Hydrology and Earth System Sciences Discussions

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

A. Viglione[1], J. Parajka[1], M. Rogger[1], J. L. Salinas[1], G. Laaha[2], M. Sivapalan[3], and G. Blöschl[1]

[1]Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria
[2]Institute of Applied Statistics and Computing, University of Natural Resources and Life Sciences, BOKU Vienna, Austria
[3]Departments of Civil and Envionmental Engineering and Geography, University of Illinois at Urbana-Champaign, Urbana, USA

HESSD

10, 449–485, 2013

Part 3: Runoff signatures in Austria

A. Viglione et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

In a three-part paper we assess the performance of runoff predictions in ungauged basins in a comparative way. While Parajka et al. (2013) and Salinas et al. (2013) assess the regionalisation of hydrographs and hydrological extremes through a literature review, in this paper we assess prediction of a range of runoff signatures for a consistent dataset. Daily runoff time series are predicted for 213 catchments in Austria by a regionalised rainfall–runoff model and by Top-Kriging, a geostatistical interpolation method that accounts for the river network hierarchy. From the runoff timeseries, six runoff signatures are extracted: annual runoff, seasonal runoff, flow duration curves, low flows, high flows and runoff hydrograph. The predictive performance is assessed by the bias, error spread and proportion of unexplained spatial variance of statistical measures of these signatures in cross-validation mode. Results of the comparative assessment show that the geostatistical approach (Top-Kriging) generally outperforms the regionalised rainfall–runoff model. The predictive performance increases with catchment area for both methods and all signatures, 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 is the most difficult to capture followed by the low flows. The relative predictive performance of the signatures depends on the selected performance measures. It is therefore essential to report performance in a consistent way by more than one performance measure.

1 Introduction

The PUB decade, i.e. the 2003–2012 initiative on Runoff Predictions in Ungauged Basins promoted by the International Association of Hydrological Sciences (Sivapalan et al., 2003), has come to an end. Much of the interest during the PUB decade has revolved on estimating runoff at locations where no runoff measurements are available. During the 10 yr, the hydrological community has worked at improving existing

models and at developing innovative approaches to predict runoff in ungauged basins. It is therefore timely to assess the performance of such predictions. This is the motivation for this and the two companion papers of Parajka et al. (2013) and Salinas et al. (2013). In these papers, the predictive performance of the methods is evaluated through leave-one-out cross-validation as a measure of the total predictive uncertainty in runoff prediction in ungauged basins. Parajka et al. (2013) assess the predicting performance of process-based methods (rainfall–runoff models with regionalised parameters) for estimating runoff hydrographs. Salinas et al. (2013) assess the predicting performance of statistical methods used to estimate extremes (low flows and floods). In this paper we consider both statistical and process-based methods for predicting runoff hydrographs and we assess the predicting performance on a range of runoff characteristics at different time scales.

Following Jothityangkoon et al. (2001), the temporal patterns of runoff response of catchments are termed here as runoff "signatures". The runoff signatures considered in this paper are annual runoff, seasonal runoff, flow duration curves, low flows, floods and runoff hydrographs. Each of them represents a set of processes at different time scales. For example, annual runoff is a reflection of the catchment dynamics at relatively long time scales, which is particularly evident in its between-year variability. Seasonal runoff reflects the within-year variability, i.e. how the catchment organises itself at the sub-annual time scale. The flow duration curve represents the full spectrum of variability in terms of their magnitudes. 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 floods are at the opposite end, when there is much water in the system. Hydrographs are the complex combination of all of these signatures. They are the most detailed signatures of how catchments respond to water and energy inputs.

Each signature is meaningful of a certain class of applications of societal relevance. Annual runoff is related to the hydrological problem of how much water is available, which is fundamental for water management purposes such as water allocation, long-term planning, groundwater recharge etc. Seasonal runoff addresses the question of

when water is available throughout the year and is necessary to plan water supply, hydropower production and river restoration measures. The flow duration curve measures for how many days in a year water is available and is the basis of studies on river ecology, hydropower potential, industrial, domestic and irrigation water supply. Low flow statistics are needed to estimate environmental flows for ecological stream health, for drought management, river restoration, dilution of effluents etc. Flood statistics, instead, are required for design of spillways, culverts, dams and levees, for reservoir management, river restoration and risk management. The entire hydrograph can be used for all the applications listed above and is specifically needed when the dynamics of runoff have to be taken into account, such as for water quality studies. Even in highly monitored areas, runoff is only measured at a few river sections. Predicting runoff signatures in ungauged basins is therefore very important for the water resources perspectives listed above.

In this paper we perform a comparative assessment of prediction performance for all six runoff signatures. We use a different approach to the one used in Parajka et al. (2013) and Salinas et al. (2013), who perform comparative assessments based on many studies from all around the world. Their assessments have the advantage of covering a wide range of climates and catchment characteristics, but the disadvantage of comparing different methods applied on different catchments with different data. In this paper, the comparison across signatures is based on one consistent data set (Austria) and two regionalisation methods for the runoff hydrograph (a statistical and a process-based one) from which the signatures are extracted. In particular, the following questions are addressed: (i) how well do we predict runoff signatures in Austria? (ii) In what way do the predictions 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 Connection across signatures

As a starting point of the comparative analysis it is useful to examine the various signatures from a process perspective. The runoff response of a catchment 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, including the effects of the landscape patterns. Since the structure of the landscape determines the heterogeneity and organisation of pathways that water can follow and associated residence times, it also governs the richness of the catchment's hydrologic responses. Runoff variability at any location is a temporal continuum covering a wide range of time scales, but the characteristics one sees depend on the temporal scale one chooses to look at. This is because catchments exhibit the characteristics of complex systems, so different patterns emerge at different time scales. At time scales of seconds one may recognise the effects of turbulence and wave action in the runoff. At time scales of millennia, if such data were available as in the case of Jefferson et al. (2010), one would recognise long-term climate and landscape evolution trends. The runoff signatures considered in this paper are some of these emergent patterns in the time domain and they are all inter-connected because they are all the result of the same complex system and co-evolutionary processes of climate, vegetation, landscape and soils.

