



# Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies

J. Parajka<sup>1,\*</sup>, A. Viglione<sup>1</sup>, M. Rogger<sup>1</sup>, J. L. Salinas<sup>1</sup>, M. Sivapalan<sup>2</sup>, and G. Blöschl<sup>1</sup>

<sup>1</sup>Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria

<sup>2</sup>Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 North Mathews Ave., Urbana, USA

\* on leave from: Institute of Hydrology, Slovak Academy of Sciences, Liptovský Mikuláš, Slovakia

Correspondence to: J. Parajka (parajka@hydro.tuwien.ac.at)

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

Revised: 10 April 2013 – Accepted: 14 April 2013 – Published: 7 May 2013

**Abstract.** The objective of this assessment is to compare studies predicting runoff hydrographs in ungauged catchments. The aim is to learn from the differences and similarities between catchments in different locations, and to interpret the differences in performance in terms of the underlying climate and landscape controls. The assessment is performed at two levels. The Level 1 assessment is a meta-analysis of 34 studies reported in the literature involving 3874 catchments. The Level 2 assessment consists of a more focused and detailed analysis of individual basins from selected studies from Level 1 in terms of how the leave-one-out cross-validation performance depends on climate and catchment characteristics as well as on the chosen regionalisation method. The results indicate that runoff-hydrograph predictions in ungauged catchments tend to be more accurate in humid than in arid catchments and more accurate in large than in small catchments. The dependence of performance on elevation differs by regions and depends on how aridity varies with elevation and air temperature. The effect of the parameter regionalisation method on model performance differs between studies. However, there is a tendency towards a somewhat lower performance of regressions than other methods in those studies that apply different methods in the same region. In humid catchments spatial proximity and similarity methods perform best while in arid catchments similarity and parameter regression methods perform slightly better. For studies with a large number of catchments (dense stream gauge network) there is a tendency for spatial proximity and geostatistics to perform better than regression or regionalisation based on simple averaging of model parameters from gauged

catchments. There was no clear relationship between predictive performance and the number of regionalised model parameters. The implications of the findings are discussed in the context of model building.

## 1 Introduction

Runoff hydrographs, i.e. the time series of river runoff, are the result of numerous interacting processes within the catchment: precipitation, runoff generation at the land surface, infiltration into the subsurface, uptake from vegetation and consequent transpiration, evaporation from the soil, water movement through various flow paths on the land surface, in the unsaturated zone and in the groundwater. Understanding the hydrograph will help understand how these processes combine. Predictions of runoff hydrographs are one way of testing our hypotheses on these processes. Predictions of runoff hydrographs are also needed for practical purposes such as obtaining design characteristics for spillways, culverts and embankments, for water resources management applications such as water allocation for irrigation, industry and human use, hydropower operation and environmental flow estimation. They are also useful for risk management such as in flood and drought forecasting. Finally, there is considerable interest in assessing the effects of environmental change (e.g. land use, hydraulic structures, climate) on the runoff hydrographs and water quality for which accurate runoff predictions are needed (Sachs and McArthur, 2005; Kovacs et al., 2012; Blöschl and Montanari, 2010). Clearly, predictions

of runoff hydrographs are important for many purposes of societal relevance. However, in most catchments of interest no runoff data are available, so the hydrographs need to be predicted from other information within that catchment or from other catchments. This is the “Prediction in Ungauged Basins” or PUB problem.

In 2003, the International Association of Hydrological Sciences (IAHS) launched a concerted effort on investigating the PUB problem, the PUB initiative. The main focus of this initiative was to advance the knowledge and understanding of climatic and landscape controls on hydrologic processes occurring at all scales and to improve the ability to predict the fluxes of water in ungauged basins, along with their uncertainties (Sivapalan et al., 2003). One of the clear tasks that the PUB initiative set out to achieve was to address the fragmentation of modelling approaches through comparative evaluation: “Classify model performances in terms of time and space scales, climate, data requirements and type of application, and explore reasons for the model performances in terms of hydrological insights and climate-soil-vegetation-topography controls” (SSG: PUB Science Steering Group, 2003, p. 18).

The objective of this and two companion papers (Salinas et al., 2013; Viglione et al., 2013) is to compare different approaches for runoff prediction in ungauged catchments. While companion papers investigate predicting performance of methods for extreme runoff estimation (Salinas et al., 2013) and compare statistical and process based methods for predicting a range of runoff characteristics at different timescales (Viglione et al., 2013), in this paper we compare studies predicting runoff hydrographs in ungauged catchments. The aim is to learn from the differences and similarities between catchments in different locations, and to interpret the differences in performance in terms of the underlying climate and landscape controls. In particular, the following research questions are addressed:

- i. How good are the runoff predictions in different climates?
- ii. Which parameter regionalisation method performs best?
- iii. How does data availability impact performance?
- iv. How does model complexity impact performance?
- v. To what extent does runoff prediction performance depend on climate and catchment characteristics?

## 2 Method of comparative assessment

For the comparative assessment of runoff-hydrograph predictions in ungauged basins, a two-step process has been adopted in this paper:

*Level 1 assessment:* in a first step, a literature survey was performed. Publications in the international refereed literature were scrutinised for results of the predictive performance of runoff hydrographs. The Level 1 assessment is a meta-analysis of prior studies performed by the hydrological community. The advantage of this type of meta-analysis is that a wide range of environments, climates and hydrological processes can be covered that go beyond what can be reasonably achieved by a single study. It is a comparative assessment that synthesises the results from the available international literature. However, the level of detail of the information provided is often limited. The results in the literature were almost always reported in an aggregated way, i.e. as average or median performance over the study region or part of the study region.

*Level 2 assessment:* to complement the Level 1 assessment, a second assessment step was performed, termed Level 2 assessment. In this step, some of the authors of the publications from Level 1 were approached to provide data on their runoff-hydrograph predictions *for individual ungauged basins*. The data they provided included information on the catchment and climate characteristics, on the method used, the data availability, and predictive performance. The overall number of catchments involved was smaller than in the Level 1 assessment, so the spectrum of hydrological processes covered in the assessment was narrower. However, the amount and detail of information available in particular catchments was much higher. As in Level 1, the cross-validation performance for ungauged basins was analysed; however, information on individual catchments was now available. The cross-validation performance was estimated by a leave-one-out strategy, where each gauged catchment was in turn considered as ungauged and estimated runoff was compared with the observed runoff hydrographs. The predictive accuracy was then described by the Nash–Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970) of daily runoff.

The comparative assessment conducted in this paper stratifies the analyses into three main groups:

1. *Analysis of process controls on the model performance.* A number of climate and catchment characteristics have been identified. A large number of catchments and modelling studies around the world have then been organised according to these climate and catchment characteristics, with a view to learning from their differences and similarities in performance in a general way.
2. *Analysis of predictive performance for different types of methods.* The methods for estimating the parameters of rainfall-runoff models in ungauged basins have been grouped into the classes discussed in Sect. 3. Rather than evaluating specific methods the focus has been on types of methods, so to be able to generalise beyond individual studies.

**Table 1.** Summary assessment of existing regionalization studies for continuous runoff simulation. Statistical evaluation indicates the leave-one-out assessment of runoff model efficiency. Methods used for transfer of hydrologic model parameters include: spatial proximity (P), similarity (S), model averaging (MA), parameter regression (R) and regional calibration (RC).

