For the evaluation of the snow model, we had to consider that the model results are available on an hourly time step whereas MODIS data are available on a daily basis. Typically, Aqua data are acquired around 1 p.m. and Terra data around 11 a.m. over Austrian territory. Therefore we used the model results at 12 noon for the evaluation. Also the model simulates uniform SWE within each elevation zone of a catchment, which can be considered either snow covered or snow free, depending on a threshold value chosen. Threshold values for
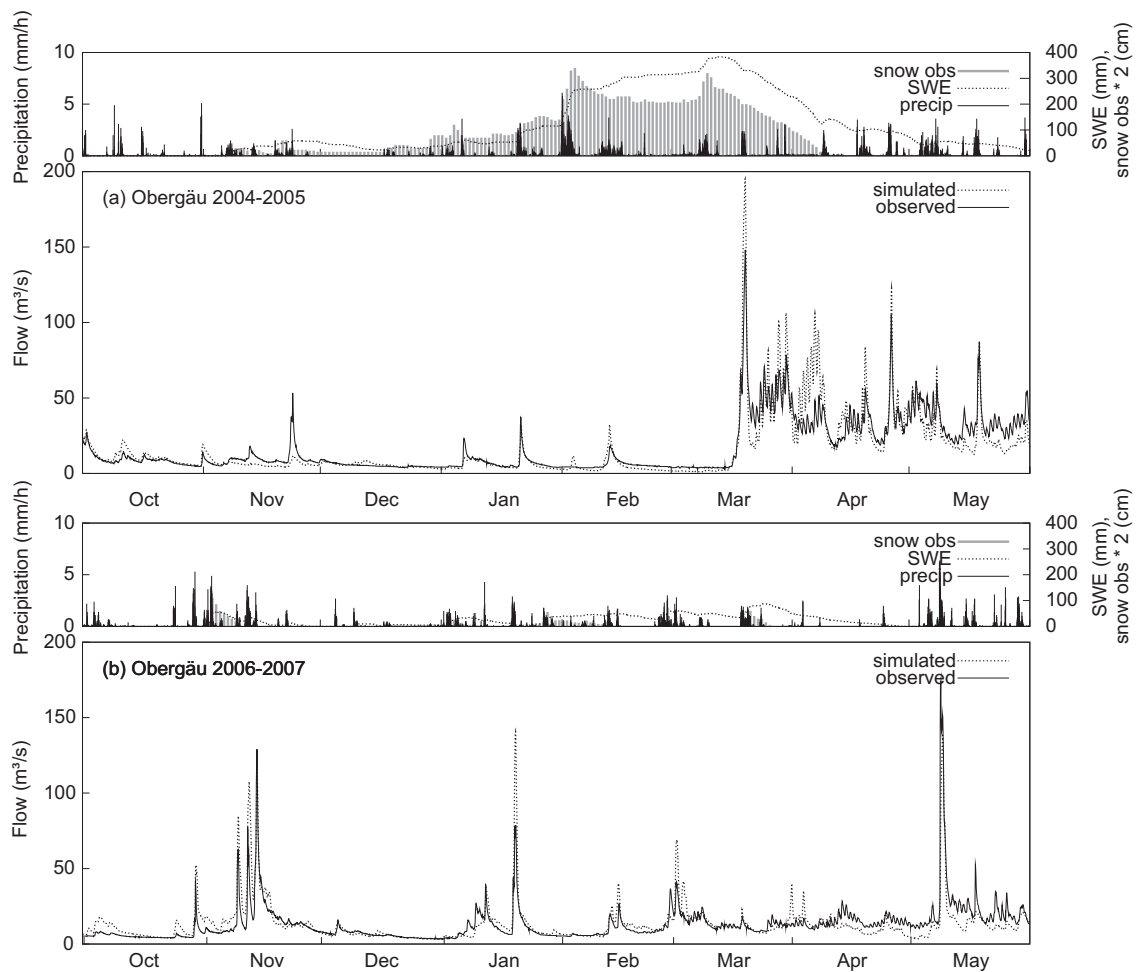
Figure 3.3: Observed and simulated hydrographs for the winter seasons 2004-2005 (calibration period of the hydrological model, top) and 2006-2007 (validation period of the hydrological model, bottom) for the gauge Obergäu/Lammer. Snow depth measurements for a single station (Annaberg, 700 m a.s.l.) in the catchment are shown.

snow depth used in the literature include 2 cm (Tong et al., 2009a), 2.5 cm (Tekeli et al., 2005), 3.5 cm (Klein and Barnett, 2003) and 4 cm (Roy et al., 2010). As a starting point, we considered simulated SWE larger than the threshold $\xi_{SWE} = 2.5$ mm as snow covered. In Figure 3.4 the same periods as in Figure 3.3 are plotted in order to evaluate the temporal evolution of the SCA. The top panels show the mean hourly catchment temperature in the range from -10 to 10°C. The lower panels show basin average SCA derived from both MODIS and the model results. MODIS SCA values are shown as circles in different sizes, the size of the markers referring to the different cloud coverage classes. Large circles denote days with cloud coverage less than 50%, when MODIS data contain a lot of information. Medium sized circles indicate less information about the snow extent (cloud coverage ranging from 50 to 80%). Small circles relate to cloud coverages ranging from 80 to 95%. Hourly model based

SCA values were calculated only from cloud free pixels for consistency of catchment area. Both figures indicate that the timing of the snow accumulation is accurately simulated. The modelled SCA values increase at the same time as the MODIS SCA. Similarly, the timing of the beginning of the snow depletion between simulation and observation matches well. However, when MODIS indicates snow in October and May, the model is underestimating SCA. Observed SCA data for the season 2004-2005 (Figure 3.4 top) indicate that the snow cover in the catchment was more or less constant (but with varying snow depths as shown in Figure 3.3 top), whereas for the season 2006-2007 (Figure 3.4 bottom) the observed SCA shows a lot of dynamics in terms of snow melt and accumulation. This is due to the fact that the snow depths observed are much smaller than in the season 2004-2005, so complete snow depletion is more likely to occur. The temporal evolution of the SCA from November to March is simulated well for both the winters shown.

Efficiency and errors for snow covered area

For snow covered area, the evaluation of the results is not straightforward as the model is based on elevation zones while the MODIS data are raster based. Additionally, the model simulates the amount (volume) of water stored in the form of snow, whereas MODIS snow cover data contains information only about the spatial extent of snow (i.e. whether a pixel is classified as snow, land or missing information). We used the method of Parajka and Blöschl (2008b) who compared MODIS snow cover data with SWE model simulations in an indirect way.

The comparison is performed in individual elevation zones of a catchment. Two types of snow errors are evaluated. The first, termed model overestimation error (S_E^O), counts the number of days m_O when the hydrologic model simulates zone SWE greater than a threshold but MODIS indicates that snow covered area less then a threshold is present in the zone, i.e.:

$$S_E^O = \frac{1}{m \cdot l} \sum_{j=1}^{l} m_O \wedge (SWE > \xi_{SWE}) \wedge (SCA < \xi_{SCA}) \qquad (3.2)$$

where SWE is the simulated snow water equivalent in one zone, SCA is the MODIS snow covered area within this zone, m is the number of days where MODIS images are available (with cloud cover less than a threshold ξ_C (%)), l is the number of zones of a particular catchment, ξ_{SWE} (mm) is a threshold that determines when a zone can be essentially considered snow free in terms of the simulations and ξ_{SCA} (%) is a threshold that determines when a zone can be essentially considered snow free in terms of the MODIS data.

The second error, termed model underestimation error (S_E^U), counts the number of days m_U when the hydrologic model simulates snow less than a threshold in a zone but MODIS indicates that snow covered area greater than a threshold is present in the zone, i.e.:

$$S_E^U = \frac{1}{m \cdot l} \sum_{j=1}^{l} m_U \wedge (SWE < \xi_{SWE}) \wedge (SCA > \xi_{SCA}) \qquad (3.3)$$
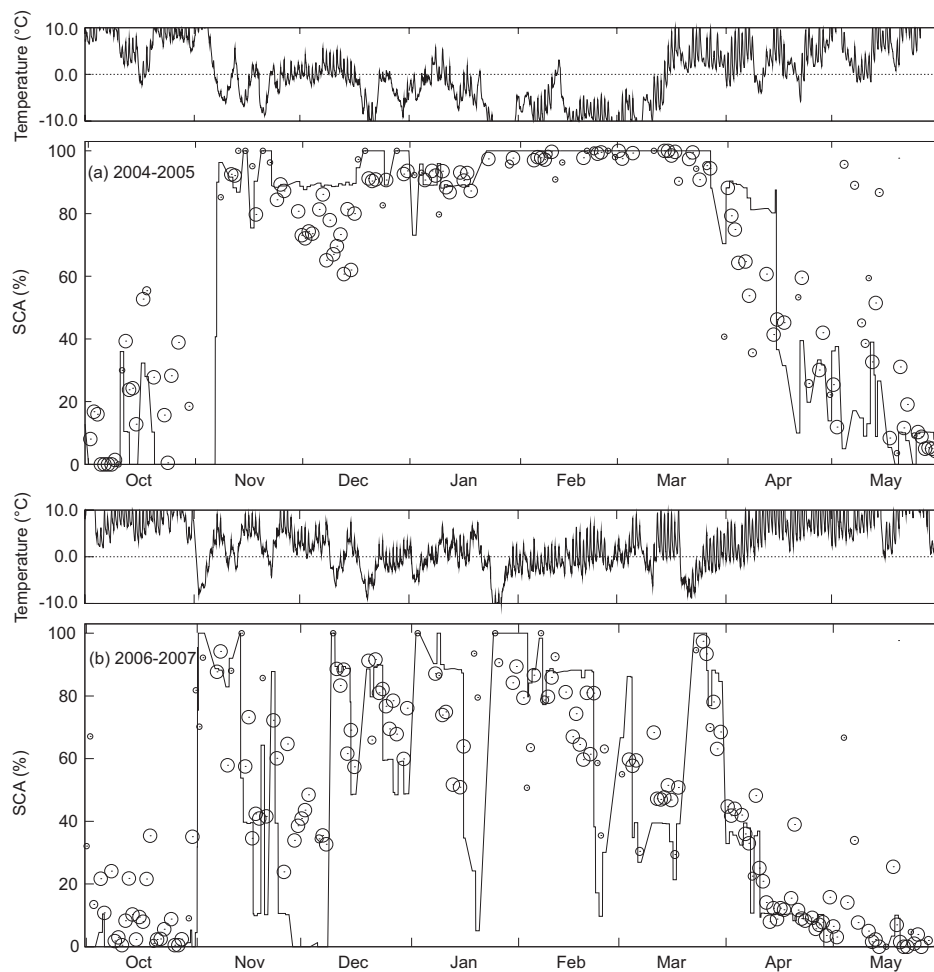
Figure 3.4: Basin average snow covered area (SCA) for the Lammer catchment, 2004-2005 (top) and 2006-2007 (bottom). MODIS data are shown as circles, model results as solid lines. Size of circles indicates the cloud coverage over the catchment (large circles - cloud coverage less than 50%, medium circles - cloud coverage 50-80% and small circles cloud coverage 80-95%).

The percent or fraction of snow covered area, SCA, within each zone was calculated from the MODIS data as:

$$SCA = \frac{S}{S + L} \tag{3.4}$$

where S and L represent the number of pixels mapped as snow and land, respectively, for a given day and a given zone. The reliability and accuracy of the SCA estimation depends on the spatial extent of clouds occurring in an elevation zone. Only those days of the SCA images were therefore used for a particular day and elevation zone if the cloud coverage, CC,

was less than a threshold ξ_C:

$$CC = \frac{C}{S + L + C} \tag{3.5}$$

where C represents the number of pixels mapped as cloud covered and CC is the fractional cloud cover for a particular day and elevation zone. The thresholds ξ_{SWE} (mm), ξ_{SCA} (%) and ξ_C (%) were chosen on the basis of a sensitivity analysis. The magnitude of the threshold ξ_C will affect the number of days for which MODIS images are available. Parajka and Blöschl (2008b) suggest a threshold of 60% of cloud coverage; Hall et al. (2010)) used a threshold of 80% to develop the MODIS cloud-gap-filled snow map product. In this study, the whole range of cloud coverage is analysed. For 50% of the MODIS data, the cloud coverage was less than 50%, for 20% of the MODIS data cloud coverage was between 50 and 80% and for 30% of the MODIS data cloud coverage was larger than 80%. Three different ranges of cloud coverage are chosen: (1) CC < 50%, (2) 50% < CC < 80% and (3) CC > 80% over a catchment. The thresholds ξ_{SWE} and ξ_{SCA} are used to compare model simulations and MODIS snow cover observations to define the snow model errors.

3.4 Results

Summary statistics of snow model performance and choice of thresholds

A sensitivity study was carried out to analyse the impact of different threshold values for ξ_{SWE} and ξ_{SCA} on Eqs. 3.2 and 3.3. Figure 3.5 shows the median overestimation (left) and underestimation (right) errors for a cloud coverage < 50%. For clarity of presentation the different cloud coverages are not shown. Overestimation errors decrease with increasing ξ_{SWE} but increase with increasing ξ_{SCA} (Fig. 5 left); underestimation errors decrease with increasing ξ_{SCA} and increase with increasing ξ_{SWE} (Figure 3.5 right). The change in overestimation errors is smaller than the change in underestimation errors. The errors for snow covered area are less sensitive to the threshold ξ_{SWE} than to the threshold ξ_{SCA}. The largest median overestimation errors occur for small thresholds for SWE and large thresholds for SCA; the largest underestimation errors occur for small thresholds for SCA and large thresholds for SWE. In order to achieve a compromise between over- and underestimation errors, we chose a threshold $\xi_{SWE} = 2.5$ mm as the median values and the percentile differences (P75% - P25%) for the overestimation errors are of the same order of magnitude for cloud coverage less than 80% (Figure 3.5 left). The threshold $\xi_{SCA} = 30\%$ was chosen as the overestimation errors increase only slightly from $\xi_{SCA} = 25$ to 30% compared to the increase of overestimation errors from $\xi_{SCA} = 30$ to 50% (Figure 3.5 left), and the underestimation errors are clearly smaller for $\xi_{SCA} = 30\%$ compared to $\xi_{SCA} = 25\%$ (Figure 3.5 right).

Table 3.2 summarises the overestimation errors for different thresholds ξ_{SWE} and different cloud coverage at a constant threshold $\xi_{SCA} = 30\%$. For cloud coverage less than 50% the S_E^O overestimation errors are not very sensitive to the choice of the threshold ξ_{SWE}. The decrease of S_E^O with increasing threshold ξ_{SWE} is small. Median values range from 1.3% for a threshold $\xi_{SWE} = 0$ mm and 0.4% for a threshold $\xi_{SWE} = 10$ mm. The percentile
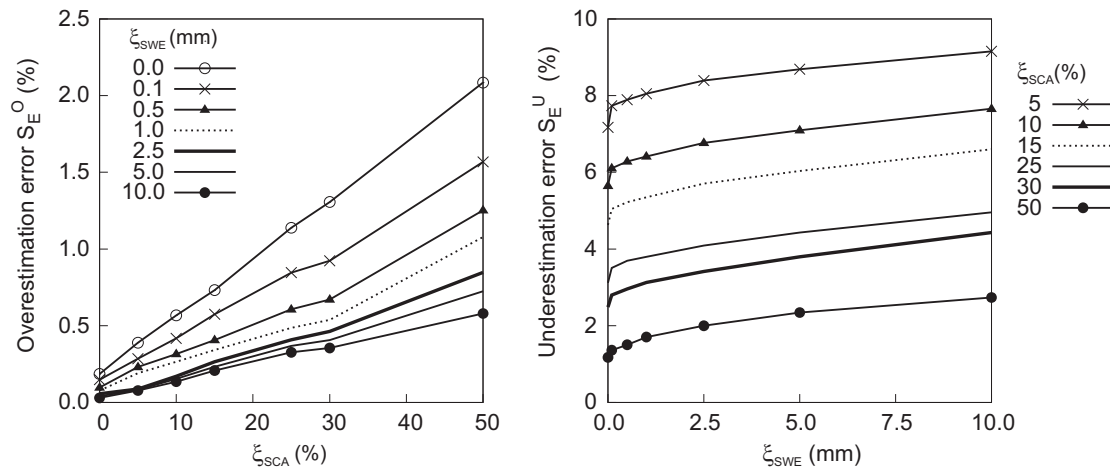
Figure 3.5: Sensitivity analysis. Medians of snow overestimation errors (left) and snow underestimation errors (right) for cloud coverage $< 50\%$. Threshold values of $\xi_{SWE} = 2.5$ mm and $\xi_{SCA} = 30\%$ were chosen for further analyses.

difference (P75% - P25%) is stable around 0.6% which indicates that the shape of the CDF does not change much with changing thresholds. For a cloud coverage between 50 and 80%, the results are slightly more sensitive, with median overestimation errors ranging from 2.0% ($\xi_{SWE} = 0$ mm) to 0.5% ($\xi_{SWE} = 10$ mm). The percentile difference (P75% - P25%) is less stable from 0.8% to 2.0%. For a cloud extent of more than 80%, the errors are larger, with medians ranging from 3.3% to 1.0% for thresholds $\xi_{SWE} = 0$ mm and $\xi_{SWE} = 10$ mm, respectively. Percentile differences (P75% - P25%) are in the range of 0.9 to 2.2%. This shows that the choice of the thresholds ξ_{SWE} is getting more important as one moves up with the threshold for ξ_C.

Table 3.2: Statistical evaluation of the snow error overestimation CDFs for different ξ_{SWE} (mm) and constant $\xi_{SCA} = 30\%$. The first value is the median; the second value is the percentile difference (P75% - P25%) over 57 catchments for the period 2003-2009.

ξ_{SWE} (mm)	$CC < 50\%$	$50\% < CC < 80\%$	$CC > 80\%$
0.0	1.3/0.8	2.0/2.0	3.3/2.2
0.1	0.9/0.6	1.6/1.6	2.6/1.9
0.5	0.7/0.6	1.4/0.9	2.2/1.3
1.0	0.5/0.6	1.2/1.1	1.8/1.4
2.5	0.5/0.6	0.8/0.9	1.5/1.2
5.0	0.4/0.7	0.7/0.8	1.2/1.0
10.0	0.4/0.6	0.5/0.8	1.0/0.9

Snow underestimation errors are evaluated for different thresholds of ξ_{SCA} but at a constant threshold $\xi_{SWE} = 2.5$ mm. Contrary to the overestimation errors, the underestimation errors S_E^U are more sensitive to the threshold which was expected from the results in Figure 3.5.

The S_E^U underestimation errors are largest for the restrictive threshold $\xi_{SCA} = 0$ (the graph for $\xi_{SCA} = 0$ is not shown in Figure 3.5, as the errors are too large). The errors decrease as the threshold gets less restrictive (increasing ξ_{SCA}), as one would expect. For example, for $\xi_{SCA} = 10\%$ and $\xi_{SWE} = 2.5$ mm (line with triangle markers) the S_E^U errors are 6.8% for half the basins, for $= 30\%$ and $\xi_{SWE} = 2.5$ mm (thick solid line) the S_E^U errors are 3.4%. The percentile difference (P75% - P25%) is not very stable and is larger than that of Parajka and Blöschl (2008b). In this study, for a threshold $\xi_{SCA} = 15\%$ the percentile difference is 3.8%, compared to 2.4% in Parajka and Blöschl (2008b) and for a threshold $\xi_{SCA} = 30\%$ the difference in this study is 3.0% compared to 1.1% in the above paper. However, they used slightly larger cloud coverage (50% vs. 60%), smaller catchments, especially in alpine areas, and fewer catchments in prealpine areas. Table 3.3 summarises the underestimation errors for different thresholds ξ_{SCA} and different cloud coverages. The results suggest that the use of a threshold of ξ_C is necessary, as the amount of information clearly decreases with the increase of cloud coverage.

Table 3.3: Statistical evaluation of snow error underestimation errors for different ξ_{SCA} (%) and cloud coverage. $\xi_{SWE} = 2.5$ mm. The first value is the median; the second value is the percentile difference (P75% - P25%) over 57 catchments for the period 2003-2009.

ξ_{SCA} (%)	CC < 50%	50% < CC < 80%	CC > 80%
0	12.5/7.5	19.5/11.2	14.1/9.6
5	8.4/5.3	15.5/9.9	13.1/8.7
10	6.8/4.4	13.1/8.7	12.5/7.4
15	5.7/3.8	10.7/7.0	11.5/6.8
25	4.1/3.2	7.2/5.8	9.5/5.7
30	3.4/3.0	6.4/5.5	8.9/5.0
50	2.0/1.6	3.1/3.3	5.9/3.8

Figure 3.6 shows the seasonal distribution of the median overestimation errors S_E^O and the median underestimation errors S_E^U for the thresholds $\xi_{SWE} = 2.5$ mm and $\xi_{SCA} = 30\%$. As expected, overestimation errors are small with the largest values in the range of 0.2% in the months February to April. Underestimation errors do have a clear seasonal cycle with peaks in March and April and October. For a SCA threshold $\xi_{SCA} = 30\%$ the largest error is around 0.8% in November. These results confirm Figure 3.4, where we showed an underestimation of snow by the model in the accumulation and depletion phases. With increasing cloud cover, the overestimation errors increase slightly whereas the increase in underestimation errors is more obvious. Interestingly, a threshold value $\xi_{SCA} = 30\%$ leads to underestimation errors in the summer months for cloud coverage larger than 50%. This indicates that some areas are marked as snow covered in MODIS whereas the model simulated no snow. Parajka and Blöschl (2006) found that the difference may be due to the cloud mask used in the snow mapping algorithm, where MODIS misclassified clouds as snow in summer months. Parajka et al. (2010) compared MODIS maps with grid maps of mean daily air temperatures in Austria and found a maximum of 1.4% of the pixels classified as snow in July and 3.2% of the pixels classified as snow in May when the mean air temperature was higher than 10°C.

