Journal of Hydrology 541 (2016) 1212–1220

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Influence of hydro-meteorological data spatial aggregation on streamflow modelling

Marc Girons Lopez [a,b,*], Jan Seibert [a,b,c]

[a] Department of Earth Sciences, Uppsala University, Villavägen 16, 752 36 Uppsala, Sweden
[b] Centre for Natural Disaster Science (CNDS), Villavägen 16, 752 36 Uppsala, Sweden
[c] Department of Geography, University of Zurich, Winterthurerstr. 190, CH-8057 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 1 March 2016
Received in revised form 10 August 2016
Accepted 14 August 2016
Available online 20 August 2016
This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief

Keywords:
Data resolution
Spatial aggregation
Discharge
Precipitation
Relative efficiency
Model calibration

ABSTRACT

Data availability is important for virtually any purpose in hydrology. While some parts of the world continue to be under-monitored, other areas are experiencing an increased availability of high-resolution data. The use of the highest available resolution has always been preferred and many efforts have been made to maximize the information content of data and thus improve its predictive power and reduce the costs of maintenance of hydrometric sensor networks. In the light of ever-increasing data resolution, however, it is important to assess the added value of using the highest resolution available.

In this study we present an assessment of the relative importance of hydro-meteorological data resolution for hydrological modelling. We used a case study with high-resolution data availability to investigate the influence of using models calibrated with different levels of spatially aggregated meteorological input data to estimate streamflow for different periods and at different locations. We found site specific variations, but model parameterizations calibrated using sub-catchment specific meteorological input data tended to produce better streamflow estimates, with model efficiency values being up to 0.35 efficiency units higher than those calibrated with catchment averaged meteorological data. We also found that basin characteristics other than catchment area have little effect on the performance of model parameterizations applied in different locations than the calibration site. Finally, we found that using an increased number of discharge data locations has a larger impact on model calibration efficiency than using spatially specific meteorological data. The results of this study contribute to improve the knowledge on assessing data needs for water management in terms of adequate data type and level of spatial aggregation.

1. Introduction

High resolution hydrometrical data is becoming increasingly available in many places (Isotta et al., 2014; Ivanov et al., 2004), enhancing the development of new approaches, such as data mining, that rely on large amounts of data to make predictions (Hall et al., 2002). For modelling purposes the general perception in the hydrological community is that it is always useful to use the best available resolution (Rauthe et al., 2013). However, with the increasing resolution of the available data this might not be so clear and the added value of using the most detailed data might be negligible. For instance, density limits in precipitation sensor networks have been found beyond which prediction skills are not improved (Girons Lopez et al., 2015; Xu et al., 2013). The issue of data-resolution requirements has also been investigated for discharge estimation in large scale basins. Findings suggest that the sensitivity to data resolution is dependent on the basin size (Bergström and Graham, 1998; Shrestha et al., 2006). At a global scale, though, data might not always be available at the desired resolution or accuracy for a specific application (Berne et al., 2004) or it might even be non-existent for the required location, which becomes a large issue due to the large geographical and scale-dependent variability and complexity of hydrological processes (Beven, 2000).

Either way hydrological data is crucial for improving societal resilience. Urban planning and fresh water supply as well as flood and drought management policies depend to a large extent on hydrological data availability (Lézine et al., 2011) and reducing the number of hydrometric stations can produce a negative impact

* Corresponding author at: Department of Earth Sciences, Uppsala University, Villavägen 16, 752 36 Uppsala, Sweden.

E-mail addresses: marc.girons_lopez@geo.uu.se (M. Girons Lopez), jan.seibert@hyd.uu.se (J. Seibert).

http://dx.doi.org/10.1016/j.jhydrol.2016.08.026
0022-1694/

on the accuracy of hydrological predictions on which such actions depend (Spence et al., 2007).

It is thus important to clearly define and quantify the data needs for different purposes including data resolution and density. Different approaches have been taken to address this issue, mainly by exploring ways to rationalize the design of hydrometric networks in order to maximize their information content and therefore their cost-benefit ratio (Beven et al., 2008). Relevant examples include findings on the value of hydrometric data (Hunger and Döll, 2008; Perrin et al., 2007), the effect of measurement location on prediction efficiency (Carrera et al., 1984), the importance of data resolution (Boyle et al., 2001; Faurès et al., 1995; Schuurmans and Bierkens, 2007), the use of remote sensing data and soft data sources to complement or replace hard data (Seibert and McDonnell, 2002; Smith et al., 2012; Wood et al., 2000), or even the use of poor quality data under certain circumstances (Boughton, 2006).

Even if there has been significant progress in the rationale behind the design of hydrometric networks much work is still needed. Mishra and Coulibaly (2009) identified several pressing issues that need to be addressed such as data processing and dissemination, increased spatial and temporal data resolution for improved sustainable water resource management under changing climatic conditions, including climatic extremes in the network design process, or the integration of different measurement approaches and techniques.

With this paper we aim to improve the knowledge on hydrometrical data needs for water management purposes by presenting an assessment of the importance of spatial resolution of meteorological data – precipitation and temperature – for streamflow simulation as well as its relative importance vis a vis other data types. More specifically we addressed the following questions:

i. Does using sub-catchment-specific meteorological time series instead of one series for the entire catchment give better model performance in the sub-catchments, if the model is calibrated for each sub-catchment?
ii. Does using sub-catchment-specific meteorological time series, instead of one series for the entire catchment, produce better model performance in the sub-catchments, if the model is calibrated for the outlet of the entire catchment only?
iii. Which modelling approach results in better model performances for sub-catchment runoff: (1) using one meteorological time series for the entire catchment, but considering sub-catchment discharge series for the calibration, or (2) using sub-catchment-specific meteorological time series, but considering only runoff at the catchment outlet for the calibration?

The third question evaluates the relative importance of hydrological and meteorological data for hydrological model calibration. These are relevant questions that can contribute to improving decisions on hydrometric measurement network deployment and management, especially in data-scarce areas.

2. Study area and dataset

The study was carried out in the Thur river basin, which is located in north-eastern Switzerland (Fig. 1). River Thur is a tributary of the Rhine and it is currently the largest non-regulated river in Switzerland. The river has a history of human intervention, mainly through the building of levees for flood defence purposes, but river restoration measures have been carried out in recent years to improve flood management as well as the ecological status of the river and its riparian zone (Woosley et al., 2007).

