Journal of Hydrology 529 (2015) 49–61

Contents lists available at ScienceDirect

Journal of Hydrology

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

Can a regionalized model parameterisation be improved with a limited number of runoff measurements?

Daniel Viviroli *, Jan Seibert

Department of Geography, University of Zürich, Switzerland

ARTICLE INFO

Article history:
Received 3 February 2015
Received in revised form 28 June 2015
Accepted 4 July 2015
Available online 9 July 2015
This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Ashish Sharma, Associate Editor

Keywords:
Hydrological modelling
Regionalisation
Value of data
Runoff measurements
Switzerland
Mountainous catchments

SUMMARY

Application of hydrological models to ungauged basins is both a highly relevant and challenging task. While research has brought forth various approaches for inferring or transferring tuneable model parameters from gauged and calibrated catchments, it has also been recently shown that a few short measurements can support predictions in an ungauged basin by constraining the acceptable range of the parameters. For the present study, we examined a combination of both parameter regionalisation and short-term runoff measurements. More precisely, we attempted to select complete parameter sets from a range of calibrated catchments using a few measurements. Then, we tested a number of ways to combine the hydrographs simulated with these parameter sets with those simulated using a well-established Nearest Neighbour scheme, in order to make use of both actually measured runoff data as well as hydrological similarity. The experimental basis for our study were 49 representative catchments in Switzerland which have been successfully calibrated and regionalised with the hydrological modelling system PREVAH. Results show that even a few short measurements during mean runoff conditions can lead to models that are more efficient than those achieved with hydrological similarity alone. The possible improvement depends largely on the regime type of the catchment examined. Also, the most suitable season to perform measurements varies: In catchments dominated by snow melt or ice melt or both, considerable improvements can be achieved with as few as two measurements during spring or summer, whereas rainfall-dominated catchments show only moderate improvements with no particular season being more suitable for the measurements. Our findings highlight the value of field measurements in mountain areas. The information gained in these regions from short measurements may act as a counterbalance to the sparse operational observation networks.

1. Introduction

Long series of hydrological measurements are the basis for tackling highly relevant questions in water resources management and water-related natural hazards, which concern both the human as well as the natural environment. Since direct measurements are available for only a limited number of sites, predictions in ungauged basins are an important albeit challenging task. This is highlighted by the attention the topic has received through the International Association of Hydrological Sciences (IAHS) decade on predictions in ungauged basins (PUB) (Hrachowitz et al., 2013).

In this context, the value of short runoff measurements gathered during a targeted field campaign has been recognised recently. Rather than relying on long data series for calibration of conceptual models or attempting a fully physical parameterisation of the catchment, using short measurements might lead to a feasible way of achieving improved predictions for ungauged sites (see Beven, 2002). Developing this approach, Seibert and Beven (2009) demonstrated that even a relatively small number of runoff measurements help in constraining model predictions, provided that these measurements are timed sensibly. The value of additional data from groundwater levels (Juston et al., 2009), glacial mass balances (Konz and Seibert, 2010) and soft data (Seibert and McDonnell, 2013) has been shown in similar studies. Related analyses were also performed by Perrin et al. (2007, 2008) who relied on short, partly continuous series of runoff data to select parameters from a vast library of predefined sets. Drogue and Plasse (2014), finally, showed that runoff measurements at random times can improve distance-based regionalisation approaches.

In the present study, we examine whether a successful regionalisation (i.e. one that performs only slightly worse than a calibrated model) can be further improved with a number of

* Corresponding author at: Hydrology and Climate, Department of Geography, University of Zürich – Irchel, Winterthurerstr. 190, CH-8057 Zürich, Switzerland.
E-mail address: daniel.viviroli@geo.uzh.ch (D. Viviroli).

http://dx.doi.org/10.1016/j.jhydrol.2015.07.009
0022-1694/

short measurements. This question was studied thoroughly for the Nearest Neighbour regionalisation developed for the conceptual, process-oriented hydrological modelling system PREVAH (Precipitation-Runoff-EVApotranspiration-HRU related model; Viviroli et al., 2009a). The Nearest Neighbour scheme was chosen since it is based on extensive model calibration for 140 sites with long gauge data series. It therefore offers a large pool of model parameter sets and, at the same time, extensive data to scrutinize the value of the measurement data introduced. The latter is of particular importance since results may differ from catchment to catchment and from year to year, resulting in misleading findings if only a few cases are examined (Seibert and Beven, 2009). All of our parameter sets are derived from calibration and provide functional and mutually adjusted parameter combinations. In contrast to purely random ("Monte-Carlo") parameter sets, they have the large advantage that they do not contain many implausible parameter combinations and are thus a computationally effective basis for our experiments (see e.g. Bárdossy, 2007; Khu and Werner, 2003; Perrin et al., 2008; Viviroli et al., 2009b). We tested our approach for 49 out of the abovementioned 140 calibrated catchments, these 49 having continuous series of at least 20 years of runoff measurements at the hourly time-step, which allowed for a thorough analysis and assessment.