Figures 1 and 2 aim to illustrate this inter-connection of runoff signatures in time and space. Figure 1 is an example of quantification (through curves) of the six runoff signatures observed in two quite diverse catchments while Fig. 2 presents snapshots of their spatial patterns across Austria. For instance, the spatial pattern of annual runoff (Fig. 2a) is controlled by the interplay of annual precipitation and evaporation. The largest precipitation rates of more than $2000\,\mathrm{mm\,yr^{-1}}$ occur in the West, mainly due to orographic lifting of north-westerly airflows at the rim of the Alps (see the elevation map in Fig. 3), which causes the highest annual runoff in the West of the country. Precipitation is lowest in 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

Part 3: Runoff signatures in Austria

A. Viglione et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

example catchments of 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. 3 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 South-East of Austria.

The seasonality of runoff, quantified with the Pardé coefficient (see Sect. 4.2 for the definition), is very pronounced in the mountainous catchments (e.g. Fig. 1b) because of snow accumulation and melt processes. Consistently, in the West the maxima occur in summer (Fig. 2b). In the lowlands of the East, the runoff seasonality is the result of the interplay between the seasonality of precipitation and evaporation, which results in a maximum in runoff in spring in the North-East or summer in the South-East (Figs. 1b and 2b).

Snowmelt in the Alpine West leads to flow duration curves that are steep in their central part (Fig. 2c). There are two regions, one in the South-East and one in the West where flow duration curves are particularly flat, which is due to the flashy nature of runoff in these regions. The flashiness is due to both convective precipitation and responsive soils, which are a result of the co-evolution of climate, landscape and soils (Gaál et al., 2012). This suggests a higher variability of the extremes, which is evident in the slope of the red (low-land catchment) flow duration curve at the extremes in Fig. 1c when compared to the black (mountainous catchment) flow duration curve. This is reflected in the steeper frequency distributions of low flows (Fig. 1d) and floods (Fig. 1e) for the low-land catchment. The small low flows in the East occur in summer and are related to the seasonality of 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 (Fig. 1b). The spatial pattern of floods is closely related to the spatial pattern of annual rainfall and therefore annual runoff. This is because of three reasons, the direct rainfall input at the event scale, the antecedent soil moisture, and landform-hydrology feedbacks, which have produced more efficient

drainage networks in high rainfall areas (Merz and Blöschl, 2008a,b; Blöschl and Merz, 2010).

Figure 1f compares the entire hydrographs at daily timestep for six years in the two example catchments. It embeds all the information already discussed for the other signatures and provides additional information. For instance, it is evident that the hydrograph of the lowland Raab at Feldbach (red line) has less memory than the one of the mountainous Lech at Steeg (black line). In Fig. 2f the hydrograph integral scale (see Sect. 4.2 for the definition) is used as a measure of the flashiness of runoff and shows slow dynamics in the snow dominated West and much faster dynamics in the lowlands in the East. A small region in the South has highly pervious rocks, the effect of which is reflected in large integral scales (slow response) as well as by relative large low flows and small floods.

This Austrian example illustrates the explanatory value of comparative hydrology across processes (through the signatures), across places (the different catchments/the regions of Austria) and across scales (small and large rivers, see Blöschl, 2006). Austria is a typical example of how the runoff signatures are connected, but the entire range of connections of signatures across the world is even richer than that. Relative phases of precipitation and soil moisture are extremely important. In winter rainfall climates, where summer rainfall lows are in phase with summer maxima of evaporation, the runoff seasonality can be much stronger with an even stronger effect on all runoff signatures.

3 Regionalisation methods

Regionalisation methods for runoff prediction in ungauged basins belong 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.

3.1 Process-based method

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 (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). 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 model concept is given, e.g. in the Appendix of Parajka et al. (2007). 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 (Merz et al., 2011). For predictions in ungauged site, 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) summarized and compared different approaches used for transfering model parameters to ungauged catchments. In this study, we apply a similarity based approach introduced in Parajka et al. (2005). This regionalisation method is based on idea to find a donor catchment that is most similar to 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.

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–Criddle method (Parajka et al., 2005) using interpolated daily air temperature and grid maps of potential sunshine duration (Mészároš et al., 2002). The model inputs are extracted for 200 m elevation zones and used for runoff model simulations in each catchment. For the rainfall–runoff regionalisation, daily runoff observations from a total of 240 stream gauges are used.

3.2 Statistical method

The main advantage of statistical methods of estimating runoff in ungauged basins is that they avoid the use of uncertain input variables such as precipitation and potential evaporation. In this paper we use Top-Kriging which is a geostatistical method that accounts for the river network hierarchy (Skøien et al., 2006; Merz et al., 2008). The method requires a variogram for local (point) runoff generation which has been taken from Skøien et al. (2006). The variogram is then integrated over the catchment areas associated with each river cross section. The assumptions of a Best Linear Unbiased estimator than give the kriging weights which are used to estimate the daily runoff for an ungauged basins from the observed daily runoff of neighbouring stations on the same day, weighted by the kriging weights. Skøien and Blöschl (2006a,b, 2007) provide discussions of the uncertainties involved. For the Top-Kriging estimation, daily runoff observations from a total of 689 stream gauges are used.

4 Method for the comparative assessment

4.1 Catchments for the blind testing and simulation efficiency

In order to assess the performance of the predictive methods, runoff hydrographs are estimated for a number of catchments without using runoff data from that basin, i.e. the catchments are treated as ungauged. Only after the runoff predictions are made, the runoff data are used for the assessment. This procedure allows for an independent cross-validation of each methodology used to provide predictions in ungauged basins, rather than enabling just a goodness of fit of a particular regionalisation method.

For cross-validation we considered a total of 213 catchments, which are representative of the hydrological variability across Austria and whose stream gauge position is shown in Fig. 3. The colour of the stream gauges in Fig. 3 indicates the aridity index (ratio of mean annual potential evaporation vs. mean annual precipitation). The wetter catchments are in the Alpine area while the dryer ones are in the northern and eastern lowlands, consistent with Figs. 1 and 2. The aridity index varies from 0.2 to 1.0 meaning that there is no really arid catchment in the data set (i.e. potential evaporation is everywhere lower than precipitation). More detailed statistics of the catchment characteristics are reported in Table 2. The dataset used for runoff prediction spans the period from 1976 to 2008 and we use the same period for the assessment. The 213 catchments are a subsample of the catchments used for the regionalisation with the process-based and geostatistical methods. They are those where both methods have been used and exclude catchments with significant anthropogenic effects.

