



Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria

A. Viglione¹, J. Parajka¹, M. Rogger¹, J. L. Salinas¹, G. Laaha², M. Sivapalan³, and G. Blöschl¹

¹Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria

²Institute of Applied Statistics and Computing, University of Natural Resources and Life Sciences, BOKU Vienna, Austria

³Departments of Civil and Environmental Engineering and Geography, University of Illinois at Urbana-Champaign, Urbana, USA

Correspondence to: A. Viglione (viglione@hydro.tuwien.ac.at)

Received: 21 December 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 14 January 2013

Revised: 15 May 2013 – Accepted: 24 May 2013 – Published: 21 June 2013

Abstract. This is the third of a three-part paper series through which we assess the performance of runoff predictions in ungauged basins in a comparative way. Whereas the two previous papers by Parajka et al. (2013) and Salinas et al. (2013) assess the regionalisation performance of hydrographs and hydrological extremes on the basis of a comprehensive literature review of thousands of case studies around the world, in this paper we jointly assess prediction performance of a range of runoff signatures for a consistent and rich dataset. Daily runoff time series are predicted for 213 catchments in Austria by a regionalised rainfall–runoff model and by Top-kriging, a geostatistical estimation method that accounts for the river network hierarchy. From the runoff time-series, six runoff signatures are extracted: annual runoff, seasonal runoff, flow duration curves, low flows, high flows and runoff hydrographs. The predictive performance is assessed in terms of the bias, error spread and proportion of unexplained spatial variance of statistical measures of these signatures in cross-validation (blind testing) mode. Results of the comparative assessment show that, in Austria, the predictive performance increases with catchment area for both methods and for most signatures, it tends to increase with elevation for the regionalised rainfall–runoff model, while the dependence on climate characteristics is weaker. Annual and seasonal runoff can be predicted more accurately than all other signatures. The spatial variability of high flows in ungauged basins is the most difficult to estimate followed by the low flows. It also turns out that in this data-rich study in Austria, the geostatistical approach (Top-kriging) generally outperforms the regionalised rainfall–runoff model.

1 Introduction

Even in highly monitored areas, only a fraction of catchments possess a stream gauge where water levels are gauged, which are then transformed into runoff, i.e. the volume of water per unit time that flows through a cross section of a stream. All other stream sections are ungauged, and yet runoff information is needed almost everywhere people live for a multitude of purposes such as water resources management, assessment of hydropower potential, design of spillways, culverts, dams and levees, for reservoir management, river restoration, water quality issues, etc. The only recourse is therefore to predict runoff in these catchments or locations using alternative data or information (Sivapalan et al., 2003; Hrachowitz et al., 2013). This is notoriously a difficult task because the predictive uncertainties tend to be large relative to the magnitude of the runoff to be predicted. These uncertainties are due to many reasons. Hydrological processes have enormous spatiotemporal variability, which is difficult to capture (Grayson and Blöschl, 2000). Any stream gauge may be far from the ungauged basin of interest and there may be uncertainties in the collected data (Montanari, 2007). Moreover predictive errors of methods arise from data uncertainties, model structure uncertainties and model parameter uncertainties (Montanari, 2011).

The objective of this and two companion papers (Parajka et al., 2013; Salinas et al., 2013) is to assess the predictive performance of methods for runoff prediction in ungauged basins. In order to estimate the total uncertainty to be expected, a “comparative assessment” is performed, i.e. predictions are tested against independent data simultaneously in many catchments through leave-one-out cross validation (Efron and Gong, 1983). Testing the predictions against independent data in many catchments provides insights on where particular methods work best and what factors control the performances. Moreover, testing the predictions against independent data is one way of testing our hypotheses on how runoff response of catchments works.

The runoff response of catchments constitutes an interesting, complex temporal pattern of water fluxes, which are the result of the collective behaviour of a great number of components of the catchment in response to precipitation and evaporation. The heterogeneity of the meteorological input, the structure of the landscape, the distribution of the vegetation and human intervention, all determine the spatial and temporal variability of the catchment’s hydrologic response. Following Jothityangkoon et al. (2001), the temporal patterns of runoff response of catchments are termed runoff “signatures” (see also Sivapalan, 2005; Wagener et al., 2007). This paper focuses on assessing how well existing methods are able to capture different runoff signatures in ungauged basins. We consider six signatures, each of them meaningful of a certain class of applications of societal relevance: annual runoff, seasonal runoff, flow duration curves, low flows, high flows and runoff hydrographs (Fig. 1).

Annual runoff is a reflection of the catchment dynamics at relatively long timescales, which is particularly evident in its between year variability (Fig. 1a). It is related to the hydrological problem of how much water is available (see e.g. McMahon et al., 2011), which is fundamental for water management purposes such as water allocation, long-term planning, groundwater recharge, etc. Seasonal runoff reflects the within-year variability (Fig. 1b). It addresses the question of when water is available throughout the year (see e.g. Sauquet et al., 2008; Hannah et al., 2011) and is necessary to plan water supply, hydropower production and river restoration measures. The flow duration curve represents the full spectrum of variability in terms of their magnitudes (Fig. 1c). It measures for how many days in a year water is available (Vogel and Fennessey, 1994, 1995) and is the basis of studies on river ecology, hydropower potential, industrial, domestic and irrigation water supply. Low flows focus on the low end of that spectrum, and so provide a window into catchment dynamics when there is little water in the system, and high flows are at the opposite end, when there is much water in the system (Fig. 1d–e). Low flow statistics (Smakhtin, 2001) are needed to estimate environmental flows for ecological stream health, for drought management, river restoration, dilution of effluents, etc. High flow (flood) statistics (Merz and Blöschl, 2008a,b), instead, are required for the design of spillways,

culverts, dams and levees, for reservoir management, river restoration and risk management. Hydrographs are a complex combination of all other signatures (Fig. 1f). They are the most detailed signatures of how catchments respond to water and energy inputs (Parajka et al., 2013). They can be used for all the applications listed above and are specifically needed when the dynamics of runoff have to be taken into account, such as for water quality studies.

Predicting runoff signatures in ungauged basins, and assessing the uncertainties of these predictions, is therefore essential for the water resources issues discussed above. While Parajka et al. (2013) and Salinas et al. (2013) assess the predictive performance of estimation of runoff hydrographs, low flows and floods separately, the focus of this paper is on the predictive performance of estimation on many runoff signatures jointly. The methodology used in this paper differs from the two companion papers also in that Parajka et al. (2013) and Salinas et al. (2013) perform the comparative assessment based on a literature review of many studies from all around the world, which has the advantage of covering a wide range of climates and catchment characteristics, but the disadvantage of comparing different methods applied to different catchments with different data. In this paper, instead, the comparison is based on one consistent dataset in a particularly data-rich region (i.e. 213 catchments in Austria, see Sect. 2). We consider one process-based and one statistical method for predicting runoff hydrographs (Sect. 3) and we assess the predictive performance on measures of the six runoff signatures discussed above (Sect. 4). Specifically, the following questions are addressed in Sect. 5: (i) how well can runoff signatures be predicted in Austria? (ii) In what way does the predictive performance depend on climate and catchment characteristics? (iii) What is the relative performance of the predictions of different signatures? (iv) What is the relative performance of statistical and process-based methods?

2 Study area

For regionalisation of runoff hydrographs, and for the assessment of method performances, daily runoff observations from a total of 213 stream gauges are used. The 213 catchments are assumed representative of the hydrological variability across Austria and their stream gauge position is shown in Fig. 2. The colour of the stream gauges in Fig. 2 indicates the aridity index (ratio of mean annual potential evaporation vs. mean annual precipitation), which varies from 0.2 to 1.0 meaning that there is no really arid catchment in the dataset (i.e. potential evaporation is everywhere lower than precipitation). The largest precipitation rates of more than 2000 mm yr^{-1} occur in the west, mainly due to orographic lifting of northwesterly airflows at the rim of the Alps (see the elevation map in Fig. 2), which causes the highest humidity in the west of the country. Precipitation is lowest in

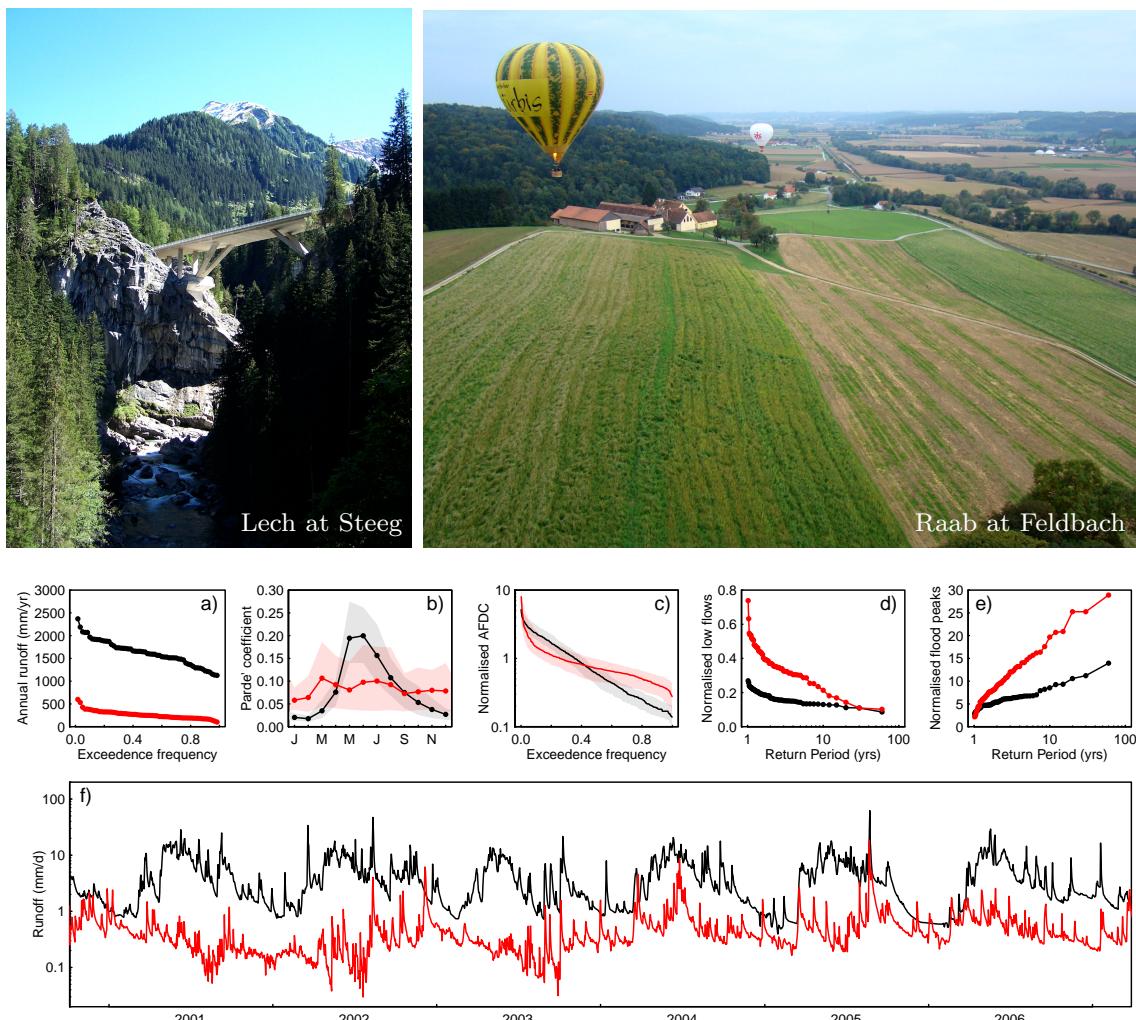


Fig. 1. Six signatures of runoff variability for two catchments: Lech at Steeg (area of 248 km^2 , median elevation of 1944 m a.s.l. and mean annual precipitation of 1520 mm yr^{-1}) in the western Alps (black lines) and Raab at Feldbach (area of 689 km^2 , median elevation of 470 m a.s.l. and mean annual precipitation of 846 mm yr^{-1}) in southeastern Austria (red lines). The six signatures are quantified by (a) distribution of annual runoff; (b) seasonal runoff regime (Pardé coefficient) and 90 % confidence bounds; (c) annual flow duration curve normalised by the mean annual runoff (and 90 % confidence bounds); (d) distribution of annual low flows normalised by the mean annual runoff; and (e) distribution of maximum annual daily flows normalised by the mean annual runoff; and (f) runoff hydrographs. The pictures are representative of the landscape of the two catchments, whose position is shown in Fig. 2.

