Journal of Hydrology 402 (2011) 340–356

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Climate and catchment controls on the performance of regional flood simulations

Thomas Nester *, Robert Kirnbauer, Dieter Gutknecht, Günter Blöschl

Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria

ARTICLE INFO

Article history:
Received 15 February 2011
Accepted 21 March 2011
Available online 12 April 2011

This manuscript was handled by K. Georgakakos, Editor-in-Chief, with the assistance of Enrique R. Vivoni, Associate Editor

Keywords:
Rainfall–runoff modelling
Calibration
Climate
Model performance

SUMMARY

Flood runoff is simulated for 57 catchments in Austria and Southern Germany. Catchment sizes range from 70 to 25,600 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 climatological variables 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.

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, 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 sub-catchments 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 (1000–1230 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

* Corresponding author. Address: Vienna University of Technology, Karlsplatz 13/223, A-1040 Wien, Austria. Tel.: +43 1 58801 22313; fax: +43 1 58801 22399.
E-mail address: nester@hydro.tuwien.ac.at (T. Nester).

0022-1694/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2011.03.028

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 nonlinear), 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, a short description of the model is given in Section 3. Section 4 presents the results found in the simulation runs, and in Section 5 the results are discussed.

2. Study region and data

The study region is hydrologically diverse covering large parts of Austria and some parts of Bavaria (Fig. 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.

Fig. 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. Fig. 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 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 25,600 km^2 and a total size of 43,800 km^2. The median size of the catchments is around 400 km^2. The small

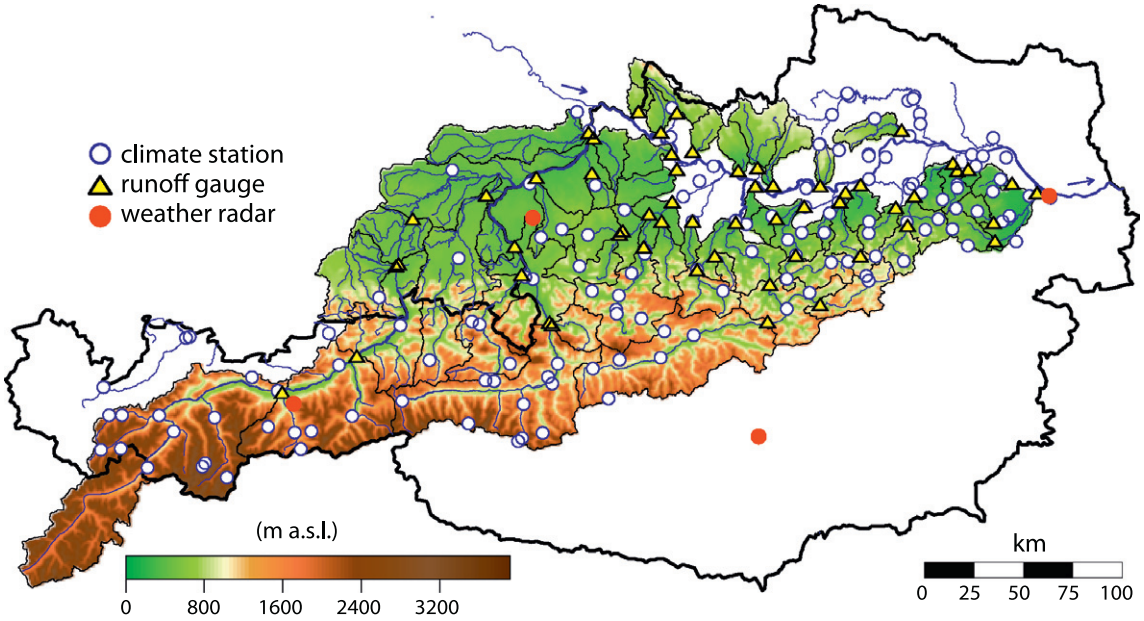

Fig. 1. Topography of Austria and parts of Southern Germany. Elevations within the model region are shown as darker colours. The stream gauges used in the study are indicated by triangles, the precipitation gauges by circles. Weather radar stations are indicated by red circles.

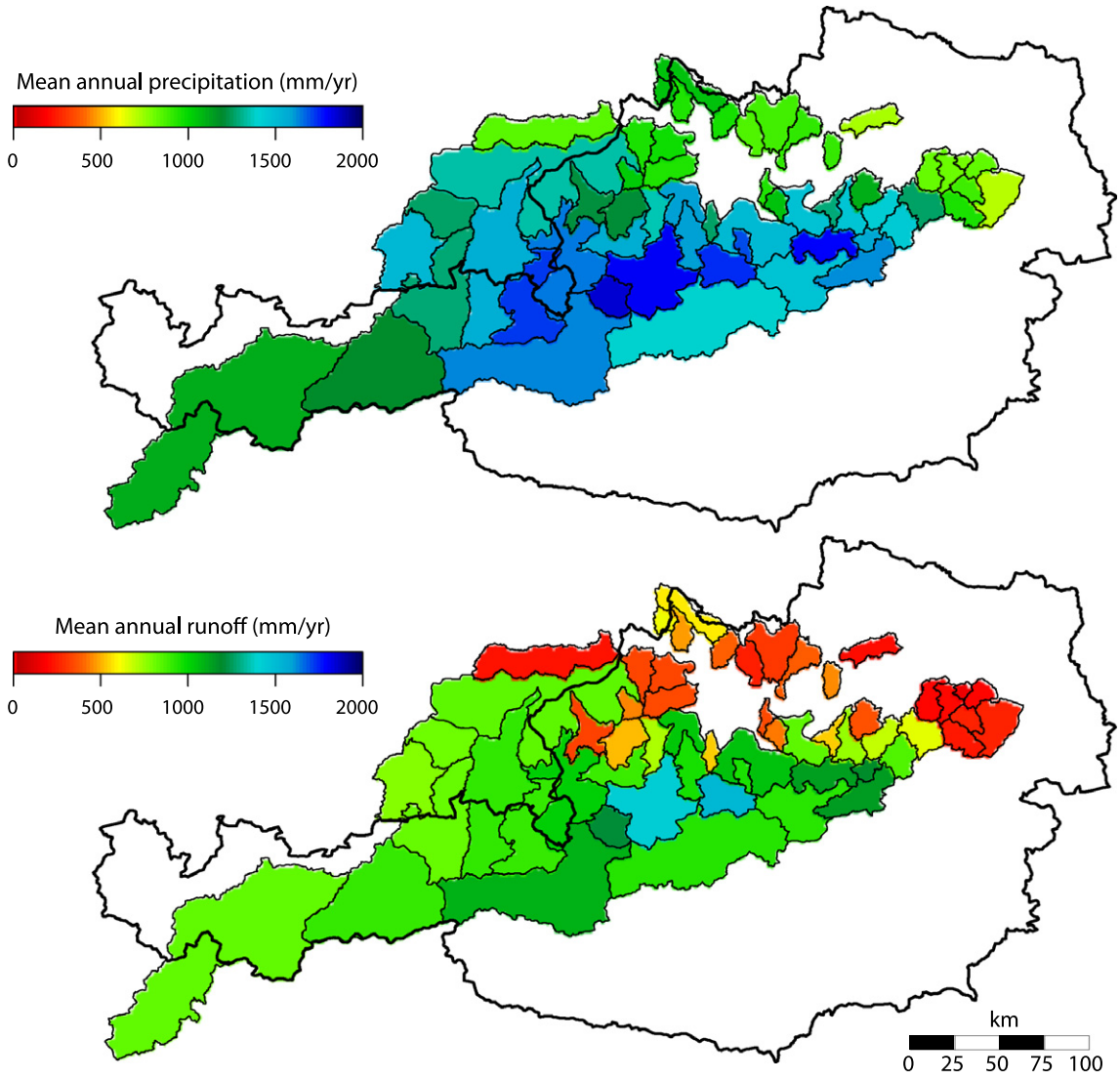

Fig. 2. Top – mean annual precipitation calculated from the precipitation data used in the study for the years 2003–2009, bottom – mean annual runoff depths calculated from discharge data for the years 2003–2009.