We find similar values in prealpine regions in this study, with 1.6% of the pixels classified as snow in July and 5.8% of the pixels classified as snow in May.
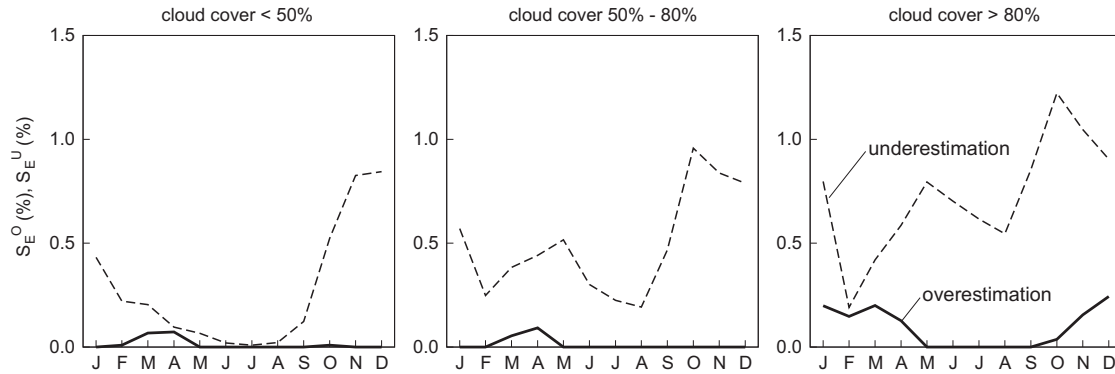

Figure 3.6: Seasonal distribution of snow overestimation S_E^O (solid lines) and underestimation S_E^U (dashed lines) errors. Median values 2003-2009 for thresholds $\xi_{SWE} = 2.5\,\text{mm}$ and $\xi_{SCA} = 30\%$.

Further insight into the snow model results is provided by Figure 3.7. For each catchment, the overestimation errors S_E^O and underestimation errors S_E^U are analysed as a function of mean catchment elevation for the period 2003-2009. We show only the errors for the thresholds $\xi_{SWE} = 2.5\,\text{mm}$ and $\xi_{SCA} = 30\%$. Open circles indicate overestimation errors and dark circles indicate underestimation errors. For a cloud cover less than 50%, the overestimation errors are in the range of 0 to 2%, whereas the underestimation errors are larger with values ranging from 1 to 11%. With increasing cloud cover, the overestimation errors increase only slightly whereas the underestimation errors increase rapidly. Interestingly, the biggest differences between over- and underestimation errors can be observed for a mean catchment elevation smaller than 1000 m a.s.l. For mean elevations larger than 1000 m a.s.l., the difference is much smaller, indicating that the performance of the snow model is much better for higher altitudes. Zappa (2008) showed similar results in his study for Switzerland. There are several reasons for larger underestimation errors. First, the poorer performance of the snow model in the lower catchments can be attributed to the use of 500 m elevation zones. There are a number of catchments in prealpine areas with only one elevation zone, resulting in a snow covered area of either 0 (snow free) or 100% (snow covered), whereas in MODIS a more precise distinction of snow cover is possible. A second reason for the underestimation errors is misclassification of clouds as snow during summer months, as stated above. Another possibility for larger underestimation errors is that the majority of the errors occur during the melt periods. Therefore, areas that experience frequent melt during the winter may tend to have poorer performance statistics than areas that have a consistent snowpack for several months.
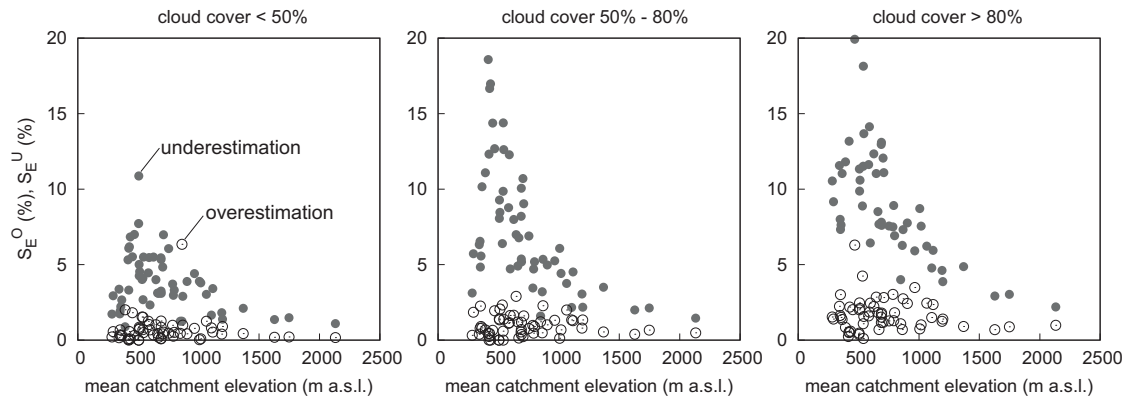
Figure 3.7: Snow overestimation S_E^O (open circles) and underestimation S_E^U errors (dark circles) as a function of elevation. The period is 2003-2009, thresholds $\xi_{SWE} = 2.5$ mm and $\xi_{SCA} = 30\%$. Both snow overestimation and underestimation errors increase as the cloud coverage increases.

Spatial analysis of snow model performance

The spatial validation of the snow routine was carried out on a pixel by pixel basis. The threshold ξ_{SWE} is used for the distinction of snow cover in this analysis; a threshold for cloud cover ξ_C is not needed as only pixels that are not cloud covered in the MODIS data are accounted for; and the threshold ξ_{SCA} is not needed as the validation is carried out on the pixel scale and not on the catchment scale.

Figure 3.8 shows the spatial validation of the snow routine carried out on a pixel basis. As an example for a day with almost no cloud coverage, the extent of the snow cover is shown for February 3, 2008, according to MODIS data (top) and model results (bottom). No data and clouds are indicated by white areas, snow is shown as grey and snow free areas are shown as green. MODIS data indicate that the mountainous regions are snow covered: the Alps are covered with snow with the exception of some valleys; also in the Northern part of Austria the higher elevation pixels are snow covered. For determining snow cover from the model results, a threshold $\xi_{SWE} = 2.5$ mm was used. The results show a similar extent of the snow cover as the MODIS data: 81.5% of the pixels are correctly classified as snow covered and snow free, respectively, with Alpine valleys and prealpine lowlands not covered by snow. 12.3% of the pixels are underestimated by the model, 4.7% of the pixels are overestimated by the model and 1.5% of the pixels do not contain data.

Figure 3.8 shows an example of a single day. To gain insight into the performance of the snow model for the entire evaluation period 2003-2009, we compared all available MODIS snow cover maps with the model results. We used two different measures. First, we compared the difference of days with simulated and observed snow cover on a pixel basis. Second, we calculated the hit rate H between MODIS and model results for snow covered days and snow free days.
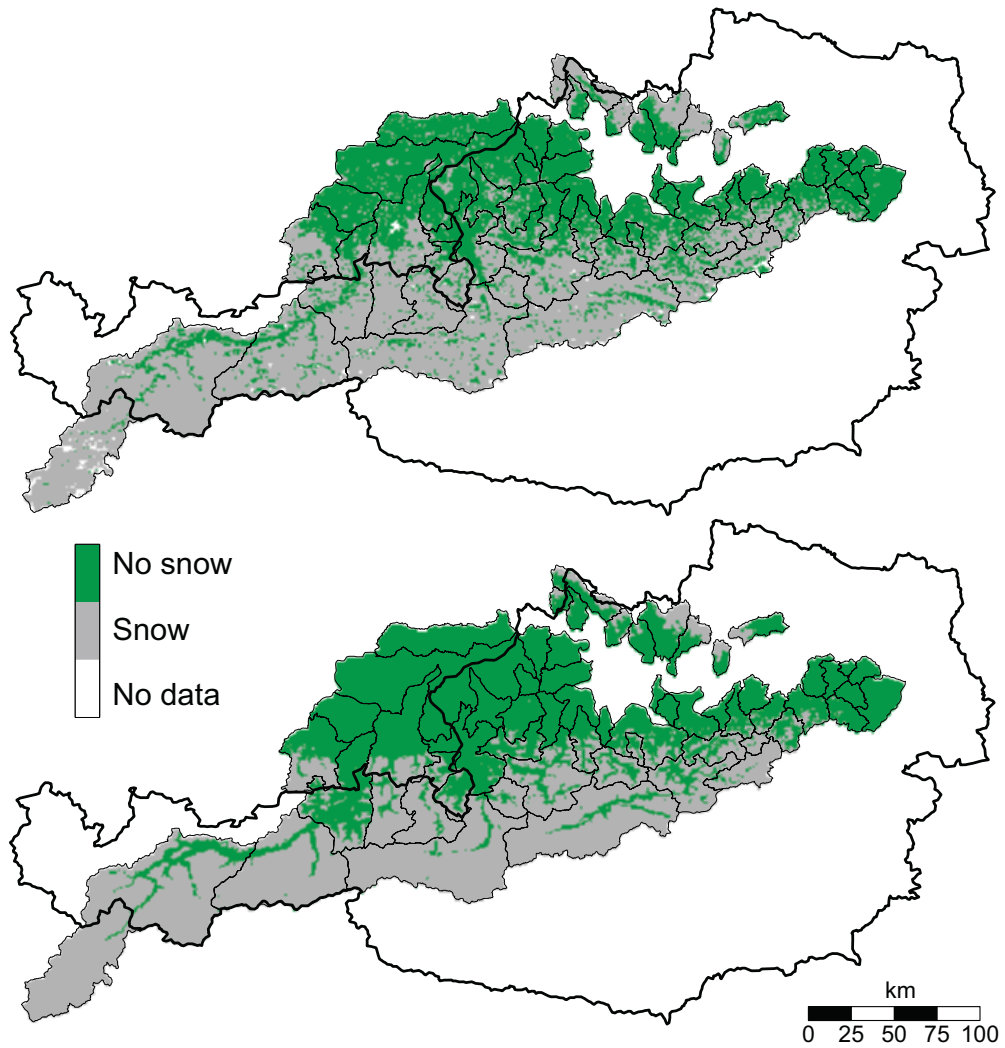
Figure 3.8: Example pattern in the snow melt phase (3 February 2008). Top refers to MODIS data, bottom to model. White indicates cloud covered (no data available), green: snow free, grey: snow covered. Only model SWE larger than a threshold $\xi_{SWE} = 2.5$ mm were considered as snow.

For every pixel the bias was calculated as

$$bias = \left(\frac{S_{MODEL}}{S_{MODEL} + L_{MODEL}} - \frac{S_{MODIS}}{S_{MODIS} + L_{MODIS}} \right) \cdot 100 \qquad (3.6)$$

where S_{MODEL} and L_{MODEL} refer to the number of days with snow cover and no snow, respectively, according to the model; and S_{MODIS} and L_{MODIS} refer to the number of days with snow cover and no snow, respectively, according to MODIS. Only days marked as cloud

free in MODIS were considered. For the distinction of S_{MODEL} and L_{MODEL}, the threshold $\xi_{SWE} = 2.5\,\mathrm{mm}$ was used. For simplicity, we used the terms positive bias when more days are snow covered in the model than MODIS and negative bias when fewer days are snow covered in the model than MODIS.
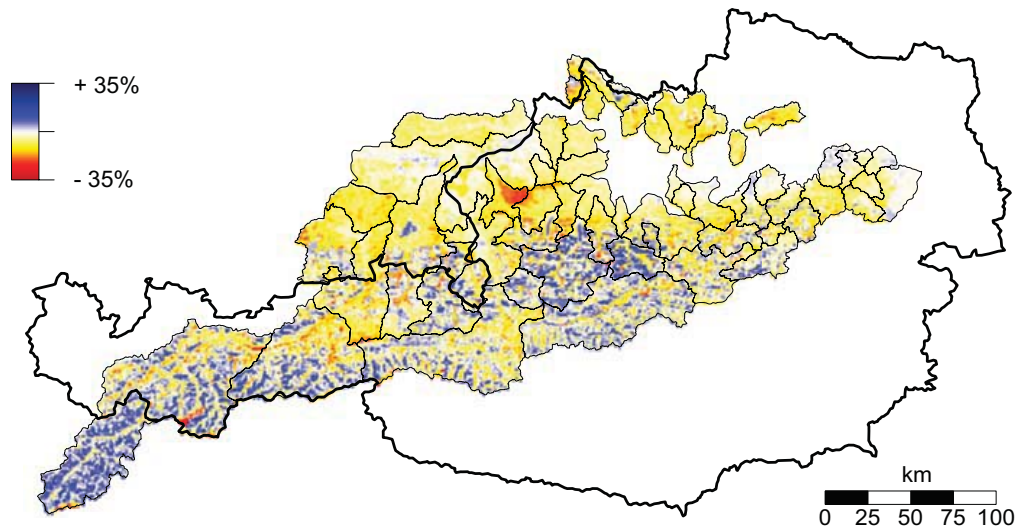
Figure 3.9: $bias$ of the model results relative to MODIS data for the period 2003-2009. Red indicates negative $bias$ (fewer days are snow covered in the model than MODIS), blue indicates positive $bias$. Only model SWE larger than 2.5 mm was considered as snow.

The spatial distribution of the $bias$ (Figure 3.9) shows that in the prealpine parts MODIS indicates snow cover on more days than the model results whereas in alpine regions the model tends to indicate snow cover on more days than MODIS. On average, the negative $bias$ is on the order of 15% of the cloud free days. The negative $bias$ is even larger in parts of a prealpine catchment in the middle of the model area. There may be two reasons for this. First, the model structure for this catchment comprises a single elevation zone. Therefore, the temperature is assumed to be the same throughout the whole catchment, resulting in uniform snow water equivalent across the catchment. The topography, however, does vary by about 400 m within this catchment, so local differences in temperatures are possible. Second, the area underestimated by the model is covered by both coniferous and deciduous forests according to a land cover map (European Environmental Agency, 2000) which can cause problems in MODIS snow detection. For example, Simic et al. (2004) showed that MODIS products give realistic snow cover maps for an average of 93% of the days, with lower percentages for evergreen forests where MODIS has a tendency to overestimate snow. Hall et al. (2002) showed that MODIS has a tendency to underestimate snow in a forested area. The influence of land cover is analysed in more detail in Figure 3.10. Error cumulative distribution functions (CDFs) with values derived from Figure 3.9 show hardly any difference for positive $bias$ between forested and open land (left) whereas there is a clear difference

for negative $bias$ between forests and open land (right) with larger values for forests. This may be related to the model structure. We estimated parameters based on the land use and geology, but the use of a semi-distributed model requires mean parameters for each elevation zone. The use of a distributed model with hydrological response units could perhaps improve the snow model, but further analysis would be required to verify this.
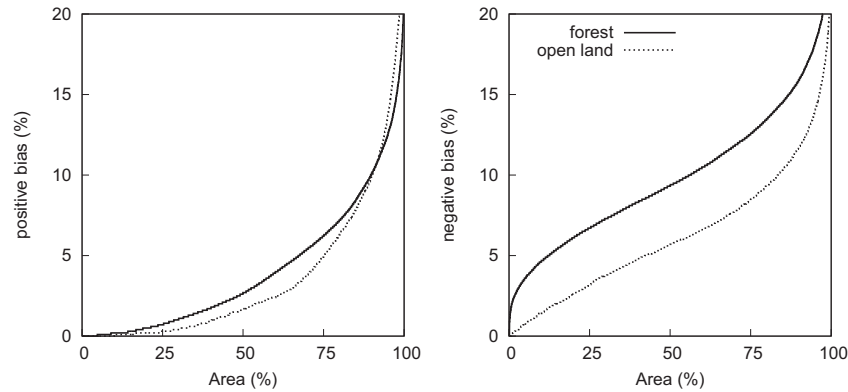
Figure 3.10: CDFs of positive and negative biases (%) for forests and open land, pixel values. Land use taken from Corine. Threshold $\xi_{SWE} = 2.5$ mm.

Figure 3.11 shows the performance of the snow model for the winter seasons (October-May) of the years 2003-2009 for all cloud free days. The overall degree of agreement between MODIS and the model results is represented by the hit rate H (Wilks, 1995):

$$H = \frac{a + d}{a + b + c + d} \tag{3.7}$$

with a, b, c, and d defined as in Table 3.4.

Table 3.4: Definitions for Eq. 3.7

	MODIS snow	MODIS no snow
model snow	a	b
model no snow	c	d

Results show that the overall accuracy of the snow model is good with 98% of the model area having a hit rate H larger than 70% for the winter seasons of 2003-2009. 88% of the model area has a hit rate H larger 80% and 1.5% of the model area has have a hit rate H between 60 to 70%. The highest hit rates occur in the high elevations in the western part of Austria, which is not surprising as the snow model simulates snow in high elevations quite accurately and MODIS indicates snow cover in these areas. In the low parts of the model area, the hit rate H is also high (around 90%). This can be attributed to the fact that the model simulates no snow and the MODIS data confirm this. The medium elevation ranges show a smaller hit rate (around 70-80%), which is related to forest cover in these areas.
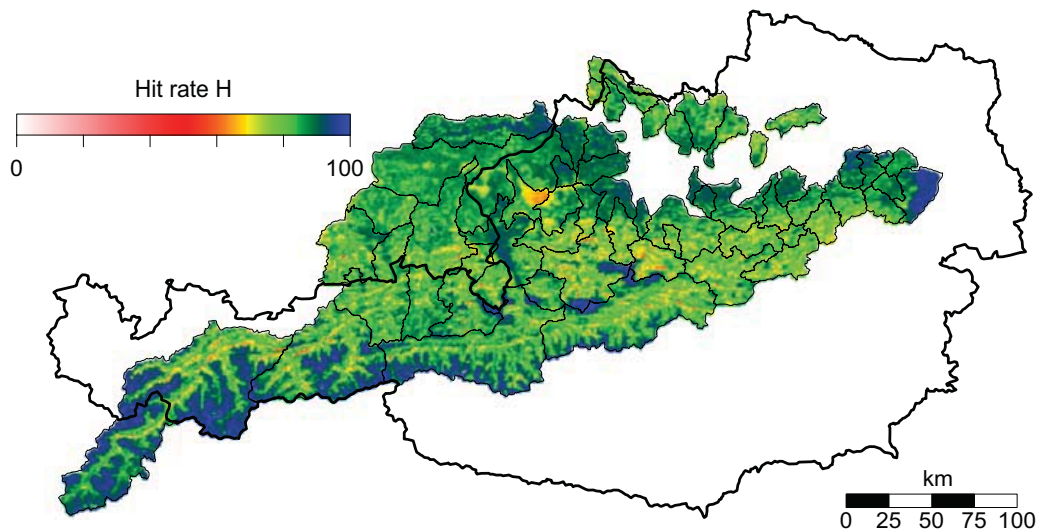
Figure 3.11: Hit rate H of snow simulations for the winter season (October-May) in the period 2003-2009. Days with clouds are not considered. Threshold $\xi_{SWE} = 2.5\,\mathrm{mm}$.

3.5 Discussion and conclusions

We have evaluated the snow model component of a conceptual semi-distributed hydrologic model run on an hourly time step. The model is used operationally for flood forecasting, but for this study we used it in a simulation mode with historical data. Parameters of the hydrologic model have been calibrated for 57 catchments manually in a three step routine. The annual water balance is approximated in a first step, seasonal patterns of streamflow were sought to be modelled correctly in a second step. The third step included the parameterization of the fast flow components to correctly estimate the timing of the rising limbs and the peaks. This approach assured to account for different hydrological situations throughout the catchments.

A comparison of the temporal evolution of the snow covered area (SCA) derived from MODIS data and SCA estimated from model results indicates good agreement between observed and simulated SCA values when cloud coverage is less than 80%. The timing of the snow accumulation and depletion periods is simulated well. However, discrepancies between model and MODIS are observed at the beginning and end of each snow season, but, as Klein and Barnett (2003) noted, *"This is perhaps not surprising as at these times, when snow would be expected to be thinnest and most patchy."*

We further evaluated the performance of the snow model using various error measures (Parajka and Blöschl, 2008b). Simulated snow water equivalents (SWE) and snow covered area estimated from a combination of MODIS (version 5) Terra and Aqua snow cover maps (Parajka and Blöschl, 2008a) were compared for each day on the catchment scale. The selection

of the cloud threshold ξ_C was found to be the most important factor for the evaluation of the snow model. Previous studies have used different thresholds for cloud coverage. E.g., Şorman et al. (2009) used a threshold of 20% to calibrate a hydrologic model on streamflow and SCA data, Hall et al. (2010) used MODIS data with cloud coverage up to 80% for the development of a cloud-gap-filled MODIS daily snow-cover product. Rodell and Houser (2004) used MODIS data when only 6% of the MODIS maps were cloud free. Results indicate that the errors are similar for CC $< 50\%$ and for $50\% < $ CC $< 80\%$, but there are clear differences for CC $> 80\%$. Therefore, we propose to use a threshold value $\xi_C = 80\%$. The remaining threshold values were selected based upon a sensitivity study. It showed that results are less sensitive to the threshold for SWE than they are to the threshold for SCA. Based on the sensitivity study we selected threshold values $\xi_{SWE} = 2.5\,\mathrm{mm}$ and $\xi_{SCA} = 30\%$ for the error analysis. We believe that the chosen thresholds can be also used as in regions with similar physiographic characteristics. The value chosen for ξ_{SWE} is within the range of threshold values in the literature (i.e., Tong et al., 2009a; Roy et al., 2010); whereas the value chosen for ξ_{SCA} is slightly larger than the threshold used in Parajka and Blöschl (2008b). Results indicate that snow underestimation errors are larger than snow overestimation errors and that the thresholds have to be chosen with care as they have a large impact on the snow model efficiency. Interestingly, overestimation errors are not as sensitive to the threshold for cloud coverage ξ_C as are underestimation errors, and the seasonal error distribution shows that the model tends to underestimate snow in the summer months. This may be due to misclassification of clouds as snow in the MODIS data (Parajka and Blöschl, 2006). Parajka and Blöschl (2006) state, *"This misclassification occurs frequently, but tends to affect only a small area"*. The error distribution as a function of elevation shows that larger underestimation errors occur in prealpine regions.

We also compared the spatial extent of simulated SCA and MODIS SCA data on a pixel basis taking into account only cloud-free pixels. Generally, the snow model performance can be classified as good for the winter periods from 2003-2009. 88% of the model area is correctly classified as snow covered or snow free on more than 80% of the days. This value is similar to Strasser and Mauser (2001) who have shown an accuracy of 84% in their study in Northern Germany and Zappa (2008) who has shown an accuracy of 87% in a study in Switzerland. However, there are some discrepancies between simulated and observed SCA. The spatial evaluation indicates that at very high altitudes, the model tends to simulate snow on more days than what MODIS observes. This is in line with Koboltschnig et al. (2008). They contributed this to the fact that ridges and steep slopes at high altitudes are snow covered in the simulation whereas snow is expected to be blown away or redistributed by avalanches. In the transition zones from lowland to alpine areas, the model tends to underestimate the quantitiy of snow covered days. Two reasons may contribute to this; elevation changes not accounted for in the model structure; and underestimation of snow cover in forested areas. Analyses indicate that 500 m elevation zones may not be detailed enough to estimate SWE in the transition zones accurately. In this context Zappa (2008) noted that the disagreement in the transition zones may be due to uncertainties in observed precipitation, local and regional temperature gradients and in the model parameters. The underestimation of the remote sensing SCA in forested areas has already been shown in Simic et al. (2004) and the MODIS

product summary page (MODIS, 2010) states that *"the maximum expected errors are 15 percent for forests, 10 percent for mixed agriculture and forest, and 5 percent for other land covers."* The performance of the snow model in the transition zones could perhaps be improved using a spatially distributed model and a process based analysis of snow distribution patterns as proposed by Sturm and Wagner (2010). To confirm this, further analyses are required.