2. Experimental basis

2.1. Hydrological model

2.1.1. General description

All simulations for this study have been performed with PREVAH (Precipitation-Runoff-EVApotranspiration-HRU related model; for definition of HRU see below) (Viviroli et al., 2009a). PREVAH is a conceptual, process-oriented hydrological modelling system which has been developed based on the HBV model (Bergström, 1972; Lindström et al. 1997) and relies on the aggregation of gridded spatial information into hydrological response units (HRUs, see Ross et al., 1979; Gurtz et al., 1999). These HRUs unite areas of a basin where similar hydrological behaviour is expected, thus representing a computationally efficient, dynamic spatial discretisation: With increasing variability of the physical catchment characteristics, the size of the HRUs decreases, while the number of HRUs increases (for details, see Viviroli et al., 2009a). Raster cells of 0.5×0.5 km^2 have proven reasonable as a basis for generating HRUs (Viviroli et al., 2009b).

PREVAH has already been used successfully in a large number of catchments and for a broad range of topics in Switzerland and abroad (see Viviroli et al., 2009a for an overview). Recent applications include flood estimation (Viviroli et al., 2009b,c; Viviroli and Weingartner, 2011), studies of climate and land use change impacts on flood and low flow hydrology (Addor et al., 2014; Bosshard et al., 2014; Köplin et al., 2012, 2013, 2014a,b; Meyer et al., 2011; Schattan et al., 2013) as well as flood, drought and water resources forecasting at various lead times (Addor et al., 2011; Fundel and Zappa, 2011; Fundel et al., 2013; Jörg-Hess et al., 2014; Liechti et al., 2013; Romang et al., 2011; Zappa et al., 2014).

The basic parameterisation of PREVAH relies on the topographic analysis of a Digital Elevation Model (DEM), on land cover characteristics and on maps of soil types. Each HRU is provided with a set of parameters based on information derived from the DEM (elevation, aspect and slope) and the soil map (plant-available soil field capacity, soil depth, hydraulic conductivity). Information on land cover provides additional values required for determining evapotranspiration (albedo, root depth, interception storage capacity, vegetation height, leaf area index and minimum stomatal resistance of the various vegetation classes). Non-vegetated surfaces (snowpack, glaciers, rock, large water bodies and urban areas) are parameterised separately (Gurtz et al., 1999). Meteorological and geophysical pre-processing is handled by a suite of comprehensive tools (Viviroli et al., 2007, 2009a).

2.1.2. Model input, model parameters and parameter estimation

For the present study, PREVAH was run at an hourly time-step, being forced by series of six observed climatic variables at the same time-step, namely precipitation, air temperature, global radiation, relative sunshine duration, wind speed and relative humidity. All of these variables were interpolated in space with Detrended Inverse Distance Weighting and Ordinary Kriging (see e.g. Garen and Marks, 2001; Isaaks and Srivastava, 1989; for elevation effects and detrending, see also Goovaerts, 2000) and averaged to 100 m elevation bands. The catchment-specific tuneable parameters of PREVAH are found in Table 1.

To calibrate the tuneable parameters against observed runoff, PREVAH provides an automatic global search algorithm based on an iterative procedure that sequentially treats the parameters pair-wise and narrows down the considered parameter space step by step (Zappa and Kan, 2007). Although being straight-forward, the algorithm leads to stable efficiencies and plausible flow components by evaluating multiple efficiency criteria (for details see Viviroli et al., 2009b). In the model version used here, the parameter for soil moisture recharge (BETA) was not calibrated, but computed from soil depth and altitude for each HRU (for details see Viviroli, 2007). Details of the model's physics, structure and parameterisation are reported in the comprehensive description by Viviroli et al. (2007).