Study	Region	Climate	Hydrologic model	Regionalisation method	Runoff model efficiency	Used in Level 2
Caballero et al. (2013)	Honduras	Tropical	SWB	A priori	−4.3–0.16*	
Petheram et al. (2012)	northern Australia	Tropical/arid	AWBM, SIMHYD, IHACRES, SMARG, Sacramento	P	0.53, 0.54, 0.55, 0.50, 0.50/0.45, 0.43, 0.45, 0.45, 0.43	X
Samuel et al. (2011)	Canada (Ontario)	Cold	MAC-HBV	P, MA, R	0.57–0.59, 0.31–0.46, 0.51–0.52	
Chiew (2010)	southeast Australia	Humid	Sacramento, IHACRES, AWBM, SMAR, SIMHYD	P	0.66, 0.68, 0.63, 0.59, 0.54	
Samaniego et al. (2010b)	Germany	Humid	MHM	P	0.48–0.75, 0.72–0.79	X
Samaniego et al. (2010a)	Germany	Humid	MHM	MA	0.78–0.83	
Zhang and Chiew (2009)	southeast Australia	Humid	SIMHYD, Xinanjiang	P, S	0.48–0.56, 0.46–0.58, 0.51–0.56, 0.48–0.52	X
Viviroli et al. (2009)	Switzerland	Cold	PREVAH	P, R	0.67–0.70, 0.65	X
Viney et al. (2009)	southeast Australia	Arid	SMAR-G, Simhyd, Sacramento, IHACRES, AWBM	P	0.60–0.62, 0.63, 0.60–0.67, 0.50–0.59, 0.60–0.61	
Seibert and Beven (2009)	Sweden	Cold	HBV	MA	0.5	
Reichl et al. (2009)	Australia	Humid	SimHyd	P, R, MA	0.63, 0.55, 0.66	
Post (2009)	Australia (Burdekin)	Arid	IHACRES	R	−0.64–0.74	
Li et al. (2009)	southeast Australia	Humid	Xinanjiang	P, S	0.50–0.52, 0.50–0.52	
Bulygina et al. (2009)	United Kingdom (Wales)	Humid	PDM	S	0.65–0.84	
Zvolenský et al. (2008)	Slovakia	Cold	HBV	S, P	0.62–0.71, 0.61–0.73	X
Oudin et al. (2008)	France	Humid	TOPMO, GR4J	MA, P, R	0.69, 0.71, 0.55, 0.71, 0.74, 0.68	X
Kim and Kaluarachchi (2008)	Ethiopia (Blue Nile)	Humid	Monthly water balance	MA, RC, R	0.56–0.60, 0.66, 0.66	
Hundecha et al. (2008)	Germany (Rhine River)	Humid	HBV	RC	0.82–0.93	
Bastola et al. (2008)	Nepal, Japan, Australia, United Kingdom, France	Humid	TOPMODEL	RC, R, R-ANN	0.56–0.87, 0.41–0.86, 0.56–0.87	
Zvolenský et al. (2007)	Slovakia	Cold	HBV	P, MA	0.60–0.72, 0.54–0.71	
Parajka et al. (2007)	Austria	Cold	HBV	RC	0.63–0.67	X
Goswami et al. (2007)	France	Humid	7 models	S, MA	0.33–0.73, 0.31–0.46	
Cutore et al. (2007)	Italy (eastern Sicily)	Humid	Rainfall-runoff regression model	R	0.48–0.81	
Boughton and Chiew (2007)	Australia and Tasmania	Humid	AWBM	R	0–0.95	
Young (2006)	United Kingdom	Humid	PDM	R	0.66	
Wagener and Wheater (2006)	United Kingdom	Humid	PDM	R	0.76–0.78	
Parajka et al. (2006)	Austria	Cold	HBV	Proxy data	0.59–0.61*	
Allasia et al. (2006)	South America (Uruguay River)	Humid	MGB-IPH	S	0.62–0.84	
Vogel (2005)	southeastern USA	Humid	ABCD	RC	0.69–0.93 ( $R^2$ )*	
Parajka et al. (2005)	Austria	Cold	HBV	S, R, P	0.61–0.67, 0.60–0.65, 0.62–0.67	X
McIntyre et al. (2005)	United Kingdom	Humid	PDM	MA	As Figure (0.08–0.49)*	X
Merz and Blöschl (2004)	Austria	Cold	HBV	R, P	0.49–0.56, 0.53–0.59	
McIntyre et al. (2004)	United Kingdom	Humid	PDM	MA, S, R	0.75, 0.75, 0.66	
Kokkonen et al. (2003)	USA (North Carolina)	Humid	IHACRES	S, R	0.68–0.88, 0.60–0.88	

The results marked by \* are not included in Figs. 2–6, because of specific regionalisation method, performance criteria or median model efficiency less 0.3.

**Table 2.** Number of studies (in brackets number of results) and number of catchments used. Arid relates to catchments with an aridity index  $> 1$ , humid to those with an aridity index  $\leq 1$ . Level 1 refers to an assessment of the average performance of studies, Level 2 to an assessment of the performance for individual catchments.

Assessment	Catchments	Number of studies	Number of catchments
Level 1	All	34 (75)	3874
Level 2	All	9	1832
Level 2	Humid	8	1479
Level 2	Arid	4	353

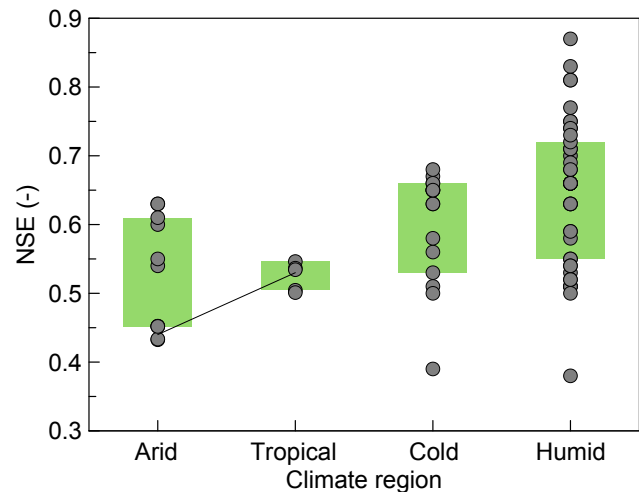
3. *Analysis of data availability.* The quality of runoff predictions in ungauged basins not only depends on the hydrological setting and the regionalisation method but also, importantly, on the data that are available for the regionalisation. The comparison therefore also examines the number of stream gauges available in a particular study as an index to characterize data availability.

### 3 Studies and datasets used

Table 1 lists the 34 studies published in the last decade that are used in this paper. It includes summary information about the study region, regionalisation method applied and the predictive runoff model efficiency. The consistency of results differs between the studies. In some papers, the results are presented only as figures, in others these are summarized by median or range of runoff model performance. Several studies compare different hydrologic models and/or regionalisation approaches, which results in a total of 75 assessments of predictive performance. These results are the base for the Level 1 assessment, which represents at total of 3874 catchments (Table 2). Nine study authors out of the Level 1 assessment provided detailed information about climate and catchment characteristics in a consistent way and reported the regionalisation performance for each catchment (Level 2 assessment). This dataset combines data from 1832 catchments. Three catchment characteristics are analysed: aridity index, mean elevation and catchment area. Aridity index (the ratio of potential evaporation and precipitation on a long-term basis, averaged across the catchment) is an indicator of the competition between energy and water affecting the water balance. Elevation (average topographic elevation within the catchment) is a composite indicator including a range of processes, such as long-term precipitation and hence soil moisture availability and air temperature. In some environments there is a relationship between elevation and aridity and elevation and snow processes. Catchment area is an indicator of the degree of aggregation of catchment processes related to scale effects; an indicator of storage within the catchment; and an indicator of the amount of rainfall data



**Fig. 1.** Map indicating the countries included in the meta-analysis of studies reported in the literature (Level 1 assessment).



**Fig. 2.** Median Nash–Sutcliffe efficiency of predicting hydrographs in ungauged basins stratified by climate (Level 1 assessment). Each symbol refers to a result from the studies in Table 1. Line indicates study (Petheram et al., 2012) where the same method was applied across different climatic regions. Boxes show 25–75 % quantiles.

that is available for runoff estimation in ungauged basins, since larger catchments tend to contain a large number of rain gauges. With increasing area also the estimation variance of areal rainfall decreases and an areal rainfall might be biased by increasing number of stations located in lower parts of the catchment (Lebel et al., 1987).