Overall, the comparison of simulated and observed snow covered area facilitated useful insights into the model performance as a function of space and time as well as others factors. Because of the usefulness in model evaluation we would expect the snow cover data to be equally useful for data assimilation in a real time mode. This will be examined in future studies.

4 Flood forecast errors and ensemble spread - a case study

Abstract

Flood forecasts are generally associated with errors, which can be attributed to uncertainties in the meteorological forecasts and the hydrologic simulations, and ensemble spreads are usually considered capable of representing them. To quantify these two components of the total forecast errors and to compare these to ensemble spreads, a four year data set of operational flood forecasts with lead times up to 48 hours is evaluated for 43 catchments in Austria and Germany. Catchment sizes range from 70 to 25600 km^2, elevations from 200 to 3800 m and mean annual precipitation from 700 to 2000 mm. A combination of ECMWF and ALADIN ensemble forecasts are used as input in a semi-distributed conceptual water balance model on an hourly time step. The results indicate that, for short lead times, the ratio of hydrologic simulation error to precipitation forecast error is 1.2 to 2.7 with increasing catchment size from 100 to 10000 km^2. For long lead times the ratio of hydrologic simulation error to precipitation forecast error decreases from 1.1 to 0.9 with increasing catchment size. Clear scaling relationships of the forecast error components with catchment area are found. A similar scaling is also found for ensemble spreads, which are shown to represent quantitatively the total forecast error when forecasting floods.

4.1 Introduction

One of the main challenges in flood forecasting and warning is to extend forecast lead times beyond the catchment response time as the forecasts will then rely on rainfall predictions (Blöschl, 2008). This is of particular concern in small and medium sized catchments where the catchment response times are short and more time is required for flood response actions.

As the forecast lead time increases the forecast errors tend to increase. For flood response actions it is therefore essential to get an indication of the magnitudes of the forecast errors to be expected at any point in time (Montanari, 2007). There are two main sources of uncertainty that contribute to the flood forecast errors: precipitation forecast errors and hydrological simulation errors (Krzysztofowicz, 2001). The precipitation forecast errors represent the differences between predicted and observed (and interpolated) precipitation. Because of the

non-linearity of the atmospheric system the errors tend to increase drastically with the forecast lead time. The common approach to quantifying the precipitation forecast uncertainty are ensemble simulations where a numerical weather prediction (NWP) model is run for a number of cases with slightly different initial conditions. The cases evolve along different trajectories which produce a range of precipitation forecasts (Buizza, 2003; Grimit and Mass, 2007). The different cases or ensemble members of precipitation are used as inputs into a hydrological model to produce a range of flood forecasts (Demeritt et al., 2007). The spread of the ensemble members in terms of flood discharge is then used as a measure of forecast uncertainty due to uncertain precipitation forecasts (e.g., Pappenberger et al., 2005). These types of ensemble forecasts have been performed for short range (around 48 hours) (e.g., Komma et al., 2007; Thirel et al., 2008), and medium-range (up to 15 days) forecasts (e.g., Gouweleeuw et al., 2005; Roulin and Vannitsem, 2005; Roulin, 2007; Verbunt et al., 2007; Thielen et al., 2009; Hopson and Webster, 2010). A review of ensemble flood forecasting systems and the additional value of using ensembles is given in Cloke and Pappenberger (2009). Most of these studies assume that the precipitation forecast uncertainty is the main source of uncertainty impacting on the flood forecasts.

Hydrological simulation errors have been the subject of numerous studies in hydrology (see e.g., Montanari and Brath, 2004; Montanari et al., 2009). The hydrological simulation errors represent the differences between predicted and observed runoff using observed (and inter-polated) precipitation. The errors are usually classified into input errors, model parameter errors and model structure errors. There are a range of methods of quantifying the first two types of errors including Monte Carlo simulations and analytical approaches (Montanari et al., 2009). A typical representative of a method of estimating the simulation error is Montanari and Grossi (2008) who infer the probability distribution of the error through a multiple regression with current forecasted discharges, past forecast error and past rainfall. Krzysztofowicz and Kelly (2000) presented Bayesian theory and a meta-Gaussian model for estimating the hydrological simulation error. A few studies have combined the precipitation forecast uncertainty and hydrological simulation uncertainty. Krzysztofowicz (2001), for example, presented an analytic-numerical of combining the two uncertainties resulting in a probability of forecast river stages that is a mixture of two distributions related to occurrence and non-occurrence of precipitation. However, most ensemble flood forecast systems focus on the precipitation forecast uncertainty alone.

In practice, the ensemble spread of the flood forecasts is often interpreted as an index of forecast errors rather than as a quantitative estimate of the errors. However, it is also of interest to understand how well the ensemble spread matches the actual forecast errors. Ensemble forecasts of precipitation are generally used based on the assumptions that the members of the ensemble are equally likely and the ensemble spread captures the precipitation forecast uncertainty. However, this is often not the case. For example, Schaake et al. (2004) analyzed precipitation ensemble forecasts of the US National Centers for Environmental Prediction (NCEP) over the period 1997-1999 over the US. They found that the ensembles were biased and the spread was insufficient to capture measured precipitation. They proposed different methods for pre-processing the precipitation forecasts in order

to remove bias and adjust the ensemble spread. Similarly, Scherrer et al. (2004) analyzed ensemble predictions of European Centre for Medium-Range Weather Forecasts (ECMWF) against precipitation observed at a rain gauge in Switzerland and found significant bias which they corrected with a neural network method. Buizza et al. (2005) compared the ensemble spreads of the methodologies used at the European Centre for Medium-Range Weather Forecasts (ECMWF), the Meteorological Service of Canada (MSC), and the US National Centers for Environmental Prediction (NCEP) for a 3-month period in 2002. Again, for all systems, the spread of ensemble forecasts was insufficient to capture reality, suggesting that postprocessing of the ensemble estimates is needed. Similar conclusions were arrived at by Hamill et al. (2008) for the ECMWF and the Global Forecast System (GFS) analyzing a longer time period but post-processing methods do not always improve the ensemble forecasts of precipitation (Schmeits and Kok, 2010).

The biases and uncertainties of precipitation forecasts may amplify when cascaded through the hydrological system. Komma et al. (2007) showed that for their flood forecasts the variability of the precipitation ensemble was amplified for lead times longer than the response time of the catchment. They used a combination of ECMWF and ALADIN ensemble forecasts as input into a distributed hydrologic model in a 620 km^2 catchment in Austria. For a lead time of 48 hours, for example, an uncertainty range of 70% in terms of precipitation translated into an uncertainty range of 200% in terms of runoff. In the context of flood forecasts is therefore important to assess the precipitation uncertainty in terms of the effect on runoff rather than in terms of comparing forecast precipitation against observed precipitation. Johnell et al. (2007) used ECMWF ensemble forecasts to estimate runoff ensemble forecasts using the HBV model for 45 catchments in Sweden with areas ranging from 6 to 6110 km^2 (mean catchment size 647 km^2). They defined the ensemble spread as the range between the upper and the lower quartile of the runoff ensemble forecasts and compared it to the mean absolute error of the median runoff ensemble forecast. They classified the forecasts into five classes representing "very small" to "very large" ensemble spread with each class containing 20% of the forecasts. The mean average forecast error increased from 2% for the class of very small ensembles to 18% for the class of very large ensembles on forecast day 1, and from 10% for very small ensembles to 75% for very large ensembles on forecast day 9. Errors for the forecast days 5 to 7 were similar, indicating that the EPS forecast has its main strength in the second part of the forecast period. Jaun and Ahrens (2009) estimated the runoff ensembles using downscaled ECMWF ensemble forecasts as input into the PREVAH model for 23 Swiss catchments with areas ranging from 610 to 34550 km^2 (mean 6000 km^2). They defined the ensemble spread as the half interquartile runoff ensemble range and compared it to the forecast error obtained by comparing the median runoff ensemble forecast error with OBS and REF, respectively. They found a tendency towards underestimation in the forecast spread when evaluating against OBS with positive forecast errors 1.5 to 5 times and negative forecast errors 1.5 to 100 times larger than the ensemble spread. When comparing against REF, the underestimation in the ensemble spread was smaller with factors around 1.5 to 2 for both positive and negative forecast errors. However, additional uncertainties during unstable weather situations were found to be captured by larger ensemble spreads during flood peaks.

A comparison with OBS and REF also allows separating the contributions of the precipitation forecast errors and the hydrological simulation errors to the total forecast errors. Understanding the relative contributions can assist in the future development of ensemble flood forecasting systems. Olsson and Lindström (2008) compared ensemble runoff forecasts in 45 Swedish catchments using forecasted ECMWF precipitation with observed runoff and a reference runoff simulation using observed precipitation. They found that 26% of the runoff forecasts were within the interquartile range of the ensemble spread when comparing the forecasts to the simulations. This means that the precipitation ensembles are too narrow as the figure should be 50%. However, only 14% of the forecasts were within the interquartile range of the ensemble spread when comparing the forecasts to observed runoff highlighting the contribution of the simulation error not accounted for in the runoff ensembles. They concluded that the contributions of the precipitation forecasts and the hydrologic simulations to the total error were of similar magnitudes but, for the rising limbs (i.e., the targets of flood forecasting systems), the precipitation forecast errors dominated. A similar analysis of separating the contributions of the precipitation forecast errors and the hydrologic simulation errors was performed by Addor et al. (2011) who analyzed ensemble runoff forecasts in a 336 km^2 catchment in Switzerland based on precipitation ensemble forecast of a regional climate model (COSMO-LEPS). They found around 14% of the forecasts were within the interquartile range of the ensemble spread when comparing forecasts to observed runoff, but close to 50% of the forecasts were within the interquartile range of the ensemble spread when comparing the forecasts to simulations. Zappa et al. (2011) superposed different sources of uncertainty in a flood forecasting system. They used COSMO-LEPS forecasts and the hydrological model PREVAH for a 186 km^2 catchment in Switzerland. The uncertainty from the meteorological forecasts was represented by the uncertainty of the COSMO-LEPS ensembles propagated through the hydrologic model. The hydrologic model uncertainty was taken into account using a Monte Carlo simulation in which seven parameters of the hydrologic model (relevant for surface runoff generation) were randomly changed. The average runoff ensemble spread for the seven events analyzed was 130 m^3/s when PREVAH was coupled with LEPS. When additionally taking the hydrological uncertainty into account the ensemble spread obtained was around 150 m^3/s, meaning that the total uncertainty increased about 15%.

The two main objectives of this study are (1) to quantify the contributions of precipitation forecast errors and hydrologic simulation errors to the total forecast error, particularly during flood events, and (2) to evaluate the capability of the runoff ensemble forecasts to represent the total runoff forecast errors as a function of lead time. We use a conceptual semi distributed hydrological model (Blöschl et al., 2008) coupled to meteorological inputs (deterministic and ensemble forecasts) based on a combination of ECMWF and ALADIN forecasts (Haiden et al., 2010). Around 4 years of forecasts and runoff data at hourly time scale for 43 catchments with areas ranging from 70 to 25600 km^2 in Austria and Germany are analyzed. Such an extended data base allows us to identify scaling properties (with catchment area and lead time) of the total forecast error, of its two components (precipitation forecast and hydrologic simulation errors) and of the runoff ensemble forecasts.

4.2 Study region, data and meteorological forecast inputs

In this study we evaluate the hydrologic forecasts of the flood forecasting system for the Austrian Danube that is currently in operational use and has been developed by the Technical University of Vienna in 2003-2005. The system consists of (1) a meteorological, (2) a hydrologic and (3) a hydraulic model part. The meteorological forecasts include deterministic and ensemble forecasts of precipitation and deterministic forecasts of air temperature for a lead time of 48 hours on an hourly time step; the output of the hydrologic model includes deterministic and ensemble runoff forecasts in the Danube tributaries which are used to run a hydraulic model to estimate runoff and water level for the Danube River.
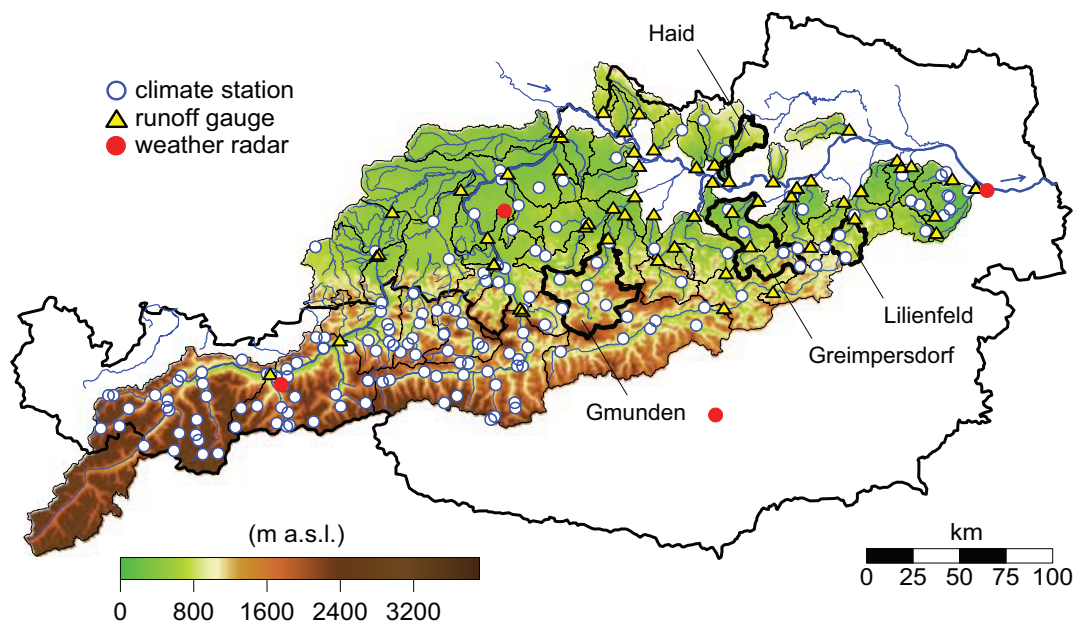
Figure 4.1: Topography of Austria and parts of Southern Germany. The stream gauges used in the study are indicated by triangles, precipitation gauges by white circles, weather radar stations by red circles. Thin black lines are catchment boundaries, the thick black lines highlight the catchments Gmunden/Traun, Greimpersdorf/Ybbs, Haid/Naarn and Lilienfeld/Traisen, used for detailed analyses.

The region is hydrologically diverse covering large parts of Austria and parts of Bavaria (Figure 4.1). The West of the region is Alpine with elevations of up to 3800 m a.s.l. while the North and East consist of prealpine terrain and lowlands with elevations between 200 and 800 m a.s.l. Mean annual precipitation is between 600 mm/yr in the East and almost 2000 mm/yr in the West. The Alpine catchments generally show much higher runoff depths with 1600 mm/yr, compared to around 100 mm/yr in the East. Runoff from 43 catchments with sizes ranging from 70 to 25600 km^2 (median size around 400 km^2) in the study region are used for the evaluation. The small catchments are mostly nested catchments. Land use

is mainly agricultural in the lowlands, forested in the medium elevation ranges and alpine vegetation, rocks and glaciers in the alpine catchments. For the calibration of the model, meteorological and hydrologic data from 2002-2009 were used. For details we refer to Nester et al. (2011a). In this paper, the analyses of the forecasts are based on a data set consisting of four years of meteorological forecasts (2006-2009). Around 35000 time steps are analyzed in each catchment. Results are presented for all catchments as well as, in more detail, for four catchments of different size (300 to 1390 km^2) in diverse hydrologic regions (wet and relatively dry). Catchment characteristics and model performance in terms of the Nash-Sutcliffe model efficiency ($nsme$) and the volume error (VE), which is used as a measure of bias, of these four catchments for the calibration and validation periods (2003-2006 and 2007-2009, respectively) are summarized in Table 4.1.

Table 4.1: Stream gauges used for detailed analyses. MAP and MAR (2002-2009) is mean annual precipitation and runoff in (mm/yr), respectively. $nsme$ stands for Nash-Sutcliffe model efficiency, VE stands for volume error; values are for calibration period/validation period (from (Nester et al., 2011a)

Gauge/catchment	area (km^2)	MAP	MAP	$nsme$	VE
Haid/Naarn	306	915	380	0.75/0.65	0.19/0.19
Lilienfeld/Traisen	333	1440	860	0.84/0.69	-0.12/-0.09
Greimpersdorf/Ybbs	1116	1480	840	0.86/0.80	-0.02/0.02
Gmunden/Traun	1390	1810	1425	0.76/0.78	-0.01/-0.11

The meteorological data and forecasts were provided by the Central Institute for Meteorology and Geodynamics (ZAMG) in Vienna and are discussed in detail in Haiden et al. (2011).

Observed precipitation fields: For each time step, rain gauge data were spatially interpolated on a 1 km grid and combined with radar data as a weighted mean. Currently, 408 online available climate stations are implemented in INCA; 169 of which lie within the model region, which equals to one climate station every 258 km^2. In small catchments, on average 0.35 stations per 100 km^2 are available whereas in large catchments on average 0.45 stations are available per 100 km^2. 70% of the stations are below 1000 m a.s.l., 24% are between 1000 and 2000 m a.s.l. and the remaining 6% are above 2000 m a.s.l. with the highest station at 3100 m a.s.l. The weights were derived from a comparison of monthly totals of radar and rain gauge data at the rain gauge locations. In areas where the visibility of the radar is low the radar weights are small while in areas with high radar visibility the weights are close to 1.

Precipitation forecasts: Deterministic precipitation forecasts are generated over a lead time of 48 hours consisting of two components. The first component, termed nowcasts, is obtained by extrapolating the interpolated precipitation field using motion vectors (Steinheimer and Haiden, 2007). The second component consists of the forecasts of the ALADIN and ECMWF numerical weather prediction (NWP) models. The two components are combined as a weighted mean. To allow for a smooth transition between nowcasts and NWP results, the weights are varied as a function of lead time

from full weight to the nowcasts during the first 2 hours, full weight to the NWP forecasts from 6 hours, and a linear transition in between.

Ensemble forecasts of precipitation: The ensemble forecasts of precipitation consist of three components. The first and second components are the nowcasts based on the motion vectors and the deterministic precipitation forecast of the ALADIN model from ZAMG, as for the precipitation forecast. On top of that, to account for small scale spatial uncertainty, the ALADIN forecast are spatially shifted in both the x and y directions to produce 25 pseudo-ensembles. The third component consists of 50 ensemble forecasts from the ECMWF model. The 50 ECMWF ensembles are randomly combined with one of the ALADIN pseudo-ensemble members and with the nowcasts. No uncertainty is assigned to the nowcasts, meaning that up to a lead time of 2 hours all ensemble members are identical (zero spread) and the spread increases at longer lead times (Komma et al., 2007). A verification of INCA precipitation forecasts with a lead time of 12 hours showed that in the first 6 hours the precipitation amount is underestimated, whereas in the second half of the forecast period the precipitation amount is overestimated which can be attributed to the increasing influence of the NWP models for longer lead times (Haiden et al., 2011).

Temperature forecasts: Temperature forecasts are based on a combination of interpolated station data and ALADIN forecasts. No temperature ensembles are generated as their effect on the flood forecasting uncertainty is deemed to be small.

We used catchment mean values of precipitation as input into the hydrologic model; for the temperature data elevation was additionally accounted for when averaging over the catchments. All forecast are generated at an hourly time interval.

For the analysis, hourly discharge data from 43 stream gauges were used. The data were checked for errors and in cases where a plausible correction could be made they were corrected. Otherwise they were marked as missing data.

4.3 Forecast model setup and evaluation methods

Forecast model setup

The rainfall-runoff model used in this paper is a typical conceptual hydrologic model (Blöschl et al., 2008; Komma et al., 2008). The model runs on an hourly time step and includes a snow routine, a soil moisture routine and a flow routing routine (Szolgay, 2004). Details about calibration and performance of the model are given in Nester et al. (2011a,b). Two real-time updating procedures are implemented to increase the accuracy of the forecasts. The first procedure is based on Ensemble Kalman filtering and is used to assimilate runoff data to update the catchment soil moisture (Komma et al., 2008). The second procedure is an additive error model that exploits the autocorrelation of the forecast error and involves an exponential decay of the correction (Komma et al., 2007). The flood forecasting system has

been in operational use for the Danube since 2006 and is operated by the state governments of Lower Austria and Upper Austria.

In this study, we emulate the real-time mode of flood forecasting in the Austrian Danube tributaries. We therefore use the same updating procedures as the operational system. For each forecasting time step t_0 the model is driven by observed precipitation and air temperature. Observed runoff for the same time step is used for the updating. The runoff forecasts are driven by (a) deterministic precipitation forecasts (and air temperature forecast) and (b) 50 ensemble members of the ensemble precipitation forecast. The latter give an uncertainty distribution of the runoff forecasts over the lead time. In line with other forecasting systems, the uncertainty in the precipitation forecasts was considered as the only source of runoff forecast uncertainty.

Examples of ensemble runoff forecasts for the catchment Greimpersdorf are given in Figure 4.2 for an event in June 2009. This event was the largest observed event in the study period with a return period of 35 years which was due to local convective storms embedded in a large scale precipitation field. The forecast calculated on June 22, 2009 at 3 a.m. is shown in the top part of Figure 4.2. The runoff at the time of the forecast was around 30 m^3/s. The ensemble forecasts give a total precipitation between 60 and 100 mm in 48 hours while the observed precipitation was 136 mm. The underestimation is due to heavy convective storms not captured by the precipitation forecasts. This leads to a significant underestimation of runoff over much of the lead time and to missing the sudden rise in the hydrograph at time 15 hours. However, at the end of the 48 hour lead time, the runoff of the deterministic forecast run is almost at the level of the observed runoff, as are some ensemble members. 24 hours later (Figure 4.2, bottom), the cumulative ensemble forecasts of precipitation range between 84 and 121 mm with an observed precipitation of 116 mm. While the fine scale structure of the event is not fully captured, the overall shape of the hydrograph is captured very well. In this case, the ensemble runs give a very good indication of the forecast errors to be expected. The example illustrates that the performance of the flood forecasts will likely differ between events and can change within a single event. In some cases the ensembles will be representative of the errors but in others they will not.