the lowlands of the east, and the contrast with the Alps is exaggerated by the higher evaporation in the east. This is clearly the case for the two example catchments in Fig. 1. The black curve in Fig. 1a represents the frequency distribution of annual runoff in a mountainous catchment in the western Austrian Alps, the Lech at Steeg (see Fig. 2 for the geographical location) indicating much higher annual runoff than the red curve referring to the Raab at Feldbach, a lowland/hilly catchment in the southeast of Austria.

The seasonality of runoff is very pronounced in the mountainous catchments (e.g. Fig. 1b, black line) where runoff maxima occur in summer because of snow accumulation and melt processes. In the lowlands of the east, the runoff sea-

sonality is the result of the interplay between the seasonality of precipitation and evaporation and is less marked (e.g. Fig. 1b, red line). Snowmelt in the Alpine west also leads to flow duration curves that are steep in their central part (Fig. 1c, black line). In the southeastern catchment the central parts of the flow duration curve is particularly flat (Fig. 1c, red line), which is due to the flashy nature of runoff in the region, due to both convective precipitation and responsive soils (Gaál et al., 2012). This suggests a higher variability of the extremes, which is reflected in the steeper frequency distributions of low flows (Fig. 1d) and floods (Fig. 1e) for the lowland catchment (red lines). The small low flows in the east occur in summer and are related to the seasonality of

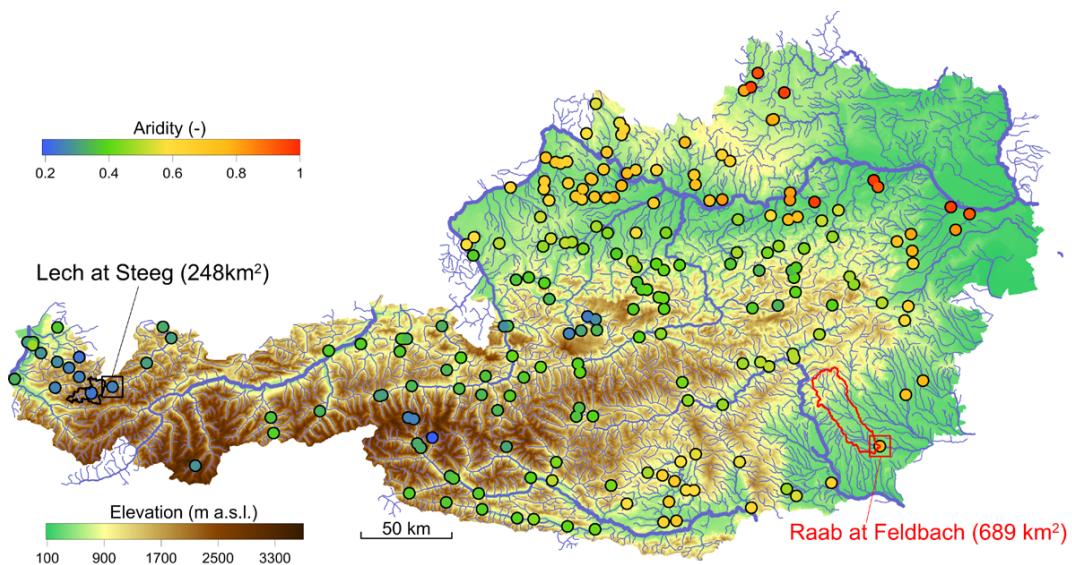


Fig. 2. Topography and river network of Austria and location of 213 streamgauges considered in this study (points) colour-coded according to catchment aridity. The two catchments in Fig. 1 are indicated by the black and red boxes and boundary lines.

runoff with minima in summer. In the Alps in the west there are also small low flows but they occur in winter and are due to snow deposition in the catchments instead of rain. More detailed statistics of the catchment characteristics and runoff signature measures are reported in Table 1.

3 Regionalisation methods

Many regionalisation methods for runoff prediction in ungauged basins exist (see e.g. Hrachowitz et al., 2013). In general terms, we consider them belonging to two different categories: statistical and process based. Statistical methods use available runoff time series data from neighbouring catchments (donor catchments) to estimate runoff signatures at ungauged locations based on one or more similarity measures and/or grouping methods. They usually do not use precipitation data in a causal way. In contrast, process-based methods use precipitation data (and other climate data) to estimate runoff based on water balance equations, i.e. they are based on some variant of rainfall–runoff models. Hereafter we present briefly the two methods used in this study.

3.1 Process-based method: rainfall–runoff model

There is a wide variety of rainfall–runoff models, ranging from physics-based models based on laboratory-scale equations to index-based models and lumped conceptual models (Singh and Frevert, 2005). As noted in Parajka et al. (2013) there are very few studies that have actually examined what model structure would be appropriate for a particular catchment or landscape, to assist in model structure selection for an ungauged catchment (Smith et al., 2004; Reed et al., 2004;

Fenicia et al., 2011). Choice of model structure is therefore usually guided by prior knowledge of the hydrologic system, the availability of data, and prior experience of the practitioner. In this paper we use a semi-distributed conceptual rainfall–runoff model which follows the structure of the Hydrologiska Byråns Vattenbalansavdelning (HBV) model (Bergström, 1995) and which has been used in Austria for quite a long time now (see e.g. Merz and Blöschl, 2004; Parajka et al., 2005, 2007b; Merz et al., 2011). The model runs on a daily time step and consists of a snow routine, a soil moisture routine and a flow routing routine. The detailed description of the model concept is given, e.g. in the Appendix of Parajka et al. (2007b). The climate model inputs (daily precipitation and air temperature) have been obtained by spatial interpolation of daily observations using elevation as auxiliary variable (see Merz et al., 2011). The potential evaporation is estimated by a modified Blaney–Ciddle method (Parajka et al., 2005) using interpolated daily air temperature and grid maps of potential sunshine duration (Mészároš et al., 2002). The dataset used in this study includes measurements of daily precipitation and snow depths at 1091 stations and daily air temperature at 212 climatic stations (Parajka et al., 2005, 2007a). The model inputs are extracted for 200 m elevation zones and used for runoff model simulations in each catchment.

From a total of 14 model parameters, 11 are estimated by using automatic model calibration against observed runoff and accounting for a priori information of the model parameters (see Merz et al., 2011, Sect. 3.2). Temporal validation has been performed by Parajka et al. (2005); Merz et al. (2009, 2011) to make sure that the chosen parameterisations give adequate results in an independent test (validation)

Table 1. Attributes and signature measures for the 213 Austrian catchments in Fig. 2. The signature measures have been calculated from observed daily runoff for the period 1976 to 2008 as explained in Sect. 4.1.

	mean	CV	min	25 %	median	75 %	max
Area (km ²)	411	2.08	13.7	75.8	167	342	6214
Median elev. (m a.s.l.)	1067	0.531	287	606	905	1449	2964
Mean ann. prec. (mm yr ⁻¹)	1201	0.268	605	945	1143	1448	2112
Aridity index (-)	0.511	0.368	0.196	0.371	0.463	0.664	0.979
(a) Mean ann. runoff \bar{Q}_m (mm yr ⁻¹)	869	0.600	170	435	790	1160	2604
(b) Range of Pardé coeff. ΔPar (-)	0.110	0.451	0.0320	0.0716	0.102	0.140	0.275
(c) Slope of FDC m_{FDC} (%/%)	1.46	0.306	0.668	1.16	1.38	1.64	3.13
(d) Normalised low flow q_{95} (-)	0.277	0.399	0.0250	0.197	0.276	0.343	0.631
(e) Normalised high flow q_{05} (-)	2.68	0.159	1.58	2.37	2.65	2.94	4.06
(f) Integral scale $\tau_{1/e}$ (days)	19.8	0.863	2	4	13	37	59

period as stressed by Klemeš (1986), Andréassian et al. (2009), and Gharari et al. (2013), among others. The median Nash–Sutcliffe efficiency (see Eq. 6 for its definition) decreases from 0.72 to 0.66 from calibration to validation period (see Table 2 in Parajka et al., 2005). Note that, in this paper, although the assessment focuses on the runoff signatures, the model has not been calibrated to them, which is what other authors do (see e.g. Yadav et al., 2007; Hingray et al., 2010).

For predictions in ungauged sites, however, calibration to observed runoff is not an option, so the model parameters need to be estimated (regionalised) by using information from other gauged catchments (Parajka et al., 2013, summarise and compare different approaches used for transferring model parameters to ungauged catchments). In this study, we apply the similarity based approach introduced in Parajka et al. (2005). This regionalisation method is based on the idea to find a donor catchment that is most similar to the ungauged site in terms of its catchment attributes (mean catchment elevation, stream network density, lake index, areal proportion of porous aquifers, land use, soils and geology). The complete parameter set from the donor catchment is then transposed to the ungauged catchment and used for modelling of water balance including runoff (see Parajka et al., 2005, 164–165 pp.). The goodness-of-fit of the rainfall–runoff model simulations in the calibration period 1976–2008 gives a median Nash–Sutcliffe efficiency of 0.72 for the entire hydrographs in the 213 sites in Fig. 2. The performance in cross-validation mode for the same period 1976–2008 gives a median Nash–Sutcliffe of 0.61. This is not much worse than the goodness-of-fit in the calibration period, even though it includes the uncertainties of the model and of the parameter regionalisation method (Montanari, 2011).