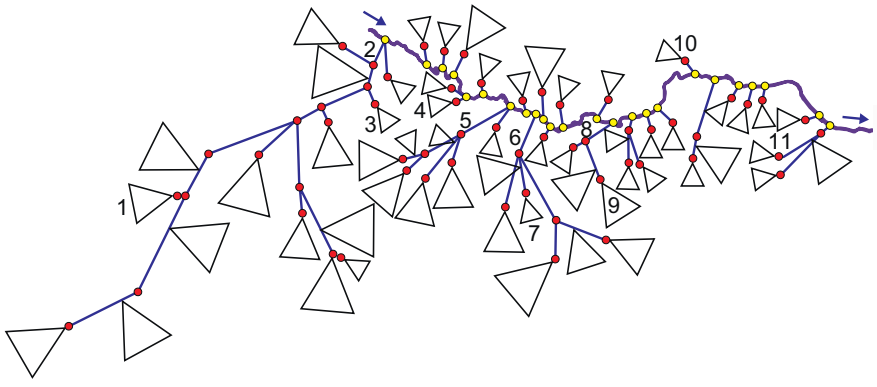

Fig. 3. Model layout. Numbers refer to the gauges and catchments in Table 1. Triangles represent catchments, lines represent routing modules. The size of the triangles indicates the size of the sub-catchments: the smallest triangles stand for catchments with less than 400 km^2, and the largest for areas larger than 1890 km^2.

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, 2003–2006 was the

Table 1
Stream gauges named in the paper. MAP and MAR (2002–2009) is mean annual precipitation and runoff, respectively.

Number	Gauge/catchment	Area (km^2)	MAP (mm/yr)	MAR (mm/yr)
1	Rosenheim/Mangfall	1100	1520	770
2	Schärding/Inn	25,600	1040	830
3	Haging/Antiesen[a]	160	1030	440
4	Fraham/Innbach[a]	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	1260	840
9	Opponitz/Ybbs	510	1800	1140
10	Krems/Imbach[a]	300	720	210
11	Cholerakapelle/Schwechat[a]	180	890	250

[a] Denotes catchments we consider to be dry.

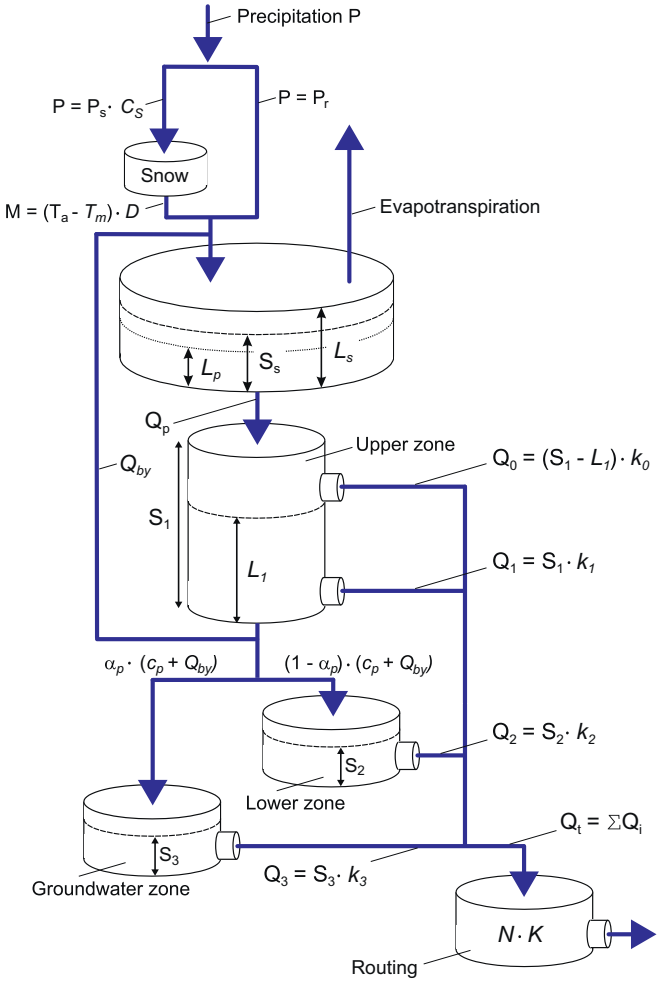

Fig. 4. Model scheme.

calibration period and 2007–2009 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 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 1×1 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).

Table 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

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.

3. Model

3.1. Model structure

Fig. 3 shows the spatial layout of the model. A total of 57 subcatchments 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[1] points and the confluences with the main stream of the

[1] For interpretation of color in Figs. 1 and 3, the reader is referred to the web version of this article.

Danube are shown as yellow points. Table 1 gives details about catchments named in the paper such as area, mean annual precipitation and runoff.

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. 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