Forecast evaluation methods

For a first overview of the performance of the ensemble runoff forecasts we used the Brier Score (BS) measure (Brier, 1950)

$$BS = \frac{1}{N} \sum_{1}^{N} (p(t\mathord{*}) - o(t\mathord{*}))^2 \tag{4.1}$$

where p is the forecast probability from the ensemble forecast for time $t\mathord{*}$ of exceeding a threshold discharge, o is a binary value depending on whether the observed discharge at the time step $t\mathord{*}$ exceeds the threshold discharge ($o = 1$) or does not ($o = 0$), and N is the
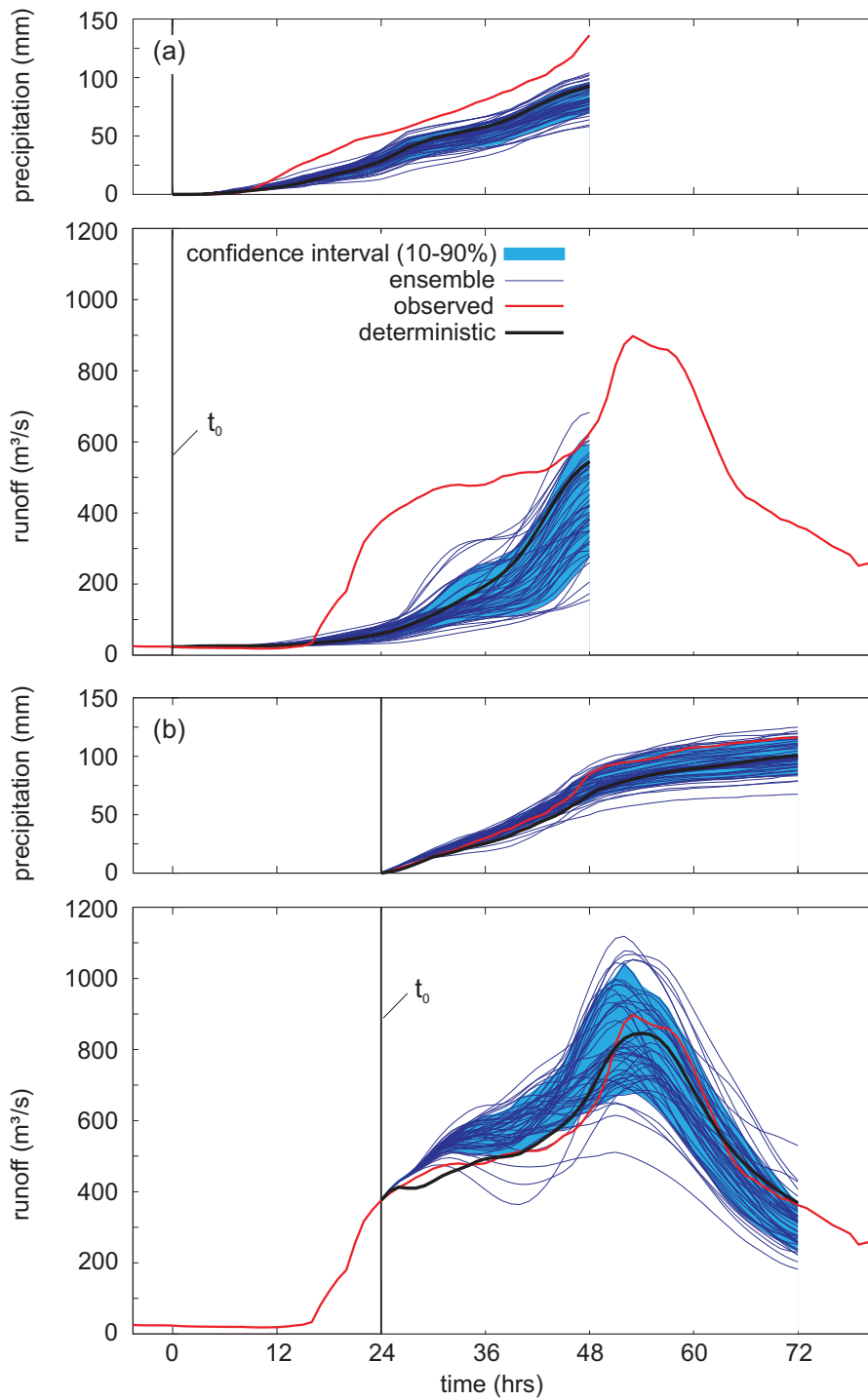
Figure 4.2: Example for ensemble forecasts at the gauge Greimpersdorf/Ybbs (1116 km^2) Return period is T=35 years. (a) forecast time t_0 is June 22, 2009, 3 a.m.; (b) t_0 is June 23, 2009, 3 a.m. Top panels - cumulative precipitation, bottom panels - runoff. Red lines are the observations; black lines are the deterministic forecasts; thin blue lines are ensemble forecasts; 80% confidence intervals are indicated in light blue.
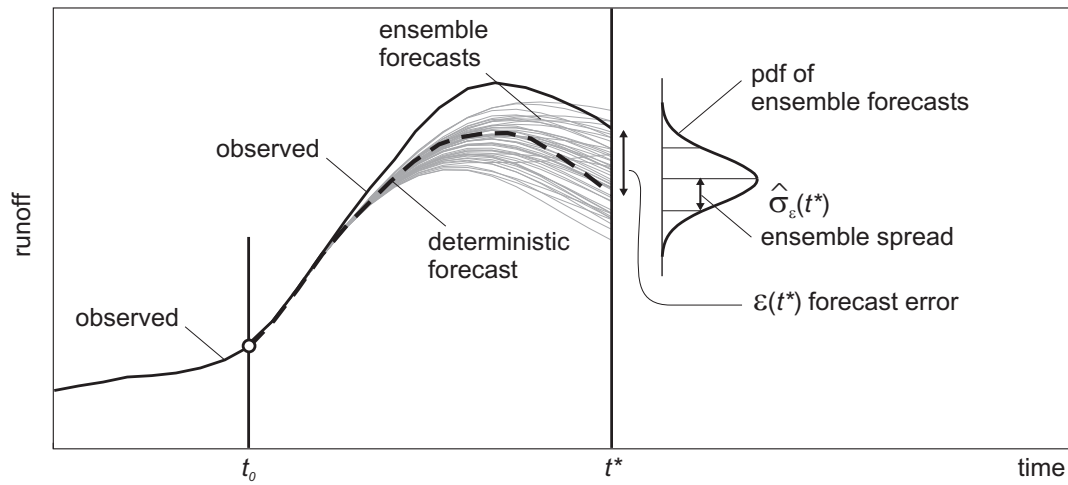
Figure 4.3: Definitions of the time t_0 the forecast is made and the time t^* for which the forecast is made. The ensemble spread at time step t^* is denoted as $\hat{\sigma}_\varepsilon(t^*)$, the forecast error as $\varepsilon(t^*)$

total number of forecasts analyzed. BS ranges from 0 to 1. $BS = 0$ indicates a perfect forecast as the probability of the ensemble forecast matches exactly the frequency of the runoff observations exceeding a certain threshold. The Brier Skill Score (BSS) measures the improvement of a probabilistic forecast relative to a reference forecast BS_{ref}

$$BSS = 1 - \frac{BS_f}{BS_{ref}} \tag{4.2}$$

The range of the BSS is $-\infty$ to 1, with the best score equal to 1. Positive scores indicate an improvement over the reference forecast. To compare our results with other studies (e.g., Rousset-Regimbeau et al., 2007; Thirel et al., 2008), we used the 50th and the 90th percentile derived from observed runoff data as a reference forecast to calculate the BSS.

For a more detailed evaluation of the ensemble runoff forecasts, we used a spread-skill analysis (e.g., Scherrer et al., 2004; Lalaurette et al., 2005). Figure 4.3 gives the definitions of the terms used. t_0 denotes the time the forecast is made and t^* refers to the time of the predicted runoff. We examined forecasts lead times ($t^* - t_0$) of 1, 3, 6, 12, 24 and 48 hours. As a measure of the ensemble spread we used the standard deviation $\hat{\sigma}_\varepsilon(t^*)$ of the runoff of the ensemble members for each point in time t^*. As a measure of skill we defined the forecast error $\varepsilon(t^*)$ as the difference between the observed runoff and the deterministic forecast.

In order to distinguish between different forecast situations, we stratify the analysis of forecast errors in classes of ensemble spreads. Small ensemble spreads are more likely to occur when the runoff is constant or falling while large ensemble spreads are more likely to occur when runoff is growing. Since floods are of interest here, the analysis of forecast errors associated

to large ensemble spreads will be emphasized in the following. We grouped the forecast time steps $t*$ (for each lead time separately) into 10 classes according to the ensemble spread $\hat{\sigma}_\varepsilon (t*)$ of that time step. For example, class 1 represents 10% of the time steps with the smallest ensemble spread and class 10 represents 10% of the time steps with the largest ensemble spread. Each class had the same number of time steps of n. For each class j we aggregated the ensemble spread $\hat{\sigma}_\varepsilon (t*)$ over n time steps:

$$\bar{\hat{\sigma}}_{\varepsilon,j} = \frac{1}{2} \cdot (\max \hat{\sigma}_{\varepsilon,j} - \min \hat{\sigma}_{\varepsilon,j}) \tag{4.3}$$

Similarly, we calculated the standard deviation $\sigma_{\varepsilon,j}$ of the forecast error $\varepsilon (t*)$ over the same time steps:

$$\sigma_{\varepsilon,j} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (\varepsilon(t*) - \bar{\varepsilon})} \tag{4.4}$$

with $\bar{\varepsilon}$ as the mean forecast error over the n time steps. If the ensemble forecasts fully portray the forecast errors, $\bar{\hat{\sigma}}_\varepsilon = \sigma_{\varepsilon,j}$ for all classes j. Any deviations from the full match indicates biases in the ensemble forecasts. It is worth noting that, differently from the BSS measure, $\sigma_{\varepsilon,j}$ and $\bar{\hat{\sigma}}_\varepsilon$ do not account for biases of the forecast estimation. By comparing $\sigma_{\varepsilon,j}$ and $\bar{\hat{\sigma}}_\varepsilon$ we aim to assess if the runoff ensemble forecasts and the total runoff forecast errors have the same spread and therefore the first can be deemed representative of the second.

The other main objective of the paper is to quantify the contributions of precipitation forecast errors and hydrologic simulation errors to the total forecast error. We analyzed the errors for two cases of runoff forecasts:

(a) In the first case we used forecasts of deterministic precipitation as an input to the runoff model.

(b) In the second case we used observed (and interpolated) precipitation data.

These two cases allow us to examine the precipitation forecast errors separately from the hydrological simulation errors. Because of the non-linearity of the runoff processes we consider it more appropriate to test the precipitation forecast errors via their effect on runoff rather than directly by comparing them against rain gauge data. This also allows us to compare the precipitation forecast errors directly with the hydrological simulation errors. This has been done looking in particular to the large of ensemble spread classes, since floods are of interest here.

4.4 Results

Forecast performance by means of the Brier Skill Score

An overall view of the performance of the ensemble runoff forecasts, irrespective of discharge and ensemble spread, is given through the diagrams in Figure 4.4, which show the Brier skill score BSS for all catchments analyzed as a function of lead time. The left panel gives the BSS for a 50% percentile of runoff as reference forecast, this panel therefore relates to medium and low runoff. The right panel gives the BSS for a 90% percentile, i.e., it relates to high flows. For both percentiles the BSS decreases with increasing lead time. Clearly, as the lead time increases, the skill of the ensemble forecasts to match the forecast errors decreases. Overall, the skill for medium runoff (50% percentile) is higher than for high runoff (90% percentile). For most of the catchments the 50% BSS ranges between 0.8 and 0.1 at a lead time of 48 hours while the 90% BSS ranges between 0.6 and 0.0. For high flows it is more difficult for the ensemble forecasts to match the forecast errors than for medium flows.

There are a small number of catchments (e.g., Greimpersdorf indicated by the black dashed line) where the BSS rapidly decreases after the first hour and increases after 12 hours (Figure 4.4, left). This is because low and medium flows in these catchments are influenced by regulations of hydropower plants which have a half-daily cycle not represented in the hydrologic model. For high flows the influence of the regulations is less apparent (Figure 4.4, right). While it would be easily possible to model these cycles provided the information is available from the hydropower operators it is not relevant for the flood forecast which are the purpose of this study.

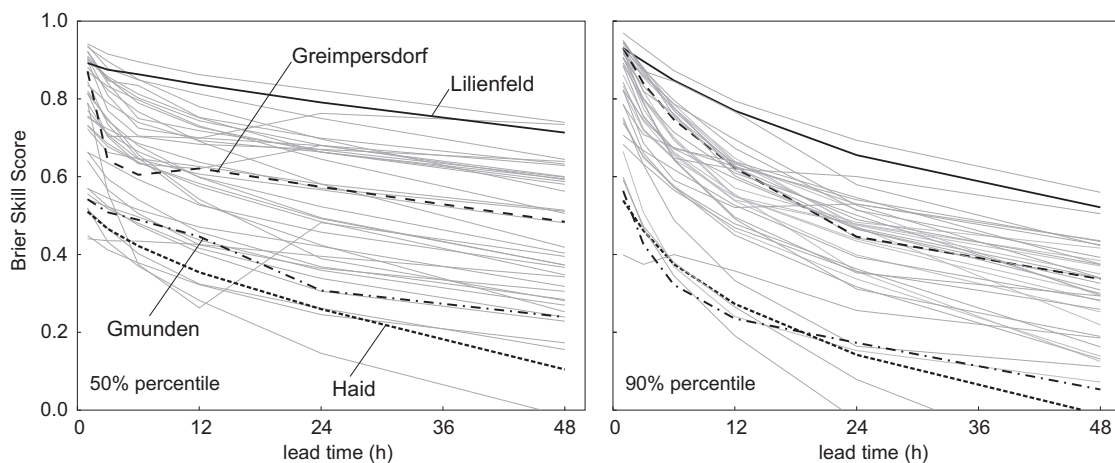Figure 4.4: Brier Skill Score BSS as a function of lead time, computed using observed river flow as reference for a runoff exceeding the 50%-percentile (left) and the 90%-percentile (right). The focus in the left figure is on predicting low and medium runoff, the focus in the right is on predicting high runoff. Thick lines refer to the catchments of Table 4.1, thin lines refer to the other catchments.

Spread-skill analysis

As discussed in Section 4.3, the spread-skill analysis provides more detailed insight into the performance of the ensemble forecasts and the contributions of the various error sources. The midpoint of each class of ensemble spread $\bar{\hat{\sigma}}_\varepsilon$ and the corresponding spread of the forecast error σ_ε are shown as a function of lead time for the four example catchments of Table 4.1. For ease of comparison, both the ensemble spread and forecast errors were scaled by the mean annual runoff of each catchment. All time steps for every lead time were assigned into one of 10 classes of equal size according to the ensemble spread $\bar{\hat{\sigma}}_\varepsilon$. Class 1 represents 10% of the time steps with the smallest ensemble spread and class 10 represents 10% of the time steps with the largest ensemble spread. Large ensemble spreads are more likely to occur when the runoff is rising, whereas small ensemble spreads can be expected when the runoff is receding or at a constant low flow level. Then we calculated the midpoint of the ensemble spreads for each class and the standard deviation of the forecast errors for each class according to Equations 4.3 and 4.4.

Figure 4.5 shows the results for the lead times 12 hours (light grey squares), 24 hours (dark grey triangles) and 48 hours (black circles) hours. If the ensemble standard deviation and the forecast error standard deviation match, the points are close to the 1:1-line. On average, the standard deviation of the forecast error is 2-3 times bigger than the ensemble standard deviation. For short lead times, in particular the ensemble spread is small. This is because the ensemble spread is related to the precipitation forecasts while for lead times shorter than the catchment response time the runoff mainly depends on observed precipitation (Komma et al., 2007). For lead times of 48 hours the errors are twice and up to five times larger than the ensembles. Several reasons contribute to this. First, parts of the errors can be explained by the fact that we used real-time data and we only corrected obvious errors whereas runoff variations on the order of a few m^3/s were not corrected. Second, short events with fast discharge increase the catchments Haid and Lilienfeld are underestimated by the forecasts. Third, the runoff at the gauge Gmunden is influenced by the retention effects of a lake in the catchments. Interestingly the shapes of the forecast errors and the ensembles are remarkably similar. In Haid both the forecast errors and the ensemble spread increases significantly for lead times beyond 24 hours. In Lilienfeld both remain constant. The similarity of the shapes suggests that the ensembles do capture the important characteristics of the errors.

In the catchment Haid (top panel left), the ensemble spread does not capture the forecast errors in the first nine classes. The ensemble spread is around 10 times smaller than the forecast errors whereas in the class with the largest ensemble forecasts the spread is larger than the forecast error by a factor of 2.0-2.5. The forecast errors almost double in this class when increasing the lead time from 24 to 48 hours. The catchments Greimpersdorf and Gmunden show a similar behavior. The ensemble spread in the first 9 classes is up to 5 times smaller than the forecast errors. With increasing ensemble spread this factor decreases and the values are much closer to the 1:1-line, indicating that the ensemble spread is almost able to capture the forecast errors when the spread is large.
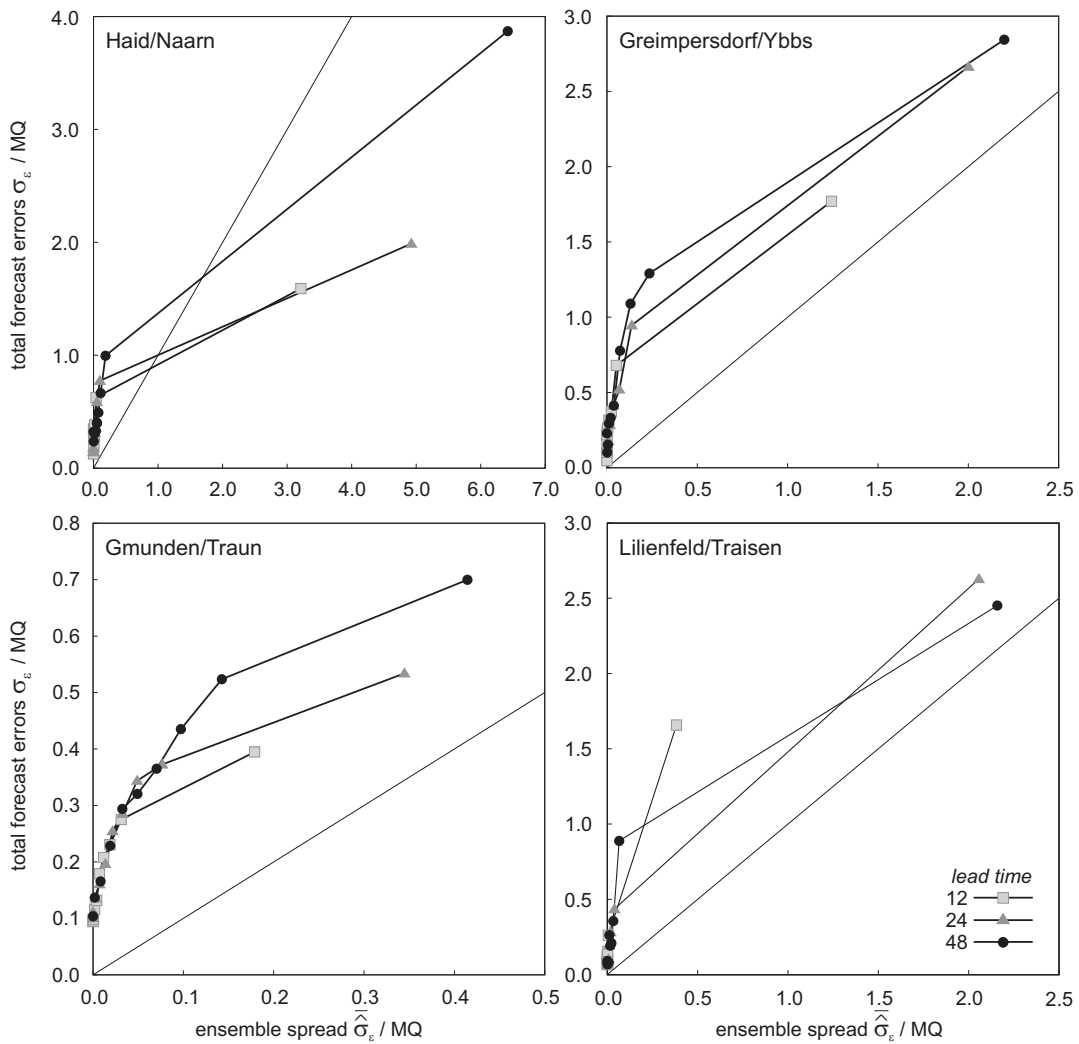
Figure 4.5: Ensemble spread $\bar{\hat{\sigma}}_\varepsilon$ vs. standard deviation of total forecast errors σ_ε, both scaled by mean runoff (MQ). Ensemble spread is plotted at the midpoint of the classes. Light grey squares indicate a lead time of 12 hours, dark grey triangles 24 hours and black circles 48 hours. The thin line is the 1:1-line.

Figure 4.6 shows the CDFs of the Spearman's rank correlation coefficient r_s between the ensemble standard deviation $\bar{\hat{\sigma}}_\varepsilon$ and the standard deviation of the total forecast error σ_ε for all 43 catchments analyzed for different lead times. It shows that for a lead time of 12 hours 75% of the Spearman's rank correlation coefficients rs are larger than 0.88, for a lead time of 24 hours 75% of the values are larger than 0.95 and for a lead time 48 hours 75% of the values are larger than 0.92. This indicates that although the ensemble spreads are too narrow to capture the total forecast errors they still are a good indicator of the forecast error.
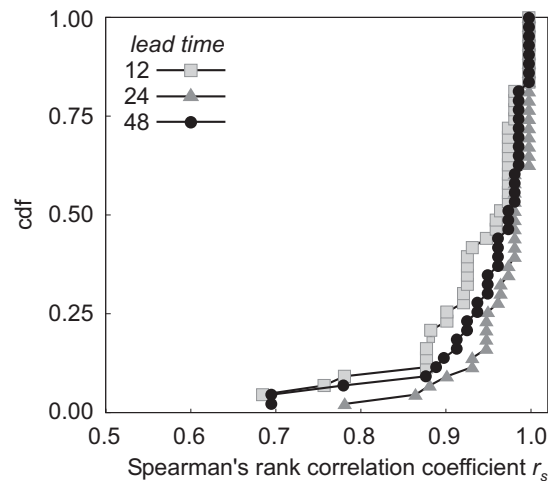
Figure 4.6: CDFs of the Spearman rank correlation coefficient r_s between ensemble spread $\bar{\hat{\sigma}}_\varepsilon$ and total forecast error σ_ε for all 43 catchments analyzed. Light grey squares indicate a lead time of 12 hours, dark grey triangles 24 hours and black circles 48 hours. Each point indicates the Spearman rank correlation coefficient r_s of a single catchment.