The goodness-of-fit of the rainfall–runoff model simulations in the calibration period 1987–1997 gives a median Nash–Sutcliffe efficiency of 0.72 (see Table 1) for the 213 observed hydrographs. The performance in cross-validation mode is not much worse (median Nash-Sutcliffe of 0.61), even though it includes the uncertainties of the model and of the parameter regionalisation method (Montanari, 2011). For the Top-Kriging interpolation, the cross-validation performance for hydrograph regionalisation is higher,

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

i.e. median Nash-Sutcliffe of 0.87 (see Table 1). However, the focus of the paper is on the signatures derived from these hydrographs.

4.2   Signatures

There are many ways of quantifying each runoff signature. For instance, in Fig. 1 the signatures are quantified by curves and in Fig. 2 they are quantified by specific statistics instead. For the comparative assessment we quantify the signatures by single values. Given the time series of observed (or simulated) specific daily runoff $Q_d(t)$ $(\mathrm{mm\,d^{-1}})$, the following statistics are calculated:

a. the mean annual specific runoff $(\mathrm{mm\,yr^{-1}})$

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

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

b. the range of Pardé coefficients (–)

$$\Delta \mathrm{Par} = \max(\mathrm{Par}) - \min(\mathrm{Par}) \qquad (2)$$

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

$$\mathrm{Par}_i = \frac{\sum_{t \in M_i} Q_d(t)}{\sum_{\forall t} Q_d(t)} \qquad (3)$$

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

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

$$m_{FDC} = 100 \cdot \frac{Q_{30\%} - Q_{70\%}}{40 \cdot \overline{Q_d}} \tag{4}$$

where $Q_{30\%}$ ($\mathrm{mm\,d^{-1}}$) 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 exceedence probability;

d. the normalised low flow statistic (–) calculated as $q_{95} = Q_{95\%}/\overline{Q_d}$ where $Q_{95\%}$ ($\mathrm{mm\,d^{-1}}$) 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 $Q_{5\%}$ ($\mathrm{mm\,d^{-1}}$) 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.

Some statistics of these signatures are listed in Table 2. 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.

4.3 Performance measures

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

1. The normalised error, which is defined as

$$NE_i = \frac{\hat{y}_i - y_i}{y_i} \qquad (5)$$

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 $\widetilde{NE}$ is a measure of (spatial) bias of estimation in Austria. A positive (negative) value of $\widetilde{NE}$ means that, on average, the method overestimates (underestimates) the signature of interest;

2. The absolute normalised error, which is defined as $ANE_i = |NE_i|$ for catchment i. The spatial median $\widetilde{ANE}$ is a measures of the average spread of the estimation error. A low value of $\widetilde{ANE}$ (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);

3. The coefficient of determination, which is defined as

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

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.

Both R^2 and $\widetilde{ANE}$ 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 $\widetilde{ANE}$ (if in at least 50 %

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

of the cases the fit is perfect, then $\widetilde{\mathrm{ANE}} = 0$); (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 $\widetilde{\mathrm{ANE}}$ is a measure of the spatial average of the error. This means that small (and therefore good) $\widetilde{\mathrm{ANE}}$ 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 $\widetilde{\mathrm{ANE}}$ the absolute errors are considered and, taking the median, the largest errors have no weight on the measure.

5   Results

5.1   How well do we predict runoff signatures in Austria?

Figure 4 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. 4a) the highest errors (in $\mathrm{mm\,yr^{-1}}$) tend to occur in the wetter catchments and the model tends to underestimate the mean annual runoff (and actually $\widetilde{\mathrm{NE}}$ is negative). The coefficient of determination R^2 is 0.86, meaning that the unexplained spatial variance is relatively low. The median absolute normalised error $\widetilde{\mathrm{ANE}}$ 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. 4b), R^2 is lower to the case of Fig. 4a and bias and spread of the points around the 1 : 1 line are wider, resulting in $\widetilde{\mathrm{NE}} = -7.2$ % and $\widetilde{\mathrm{ANE}} = 13$ %. A slightly lower performance is obtained for the slope of the flow duration curves (Fig. 4c) 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 ($\widetilde{\mathrm{NE}} = -8.1$ % and $\widetilde{\mathrm{ANE}} = 14$ %). The process-based method tends to underestimate
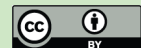
the slope of the flow duration curves likely because an automatic model calibration has been used, more focused on timing of runoff peaks and low flow recession rather than to flows representing the central part of the flow duration curve. Other results would have been obtained from other objective functions in the calibration stage (Kollat et al.,
5 2012; Montanari and Toth, 2007; Wagener and Montanari, 2011).

Compared to all other signatures, R^2 are much lower for low flows (Fig. 4d) and high flows (Fig. 4e). 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 4f shows observed vs. estimated
10 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.
15 Figure 5 is analogous to Fig. 4 but, here, the statistical method (Top-Kriging) is used for regionalisation. For the case of annual specific runoff (Fig. 5a) the method is essentially unbiased (and $\widetilde{\text{NE}}$ is very small) and the highest errors (in mm yr^{-1}) occur in the wetter catchments. The coefficient of determination R^2 is 0.84, meaning that the unexplained spatial variance is relatively low. The median absolute normalised error $\widetilde{\text{ANE}}$
20 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. 5b) 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 (and, indeed, $\widetilde{\text{NE}}$ is slightly negative). Similar results are obtained for the slope of the flow duration curves (Fig. 5c) while for
25 low flows (Fig. 5d) R^2 is significantly lower (0.68), there is a positive bias ($\widetilde{\text{NE}}$ = 7.3 %) 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. The picture for high flows (Fig. 5e) looks similar to the one for low flows. R^2 is low, but $\widetilde{\text{ANE}}$ is

much lower than for low flows because there is little bias ($\widetilde{NE} = -1.9\%$) and the errors are normalised by the higher observed q_{05} values. For the integral scale (Fig. 5f), the spread around the line is quite large (and therefore $\widetilde{ANE}$ is high), but R^2 is relatively high because of the large observed spatial range. The main difference between Fig. 5 and Fig. 4 resulting from a visual inspection is that the process-based method produces more biased estimates than the statistical method and the scatter is larger for some of the signatures.