The drainage basin of river Thur has an area of 1696 km^2 and covers the front ranges of the Swiss Limestone Alps. The elevation of the basin ranges between 356 to 2503 m a.s.l. with an average elevation of 770 m a.s.l. The catchment headwaters are dominated by alpine limestone whereas the lowlands are mainly composed of Molasse-sandstones and Pleistocene unconsolidated sediments (PEER, 2010). Agriculture is the main type of land use in the lowlands while the highlands are dominated by pastures. Overall, approximately a quarter of the catchment area is covered by forest. Population is mainly composed of scattered settlements and some larger agglomerations, the largest of which are St. Gallen (72,000 inhabitants) and Frauenfeld (23,000 inhabitants).

The climate of the basin is classified as alpine and pre-alpine (Yang et al., 2007) and the flow regime of the river Thur is dominated by snowmelt (nivo-pluvial). The average flow at the outlet for the period 1904–2008 is 47 $m^3\,s^{-1}$, the 100-year high flow is 1071 $m^3\,s^{-1}$, and the low flow for the same return period is 3.16 $m^3\,s^{-1}$ (FOEN, http://www.hydrodaten.admin.ch/de/2044.html, 22 January, 2016). The average annual rainfall is 1350 mm and is distributed over the year with a peak during the summer months and a positive elevation gradient. Large precipitation events in the headwaters might cause a rapid discharge build-up in the basin due to the steep terrain and short concentration times.

Several data sources were used in this study. First, hourly gridded precipitation data was obtained from the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss). This is a high-resolution experimental dataset with a resolution of 1 km^2, called RdisaggH, which is the result of aggregating rain gauge and radar data (Wüest et al., 2010), and which is available for the period 2003–2010. A total of 29 rain gauges within the Thur basin were used for the precipitation correction procedure of RdisaggH (Fig. 1). The high temporal resolution provided by this product was important to capture the short concentration times in the smaller sub-basins and the combination of data sources ensured a good representation of the spatial variability of the precipitation.

Temperature data was also obtained from MeteoSwiss. The data consisted of daily gridded mean, minimum, and maximum temperature (named TmeanD, TminD, and TmaxD, respectively) based on an interpolation procedure between meteorological stations (Frei, 2014). Annual average values of potential evaporation for the Thur river catchment were obtained from the Hydrological Atlas of Switzerland (HADES) and distributed over the 12 months using a sine curve. A digital elevation model at a resolution of 25 m provided by the Swiss Federal Office of Topography (swisstopo) was used together with the gridded precipitation and temperature data to calculate elevation precipitation and temperature lapse rates respectively. The DEM was further used to compute the elevation range areas for each of the (sub) basins in the analysis as required by the hydrological model (see Section 3).

Discharge data from the Swiss Federal Office for the Environment (FOEN) were used for validating the results of this study. FOEN currently has nine operational stream discharge monitoring stations within the Thur River Basin providing hourly cumulative discharge information (Table 1). Among the different sub-basins, Rietholzbach – Mosnang (S9) is a research catchment that has been used for about 40 years as it is representative of pre-alpine countryside (Gurtz et al., 1999). The contributing areas as well as elevation range and land cover differ significantly from one station to another. All the sub-catchments have the annual peak flow during the spring months, but varying from March to June depending on their respective elevation. Despite differences in the annual mean runoff, flow duration curves for the different sub-catchments have a similar shape. This is supported by a two-sample Kolmogorov-Smirnov test among all the possible sub-catchment pairs. The test

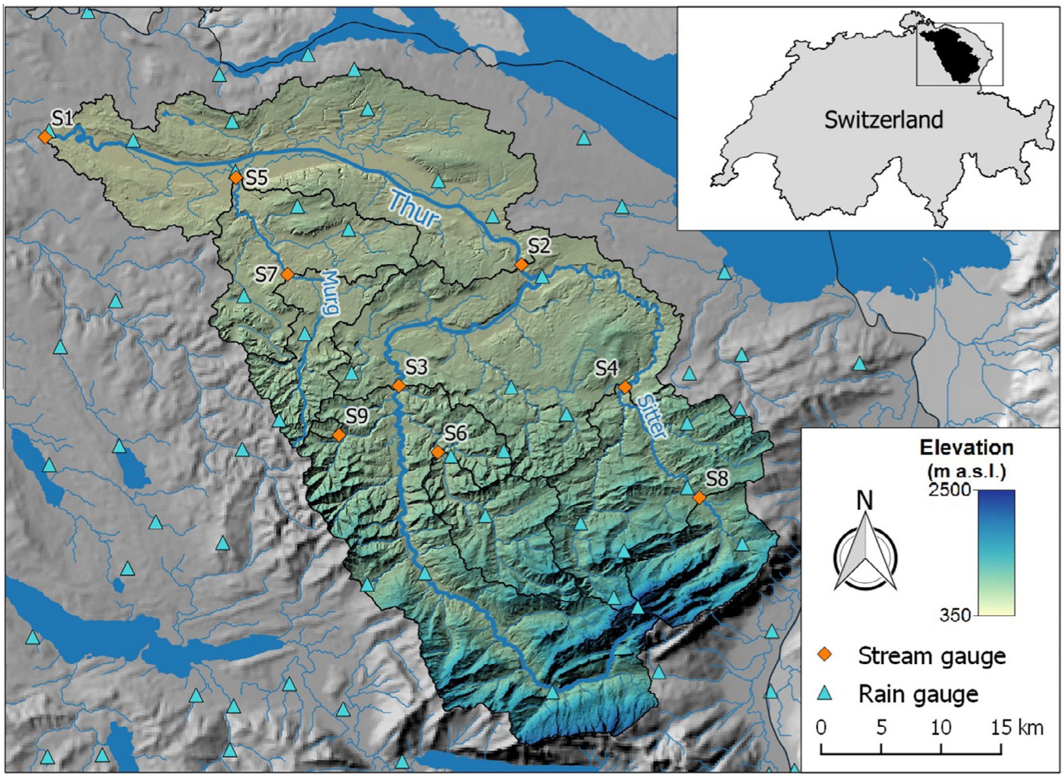

Fig. 1. Thur River Basin location and overview including the topography, stream network, and location of the FOEN stream gauging stations. Additionally, the location of the MeteoSwiss rain gauges used for developing the precipitation data products used for this study is also presented. *Source:* Swiss Federal Office of Topography.

Table 1
List of FOEN stream gauge stations with their corresponding location, identification code, contributing area, mean elevation and elevation range, and forest percentage.