Contributions to the forecast error

Figure 4.5 shows the total forecast errors where no distinction between the individual error sources is made. It is now of interest to examine the contributions of the precipitation forecasts and the hydrologic simulations to the total forecast errors. We analyzed the errors for two cases of runoff forecasts: (a) In the first case we used forecasts of deterministic precipitation as input in the runoff model as in Figures 4.5 and 4.6. (b) In the second case we used observed (and interpolated) precipitation data. In the second case, the precipitation forecast error is absent and the entire error is what we term hydrological simulation error. This error is due to precipitation measurement and interpolation, and the structure and the parameters of the runoff model (Table 4.3). The first case also includes error components from the parameters and the structure of the atmospheric model and the initial conditions.

Table 4.2: Contributions to the hydrologic simulation error and precipitation forecast error

Hydrologic simulation error	Precipitation forecast error
Parameters of runoff model	Parameters of atmospheric model
Structure of runoff model	Structure of atmospheric model
Precipitation measurement	Initial conditions
Precipitation interpolation	

For each time step we calculated the differences between observed runoff and the two cases of runoff forecasts. Using the same 10 classes of ensemble spread, we calculated the standard deviation of the hydrological simulation error, σ_{hysim}. Assuming that the precipitation forecast errors and the hydrological simulation errors are independent, the variances are additive and the precipitation forecast error standard deviation σ_{pfor} can be calculated as

$$\sigma_{pforec} = \sqrt{\sigma_{total}^2 - \sigma_{hydsim}^2} \qquad (4.5)$$

The main reason for not directly comparing precipitation forecasts with precipitation measurements was that the time scales relevant for the comparison depend on the catchment response. For example, in a catchment with a fast response, one would have to compare, say, precipitation forecast with a 6 hour aggregation level, while for slowly responding catchments the aggregation level would have to be 24 hours or more. Also, the non-linearity of the runoff processes does not allow a direct comparison of rainfall errors (unit mm) with runoff errors (unit m^3/s). We therefore consider it more appropriate to back-calculate the contribution of the precipitation forecasts from Equation 4.5.

Figure 4.7 shows the contributions to the total forecast errors for the four catchments. For clarity, only the results for the lead time of 48 hours are shown. For small ensemble spreads the entire error is made up of hydrological simulation error. For larger ensemble spreads, the contribution of the precipitation forecast error increases and for the largest ensemble class, it is larger than the hydrologic simulation error. This is the class of major interest because the large ensemble spreads typically occur during the rising limbs of events, which could reveal to be the flood events that we want to forecast, while the small ensemble spreads typically occur during recessions or constant runoff periods. There are some apparent differences between the catchments. In the smaller catchments (Haid and Lilienfeld) the hydrologic simulation error is larger than the precipitation forecast errors for all but the largest ensemble class. For Haid the precipitation forecast errors for of the largest ensemble class is about twice the hydrological simulation error, while for Lilienfeld they are similar. This is because Haid is drier than Lilienfeld (380 mm mean annual runoff as opposed to 860 mm in Lilienfeld), so one would expect larger hydrologic simulation errors (Nester et al., 2011a). The errors in Greimpersdorf are similar to the errors in Lilienfeld. Both catchments have similar mean annual runoff depths and mean annual precipitation, however, Greimpersdorf is 3.5 times the size of Lilienfeld. The precipitation forecast errors are somewhat larger in Greimpersdorf, which can be attributed to the smaller number of precipitation stations per 100 km^2 in the catchments (1.2 in Lilienfeld and 0.4 in Greimpersdorf). The runoff in Gmunden is influenced by the retention effects of a lake, which is not explicitly modeled. However, the lake retention affects mainly low runoff situations, and flood peaks are not affected. This explains that for the small classes of ensemble spread the precipitation forecast error is zero, and the hydrologic simulation error makes up 100% of the total forecast error.

To assess the error contributions for all catchments, the forecast errors scaled with mean catchment runoff have been plotted against catchment size in Figure 4.8 for different lead times. As the main interest in this study is on the forecasting of floods the values of the
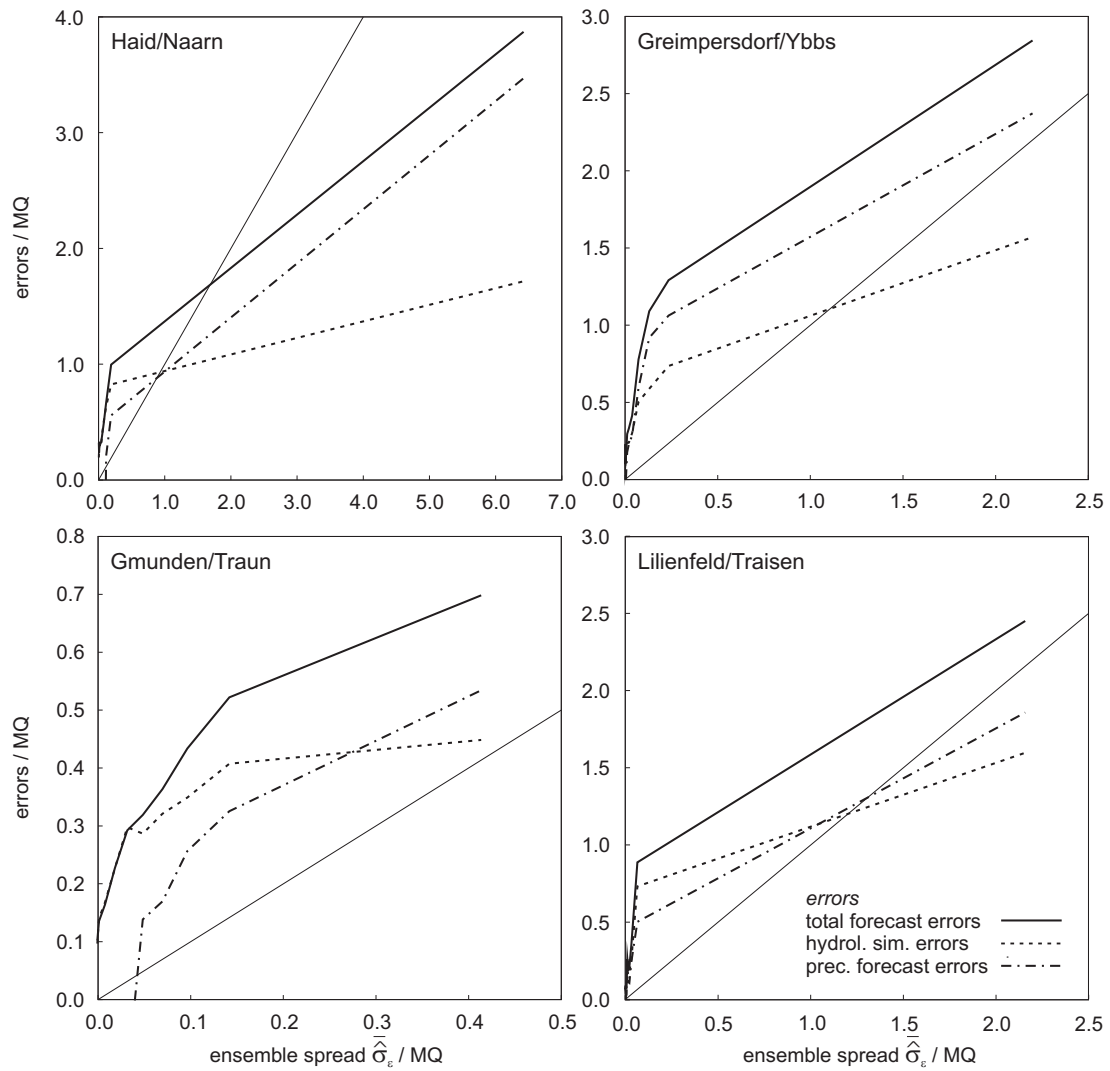
Figure 4.7: Contributions to the forecast error for a lead time of 48 hours for four catchments. Total forecast errors, σ_ε, are indicated by solid lines (same as line with black circles in Fig. 4.5), hydrologic simulation errors σ_{hysim} are indicated by dashed lines and precipitation errors σ_{pfor} by dash-dotted lines. Ensemble spread is plotted at the midpoint of the classes. The thin line is the 1:1-line.

top ensemble spread (largest 10%) are shown. Light grey squares indicate a lead time of 12 hours, dark grey triangles represent a lead time of 24 hours and black circles stand for a lead time of 48 hours. A linear regression was fitted to the errors of the individual catchments in the logarithmic domain:

$$\sigma_\varepsilon = \alpha \cdot A^\beta + \varsigma \tag{4.6}$$

Figure 4.8: Errors scaled by mean catchment runoff vs. catchment area for the top class (largest 10%) of ensemble spreads for 43 catchments. From left to right: Total forecast errors σ_ε, hydrologic simulation errors, σ_{hysim}, and precipitation forecast errors, σ_{pfor}. The regression lines relate to different forecast lead times according to the grey scale. (*) the hydrologic simulation error includes precipitation measurement and interpolation errors.

where α and β are coefficients, A is the catchment area and ς is the error of the regression. The grey shades of the regression lines in Figure 4.8 match those of the symbols for different lead times. All errors decrease very clearly with catchment area, although the rate of decrease differs with the error component and the lead time (Table). The precipitation forecast errors (right panel in Figure 4.8) decrease with catchment area. The precipitation forecast errors also decrease with decreasing lead time (from 48 to 12 hours). As the lead times get close to the catchment response time, any errors of forecasted precipitation will no longer affect the runoff forecasts. This is particularly the case for the large catchments where the response times are longer than in the small catchments. Because of this the 12 hour precipitation forecast errors in the large catchments are very small and therefore the dependence on area is stronger ($\beta = $ -0.695) than for 48 hours ($\beta = $ -0.433). The hydrological simulation errors (middle panel in Figure 4.8) also decrease with catchment area due to the aggregation effects. However, there is much less dependence on the lead time. There is some dependence which is related to the updating of the model. Komma et al. (2008) showed that updating procedures can reduce the error in particular for short lead times. Without updating the model states one would not expect any dependence as these are strictly simulations. The decrease in the errors from 48 to 12 hours (e.g. 1.18 to 1.02 for catchment areas of 1000 km^2) points to the value of the updating procedure for cases when the ensemble spread is large (top 10% of ensemble spreads). For the catchment areas of 10000 km^2 the relative effect of the updating is about twice as big (0.40 to 0.31 for catchment areas of 1000 km^2) which is related to the longer response times and therefore longer autocorrelation in the hydrographs of the large

catchments. The total forecast errors (left panel in Figure 4.8) are the combined results of the two error components. There is again a strong dependence on catchment area and a moderate dependence on the forecast lead time.

It is now of interest to compare the error components as a function of catchment scale and lead time. For the 48 hour lead time the precipitation forecast errors and hydrological simulation errors are of similar magnitudes. As the lead time decreases, the hydrological simulation errors change little while the precipitation forecast errors do, in particular in the large catchments. Obviously, for very short lead times the precipitation forecast errors would be zero. It is important to note, however, that this analysis is for those 10% of the time steps with the largest ensemble spreads, i.e. for a total of 36 days per year which not only includes floods. If individual large events were examined, and in particular the rising limbs, the relative magnitudes of the two error sources may change with the precipitation forecast errors becoming much more important than the hydrological simulation errors (see e.g., Blöschl et al., 2008).

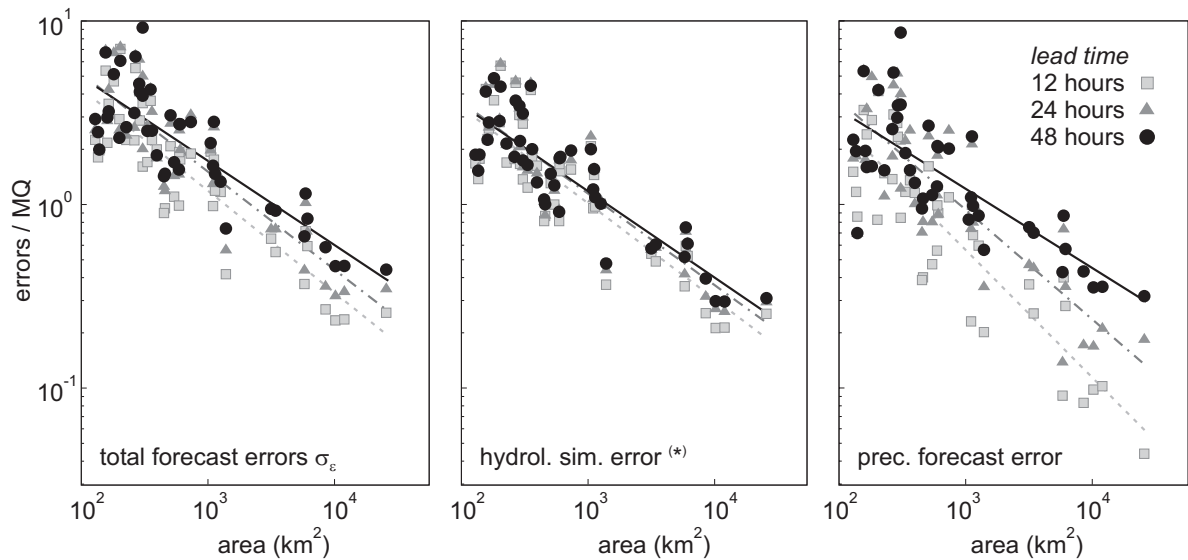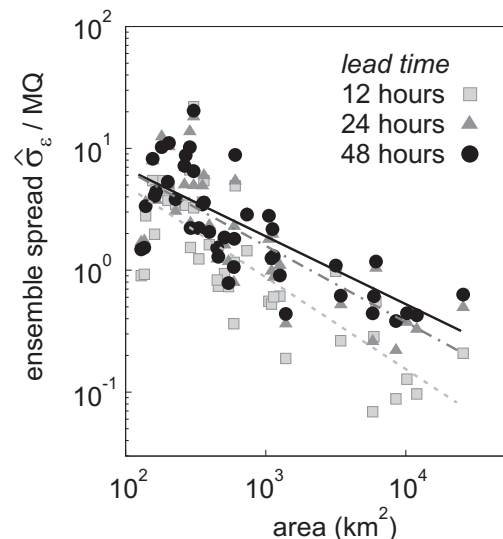Figure 4.9: Ensemble spread $\bar{\hat{\sigma}}_\varepsilon$ scaled by mean catchment runoff vs. catchment area for the top class (largest 10%) of ensemble spreads for 43 catchments.

Figure 4.9 shows a similar analysis as Figure 4.8, however with the ensemble spread plotted against the catchment area. Table gives the associated slopes and magnitudes of the ensemble spread along with those of the total forecast, the hydrologic simulation and the precipitation forecast errors. Overall, the scaling characteristics of the ensemble spreads are very similar as those of the forecast errors. The decrease with catchment area is very similar. The slope of the dependency between ensemble spread and area also increases with decreasing lead time, similar to that of the forecast error, although it is somewhat steeper for the shortest lead time. Similarly the magnitudes of the ensemble spread and the total forecast errors compare well for all lead times and catchment areas with a tendency of underestimates for shorter lead times an larger catchment areas.

Table 4.3: Mean ensemble spread $\bar{\hat{\sigma}}_\varepsilon$, total forecast errors σ_ε, hydrologic simulation errors σ_{hysim} and precipitation forecast errors σ_{pfor} (all scaled by mean catchment runoff) from the regressions against catchment area as in Figures 4.8 and 4.9 for the top class (largest 10%) of ensemble spreads. β is the slope of the regression in Eq. 4.6

Lead time (hrs)	Slope β of $\bar{\hat{\sigma}}_\varepsilon/\sigma_\varepsilon/\sigma_{hysim}/\sigma_{pfor}$	10^2 km² $\bar{\hat{\sigma}}_\varepsilon/\sigma_\varepsilon/\sigma_{hysim}/\sigma_{pfor}$	10^3 km² $\bar{\hat{\sigma}}_\varepsilon/\sigma_\varepsilon/\sigma_{hysim}/\sigma_{pfor}$	10^4 km² $\bar{\hat{\sigma}}_\varepsilon/\sigma_\varepsilon/\sigma_{hysim}/\sigma_{pfor}$
12	-0.754/-0.555/-0.522/-0.695	4.97/4.27/3.38/2.80	0.88/1.19/1.02/0.56	0.15/0.33/0.31/0.11
24	-0.626/-0.534/-0.499/-0.593	6.71/5.17/3.61/3.66	1.59/1.51/1.15/0.93	0.38/0.44/0.36/0.24
48	-0.557/-0.458/-0.471/-0.433	6.32/4.97/3.49/3.27	1.75/1.73/1.18/1.21	0.49/0.60/0.40/0.45

Figure 4.10 summarizes the spread-skill relationship for a lead time of 48 hours for all analyzed catchments with a runoff record longer than 2.5 years. Only the time steps corresponding to the top 10% of the ensemble spreads are represented. The left panel of Figure 4.10 shows the relation between ensemble spread and total forecast error, the right panel shows the relation between ensemble spread and precipitation forecast error, estimated according to Equation 4.5. The size of the circles indicates the catchment area. The thin line indicates the 1:1-line.
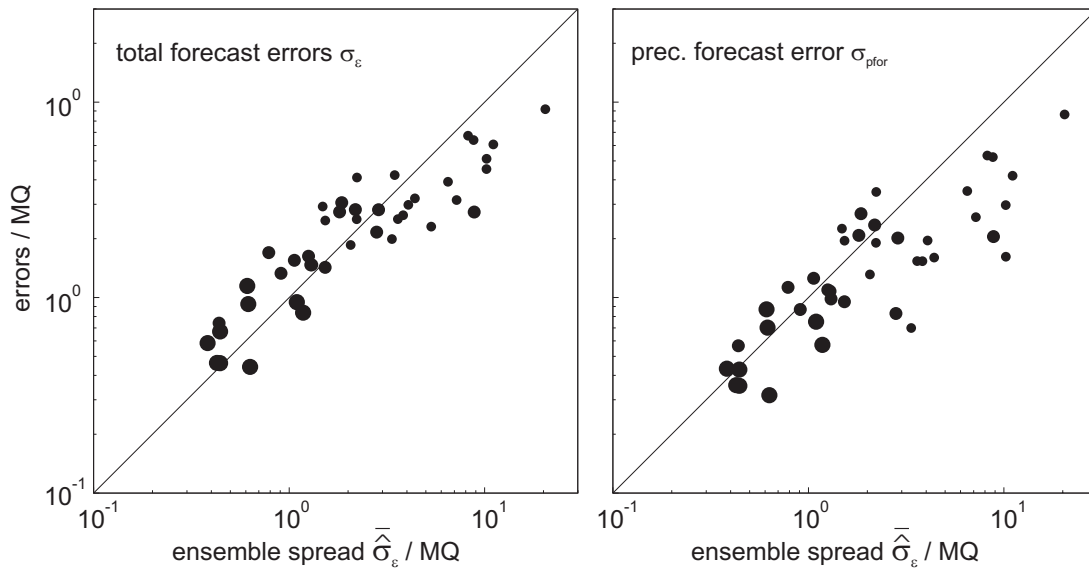
Figure 4.10: Ensemble spread $\bar{\hat{\sigma}}_\varepsilon$, scaled by mean runoff vs. total forecast errors σ_ε, scaled by mean runoff for the top 10% of the ensemble spreads and different classes of runoff (indicated by the grayscale of the circles). Right: Ensemble spread $\bar{\hat{\sigma}}_\varepsilon$, scaled by mean runoff vs. precipitation forecast errors σ_{pfor}, scaled by mean runoff for the top 10% of the ensemble spreads and different classes of runoff. Only the lead time of 48 hours is represented. The size of the circles indicates the size of the catchments.

From Figure 4.10 (left panel) it becomes apparent that on average the total forecast errors and the ensemble spreads are of similar magnitude for all catchments as the points are close to the 1:1-line. For small catchments the ensemble spread is on average larger than the total forecast error by a factor of 1.4, and for large catchments the factor is 0.9. In Figure 4.10 (right panel), the points are somewhat farther from the 1:1-line, especially for the small catchments where the ensemble spread is bigger than the precipitation forecast error by an average factor of 2.3. For the large catchments, the factor is 1.3. This indicates that on average, the ensemble spread is representing the precipitation forecast errors and the total forecast errors for the top 10% of the ensemble spreads, meaning that the model uncertainty has not to be considered for the case of large ensemble spreads.

4.5 Discussion and conclusion

In this study we perform an error analysis on the forecasts of an operational flood forecasting system for the Danube tributaries in Austria and Germany. Regarding the overall performance of the flood forecasting system, for a lead time of 24 hours we obtain mean Brier Skill Scores BSS of 0.40 and, for a lead time of 48 hours, the mean value was 0.25 when using the 90% percentile of runoff as reference forecast. These values are consistent with performances of other forecasting systems reported in literature. Addor et al. (2011) showed similar BSS values in the range of 0.30-0.50 for a lead time of one day and 0.10-0.30 for a lead time of 2 days, when evaluating the PREVAH model with COSMO-LEPS ensembles run on an hourly time step. Rousset-Regimbeau et al. (2007) and Thirel et al. (2008) evaluated ensemble runoff forecasts for 900 French catchments with areas ranging from 240 to 112000 km^2. They used ECMWF forecasts as input into a coupled land surface and hydrogeological model. Rousset-Regimbeau et al. (2007) analyzed ensemble forecasts with a lead time up to 10 days and found BSS in the range of 0.4 to 1.0 for one-day runoff forecasts and in the range of 0.3 to 1 for five-day forecasts using the 90% percentile of runoff as reference forecast. Thirel et al. (2008) focused on short range forecasts up to 2 days. They reported mean BSS of 0.90 for the first day of the forecast and 0.85 for the second day of the forecast. The larger BSS values in the studies of Rousset-Regimbeau et al. (2007) and Thirel et al. (2008) can be attributed to the larger time step of the model (1 day) which increases the model performance due to averaging effects (e.g., Skøien et al., 2003).

However, the objective of this study is not to evaluate the overall performance of the flood forecast but (1) to quantify the contributions of precipitation forecast errors and hydrologic simulation errors to the total forecast error, particularly during flood events, and (2) to evaluate the capability of the runoff ensemble forecasts to represent the total runoff forecast errors as a function of lead time.