5.2 In what way do the predictions depend on climate and catchment characteristics?

Table 3 reports the Spearman correlation coefficients between the absolute normalised error and four catchment attributes of the 213 Austrian catchments for each runoff signature and for the two methods used. The correlations that are significant at 5% significance level are indicated in bold. Table 3 examines whether the runoff prediction performances vary depending on the size, elevation, available water and/or aridity of the catchments.

The highest correlations are obtained with catchment area. There are a number of reasons why catchment area is a good predictor. The landscape characteristics and behaviour and dominant processes change with increasing catchment area. For example, headwater catchments tend to be steep with landslides being dominant, whereas flatland catchments, which are larger, tend to be dominated by groundwater aquifers, wide floodplains and frequent inundations etc., and therefore exhibit very different flow paths. Also, catchment area is a key variable in the aggregation behaviour of rainfall runoff generation processes (Blöschl et al., 1995; Viglione et al., 2010a,b). With increasing catchment area new processes take over which may depend on climate. All of these changes or transitions make area a holistic similarity index for the runoff signatures.

Figure 6 shows the absolute normalised error ANE for the 213 catchments plotted vs. catchment area for the signatures regionalised using the process-based model (PB). Each point corresponds to a catchment. The black line represents the moving window median ANE (considering 10 neighboring catchments in terms of area) and the grey shading its moving window 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. 6e) and particularly for low flows (Fig. 6d), consistently with Table 3. 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 flow duration curves, instead, there is no evident relationship of the estimation performance with catchment area (Fig. 6a–c). Figure 7 is analogous to Fig. 6 but for Top-Kriging (TK). For all signatures the performance increases for increasing catchment area. Overall, this very clear pattern of an increase of the performance with catchment scale may be due to two reasons. The first is a trend for an increasing number of nested gauged subcatchments within a catchment as the catchment size increases (and of rain gauges when the rainfall–runoff model is used). The second may be related to the aggregation effect of runoff (Skøien et al., 2003). As the catchment size increases some of the hydrological variability is averaged out due to an interplay of space-time scale processes that will improve hydrological simulation. Both 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).

The second column of Table 3 shows the correlation between the absolute normalised error for the 213 catchments and the median catchment elevation. Topographic elevation, averaged over the catchment, is a composite indicator including a range of processes that are related to elevation, such as long term precipitation, evaporation and soil moisture availability. In cold regions such as Austria snow will also be important. One would expect that the predictive performance improves with increasing elevation, since higher elevations may be wetter and more snow dominated, both of

which may contribute to an improvement of predicting performance. Indeed 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. Table 2 indicates very little dependence between performance and mean an-
5 nual precipitation for the case of process based method. For the geostatistical method, there is a decreasing performance with precipitation for four out of six signatures. High precipitation catchments are the mountainous headwater which are also small.

The aridity index (the ratio of potential evaporation and precipitation on a long term basis, averaged across the catchment) is an indicator of the relative availability of en-
10 ergy and water affecting the water balance and therefore all runoff signatures. For the mean annual runoff the performances increase for increasing aridity index, consistently with their decrease with catchment elevation and mean annual precipitation. Regarding the other signatures, there is hardly any dependence on aridity for any of the signatures (except mean annual runoff). Only the performance of estimation of the integral scale
15 significantly decreases with increasing aridity when the process-based model is used. 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 pro-
20 cesses 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
25 catchments are really arid (aridity index 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.

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

Figure 8 shows a comparative summary of the results from Sects. 5.1 and 5.2. Figure 8a and b shows 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 50 % of the values of NE and ANE while the lines connect the median values $\widetilde{NE}$ and $\widetilde{ANE}$. 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. $\widetilde{NE}$) are very close to zero for the statistical method (Fig. 8a), which indeed is optimised in a way to minimise biases (which still remain, since the performances are calculated in cross-validation mode). The biases of the process-based method may be related not only to the method of parameter regionalisation, but also to biases in the inputs (i.e. precipitation and air temperature), in the model structure and parameter estimation.

When the runoff signatures are compared among themselves, one sees that the lowest performances 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 floods are much lower than, for example, for low flows.

Figure 8c 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. 8c 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. 8a and b. The higher predictability of mean annual runoff and seasonal runoff is due to the aggregation of

runoff variation over a relatively long time period. They therefore vary more smoothly in space, which enhances their predictability. In contrast, the extremes (low flows and floods) have lower R^2. Extremes are generally harder to predict than averages, because of their more heterogeneous nature. In terms of R^2 low flows are easier to predict than floods because droughts tend to persist over larger areas and longer time scales, making the estimation of low flows from other stream gauges in the area fairly robust. The low R^2 for floods contrasts with Fig. 8a and b, where the floods were the ones with highest prediction performance. The R^2 for floods is so low because their spatial variability is low and relatively little explained by the PUB methods. The spatial variability of the integral scale can be predicted with more confidence. This is because most parts of the hydrographs (recession) are easy to predict. Although the extremes are harder to predict, the model efficiency metric treats all time steps with the same weight, reducing the impact of poorer predictive capacity for the flow extrema.

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 optimized to predict specific signatures and their characteristics. In the Austrian example, the targeted method for low flows estimation (see e.g. Laaha and Blöschl, 2007) gives significantly better performances (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 floods have a median regionalisation 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.

6   Conclusions

An assessment of the performance of predicting six runoff signatures in ungauged basins has been conducted using two methods for hydrograph regionalisation. 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 the statistical approach (Top-Kriging) generally outperforms the rainfall–runoff model. On average, the biases are small ($< 10\%$ for most of the signatures), but not negligible,
5   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. The average error spread is lower than 10% of the observed values for the statistical region-
10 alisation method while it is somewhat higher for the process-based method. The better performance of Top-Kriging is due to a number of reasons. First, the stream gauge density of the study region is quite high (689 stations over an area of $80\,000\,\text{km}^2$), so there is a lot of runoff information available for Top-Kriging which uses correlations along the stream network. In countries where runoff measurements are more sparse,
15 process-based methods or other statistical methods based on catchment attributes may perform relatively better than geostastical 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 identiability associated with rainfall
20 runoff models. 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.