3.2 Statistical method: Top-kriging

The main advantage of statistical methods of estimating runoff in ungauged basins is that they avoid the use of un-

certain input variables such as precipitation and potential evaporation. In this paper we use Top-kriging (Skøien et al., 2006), which is a geostatistical method that accounts for the river network hierarchy (see also Gottschalk, 1993; Sauquet et al., 2000; Gottschalk et al., 2006, for similar methods). Top-kriging combines two processes: local runoff generation, which is continuous in space, and runoff aggregation and routing along the stream network. The method requires a variogram for local (point) runoff generation. The variogram is then integrated over the catchment areas associated with each river cross section (see e.g. Skøien et al., 2006; Merz et al., 2008). This integrated variogram depends on the point variogram as well as the sizes and the relative positions and nestedness of the catchments. The assumption of a best linear unbiased estimator then gives the kriging weights which are used to estimate the daily runoff for each ungauged basin from the observed daily runoff of neighbouring stations on the same day, weighted by the kriging weights. Top-kriging also provides estimates of the kriging variance. The uncertainties involved are discussed in Skøien and Blöschl (2006a,b, 2007).

For the Top-kriging estimation, daily runoff observations from the 213 stream gauges in Fig. 2 and Table 1 are used. A number of variograms have been tested and the variograms used in Skøien et al. (2006) and Merz et al. (2008) fit the case of runoff time series well. The cross-validation performance for hydrograph regionalisation in the period 1976–2008 gives a median Nash–Sutcliffe of 0.87, higher than for the process-based method.

As an illustration, Fig. 3 shows the result of Top-kriging for the extreme August 2002 flood (see Gutknecht et al., 2002; Komma et al., 2007; Blöschl et al., 2008; Reszler et al., 2008; Viglione et al., 2010a, 2013, for details on the event). The 2002 event covered a large area of northern Austria. It consisted of two frontal type storms, both produced by Vb cyclones (Ulbrich et al., 2003; Mudelsee et al., 2004), which are a typical meteorological situation for long rain events in the region. Even though the rainfall depth associated with the

second storm was lower than in the first storm in most parts of the region, the second storm produced larger flood discharges (see 12 and 13 August in Fig. 3) because of catchment saturation due to the previous event. The maps in Fig. 3 allow one to appreciate the spatiotemporal evolution of the 2002 flood event, in particular the movement of the runoff-contributing areas from west to east on the 12 and 13 of August where most of the damage in Austria occurred.

4 Method for the comparative assessment

In order to assess the performance of the predictive methods in capturing the temporal and spatial variability of runoff in Austria, runoff hydrographs are estimated for each of the 213 catchments in Fig. 2 without using runoff data from that basin, i.e. the catchments are treated as ungauged. Statistical measures of the six signatures discussed above are then extracted from the observed and predicted hydrographs (Sect. 4.1), and are compared through different performance measures (Sect. 4.2). This procedure is known as leave-one-out cross validation (Efron and Gong, 1983) and allows for an independent validation of each methodology used to provide predictions in ungauged basins, rather than enabling just a goodness of fit of a particular regionalisation method. It can be therefore seen as a measure of the total predictive uncertainty in runoff prediction in ungauged basins.

4.1 Signature measures

There are many ways of quantifying each runoff signature. For instance, in Fig. 1 the signatures are quantified by curves. For the comparative assessment, in this paper we quantify the signatures by single values. Given the time series of observed (or simulated) specific daily runoff $Q_d(t)$ (mm d^{-1}), the following statistics are calculated:

- a. the mean annual specific runoff (mm yr^{-1}):

$$\overline{Q_m} = 365 \cdot \overline{Q_d} = \frac{365}{T} \sum_{t=1}^T Q_d(t), \quad (1)$$

where $\overline{Q_d}$ is the mean daily specific runoff (mm d^{-1}) and T (days) is the record length (corresponding to 33 yr in our case);

- b. the range of Pardé coefficients (-):

$$\Delta \text{Par} = \max(\text{Par}_i) - \min(\text{Par}_i), \quad (2)$$

where Par_i , the Pardé coefficient for month i , is defined as the mean monthly runoff volume for the month i divided by the mean annual runoff volume ($\sum_{i=1}^{12} \text{Par}_i = 1$). We calculate it as

$$\text{Par}_i = \frac{\sum_{t \in M_i} Q_d(t)}{\sum_{\forall t} Q_d(t)}, \quad (3)$$

A. Viglione et al.: Part 3: Runoff signatures in Austria

where $t \in M_i$ means all time steps (days) belonging to the month i and $\forall t$ means all time steps, from 1 to T ;

- c. the slope of the flow duration curve (%/%):

$$m_{\text{FDC}} = 100 \cdot \frac{Q_{30\%} - Q_{70\%}}{40 \cdot \overline{Q_d}}, \quad (4)$$

where $Q_{30\%}$ (mm d^{-1}) is the value of daily runoff which is exceeded 30 % of the time (on average, around 110 days a year) and $Q_{70\%}$ 70 % of the time (on average, around 255 days a year). m_{FDC} is a measure of slope of the central part of the flow duration curve and indicates the percentage of increase of runoff, with respect to the annual mean, for 1 % decrease of exceedance probability;

- d. the normalised low flow statistic (-) calculated as $q_{95} = Q_{95\%}/\overline{Q_d}$ where $Q_{95\%}$ (mm d^{-1}) is the value of daily runoff which is exceeded 95 % of the time (on average, around 347 days a year);
- e. the normalised high flow statistic (-) calculated as $q_{05} = Q_{5\%}/\overline{Q_d}$ where $Q_{5\%}$ (mm d^{-1}) is the value of daily runoff which is exceeded 5 % of the time (on average, around 18 days a year);
- f. the integral scale $\tau_{1/e}$ (days) calculated as the time lag at which the autocorrelation function drops below $1/e \sim 0.368$. The autocorrelation function has been estimated with the function "acf" in R (R Core Team, 2012). The integral scale is a raw measure of the runoff hydrograph memory (see e.g. Blöschl and Sivapalan, 1995, p. 255 and reference therein).

Some aggregated statistics of these signature measures are listed in Table 1. Most of the signatures are normalised by the mean (daily) runoff. The rationale for this normalisation is that we aim at assessing the capability of the methods to estimate the volume of runoff once (i.e. $\overline{Q_m}$) and the variability of runoff independently of the volume for the other signatures.

Figure 4 illustrates the spatial variability of the six signature measures in Austria. The figure has been obtained by Top-kriging starting from 213 observed runoff time series in the locations in Fig. 2. In Austria, the spatial patterns observed in these runoff signatures can be traced back to a fairly small subset of key processes, particularly: the role of snow, the absolute volume of precipitation, the seasonality of precipitation and evaporation, and subsurface storage. For example, the snow dynamics in the western Alpine part of Austria are responsible for the pronounced seasonality, the steep flow-duration curves, the winter minima in low flows, and finally for the long integral timescales of runoff. The large volumes of runoff in western Austria, however, relate not to snow, but to the effects of orographic lifting of northwesterly airflows at the rim of the Alps, leading to precipitation rates of more than 2000 mm yr^{-1} . Precipitation is lowest

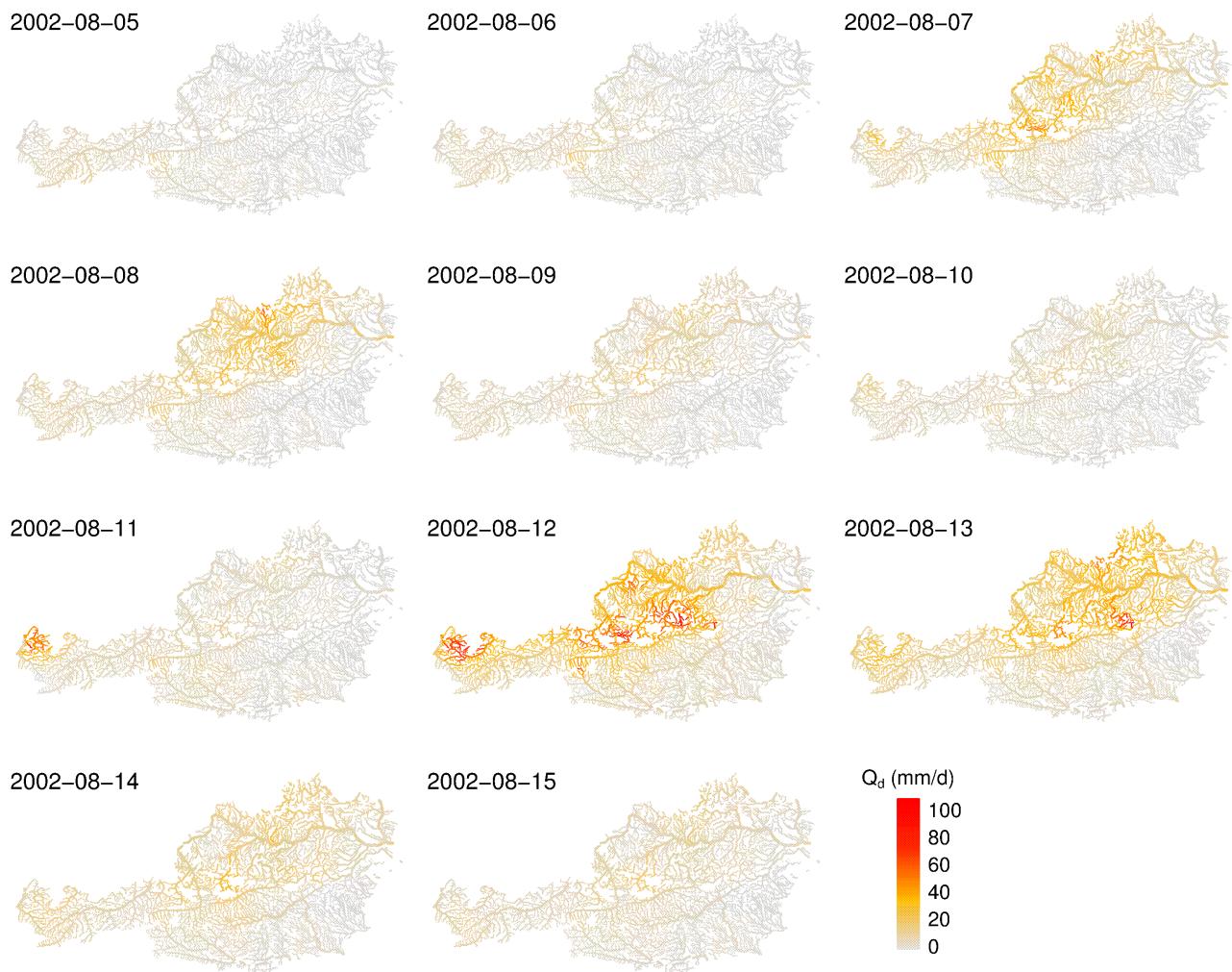


Fig. 3. Maps of daily runoff obtained by regionalising observed runoff time series through Top-kriging (Sect. 3.2). The figure illustrates the extreme flood of August 2002 in northern Austria.

in the lowlands of the east, and the contrast with the Alps is exaggerated by higher evaporation (see also Fig. 1). The role of evaporation in the east is in phase with precipitation maxima in summer, leading to low seasonality of runoff in the absence of snow processes. Otherwise, the role of the catchment morphology and geology in shaping hydrological signatures is most obvious in the flow duration curve and the runoff hydrograph. Aside from snow-dominated areas, flashy locations are associated with convective precipitation and rapidly draining soils: in these regions the integral timescale of runoff is short, and the duration curves are flat. Slow dynamics in the hydrograph also arise in regions with highly pervious geology (as in the south of Austria), and are also reflected in large low flows and small floods.