2.2. Regionalisation

The baseline parameterisation was derived from a Nearest Neighbour regionalisation approach. This approach essentially consists in identifying a calibrated donor catchment that is as similar as possible to the ungauged target basin in question. All tuneable model parameters are then transferred from the donor to the target as a complete, unaltered set, preserving the mutual adjustment of the calibrated model parameters (Kokkonen et al., 2003; Young, 2006). Catchment similarity can be determined, for example, from spatial proximity (see Patil and Stieglitz, 2012 and references therein) or, as done in this study and explained in more

Table 1
Catchment-specific tuneable parameters of PREVAH as used in the present study (for details see Viviroli et al., 2007).

Abbreviation	Description	Unit
BETA	Non-linearity exponent for soil moisture recharge	(–)
CG1H	Storage time for quick baseflow	(h)
ICERMF	Radiation melt factor for ice	(mm h^{-1} K^{-1} W^{-1} m^2)
ICETMF	Temperature melt factor for ice	(mm d^{-1} K^{-1})
K0H	Storage time for surface runoff	(h)
K1H	Storage time for interflow	(h)
K2H	Storage time for slow baseflow	(h)
PERC	Percolation rate	(mm h^{-1})
PKOR	Water balance adjustment factor for rainfall	(%)
RMFSNOW	Radiation melt factor for snow	(mm h^{-1} K^{-1} W^{-1} m^2)
SGR	Threshold for generation of surface runoff	(mm)
SLZ1MAX	Maximum storage available for fast baseflow	(mm)
SNOKOR	Water balance adjustment factor for snowfall	(%)
T0	Threshold temperature for snowmelt	(°C)
TMFSNOW	Temperature melt factor for snow	(mm d^{-1} K^{-1})

Table 2
Catchment attributes used for determining hydrological similarity in the Nearest Neighbour approach, grouped by topic. The set was chosen from a total of 82 attributes by retaining the two attributes with highest correlation to each of the 14 tuneable model parameters (see main text for details; see Section 2.1.2 and Table 1 for explanation of model parameters).

Attribute	Unit	Thematic group	Model parameter(s) with highest correlation
Catchment maximum elevation	(m a. s. l.)	Physiography[a]	RMFSNOW
Catchment perimeter	(m)	Physiography	SLZ1MAX
Surfaces with inclination >15°	(%)	Physiography	K0H
Contributing areas[b]	(%)	Physiography	T0
Soil-covered areas	(%)	Land use[c]	K1H
Forested areas	(%)	Land use	K2H
Forested area in contributing areas[b]	(%)	Land use	T0
Hard rock, generic	(%)	Hydrogeology[d]	TMFSNOW, SLZ1MAX
Unconsolidated rock, intermediate permeability	(%)	Hydrogeology	K2H, ICETMF
Hydraulic conductivity, average	(mm h^{-1})	Physiography	K1H, CG1H
Hydraulic conductivity, standard deviation	(mm h^{-1})	Physiography	CG1H, ICERMF
Hydraulic conductivity, kurtosis	(–)	Physiography	ICERMF
Net field capacity, average	(%)	Physiography	SGR, K0H
Net field capacity, standard deviation	(%)	Physiography	PERC
1 h precipitation intensity, observed maximum	(mm h^{-1})	Precipitation[e]	PERC
15 min precipitation intensity with return period 2.33 years	(mm)	Precipitation	RMFSNOW
1 h precipitation intensity with return period 100 years	(mm)	Precipitation	PKOR, ICETMF
24 h precipitation intensity with return period 100 years	(mm)	Precipitation	PKOR
24 h precipitation intensity, variability of Julian Date of yearly maxima[f]	(–)	Precipitation	TMFSNOW
Soil-topographic index,[g] average	(–)	Soil[c]	SNOKOR
Soil-topographic index, standard deviation	(–)	Soil	SGR
Soil-topographic index, skewness	(–)	Soil	SNOKOR

[a] Based on $100 \times 100\ m^2$ digital elevation model from SFSO (2003).
[b] Contributing areas defined as having an average distance to the channel of no more than 250 m (see Kölla, 1987).
[c] See SFSO (2003) for details.
[d] See Bitterli et al. (2004) for details.
[e] Based on Geiger et al. (1992), Jensen et al. (1997) and MeteoSwiss (2013).
[f] See Burn (1997).
[g] See e.g. Ambroise et al. (1996).

detail below, from distance in a multi-dimensional space of catchment attributes (see e.g. Young, 2006).