    The predictive performance in ungauged basins is correlated with a number of climate and catchment characteristics. The predictive performance of Top-Kriging in-
25 creases with increasing catchment area for all six signatures significantly, while the dependence is less pronounced for the case of regionalised rainfall–runoff models. 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, runoff processes

tend to become more linear as catchment are increases due to aggregation effects which may increase the predictability (see e.g. Sivapalan, 2003).

For the rainfall–runoff model the regionalation performance tends to increase with elevation as a result of snow processes in the mountainous catchments, which are easier to predict. For the statistical method no clear dependencies with elevation or aridity are apparent, but there is a tendency for the performances to decrease with mean annual precipitation. This is probably because of the higher mean annual precipitation in the mountains where the catchments are smaller.

Annual and seasonal runoff can be predicted more accurately than all other signatures. This is because they vary more smoothly in space than the other signatures. 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 because of the fact that they are extremes, so their spatial patterns may involve a lot of small scale heterogeneiy 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 because the processes associated with low flows (in particular climate, longer time scale dry spells) vary more smoothly in space than do the processes associate with high flows and floods.

The relative performance of runoff prediction for different signatures depends on the performance measures selected for the assessment. In this paper we use the coefficient of determination (as a measure of the ability to capture the spatial patterns of the signatures) and the absolute normalised error (as a measure of the local predictive performance). For some signatures such as high flows these error measures give different results because they measure different things. In the literature on runoff regionalisation there is a wide variety of performance measures used. It would be useful to report performance in a consistent way by more than one performance measure in order to be able to generalise the findings beyond individual studies.

Part 3: Runoff signatures in Austria

A. Viglione et al.

References

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. 466

Bergström, S.: The HBV model, in: Computer Models of Watershed Hydrology, edited by: Singh, V. P., Water Resources Publications, Highlands Ranch, Colorado, 443–476, 1995. 456

Blöschl, G.: Hydrologic synthesis: across processes, places, and scales, Water Resour. Res., 42, W03S02, doi:10.1029/2005WR004319, 2006. 455

Blöschl, G. and Merz, R.: Landform – Hydrology Feedbacks, in: Landform – Structure, Evolution, Process Control, edited by: Otto, C. and Dikau, R., Springer, Wien, Heidelberg, 117–126, 2010. 455

Blöschl, G., Grayson, R. B., and Sivapalan, M.: On the representative elementary area (REA) concept and its utility for distributed rainfall–runoff modelling, Hydrol. Process., 9, 313–330, 1995. 464

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. 466

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, W11510, doi:10.1029/2010WR010174, 2011. 456

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

10, 449–485, 2013

Part 3: Runoff signatures in Austria

A. Viglione et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Gaál, 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. 454

Jefferson, A., Grant, G., Lewis, S., and Lancaster, S.: Coevolution of hydrology and topography on a basalt landscape in the Oregon Cascade Range, USA, Earth Surf. Proc. Land., 35, 803–816, doi:10.1002/esp.1976, 2010. 453

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. 451

Kollat, J., Reed, P. M., and Wagener, T.: When are multiobjective calibration trade-offs in hydrologic models meaningful?, Water Resour. Res., 48, W03520, doi:10.1029/2011WR011534, 2012. 463

Laaha, G. and Blöschl, G.: A national low flow estimation procedure for Austria, Hydrol. Sci., 52, 625–644, 2007. 468

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. 455

Merz, R. and Blöschl, G.: Flood frequency hydrology: 2. Combining data evidence, Water Resour. Res., 44, W08433, doi:10.1029/2007WR006745, 2008b. 455

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. 457

Merz, R., Parajka, J., and Blöschl, G.: Scale effects in conceptual hydrological modeling, Water Resour. Res., 45, W09405, doi:10.1029/2009WR007872, 2009. 465

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. 456, 457

Mészáros, 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., Slovak NC IHP UNESCO/UH SAV, Bratislava, Slovakia, 127–135, 2002. 457

Montanari, A.: Uncertainty of hydrological predictions, in: Treatise on Water Science, vol. 2, edited by: Wilderer, P., Elsevier, chapter 2.17, 459–478, 2011. 458

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀ | ▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Montanari, A. and Toth, E.: Calibration of hydrological models in the spectral domain: an opportunity for scarcely gauged basins?, Water Resour. Res., 43, W05434, doi:10.1029/2006WR005184, 2007. 463

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. 465

Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, Hydrol. Earth Syst. Sci., 9, 157–171, doi:10.5194/hess-9-157-2005, 2005. 456, 457

Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, Hydrol. Process., 21, 435–446, doi:10.1002/hyp.6253, 2007. 456

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, Hydrol. Earth Syst. Sci. Discuss., 10, 375–409, doi:10.5194/hessd-10-375-2013, 2013. 450, 451, 452, 456, 466

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2012. 460, 465, 477

Salinas, J. L., Parajka, J., Viglione, A., Rogger, M., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 2: Flood and low flow studies, Hydrol. Earth Syst. Sci. Discuss., 10, 411–447, doi:10.5194/hessd-10-411-2013, 2013. 450, 451, 452, 466

Singh, V. and Frevert, D.: Watershed Models, Taylor & Francis, 680 pp., 2005. 456

Sivapalan, M.: Process complexity at hillslope scale, process simplicity at the watershed scale: is there a connection?, Hydrol. Process., 17, 1037–1041, doi:10.1002/hyp.5109, 2003. 470

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., 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, Hydrol. Sci., 48, 857–880, doi:10.1623/hysj.48.6.857.51421, 2003. 450

Skøien, J. O. and Blöschl, G.: Sampling scale effects in random fields and implications for environmental monitoring, Environ. Model. Assess., 114, 521–552, doi:10.1007/s10661-006-4939-z, 2006a. 457

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Skøien, J. O. and Blöschl, G.: Scale effects in estimating the variogram and implications for soil hydrology, Vadose Zone J., 5, 153–167, 2006b. 457

Skøien, J. O. and Blöschl, G.: Spatiotemporal topological kriging of runoff time series, Water Resour. Res., 43, W09419, doi:10.1029/2006WR005760, 2007. 457

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. 465

Skøien, J. O., Merz, R., and Blöschl, G.: Top-kriging – geostatistics on stream networks, Hydrol. Earth Syst. Sci., 10, 277–287, doi:10.5194/hess-10-277-2006, 2006. 457

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. 464

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. 464

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. 463

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 1. Median Nash–Sutcliffe efficiency measure calculated for the time series simulated vs. measured in the 213 stations of Fig. 3. The Nash–Sutcliffe efficiency measure is analogous to the coefficient of determination in Eq. (6) but calculated in time (i referring to time-steps rather than catchments).