In order to quantify the contributions of precipitation forecast errors and hydrologic simulation errors to the total forecast error we compare estimated runoff obtained by using forecasted precipitation and observed precipitation as input into the hydrologic model. To distinguish between different forecast situations, we stratify the analysis of forecast errors in classes of ensemble spreads. Analyses revealed that for small ensemble spreads, which indicate a small meteorological uncertainty and are more likely to occur when the runoff is constant or falling, the hydrologic simulation error accounts for almost 100% of the total errors. With increasing ensemble spread the uncertainty from meteorology increases and is the main source of uncertainty for large ensemble spreads, which are more likely to occur when runoff is growing and therefore in situations of flood prognosis. For the largest 10% of the ensemble spreads in the four focus catchments, the contributions of the precipitation forecast errors account for 60-85% of the error variance, whereas the hydrologic simulation errors account for 15-40% of the error variance. Olsson and Lindström (2008) found that for all phases of runoff the contributions of the meteorological and hydrologic simulation errors are similar, and only for the rising limb the uncertainty in the meteorological forecasts dominated. However, Olsson and Lindström (2008) used a daily time step which reduces the

anthropogenic variability introduced by the operation of reservoirs and lakes which are not taken into account in this study.

For short lead times, the hydrologic simulation error is the main source of uncertainty. For a lead time of 12 hours, the ratio of the hydrologic simulation error and the precipitation forecast error increases from 1.2 to 2.7 with the catchment size increasing from 100 to 10000 km^2. For long lead times, the precipitation forecast error is dominant. For lead times of 48 hours, the ratio of hydrologic simulation error to precipitation forecast error decreases from 1.1 to 0.9 with increasing catchment size. This is due to two reasons: (1) for short lead times the uncertainty in the precipitation forecasts is small as no uncertainty is attributed to the meteorological ensemble in the first two hours (Komma et al., 2007) and (2) the response time of the catchments is longer than the lead time of the forecasts, meaning that it takes the input variability from the meteorological ensemble forecasts longer to reach the basin outlet than the lead time of the forecast (Renner et al., 2009).

All errors decrease clearly with increasing catchment area and decreasing lead time. The decrease of precipitation forecast errors with catchment area can be attributed to averaging effects as discussed by Sivapalan (2003) and Skøien and Blöschl (2006). The precipitation forecast error scales with $\beta = $ -0.695 for a lead time of 12 hours, and $\beta = $ -0.433 for a lead time of 48 hours. The smaller errors for short lead times and large catchments can be attributed to the fact that runoff in large catchments does not depend much on the future precipitation at short lead times, but on the observed precipitation. The runoff in small catchments, which have shorter response times, are more dependent on the future precipitation, even at short lead times. In fact, a lead time of 12 hours can be deemed large in a small catchment, characterised by a small response time. For the 48 hour lead time, the runoff in large catchments is affected more by the future precipitation, which is reflected in the smaller scaling factor, because 48 h can be deemed as short in comparison to large response times. The hydrologic simulation error also decreases with catchment area, but the dependence on the lead time is smaller ($\beta = $ -0.522 for 12 hour lead time and $\beta = $ -0.471 for 48 hour lead time). This is because, as showed in Komma et al. (2008), updating procedures can reduce the error of forecasts in particular for short lead times and large catchments. Komma et al. (2008) analyzed rising limbs and showed errors 12% smaller compared to forecasts without updating for a lead time of 12 hours. The total forecast errors as a combination of the two components show again a strong dependence on catchment area and a moderate dependence on the forecast lead time ($\beta = $ -0.555 and -0.458 for lead times of 12 and 48 hours, respectively). Having the variability of forecast errors decreasing with catchment size is partly due to the fact that the variability of streamflow is lower in large catchments. In fact the coefficient of variation of the entire runoff hydrographs scales with -0.254 with catchment area for the 43 catchments in this study. This slope is much lower than to the one of a random field (scaling factor -0.5) because of the spatial and temporal correlation of rainfall and runoff production over the catchments (Viglione et al., 2010a,b). Similar values are found by Merz and Blöschl (2003) who focused on the evaluation of mean annual flood and further distinguished different runoff situations. They found CVs in the range of -0.205 for snowmelt induced floods and -0.413 for flash floods, which are less spatially

and temporally organized (Viglione et al., 2010b). Since the scaling factors for the forecast errors are higher (in absolute value), we can conclude that the performance of the forecasting system increases with catchment size.

The second objective of this study is to evaluate the capability of the runoff ensemble forecasts to represent the total forecast error as a function of lead time. Ensemble forecasts have been considered a suitable tool for quantifying and communicating the uncertainties of forecasts (see e.g., Hlavcova et al., 2006; Demeritt et al., 2007), as the spread of the ensemble members can be used as a measure of forecast uncertainty (Buizza, 2003). In the studies of Johnell et al. (2007), which is based on ECMWF ensemble forecasts and Jaun and Ahrens (2009), which is based on downscaled ECMWF ensemble forecasts, forecast errors increased with increasing ensemble spread and with increasing runoff for all catchments. Johnell et al. (2007) showed an increase of the mean absolute error of the ensemble median with increasing ensemble spread class from 5 to 30%, averaged over all catchments which is smaller than the mean absolute error of the deterministic forecasts in this study. For a lead time of 48 hours, we have estimated values in the order of 8 to 40% for the large catchments, and values in the order of 15 to 140% for small catchments. The larger errors can be attributed to the facts that we use an hourly time step and Johnell et al. (2007) used a daily time step for estimating runoff. Komma et al. (2007) analyzed the forecasts of 5 flood events in a 600 km^2 catchment in Austria and found mean normalized absolute errors of about 40% when evaluating the entire events at a lead time of 48 hours, which is somewhat lower than the values in this study for medium catchments. In the study of Komma et al. (2007) the number of precipitation stations per 100 km^2 is 1.28, while in this study it is on average 35 stations per 100 km^2 are available in small catchments and 0.45 stations per 100 km^2 in large catchments. An increasing number of precipitation stations per catchment as discussed by Merz et al. (2009) allows better estimates of catchment precipitation, which further reduces the forecast errors in larger catchments. Jaun and Ahrens (2009) who evaluated daily forecasts for 23 catchments in Switzerland show ensemble spreads and forecast errors of similar magnitude for large ensemble spreads. For small ensemble spreads positive forecast errors were 1.5 to 5 times larger than the (small) ensemble spread, and with increasing ensemble spread the factor between error and spread decreased to a value of 0.9 for large ensemble spreads. For negative forecast errors, the factor was larger in the range of 1.5 to 100 for small ensemble spreads and decreased to a factor of 1.3 for large ensemble spreads. Jaun and Ahrens (2009) concluded that the uncertainty is covered by the ensemble with appropriate spread. For small ensemble spreads we observe total forecast errors which are larger than the ensemble spread by a factor in the range of 10 to 100 for all lead times. For large ensemble spreads the total forecast errors and the ensemble spreads are much closer to the 1:1-line. On average, for a lead time of 48 hours the ensemble spread is larger than the total forecast error by a factor of 1.4 for small catchments, and for large catchments the factor is 0.9. This indicates that the ensemble spread is representative of the total forecast errors in situations with large ensemble spreads, which are of particular interest for flood forecasting, but there is still potential to improve the spread-skill relationship also for small ensemble spreads (see e.g., Schaake et al., 2004; Olsson and Lindström, 2008). Even if the ensemble spread does not always capture the magnitude of the forecast error, there is a clear correlation between the two, as shown by

calculating the Spearman's rank correlation, meaning that the ensemble spread can always be used as an index of forecast errors.

As for the ensemble spread the scaling factor is large in absolute values for short lead times (-0.754 for 12 hours), for the same reason, i.e., the runoff in large catchments does not depend much on the future precipitation at short lead times, but on the observed precipitation. We believe that this kind of scaling analysis of the forecast errors should be performed in other case studies as well, e.g., in other climates and using different models. Comparing different studies is needed for understanding, which is the idea underlying the so called comparative hydrology, in which simple indices are used for quantifying similarities of processes and models across scales (McDonnell and Woods, 2004; Blöschl, 2006).

5 Summary and overall conclusions

The aim of this study was to evaluate the performance of the hydrological model used for operational flood forecasting in the Danube tributaries. The model is an important part of the flood forecasting system for the Danube River. The forecasting chain consists of meteorological forecasts, hydrological forecasts and hydraulic forecasts. Meteorological forecasts comprise deterministic and ensemble forecasts of precipitation and deterministic forecasts of air temperature for a lead time of 48 hours on an hourly time step. These forecasts are used to drive the hydrological model to estimate future deterministic and ensemble runoff in the tributaries. Runoff and water level in the Danube River are estimated using the runoff estimates from the tributaries. For the development of the hydrological model, meteorological and hydrological data were used from the years 2003-2009. The model area covers 57 catchments with sizes ranging from 70 to 25600 km^2 in Austria and Bavaria. For the evaluation of the snow model, remote sensing data were used from the years 2003-2009. For the evaluation of the runoff forecasts, meteorological forecasts from 2006-2009 were used.

The aim of the first part of the study (chapter 2) was to analyse of the controls of climatic and hydrological catchment characteristics on the performance of flood simulations. A semi-distributed conceptual hydrological model was calibrated manually on runoff using observed meteorological data on an hourly time step. Manual calibration based on hydrological reasoning yields model parameters that are more suitable for the extrapolation of extreme conditions which was one of the main interests in this study. The model performance was evaluated using four different statistical measures: (1) the Nash-Sutcliffe model efficiency, (2) the volume error which was used as a measure for the bias, (3) peak discharge errors and (4) peak timing errors. The first two error measures give insight into the overall performance of the model, as the entire range of runoff from low flow conditions to floods is considered. In contrast, the latter two measures assess the ability of the model to simulate flood peaks. The analyses for each statisical measures show that the performance of the model increases with increasing catchment scale.

The Nash-Sutcliffe model efficiencies show a clear tendency to increase with catchment scale which is reflected in the Spearman's rank correlation coefficient ($r_s = 0.43$). The mean absolute peak discharge errors clearly decreases with catchment scale ($r_s = $-0.60). There are differences in the performance of the model depending on the season. In the summer, the Spearman's rank correlation coefficient between the mean peak discharge error and the catchment area is $r_s = $-0.58, whereas in winter $r_s = $-0.37. The tendency of the peak time errors to decrease with catchment scale is somewhat weaker, and there is no real trend for the volume error. The better model performance in large catchments can be attributed to

(1) averaging effects, (2) the decreasing variability in streamflow with increasing catchment scale and (3) the increasing number of precipitation stations per catchment (0.35 stations per 100 km^2 in small catchments and 0.45 stations per 100 km^2 in large catchments) which allow better estimation of catchment precipitation in the larger catchments.

The analyses further revealed that climatological catchment characteristics impact the model performance. Climatological catchment characteristics taken into account include the mean annual precipitation (MAP), mean annual runoff (MAR) and the ratio of rain in total precipitation. Similar results are obtained when comparing the model performance measures with MAP and MAR. The model performance increased with increasing catchment wetness. The model performance is somewhat better in snow dominated catchments than in rain dominated regimes. This is because catchments that follow a distinct annual hydrological cycle with snow accumulation and snow melt phases tend to be more easily simulated.

When looking at the climatological indices, it has to be kept in mind that there is a strong correlation between MAP, MAR, the ratio of rain to total precipitation and the catchment area. The model performance in terms of all performance indices tends to increase with catchment size, mean annual precipitation, and mean annual runoff and the long-term ratio of snowfall and precipitation. This is confirmed by the correlation coefficients; however, the latter are mainly due to the fact that there is correlation between catchment size and the climatological indices, indicating that the catchment size is the most important control on model performance.

In most of the basins in the study region springtime streamflow is influenced by the water stored in the snow pack during the winter period. For flood forecasting it is especially important to estimate the available water storage as accurately as possible. In the second part of the study (chapter 3) the focus of the analyses is on the snow model. The performance of the snow module is evaluated on a temporal and spatial scale and with statistical measures by comparing the simulated snow covered area (SCA) with an independent observed SCA set derived from MODIS satellite data. The accuracy of snow detection by the MODIS satellites depends on land cover, snow conditions and snow depth. The comparison of simulated and observed SCA is not straightforward for two reasons: (1) the satellite data are based on pixel values which represent the land cover (snow, no snow, no data), i.e., qualitative values and (2) the simulated snow is based on values of snow water equivalent (SWE) for elevation zones, i.e., quantitative values. To compare the two data sets, thresholds values were defined based on a sensitivity study. A threshold value of $\xi_{SCA} = 30\%$ was selected for SCA. This means that a catchment is considered as snow covered if 30% of the area is snow covered according to MODIS. A threshold value of $\xi_{SWE} = 2.5$ mm was chosen for SWE. This means that if SWE is bigger that 2.5 mm, the elevation zone is snow covered according to the simulation. The most important factor for the evaluation of the snow model is the threshold value for cloud coverage, as with increasing cloud cover the information in the MODIS data decreases. A threshold value of $\xi_C = 80\%$ was selected for cloud coverage.

The space borne snow cover data proved to be very useful for evaluating the snow model, which was found to simulate the snow well. The temporal comparison of SCA derived from

MODIS data and SCA estimated from model results indicates good agreement between observed and simulated SCA. The timing of the snow accumulation and depletion periods is simulated well. Discrepancies between model and MODIS are observed at the beginning and end of each snow season which are the times when snow conditions and depth are difficult to map by means of the satellites. A spatial evaluation of simulated SCA and MODIS SCA data on a pixel basis taking into account only cloud-free pixels shows good performance of the snow model for the winter periods from 2003-2009. 88% of the model area is correctly classified as snow covered or snow free on more than 80% of the days. The model overestimates the snow cover for ridges and steep slopes at high altitudes, where snow can be blown away or redistributed by avalanches. The model tends to underestimate the snow cover in the transition zones from lowland to alpine areas, which may be due to two reasons: there is an elevation change which is not accounted for in the model structure and the remote sensing product is underestimating SCA in forested areas.

Flood forecasts are generally associated with errors, which can be attributed to uncertainties in meteorological forecasts and hydrologic simulations. In the third part of the study (chapter 4) the focus of the analyses is on the relationship between flood forecast errors and the ensemble spread. So called ensemble forecasts consist of a set of equally likely precipitation fields to describe possible future conditions of rainfall. The spread of the ensembles is used to account for the forecast uncertainty. The comparison of forecast errors and ensemble spread shows, that the total forecast error is larger than the ensemble spread by a factor of 10 to 100 for all lead times, when the ensemble spread is small. For large ensemble spreads, the spread and the errors are of similar magnitude. On average, for a lead time of 48 hours the ensemble spread is larger than the total forecast error by a factor of 1.4 for small catchments, and for large catchments the factor is 0.9. This indicates that the ensemble spread is representative of the total forecast errors in situations with large ensemble spreads, which are of particular interest for flood forecasting. However, there remains potential to improve the spread-skill relationship also for small ensemble spreads. Even if the ensemble spread does not always capture the magnitude of the forecast error, the Spearman's rank correlation coefficient shows a clear correlation between the two, meaning that the ensemble spread can always be used as an index of forecast errors.

Using observed meteorological data as input in the hydrologic model allows separation of the contribution of each of the uncertainties in the meteorological forecasts and the hydrologic simulations to the total forecast error. Due to the extended data base, scaling properties with catchment area and lead time can be identified. For short lead times and small ensemble spreads, the contribution of the hydrologic simulation error to the total forecast error is dominating. For long lead times and large ensemble spreads, the contributions of the precipitation forecast error is more important. For the identification of scaling properties only the largest 10% of the ensemble spreads are considered as these are more likely to occur when runoff is increasing and therefore are the main interest when forecasting floods. There is a clear scaling relationship of the forecast error components with catchment area. All errors decrease clearly with increasing catchment area and decreasing lead time which can be attributed to averaging effects. For short lead times, the precipitation forcast errors in

small catchments are large, in large catchments the precipitation forecast errors are small. This can be attributed to the fact that runoff in large catchments does depend more on observed precipitation than on the future precipitation at short lead times. Runoff in small catchments (which have shorter response times) are more dependent on future precipitation, even at short lead times. The hydrologic simulation error also decreases with catchment area, but the dependence on the lead time is smaller. The total forecast error as a combination of the two error components show a strong dependence on catchment area and a moderate dependence on the forecast lead time. A similar scaling behaviour can be also observed for the ensemble spreads, which are shown to represent quantitatively the total forecast error when forecasting floods.

There will always be potential to improve flood simulations and forecasts, not only in small catchments where the accuracy of the runoff simulation is reasonable, but also in the large catchments by (1) taking into account anthropogenic effects like retention effects of lakes and daily runoff fluctuations due to hydroelectric power plants, (2) assimilating snow cover data, and (3) extending the ensemble spread for runoff forecasting. The analyses of the operational flood forecasting system for the Danube tributaries show that the hydrological model is able to reproduce runoff in different hydrological situations well. In the context of the flood framework directive of the European Union, the flood forecasting system for the Danube tributaries is an important component of the flood risk management to reduce the consequences for human health, the environment, cultural heritage and economic activity associated with floods.

A Statistical measures

Statistical measures used to evaluate the model performance include the Nash and Sutcliffe (1970) coefficient of efficiency ($nsme$):

$$nsme = 1 - \frac{\sum\limits_{i=1}^{n} \left(Q_{sim,i} - Q_{obs,i}\right)^2}{\sum\limits_{i=1}^{n} \left(Q_{obs,i} - \overline{Q_{obs}}\right)^2} \tag{A.1}$$

where $Q_{obs,i}$ and $Q_{sim,i}$ are observed and simulated runoff at hour i, respectively, and $\overline{Q_{obs}}$ is the mean observed runoff over the calibration or validation period of n hours. $nsme$ values can range from ∞ to 1. A perfect match between simulation and observation implies $nsme$ = 1; $nsme = 0$ indicates that the model predictions are as accurate as the mean of the observed data, and $nsme < 0$ occurs when the observed mean is a better predictor than the model.

As a measure of bias the volume error, VE, was used:

$$VE = \frac{\sum\limits_{i=1}^{n} Q_{sim,i} - \sum\limits_{i=1}^{n} Q_{obs,i}}{\sum\limits_{i=1}^{n} Q_{obs,i}} \tag{A.2}$$

The value can be positive or negative, with a VE of an unbiased model being 0. Values larger and smaller than 0 imply over- and underestimation, respectively.

Peak discharge errors were estimated as

$$pde = \frac{Q_{sim,peak} - Q_{obs,peak}}{Q_{obs,peak}} \tag{A.3}$$

where $Q_{obs,peak}$ and $Q_{sim,peak}$ are the observed and simulated peak discharges, respectively. Based on the peak discharge errors, the mean absolute peak discharge errors $mapde$ (%) were calculated as

$$mapde = \frac{1}{m} \cdot \sum\limits_{i=1}^{m} |pde_i| \cdot 100 \tag{A.4}$$

where m is the total number of peaks analysed for the calibration (or validation) period of the catchment.

Analogue to the peak discharge error, peak time errors were estimated as

$$pte = \frac{t_{0-peak,sim} - t_{0-peak,obs}}{t_{0-peak,obs}} \tag{A.5}$$

where $t_{0-peak,obs}$ and $t_{0-peak,sim}$ are the observed and simulated duration of the rising limb, respectively. Based on the peak time errors, the mean absolute peak time errors $mapte$ (%) were calculated as

$$mapte = \frac{1}{m} \cdot \sum_{i=1}^{m} |pte_i| \cdot 100 \tag{A.6}$$

where m is the total number of peaks analysed for the calibration (or validation) period of the catchment.

Spearman's rank correlation coefficient r_s is calculated as

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot (n^2 - 1)} \tag{A.7}$$

with

$$d_i = rk(x_i) - rk(y_i) \tag{A.8}$$

with $rk(x_i)$ as the rank of x_i, where the highest value has rank 1 and the lowest value has rank n. Spearman's r_s can vary between -1 and 1, where -1 represents a completely negative correlation and 1 represents a completely positive correlation. Completely uncorrelated pairs of data have a Spearman's r_s of 0. The partial correlation coefficient is calculated as

$$r_{xy,z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{\left(1 - r_{xz}^2\right) \cdot \left(1 - r_{yz}^2\right)}} \tag{A.9}$$

with r_{XY}, r_{XZ} and r_{YZ} as Spearman's rank correlation coefficient between variables X and Y, X and Z, and Y and Z, respectively, and $r_{XY,Z}$ as the partial correlation of X and Y adjusted for Z. For $r_{XY,Z} = 0$ and $r_{XY} \neq 0$ the correlation is highly influenced by Z, for $r_{XY,Z} = r_{XY}$ the third variable Z has no influence on the correlation of X and Y.

B Model description

A Snow routine

Snow accumulation and snow melt are represented by a simple degree day concept. The precipitation input P into each catchment is split into rain P_r and snowfall P_s based on the air temperature T_a:

$$P_r = P, \text{ if } T_a \geq T_r \tag{B.1}$$

$$P_r = P \cdot \frac{(T_a - T_s)}{(T_r - Ts)} \text{ and } P_S = P - P_R, \text{ if } T_s < T_a < T_r \tag{B.2}$$

$$P_r = 0, \text{ if } T_a < T_s \tag{B.3}$$

with T_s and T_r as the lower and upper threshold temperatures, respectively. The correction factor C_S of snow catch deficit is set to 1 as snow is corrected in the INCA system (Haiden and Pistotnik, 2009; Haiden et al., 2010). Snow melt starts at air temperatures above a threshold T_m:

$$M = (T_a - T_m) \cdot D, \text{ if } T_a > T_m \text{ and } SWE > 0 \tag{B.4}$$

$$M = 0, \text{ otherwise} \tag{B.5}$$

where M is the amount of melt water per time step, D is a melt factor and SWE is the snow water equivalent. During rain-on-snow events large melt rates are likely to occur in northern Austria (Sui and Koehler, 2001). This enhanced melting is represented in the model by increasing D by a factor of 2 if rain falls on an existing snow pack. Changes in the snow water equivalent from time step $i - 1$ to i are then:

$$SWE_i = SWE_{i-1} + (C_S \cdot P_s - M) \cdot \Delta t \tag{B.6}$$

where t is the time step of 1 hour.