Prediction of runoff hydrographs in ungauged catchments is traditionally based on hydrologic model simulations. Almost all the studies reported in Table 1 used lumped conceptual models, a few studies used semi-distributed (Parajka et al., 2005), HRU-based (Viviroli et al., 2009) or distributed models (Allasia et al., 2006; Samaniego et al., 2010a, b). Most of the models predict the hydrographs at a daily time step. In the case of conceptual models, the model parameters cannot usually be measured or inferred from measurements. The parameters therefore need to be transferred (regionalised) from gauged catchments in the region, termed donor catchments (Blöschl, 2005). There is a plethora of

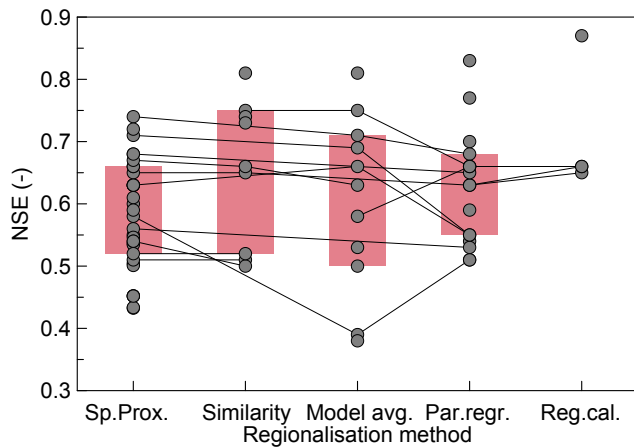
different methods used for parameter regionalisation. In the Level 1 and Level 2 assessments we assigned them into *five* groups: spatial proximity, similarity, model averaging, parameter regression and regional calibration. While the spatial proximity, similarity and model averaging methods, assume that the *entire* parameter set of a gauged basin is also valid in the ungauged basins, parameter regression and regional calibration methods relate *individual* model parameters to catchment characteristics. A more detailed description of each group of regionalisation methods is as follows.

- *Spatial proximity*: if one assumes that climate and catchment characteristics vary only smoothly in space then spatial proximity between the catchments may be a suitable similarity measure to select the donor catchment. Proximity is usually defined on the basis of distances between the catchment outlets or catchment centroids (Randrianasolo et al., 2011; Zvolenský et al., 2008; Li et al., 2009). It is also possible to use the geostatistical distances, which account for the nestedness of the catchments (e.g. Skøien et al., 2006; Skøien and Blöschl, 2007).
- *Similarity*: an alternative is to choose the donor on the basis of the similarity of the climate and catchment characteristics in the two catchments. Similarity is usually measured by the root mean square difference of all the characteristics in a pair of catchments. The characteristics are usually standardised by their standard deviation or transformed in another way to make them comparable. Studies which choose a donor on the basis of this method use a wide range of climate and catchment characteristics. Kokkonen et al. (2003) transferred the entire parameter set from the catchment with the most similar elevation of the catchment outlet. McIntyre et al. (2004) defined the most similar catchment in terms of the catchment area, standardised annual average precipitation, and baseflow index. Other studies used a larger number of characteristics, such as Parajka et al. (2005) who defined the similarity by mean catchment elevation, areal proportion of porous aquifers, lake index, stream network density, soils, geology and land use, and Zhang and Chiew (2009) who identified the most similar catchments in terms of catchment area, mean elevation, slope, stream length, aridity, woody vegetation fraction and plant-available water-holding capacity. As it is discussed in Oudin et al. (2010), more relevant catchment characteristics should be sought to better describe the geological and lithological conditions from a hydrological perspective.
- *Model averaging*: sometimes a weighted combination of the parameter sets from more than one donor catchment is used, where the catchments are selected either based on proximity, catchment characteristics or both (Goswami et al., 2007; Kim and Kaluarachchi, 2008; Seibert and Beven, 2009). One can either assume a fixed subdivision of the region into groups of catchments or, alternatively, allow each catchment to have its own group of donor catchments (Burn and Boorman, 1993).
- *Parameter regression*: alternatively, the calibrated model parameters can be related *individually* to catchment characteristics in the gauged catchments through empirical relationships, and these can be used to estimate the model parameters in the ungauged basin. The most common method of this type is the parameter regression method. For example, Kokkonen et al. (2003) found the drying parameter of the IHACRES (Identification of unit Hydrographs And Component flows from Rainfall, Evaporation and Streamflow data) model to be negatively related to mean overland-flow distance and the time constant governing the rate of recession in the slow store to be related to topographic slope in the Coweeta catchment, North Carolina. Merz and Blöschl (2004) found the very fast storage coefficient to be negatively correlated with topographic slope and elevation. This implies that runoff response may be particularly flashy in the high elevation catchments in Austria. Ideally, the relationship between the model parameters and the catchment characteristics should be hydrologically justifiable to give confidence for extrapolation to ungauged basins. However, this is not always the case (e.g. Sefton and Howarth, 1998; Peel et al., 2000; Fernandez et al., 2000) due to unrepresentative catchment characteristics and identifiability issues of the model parameters (Blöschl, 2005).
- *Regional calibration*: instead of first estimating model parameters at each (gauged) site and then relating them to catchment characteristics by an empirical relationship as in the above methods, these two steps are implemented concurrently in the regional calibration method by calibrating the coefficients of these relationships. The main motivation for doing this is to find more reliable parameters than is possible by calibrating the model parameters themselves and to make use of the spatial information contained in the catchment characteristics. The studies differ in the empirical relationships used such as regressions (Fernandez et al., 2000), homogeneous groups (Hlavčová et al., 2000; Szolgay et al., 2003), and geostatistical methods (Parajka et al., 2007; Hundecha et al., 2008).

## 4 Results and discussion

### 4.1 How good are the predictions in different climates?

The synthesis of the results of the existing studies (Level 1) indicates that most of the studies were performed in Europe and Australia, and more studies were performed in humid



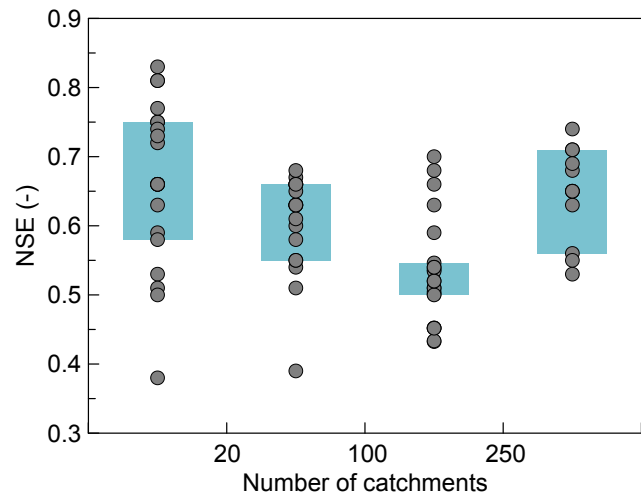
**Fig. 3.** Median Nash–Sutcliffe efficiency of predicting hydrographs in ungauged basins stratified by regionalisation method (Level 1 assessment). Each symbol refers to a result from the studies in Table 1. Lines indicate studies that compared different methods for the same set of catchments. Boxes show 25–75 % quantiles.

than in tropical and arid climates (Fig. 1 and Table 1). In total, there are 11, 5, 16 and 43 studies in arid, tropical, cold and humid climates, respectively.