Station code	River	Location	FOEN code	Area (km^2)	Elevation average (range) (m a.s.l.)	Forest (%)
S1	Thur	Andelfingen	2044	1696	770 (356–2503)	28
S2	Thur	Halden	2181	1085	910 (458–2503)	31
S3	Thur	Jonschwil	2303	493	1030 (534–2503)	35
S4	Murg	Frauenfeld	2386	212	580 (390–1036)	30
S5	Murg	Wängi	2126	78.9	650 (469–1036)	33
S6	Sitter	St. Gallen	2468	261	1040 (576–2503)	32
S7	Sitter	Appenzell	2112	74.2	1252 (769–2503)	32
S8	Necker	Mogelsberg	2374	88.2	959 (606–1531)	40
S9	Rietholzbach	Mosnang	2414	3.31	795 (673–940)	22

produced D values median, 10th, and 90th percentiles equaling 0.14, 0.07, and 0.27 respectively with a critical value of 0.19 for a significance level (α) of 0.05.

3. Methodology

In order to accomplish the objectives a number of steps were performed. First, the HBV model (Bergström, 1995; Lindström et al., 1997) was selected as hydrological model for performing the analysis. The selection was motivated by the model's simplicity, computational efficiency and low input data requirements. The HBV model is a semi-distributed rainfall-runoff model that was originally developed by the Swedish Meteorological and Hydrological Institute (SMHI) and which has been successfully applied in many areas of the world under a range of climatological and environmental conditions. Furthermore, the HBV model is also currently being used for operational flood forecasting in the river Rhine – of which the Thur is a tributary – by the German Federal Institute of Hydrology (BfG) and the Dutch Rijkswaterstaat (Berglöv et al., 2009).

MeteoSwiss, swisstopo, and FOEN perform quality controls before releasing their data products (see e.g. Wüest et al., 2010). Nonetheless, we performed an additional check for data completeness. All data products were found to be complete except for the gridded precipitation data, which was incomplete for 3% of the study period. These times were filled with corresponding daily data interpolated from rain gauge stations which were equally distributed among the time steps where hourly data were missing. These periods, including the following 5 days, were then flagged and not taken into account for the evaluation of the hydrological model. Due to the short concentration times of the Thur basin, the 5-day window after the missing data periods ensures that any potential missed precipitation events have no impact on the model performance. A consistent temporal resolution of one hour was enforced for the modelling. Temperature data was only available at daily resolution so a sine curve was fitted between the daily minimum and maximum values to represent the hourly temperature variability. Finally, the precipitation and temperature data were aggregated into two different time series for the analyses: (1) average values for the entire Thur river basin (referred to as

PT_{Thur} in this study) and (2) average values for each of the nine non-overlapping areas as defined by the different stream gauges (referred to collectively as PT_{sub}). Finally, precipitation and temperature lapse rates were obtained by computing the precipitation and temperature averages for the study period for each DEM cell in the Thur basin and applying a linear regression.

The pre-processed data were used to run the HBV model. We divided the study period into two equal independent periods: October 2004 – September 2007 and October 2007 – September 2010 (referred to as "period 1" and "period 2" respectively). The two periods were hydrologically comparable with almost similar flow duration curves. In total 36 set-ups were analysed, resulting from the combination of two input data aggregations (PT_{Thur} and PT_{sub}), two independent analysis periods, and nine calibration locations (FOEN stream gauges). The model was then calibrated for each of the 36 set-ups using a genetic algorithm followed by a fine tuning using a steepest gradient method (Seibert, 2000). Parameter uncertainty (Beven, 2006) was considered by using independent calibration trials. Each calibration trial consisted of 5000 model runs for the genetic algorithm, which includes stochastic elements and thus might result in different parameter sets if multiple, almost similar good sets exist, and 1000 runs for the subsequent fine-tuning, for which an implementation of Powell's quadratically convergent method was used. For each set-up we obtained 10 different parameter sets and computed median values of the model efficiency measures. This ensured robust results despite the potential problem of parameter uncertainty.

To address the aims of the study we performed modelling experiments in three different steps. In a first step we studied the effect of different levels of aggregation of meteorological input data on streamflow. To achieve this we compared the effect of using PT_{sub} and PT_{Thur} input data to calibrate the hydrological model in each of the nine calibration locations. We further investigated the geographic variability of the effect of the meteorological input by subtracting the performance obtained from calibrating the model with PT_{Thur} data from that of calibrating it using PT_{sub} data. Obtaining positive values would thus imply that better performance is achieved by calibrating the model with the PT_{sub} dataset at the specific location.

In a second step we studied the transferability of meteorological data across spatial scales and locations. We considered two different cases: in the first only PT_{Thur} was used and in the second only PT_{sub} was used. We calibrated the model using both input data aggregates at the Thur River Basin outlet (S1) and then transferred and evaluated the model performance at the different internal sites (stations S2 to S9). Thereafter we investigated possible correlations between the performance of the model and several catchment characteristics such as contributing area, mean elevation, and forest area fraction.

In the last step we studied the relative importance of meteorological and streamflow data. We considered two different cases again: for the first case we calibrated the model at the basin outlet (S1) using PT_{sub} data. Conversely, for the second case we used PT_{Thur} data and streamflow data for the nine stream gauging locations. For this case, we calibrated the model at the nine available stream gauges but using only PT_{Thur} input data. This approach allowed us to assess which of the two data types (meteorological data or discharge data) has a larger impact on the model performance for streamflow prediction.

The different analyses were evaluated using two different model performance metrics. First, the model efficiency, R_{eff} (Nash and Sutcliffe, 1970), which is one of the most widely used metrics in hydrological modelling, was used. R_{eff} is a good measure for assessing model performance for individual set-ups but it is not equally informative for assessing the transference of parameter sets across a number of set-ups involving different independent catchments and time periods. To accomplish this we introduced a new metric, the scaled model efficiency (R^*_{eff}) which scales R_{eff} based on an upper and a lower benchmark for each catchment and period. Such a relative performance measure makes the comparison of values more meaningful as it relates model performances to (i) the best achievable performance and (ii) the performance achieved without any prior information for any given set-up. In this way the performance of a certain model parameterization is evaluated based on what would be achiveable relative to the best and worst possible cases. A (normal) model efficiency of, for instance, 0.7 could be considered to be reasonably good if the best achievable value was 0.72 whereas it should be considered poor if the best achievable value was 0.85. The scaled efficiency makes efficiency values comparable in such cases.