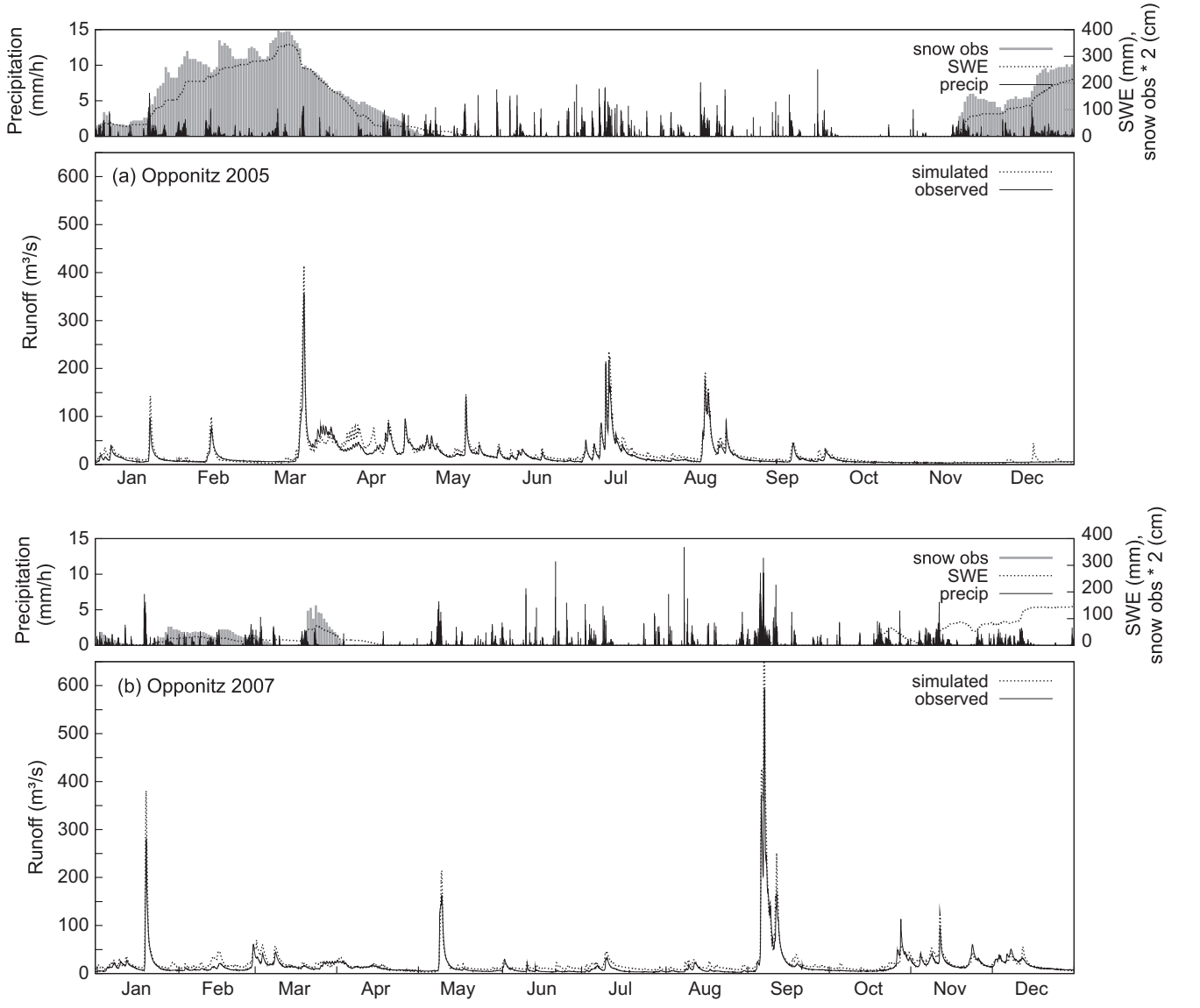

Fig. 5. Calibration (top) and validation (bottom) results for a climatologically wet catchment: Gauge Opponitz/Ybbs. Mean annual runoff depth is 1100 mm/yr, catchment size is 510 km^2. Gauge number 9 in Fig. 3. SWE is the snow water equivalent. Snow depth measurement for a single representative station (approx. mean catchment elevation) in the catchment is shown (data only available until September 2007).

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).

3.2. 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, 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

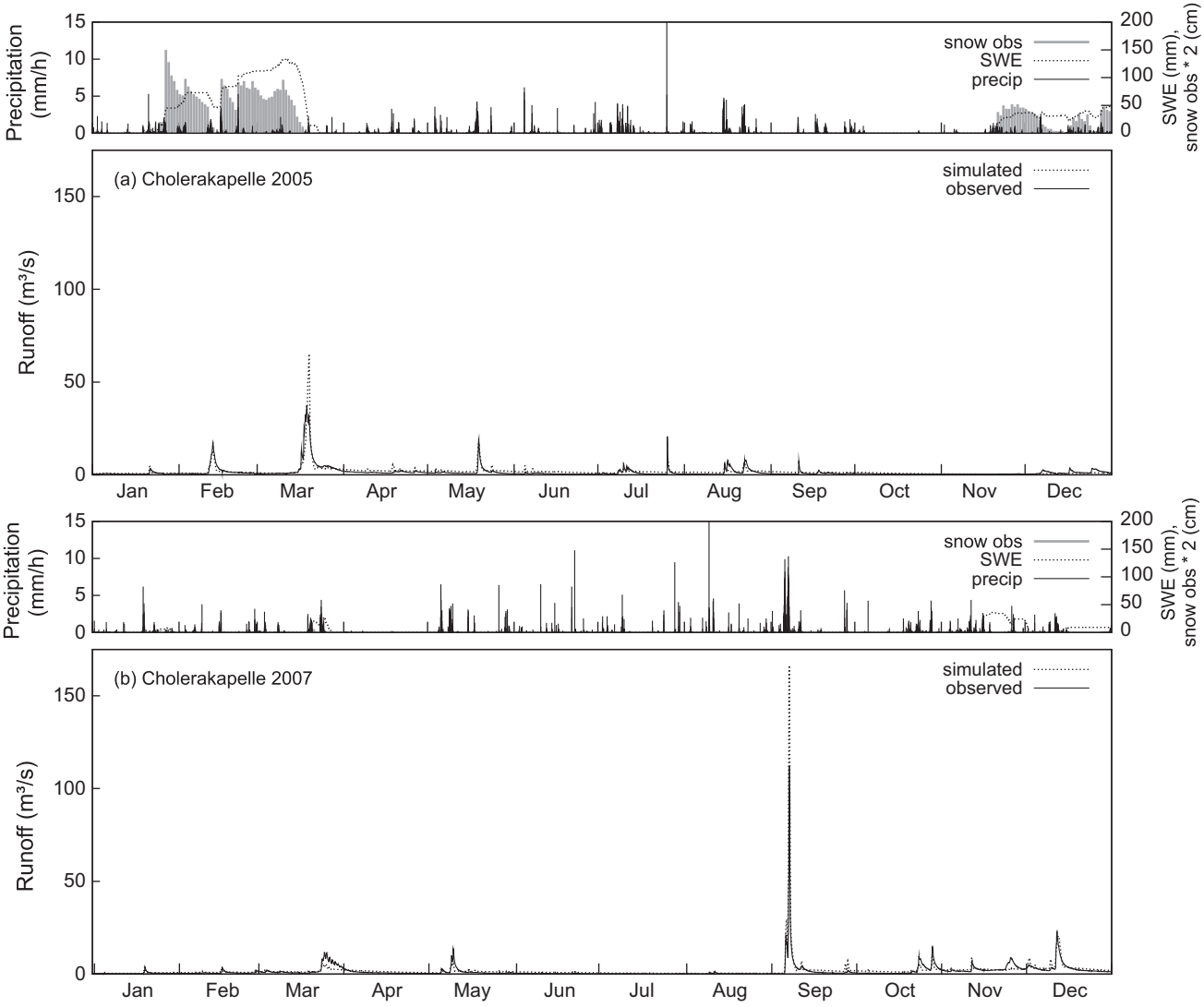

Fig. 6. Calibration (top) and validation (bottom) results for a dry catchment: Gauge Cholerakapelle/Schwechat. Mean annual runoff depth is 250 mm/yr, catchment size is 181 km^2. Gauge is labelled as number 11 in Fig. 3. Snow depth measurement for a single representative station (approx. mean catchment elevation) in the catchment is shown (data only available until September 2007).

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.

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 and Pistotnik, 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 0.8–1.6 °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 1.3–2.3 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.