4.2 Performance measures

We assess the performance of the methods by three statistical metrics.

1. The normalised error, which is defined as

$$\text{NE}_i = \frac{\hat{y}_i - y_i}{y_i}, \quad (5)$$

where y_i is the observed signature at the i th catchment (i from 1 to 213) and \hat{y}_i is the estimated signature. It expresses the error of estimation relative to the observed signature for catchment i . Its spatial median $\widetilde{\text{NE}}$ is a measure of (spatial) bias of estimation in Austria. A positive (negative) value of $\widetilde{\text{NE}}$ means that, on average, the method overestimates (underestimates) the signature of interest.

2. The absolute normalised error, which is defined as $\text{ANE}_i = |\text{NE}_i|$ for catchment i . The spatial median $\widetilde{\text{ANE}}$ is a measure of the average spread of the estimation error. A low value of $\widetilde{\text{ANE}}$ (close to 0) means that, on average, the percentage error of estimation at a catchment is low (i.e. the efficiency of the method is high).

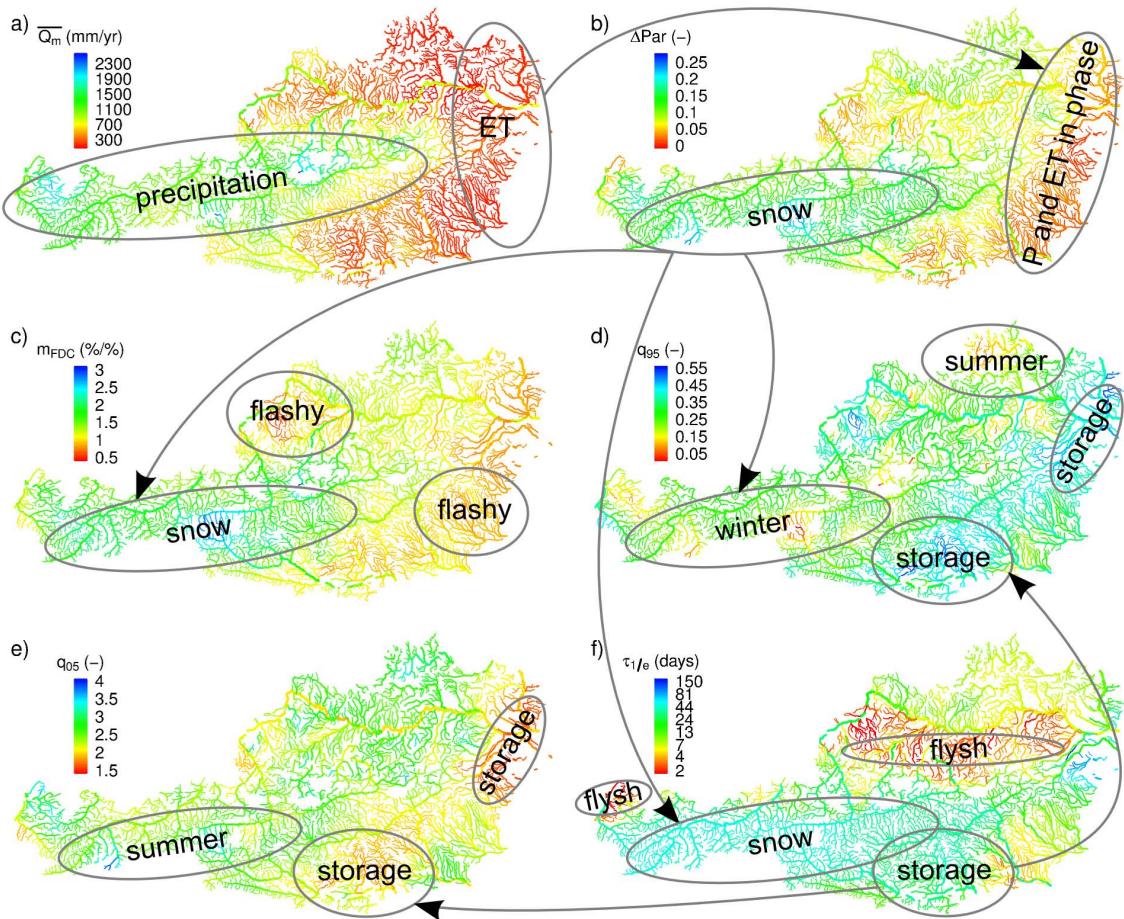


Fig. 4. Connection of runoff signatures for ungauged basins in Austria. The maps have been obtained by regionalising daily runoff time series through Top-kriging (Sect. 3.2) and extracting the six signatures of Sect. 4.1: (a) mean annual runoff, (b) range of Pardé coefficients, (c) slope of the flow duration curve, (d) normalised daily runoff value which is exceeded 95 % of the time as a measure of low flows, (e) normalised daily runoff value which is exceeded 5 % of the time as a measure of high flows, (f) and integral timescale. Ellipses and arrows help the description of connectivity between process and response provided in the text.

3. The coefficient of determination, which is defined as

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad (6)$$

where \bar{y} is the spatial average of the observed signature y_i over the 213 catchments. A high R^2 (close to 1) means that the method captures well the spatial variability of the signature in Austria. Note that Eq. (6) is a general definition of the coefficient of determination and corresponds to the squared Pearson correlation coefficient if \hat{y}_i are estimated through linear least squares regression. When applied to time series (e.g. y_i and \hat{y}_i are observed and estimated runoff at time i) the coefficient of determination calculated with Eq. (6) is known as Nash–Sutcliffe efficiency (Nash and Sutcliffe, 1970; Schaeffli and Gupta, 2007).

Both R^2 and $\widetilde{\text{ANE}}$ measure the performance of the methods. The main differences between these two efficiency measures

are (i) the methods' efficiency increases with increasing R^2 ($R^2 = 1$ means perfect fit) and with decreasing $\widetilde{\text{ANE}}$ (if in at least 50 % of the cases the fit is perfect, then $\widetilde{\text{ANE}} = 0$); (ii) in R^2 the errors are scaled by the spatial variance of the signature, while ANE_i scales the errors locally by the observed value and $\widetilde{\text{ANE}}$ is a measure of the spatial average of the error. This means that small (and therefore good) $\widetilde{\text{ANE}}$ could correspond to small (and therefore bad) R^2 if the spatial variability of the signature is small, but the spatial average is large; (iii) in R^2 the errors are squared, therefore a big weight is given to the largest errors, while in $\widetilde{\text{ANE}}$ the absolute errors are considered and, taking the median, the largest errors have no weight on the measure.

While $\widetilde{\text{ANE}}$ refers to the expected error for the estimation in a particular ungauged catchment (which is of interest for local studies), R^2 measures as well the regional pattern that

is captured by the method (which is of interest for regional studies).

5 Results

5.1 How well can runoff signatures be predicted in Austria?

Figure 5 shows the simulated runoff signatures for the 213 catchments using the process-based method (rainfall–runoff model with parameters regionalised by the similarity method). The spread around the 1 : 1 line is a measure of how well the runoff signatures are estimated in ungauged catchments. For the case of mean annual specific runoff (Fig. 5a) the highest errors (in mm yr^{-1}) tend to occur in the wetter catchments and the model tends to underestimate the mean annual runoff. The coefficient of determination R^2 is 0.86, meaning that the unexplained spatial variance is relatively low. The median absolute normalised error $\widetilde{\text{ANE}}$ is less than 10 %, meaning that, on average, the local error of estimation of mean annual runoff is relatively low. For the range of Pardé coefficients (Fig. 5b), R^2 is lower than in the case of Fig. 5a and bias and spread of the points around the 1 : 1 line are wider, resulting in $\widetilde{\text{NE}} = -7.2\%$ and $\widetilde{\text{ANE}} = 13\%$. A slightly lower performance is obtained for the slope of the flow duration curves (Fig. 5c) for which $R^2 = 0.63$ and bias and average spread of the errors are similar to the ones for the range of Pardé coefficients ($\widetilde{\text{NE}} = -8.1\%$ and $\widetilde{\text{ANE}} = 14\%$). The process-based method tends to underestimate the slope of the flow duration curves likely because an automatic model calibration has been used (Merz et al., 2011), which is more focused on timing of runoff peaks and low flow recession rather than to flows representing the central part of the flow duration curve. Other results would have probably been obtained from other objective functions in the calibration stage (Kollat et al., 2012; Montanari and Toth, 2007; Wagener and Montanari, 2011).

Compared to all other signatures, R^2 is much lower for low flows (Fig. 5d) and high flows (Fig. 5e). Even though R^2 is lower for high flows than for low flows, $\widetilde{\text{ANE}}$ is much lower (the performance is higher) for high flows probably because errors are normalised by the higher observed q_{05} values. Figure 5f shows observed vs. estimated integral scales of runoff time series in log–log scale. The integral scale is significantly overestimated for flashier catchments, i.e. where the observed integral scale is small, and, overall, $\widetilde{\text{NE}}$ and $\widetilde{\text{ANE}}$ have the greatest values encountered so far. However, R^2 is relatively high because the observed spatial range is high and therefore easily captured by the model.

Figure 6 is analogous to Fig. 5 but, here, the statistical method (Top-kriging) is used for regionalisation. For the case of annual specific runoff (Fig. 6a) the method is slightly biased ($\widetilde{\text{NE}} = 3\%$) and the highest errors (in mm yr^{-1}) occur in the wetter catchments. The coefficient of determination

R^2 is 0.88, meaning that the unexplained spatial variance is relatively low. The median absolute normalised error $\widetilde{\text{ANE}}$ is below 10 %, meaning that, on average, the local error of estimation of mean annual runoff is relatively low. For the range of Pardé coefficients (Fig. 6b) the values of R^2 and $\widetilde{\text{ANE}}$ are similar, and also in this case the highest errors occur in wet catchments (blue points) where the method underestimates ΔPar . Similar results are obtained for the slope of the flow duration curves (Fig. 6c) while for low flows (Fig. 6d) R^2 is significantly lower (0.61), there is a positive bias ($\widetilde{\text{NE}} = 6.1\%$), and $\widetilde{\text{ANE}}$ is significantly higher (greater than 10 %). This means that the unexplained spatial variance of q_{95} is relatively high and that the percentage error one makes for individual estimations (relative to the observed q_{95}) is on average also high. Also for high flows (Fig. 6e) R^2 is low, but $\widetilde{\text{ANE}}$ is much lower than for low flows because there is little bias ($\widetilde{\text{NE}} = -1.1\%$) and the errors are normalised by the higher observed q_{05} values. For the integral scale (Fig. 6f), the estimation is unbiased but the spread around the line is quite large (and therefore $\widetilde{\text{ANE}}$ is high). R^2 is relatively high because of the large observed spatial range. The main difference between the two regionalisation methods is that the process-based method produces more biased estimates than the statistical method and the scatter is larger for most of the signatures, which is evident from visual inspection of Figs. 5 and 6.