Figure 2 shows that the performance of runoff predictions tends to be lower in arid than in cold and humid regions. The range of NSE varies between less than 0.4 (Goswami et al., 2007; McIntyre et al., 2005) to 0.87 (Hundechea et al., 2008). The median NSE is 0.54, 0.64 and 0.66 in arid, cold and humid regions, respectively. There is only one study that compares the performance of the same method for different climatic conditions, the study of Petheram et al. (2012), indicated as a grey line in Fig. 2. Their results show that, in Australia, the NSE runoff efficiency is higher in tropical than in arid catchments. The main reason that the methods perform less well in arid regions appears to be that arid regions tend to be spatially more heterogeneous and the hydrological processes more non-linear.

#### 4.2 Which method performs best?

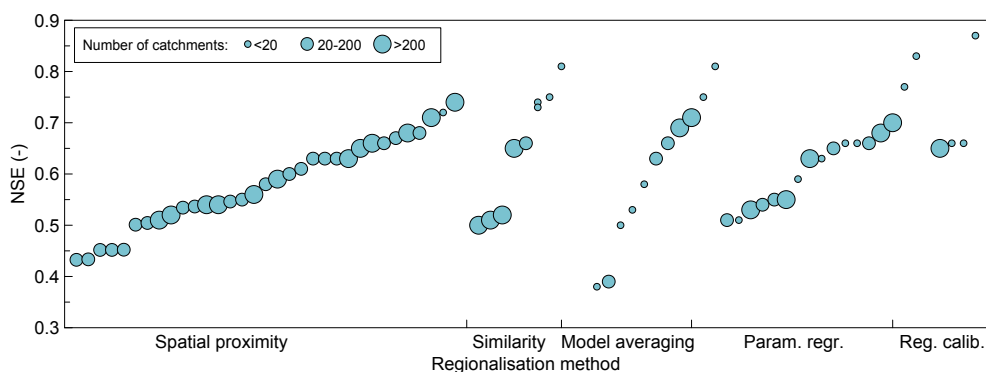
The parameter-regionalisation methods used in the Level 1 assessment include spatial proximity, similarity, model averaging, parameter regression and regional calibration. The assessments in each group are not based on exactly the same regionalisation approach, but the methodology is similar. The spatial proximity group consists of 33 results that include the nearest neighbour, kriging and inverse distance weighting interpolation methods. The similarity group (9 results) uses parameters from those catchments that are most similar in terms of catchment and/or climate characteristics. The parameter regression group includes 17 results with different regression models used for transfer of model parameters and one study (Boughton and Chiew, 2007) in which a hydrologic model is calibrated to mean annual runoff estimated by a regression



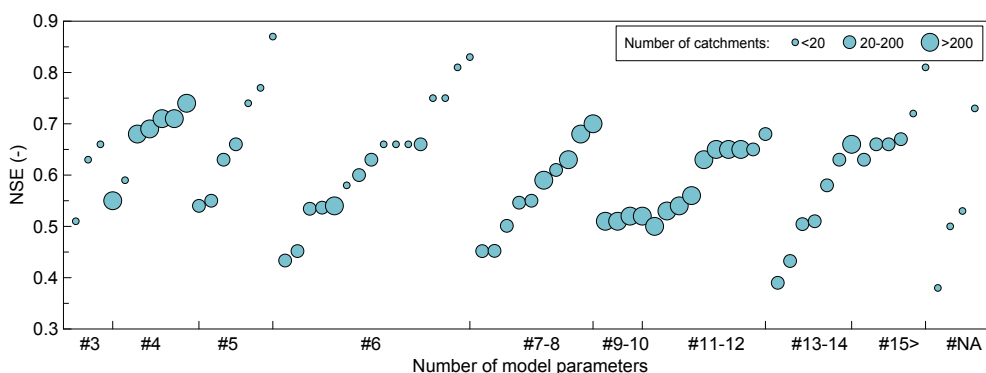
**Fig. 4.** Median Nash–Sutcliffe efficiency of predicting hydrographs in ungauged basins stratified by the number of catchments within each study (Level 1 assessment). Each symbol refers to a result from the studies in Table 1. Boxes show 25–75 % quantiles.

model. The model averaging group includes 12 results from either a regional pooling (averaging) of model parameters or ensemble runoff simulations for ungauged catchments. Finally, the regional calibration group includes 4 results from parameter estimation and model calibration simultaneously in a number of gauged catchments in a region.

The comparison of the methods (Fig. 3) indicates that the difference between the studies within each group is larger than between the groups. The NSE performance within each group is, for most of the assessments, within the range 0.5 and 0.75, while the median NSE for each group varies between 0.58 (spatial proximity) to 0.66 (similarity). The results of studies that compare different approaches (shown as grey lines in the figure) indicate that the predictive performance of parameter regression is poorer than the other methods, with the exception of one study (Samuel et al., 2011) where the simple average of model parameters performed the worst. In this case, however, the predictive performance is generally lower than in other published studies. The reasons why one approach to regionalisation may work better than others are discussed within several inter-comparison studies and other reviews (Merz and Blöschl, 2004; Oudin et al., 2008; Parajka et al., 2005; Vogel, 2005). Oudin et al. (2008), for example, reported that spatial proximity slightly outperformed the similarity method in regions with a dense stream gauge network. They reported that the predictive performance of these two approaches becomes similar when the density of stream gauges decreases to less than 60 gauges per 100 000 km<sup>2</sup>. Parajka et al. (2005) reported that a significant similarity in catchment characteristics over relatively short distances in Austria may contribute to the relatively good performance of the spatial proximity and similarity regionalisation methods.



**Fig. 5.** Median Nash–Sutcliffe efficiency of predicting hydrographs in ungauged basins stratified by the regionalisation method and ranked by performance (Level 1 assessment). Each symbol refers to a result from the studies in Table 1. Circle size indicates number of catchments per study.



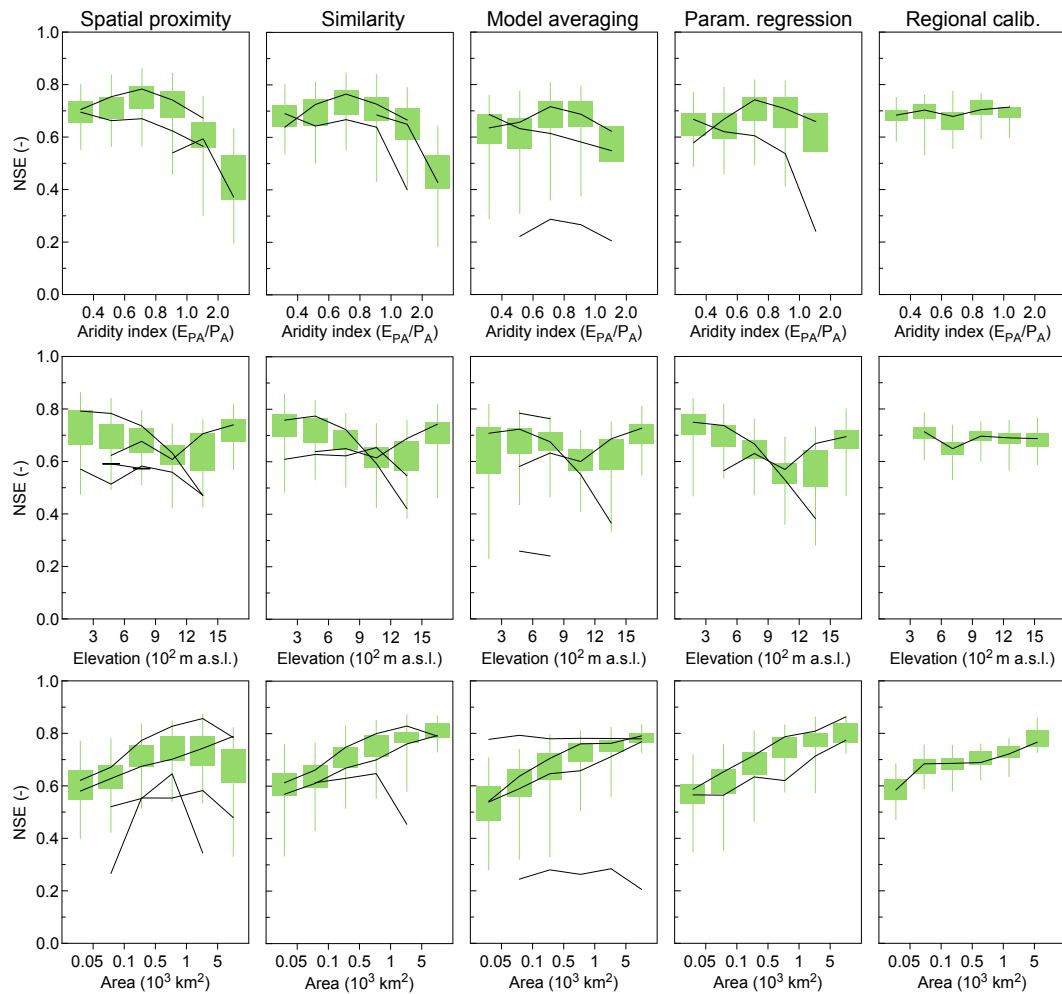
**Fig. 6.** Median Nash–Sutcliffe efficiency of predicting hydrographs in ungauged basins stratified by model complexity (number of transferred model parameters) and ranked by performance (Level 1 assessment). Each symbol refers to a result from the studies in Table 1. Circle size indicates number of catchments per study.