	Calibration	Cross-validation
Process based	0.72	0.61
Top-Kriging	–	0.87

Table 2. Attributes and signature values for the 213 Austrian catchments. The signatures have been calculated from observed daily runoff from 1976 to 2008.

	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 $\overline{Q_m}$ ($mm\,yr^{-1}$)	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

Table 3. Spearman correlation coefficient between the absolute normalised error (ANE_i) 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 ann. prec. PB/TK	Aridity index PB/TK
(a) Mean ann. runoff $\overline{Q_m}$	−0.09/**−0.22**	**0.18**/0.13	**0.20**/**0.21**	**−0.21**/**−0.17**
(b) Range of Pardé coeff. ΔPar	−0.02/**−0.33**	**−0.15**/−0.03	0.05/**0.20**	0.07/−0.08
(c) Slope of FDC m_{FDC}	0.02/**−0.20**	−0.13/−0.05	−0.05/0.03	0.12/0.02
(d) Normalised low flow q_{95}	**−0.29**/**−0.35**	0.00/−0.06	0.09/0.07	−0.07/−0.06
(e) Normalised high flow q_{05}	**−0.23**/**−0.34**	**−0.20**/−0.09	0.06/**0.19**	0.04/−0.07
(f) Integral scale $\tau_{1/e}$	**−0.16**/**−0.23**	**−0.64**/0.02	−0.03/**0.19**	**0.38**/−0.07
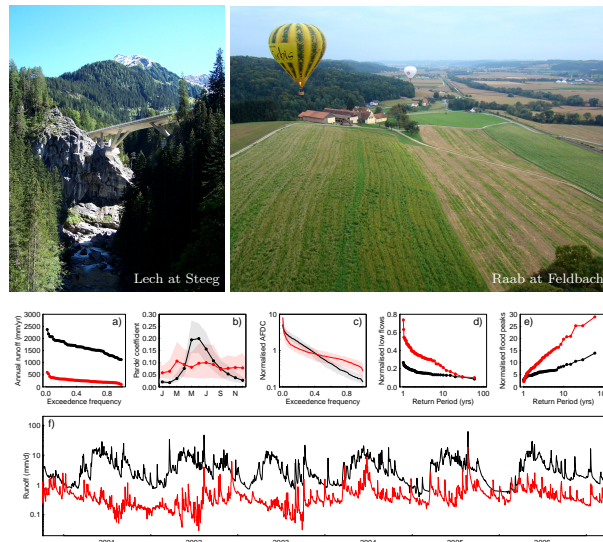
Fig. 1. Signatures of runoff variability for two catchments: Lech at Steeg (area of $248\,km^2$, median elevation of $1944\,m$ a.s.l. and mean annual precipitation of $1520\,mm\,yr^{-1}$) in the western Alps (black lines) and Raab at Feldbach (area of $689\,km^2$, median elevation of $470\,m$ a.s.l. and mean annual precipitation of $846\,mm\,yr^{-1}$) in south-eastern Austria (red lines). The six signatures are: **(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 (i.e. q_{95} defined in Sect. 4.2); **(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. Lech_von_alter_Passtrasse_zur_neuen_Lechbruecke.jpg and Raab_valley_near_Morgensdorf.jpg are from https://ssl.panoramio.com/photo/58708762 and https://ssl.panoramio.com/photo/43726186, respectively.