5.2 In what way does the predictive performance depend on climate and catchment characteristics?

Table 2 reports the Spearman correlation coefficients (see e.g., Kottek and Rosso, 1997, p. 281) between the absolute normalised error and several catchment attributes of the 213 Austrian catchments for each runoff signature and for the two methods used. Through this analysis the “dependence” of predictions on climate and catchment characteristics is meant not necessarily as causality but as correlation. The correlations that are significant at 5 % significance level are indicated in bold.

High correlations are obtained with catchment area. Figure 7 shows the absolute normalised error ANE for the 213 catchments plotted vs. catchment area for the signatures regionalised using the process-based model. Each point corresponds to a catchment. The black line represents the moving window median ANE (considering 10 neighbouring catchments in terms of area) and the grey shading its moving window’s 25 and 75 % quantiles, all smoothed through a cubic smoothing spline (function “smooth.spline” in R; R Core Team, 2012). The increase of performance with area is clear for high flows (Fig. 7e) and particularly for low flows (Fig. 7d), consistently with Table 2. An increase of performance can be noticed for the integral scale as well, even though the errors are much more scattered. For mean annual runoff, the range of the Pardé coefficients and the slope of the

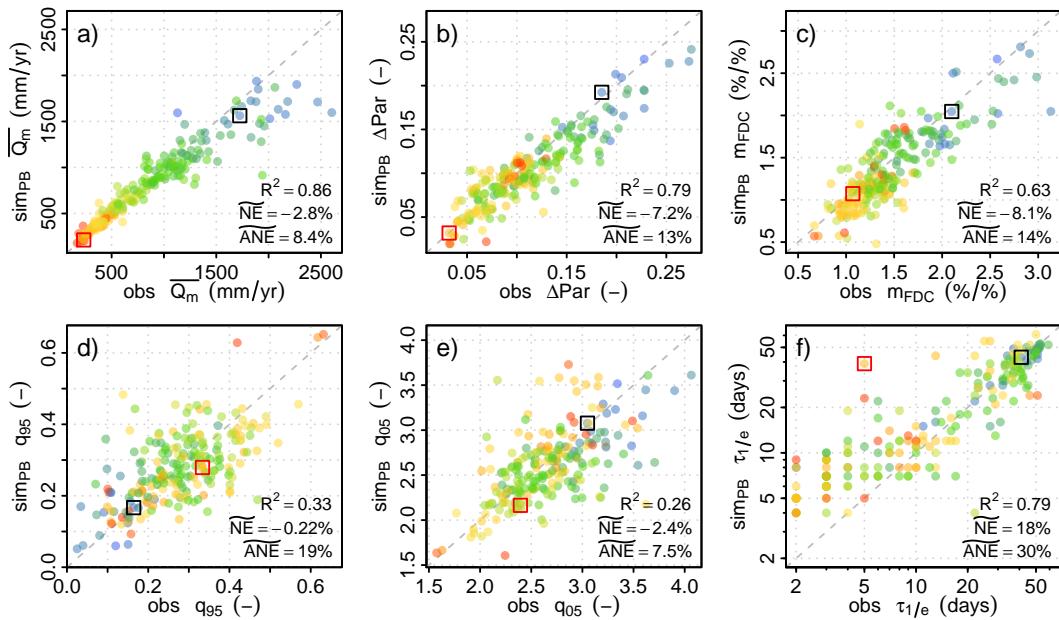


Fig. 5. Observed vs. simulated signatures using the process-based (PB) method in cross-validation mode for the period 1976–2008: (a) mean annual specific runoff (mm yr^{-1}), (b) range of the Pardé coefficient (-), (c) slope of the normalised flow duration curve (%/%), (d) normalised flow duration curve value which is exceeded 95 % of the time (-), (e) normalised flow duration curve value which is exceeded 5 % of the time (-), (f) integral scale (days) in log–log scale. The colour of the points indicates the catchment aridity (blue-wet vs. red-dry) as in Fig. 2. The coefficient of determination R^2 , the median normalised error \widetilde{NE} and the median absolute normalised error \widetilde{ANE} (as percentages) are given. The two catchments in Fig. 1 are indicated by the black and red boxes.

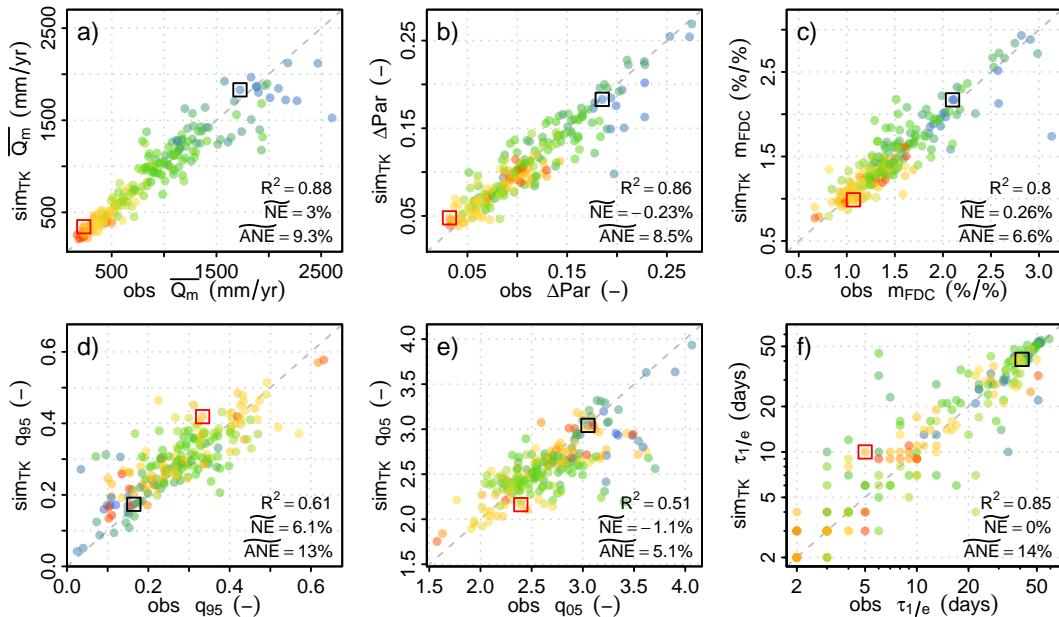


Fig. 6. Observed vs. simulated signatures using Top-kriging (TK) in cross-validation mode for the period 1976–2008: (a) mean annual specific runoff (mm yr^{-1}), (b) range of the Pardé coefficient (-), (c) slope of the normalised flow duration curve (%/%), (d) normalised flow duration curve value which is exceeded 95 % of the time (-), (e) normalised flow duration curve value which is exceeded 5 % of the time (-), (f) integral scale (days) in log–log scale. The colour of the points indicates the catchment aridity (blue-wet vs. red-dry) as in Fig. 2. The coefficient of determination R^2 , the median normalised error \widetilde{NE} and the median absolute normalised error \widetilde{ANE} (as percentages) are given. The two catchments in Fig. 1 are indicated by the black and red boxes.

Table 2. Spearman correlation coefficient between the absolute normalised error and catchment attributes for the 213 Austrian catchments. In bold are marked the significant correlations at 5 % significance level (two sided test with the function “cor.test” in R; R Core Team, 2012). The process-based (PB) method and the Top-kriging (TK) method results are printed on the same columns separated by a slash.

	Area PB/TK	Median elev. PB/TK	Mean slope PB/TK	Network dens. PB/TK	Topo. index PB/TK	Mean ann. prec. PB/TK	Aridity index PB/TK
(a) Mean ann. runoff \bar{Q}_m	-0.09/-0.08	0.18 /0.03	0.21 /0.09	-0.26/-0.15	-0.21/-0.05	0.20 /0.05	-0.21/0.00
(b) Range of Pardé coeff. Δ_{Par}	-0.02/ -0.20	-0.15 /-0.11	-0.13/-0.05	-0.12/0.04	0.13/0.09	0.05/0.10	0.07/0.01
(c) Slope of FDC m_{FDC}	0.02/-0.11	-0.13/-0.06	-0.14/-0.07	-0.03/0.03	0.12/0.11	-0.05/0.01	0.12/0.05
(d) Normalised low flow q_{95}	-0.29 / -0.39	0.00/-0.09	0.00/-0.05	-0.13/-0.02	-0.07/0.01	0.09/0.05	-0.07/-0.02
(e) Normalised high flow q_{05}	-0.23 / -0.30	-0.20 /-0.13	-0.15/-0.05	-0.05/-0.03	0.11/0.05	0.06/0.12	0.04/-0.01
(f) Integral scale $\tau_{1/e}$	-0.16/-0.11	-0.64 /-0.09	-0.48 /-0.1	0.32 /0.01	0.40 / 0.18	-0.03/0.11	0.38 /0.03

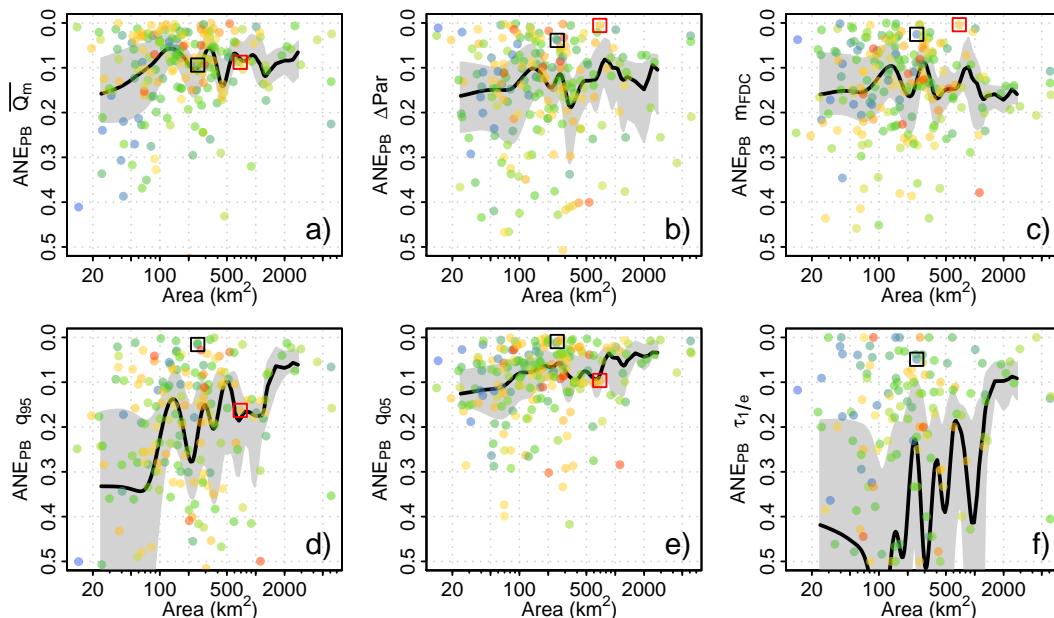


Fig. 7. Absolute normalised error vs. catchment area (km^2) for the signatures regionalised using the process-based (PB) model in cross-validation mode: (a) mean annual specific runoff (mm yr^{-1}), (b) range of the Pardé coefficient (–), (c) slope of the normalised flow duration curve (%/%), (d) normalised flow duration curve value which is exceeded 95 % of the time (–), (e) normalised flow duration curve value which is exceeded 5 % of the time (–), (f) integral scale (days) in log–log scale. The colour of the points indicates the catchment aridity (blue-wet vs. red-dry) as in Fig. 2. The two catchments in Fig. 1 are indicated by the black and red boxes. See text for further details.

flow duration curves, instead, there is no evident relationship of the estimation performance with catchment area (Fig. 7a–c). Figure 8 is analogous to Fig. 7 but for Top-kriging. Also in this case for most signatures the performance increases for increasing catchment area.