### 4.3 How does data availability impact performance?

Figure 4 shows the median Nash–Sutcliffe performance as a function of the number of catchments analysed in each study. As would be expected, the 21 studies with less than 20 catchments have the largest scatter in the performance because of the smallest sample size. As the number of catchments increases the performance tends to decrease. It is possible that, in some of the studies with few catchments, these catchments were hand picked in terms of suitability for regionalisation and this happens less frequently in the studies with more catchments. For 12 studies with more than 250 catchments the performance however tends to increase. Again, some selection of catchments based on automated methods may have been performed at that scale.

More detailed information on the dependence of performance on both method and number of catchments per study is shown in Fig. 5. Figure 5 summarizes 33, 9, 12, 17 and 4 results for spatial proximity, similarity, model averaging, parameter regression and regional calibration methods, respectively. The maximum performance exceeds 0.8 for the

similarity, regression and model averaging methods, but this performance is documented only for the small datasets. Interestingly, the performance of similarity-based regionalisation is clearly lower for assessments with large datasets. There are only a few studies that compare runoff-hydrograph predictions obtained by different groups of methods over large datasets (e.g. three or more groups of methods and validation in more than 25 catchments). These studies suggest that for regions with dense networks of gauging stations (e.g. France and Austria) the spatial proximity approach performed best. Oudin et al. (2008) concluded that spatial proximity was the best regionalisation method in France while the regression approach was the least satisfactory. The results of Parajka et al. (2005) indicate that, for Austria, kriging and similarity-based approaches performed equally well, and significantly better than regressions or global or regional parameter averages. The results of Samuel et al. (2011) showed that also for the less dense stream-gauge network in Ontario (Canada) spatial proximity methods can perform more favourably than methods that use catchment characteristics, and coupling of



**Fig. 7.** Nash–Sutcliffe efficiency of predicting hydrographs in ungauged basins as a function of aridity ( $E_{PA}/P_A$ ), mean elevation and catchment area for different parameter regionalisation methods (Level 2 assessment). Lines connect median efficiencies for the same study. Boxes are 40–60 % quantiles, whiskers are 20–80 % quantiles.

spatial proximity and similarity methods provided better performance than regression and model-averaging approaches.

#### 4.4 How does model complexity impact performance?

To assess the effect of model complexity, the studies were grouped in terms of the number of model parameters that were regionalised (Fig. 6). The results indicate that, overall, there is no strong dependency of the performance on model complexity. The median of the performance for each group of models is around 0.65, with the exception of the group with 9–10 model parameters, which is lower. The largest variability (between 0.5 and 0.88) is found for models with 11–12 parameters. Studies that explored regionalisation performance of models with different complexity (Petheram et al., 2012; Chiew, 2010; Viney et al., 2009) suggest that whilst an increasing number of free parameters may lead to increased calibration performance, the difference in runoff-prediction

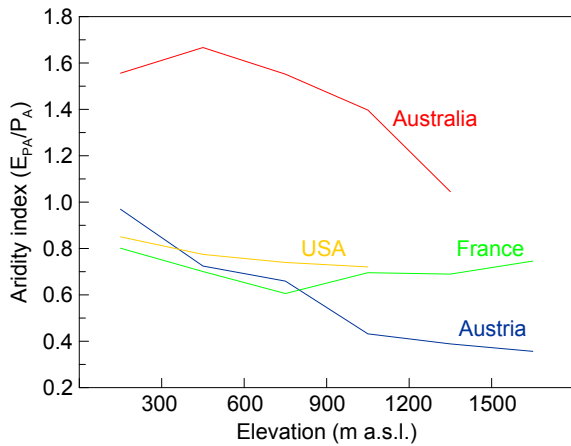
performance was small or negligible (Viney et al., 2009; Petheram et al., 2012). The results of Oudin et al. (2008) showed that simpler models may slightly outperform more complex models in the predictive mode.

It is also interesting to compare what regionalisation methods have been used in the different studies. The spatial proximity approach tends to be used for more complex models (more than 9 transferred parameters). There is a tendency of applying simpler models in arid and mixed arid and humid catchments, while in humid and cold regions more complex models have been used.

#### 4.5 To what extent does runoff prediction performance depend on climate and catchment characteristics?

The assessment of NSE predictive performance with respect to the four climate and catchment characteristics (Level 2 assessment) is presented in Fig. 7. The top panel shows a very

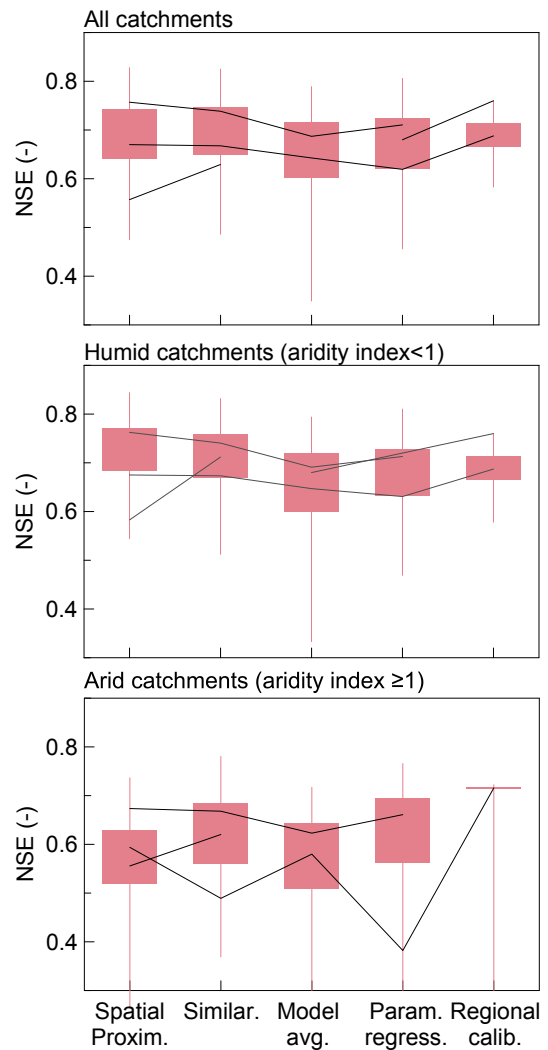




**Fig. 8.** Aridity index as a function of mean catchment elevation for the studies used in Level 2 assessment (Table 1). The aridity represents the median over all catchments in a particular elevation class.

clear pattern of decreasing performance with aridity index for catchments with an aridity larger than 0.6. The performance in the humid catchments is generally above 0.6, while it decreases to 0.5 or less in more arid catchments. It appears that in humid catchments, the rainfall–runoff processes are more linear, the hydrologic states tend to be less variable and the controls on runoff are spatially less variable, so a better performance would be expected. For the regional calibration method there is little dependency of the performance on aridity, but these studies are from Germany and Austria, where the catchments are never very arid.