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

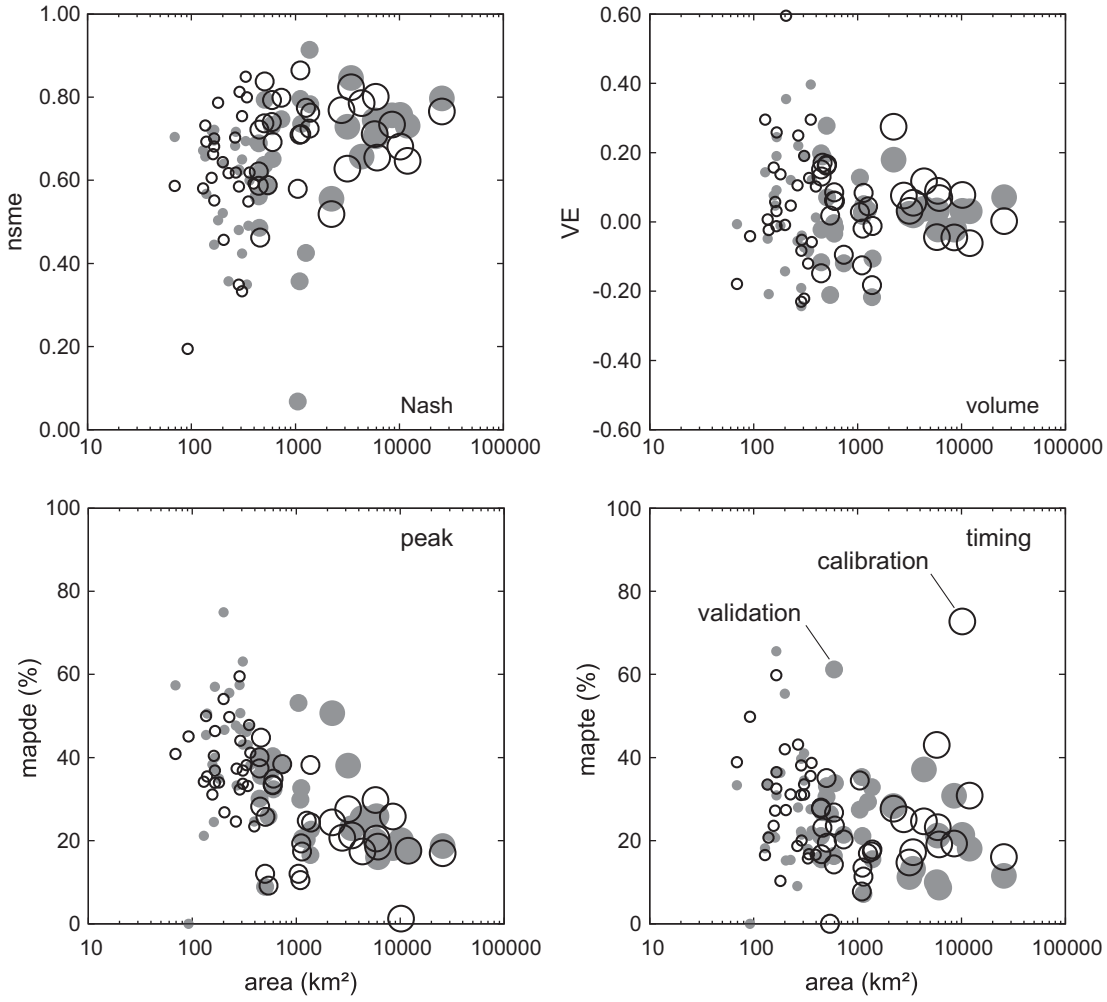

Fig. 7. Nash–Sutcliffe model efficiency (*nsme*), volume error (*VE*), mean absolute peak discharge error (*mapde*) and mean absolute peak time error (*mapte*) plotted versus catchment area. The marker sizes indicate the catchment size. The open circles relate to the calibration period, the full circles to the validation period.

routine; values for the maximum soil moisture L_S are in the range of 0–600 mm and the nonlinearity parameter β is varied in the range of 0–20 (–); values for storage parameters are in the range of 0–2 days for k_0, 2–30 days for k_1 and 30–250 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 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.

4. Results

4.1. 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. 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. 3) has a mean annual precipitation of around 890 mm/yr and mean annual runoff of around 250 mm/yr. Fig. 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 (Fig. 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 Figs. 5 and 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

Table 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**	**0.50**	**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