B Soil moisture accounting

For the soil moisture accounting routine the sum of rain and melt, $P_r + M$, is split into a component dS that increases the 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:

$$Q_p = \left(\frac{S_S}{L_S}\right)^{\beta} \cdot (P_r + M) \tag{B.7}$$

where L_s is the maximum soil moisture storage (Bergström, 1976); 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 contribute 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 $dS > 0$:

$$dS = P_r + M - Q_p - Q_{by}, \text{ if } P_r + M - Q_p - Q_{by} > 0 \tag{B.8}$$

$$dS = 0, \text{ otherwise} \tag{B.9}$$

Additionally, bypass flow Q_by is accounted for. Analysis of the runoff data indicated that in some catchments 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 $\xi_1 = 0.4$, $\xi_2 = 0.9$) bypass flow was assumed to occur as

$$Q_{by} = \alpha_{by} \cdot (P_r + M), \text{ if } \alpha_{by} \cdot (P_r + M) < L_{by}, \text{ and } Q_{by} = L_{by}, \text{ otherwise} \tag{B.10}$$

while no bypass 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

$$S_{S,i} = S_{S,i-1} + (dS - E_A) \cdot \Delta t, \; S_{S,i} > 0 \tag{B.11}$$

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:

$$E_A = E_P \cdot \left(\frac{S_S}{L_P}\right), \text{ if } S_S < L_P \tag{B.12}$$

$$E_A = E_P, \text{ otherwise} \tag{B.13}$$

where L_P is a parameter termed the limit for potential evaporation.

C Catchment routing

Elevation zone scale routing is represented by three reservoirs. The fraction Q_p of rain and snowmelt that contributes to runoff enters the upper zone reservoir and leaves this reservoir through three paths: percolation to the lower and groundwater zones with a given percolation rate c_p, outflow from the reservoir with a fast storage coefficient of k_1, and, if a threshold L_1 of the storage state is exceeded, through an additional outlet with a very fast storage coefficient of k_0. The analysis of observed discharge data suggested that k_1 and k_2 should be related to S_S. A linear relationship was assumed:

$$k_1 = k_1 * \cdot \left(1 + \frac{\delta_1 \cdot S_S}{L_S}\right) \tag{B.14}$$

with k_1* being a storage coefficient and δ_1 a free parameter. An analogous relationship for k_2 was used. The percolation rate c_p changes with soil moisture and was related to the storage of the top soil S_S by

$$c_p = \left(\frac{S_S}{L_S}\right)^{\gamma} \cdot L_{cp} \tag{B.15}$$

with L_{cp} as the maximum percolation rate. Both L_{cp} and γ are free parameters. The percolation rate c_p is split into two components by a fraction α_p which flow into the lower zone reservoir and the groundwater reservoir. Bypass flow Q_{by} is also split into two components by a fraction α_p and is added to the lower zone reservoir and the groundwater reservoir. Q_0 represents the fast surface or near surface runoff. Q_1 is a somewhat slower component representing interflow. Q_2 is the contribution to total runoff from the lower zone. Q_3 is the slowest component representing groundwater flow. For those catchments where part of the discharge is in the deep subsurface and not captured by the stream gauge, the slowest groundwater component is reduced by a factor $f_3 < 1$ to account for deep percolation. Total runoff from an elevation zone then consists of the following components:

$$Q_t = Q_0 + Q_1 + Q_2 + Q_3 \cdot f_3 \tag{B.16}$$

For every elevation zone there are a total of 21 parameters. The snow model parameters are T_s, T_r, T_m, D, C_S; the soil moisture accounting parameters are L_S, β, α_{by}, L_{by}, L_P; and the elevation zone scale routing parameters are k_0, k_1, δ_1, k_2, δ_2, k_3, L_1, L_{cp}, γ, α_p, f_3.

The outflow from the reservoirs, Q_t, representing a single elevation zone of a catchment is convoluted by a transfer function which represents the runoff routing in the streams within a catchment. As a transfer function, a linear storage cascade with the parameters N (number of reservoirs) and K (time parameter of each reservoir) is used. The convolution is performed in the state space notation in a similar way as stream routing. The sum of this convoluted runoff over each direct catchment is used as the lateral inflow to the stream routing model

of each river reach. There are two within-catchment routing parameters for each catchment, N, K.

D Stream routing

A linear storage cascade in the state space notation of Szolgay (2004) is used here. If one assumes that the input vector $\boldsymbol{U}$ to each reservoir is constant within a time interval $(i, i-1)$ of duration Δt,

$$\boldsymbol{S}_i = \boldsymbol{F}_{i,i-1} \cdot \boldsymbol{S}_{i-1} + \boldsymbol{G}_{i,i-1} \cdot \boldsymbol{U}_{i,i-1} \tag{B.17}$$

$$\boldsymbol{Q}_i = \boldsymbol{H}_i \cdot \boldsymbol{S}_i \tag{B.18}$$

where $\boldsymbol{S}$ and $\boldsymbol{Q}$ are the $(n_r \cdot 1)$ state vectors of reservoir storages and outflow with n_r being the number of reservoirs. The number of reservoirs n_r is constant for every reach, as is the time parameter k_r. $\boldsymbol{H}$ is an $(n_r \cdot n_r)$ matrix that contains the inverse of the time parameter k_r in the diagonal

$$\boldsymbol{H} = \begin{bmatrix} 1/k_r & 0 & ... & 0 \\ 0 & 1/k_r & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1/k_r \end{bmatrix} \cdot \boldsymbol{I} \tag{B.19}$$

where $\boldsymbol{I}$ is the identity matrix. The transition matrices $\boldsymbol{F}$ and $\boldsymbol{G}$ (dimension $n_r \cdot n_r$) are defined as:

$$\boldsymbol{F}(\iota, \zeta) = e^{-\Delta t/k_r} \cdot \frac{\Delta t^{\iota-\zeta}}{(\iota - \zeta)! k_r^{\iota-\zeta}} \tag{B.20}$$

$$\boldsymbol{G}(\iota, \zeta) = k_r - e^{-\Delta t/k_r} \cdot \sum_{v=0}^{\iota-\zeta} \frac{\Delta t^v}{v! k_r^{v-1}} \tag{B.21}$$

for ι greater than or equal to ζ, and $\boldsymbol{F} = 0$, $\boldsymbol{G} = 0$ for ι less than ζ, where ι and ζ relate to the rows and columns of the matrices, respectively. The duration Δt of the time interval is 1 hour. Inflow $\boldsymbol{U}^{(1)}_{i,i-1}$ to each reach is the outflow from the upstream reach. Lateral inflow from the direct catchments is added to the downstream node.

C Catchments

This appendix gives an overview of the gauges used in the study. Each table includes the gauge ID of the Hydrographical Service, the names of the gauges, the rivers and the countries and the area of the catchments. Further information includes are mean annual precipitation (mm) as estimated from the available precipitation data, mean annual runoff (mm) as estimated from the available runoff data, the largest runoff (m^3/s) observed in the period 2002-2009, the mean observed runoff (m^3/s) in the period 2002-2009 and the Nash-Sutcliffe model effiencies ($nsme$) for the calibration and validation periods.

The areas of the catchments are ranging from 69 to 25600 km^2, with a median catchment size of 444 km^2 and a medium catchment area of 1960 km^2. Table C.1 includes the catchments with areas below the median catchment area, table C.2 the catchments with areas between the median and the medium catchment area, and table C.3 the catchments with areas larger than the medium catchment area.

Table C.1: Small catchments in the study region. Table includes ID, name of the gauge and name of the river, catchment area (km²), mean annual precipitation (MAP) (mm) and runoff (MAR) (mm), the highest observed runoff in the data record 2002-2009 $Q_{obs,max}$ (m³/s), the mean observed catchment runoff MQ_{obs} (m³/s) and the Nash–Sutcliffe model efficiencies ($nsme$) for the calibration and validation period (2003-2006, 2007-2009)

ID	Gauge	River	Country	Area	MAP	MAR	$Q_{obs,max}$	MQ_{obs}	$nsme_{calib}$	$nsme_{valid}$
209569	Sieghartskirchen	Kleine Tulln	Lower Austria	69	797	142	33	0.3	0.587	0.704
209247	St. Pantaleon	Erlabach	Lower Austria	92	946	340	16	0.9	0.194	NO DATA
205898	Molln	Krumme Steyrling	Upper Austria	129	1705	865	108	3.5	0.581	0.672
208850	Erlaufboden	Große Erlauf	Lower Austria	136	1587	1187	113	5.3	0.732	0.657
204875	Oberkappel	Ranna	Upper Austria	139	1055	622	52	2.7	0.693	0.568
207696	Krenstetten	Url	Lower Austria	156	1053	429	145	2.1	0.606	0.719
205641	Kremsmünster	Krems	Upper Austria	161	1263	537	77	2.7	0.663	0.716
204750	Haging	Antiesen	Upper Austria	165	1031	443	97	2.2	0.701	0.696
207613	Isperdorf	Isper	Lower Austria	165	906	453	39	2.3	0.681	0.445
207795	Wieselburg	Kleine Erlauf	Lower Austria	166	1287	536	171	2.8	0.552	0.722
208090	Cholerakapelle	Schwechat	Lower Austria	181	891	257	115	1.4	0.787	0.504
204883	Obermühl	Kleine Mühl	Upper Austria	200	948	476	85	3.0	0.644	0.521
208017	Siegersdorf	Große Tulln	Lower Austria	204	814	197	125	1.2	0.457	0.640
205088	Rottenegg	Große Rodl	Upper Austria	227	962	385	90	2.7	0.617	0.357
205716	St. Georgen	Gusen	Upper Austria	263	861	250	52	2.0	0.703	0.683
208009	Atzenbrugg	Perschling	Lower Austria	268	822	196	107	1.6	0.619	0.717
207837	Matzleinsdorf	Melk	Lower Austria	285	1103	353	229	3.1	0.349	0.480
208124	Hirtenberg	Triesting	Lower Austria	287	956	230	127	2.3	0.585	0.626
207852	Hofstetten	Pielach	Lower Austria	290	1424	687	275	6.6	0.813	0.623
207878	Imbach	Krems	Lower Austria	306	722	210	128	2.6	0.333	0.424
206029	Haid	Naarn	Upper Austria	306	914	377	65	3.7	0.754	0.650
207894	Lilienfeld	Traisen	Lower Austria	333	1441	858	227	9.3	0.849	0.694
204867	Pramerdorf	Pram	Upper Austria	341	950	330	176	4.8	0.799	0.349
205070	Pfaffing	Aschach	Upper Austria	353	976	334	144	3.7	0.549	0.491
205039	Fraham	Innbach	Upper Austria	362	942	340	85	3.9	0.619	0.600
203307	Obergäu	Lammer	Salzburg	395	1856	1215	270	16.9	0.593	0.692

Table C.2: Medium catchments in the study region. Table includes ID, name of the gauge and name of the river, catchment area (km^2), mean annual precipitation (MAP) and runoff (MAR) (mm), the highest observed runoff in the data record 2002-2009 $Q_{obs,max}$ (m^3/s), the mean observed catchment runoff MQ_{obs} (m^3/s) and the Nash-Sutcliffe model efficiencies ($nsme$) for the calibration and validation period (2003-2006, 2007-2009)

ID	Gauge	River	Country	Area	MAP	MAR	$Q_{obs,max}$	MQ_{obs}	$nsme_{calib}$	$nsme_{valid}$
204677	Jahrsdorf	Mattig	Upper Austria	444	1221	324	23	8.2	0.621	0.689
205419	Vöcklabruck	Vöckla	Upper Austria	444	1209	512	151	8.7	0.587	0.562
204933	Teufelmühle	Große Mühl	Upper Austria	450	1054	576	114	8.7	0.722	0.486
205518	Penningersteg	Alm	Upper Austria	459	1601	964	277	14.6	0.462	0.620
206276	Dürnau	Ager	Upper Austria	502	1567	900	70	15.6	0.737	0.637
207654	Opponitz	Ybbs	Lower Austria	507	1802	1140	614	18.2	0.838	0.794
205864	Klaus	Steyr	Upper Austria	542	1749	1527	330	28.9	0.588	0.585
210898	Wildalpen	Salza	Styria	592	1630	1146	310	21.5	0.740	0.733
207803	Niederndorf	Erlauf	Lower Austria	595	1459	752	440	18.0	0.793	0.730
205989	Schwertberg	Aist	Upper Austria	605	885	305	96	5.8	0.692	0.651
207910	Windpassing	Traisen	Lower Austria	735	1268	637	426	15.8	0.799	0.747
18808004	Ruhstorf	Rott	Bavaria	1052	841	221	159	8.0	0.579	0.068
18209000	Rosenheim	Mangfall	Bavaria	1099	1520	429	350	27.0	0.710	0.357
207688	Greimpersdorf	Ybbs	Lower Austria	1116	1478	837	898	29.3	0.865	0.795
203570	Siezenheim	Saalach	Salzburg	1139	1727	910	569	35.4	0.711	0.736
205450	Fischerau	Ager	Upper Austria	1260	1357	720	361	32.7	0.774	0.425
208157	Schwechat	Schwechat	Lower Austria	1370	701	255	281	8.0	0.725	0.914
206128	Gmunden–Theresienthal	Traun	Upper Austria	1390	1811	1424	545	66.3	0.763	0.783

Table C.3: Large catchments in the study region. Table includes ID, name of the gauge and name of the river, catchment area (km^2), mean annual precipitation (MAP) and runoff (MAR) (mm), the highest observed runoff in the data record 2002-2009 $Q_{obs,max}$ (m^3/s), the mean observed catchment runoff MQ_{obs} (m^3/s) and the Nash-Sutcliffe model efficiencies ($nsme$) for the calibration and validation period (2003-2006, 2007-2009)

ID	Gauge	River	Country	Area	MAP	MAR	$Q_{obs,max}$	MQ_{obs}	$nsme_{calib}$	$nsme_{valid}$
18408200	Burgkirchen	Alz (inkl. Alzkanal)	Bavaria	2222	1555	917	317	68.0	0.519	0.556
210849	Gstatterboden	Enns	Styria	2777	1393	939	470	67.1	0.769	NO DATA
203323	Golling	Salzach	Salzburg	3161	1634	1098	737	133.1	0.628	0.729
206391	Wels-Lichtenegg	Wels	Upper Austria	3426	1601	854	1241	118.1	0.824	0.846
206342	KW Schönau	Enns	Styria	4341	1474	933	1191	141.4	0.786	0.657
201525	Innsbruck	Inn	Tyrol	5792	1113	825	1489	152.6	0.711	0.745
205922	Steyr-Ortskai	Enns	Upper Austria	5915	1506	1058	2196	198.3	0.801	0.706
203539	Oberndorf	Salzach	Salzburg	6120	1651	1022	1665	236.0	0.655	0.738
201806	Brixlegg	Inn	Bavaria	8503	1191	903	1934	245.1	0.734	0.756
18001508	Rosenheim	Inn	Bavaria	10186	1285	816	1160	314.0	0.681	0.757
18003004	Wasserburg	Inn	Bavaria	11977	1270	801	2600	358.0	0.647	0.731
204776	Schärding	Inn	Upper Austria	25664	1342	832	4380	683.0	0.766	0.797

References

Addor, N., Jaun, S., Fundel, F., and Zappa, M. (2011). An operational hydrological ensemble prediction system for the city of zurich (switzerland): Skill, case studies and scenarios. *Hydrol. Earth Syst. Sci.*, 15:2327–2347. doi:10.5194/hess-15-2327-2011.

Andreadis, K. M. and Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Adv. Water Res.*, 29(6):872–886. doi:10.1016/j.advwatres.2005.08.004.

Bergström, S. (1976). Development and application of a conceputual runoff model for Scandinavian catchments. *Dept. Water Resou. Engng, Lund Inst. Techol./Univ. Lund, Bull. Ser. A52, 134.*

Blöschl, G. (2006). Hydrologic synthesis: across processes, places and scales. *Water Resour. Res.*, 42(3):3. doi:10.1029/2005WR004319.

Blöschl, G. (2008). Flood warning - on the value of local information. *Intl. J. River Basin Management*, 6(1):41–50. doi:10.1080/15715124.2008.9635336.

Blöschl, G. and Kirnbauer, R. (1991). Point snowmelt models with different degrees of complexity - Internal processes. *J. Hydrol.*, 129(1-4):127–147. doi:10.1016/0022-1694(91)90048-M.

Blöschl, G. and Kirnbauer, R. (1992). An analysis of snow cover patterns in a small alpine catchment. *Hydrol. Process.*, 6(1):99–109. doi:10.1002/hyp.3360060109.

Blöschl, G., Reszler, C., and Komma, J. (2008). A spatially distributed flash flood forecasting model. *Environ. Modell. Softw.*, 23(4):464 – 478. doi:10.1016/j.envsoft.2007.06.010.

Blöschl, G. and Zehe, E. (2005). On hydrological predictability. *Invited commentary. Hydrol. Process.*, 19(19):3923–3929. doi:10.1002/hyp.6075.

Blöschl, G. (2002). Ingenieurhydrologie, Vorlesungsunterlagen, Institut für Hydraulik, Gewässerkunde und Wasserwirtschaft, TU Wien.

BMLFUW (2009). Das Hochwasser in Österreich vom 22. bis 30. Juni 2009 - Beschreibung der Hydrologischen Situation. Technical report. Accessed September 29, 2010. `http://wasser.lebensministerium.at/filemanager/download/51869/`.

Braun, L. N. (1985). *Simulation of Snowmelt-Runoff in Lowland and Lower Alpine Regions in Switzerland.* PhD thesis, ETH Zürich.

Braun, L. N. and Renner, C. B. (1992). Application of a conceptual runoff model in different physiographic regions of Switzerland/Application d'un modèle conceptual d'écoulement à différentes régions physiographiques de la suisse. *Hydrolog. Sci. J.*, 37(3):217–231. doi:10.1080/02626669209492583.

Brier, G. W. (1950). Verification of Forecasts Expressed in Terms of Probability. *Mon. Weather Rev.*, 78(1):1–3. doi:10.1175/1520-0493(1950)078¡0001:VOFEIT¿2.0.CO;2.

Buizza, R. (2003). *Encyclopaedia of Atmospheric Sciences*, chapter Weather Prediction: Ensemble Prediction, pages 2546–2557. Academic Press, London.

Buizza, R., Houtekamer, P., Toth, Z., Pellerin, G., Wei, M., and Zhu, Y. (2005). A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Weather Review*, 133(5):1076–1097.

Cloke, H. L. and Pappenberger, F. (2009). Ensemble Flood Forecasting: A Review. *J. Hydrol.*, 375:613–626. doi:10.1016/j.jhydrol.2009.06.005.

Das, T., Bárdossy, A., Zehe, E., and He, Y. (2008). Comparison of conceptual model performance using different representations of spatial variability. *J. Hydrol.*, 356(1-2):106–118. doi:10.1016/j.jhydrol.2008.04.008.

Demeritt, D., Cloke, H., Pappenberger, F., Thielen, J., Bartholmes, J., and Ramos, M.-H. (2007). Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting. *Environmental Hazards*, 7:115–127. doi:10.1016/j.envhaz.2007.05.001.

Déry, S. J., Salomonson, V. V., Stieglitz, M., Hall, D. K., and Appel, I. (2005). An approach to using snow areal depletion curves inferred from MODIS and its application to land-surface modelling in Alaska. *Hydrol. Process.*, 19(14):2755–2774. doi:10.1002/hyp.5784.

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

European Environmental Agency (2000). CORINE Land Cover 2000. `http://www.eea.europa.int`.

European Union (2007). Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks. *Official Journal of 6 November 2007, L288/page27*. Available online at `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:288:0027:0034:EN:PDF`, accessed 10 November 2011.

Ewen, J., O'Donnell, G., Burton, A., and O'Connell, E. (2006). Errors and uncertainty in physically-based rainfall-runoff modelling of catchment change effects. *J. Hydrol.*, 330:641–650. doi:10/1016/j.jhydrol.2006.04.024.

Franchini, M. and Pacciani, M. (1991). Comparative analysis of several conceptual rainfall-runoff models. *J. Hydrol.*, 122(1-4):161–219. doi:10.1016/0022-1694(91)90178-K.

Gafurov, A. and Bárdossy, A. (2009). Cloud removal methodology from MODIS snow cover product. *Hydrol. Earth Syst. Sci.*, 13(7):1361–1373.

Garen, D. C. and Marks, D. (2005). Spatially distributed energy balance snowmelt modelling in a mountainous river basin: estimation of meteorological inputs and verification of model results. *J. Hydrol.*, 315(1-4):126–153. doi:10.1016/j.jhydrol.2005.03.026.

Goodrich, D. C., Lane, L. J., Shillito, R. M., Miller, S. N., Syed, K. H., and Woolhiser, D. A. (1997). Linearity of basin response as a function of scale in a semiarid watershed. *Water Resour. Res.*, 33(12):2951–2965. doi:10.1029/97WR01422.

Gouweleeuw, B. T., Thielen, J., Franchello, G., De Roo, A. P. J., and Buizza, R. (2005). Flood Forecasting Using Medium-Range Probabilistic Weather Prediction. *Hydrol. Earth Syst. Sci.*, 9(4):365–380.

Grayson, R., Blöschl, G., Western, A., and McMahon, T. (2002). Advances in the use of observed spatial patterns of catchment hydrological response. *Adv. Water Res.*, 25(8-12):1313–1334. doi:10.1016/S0309-1708(02)00060-X.

Grimit, E. P. and Mass, C. F. (2007). Measuring the Ensemble Spread-Error Relationship with a Probabilistic Approach: Stochastic Ensemble Results. *Mon. Weather Rev.*, 135:203–221. DOI: 10.1175/MWR3262.1.

Gupta, H. V., Sorooshian, S., and Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrol. Eng.*, 4(2):135–143. doi:10.1061/(ASCE)1084-0699(1999)4:2(135).

Gutknecht, D. (1988). Hochwasser in Österreichs Alpen- und Donauraum: Prognosemöglichkeiten und -modelle. *ÖZE*, 41(8):217–229.

Haiden, T. (2009). Meteorologische Analyse des Niederschlags von 22.-25. Juni 2009. `http://www.zamg.ac.at/docs/aktuell/2009-06-30_Meteorologische%20Analyse%20HOWA2009.pdf` (accessed September 29, 2010).

Haiden, T., Kann, A., Stadlbacher, K., Steinheimer, M., and Wittmann, C. (2010). Integrated nowcasting through comprehensive analysis (INCA) - system overview. Technical report, ZAMG. `http://www.zamg.ac.at/fix/INCA_system.doc` (accessed February 22, 2010).

Haiden, T., Kann, A., Wittmann, C., Pistotnik, G., Bica, B., and Gruber, C. (2011). The Integrated Nowcasting through Comprehensive Analysis (INCA) System and its Validation over the Eastern Alpine Region. *Weather Forecast.*, 26:166–183. doi:10.1175/2010WAF2222451.1.