The relationship between performance and elevation is more complex and depends on the region used for the assessment. There is a decrease of performance with increasing elevation in France (Oudin et al., 2008) and Australia (Zhang et al., 2009), and an increase of performance with increasing elevation in Austria (Parajka et al., 2005). These differences are due to the different dependencies of aridity with elevation (Fig. 8). Figure 8 summarizes the aridity in 320, 912, 76 and 210 catchments in Austria, France, USA and Australia, respectively. While in Austria the aridity is less than 0.5 in catchments above 900 m a.s.l. and strongly decreases with increasing elevation, in France the aridity index exceeds 0.75 and actually increases with elevation. In Australia the aridity index is always larger than in the other regions. This pattern is consistent for all regionalisation approaches, except regional calibration, which was applied in Germany (9 catchments) and Austria (320 catchments) where catchments are never very arid. The pattern for air temperature (not shown here) is similar with a clear tendency of decreasing performance with increasing temperature in Austria and the opposite in France. Interestingly, the model averaging method has a low median and large scatter of performance in colder catchments, which may be due to snow processes. Similarly,



**Fig. 9.** Nash–Sutcliffe efficiency of predicting hydrographs in ungauged basins for different regionalisation methods, stratified by aridity (Level 2 assessment). Top: all catchments. Centre: humid catchments (aridity index < 1). Bottom: arid catchments (aridity index ≥ 1). Lines connect median efficiencies for the same study. Boxes are 40–60 % quantiles, whiskers are 20–80 % quantiles.

as for other characteristics, the regional calibration is less sensitive to air temperature than the other methods.

The results in Fig. 7 show a very clear increase of the performance with catchment scale for all approaches and essentially all regions. The median performance is around 0.60 in small catchments (0–300 km<sup>2</sup>) and increases to around 0.80 for larger catchments. Also, the variability in performance between the catchments decreases with catchment scale, i.e. the large catchments never give a very low performance. An exception is a slight increase of performance variability for the spatial-proximity method in the largest catchments in Australia and France, but this is only for a small group of catchments. Overall, this very clear pattern of an

**Table 3.** Methods with the highest and lowest cross-validation performance of runoff predictions in ungauged basins. Arid relates to catchments with an aridity index  $> 1$ , humid to those with an aridity index  $\leq 1$ . Level 1 refers to an assessment of the average performance of studies, Level 2 to an assessment of the performance for individual catchments. Number of studies and catchments see Table 2.

Assessment	Catchments	Highest cross-validation performance	Lowest cross-validation performance
Level 1	All	*	*
Level 2	All	similarity, spatial proximity	model average
Level 2	Humid	spatial proximity, similarity	model average
Level 2	Arid	similarity, regression	model average

\* Indicates more than two methods with similar performance.

increase of the performance with catchment scale may be due to two reasons. The first is a trend for an increasing number of raingauges within a catchment as the catchment size increases. This trend likely reflects the relation between raingauge density relative to the correlation length scale of the rainfall (Schaake, 1981). The second may be related to the aggregation effect of runoff. As the catchment size increases, some of the hydrological variability is averaged out due to an interplay of space–time scale processes, which will improve hydrological simulation. Both effects are consistent with the scale effects of performance in gauged catchments (see, e.g. Merz et al., 2009; 2011; and Nester et al., 2011).

Figure 9 summarizes the performance for different regionalisation approaches, stratified by the aridity index. The total number of catchments is 1570, 1466, 1507, 1241 and 329 for spatial proximity, similarity, model averaging, parameter regression and regional calibration methods, respectively. The top, middle and bottom panels show the performance for all catchments in Table 2, and catchments with an aridity index below and above 1, respectively. Overall, in all catchments the spatial proximity and similarity methods perform slightly better than the parameter regression and model averaging approaches. In arid catchments, however, similarity and parameter regression tend to perform slightly better than spatial proximity and model averaging. These results suggest that climate characteristics more strongly impact the runoff-prediction performance in ungauged basins than the regionalisation method.

## 5 Discussion and conclusions

This paper has compared the performance of predicting daily runoff hydrographs in ungauged basins using conceptual runoff models with regionalised model parameters. Two kinds of assessments were performed; a Level 1 assessment, which constitutes a meta-analysis from the literature; and a Level 2 assessment, which analyses individual catchments in more detail. The results indicate that the Level 1 and Level 2 assessments are consistent while shedding light on different aspects of the prediction problem. The Level 1 assessment suggests that in humid and cold regions the performance of

predicting daily runoff hydrographs in ungauged basins tends to be better than in arid regions. All regionalisation methods analysed (spatial proximity, similarity, model averaging, parameter regression and regional calibration) show a similar performance with considerable scatter within each method. There is a tendency towards a somewhat lower performance of regressions than other methods in those studies that apply different methods in the same region. Studies with few catchments and studies with a large number of catchments tend to exhibit better performance than studies with an intermediate number of catchments. For studies with a large number of catchments (dense stream-gauge network) there is a tendency for spatial proximity and geostatistics to perform better than regression or regionalisation based on simple averaging of the model parameters. There is no clear dependence of the model performance on the number of model parameters regionalised. The Level 2 assessment suggests that the performance of all methods decreases with increasing aridity. The dependence of performance on elevation and air temperature differs by region and depends on how aridity varies with elevation and air temperature. The performance of all methods increases with catchment area. In humid conditions spatial proximity and similarity methods perform best, while in arid catchments similarity and parameter regression perform slightly better than the other methods (Table 3).

The predictive accuracy of different regionalisation methods was quantified in terms of Nash–Sutcliffe efficiency (NSE). Since it is a traditional performance measure used in hydrology, it has an advantage that almost all reviewed studies evaluate the predictive accuracy by using NSE (an exception is the study of Vogel (2005) that uses  $R^2$ ). On the other hand, NSE is a normalized skill score that measures runoff model performance relative to a baseline model, which is in this case mean of observed runoff values. This can lead to overestimation of NSE in catchments with strong seasonal runoff regime (see e.g. discussion in Schaeffli and Gupta, 2007). As pointed out in Gupta et al. (2009), a comparison of NSE across basins with different seasonality should therefore be interpreted with caution. For future comparative evaluations, we would hence suggest to use additional information and performance measures that will also enable

evaluation of different parts of runoff hydrographs, i.e. peaks, times to peak (Nester et al., 2011) or event recessions. This will help shed more light on the ability of different regionalisation methods to predict different hydrograph signatures across different runoff regimes.

Most of the studies analysed in this assessment applied lumped hydrologic models for runoff-hydrograph predictions. There are only a few distributed-modelling studies available for the assessment. Distributed models are harder to compare because of the added complexity in parameter estimation (see Blöschl et al., 2008). As distributed models are increasingly used for a range of purposes, it will be ever more important to also compare and cross-validate the prediction accuracy of distributed models in the future. An example of such comparisons is presented in the results of the Distributed Model Intercomparison Project (Smith et al., 2004, 2012), which focuses mainly on operational flood and water resources forecasting. A cross-validation in terms of prediction accuracy in ungauged catchments will help to further improve the understanding of how to effectively parameterize the climate/landscape relationship with runoff generation at different scales. Also, it may be useful, to compare distributed models not only on the basis of runoff data but also on the basis of other hydrological response variables such as snow patterns using snow models of different complexity (Blöschl and Kirnbauer, 1991, 1992; Blöschl et al., 1991; Nester et al., 2012).