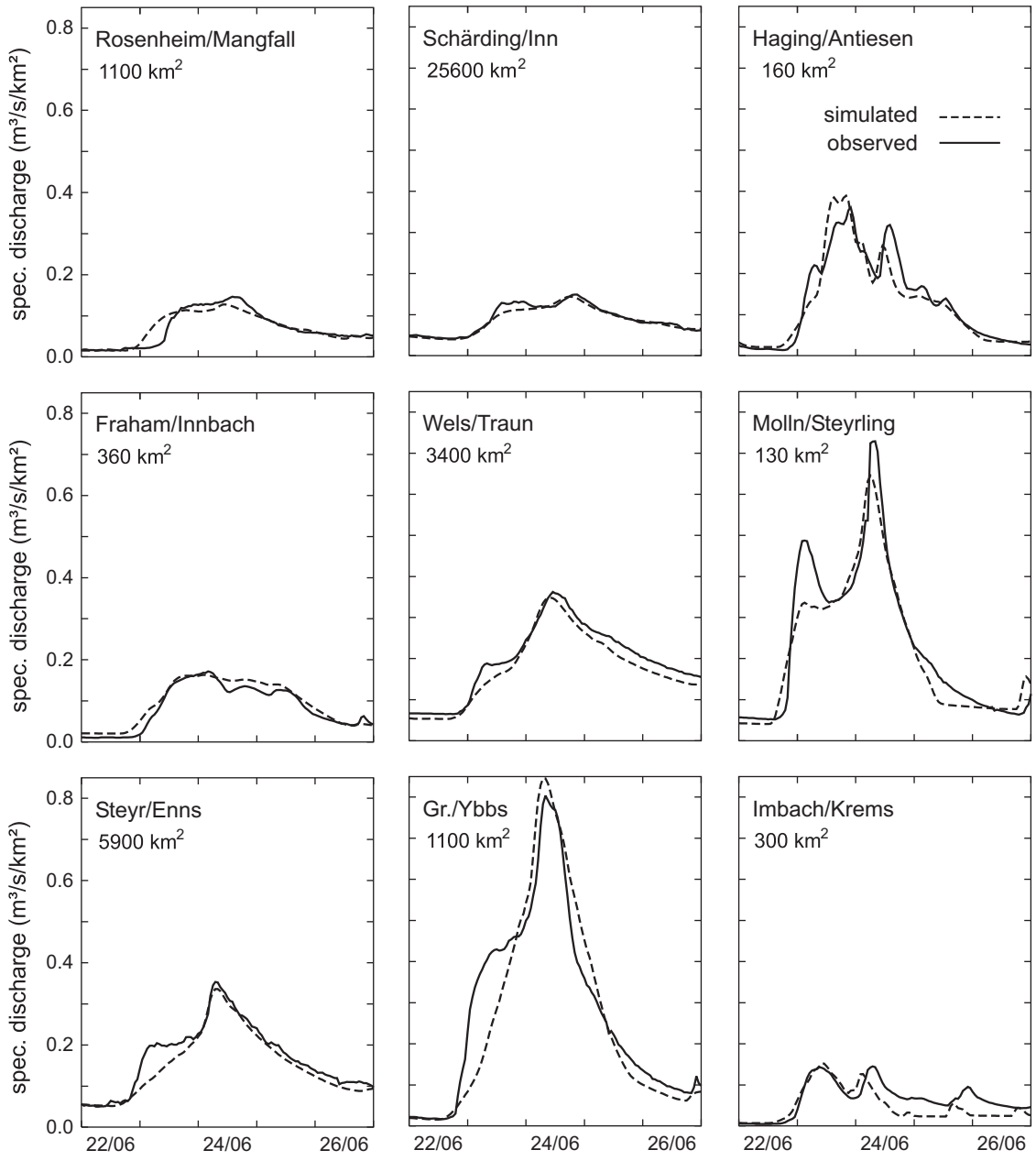

Fig. 8. Event hydrographs for the June 2009 event during the validation period. Gr./Ybbs stands for Greimpersdorf/Ybbs.

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.

Fig. 6 provides the results of a drier catchment in the forelands of the Alps (Cholerakapelle/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 Fig. 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

Table 4
Catchment precipitation totals for the event June 22–26, 2009 shown in Fig. 8.

Gauge	Number in Fig. 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

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 (Figs. 5 and 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.

4.2. 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. 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 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* for the entire period (Table 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.

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

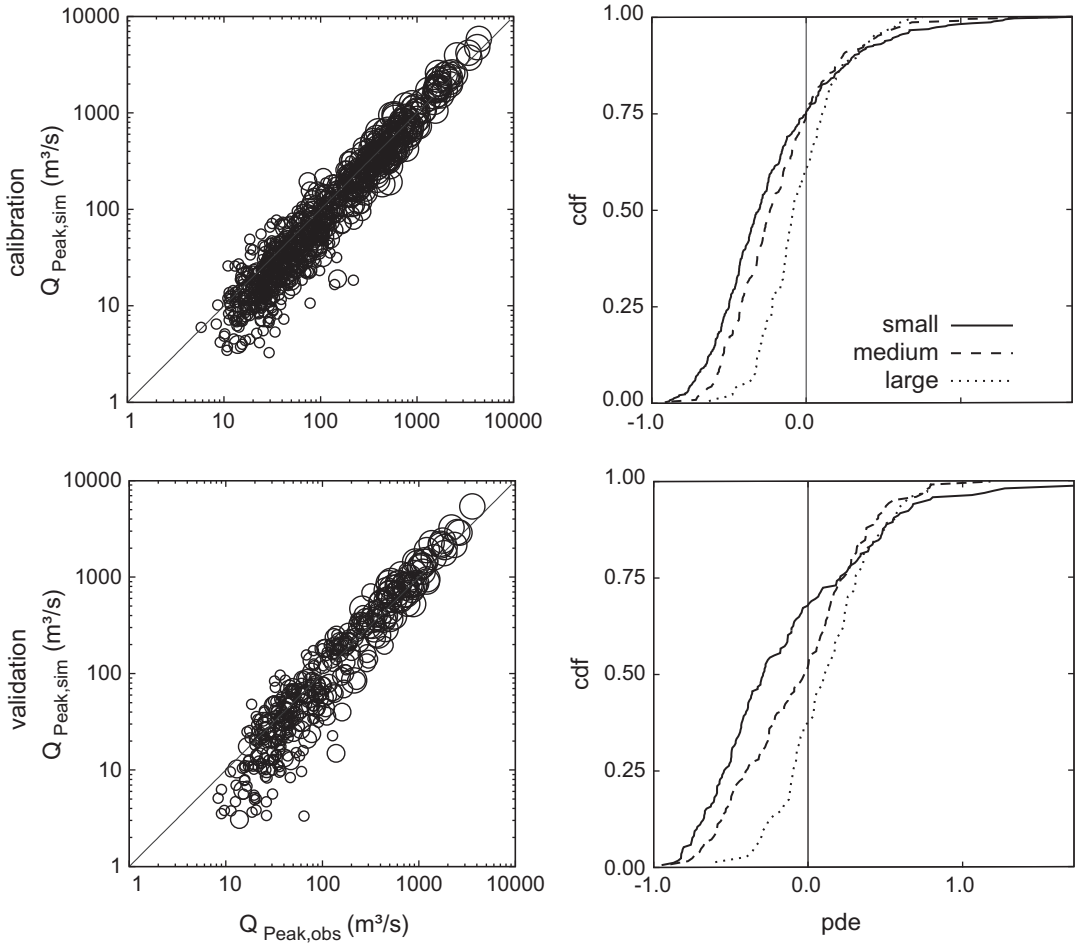

Fig. 9. Scatter plots of $Q_{obs,peak}$ vs. $Q_{sim,peak}$ (left side) and cdf of peak discharge errors (pde) (right side). Size of markers represents catchment size. Calibration (top), validation (bottom). The line types in the right panels relate to three groups of catchment sizes (<400 km^2, 400–1890 km^2, and >1890 km^2).

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 Fig. 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 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 h, 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.

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

Table 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 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

Table 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.

	A (km^2)	N	Average									Median								
			nsme			*VE*			*mapde*		*mapte*	*nsme*			*VE*			*mapde*		*mapte*	
			Calib.	Valid	Total	Calib.	Valid	Total	Calib.	Valid	Total	Calib.	Valid	Total	Calib.	Valid	Total	Calib.	Valid	Total	Calib.	Valid	Total
Small catchments	<400	27	0.62	0.60	0.62	0.06	0.04	0.05	38	43	42	30	27	30	0.62	0.64	0.64	0.04	0.01	0.03	37	46	42	31	22	30
Medium catchments	400–1890	18	0.71	0.63	0.67	0.03	0.01	0.01	27	31	31	20	27	24	0.72	0.67	0.69	0.05	0.01	0.03	27	33	34	19	27	23
Large catchments	>1890	13	0.71	0.73	0.70	0.05	0.03	0.04	20	25	23	28	19	22	0.72	0.74	0.73	0.06	0.03	0.04	21	20	21	24	18	25
All catchments	70–25,600	57	0.66	0.63	0.65	0.05	0.03	0.03	31	36	34	26	25	26	0.69	0.68	0.67	0.05	0.02	0.03	33	37	35	24	22	25

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 30–50% 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 11–41% for the medium absolute percentage error of peak discharges.