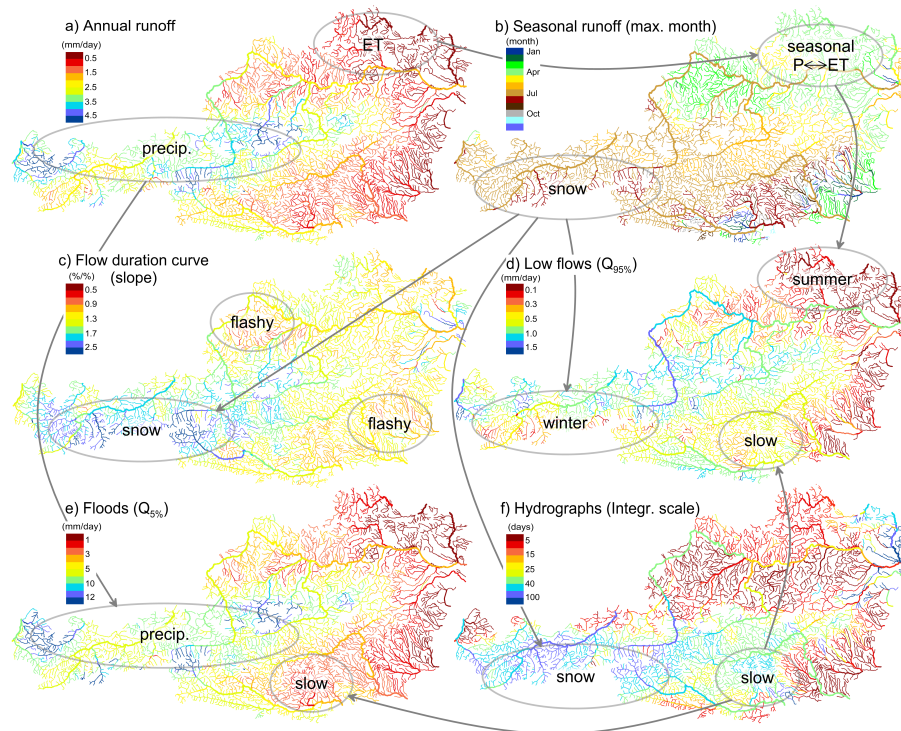
Fig. 2. Connection of runoff signatures for ungauged basins in Austria. The maps have been obtained by regionalising daily runoff time series through Top-Kriging (Sect. 3.2) and extracting then the following signatures: **(a)** mean annual runoff, **(b)** timing of monthly runoff maxima as seasonality signature, **(c)** slope of the flow duration curve, **(d)** daily runoff value which is exceeded 95 % of the time as a measure of low flows, **(e)** daily runoff value which is exceeded 5 % of the time as a measure of high flows, **(f)** integral time scale as a measure of catchment response memory (see Sect. 4.2 for the definition). Ellipses and arrows help the description of connectivity between process and response provided in the text.
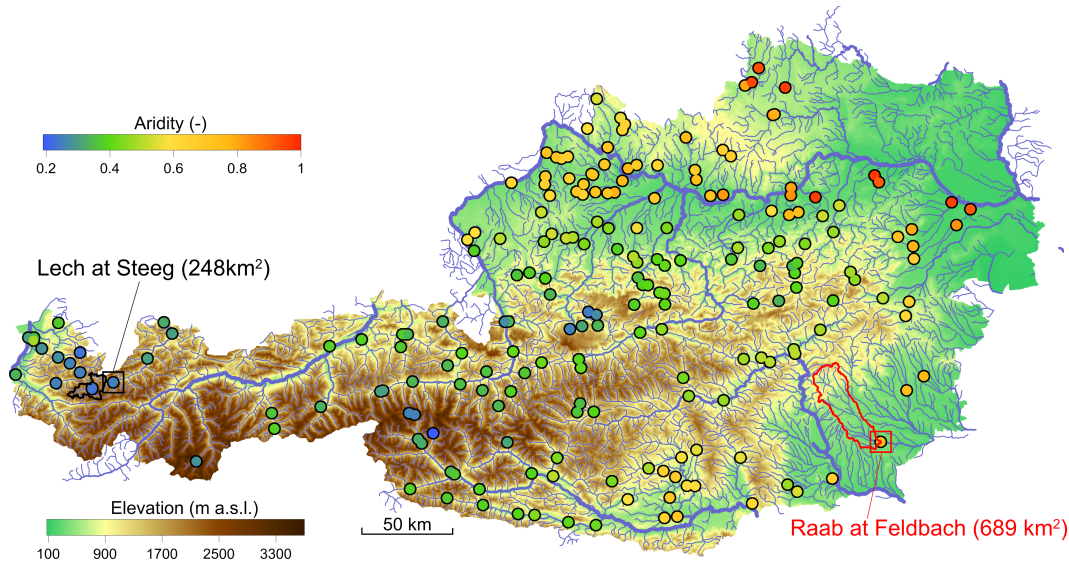
Fig. 3. Topography and river network of Austria and location of 213 streamgauges considered for the cross-validation (points) colour-coded according to catchment aridity.
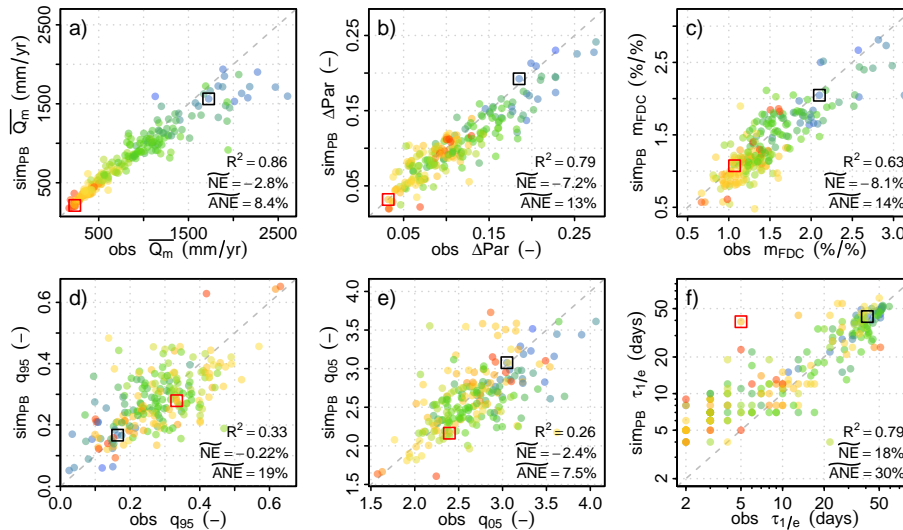
Fig. 4. Observed vs. simulated signatures using the process-based method (PB) in cross-validation mode for the period 1976–2008: **(a)** mean annual specific runoff (mm yr^{-1}), **(b)** range of the Pardé coefficient (–), **(c)** slope of the normalised flow duration curve (%/%), **(d)** normalised 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. 3. The coefficient of determination R^2, the median normalised error $\widetilde{\text{NE}}$ and the median absolute normalised error $\widetilde{\text{ANE}}$ (as percentages) are given. The two catchments of Fig. 1 are indicated by the black and red boxes.
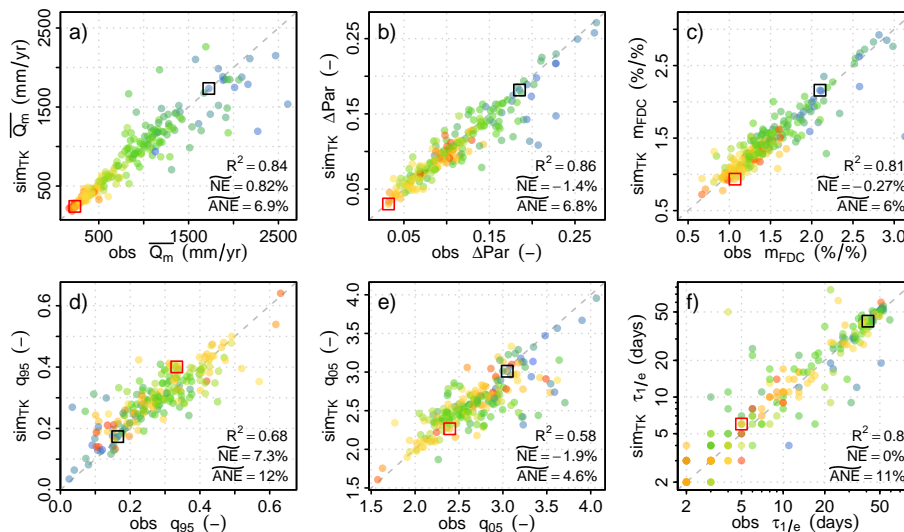