The comparative assessment indicates two main implications for hydrological modelling. The first implication relates to the selection of model structure for runoff prediction. There are very few studies that have actually analysed what model structure would be appropriate for a particular catchment or landscape, yet it is likely that not all models will work equally well in all environments (e.g. Fenicia et al., 2011). Choice of model structure is usually guided by prior knowledge of the hydrologic system, the availability of data, and prior experience of the practitioner. This has led to a plurality of models being used. To avoid fragmentation and duplication, it might be valuable to group the world into classes of similar behaviour, based on some kind of classification scheme, and then to narrow down the number of models adopted. This will increase the experience with all such models and, through the sharing of this experience, it can lead to improvement of the models themselves and also improved predictive performance. van Werkhoven et al. (2009) found that an appropriate choice of the model structure simplified parameter estimation as the plausible parameter range is narrower if the model structure corresponds to the actual controls. The model structure of a catchment should hence be selected in the context of the particular hydro-climatic situation that controls the water balance through the soil–vegetation–atmosphere system. Depending on the setting, model structures should differ because the important hydrological processes may differ vastly between different landscapes.

The second implication stems from the fact that there is still a great potential of what can be learned from the synthesis of existing studies. Presently, however, it is not straightforward to compare the results of different studies. Many studies combine and aggregate results from different climate and physiographic settings and report only summary statistics of regionalisation performance and/or catchment characteristics. For future synthesis assessments, it would be useful to develop a universal protocol on reporting scientific results in the hydrological literature. In addition, the establishment of freely accessible data repositories to improve the synthesis and repeatability of studies would significantly contribute to making hydrology more coherent around the globe.

*Acknowledgements.* We are very grateful to S. Archfield, K. Hlavčová, N. McIntyre, L. Oudin, C. Petheram, L. Samaniego, D. Viviroli and Y. Zhang for providing the hydrologic data and model results of their studies. Without their very useful support the Level 2 analysis of this study would not have been possible. We would also like to thank the Austrian Science Fund (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

- Allasia, D. G., Da Silva, B. C., Collischinn, W., and Tucci, C. E. M.: Large basin simulation experience in South America, in: Predictions in Ungauged Basins: Promise and Progress (Proceedings of symposium S7 held during the Seventh IAHS Scientific Assembly at Foz do Iguaçu, Brazil, April 2005), IAHS Publ. 303, 360–370, 2006.
- Bastola, S., Ishidaira, H., and Takeuchi, K.: Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe, *J. Hydrol.*, 357, 188–206, 2008.
- Blöschl, G.: Rainfall-runoff modeling of ungauged catchments, in: *Encyclopedia of Hydrological Sciences*, edited by: Anderson, M. G., John Wiley & Sons, Chichester, Article 133, 2061–2080, 2005.
- Blöschl, G. and Kirnbauer, R.: Point snowmelt models with different degrees of complexity – internal processes, *J. Hydrol.*, 129, 127–147, 1991.
- Blöschl, G. and Kirnbauer, R.: An analysis of snow cover patterns in a small Alpine catchment, *Hydrol. Process.*, 6, 99–109, 1992.
- Blöschl, G. and Montanari, A.: Climate change impacts – throwing the dice?, *Hydrol. Process.*, 24, 374–381, doi:10.1002/hyp.6075, 2010.
- Blöschl, G., Gutknecht, D., and Kirnbauer, R.: Distributed snowmelt simulations in an Alpine catchment. 2. Parameter study and model predictions, *Water Resour. Res.*, 27, 3181–3188, 1991.
- Blöschl, G., Reszler, C., and Komma, J.: A spatially distributed flash flood forecasting model, *Environ. Modell. Softw.*, 23, 464–478, 2008.
- Boughton, W. and Chiew, F.: Estimating runoff in ungauged catchments from rainfall, PET and the AWBM model, *Environm.*

- Modell. Softw., 22, 476–487, doi:10.1016/j.envsoft.2006.01.009, 2007.
- Bulygina, N., McIntyre, N., and Wheeler, H.: Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis, *Hydrol. Earth Syst. Sci.*, 13, 893–904, doi:10.5194/hess-13-893-2009, 2009.
- Burn, D. H. and Boorman, D. B.: Estimation of hydrological parameters at ungauged catchments, *J. Hydrol.*, 143, 429–454, doi:10.1016/0022-1694(93)90203-L, 1993.
- Caballero, L. A., Easton, Z. M., Richards, B. K., and Steenhuis, T. S.: Evaluating the bio-hydrological impact of a cloud forest in Central America using a semi-distributed water balance model, *J. Hydrol. Hydromech.*, 61, 9–20, doi:10.2478/jhh-2013-0003, 2013.
- Chiew, F. H. S.: Lumped Conceptual Rainfall-Runoff Models and Simple Water Balance Methods: Overview and Applications in Ungauged and Data Limited Regions, *Geogr. Compass*, 4/3, 206–225, doi:10.1111/j.1749-8198.2009.00318.x, 2010.
- Cutore, P., Cristaudo, G., Campisano, A., Modica, C., Cancelliere, A., and Rossi, G.: Regional Models for the Estimation of Streamflow Series in Ungauged Basins, *Water Resour. Manage.*, 21, 789–800, 2007.
- Fenicia, F., Kavetski, D., and Savenije, 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.
- Fernandez, W., Vogel, R. M., and Sankarasubramanian, A.: Regional calibration of a watershed model, *Hydrol. Sci. J.*, 45, 689–707, doi:10.1080/02626660009492371, 2000.
- Goswami, M., O'Connor, K. M., and Bhattarai, K. P.: Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment, *J. Hydrol.*, 333, 517–531, 2007.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377, 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.
- Hlavčová, K., Szolgay, J., Čistý, M., Kóhnová, S., and Kalaš, M.: Estimation of mean monthly flows in small ungauged catchments, *Slovak Journal of Civil Engineering*, VIII, 21–29, 2000.
- Hundecha, Y., Ouarda, T. B. M. J., and Bárdossy, A.: Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the “spatial” structures of the parameters within a canonical physiographic-climatic space, *Water Resour. Res.*, 44, W01427, doi:10.1029/2006WR005439, 2008.
- Kim, U. and Kaluarachchi, J.: Application of parameter estimation and regionalization methodologies to ungauged basins of the Upper Blue Nile River Basin, Ethiopia, *J. Hydrol.*, 362, 39–56, doi:10.1016/j.jhydrol.2008.08.016, 2008.
- Kokkonen, T. S., Jakeman, A. J., Young, P. C., and Koivusalo, H. J.: Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina, *Hydrol. Process.*, 17, 2219–2238, doi:10.1002/hyp.1329, 2003.
- Kovacs, A., Honti, M., Eder, A., Zessner, M., Clement, A., and Blöschl, G.: Identification of phosphorus emission hotspots in agricultural catchments, *Sci. Total Environ.*, 433, 74–88, doi:10.1016/j.scitotenv.2012.06.024, 2012.
- Lebel, T., Bastin, G., Obled, C., and Creutin, J. D.: On the accuracy of areal rainfall estimation: a case study (Cevennes, France), *Water Resour. Res.*, 23, 2123–2134, 1987.
- Li, H., Zhang, Y., Chiew, F. H. S., and Xu, S.: Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index, *J. Hydrol.*, 370, 155–162, doi:10.1016/j.jhydrol.2009.03.003, 2009.
- McIntyre, N. R., Lee, H., Wheeler, H. S., and Young, A. R.: Tools and approaches for evaluating uncertainty in streamflow predictions in ungauged UK catchments, in: *Complexity and Integrated Resources Management*, edited by: Pahl-Wostl, C., Schmidt, S., and Jakeman, T., Proc. IEMSS International Congress, Osnabrueck, Germany, June, 2004.
- McIntyre, N., Lee, H., Wheeler, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, 41, W12434, doi:10.1029/2005WR004289, 2005.
- Merz, R. and Blöschl, G.: Regionalization of catchment model parameters, *J. Hydrol.*, 287, 95–123, 2004.
- 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.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models, *J. Hydrol.*, 10, 282–290, 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, 2011.
- Nester, T., Kirnbauer, R., Parajka, J., and Blöschl, G.: Evaluating the snow component of a flood forecasting model, *Hydrol. Res.*, 43, 762–779, 2012.
- Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, 44, 1–15, doi:10.1029/2007WR006240, 2008.
- Oudin, L., Kay, A., Andréassian, V., and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, *Water Resour. Res.*, 46, W11558, doi:10.1029/2009WR008887, 2010.
- Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalization methods for catchment model parameters, *Hydrol. Earth Syst. Sci.*, 9, 157–171, doi:10.5194/hess-9-157-2005, 2005.
- Parajka, J., Naeimi, V., Blöschl, G., Wagner, W., Merz, R., and Scipal, K.: Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale, *Hydrol. Earth Syst. Sci.*, 10, 353–368, doi:10.5194/hess-10-353-2006, 2006.
- Parajka, J., Blöschl, G., and Merz, R.: Regional calibration of catchment models: Potential for ungauged catchments, *Water Resour. Res.*, 43, 1–16, doi:10.1029/2006WR005271, 2007.
- Peel, M. C., Chiew, F. H. S., Western, A. W., and McMahon, T. A.: Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in 5 ungauged catchments. Report prepared for the National Land and Water Resources Audit, in: *Australian Natural Resources Atlas website*, 37 pp., available at : <http://anra.gov.au/topics/water/pubs/national/streamflow/streamflow.pdf> (last access: 2 May 2013),