As a basis for our experiments, we used the Nearest Neighbour approach by Viviroli et al. (2009c) that has been developed and thoroughly evaluated for 140 meso-scale catchments in Northern-alpine Switzerland. The approach combines five independent simulations, based on parameter sets from five Nearest Neighbour catchments, by computing the median runoff value for each (hourly) time-step. The underlying similarity measure was derived from a statistical analysis of catchment attributes: From a total of 82 attributes (Viviroli et al., 2009c), the two with the highest correlation to each of the 14 tuneable model parameters were retained. Since some attributes occurred more than once in the analysis, the resulting set consists of 22 attributes only instead of the 28 theoretically possible (see Table 2). These attributes were used with equal weight for computing the Euclidean Distance in the resulting 22-dimensional space. As a slight deviation from the original method, we did not restrict the search for Nearest Neighbours to zones of similar mean elevation, although this distinction leads to slightly improved results. This was done to keep the amount of potential donor catchments as large as possible in the experiments that follow.

Note that Viviroli et al. (2009c) came up with an extended regionalisation by combining the Nearest Neighbour approach with two additional approaches. This led to more stable results, as was also shown in a similar study by Oudin et al. (2008). The first additional approach consists of interpolating model parameters in space with Kriging, the second one in deriving them from catchment attributes with a regression analysis. Both additions treat the model parameters individually – rather than as a set – and are used for a simulation each. Finally, the median runoff value is calculated from the Nearest Neighbour, Kriging and regression hydrographs for each time-step (as described above, the Nearest Neighbour hydrograph itself is a combination of five simulations). We will refer to this refined approach as "extended regionalisation" and use it as an additional benchmark in assessing our experiments.

2.3. Study catchments and investigation period

As a data basis, we used a set of 49 study catchments in Switzerland (Fig. 1). This set consists of meso-scale catchments with long-term measurements (at least 20 years) and no significant direct human flow alterations, and covers the most relevant hydrological conditions in the Northern Alpine, Plateau and Jura regions of Switzerland. Catchments south of the main alpine ridge were not considered because there, hydropower production leads to a marked impairment of most observed hydrographs (see Margot et al., 1992; Köplin et al., 2010).

For these 49 catchments, extended by an additional set of 91 catchments with shorter measurements (at least 5 years), the regionalisation explained above (Section 2.2) has been examined comprehensively by Viviroli et al. (2009c). For comparability and consistency reasons, we use the same 20-year investigation period as Viviroli et al. (2009c), namely 1984–2003, with 1983 serving as warm-up year to initialise the model storages, and 1994–1997 serving for model calibration. Before 1983, applications are restricted by a lack of meteorological data at hourly resolution, while after 2003, consistency problems arise in the interpolation of model input due to the reorganisation of the Swiss meteorological observation network (MeteoSwiss, 2010).

3. Methods

3.1. Sampling strategies

Similar to the approach used in previous studies (Seibert and Beven, 2009; Seibert and McDonnell, 2013), we pretended that runoff data were available only for selected points in time