We defined the scaled model efficiency as the model efficiency relative to the best possible efficiency, R^{max}_{eff}, that can be achieved for a certain catchment and time period, and the efficiency obtained on average by randomly chosen parameter value sets, R^{min}_{eff} (Eq. (1)). The upper benchmark R^{max}_{eff} was determined by the model efficiency that was achieved when calibrating the HBV model for the respective catchment, data type, and period. The estimation of the lower benchmark R^{min}_{eff}, is less obvious. Here we assumed no prior information and started with simulating runoff time series with 500 parameter sets randomly chosen within typical parameter ranges (Seibert, 1999). The ensemble of these 500 simulated time series was then averaged into a single mean time series and the model efficiency of this ensemble mean time series was used as R^{min}_{eff}. Preliminary tests showed that this procedure resulted in higher values for R^{min}_{eff} than the average of the model efficiencies for the individual simulated time series, which confirmed that the ensemble mean usually outperforms the individual ensemble members (e.g., Seibert and Beven, 2009).

$$R^*_{eff} = \frac{R_{eff} - R^{min}_{eff}}{R^{max}_{eff} - R^{min}_{eff}} \tag{1}$$

Similar to R_{eff}, the range of R^*_{eff} is (−inf, 1], but the values one and zero have a different meaning: a value of one indicates a performance as good as the best possible fit (i.e., performance when calibrated) while a value of zero corresponds to a performance similar to that of the lower benchmark using random parameter values.

4. Results

The general purpose of studying the importance of spatial resolution of meteorological data for streamflow simulation was subdivided into smaller aims in order to provide an adequate framework for obtaining relevant information. In the following sections we present significant results for each of the three specific steps.

4.1. Effect of spatial meteorological data resolution on streamflow estimation

The hydrological model performance when calibrated with PT_{Thur} data compared to when calibrated with PT_{sub} data is presented for the different calibration locations and study periods (Fig. 2). Each point represents a comparison between the same model set-up (location and period) using different meteorological input data. A large range of efficiency values is observed – between 0.4 and 0.8 – but most points fall above the dashed identity line, which indicates that a better performance was achieved when using PT_{sub} data for calibrating the model. No clear differences could be observed between the different calibration and validation periods. The same trend could be observed for both efficiency metrics except for the consistently higher values observed when using

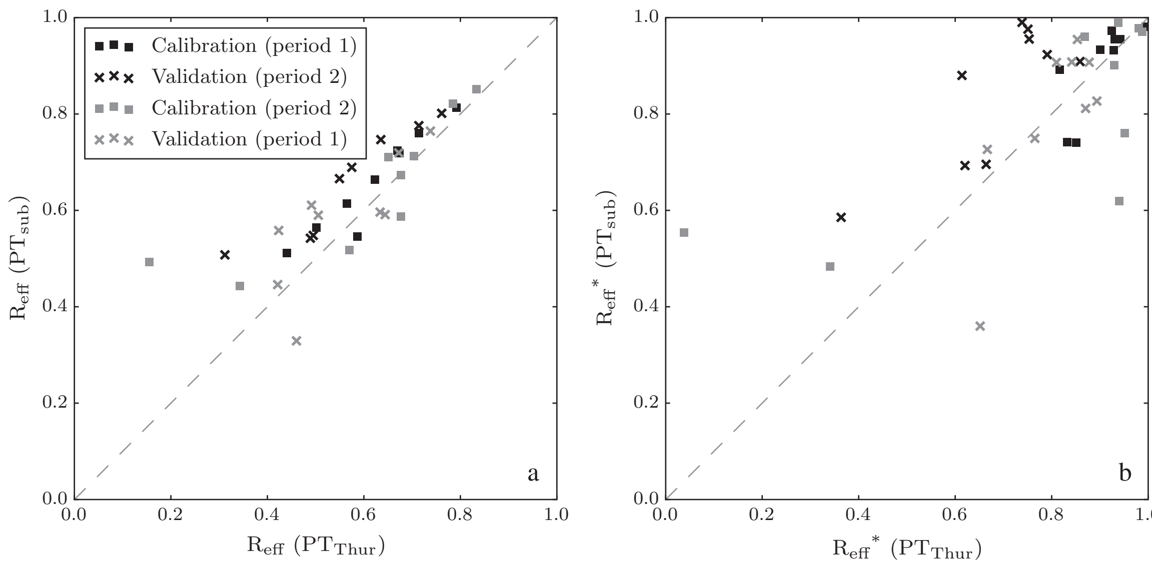

Fig. 2. Model performance comparison between model calibrations in all sub-basins using PT_{Thur} input data (x-axis) respect to using PT_{sub} input data (y-axis). Each point represents one of the 36 set-ups considered in the study. Two different efficiency measures were considered: R_{eff} (a) and R^*_{eff} (b). The dashed lines represent an equal performance for both data aggregates.

R^*_{eff}, which are generally above 0.8 and slightly more spread out on both sides of the dashed line.

To illustrate the spatial distribution of the results from Fig. 2 we distributed the data points according to their corresponding calibration location (Fig. 3). The stations located at the lower Thur River (S1 and S2) as well as those located at the Sitter River (S6 and S7) had generally positive values – up to +0.35 units – across the different calibration and validation periods indicating that using PT_{sub} data produces better results. Conversely, the stations located at the Murg River and those at the upper reaches of the Thur River showed a larger variability and ranging from +0.10 to −0.15 (reaching even values of up to −40 when using the R^*_{eff} measure), indicating that the data aggregate producing the best results depends on the analysis period. Both efficiency measures generally agree on the sign, with R^*_{eff} providing higher values (either positive or negative). In some cases, though, the type of input data producing better results depends on the selected efficiency measure, most notably in station S4 when calibrating in the first period (Fig. 3a).

4.2. Effect of spatial meteorological data resolution on streamflow estimation at different locations

The hydrological model performance when calibrated at the outlet (S1) with PT_{Thur} data compared to PT_{sub} data and evaluated at the internal sites (S2 to S9) is presented for the different study periods (Fig. 4). The overall distribution of efficiencies is similar to that of Fig. 2 showing that model calibrations performed with PT_{sub} input data perform better at the internal sites than those calibrations made using PT_{Thur} data. The evaluation using the R^*_{eff} metric shows that efficiency values are consistently higher compared to R_{eff} and that they are also more spread out along both sides of the equal performance line.

We then investigated the existence of possible correlations between the performance of using the model calibrated at the basin outlet and evaluated at the internal sites and several characteristics of the different catchments such as contributing area, mean elevation and forest area fraction (Fig. 5). Other catchment characteristics such as mean temperature and discharge, slope, urban area or open field area were also investigated but since they do not bring significantly different results they are not presented here. No clear trends were observed for the parameters that were analysed using neither efficiency measures. However, slightly positive trends are observed for contributing area. The location with the lowest overall model performance for all the analysed catchment characteristics is S8, which has one of the smallest contributing areas and the highest mean elevation. When performing the analysis using the R^*_{eff} metric the same slightly positive trends can be observed but with flatter slopes since most efficiency values are above 0.4 and median values above 0.7.