Haiden, T. and Pistotnik, G. (2009). Intensity-dependent parameterization of elevation effects in precipitation analysis. *Adv. Geosci.*, 20:33–38. 2009 www.adv-geosci.net/20/33/2009/.

Hall, D. K. and Riggs, G. A. (2007). Accuracy assessment of the MODIS snow products. *Hydrol. Process.*, 21:1534–1547. doi:10.1002/hyp.6715.

Hall, D. K., Riggs, G. A., Foster, J. L., and Kumar, S. V. (2010). Development and evaluation of a cloud-gap-filled MODIS daily snow-cover product. *Remote Sens. Environ.*, 114:496–503. doi:10.1016/j.rse.2009.10.007.

Hall, D. K., Riggs, G. A., and Salomonson, V. V. (2006). MODIS/Terra Snow Cover Daily L3 Global 500m Grid V005, January 2003 to December 2009. Digital media, updated daily.

Hall, D. K., Riggs, G. A., and Salomonson, V. V. (2007). MODIS/Aqua Snow Cover Daily L3 Global 500m Grid V005, January 2003 to December 2009. Digital media, updated daily.

Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E., and Bayr, K. J. (2002). MODIS snow-cover products. *Remote Sens. Environ.*, 83:181–194. doi:10.1016/S0034-4257(02)00095-0.

Hamill, T., Hagedorn, R., and Whitaker, J. (2008). Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part II: Precipitation. *Mon. Weather Review*, 136:2620–2632. doi:10.1175/2007MWR2411.1.

Hellebrand, H. and van den Bos, R. (2008). Investigating the use of spatial discretization of hydrological processes in conceptual rainfall runoff modelling: a case study for the meso-scale. *Hydrol. Process*, 22(16):2943–2952. doi:10.1002/hyp.6909.

Hlavcova, K., Szolgay, J., Kubes, R., Kohnova, S., and Zvolensky, M. (2006). *Routing of numerical weather predictions through a rainfall-runoff model*, pages 79–90. NATO Science Series: IV: Earth and Environmental Sciences. Springer.

Hopson, T. M. and Webster, P. J. (2010). A 1-10-day ensemble forecasting scheme for the major river basins of Bangladesh: Forecasting severe floods of 2003-07. *J. Hydrometeorol.*, 11(3):618–641.

Immerzeel, W. W., Droogers, P., de Jong, S. M., and Bierkens, M. F. P. (2009). Large-scale monitoring of snow cover and runoff simulation in Himalayan river basins using remote sensing. *Remote Sens. Environ.*, 113(1):40–49. doi:10.1016/j.rse.2008.08.010.

Ivanov, V. Y., Vivoni, E. R., Bras, R. L., and Entekhabi, D. (2004). Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: a fully-distributed physically-based approach. *J. Hydrol.*, 298(1-4):80–111. doi:10.1016/j.jhydrol.2004.03.041.

Jakeman, A. J. and Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resour. Res.*, 29(8):2637–2649. doi:10.1029/93WR00877.

Jaun, S. and Ahrens, B. (2009). Evaluation of a probabilistic hydrometeorological forecast system. *Hydrol. Earth Syst. Sci.*, 13(7):1031–1043.

Johnell, A., Lindström, G., and Olsson, J. (2007). Deterministic evaluation of ensemble streamflow predictions in sweden. *Nord. Hydrol.*, 38(4):441–450. doi:10.2166/nh.2007.022.

Kienzle, S. W. (2008). A new temperature based method to separate rain and snow. *Hydrol. Process.*, 22(26):2067–2085. doi:10.1002/hyp.7131.

Kirnbauer, R. and Schönlaub, H. (2006). Vorhersage für den Inn. In *Wiener Mitteilungen 199 - Hochwasserprognose. Erfahrungen, Entwicklungen & Realität*.

Klein, A. G. and Barnett, A. C. (2003). Validation of daily MODIS snow cover maps of the Upper Rio Grande River Basin for the 2000-2001 snow year. *Remote Sens. Environ.*, 86(2):162–176. doi:10.1016/S0034-4257(03)00097-X.

Koboltschnig, G. R., Schöner, W., Zappa, M., Kroisleitner, C., and Holzmann, H. (2008). Runoff modelling of the glacierized Alpine Upper Salzach basin (Austria): multi-criteria result validation. *Hydrol. Process.*, 22(19):3950–3964. doi:10.1002/hyp.7112.

Komma, J., Blöschl, G., and Reszler, C. (2008). Soil Moisture Updating by Ensemble Kalman Filtering in Real-Time Flood Forecasting. *J. Hydrol.*, 357(3-4):228 – 242. doi:10.1016/j.jhydrol.2008.05.020.

Komma, J., Drabek, U., and Blöschl, G. (2009). Aktuelle Methoden der Hochwasservorhersage. *Wiener Mitteilungen*, 216:181–212.

Komma, J., Reszler, C., Blöschl, G., and Haiden, T. (2007). Ensemble prediction of floods - catchment non-linearity and forecast probabilities. *Nat. Hazards Earth Sys. Sci.*, 7:431–444. doi:10.5194/nhess-7-431-2007.

Krzysztofowicz, R. (2001). Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model. *J. Hydrol.*, 249:69–85. doi:10.1016/S0022-1694(01)00413-9.

Krzysztofowicz, R. and Kelly, K. (2000). Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.*, 36(11):3265–3277. doi:10.1029/2000WR900108.

Lalaurette, F., Bidlot, J., Ferranti, L., Ghelli, A., Grazzini, F., Leutbecher, M., Paulsen, J.-E., and Viterbo, P. (2005). Verification statistics and evaluations of ECMWF forecasts in 2003-2004. Technical report, ECMWF, Shinfield Park Reading, Berks RG2 9AX. online available at: `http://www.ecmwf.int/publications/library/ecpublications/pdf/tm/401-500/tm463.pdf`.

Lidén, R. and Harlin, J. (2000). Analysis of conceptual rainfall–runoff modelling performance in different climates. *J. Hydrol.*, 238(3-4):231–247. doi:10.1016/S0022-1694(00)00330-9.

Madsen, H., Wilson, G., and Ammentorp, H. C. (2002). Comparison of different automated strategies for calibration of rainfall-runoff models. *J. Hydrol.*, 261(1-4):48–59. doi:10.1016/S0022-1694(01)00619-9.

McDonnell, J. and Woods, R. (2004). On the need for catchment classification. *J. Hydrol.*, 299(1-2):2–3. doi:10.1016/j.jhydrol.2004.09.003.

Merz, R. and Blöschl, G. (2003). A process typology of regional floods. *Water Resour. Res.*, 39(12):SWC51–SWC520.

Merz, R. and Blöschl, G. (2004). Regionalisation of catchment model parameters. *J. Hydrol.*, 287(1-4):95–123. doi:10.1016/j.jhydrol.2003.09.028.

Merz, R. and Blöschl, G. (2008). Process controls on the statistical flood moments - a data based analysis. *Hydrol. Process.*, 23(5):675–696. doi:10.1002/hyp.7168.

Merz, R., Parajka, J., and Blöschl, G. (2009). Scale effects in conceptual hydrological modeling. *Water Resour. Res.*, 45(9). doi:10.1029/2009WR007872.

Micovic, Z. and Quick, M. C. (2009). Investigation of the model complexity required in runoff simulation at different time scales. *Hydrol. Sci. J.*, 54(5):872–885. DOI: 10.1623/hysj.54.5.872.

Mimikou, M. A., Hatjisava, P. S., Kouvopoulos, Y. S., and Anagnostou, E. N. (1992). The Influence of Basin Aridity on the Efficiency of Runoff Predicting Models. *Nord. Hydrol.*, 23(2):105–120.

Modarres, R. (2009). Multi-criteria validation of artificial neural network rainfall-runoff modeling. *Hydrol. Earth Syst. Sci.*, 13:411–421.

MODIS (2010). web page information. Technical report. `http://nsidc.org/data/docs/daac/modis_v5/mod10_l2_modis_terra_snow_cover_5min_swath.gd.html`.

Montanari, A. (2007). What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology. *Hydrol. Process.*, 21:841–845. doi:10.1002/hyp.6623.

Montanari, A. and Brath, A. (2004). A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resour. Res.*, 40. doi:10.1029/2003WR002540.

Montanari, A. and Grossi, G. (2008). Estimating the uncertainty of hydrological forecasts: A statistical approch. *Water Resour. Res.*, 44:W00B08. doi:10.1029/2008WR006897.

Montanari, A., Shoemaker, C., and Van De Giesen, N. (2009). Introduction to special section on Uncertainty Assessment in Surface and Subsurface Hydrology: An overview of issues and challenges. *Water Resour. Res.*, 45. doi:10.1029/2009WR008471.

Moore, R., Bell, V., and Jones, D. (2005). Forecasting for flood warning. *C.R. Geoscience*, 337:203–217. doi:10.1016/j.crte.2004.10.017.

Nash, J. E. and Sutcliffe, J. V. (1970). River Flow Forecasting through Conceptual Models - Part I: A Discussion of Principles. *J. Hydrol.*, 10(3):282–290. doi:10.1016/0022-1694(70)90255-6.

Nester, T., Kirnbauer, R., Gutknecht, D., and Blöschl, G. (2011a). Climate and catchment controls on the performance of regional flood simulations. *J. Hydrol.*, 402(3-4):340–356. doi:10.1016/j.jhydrol.2011.03.028.

Nester, T., Kirnbauer, R., Parajka, J., and Blöschl, G. (2011b). Evaluating the snow component of a flood forecasting model. *Hydrol. Res.* in press.

Ogden, F. L. and Dawdy, D. R. (2003). Peak discharge scaling in small Hortonian watershed. *J. Hydrol. Eng.*, 8(2):64–73. doi:10.1061/(ASCE)1084-0699(2003)8:2(64).

Olsson, J. and Lindström, G. (2008). Evaluation and calibration of operational hydrological ensemble forecasts in Sweden. *J. Hydrol.*, 350:14–24. doi:10.1016/j.jhydrol.2007.11.010.

Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N. (2008). Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 french catchments. *Water Resour. Res.*, 44(3). doi:10.1029/2007WR006240.

Pappenberger, F., Beven, K. J., Hunter, N. M., Bates, P. D., Gouweleeuw, B. T., Thielen, J., and de Roo, A. P. J. (2005). Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS). *Hydrol. Earth Syst. Sci.*, 9(4):381–393. doi:10.5194/hess-9-381-2005.

Parajka, J. and Blöschl, G. (2006). Validation of MODIS snow cover images over Austria. *Hydrol. Earth Syst. Sci.*, 10(5):679–689.

Parajka, J. and Blöschl, G. (2008a). Spatio-temporal combination of MODIS images - potential for snow cover mapping. *Water Resour. Res.*, 44(3):W03406. doi:10.1029/2007WR006204.

Parajka, J. and Blöschl, G. (2008b). The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models. *J. Hydrol.*, 358(3-4):240–258. doi:10.1016/j.jhydrol.2008.06.006.

Parajka, J. and Blöschl, G. (2012). *MODIS-Based Snow Cover Products, Validation, and Potential Hydrologic Applications*, chapter 9. CRC Press. in press, ISBN 9781439877456.

Parajka, J., Merz, R., and Blöschl, G. (2003). Estimation of daily potential evapotranspiration for regional water balance modeling in Austria. In *In: 11th International Poster Day and Institute of Hydrology Open Day "Transport of Water, Chemicals and Energy in the Soil - Crop Canopy - Atmosphere System", Slovak Academy of Sciences, Bratislava, 299-306*.

Parajka, J., Merz, R., and Blöschl, G. (2007). Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments. *Hydrol. Process.*, 21(4):435–446. doi:10.1002/hyp.6253.

Parajka, J., Pepe, M., Rampini, A., Rossi, S., and Blöschl, G. (2010). A regional snow-line method for estimating snow cover from MODIS during cloud cover. *J. Hydrol.*, 381(3-4):203–212. doi:10.1016/j.jhydrol.2009.11.042.

Pu, Z., Xu, L., and Salomonson, V. V. (2007). MODIS/Terra observed seasonal variations of snow cover over the Tibetan Plateau. *Geophys. Res. Lett.*, 34(6):L06706. doi:10.1029/2007GL029262.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., and Seo, D.-J. (2004). Overall distributed model intercomparison project results. *J. Hydrol.*, 298(1-4):27–60. doi:10.1016/j.jhydrol.2004.03.031.

Reed, S., Schaake, J., and Zhang, Z. (2007). A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. *J. Hydrol.*, 337(3-4):402–420. doi:10.1016/j.jhydrol.2007.02.015.

Renner, M., Werner, M. G. F., Rademacher, S., and Sprokkereef, E. (2009). Verification of ensemble flow forecasts for the River Rhine. *J. Hydrol.*, 376:463–475. doi:10.1016/j.jhydrol.2009.07.059.

Riggs, G. A., Hall, D. K., and Salomonson, V. V. (2006). MODIS Snow Products. User Guide to Collection. Technical report. `http://modis-snow-ice.gsfc.nasa.gov/sug_c5.pdf`.

Robinson, J. S., Sivapalan, M., and Snell, J. D. (1995). On the relative roles of hillslope processes, channel routing, and network geomorphology in the hydrologic response of natural catchments. *Water Resour. Res.*, 31(12):3089–3101. doi:10.1029/95WR01948.

Rodell, M. and Houser, P. R. (2004). Updating a land surface model with MODIS-derived snow cover. *J. Hydrometeorol.*, 5(6):1064–1075.

Roulin, E. (2007). Skill and relative economic value of medium-range hydrological ensemble predictions. *Hydrol. Earth Syst. Sci.*, 11:725–737. doi:10.5194/hess-11-725-2007.

Roulin, E. and Vannitsem, S. (2005). Skill of Medium-Range Hydrological Ensemble Predictions. *J. Hydrometeorol.*, 6(5):729–744.

Rousset-Regimbeau, F., Habets, F., Martin, E., and Noilhan, J. (2007). Ensemble runoff forecasts over France. *ECMWF Newsletter*, 111. `http://www.ecmwf.int/publications/newsletters/pdf/111.pdf`, accessed September 29, 2011.

Roy, A., Royer, A., and Turcotte, R. (2010). Improvement of springtime streamflow simulations in a boreal environment by incorporating snow-covered area derived from remote sensing data. *J. Hydrol.*, 390(1-2):35–44. doi:10.5194/adgeo-27-99-2010.

Schaake, J., Perica, S., Mullusky, M., Demargne, D., Welles, E., and Wu, L. (2004). Pre-processing of Atmospheric Forcing for Ensemble Runoff Prediction. In *Proceedings of the the 84th AMS Annual Meeting Held in Seattle*. `http://ams.confex.com/ams/pdfpapers/72172.pdf`, accessed 29 September 2011.

Schatzl, R. and Ruch, C. (2006). Internationales Hochwasserprognosemodell Mur. In *Wiener Mitteilungen 199 - Hochwasserprognose. Erfahrungen, Entwicklungen & Realität.*

Scherrer, S. C., Appenzeller, C., Eckert, P., and Cattani, D. (2004). Analysis of the Spread-Skill Relations Using the ECMWF Ensemble Prediction System over Europe. *Weather Forecast.*, 19:552–565.

Schmeits, M. and Kok, J. (2010). A Comparison between Raw Ensemble Output, (Modified) Bayesian Model Averaging, and Extended Logistic Regression Using ECMWF Ensemble Precipitation Reforecasts. *Mon. Weather Rev.*, 138:4199–4211. doi:10.1175/2010MWR3285.1.

Schöber, J., Achleitner, S., Kirnbauer, R., Schöberl, F., and Schönlaub, H. (2010). Hydrological modelling of glacierized catchments focussing on the validation of simulated snow patterns - applications within the flood forecasting system of the tyrolean river inn. *Adv. Geosci.*, 27:99–109.

Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. *Agr. Forest Meteorol.*, 98-99:279–293. doi:10.1016/S0168-1923(99)00105-7.

Senarath, S. U. S., Ogden, F. L., Downer, C. W., and Sharif, H. O. (2000). On the calibration and verification of two-dimensional, distributed, hortonian, continous watershed models. *Water Resourc. Res.*, 36(6):1495–1510. doi:10.1029/2000WR900039.

Simic, A., Fernandes, R., Brown, R., Romanov, P., and Park, W. (2004). Validation of VEGETATION, MODIS, and GOES + SSM/I snow-cover products over canada based on surface snow depth observations. *Hydrol. Process.*, 18(6):1089–1104. doi:10.1002/hyp.5509.

Sivapalan, M. (2003). Process complexity at hillslope scale, process simplicity at the watershed scale: Is there a connection? *Hydrol. Process.*, 17:1037–1041.

Skøien, J. O. and Blöschl, G. (2006). Sampling scale effects in random fields and implications for environmental monitoring. *Environ. Monit. Assess.*, 114(1-3):521–552. doi:10.1007/s10661-006-4939-z.

Skøien, J. O., Blöschl, G., and Western, A. W. (2003). Characteristic space scales and timescales in hydrology. *Water Resourc. Res.*, 39(10):SWC111–SWC1119.

Şorman, A. A., Şensoy, A., Tekeli, A. E., Şorman, A. Ü., and Akyürek, Z. (2009). Modelling and forecasting snowmelt runoff process using the HBV model in the eastern part of turkey. *Hydrol. Process.*, 23(7):1031–1040. doi:10.1002/hyp.7204.

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

Strasser, U. and Mauser, W. (2001). Modelling the spatial and temporal variations of the water balance for the weser catchment 1965-1994. *J. Hydrol.*, 254(1-4):199–214. doi:10.1016/S0022-1694(01)00492-9.

Sturm, M. and Wagner, A. M. (2010). Using repeated patterns in snow distribution modeling: An arctic example. *Water Resour. Res.*, 46(12). doi:10.1029/2010WR009434.

Sui, J. and Koehler, G. (2001). Rain-on-snow induced flood events in southern germany. *J. Hydrol.*, 252(1-4):205–220. doi:10.1016/S0022-1694(01)00460-7.

Szolgay, J. (2004). Multilinear flood routing using variable travel-time discharge relationships on the Hron River. *J. Hydro. Hydromech.*, 52(4):303–316.

Tekeli, A. E., Akyürek, Z., Şorman, A. A., Şensoy, A., and Şorman, A. Ü. (2005). Using MODIS snow cover maps in modeling snowmelt runoff process in the eastern part of turkey. *Remote Sens. Environ.*, 97(2):216–230. doi:10.1016/j.rse.2005.03.013.

Thielen, J., Bartholmes, J. C., Ramos, M., and de Roo, A. P. J. (2009). The European Flood Alert System - Part 1: Concept and development. *Hydrol. Earth Syst. Sci.*, 13:125–140. www.hydrol-earth-syst-sci.net/13/125/2009.

Thirel, G., Rousset-Regimbeau, F., Martin, E., and Habets, F. (2008). On the Impact of Short-Range Meteorological Forecasts for Ensemble Runoff Predictions. *J. Hydrometeorol.*, 9:1301–1317.

Thirel, G., Salamon, P., Burek, P., and Kalas, M. (2011). Assimilation of MODIS snow cover area data in a distributed hydrological model. *Hydrol. Earth Syst. Sci. Discuss.*, 8:1329–1364.

Tong, J., Déry, S. J., and Jackson, P. L. (2009a). Interrelationships between MODIS/Terra remotely sensed snow cover and the hydrometeorology of the Quesnel River Basin, British Columbia, Canada. *Hydrol. Earth Syst. Sci.*, 13(8):1439–1452.

Tong, J., Déry, S. J., and Jackson, P. L. (2009b). Topographic control of snow distribution in an alpine watershed of western Canada inferred from spatially-filtered MODIS snow products. *Hydrol. Earth Syst. Sci.*, 13(3):319–326.

Udnæs, H.-C., Alfnes, E., and Andreassen, L. M. (2007). Improving runoff modelling using satellite-derived snow covered area? *Nord. Hydrol.*, 38(1):21–32. doi:10.2166/nh.2007.032.

Verbunt, M., Walser, A., Gurtz, J., Montani, A., and Schär, C. (2007). Probabilistic flood forecasting with a limited-area ensemble prediction system: Selected case studies. *J. Hydrometeorol.*, 8:897–909. doi:10.1175/JHM594.1.

Viglione, A., Chirico, G. B., Komma, J., Woods, R., Borga, M., and Blöschl, G. (2010a). Quantifying space-time dynamics of flood event types. *J. Hydrol.*, 394(1-2):213–229. doi:10.1016/j.jhydrol.2010.05.041.

Viglione, A., Chirico, G. B., Woods, R., and Blöschl, G. (2010b). Generalised synthesis of space-time variability in flood response: An analytical framework. *J. Hydrol.*, 394(1-2):198–212. doi:10.1016/j.jhydrol.2010.05.047.

Vivoni, E. R., Entekhabi, D., Bras, R. L., and Ivanov, V. Y. (2007). Controls on runoff generation and scale dependence in a distributed hydrologic model. *Hydrol. Earth. Sys. Sci.*, 11(5):1683–1701.

Wigmosta, M. S., Vail, L. W., and Lettenmaier, D. P. (1994). A distributed hydrology-vegetation model for complex terrain. *Water Resour. Res.*, 30(6):1665–1679. doi:10.1029/94WR00436.

Wilks, D. S. (1995). *Statistical Methods in the Atmospheric Sciences*, volume 59 of *International Geophysics Series*. Academic Press, San Diego.

Xiong, L. and Guo, S. (2004). Effects of the catchment runoff coefficient on the performance of TOPMODEL in rainfall-runoff modelling. *Hydrol. Process.*, 18(10):1823–1836. doi:10.1002/hyp.1449.

Yatheendradas, S., Wagener, T., Gupta, H., Unkrich, C., Goodrich, D., Schaffner, M., and Stewart, A. (2008). Understanding uncertainty in distributed flash flood forecasting for semiarid regions. *Water Resour. Res.*, 44(5). doi:10.1029/2007WR005940.

Zappa, M. (2008). Objective quantitative spatial verification of distributed snow cover simulations - an experiment for the whole of switzerland. *Hydrolog. Sci. J.*, 53(1):179–191. doi:10.1623/hysj.53.1.179.

Zappa, M., Jaun, S., Germann, U., Walser, A., and Fundel, F. (2011). Superposition of three sources of uncertainties in operational flood forecasting chains. *Atmos. Res.*, 100:246–262. doi:10.1016/j.atmosres.2010.12.005.

Zehe, E. and Blöschl, G. (2004). Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions. *Water Resour. Res.*, 40:W10202. doi:10.1029/2003WR002869.

Zehe, E., Elsenbeer, H., Lindenmaier, F., Schulz, K., and Blöschl, G. (2007). Patterns of predictability in hydrological threshold systems. *Water Resour. Res.*, 43(7). doi:10.1029/2006WR005589.