The second column of Table 2 shows the correlation between the absolute normalised error for the 213 catchments and the median catchment elevation. For most signatures the Spearman correlation coefficient is negative (significantly for the case of process based method) meaning that the error decreases with increasing catchment elevation. For the mean annual runoff prediction with the rainfall–runoff model, instead, ANE increases with catchment elevation.

The following three columns report the correlation of regionalisation performances to other average catchment at-

tributes: to the mean catchment slope; to the network density (from the digital river network map at the 1 : 50 000 scale, as in Merz and Blöschl, 2004); and to the mean topography index (calculated as λ in Beven and Kirkby, 1979, p. 48). Table 2 indicates little dependence between performance and these characteristics, with the exception of the mean annual runoff and the integral scale for the process based method. For mean slope and topography index, this is expected because the first is positively correlated with catchment elevation and the second is negatively correlated with the slope. This is reflected in the values of the Spearman correlation coefficients in Table 2. For the network density, instead, one reason for the higher errors in predicting annual runoff (and therefore runoff volumes) in low network density catchments

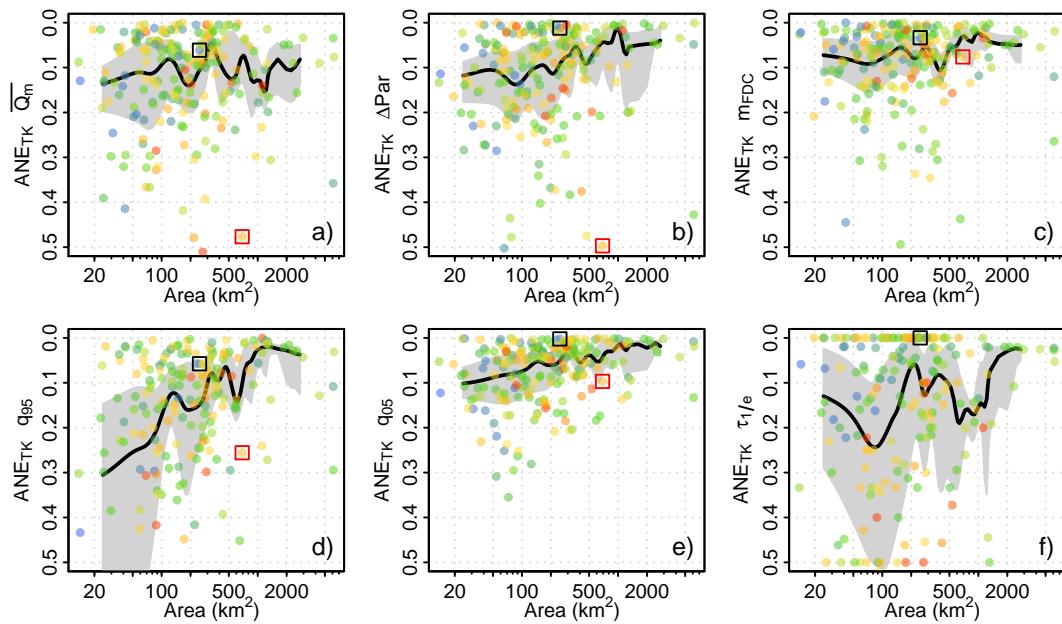


Fig. 8. Absolute normalised error vs. catchment area (km^2) for the signatures regionalised using Top-kriging (TK) in cross-validation mode: (a) mean annual specific runoff (mm yr^{-1}), (b) range of the Pardé coefficient (–), (c) slope of the normalised flow duration curve (%/%), (d) normalised flow duration curve value which is exceeded 95 % of the time (–), (e) normalised flow duration curve value which is exceeded 5 % of the time (–), (f) integral scale (days) in log–log scale. The colour of the points indicates the catchment aridity (blue-wet vs. red-dry) as in Fig. 2. The two catchments in Fig. 1 are indicated by the black and red boxes. See text for further details.

may be the fact that in many cases their geology is partly karstic, which is notably hard to model.

Table 2 indicates very little dependence between performance and mean annual precipitation for both methods. The rainfall–runoff model performance in predicting annual runoff decreases with mean annual precipitation, which is highly correlated with elevation in Austria. Similarly, for the mean annual runoff the performances increase for increasing aridity index when the process-based model is used. Regarding the other signatures, there is hardly any dependence on aridity for any of the signatures. Only the performance of estimation of the integral scale significantly decreases with increasing aridity when using the process-based model.

5.3 What is the relative performance of the predictions of different signatures?

Figure 9 shows a comparative summary of the results from Sects. 5.1 and 5.2. Figure 9a and b show the performances of the process-based (fuchsia) and statistical (beige) methods in terms of normalised error (NE) and absolute normalised error (ANE), respectively. The bars contain the interquartile range (i.e. 50 %) of the values of NE and ANE while the lines connect the median values $\widetilde{\text{NE}}$ and $\widetilde{\text{ANE}}$. These two graphs show that, both in terms of bias (NE) and error spread (ANE), the statistical method (Top-kriging) outperforms the regionalised rainfall–runoff model for essentially all signatures. In particular, with the exception of the low flows, the overall

spatial biases (i.e. $\widetilde{\text{NE}}$) are very close to zero for the statistical method (Fig. 9a), which indeed is optimised in a way to minimise biases (which still remain, since the performances are calculated in cross-validation mode). When the runoff signatures are compared among themselves, one sees that the lowest performances in local prediction are obtained for the integral scale and the low flow statistic q_{95} . Quite surprisingly, the highest performance is obtained for the high flow statistic q_{05} . Since the errors are normalised by the observed values, which are high, ANEs for high flows are much lower than, for example, for low flows.

Figure 9c shows the R^2 for the six signatures regionalised through the process-based model (fuchsia lines and points) and Top-kriging (beige lines and points). Also Fig. 9c shows that Top-kriging generally outperforms the regionalised rainfall–runoff model in estimating the signatures in ungauged basins. The figure indicates that the performance in terms of R^2 , i.e. the ability of the methods to explain the spatial variability of the signatures, is best for seasonal runoff, annual runoff and runoff hydrographs, and is poorer for the prediction of low flows and floods. For most of the signatures the relative performance in terms of R^2 is consistent with Fig. 9a and b. In contrast, the extremes have lower R^2 , which is minimum for high flow prediction. The low R^2 for high flows contrasts with Fig. 9a and b, where they were the ones with highest prediction performance. The spatial variability of the integral scale can be predicted with more confidence.

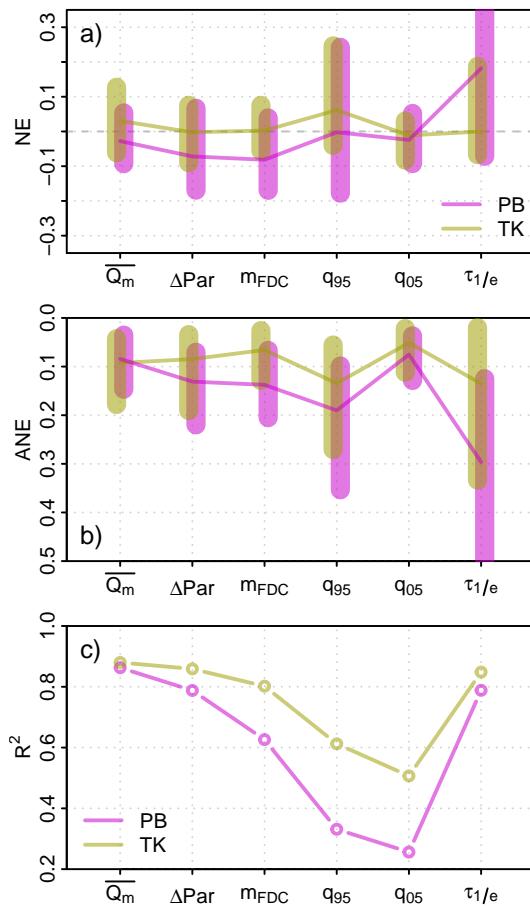


Fig. 9. Comparison of cross-validation performance of prediction methods for different runoff signatures in ungauged basins in Austria. The performance measures are (a) normalised error, (b) absolute normalised error (the bars contain 50 % of the values and the lines connect the medians). (c) coefficient of determination R^2 . The prediction methods are (fuchsia) process-based (PB) method – conceptual rainfall–runoff model whose parameters are regionalised. (beige) statistical method – Top-kriging (TK). The signatures are \bar{Q}_m – mean annual specific runoff (mm yr^{-1}), ΔPar – range of the Pardé coefficient (–), m_{FDC} – slope of the normalised flow duration curve (%/%), q_{95} – normalised flow duration curve value which is exceeded 95 % of the time (–), q_{05} – normalised flow duration curve value which is exceeded 5 % of the time (–), $\tau_{1/e}$ – integral scale (days).

6 Discussion and conclusions

An assessment of the performance of predicting six runoff signatures in ungauged basins has been conducted using two methods for hydrograph regionalisation: a statistical approach (Skøien et al., 2006, Top-kriging) and a regionalised rainfall–runoff model (Parajka et al., 2007b). The assessment has been performed in cross-validation mode for 213 catchments in Austria, representative of the hydrologic diversity in the country. The results show that, on average, the biases are small (< 10 % for most of the signatures), but not neg-

ligible, when the process-based method is used, while they are very close to 0 % when the geostatistical method is used. This is because Top-kriging is an unbiased estimator while the rainfall–runoff model involves biases in the input variables (precipitation and temperature) on top of biases due to model structure and regionalised parameters (Montanari, 2011). The average error spread is lower than 10 % of the observed values for the statistical regionalisation method while it is somewhat higher for the process-based method. The better performance of Top-kriging in Austria is due to a number of reasons. First, the stream gauge density of the study region is quite high, so there is a lot of runoff information available for Top-kriging which uses correlations along the stream network (Laaha et al., 2013). In countries where runoff measurements are more sparse, process-based methods or other statistical methods based on catchment attributes may perform relatively better than geostatistical methods based on spatial proximity. Second, Top-kriging avoids the use of uncertain input variables such as precipitation and potential evaporation. Third, Top-kriging is a linear estimator so it may avoid some of the issues with model structure and parameter identifiability associated with rainfall runoff models (Beven, 1993). However, geostatistical methods such as Top-kriging cannot be used for forecasts in time and/or assessment of changes in the catchment which is one of the main applications of rainfall–runoff models (Blöschl and Montanari, 2010).