4.3. Relative importance of meteorological data and streamflow data

The hydrological model performance when calibrated at the outlet (S1) with PT_{sub} data compared to when calibrated at the internal sites (S2 to S9) using PT_{Thur} data is presented for the different study periods (Fig. 6). Most points fall below the equal performance line, pointing at a larger importance of local calibration based on sub-catchment streamflow data compared to correct areal precipitation for the individual sub-catchments. For the R^*_{eff} efficiency measure the same tendency can be observed as all cases resulted in high efficiency values when using PT_{Thur} and Q_{sub} data (>0.8) also for cases with low efficiency using PT_{sub} and Q_{Thur} data (<0.6).

5. Discussion

The results show that, in general, using PT_{sub} data produces better results than using PT_{Thur} data for streamflow simulation, which is in agreement with the findings of e.g. Schuurmans and Bierkens (2007). This trend can be observed when transferring calibrated models to independent periods (Fig. 2) and to different locations (Fig. 4). The differences between using PT_{Thur} and PT_{sub} data are however relatively small as most values are very close to the equal performance line. Performance variability range is, in fact, much larger when focusing on a single input data aggregate. For instance, model efficiency ranges between around 0.3 to 0.8 when using either data aggregate (Fig. 2).

Geographical location (Fig. 3) as well as contributing area (Fig. 5a) appear to be the most relevant factors affecting model performance. Efficiency differences between using PT_{sub} and PT_{Thur} range between +0.35 and −0.15 efficiency units for different locations, generally favouring the use of PT_{sub} (Fig. 3). This behaviour is

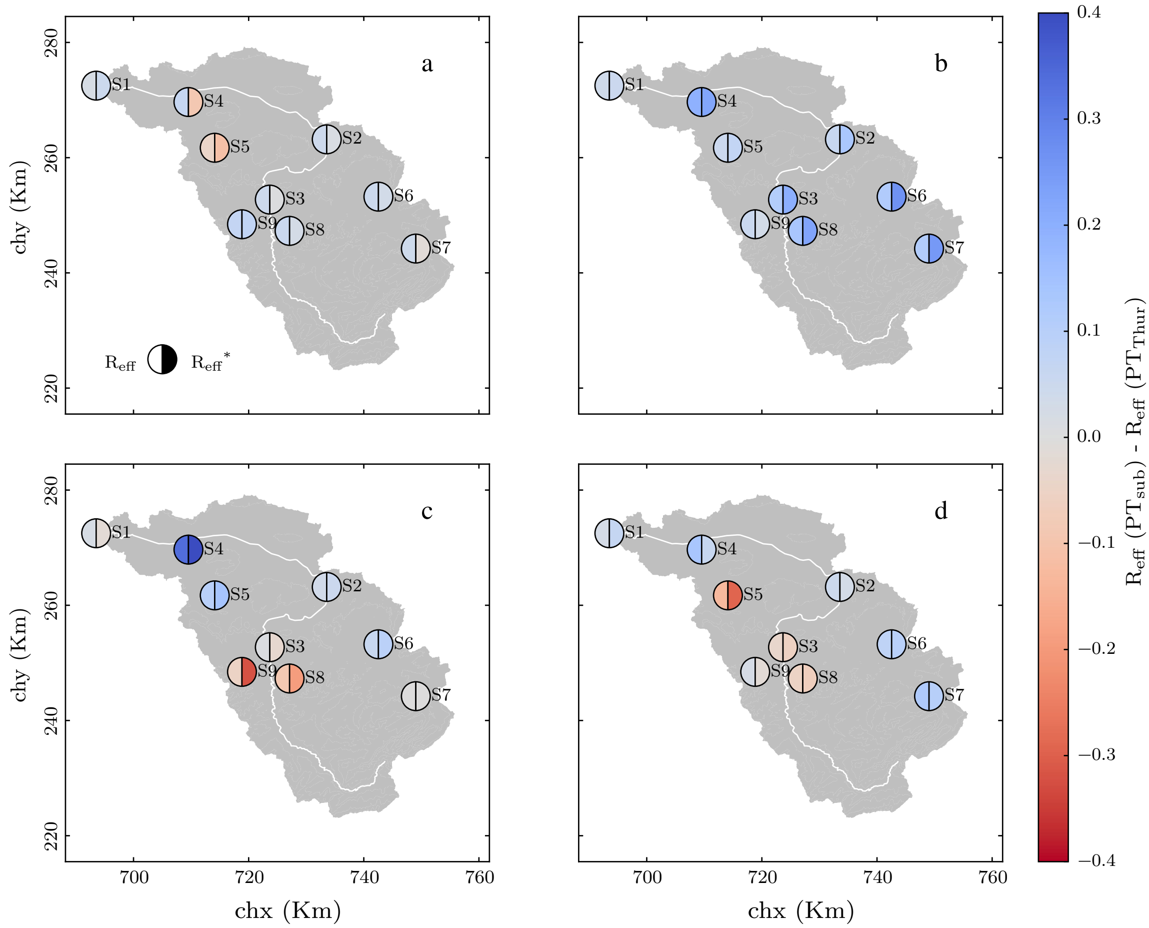

Fig. 3. Geographical distribution of the model results from Fig. 2 for the different set-ups: (a) calibration in period 1, (b) validation in period 2, (c) calibration in period 2, and (d) validation in period 1. Two different performance measures are considered: R_{eff} on the left side of the circle and R^*_{eff} on the right side. The colour and shade indicate the relative model performance when calibrated with the different input data aggregations (PT_{sub} and PT_{Thur}). The stronger the shade of the colour, the more pronounced the performance difference. Map coordinates – chx and chy – correspond to the CH1903 Swiss coordinate system (swisstopo, 2008). The labels correspond to the FOEN stream gauges within the Thur River Basin used for the calibrations (Table 1).