4.3. 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 Fig. 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 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 Fig. 11. The patterns are similar to those in Fig. 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 r_s to separate the two effects. Correlation coefficients and partial correlation coefficients variables are summarized in Table 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 20–36%

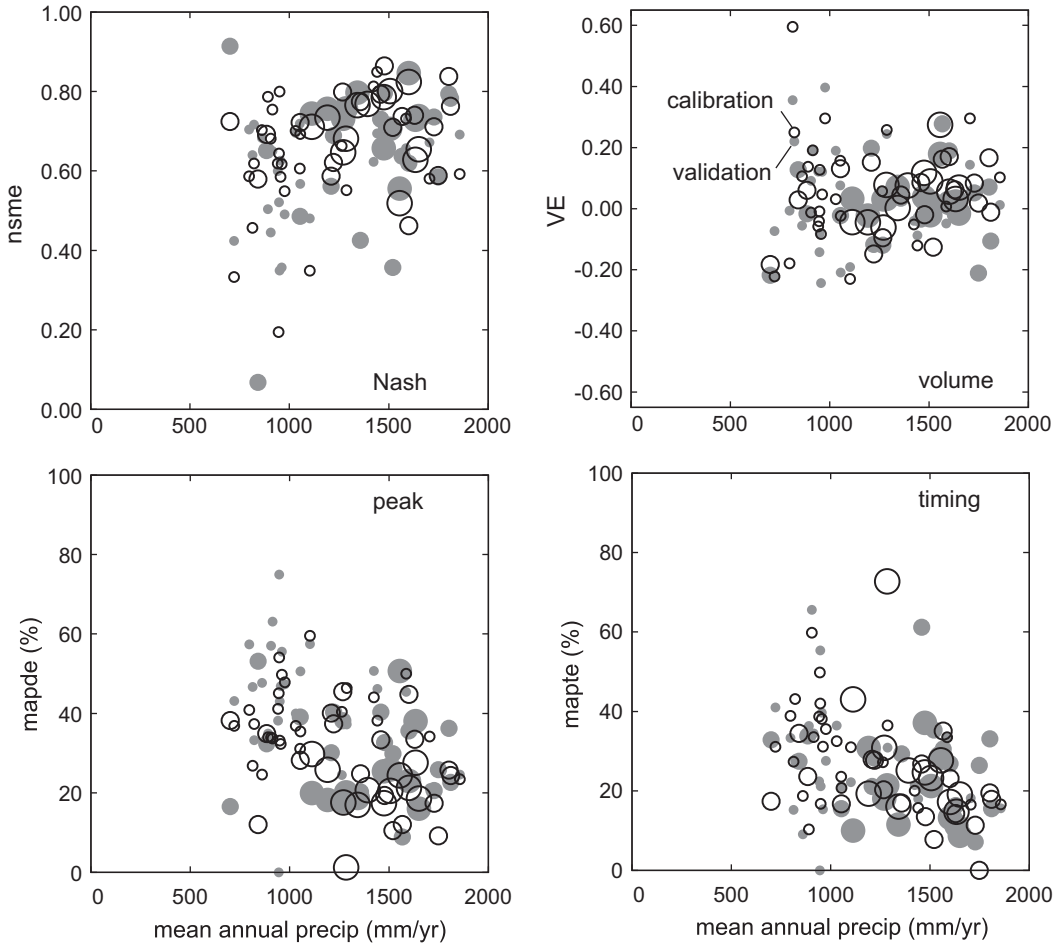

Fig. 10. Nash–Sutcliffe model efficiency (*nsme*), volume error (*VE*), mean absolute peak discharge error (*mapde*) and mean absolute peak time error (*mapte*) plotted versus mean annual precipitation. The marker sizes indicate the catchment size. The open circles relate to the calibration period, the full circles to the validation period.

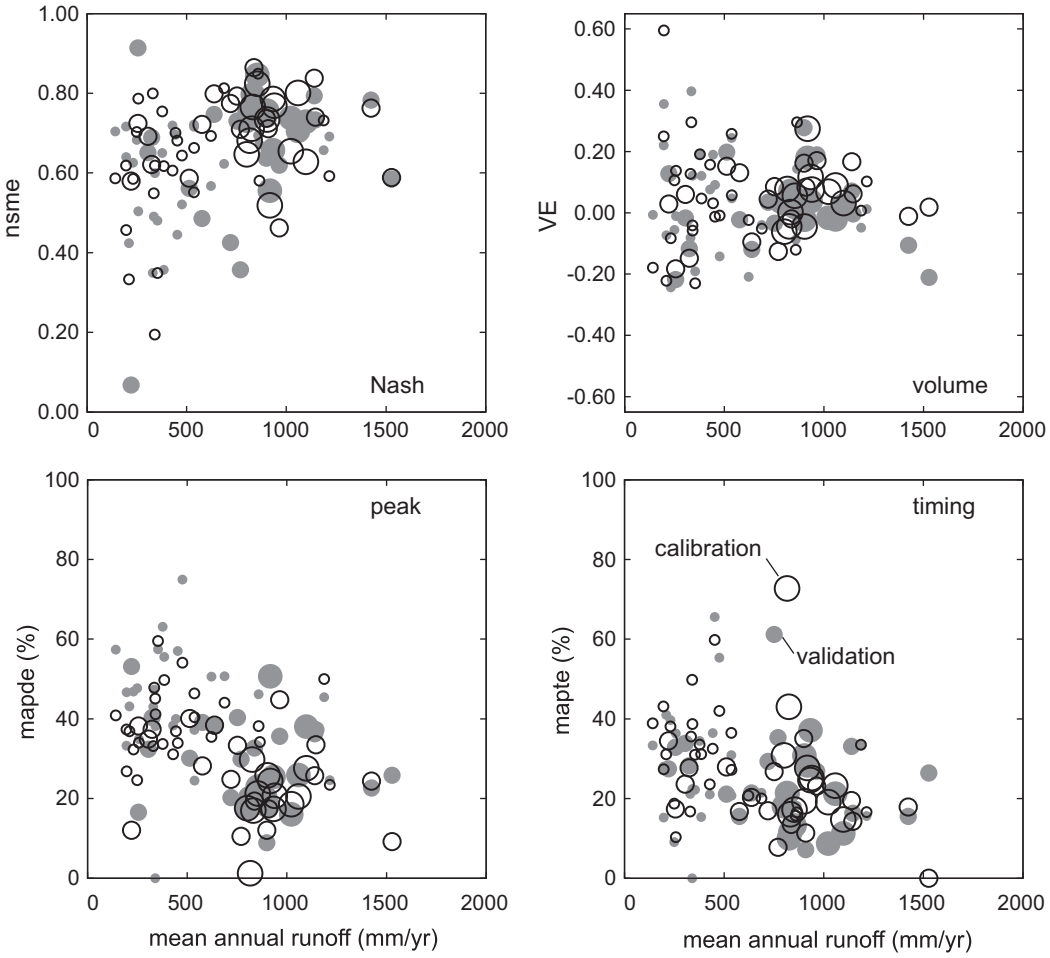

Fig. 11. Nash–Sutcliffe model efficiency (*nsme*), volume error (*VE*), mean absolute peak discharge error (*mapde*) and mean absolute peak time error (*mapte*) plotted versus mean annual runoff. The marker sizes indicate the catchment size. The open circles relate to the calibration period, the full circles to the validation period.

whereas taking away the effects of the area causes a decrease on the order of 30–55% 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 15–25% 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 between the correlations. The influence of the ratio rain/precip. has the largest influence on the correlations between timing error and area.