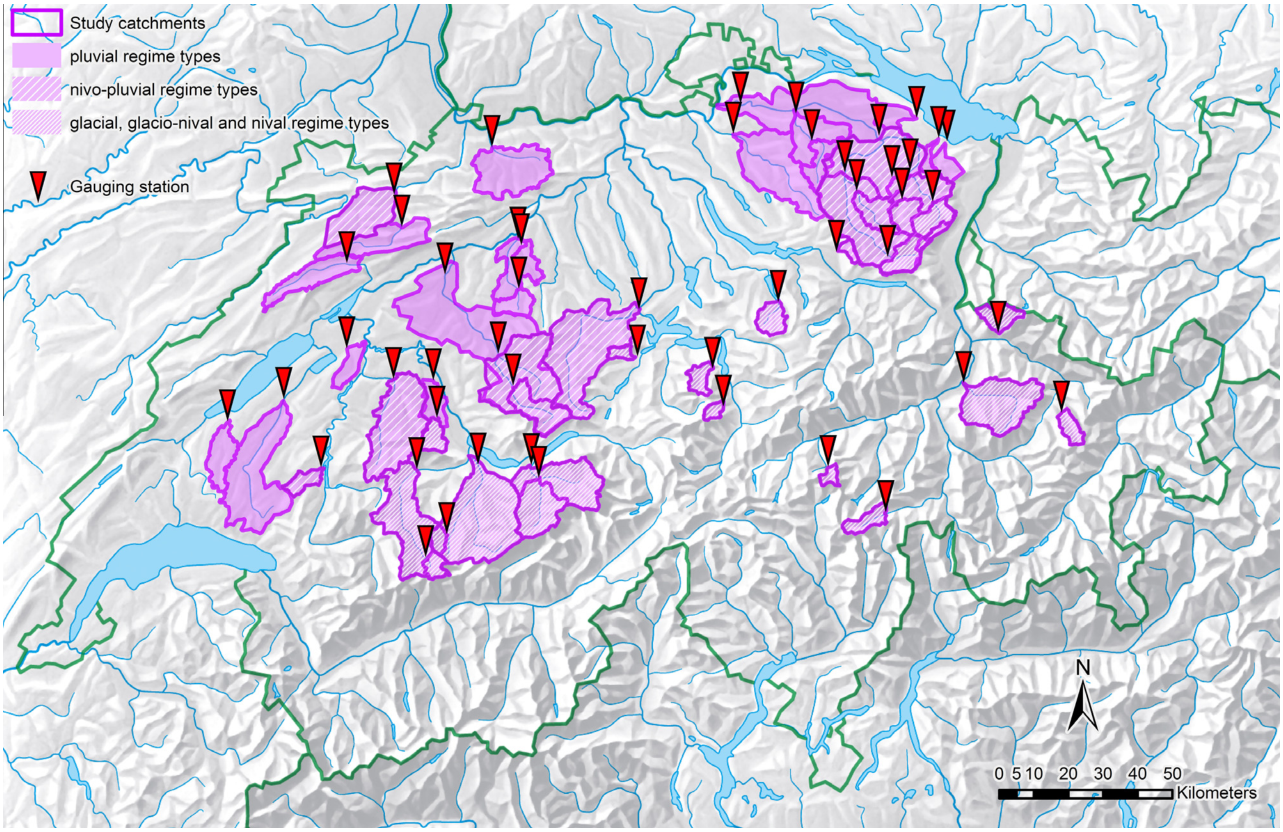

Fig. 1. Map of Switzerland showing the location of the 49 study catchments and the corresponding gauging stations (basemap: Swiss Federal Office of Topography).

distributed over a 'field campaign period' of three months. In such a field campaign, hourly runoff would be measured a number of times (up to 256 times) for a certain duration (up to 8 h). The complete observed hydrograph was used only later for evaluating the various experiments.

We assumed that the time of the year during which a field campaign takes place has an influence on the results. Therefore, we divided the calendar year into four seasons. October 1st, the beginning of the hydrological year in Switzerland, lies at the start of the fall season (October, November and December), with the seasons winter (January, February and March), spring (April, May and June) and summer (July, August and September) following. All further analyses were done for each of these four seasons individually. In addition, we examined the results of performing the measurements in the course of one full hydrological year (October to September), and, in selected cases, also distributed the measurements across half a year instead of three months only (spring/summer: April to September, fall/winter: October to March).

For determining the time at which an individual measurement is carried out, a number of different sampling strategies were examined. The basic idea for these strategies was that during a field campaign measurements would be attempted to be taken during average flow conditions. Apparently, the exact time of occurrence of average flow is unknown in an ungauged catchment. However, we here focus on the potential value of such data if it were available, and this more theoretical strategy would have to be adapted in practice. In initial tests, best results were achieved with first subdividing the field campaign period into as many time segments of identical length as there would be measurements. Then, average runoff was determined for each segment. Finally, one measurement was made per segment, at exactly the time when the hydrograph reached average runoff (see Fig. 2). Where average runoff occurred more than once in a time segment, the instance with the most steady runoff (i.e. the smallest gradient in the hydrograph) was used. Strategies with focus on low flow and on peak flow might be more favourable for parameter sets referring to flood or low flow conditions (Viviroli et al., 2009b; Meyer et al., 2011) which were, however, not in the scope of this study.

3.2. Introduction of field measurements

For each of the 49 study catchments, 48 catchments were available as potential donors of model parameters. Each potential donor parameter set was then used for simulating a hydrograph. Ordered by decreasing hydrological similarity (cf. Table 2) of the corresponding donor catchments, these Nearest Neighbour hydrographs were called $NNBR_{01-48}$, where $NNBR_{01}$ is the hydrograph simulated with the parameters from the hydrologically most similar catchment.