The predictive performance in ungauged basins is correlated with a number of climate and catchment characteristics. The predictive performance increases with increasing catchment area for most of the signatures significantly. The dependence of the performance on catchment area may be due to two reasons. First, larger catchments tend to contain a large number of data points (both runoff and rainfall), so more information is available for the predictions. Second, catchment area is a key variable in the aggregation behaviour of rainfall–runoff generation processes (Blöschl et al., 1995; Robinson et al., 1995; Viglione et al., 2010a,b). As the catchment size increases some of the hydrological variability is averaged out due to an interplay of space–time scale processes, thus improving hydrological simulation (see e.g. Sivapalan, 2003; Skøien et al., 2003). These two effects are consistent with the scale effects of performance of rainfall–runoff models in gauged catchments (see e.g. Merz et al., 2009; Nester et al., 2011). They are also consistent with the findings in Parajka et al. (2013) and Salinas et al. (2013) in which the performance of all methods increases with catchment area (particularly for hydrograph and flood predictions).

Interestingly, Top-kriging performance is significantly correlated with catchment area only, which is consistent with the findings in Laaha et al. (2013), i.e. that Top-kriging performs much better for locations with upstream data points. For the rainfall–runoff model the regionalisation performance tends to increase with elevation. This may be due to snow processes in the mountainous catchments, which are

easier to predict because runoff variability is more deterministic being temperature driven (see also Parajka et al., 2005, 2013). This is the case for all signatures with the exception of the mean annual runoff. As the signature is the only one related to the volume of water, a possible explanation could be the sparseness of precipitation stations at high elevations and their undercatch (Frei and Schär, 1998; Bartolini et al., 2011).

Parajka et al. (2013) and Salinas et al. (2013) found a clear pattern of decreasing performance of predicting signatures with aridity from a synthesis of many studies around the world. One would expect that, as the climate gets more arid, the runoff processes tend to become more non-linear (Atkinson et al., 2002; Farmer et al., 2003). Runoff processes in arid climates therefore tend to be spatially more heterogeneous than in humid or cold climates. Similarly the temporal dynamics of runoff tend to be more episodic in arid climates. The relatively larger space–time variability results in lower predictability of runoff in ungauged basins in arid catchments around the world (Parajka et al., 2013; Salinas et al., 2013). This does not appear to be the case in Austria since none of the catchments are really arid (i.e. the aridity index is never greater than unity), while in the studies of Parajka et al. (2013) and Salinas et al. (2013) the aridity index may be as large as 3.

Annual and seasonal runoff can be predicted more accurately than all other signatures. This is likely because of the aggregation of runoff variation over a relatively long time period. They therefore vary more smoothly in space, which enhances their predictability. The spatial variability of high flows and low flows in Austria are harder to predict than the spatial variability of the other signatures. This is likely because they are extremes, so their spatial patterns may involve a lot of small-scale heterogeneity as a result of small-scale variation of precipitation and soil/land use characteristics. The spatial variability of low flows is slightly easier to predict than that of high flows. One reason could be that the processes associated with low flows (in particular climate, longer timescale dry spells) vary more smoothly in space than do the processes associate with high flows and floods. The local relative error of estimation, instead, is higher for low flows than for high flows. Also, extremes are harder to estimate with process-based methods than with statistical methods. This is reflected in the fact that all studies reviewed in Salinas et al. (2013) on regionalisation of extremes use statistical methods.

The distinction between the different methods of predicting flood and low flow behaviour highlights the important point that improved hydrograph fitting should not be the ultimate goal of predictions in ungauged basins. Instead, methods must be optimised to predict specific signatures and their characteristics (Yadav et al., 2007; Hingray et al., 2010; Singh et al., 2011; Euser et al., 2013). In the Austrian example, the targeted method for low-flow estimations (see e.g. Laaha and Blöschl, 2007) gives significantly better perfor-

mances (e.g. $R^2 = 0.75$) than those from the regionalised hydrographs ($R^2 = 0.68$ with Top-kriging) even though the hydrographs used to estimate these flows have a median regional Nash–Sutcliffe efficiency of 0.87. A detailed comparative approach focused on understanding individual signatures and how they are connected may provide more insights and eventually lead to better predictions than solely focusing on reproducing the full hydrograph. This connectivity is underpinned by the driving processes. The fact that multiple runoff signatures in Austria respond to individual process controls illustrates the complex connectivity between process and response. This Austrian example illustrates the explanatory value of comparative hydrology (Falkenmark and Chapman, 1989) across processes (through the connection among signatures), across places (the different catchments/the regions of Austria) and across scales (small and large rivers, see Blöschl, 2006).

Acknowledgements. Financial support for the project “Mountain floods – regional joint probability estimation of extreme events” funded by the Austrian Academy of Sciences, the Doctoral program DK-plus W1219-N22 funded by the Austrian Science Funds and the “FloodChange” project funded by the European Research Council are acknowledged. We would also like to thank the Austria Science Funds (project P 23723-N21) and the Austrian Climate and Energy Fund (Project No. K10AC0K00003, CILFAD) for financial support.

Edited by: H. H. G. Savenije

References

- Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.-H., and Valéry, A.: HESS Opinions “Crash tests for a standardized evaluation of hydrological models”, *Hydrol. Earth Syst. Sci.*, 13, 1757–1764, doi:10.5194/hess-13-1757-2009, 2009.
- Atkinson, S., Woods, R. A., and Sivapalan, M.: Climate and landscape controls on water balance model complexity over changing timescales, *Water Resour. Res.*, 38, 1314, doi:10.1029/2002WR001487, 2002.
- Bartolini, E., Allamano, P., Laio, F., and Claps, P.: Runoff regime estimation at high-elevation sites: a parsimonious water balance approach, *Hydrol. Earth Syst. Sci.*, 15, 1661–1673, doi:10.5194/hess-15-1661-2011, 2011.
- Bergström, S.: The HBV model, in: *Computer Models of Watershed Hydrology*, edited by: Singh, V. P., Water Resources Publications, 443–476, Highlands Ranch, Colorado, 1995.
- Beven, K. J.: Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, 16, 41–51, doi:10.1016/0309-1708(93)90028-E, 1993.
- Beven, K. J. and Kirkby, M. J.: A physically-based, variable contributing area model of basin hydrology, *Hydrolog. Sci. Bull.*, 24, 43–69, doi:10.1080/02626667909491834, 1979.
- Blöschl, G.: Hydrologic synthesis: Across processes, places, and scales, *Water Resour. Res.*, 42, W03S02, doi:10.1029/2005WR004319, 2006.

- Blöschl, G. and Montanari, A.: Climate change impacts – throwing the dice?, *Hydrolog. Process.*, 24, 374–381, doi:10.1002/hyp.7574, 2010.
- Blöschl, G. and Sivapalan, M.: Scale issues in Hydrological Modeling – a review, *Hydrolog. Process.*, 9, 251–290, doi:10.1002/hyp.3360090305, <http://www3.interscience.wiley.com/journal/112736521/abstract>, 1995.
- Blöschl, G., Grayson, R. B., and Sivapalan, M.: On the representative elementary area (REA) concept and its utility for distributed rainfall-runoff modelling, *Hydrolog. Process.*, 9, 313–330, 1995.
- Blöschl, G., Reszler, C., and Komma, J.: A spatially distributed flash flood forecasting model, *Environ. Modell. Softw.*, 23, 464–478, 2008.
- Efron, B. and Gong, G.: A leisurely look at the bootstrap, the jackknife, and cross-validation, *The American Statistician*, 37, 36–48, 1983.
- Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, *Hydrol. Earth Syst. Sci.*, 17, 1893–1912, doi:10.5194/hess-17-1893-2013, 2013.
- Falkenmark, M. and Chapman, T.: Comparative hydrology: An ecological approach to land and water resources, The Unesco Press, Paris, 1989.
- Farmer, D., Sivapalan, M., and Jothityangkoon, C.: Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: Downward approach to water balance analysis, *Water Resour. Res.*, 39, 1035, doi:10.1029/2001WR000328, 2003.
- Fenicia, F., Kavetski, D., and Savenije, H. H.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, *Water Resour. Res.*, 47, 13, doi:10.1029/2010WR010174, 2011.
- Frei, C. and Schär, C.: A precipitation climatology of the Alps from high-resolution rain-gauge observations, *Int. J. Climatol.*, 18, 873–900, doi:10.1002/(SICI)1097-0088(19980630)18:8, 1998.
- Gáál, L., Szolgay, J., Kohnová, S., Parajka, J., Merz, R., Viglione, A., and Blöschl, G.: Flood timescales: Understanding the interplay of climate and catchment processes through comparative hydrology, *Water Resour. Res.*, 48, W04511, doi:10.1029/2011WR011509, 2012.
- Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: An approach to identify time consistent model parameters: sub-period calibration, *Hydrol. Earth Syst. Sci.*, 17, 149–161, doi:10.5194/hess-17-149-2013, 2013.
- Gottschalk, L.: Interpolation of runoff applying objective methods, *Stoch. Hydraul. Hydraul.*, 7, 269–281, doi:10.1007/BF01581615, 1993.
- Gottschalk, L., Krasovskaja, I., Leblois, E., and Sauquet, E.: Mapping mean and variance of runoff in a river basin, *Hydrol. Earth Syst. Sci.*, 10, 469–484, doi:10.5194/hess-10-469-2006, 2006.
- Grayson, R. B. and Blöschl, G. (Eds.): Spatial Patterns in Catchment Hydrology: Observation and Modelling, Cambridge University Press, Cambridge, 404 pp., ISBN:0-521-63316-8, 2000.
- Gutknecht, D., Reszler, C., and Blöschl, G.: Das Katastrophenhochwasser vom 7. August 2002 am Kamp – eine erste Einschätzung (The August 7, 2002 – flood of the Kamp – a first assessment), *Elektrotechnik und Informationstechnik*, 119, 411–413, 2002.
- Hannah, D. M., Demuth, S., van Lanen, H. A. J., Loosser, U., Prudhomme, C., Rees, G., Stahl, K., and Tallaksen, L. M.: Large-scale river flow archives: importance, current status and future needs, *Hydrolog. Process.*, 25, 1191–1200, doi:10.1002/hyp.7794, 2011.
- Hingray, B., Schaefli, B., Mezghani, A., and Hamdi, Y.: Signature-based model calibration for hydrological prediction in mesoscale Alpine catchments, *Hydrolog. Sci. J.*, 55, 1002–1016, doi:10.1080/02626667.2010.505572, 2010.
- Hrachowitz, M., Savenije, H. H., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R., Montanari, A., Pande, S., Tetzelaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB) – a review, *Hydrolog. Sci. J.*, doi:10.1080/02626667.2013.803183, in press, 2013.
- Jothityangkoon, C., Sivapalan, M., and Farmer, D.: Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development, *J. Hydrol.*, 254, 174–198, 2001.
- Klemeš, V.: Operational testing of hydrological simulation models, *Hydrolog. Sci. J.*, 31, 13–24, doi:10.1080/02626668609491024, 1986.
- Kollat, J., Reed, P. M., and Wagener, T.: When are multiobjective calibration trade-offs in hydrologic models meaningful?, *Water Resour. Res.*, 48, 19, doi:10.1029/2011WR011534, 2012.
- Komma, J., Reszler, C., Blöschl, G., and Haiden, T.: Ensemble prediction of floods – catchment non-linearity and forecast probabilities, *Nat. Hazards Earth Syst. Sci.*, 7, 431–444, doi:10.5194/nhess-7-431-2007, 2007.
- Kottegoda, N. T. and Rosso, R.: Statistics, Probability, and Reliability for Civil and Environmental Engineers, McGraw-Hill Companies, International Edn., 1997.
- Laaha, G. and Blöschl, G.: A national low flow estimation procedure for Austria, *Hydrolog. Sci. J.*, 52, 625–644, 2007.
- Laaha, G., Skøien, J. O., and Blöschl, G.: Spatial prediction on river networks: comparison of top-kriging with regional regression, *Hydrolog. Process.*, doi:10.1002/hyp.9578, in press, 2013.
- McMahon, T. A., Peel, M. C., Pegram, G. G. S., and Smith, I. N.: A Simple Methodology for Estimating Mean and Variability of Annual Runoff and Reservoir Yield under Present and Future Climates, *J. Hydrometeorol.*, 12, 135–146, doi:10.1175/2010JHM1288.1, 2011.
- Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, *J. Hydrol.*, 287, 95–123, 2004.
- Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, *Water Resour. Res.*, 44, W08432, doi:10.1029/2007WR006744, 2008a.
- Merz, R. and Blöschl, G.: Flood frequency hydrology: 2. Combining data evidence, *Water Resour. Res.*, 44, W08433, doi:10.1029/2007WR006745, 2008b.
- Merz, R., Blöschl, G., and Humer, G.: National flood discharge mapping in Austria, *Nat. Hazards*, 46, 53–72, doi:10.1007/s11069-007-9181-7, 2008.
- Merz, R., Parajka, J., and Blöschl, G.: Scale effects in conceptual hydrological modeling, *Water Resour. Res.*, 45, W09405, doi:10.1029/2009WR007872, 2009.