Fig. 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. Fig. 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% to 90%, with a peak in model performance for a ratio of 75–80%. 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 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 75–80% of the total precipitation is recorded as liquid rain with *nsme* values on the order of 0.70–0.85. These catchments comprise mainly Alpine catchments with a mean annual precipitation of more than 1500 mm/yr. obviously, there is a much larger

Table 6
Partial correlation coefficient based on Spearman's rank correlation coefficient r_s from Table 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	−0.40	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

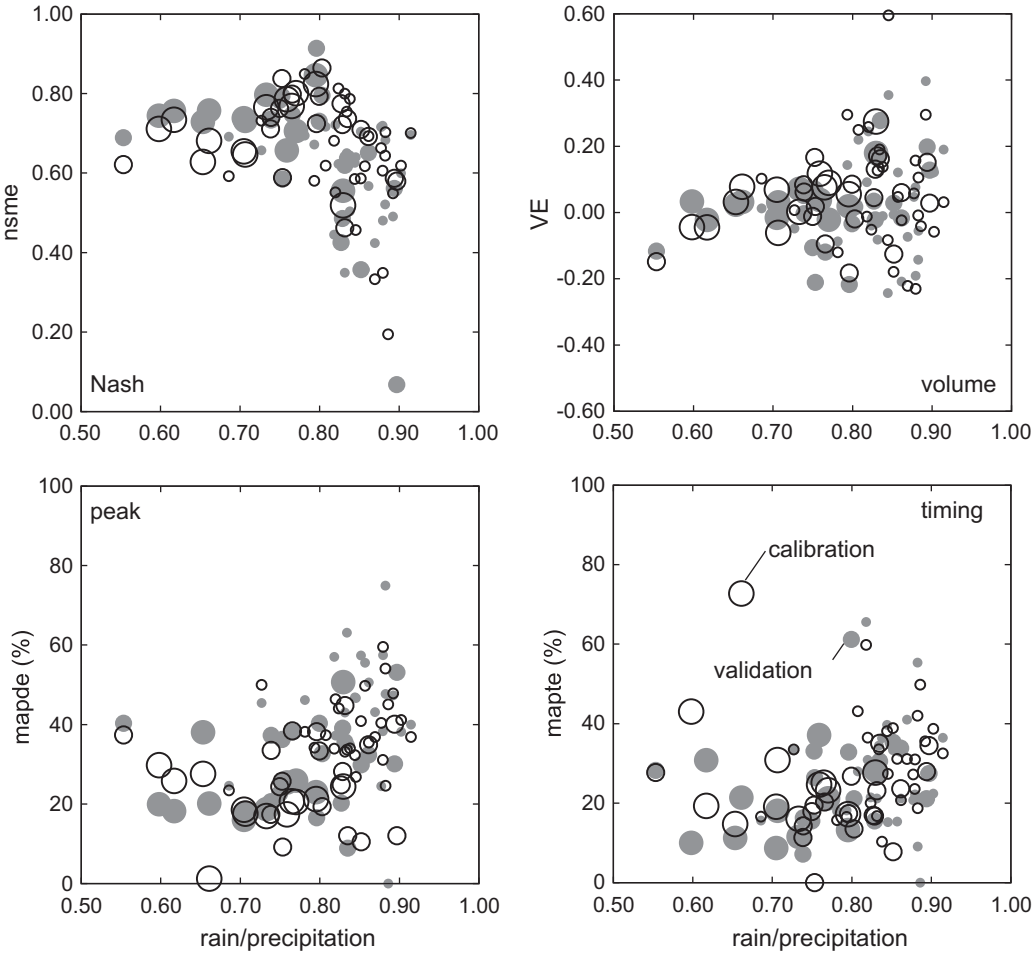

Fig. 12. Nash–Sutcliffe model efficiency (*nsme*), volume error (*VE*), mean absolute peak discharge error (*mapde*) and mean absolute peak time error (*mapte*) plotted versus the long-term ratio of rainfall and total precipitation. The marker sizes indicate the catchment size. The open circles relate to the calibration period, the full circles to the validation period.

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 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*.

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 h 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.,

Table 7
Error statistics of runoff: Nash–Sutcliffe model efficiency (*nsme*), volume error (*VE*), mean absolute peak discharge errors (*mapde*) and mean absolute peak timing errors (*mapte*). *R*/*P* is the long-term ratio of rainfall and total precipitation, *N* is the number of catchments.

	R/*P* (–)	*N*	Average								Median							
			nsme			*VE*			*mapde*			*mapte*			*nsme*			*VE*			*mapde*			*mapte*		
			Calib.	Valid	Total	Calib.	Valid	Total	Calib.	Valid	Total	Calib.	Valid	Total	Calib.	Valid	Total	Calib.	Valid	Total	Calib.	Valid	Total	Calib.	Valid	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	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	0.71	0.67	0.69	0.08	0.03	0.06	28	33	31	25	26	25	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	0.73	0.69	−0.05	−0.02	−0.05	30	29	29	30	26	30	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	0.66	0.63	0.65	0.05	0.03	0.03	31	36	34	26	25	26	0.69	0.68	0.67	0.05	0.02	0.03	33	37	35	24	22	25

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 30–50% 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 $R^2_V = nsme - 0.1\,|nme|$. For the wet catchments they obtained $R^2_V = 0.80$, 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 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 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.

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, 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.

Acknowledgements

We 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.

Appendix A

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

$$nsme = 1 - \frac{\sum_{i=1}^{n}(Q_{sim,i} - Q_{obs,i})^2}{\sum_{i=1}^{n}(Q_{obs,i} - \overline{Q_{obs}})^2}, \quad (1)$$

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 $-\infty$ 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:

$$VE = \frac{\sum_{i=1}^{n} Q_{sim,i} - \sum_{i=1}^{n} Q_{obs,i}}{\sum_{i=1}^{n} Q_{obs,i}}, \quad (2)$$

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

$$pde = \frac{Q_{sim,peak} - Q_{obs,peak}}{Q_{obs,peak}}, \quad (3)$$

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

$$mapde = \frac{1}{m} \cdot \sum_{i=1}^{m} |pde_i| \cdot 100, \quad (4)$$

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

$$pte = \frac{t_{0-peak,sim} - t_{0-peak,obs}}{t_{0-peak,obs}} \quad (5)$$

where $t_{0\text{-}peak,obs}$ and $t_{0\text{-}peak,sim}$ 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

$$mapte = \frac{1}{m} \cdot \sum_{i=1}^{m} |pte_i| \cdot 100, \quad (6)$$

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

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot (n^2 - 1)} \quad \text{with } d_i = rk(x_i) - rk(y_i) \quad (7)$$

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

$$r_{XY,Z} = \frac{r_{XY} - r_{XZ} \cdot r_{YZ}}{\sqrt{(1 - r_{XZ}^2) \cdot (1 - r_{YZ}^2)}} \quad (8)$$

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.

References

Bergström, S., 1976. Development and application of a conceptual runoff model for Scandinavian catchments. Dept. Water Resou. Engng, Lund Inst. Technol./Univ. Lund, Bull. Ser. A52, 134.

Blöschl, G., Kirnbauer, R., 1991. Point snowmelt models with different degrees of complexity – internal processes. J. Hydrol. 129, 127–147.

Blöschl, G., Kirnbauer, R., 1992. An analysis of snow cover patterns in a small Alpine catchment. Hydrol. Process. 6, 99–109.

Blöschl, G., Zehe, E., 2005. On hydrological predictability. Invited commentary. Hydrol. Process. 19, 3923–3929.