Runoff measurement data were introduced subsequently as follows: For the small number of time steps with (assumed) measurements, the root mean square error (RMSE) of observation and simulation was computed for each donor catchment. Sorted by increasing RMSE, these hydrographs were called $field_{01-48}$, where $field_{01}$ has the smallest RMSE. $field_{01-48}$ is thus essentially a re-ordering of $NNBR_{01-48}$ using runoff measurements gained in the field campaign.

A number of strategies were then tested to make use of the *field* hydrographs (Table 3). In a first straight-forward approach, the $field_{01}$ hydrograph was used directly. Aiming at more robust results, the simulated hydrographs $field_{01-05}$ were then also combined, in a first variant by computing their median at each time step, and in a second variant by computing their weighted mean at each time step (with weights $^5/_{15}$ for $field_{01}$, $^4/_{15}$ for $field_{02}$, $^3/_{15}$

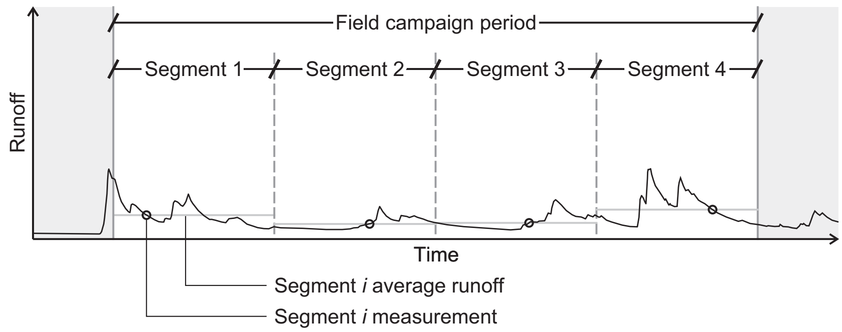

Fig. 2. Visualisation of the sampling strategy used, example for 4 measurements: first, the field campaign period was divided into time segments of identical length (as many segments as there would be measurements). Second, average runoff was determined for each segment. Third, one measurement was made per segment at the time of average runoff. Where average runoff occurred more than once in a time segment, the instance with the smallest gradient in the hydrograph was used.

Table 3
Strategies for using field-based and similarity-based Nearest Neighbours in the present article.

Strategy	Explanation
$field_{01}$	From all hydrographs simulated with parameter sets from the 48 potential donor catchments, use the one hydrograph with the smallest root mean square error (RMSE) relative to the field measurements performed
$field_{01-05}$	Same as $field_{01}$, but combining the five hydrographs with smallest RMSE by computing their median for each time-step
$field_{01-05}$, weighted	Same as $field_{01-05}$, but mean with weights $^5/_{15}$, $^4/_{15}$, ..., and $^1/_{15}$ assigned to the hydrographs ordered by increasing RMSE
$field_{01} \otimes NNBR_{01}$	Same as $field_{01}$, but in combination with the hydrograph simulated using the parameter set from the hydrologically most similar donor catchment (compute mean for each time-step)
$field_{01-05} \otimes NNBR_{01-05}$	Same as $field_{01} \otimes NNBR_{01}$, but combining the five hydrographs with smallest RMSE and the hydrographs simulated with parameter sets from the five hydrologically most similar donor catchments (compute median for each time-step)

for $field_{03}$, $^2/_{15}$ for $field_{04}$, and $^1/_{15}$ for $field_{05}$). The use of five donor catchments was intended as a compromise between achieving robust results and including too many and potentially unsuitable parameter sets (see Viviroli et al., 2009c).

In addition, the hydrographs selected on the basis of runoff measurements were combined with the similarity-oriented Nearest Neighbour hydrographs, on the one hand by computing the mean of $field_{01}$ and $NNBR_{01}$ at each time step ($field_{01} \otimes NNBR_{01}$), and on the other hand by computing the median of $field_{01-05}$ and $NNBR_{01-05}$ at each time step ($field_{01-05} \otimes NNBR_{01-05}$). These combinations aim at pairing hydrological information from the integrated catchment response (field measurements) with hydrological information from catchment attributes (Nearest Neighbours).

The approach of combining a number of hydrographs based both on field measurements as well as Nearest Neighbour relations slightly resembles the methods used by Drogue and Plasse (2014). These authors selected their donor catchments on the basis of a hybrid distance that combines distance in space (rather than hydrological dissimilarity as in our case) with the distance between short observations and simulations with donor catchment parameter sets (using square root transformed runoff values for computing RMSE). Furthermore, they combined this hybrid distance with the map correlation method (Archfield and Vogel, 2010).