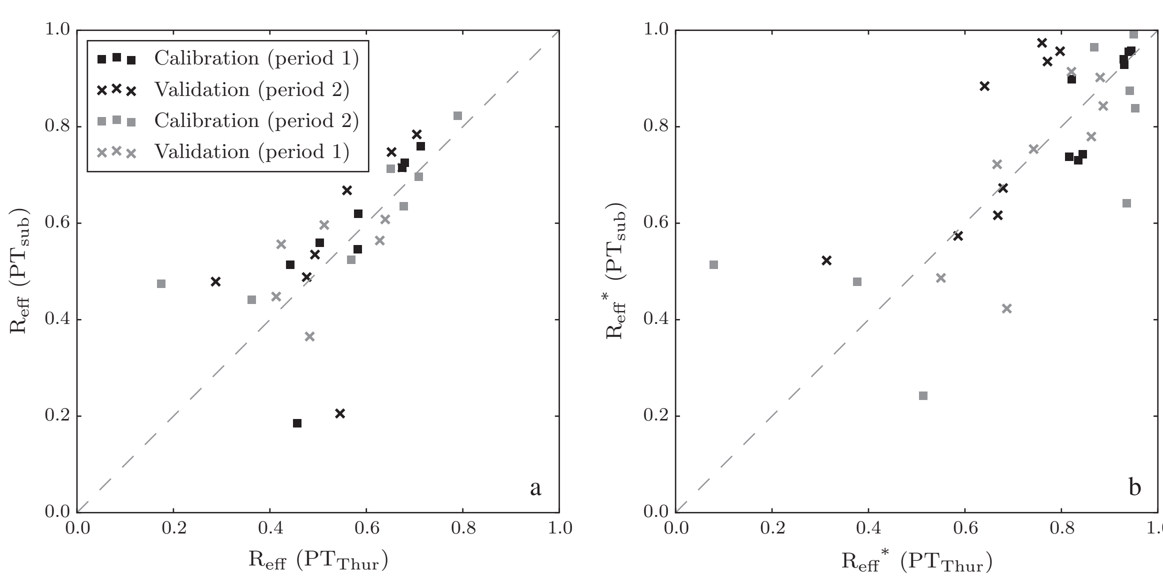

Fig. 4. Comparison of model performance when calibrating at the Thur river basin outlet using different data aggregates and then being transferred to the internal sites. Each point represents one of the 36 set-ups considered in the study. Two alternative efficiency measures were used: R_{eff} (a) and R^*_{eff} (b). The dashed lines represent an equal performance for both data aggregates.

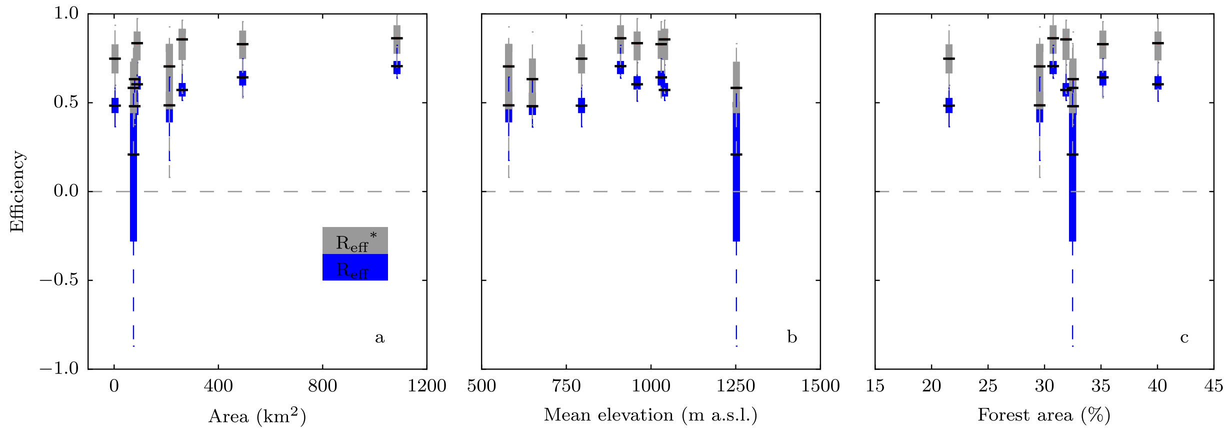

Fig. 5. Box plots showing the performance of parameter sets calibrated at the basin outlet and transferred to the internal sites for different attributes of the sub-basins, namely contributing area (a), mean elevation (b), and forest area fraction (c). Two performance measures were considered: R_{eff} (blue boxes) and R^*_{eff} (grey boxes). Horizontal black lines represent median performances. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

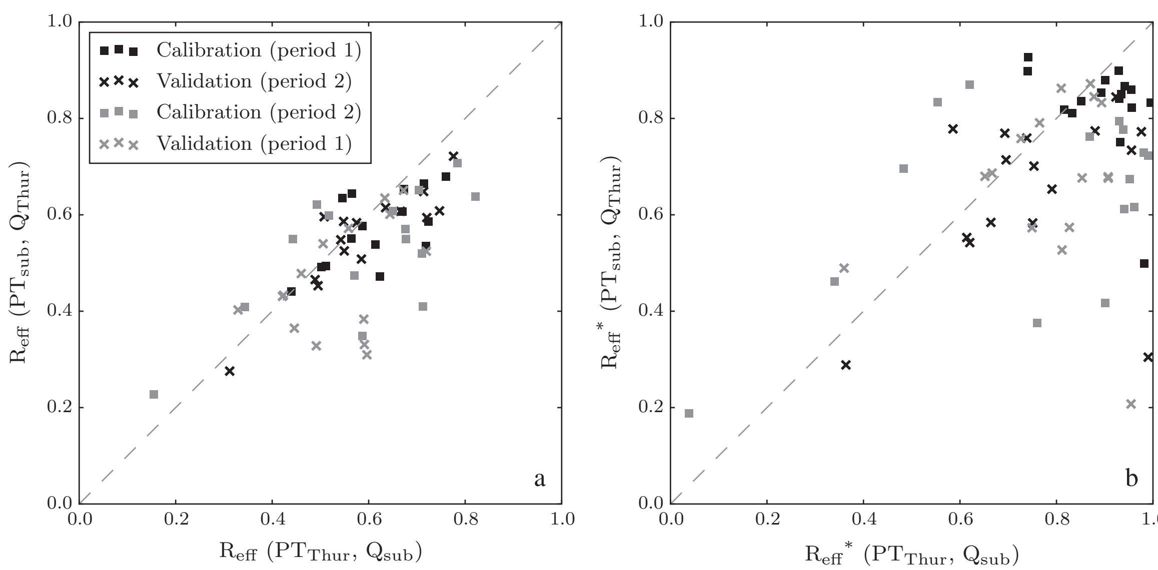

Fig. 6. Model performance comparison between two different data availability cases: PT_{Thur} and Q_{sub} (x-axis), and PT_{sub} and Q_{Thur} (y-axis). Model performance values were obtained using the R_{eff} (a) and R^*_{eff} (b) efficiency measures. The dashed lines represent an equal performance for the different data aggregate combinations.

however very site specific: while for some of the catchments using PT_{sub} produces consistently better results (e.g. S2, S7), for some others it is not clear which data aggregate produces the best results as it varies for different periods (e.g. S5, S8). Even if these different sub-catchments have different overall settings and characteristics, the effect of different basin characteristics is difficult to quantify as no clear trends can be identified (Fig. 5). That being said, contributing area presents a slight positive trend, meaning that parameter sets calibrated for the basin outlet can be used to successfully predict streamflow in the larger sub-catchments, which is in accordance with previous findings on the effect of data resolution on different spatial scales (Parajka et al., 2013; Shrestha et al., 2006). Regarding elevation differences, reverse trends can be obtained by either including or removing S8, making the interpretation of the impact of this parameter unclear. Understanding the sign of the impact of catchment mean elevation on runoff estimation is however important for water management as related characteristics such as increasing precipitation volumes with elevation (through the lapse rate) and steeper slopes have a large influence on the magnitude and concentration times of high flow events.