Blöschl, G., 2008. Flood warning – on the value of local information. International Journal of River Basin Management 6 (1), 41–50.

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

BMLFUW, 2009. Das Hochwasser in Österreich vom 22. bis 30 Juni, 2009 – Beschreibung der hydrologischen Situation (in German). <http://wasser.lebensministerium.at/filemanager/download/51869/> (accessed 29.09.10).

Braun, L.N., 1985. Simulation of Snowmelt-runoff in Lowland and Lower Alpine Regions in Switzerland. Dissert. Zürcher Geographische Schriften, Heft 21. (ETH) Zürich.

Braun, L.N., 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.

Das, T., Bárdossy, A., Zehe, E., He, Y., 2008. Comparison of conceptual model performance using different representations of spatial variability. J. Hydrol. 356, 106–118.

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

Franchini, M., Pacciani, M., 1991. Comparative analysis of several conceptual rainfall–runoff models. J. Hydrol. 122, 161–219.

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

Grayson, R., Blöschl, G., Western, A., McMahon, T., 2002. Advances in the use of observed spatial patterns of catchment hydrological response. Adv. Water Res. 25, 1313–1334.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1999. Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. J. Hydrol. Eng. 4 (2), 135–143.

Haiden, T., Kann, A., Stadlbacher, K., Steinheimer, M., Wittmann, C., 2010. Integrated Nowcasting Through Comprehensive Analysis (INCA) – System Overview. ZAMG Report, 60 p. <http://www.zamg.ac.at/fix/INCA_system.doc> (accessed 22.02.10).

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

Haiden, T., 2009. Meteorologische Analyse des Niederschlags von 22–25 Juni, 2009 (in German). <http://www.zamg.ac.at/docs/aktuell/2009-06-30_Meteorologische%20Analyse%20HOWA2009.pdf> (accessed 29.09.10).

Hellebrand, H., van den Bos, R., 2008. Investigating the use of spatial discretization of hydrological processes in conceptual rainfall–runoff modelling: a case study for the mesoscale. Hydrol. Process. 22, 2943–2952.
Ivanov, V.Y., Vivoni, E.R., Bras, R.L., Entekhabi, D., 2004. Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: a fully-distributed physically-based approach. J. Hydrol. 298, 80–111.
Jakeman, A.J., Hornberger, G.M., 1993. How much complexity is warranted in a rainfall–runoff model? Water Resour. Res. 29, 2637–2649.
Kienzle, S.W., 2008. A new temperature based method to separate rain and snow. Hydrol. Process. 22, 2067–2085. doi:10.1002/hyp.7131.
Komma, J., Blöschl, G., Reszler, C., 2008. Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting. J. Hydrol. 357, 228–242.
Lidén, R., Harlin, J., 2000. Analysis of conceptual rainfall–runoff modelling performance in different climates. J. Hydrol. 238, 231–247.
Madsen, H., Wilson, G., Ammentorp, H.C., 2002. Comparison of different automated strategies for calibration of rainfall–runoff models. J. Hydrol. 261, 48–59.
Merz, R., Blöschl, G., 2004. Regionalisation of catchment model parameters. J. Hydrol. 287, 95–123.
Merz, R., 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., Blöschl, G., 2009. Scale effects in conceptual hydrological modeling. Water Resour. Res. 45, W09405. doi:10.1029/2009WR007872.
Micovic, Z., Quick, M.C., 2009. Investigation of the model complexity required in runoff simulation at different time scales. Hydrol. Sci. J. 54, 872–885.
Mimikou, M.A., Hatjisava, P.S., Kouvopoulos, Y.S., Anagnostou, E.N., 1992. The influence of basin aridity on the efficiency of runoff predicting models. Nord. Hydrol. 23, 105–120.
Modarres, R., 2009. Multi-criteria validation of artificial neural network rainfall–runoff modeling. Hydrol. Earth Syst. Sci. 13, 411–421.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models: Part I. A discussion of principles. J. Hydrol. 10, 282–290. doi:10.1016/0022-1694(70)90255-6.
Ogden, F.L., Dawdy, D.R., 2003. Peak discharge scaling in small Hortonian watershed. J. Hydrol. Eng. 8 (2), 64–73.
Oudin, L., Andréassian, V., Perrin, C., Michel, C., 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, W03413. doi:10.1029/2007WR006240.
Parajka, J., Merz, R., Blöschl, G., 2003. Estimation of daily potential evapotranspiration for regional water balance modeling in Austria. In: 11th International Poster Day and Institute of Hydrology Open Day "Transport of Water, Chemicals and Energy in the Soil – Crop Canopy – Atmosphere System". Slovak Academy of Sciences, Bratislava, pp. 299–306.
Parajka, J., Merz, R., Blöschl, G., 2007. Uncertainty and multiple objective calibration in regional water balance modeling – case study in 320 Austrian catchments. Hydrol. Process. 21, 435–446.
Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., 2004. Overall distributed model intercomparison project results. J. Hydrol. 298, 27–60.
Reed, S., Schaake, J., Zhang, Z., 2007. A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. J. Hydrol. 337, 402–420. doi:10.1016/j.jhydrol.2007.02.015.
Robinson, J.S., Sivapalan, M., 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 (2), 3089–3101.
Seibert, J., 1999. Regionalisation of parameters for a conceptual rainfall–runoff model. Agr. Forest Meteorol. 98–99, 279–293.
Senarath, S.U.S., Ogden, F.L., Downer, C.W., Sharif, H.O., 2000. On the calibration and verification of two-dimensional, distributed, Hortonian, continuous watershed models. Water Resour. Res. 36 (6), 1495–1510.
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., Blöschl, G., Western, A.W., 2003. Characteristic space scales and timescales in hydrology. Water Resour. Res. 39 (10) (article number 1304).
Skøien, J.O., Blöschl, G., 2006. Sampling scale effects in random fields and implications for environmental monitoring. Environ. Monit. Assess. 114 (1–3), 521–552.
Sui, J., Koehler, G., 2001. Rain-on-snow induced flood events in Southern Germany. J. Hydrol. 252, 205–220.
Szolgay, J., 2004. Multilinear flood routing using variable travel-time discharge relationships on the Hron river. J. Hydro. Hydromech. 52, 4.
Viglione, A., Chirico, G.B., Komma, J., Woods, R., Borga, M., 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., 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., Ivanov, V.Y., 2007. Controls on runoff generation and scale dependence in a distributed hydrologic model. Hydrol. Earth Sys. Sci. 11, 1683–1701.
Xiong, L., Guo, S., 2004. Effects of the catchment runoff coefficient on the performance of TOPMODEL in rainfall–runoff modelling. Hydrol. Process. 18, 1823–1836.
Yatheendradas, S., Wagener, T., Gupta, H., Unkrich, C., Goodrich, D., Schaffner, M., Stewart, A., 2008. Understanding uncertainty in distributed flash flood forecasting for semiarid regions. Water Resour. Res. 44, W05S19. doi:10.1029/2007WR005940.
Zehe, E., Elsenbeer, H., Lindenmaier, F., Schulz, K., Blöschl, G., 2007. Patterns of predictability in hydrological threshold systems. Water Resour. Res. 43, W07434. doi:10.1029/2006WR005589.