- 2000.
- Petheram, C., Rustomji, P., Chiew, F. H. S., and Vleeshouwer, J.: Rainfall-runoff modelling in northern Australia: a guide to modelling strategies in the tropics, *J. Hydrol.*, 462–463, 28–41, 2012.
- Post, D. A.: Regionalizing rainfall-runoff model parameters to predict the daily streamflow of ungauged catchments in the dry tropics, *Hydrol. Res.*, 40, 433–444, 2009.
- Randrianasolo, A., Ramos, M. H., and Andréassian, V.: Hydrological ensemble forecasting at ungauged basins: using neighbour catchments for model setup and updating, *Adv. Geosci.*, 29, 1–11, doi:10.5194/adgeo-29-1-2011, 2011.
- Reichl, J. P. C., Western, A. W., McIntyre, N., and Chiew, F. H. S.: Optimisation of a similarity measure for estimating ungauged streamflow, *Water Resour. Res.*, 45, W10423, doi:10.1029/2008WR007248, 2009.
- Sachs, J. D. and McArthur, J. W.: The Millennium Project: a plan for meeting the Millennium Development Goals, *Lancet*, 365, 347–53, 2005.
- 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, *Hydrol. Earth Syst. Sci. Discuss.*, 10, 411–447, doi:10.5194/hessd-10-411-2013, 2013.
- Samaniego, L., Bardossy, A., and Kumar, R.: Streamflow prediction in ungauged catchments using copula-based dissimilarity measures, *Water Resour. Res.*, 46, W02506, doi:10.1029/2008WR007695, 2010a.
- Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resour. Res.*, 46, W05523, doi:10.1029/2008WR007327, 2010b.
- Samuel, J., Coulibaly, P., and Metcalfe, R.: Estimation of continuous streamflow in Ontario ungauged basins: comparison of regionalization methods, *J. Hydrol. Eng.*, 16, 447–459, doi:10.1061/(ASCE)HE.1943-5584.0000338, 2011.
- Schaake, J. C.: Raingage Network Density Requirement for Hydrologic Modeling, available at: <http://www.nws.noaa.gov/oh/mopex/raingage%20density%20requirement.htm> (last access: 2 May 2013), 1981.
- Schaeffli, B. and Gupta, H. V.: Do Nash values have value?, *Hydrol. Process.*, 21, 2075–2080, doi:10.1002/hyp.6825, 2007.
- Sefton, C. E. M. and Howarth, S. M.: Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales, *J. Hydrol.*, 211, 1–16, doi:10.1016/S0022-1694(98)00163-2, 1998.
- Seibert, J. and Beven, K. J.: Gauging the ungauged basin: how many discharge measurements are needed?, *Hydrol. Earth Syst. Sci.*, 13, 883–892, doi:10.5194/hess-13-883-2009, 2009.
- 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. J.*, 48, 857–880, 2003.
- 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., 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.
- Smith, M. B., Seo, D., Koren, V., Reed, S., Zhang, Z., Duan, Q., Moreda, F., and Cong, S.: The distributed model intercomparison project (DMIP): motivation and experiment design, *J. Hydrol.*, 298, 4–26, doi:10.1016/j.jhydrol.2004.03.040, 2004.
- Smith, M. B., Koren, V., Zhang, Z., Zhang, Y., Reed, S., Cui, Z., Moreda, F., Anderson, E., Cosgrove, B., Mizukami, N., and DMIP 2 participants: Results from the DMIP 2 Oklahoma experiments, *J. Hydrol.*, 418–419, 17–48, 2012.
- SSG: PUB Science Steering Group: PUB Science and Implementation Plan IAHS Decade on Predictions in Ungauged Basins (PUB): 2003–2012, International Association of Hydrological Sciences, 45 pp., available at: [http://pub.iahs.info/download/PUB\\_Science\\_Plan\\_V\\_5.pdf](http://pub.iahs.info/download/PUB_Science_Plan_V_5.pdf) (last access: 2 May 2013), 2003.
- Szolgay, J., Hlavčová, K., Kohnová, S., and Danihlík, R.: Regional Estimation of Parameters of a Monthly Water Balance Model, *J. Hydrol. Hydromech.*, 51, 256–273, 2003.
- Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria, *Hydrol. Earth Syst. Sci. Discuss.*, 10, 449–485, doi:10.5194/hessd-10-449-2013, 2013.
- Viney, N. R., Perraud, J., Vaze, J., Chiew, F. H. S., Post, D. A., and Yang, A.: The usefulness of bias constraints in model calibration for regionalisation of ungauged catchments, in: Proceed, 18th World IMACS/MODSIM Congress, Cairns, Australia 13–17 July, 2009.
- Viviroli, D., Mittelbach, H., Gurtz, J., and Weingartner, R.: Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part II: Parameter regionalisation and flood estimation results, *J. Hydrol.*, 377, 208–225, doi:10.1016/j.jhydrol.2009.08.022, 2009.
- Vogel, R.: Regional calibration of watershed models, in: *Watershed Models*, edited by: Singh, V. and Frevert, D., CRC Press, Boca Raton, FL, 47–71, 2005.
- Wagener, T. and Wheatler, H. S.: Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *J. Hydrol.*, 320, 132–154, doi:10.1016/j.jhydrol.2005.07.015, 2006.
- van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y.: Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models, *Adv. Water Resour.*, 32, 1154–1169, 2009.
- Young, A. R.: Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model, *J. Hydrol.*, 320, 155–172, doi:10.1016/j.jhydrol.2005.07.017, 2006.
- Zhang, Y. Q. and Chiew, F. H. S.: Relative merits of different methods for runoff predictions in ungauged catchments, *Water Resour. Res.*, 45, W07412, doi:10.1029/2008WR007504, 2009.
- Zvolenský, M., Kohnová, S., Hlavčová, K., Szolgay, J., and Parajka, J.: To the Application of Spatial Interpolation Methods for Estimation of Rainfall-Runoff Model Parameters in Ungauged Catchments, *Meteorolog. J.*, 10, 145–152, 2007.
- Zvolenský, M., Kohnová, S., Hlavčová, K., Szolgay, J., and Parajka, J.: Regionalisation of rainfall-runoff model parameters based on geographical location of gauged catchments, *J. Hydrol. Hydromech.*, 56, 176–189, 2008.