The larger influence of geographical location and basin characteristics over input data resolution might be a consequence of the geographical and scale dependent variability of hydrological processes (Beven, 2000). Even if the study area was constrained to a single mid-sized basin and the study period was relatively short, the wide topographical and climatological range of the basin may influence the performance obtained when transferring model parameterizations to a significant extent. Climatological and topographical characteristics may partially explain the performance variability for different sub-basins in Fig. 2.

An immediate consequence of this reasoning is that PT_{Thur} data can, under certain circumstances, produce adequate streamflow estimations if compared to using PT_{sub} data. This is especially true for some of the locations (e.g. S5, S8) and it is further supported by the R^*_{eff} values. Most points in Fig. 2 as well as in Fig. 4 have R^*_{eff} values larger than 0.8 for both PT_{Thur} and PT_{sub} input, meaning that models calibrated at a different location and period perform only marginally worse than those calibrated for that specific set-up, regardless of the input data resolution.

In addition to the importance of basin characteristics, streamflow data was found to be more significant for calibrating the HBV model than meteorological input data (Fig. 6). Higher model efficiency values were obtained by using different discharge data at more locations than by using different meteorological data aggregates. This is especially relevant for the sub-catchments with steepest slopes – leading to shorter concentration times – and highly variable precipitation patterns. Precipitation variability is especially important at the temporal resolution of the analysis (Faurès et al., 1995; Frei and Schär, 1998) but it is found to have a smaller influence on model calibration and operation than discharge information.

The introduction of the scaled model efficiency, R^*_{eff}, allowed a clearer comparison of the relative model performances obtained by using different modelling set-ups with parameter sets derived at different catchments and time periods compared to the standard model efficiency, R_{eff}. In contrast to R_{eff}, which generally provided near-neutral comparisons between the different data aggregations, as indicated by the faded colours in Fig. 3, R^*_{eff}, provided clearer indications of the most suitable data aggregation. Only in a few cases, such as for S4 when calibrating for period 1 (Fig. 3a), it indicated a different optimal data aggregate than R_{eff}. Additionally, most of the set-ups analysed in this study resulted in high R^*_{eff} values with the exception of some few cases (Fig. 6). As a result, using different input data resolutions as well as transferring model calibrations to other periods and locations does not seriously compromise the model performance in most cases for the study location.

Overall, this analysis helped to improve the understanding of data type and resolution needs for adequately estimating streamflow. This is relevant for maximizing the cost-efficiency of hydrometric networks and shows that the expected benefits of increased data resolution need to be assessed before investing in them. The findings also contribute to constrain the relevant factors impacting model performance when transferring calibrated models in time and space and to quantify the relative importance of streamflow and meteorological data for designing hydrometric sensor networks.

6. Conclusions

This study focused on assessing the level of spatial aggregation of meteorological input data for streamflow estimation as well as its relative importance compared to other data types for calibrating hydrological models such as streamflow data. The main findings can be summarized as follows:

i. In general, sub-catchment-specific meteorological data contributes to improve streamflow estimation performance in the different sub-catchments. Efficiency improvement magnitudes range between +0.35 and −0.15 units and are site specific.
ii. Sub-catchment-specific meteorological data produces better streamflow estimations for the internal sites when the model is calibrated for the catchment outlet only. Streamflow seems to be most successfully estimated for sub-catchments with larger contributing areas. No trends could be identified for other catchment characteristics.
iii. Better model performances are achieved when using sub-catchment discharge series for model calibration compared to using sub-catchment-specific meteorological data series.
iv. The use of the scaled efficiency measure for model evaluation, R^*_{eff}, can give valuable insight into the potential of applying model parameterizations in different locations and periods.

Acknowledgements

The authors thank FOEN, MeteoSwiss and swisstopo for providing the necessary data for the study. Marc Vis is thanked for support with HBV. Tracy Ewen provided helpful comments to improve the language.