3.3. Evaluation of simulations

All of the combination strategies described above were tested for each of the 20 years 1984–2003, and for each year used, efficiency scores were computed for the entire period 1984–2003. As this was done in each of the 49 study basins, $49 \cdot 20 = 980$ evaluations were available per strategy.

3.3.1. Number and duration of measurements examined

The various strategies for using sampled data were evaluated for 1, 2, 4, 8, 16, 32, 64, 128 and 256 measurements, each with a duration of 1, 2, 4 and 8 h (one measurement per hour). Note that 256 measurements of 8 h duration result in a total measurement time of 2048 h or 85.3 days. When performed in the course of one season (i.e., 3 months), this is almost tantamount to a continuous measurement. Although it is not realistic to perform such a high number of measurements in the course of a short field campaign, we chose to examine this as well to identify the theoretical limits of the approach.

In addition to the above, all analyses were also done without using any field data. For this, the order of $field_{01-48}$ was determined randomly, while the order of the Nearest Neighbour hydrographs was left unchanged. The hydrographs $field_{01}$ and $field_{01-05}$, now unrelated to our assumed measurement campaign, were then used just as described in Section 3.2. For each combination strategy, these random experiments were performed 10,000 times, and the median of the corresponding efficiency scores was retained. In presenting and discussing the results (Section 4), these scores will be referred to as having a measurement duration of 0 h.

3.3.2. Efficiency score

In all our experiments, model performance was evaluated for the entire period using the so-called "Lindström efficiency" E_V (Lindström, 1997; Lindström et al., 1997). E_V expands the popular Nash–Sutcliffe efficiency (E, Nash and Sutcliffe, 1970) by a term for the relative volume error (D_V), namely

$$E = 1 - \frac{\sum_{t=1}^{n}(Q_t - q_t)^2}{\sum_{t=1}^{n}(Q_t - \overline{Q})^2}; \quad E \in]-\infty, 1]$$

and

$$D_V = \frac{\sum_{t=1}^{n}(q_t - Q_t)}{\sum_{t=1}^{n} Q_t}; \quad D_V \in]-\infty, \infty[$$

to

$$E_V = E - \omega|D_V| \text{ with } \omega = 0.1; \quad E_V \in]-\infty, 1],$$

where Q_t is observed runoff at time step t, q_t simulated runoff at time step t, $\overline{Q}$ the average of observed runoff, n the number of time steps and ω a weight. E_V is very similar to E, but evades some of its deficits (see e.g. Gupta et al., 2009; Schaefli and Gupta, 2007) by adding a penalty for volume error. Following the recommendation of Lindström (1997), we set ω to 0.1 to achieve a good compromise between E and D_V. Note that E and E_V put more weight on high flows than on low flows (Legates and McCabe, 1999; Krause et al., 2005), and that the characteristics of the chosen efficiency criterion will influence the conclusions on our sampling strategies.

As a reference for assessing our strategies, we used the median of the E_V score of the Nearest Neighbour regionalisation (see Section 2.2) in all of our 49 study catchments. The gain or loss in efficiency that results from introducing field measurements is then expressed as ΔE_V. To ensure that the reference score is comparable, however, it has to combine as many hydrographs as the strategy does. It follows, for example, that $field_{01}$ must be assessed against $NNBR_{01}$, while $field_{01-05}$ is assessed against $NNBR_{01-05}$ and $field_{01-05} \otimes NNBR_{01-05}$ against $NNBR_{01-10}$ (see Table 4 for details). Note that the standard Nearest Neighbour scheme used by Viviroli et al. (2009c), equivalent to $NNBR_{01-05}$, achieves a median E_V score of 0.70 over all study catchments. The significance of the changes in the median of ΔE_V was assessed with a two-sided Wilcoxon signed-rank test (Hollander and Wolfe, 1999) at a level of $\alpha = 0.05$.

Fig. 3 summarises the methods and data used in a flowchart.

Table 4
Reference hydrographs and corresponding efficiencies (E_V score, median over all 49 study catchments) for all strategies evaluated.