- Merz, R., Parajka, J., and Blöschl, G.: Time stability of catchment model parameters: Implications for climate impact analyses, *Water Resour. Res.*, 47, W02531, doi:10.1029/2010WR009505, 2011.
- Mészároš, I., Miklánek, P., and Parajka, J.: Solar energy income modelling in mountainous areas, in: ERB and NEFRIEND Proj.5 Conf. Interdisciplinary Approaches in Small Catchment Hydrology: Monitoring and Research, edited by: Holko, L., Miklánek, P., Parajka, J., and Kostka, Z., 127–135, Slovak NC IHP UNESCO/UH SAV, Bratislava, Slovakia, 2002.
- Montanari, A.: What do we mean by “uncertainty”? The need for a consistent wording about uncertainty assessment in hydrology, *Hydrolog. Process.*, 21, 841–845, doi:10.1002/hyp.6623, 2007.
- Montanari, A.: Uncertainty of hydrological predictions, in: Treatise on Water Science, edited by: Wilderer, P., 2, Chap. 2.17, 459–478, Elsevier, 2011.
- Montanari, A. and Toth, E.: Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, *Water Resour. Res.*, 43, 10, doi:10.1029/2006WR005184, 2007.
- Mudelsee, M., Börngen, M., Tetzlaff, G., and Grünwald, U.: Extreme floods in central Europe over the past 500 years: Role of cyclone pathway “Zugstrasse Vb”, *J. Geophys. Res.-Atmos.*, 109, D23101, doi:10.1029/2004JD005034, 2004.
- Nash, J. and Sutcliffe, J.: 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, 1970.
- Nester, T., Kirnbauer, R., Gutknecht, D., and Blöschl, G.: Climate and catchment controls on the performance of regional flood simulations, *J. Hydrol.*, 402, 340–356, doi:10.1016/j.jhydrol.2011.03.028, 2011.
- Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, *Hydrolog. Earth Syst. Sci.*, 9, 157–171, doi:10.5194/hess-9-157-2005, 2005.
- Parajka, J., Blöschl, G., and Merz, R.: Regional calibration of catchment models: Potential for ungauged catchments, *Water Resour. Res.*, 43, W06406, doi:10.1029/2006WR005271, 2007a.
- Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, *Hydrolog. Process.*, 21, 435–446, doi:10.1002/hyp.6253, 2007b.
- Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies, *Hydrolog. Earth Syst. Sci.*, 17, 1783–1795, doi:10.5194/hess-17-1783-2013, 2013.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, available at: <http://www.r-project.org>, ISBN 3-900051-07-0, 2012.
- Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., and DMIP Participants: Overall distributed model intercomparison project results, *J. Hydrol.*, 298, 27–60, doi:10.1016/j.jhydrol.2004.03.031, 2004.
- Reszler, C., Blöschl, G., and Komma, J.: Identifying runoff routing parameters for operational flood forecasting in small to medium sized catchments, *Hydrolog. Sci. J.*, 53, 112–129, 2008.
- Robinson, J., Sivapalan, M., and Snell, J.: On the relative roles of hillslope processes, channel routing, and network geomorphology in the hydrologic response of natural catchments, *Water Resour. Res.*, 31, 3089–3101, 1995.
- Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, *Hydrolog. Earth Syst. Sci. Discuss.*, 10, 411–447, doi:10.5194/hessd-10-411-2013, 2013.
- Sauquet, E., Gottschalk, L., and Leblois, E.: Mapping average annual runoff: a hierarchical approach applying a stochastic interpolation scheme, *Hydrolog. Sci. J.*, 45, 799–815, doi:10.1080/02626660009492385, 2000.
- Sauquet, E., Gottschalk, L., and Krasovskia, I.: Estimating mean monthly runoff at ungauged locations: an application to France, *Hydrolog. Res.*, 39, 403–423, doi:10.2166/nh.2008.331, 2008.
- Schaeffli, B. and Gupta, H. V.: Do Nash values have value?, *Hydrolog. Process.*, 21, 2075–2080, doi:10.1002/hyp.6825, 2007.
- Singh, R., Wagener, T., van Werkhoven, K., Mann, M. E., and Crane, R.: A trading-space-for-time approach to probabilistic continuous streamflow predictions in a changing climate – accounting for changing watershed behavior, *Hydrolog. Earth Syst. Sci.*, 15, 3591–3603, doi:10.5194/hess-15-3591-2011, 2011.
- Singh, V. and Frevert, D.: Watershed models, Taylor & Francis, 2005.
- Sivapalan, M.: Process complexity at hillslope scale, process simplicity at the watershed scale: is there a connection?, *Hydrolog. Process.*, 17, 1037–1041, doi:10.1002/hyp.5109, 2003.
- Sivapalan, M.: Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, Chap. 13, 193–220, Encyclopedia of Hydrological Sciences, John Wiley & Sons, Chichester, available at: <http://www.mrw.interscience.wiley.com/emrw/9780470848944/ehs/article/hsa012/current/html>, 2005.
- Sivapalan, M., Takeuchi, K., Franks, S., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E., O’Connell, P., Oki, T., Pomeroy, J., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrolog. Sci. J.*, 48, 857–880, doi:10.1623/hysj.48.6.857.51421, 2003.
- Skøien, J. O. and Blöschl, G.: Sampling scale effects in random fields and implications for environmental monitoring, *Environ. Modell. Assess.*, 114, 521–552, doi:10.1007/s10661-006-4939-z, 2006a.
- Skøien, J. O. and Blöschl, G.: Scale effects in estimating the variogram and implications for soil hydrology, *Vadose Zone J.*, 5, 153–167, 2006b.
- Skøien, J. O. and Blöschl, G.: Spatiotemporal topological kriging of runoff time series, *Water Resour. Res.*, 43, W09419, doi:10.1029/2006WR005760, 2007.
- Skøien, J. O., Blöschl, G., and Western, A.: Characteristic space scales and timescales in hydrology, *Water Resour. Res.*, 39, 1304, doi:10.1029/2002WR001736, 2003.
- Skøien, J. O., Merz, R., and Blöschl, G.: Top-kriging – geostatistics on stream networks, *Hydrolog. Earth Syst. Sci.*, 10, 277–287, doi:10.5194/hess-10-277-2006, 2006.
- Smakhtin, V.: Low flow hydrology: a review, *J. Hydrol.*, 240, 147–186, doi:10.1016/S0022-1694(00)00340-1, 2001.
- Smith, M. B., Seo, D.-J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F., and Cong, S.: The distributed model intercomparison project (DMIP): motivation and experiment design, *J. Hydrol. Earth Syst. Sci.*, 17, 2263–2279, 2013.

- Hydrol., 298, 4–26, doi:10.1016/j.jhydrol.2004.03.040, 2004.
- Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., and Pinto, J. G.: The central European floods of August 2002: Part 2 – Synoptic causes and considerations with respect to climatic change, *Weather*, 58, 434–442, doi:10.1256/wea.61.03B, 2003.
- Viglione, A., Chirico, G. B., Komma, J., Woods, R. A., Borga, M., and Blöschl, G.: Quantifying space-time dynamics of flood event types, *J. Hydrol.*, 394, 213–229, doi:10.1016/j.jhydrol.2010.05.041, 2010a.
- Viglione, A., Chirico, G. B., Woods, R. A., and Blöschl, G.: Generalised synthesis of space-time variability in flood response: An analytical framework, *J. Hydrol.*, 394, 198–212, doi:10.1016/j.jhydrol.2010.05.047, 2010b.
- Viglione, A., Merz, R., Salinas, J. L., and Blöschl, G.: Flood frequency hydrology: 3. A Bayesian analysis, *Water Resour. Res.*, 49, 1–18, doi:10.1029/2011WR010782, 2013.
- Vogel, R. M. and Fennessey, N. M.: Flow-duration Curves. I: New Interpretation and Confidence Intervals, *J. Water Resour. Plan. Manage.-ASCE*, 120, 485–504, 1994.
- Vogel, R. M. and Fennessey, N. M.: Flow Duration Curves II: A Review of Applications in Water Resources Planning, *J. Am. Water Resour. As.*, 31, 1029–1039, doi:10.1111/j.1752-1688.1995.tb03419.x, 1995.
- Wagener, T. and Montanari, A.: Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, *Water Resour. Res.*, 47, W06301, doi:10.1029/2010WR009469, 2011.
- Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment Classification and Hydrologic Similarity, *Geogr. Compass*, 1, 901–931, doi:10.1111/j.1749-8198.2007.00039.x, 2007.
- Yadav, M., Wagener, T., and Gupta, H. V.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. Water Resour.*, 30, 1756–1774, doi:10.1016/j.advwatres.2007.01.005, 2007.