References

Berglöv, G., German, J., Gustavsson, H., Harbman, U., Johansson, B., 2009. Improvement HBV Model Rhine in FEWS: Final Report. Norrköping, Sweden.
Bergström, S., 1995. The HBV model. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, CO, pp. 443–476.
Bergström, S., Graham, L.P., 1998. On the scale problem in hydrological modelling. J. Hydrol. 211, 253–265. http://dx.doi.org/10.1016/S0022-1694(98)00248-0.
Berne, A., Delrieu, G., Creutin, J.-D., Obled, C., 2004. Temporal and spatial resolution of rainfall measurements required for urban hydrology. J. Hydrol. 299, 166–179. http://dx.doi.org/10.1016/j.jhydrol.2004.08.002.
Beven, K.J., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36. http://dx.doi.org/10.1016/j.jhydrol.2005.07.007.
Beven, K.J., 2000. Uniqueness of place and process representations in hydrological modelling. Hydrol. Earth Syst. Sci. 4, 203–213. http://dx.doi.org/10.5194/hess-4-203-2000.
Beven, K.J., Smith, P.J., Freer, J.E., 2008. So just why would a modeller choose to be incoherent? J. Hydrol. 354, 15–32. http://dx.doi.org/10.1016/j.jhydrol.2008.02.007.
Boughton, W., 2006. Calibrations of a daily rainfall-runoff model with poor quality data. Environ. Model. Softw. 21, 1114–1128. http://dx.doi.org/10.1016/j.envsoft.2005.05.011.
Boyle, D.P., Gupta, H.V., Sorooshian, S., Koren, V., Zhang, Z., Smith, M., 2001. Toward improved streamflow forecasts: value of semidistributed modeling. Water Resour. Res. 37, 2749–2759. http://dx.doi.org/10.1029/2000WR000207.
Carrera, J., Usunoff, E., Szidarovszky, F., 1984. A method for optimal observation network design for groundwater management. J. Hydrol. 73, 147–163. http://dx.doi.org/10.1016/0022-1694(84)90037-4.
Faurès, J.-M., Goodrich, D.C., Woolhiser, D.A., Sorooshian, S., 1995. Impact of small-scale spatial rainfall variability on runoff modeling. J. Hydrol. 173, 309–326. http://dx.doi.org/10.1016/0022-1694(95)02704-S.
Frei, C., 2014. Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances. Int. J. Climatol. 34, 1585–1605. http://dx.doi.org/10.1002/joc.3786.
Frei, C., Schär, C., 1998. A precipitation climatology of the Alps from high-resolution rain-gauge observations. Int. J. Climatol. 18, 873–900. http://dx.doi.org/10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9.
Girons Lopez, M., Wennerström, H., Nordén, L.-Å., Seibert, J., 2015. Location and density of rain gauges for the estimation of spatial varying precipitation. Geogr. Ann. Ser. A Phys. Geogr. 97, 167–179. http://dx.doi.org/10.1111/geoa.12094.
Gurtz, J., Baltensweiler, A., Lang, H., 1999. Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins. Hydrol. Process. 13, 2751–2768. http://dx.doi.org/10.1002/(SICI)1099-1085(19991215)13:17<2751::AID-HYP897>3.0.CO;2-O.
Hall, M.J., Minns, A.W., Ashrafuzzaman, A.K.M., 2002. The application of data mining techniques for the regionalisation of hydrological variables. Hydrol. Earth Syst. Sci. 6, 685–694. http://dx.doi.org/10.5194/hess-6-685-2002.
Hunger, M., Döll, P., 2008. Value of river discharge data for global-scale hydrological modeling. Hydrol. Earth Syst. Sci. 12, 841–861. http://dx.doi.org/10.5194/hess-12-841-2008.
Isotta, F.A., Frei, C., Weilguni, V., Percec Tadic, M., Lassègues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S.M., Munari, M., Micheletti, S., Bonati, V., Lussana, C., Ronchi, C., Panettieri, E., Marigo, G., Vertacnik, G., 2014. The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. Int. J. Climatol. 34, 1657–1675. http://dx.doi.org/10.1002/joc.3794.
Ivanov, V.Y., Vivoni, E.R., Bras, R.L., Entekhabi, D., 2004. Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: a fully-distributed physically-based approach. J. Hydrol. 298, 80–111. http://dx.doi.org/10.1016/j.jhydrol.2004.03.041.
Lézine, A.M., Hély, C., Grenier, C., Braconnot, P., Krinner, G., 2011. Sahara and Sahel vulnerability to climate changes, lessons from Holocene hydrological data. Quat. Sci. Rev. 30, 3001–3012. http://dx.doi.org/10.1016/j.quascirev.2011.07.006.
Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. J. Hydrol. 201, 272–288. http://dx.doi.org/10.1016/S0022-1694(97)00041-3.
Mishra, A.K., Coulibaly, P., 2009. Developments in hydrometric network design: a review. Rev. Geophys. 47. http://dx.doi.org/10.1029/2007RG000243.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I: a discussion of principles. J. Hydrol. 10, 282–290. http://dx.doi.org/10.1016/0022-1694(70)90255-6.
Parajka, J., Viglione, A., Rogger, M., Salinas, J.L., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins-Part 1: runoff-hydrograph studies. Hydrol. Earth Syst. Sci. 17, 1783–1795. http://dx.doi.org/10.5194/hess-17-1783-2013.
PEER, 2010. River Thur, Switzerland – Hydrological Observatory Description. Leipzig, Germany.
Perrin, C., Oudin, L., Andreassian, V., Rojas-Serna, C., Michel, C., Mathevet, T., 2007. Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models. Hydrol. Sci. J. 52, 131–151. http://dx.doi.org/10.1623/hysj.52.1.131.
Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., Gratzki, A., 2013. A Central European precipitation climatology – Part I: generation and validation of a high-

resolution gridded daily data set (HYRAS). Meteorol. Zeitschrift 22, 235–256. http://dx.doi.org/10.1127/0941-2948/2013/0436.

Schuurmans, J.M., Bierkens, M.F.P., 2007. Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model. Hydrol. Earth Syst. Sci. 11, 677–693. http://dx.doi.org/10.5194/hessd-3-2175-2006.

Seibert, J., 1999. Regionalisation of parameters for a conceptual rainfall-runoff model. Agric. For. Meteorol. 98–99, 279–293.

Seibert, J., 2000. Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. Hydrol. Earth Syst. Sci. 4, 215–224.

Seibert, J., Beven, K.J., 2009. Gauging the ungauged basin: how many discharge measurements are needed? Hydrol. Earth Syst. Sci. 13, 883–892. http://dx.doi.org/10.5194/hessd-6-2275-2009.

Seibert, J., McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. Water Resour. Res. 38. http://dx.doi.org/10.1029/2001WR000978.

Shrestha, R., Tachikawa, Y., Takara, K., 2006. Input data resolution analysis for distributed hydrological modeling. J. Hydrol. 319, 36–50. http://dx.doi.org/10.1016/j.jhydrol.2005.04.025.

Smith, R.A.E., Bates, P.D., Hayes, C., 2012. Evaluation of a coastal flood inundation model using hard and soft data. Environ. Model. Softw. 30, 35–46. http://dx.doi.org/10.1016/j.envsoft.2011.11.008.

Spence, C., Saso, P., Rausch, J., 2007. Quantifying the impact of hydrometric network reductions on regional streamflow prediction in Northern Canada. Can. Water Resour. J. 32, 1–20. http://dx.doi.org/10.4296/cwrj3201001.

Wood, S.J., Jones, D.A., Moore, R.J., 2000. Accuracy of rainfall measurement for scales of hydrological interest. Hydrol. Earth Syst. Sci. 4, 531–543. http://dx.doi.org/10.5194/hess-4-531-2000.

Woosley, S., Capelli, F., Gonser, T., Hoehn, E., Hostmann, M., Junker, B., Paetzold, A., Roulier, C., Schweizer, S., Tiegs, S.D., Tockner, K., Weber, C., Peter, A., 2007. A strategy to assess river restoration success. Freshw. Biol. 52, 752–769. http://dx.doi.org/10.1111/j.1365-2427.2007.01740.x.

Wüest, M., Frei, C., Altenhoff, A., Hagen, M., Litschi, M., Schär, C., 2010. A gridded hourly precipitation dataset for Switzerland using rain-gauge analysis and radar-based disaggregation. Int. J. Climatol. 30, 1764–1775. http://dx.doi.org/10.1002/joc.2025.

Xu, H., Xu, C.-Y., Chen, H., Zhang, Z., Li, L., 2013. Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China. J. Hydrol. 505, 1–12. http://dx.doi.org/10.1016/j.jhydrol.2013.09.004.

Yang, J., Reichert, P., Abbaspour, K.C., 2007. Bayesian uncertainty analysis in distributed hydrologic modeling: a case study in the Thur River basin (Switzerland). Water Resour. Res. 43. http://dx.doi.org/10.1029/2006WR005497.