Fig. 5. Observed vs. simulated signatures using Top-Kriging (TK) in cross-validation mode for the period 1976–2008: **(a)** mean annual specific runoff (mm yr^{-1}), **(b)** range of the Pardé coefficient (–), **(c)** slope of the normalised flow duration curve (%/%), **(d)** normalised flow duration 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. 3. 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 of Fig. 1 are indicated by the black and red boxes.
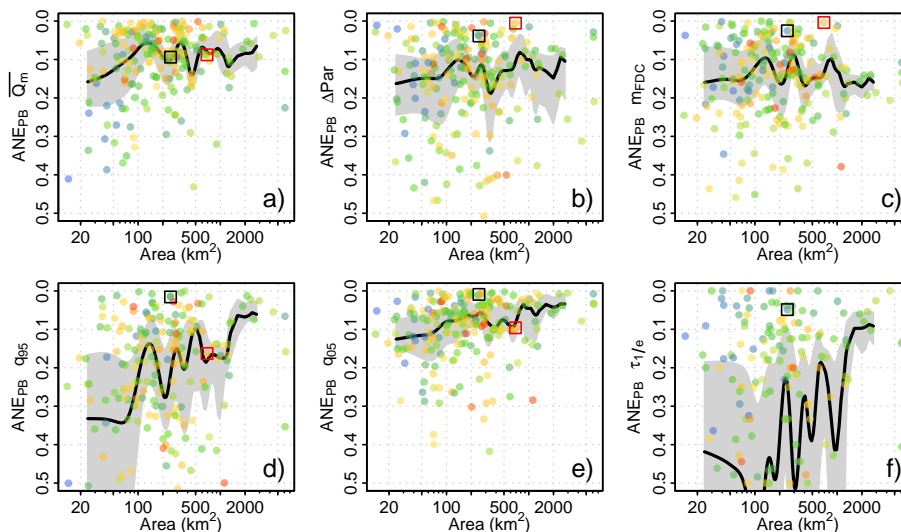
Fig. 6. Absolute Normalised Error vs. catchment area (km^2) for the signatures regionalised using the process-based model (PB) in cross-validation mode: **(a)** mean annual specific runoff ($mm\,yr^{-1}$), **(b)** range of the Pardé coefficient (–), **(c)** slope of the normalised flow duration curve (%/%), **(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. 3. The two catchments of Fig. 1 are indicated by the black and red boxes.
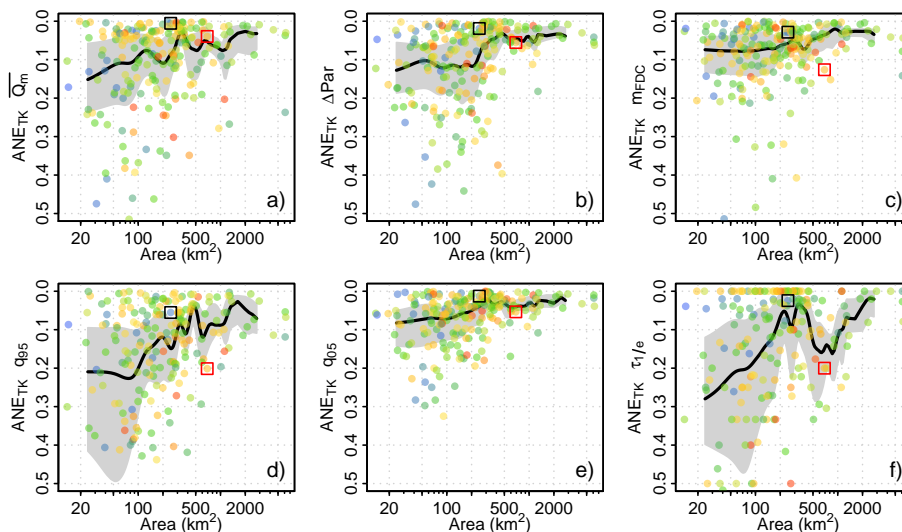
Fig. 7. Absolute Normalised Error vs. catchment area (km^2) for the signatures regionalised using Top-Kriging (TK) in cross-validation mode: **(a)** mean annual specific runoff (mm yr^{-1}), **(b)** range of the Pardé coefficient (–), **(c)** slope of the normalised flow duration curve (%/%), **(d)** 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. 3. The two catchments of Fig. 1 are indicated by the black and red boxes.
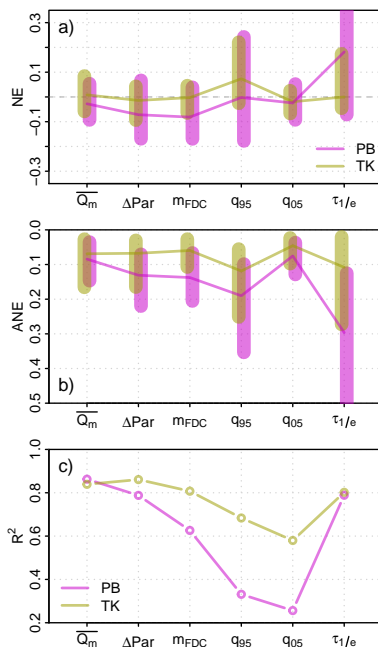
Fig. 8. Comparison of cross-validation performance of prediction methods for different runoff signatures in ungauged basins in Austria. The performance measures are: **(a)** normalised error NE, **(b)** absolute normalised error ANE (the bars contain 50 % of the values and the lines connect the medians). **(c)** coefficient of determination R^2. The prediction methods are: (fuchsia) process-based method – conceptual rainfall–runoff model whose parameters are regionalised. (beige) statistical method – Top-Kriging. The signatures are: $\overline{Q}_m$ – 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).