Hydrograph evaluated	Reference hydrograph	Reference efficiency (E_V)
$field_{01}$	$NNBR_{01}$	0.63
$field_{01-05}$	$NNBR_{01-05}$	0.70
$field_{01-05}$, weighted	$NNBR_{01-05}$	0.70
$field_{01} \otimes NNBR_{01}$	$NNBR_{01-02}$	0.69
$field_{01-05} \otimes NNBR_{01-05}$	$NNBR_{01-10}$	0.70

3.3.3. Grouping by runoff regime type

Besides evaluating our strategies for all study catchments (Sections 4.1 and 4.2), we also performed analyses on the catchments grouped by their dominant regime type. This served to examine the relevance of snow and ice melt processes in our considerations (Section 4.3). We devised three groups of regime types, namely (i) glacial, glacionival and nival types (18 catchments, dominated by snow melt or ice melt or both), (ii) nivopluvial types (14 catchments, dominated by a mixture of snow melt and rainfall) and (iii) pluvial types (17 catchments, dominated by rainfall). For more details on the individual regime types, we refer to the work of Aschwanden and Weingartner (1985) and Weingartner and Aschwanden (1992). A further refinement of the regime types was not envisaged since the number of study catchments per regime would become too small.

The above analysis by regime type was done for the hydrograph $field_{01}$ (i.e., the simulated hydrograph with the best correspondence to the field data) since this proved to be the one most sensitive to the introduction of field measurements. Note that the median of the improvements achieved with calibration and

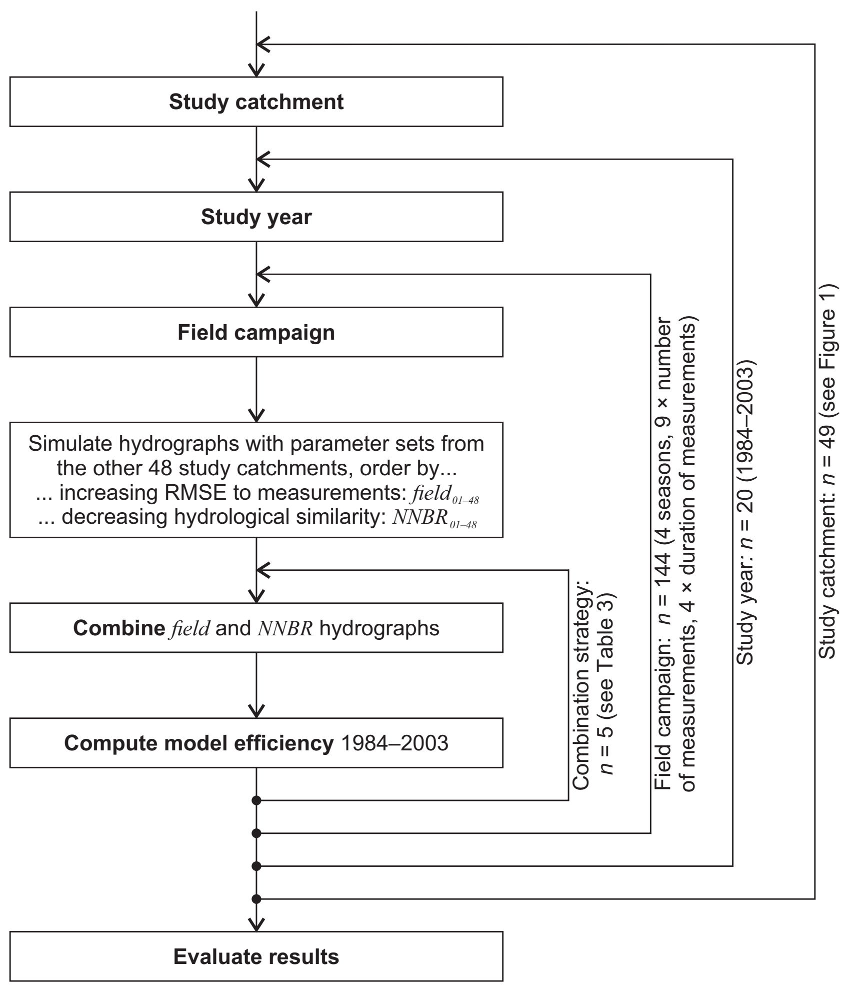

Fig. 3. Flowchart of methods and data used.

extended regionalisation were recomputed separately for each group of regime types. However, the reference value of E_V, and thus also the level of $\Delta E_V = 0$, still applies to the entire set of 49 study catchments (as reported in Table 4) for better comparison between the experiments.

4. Results

4.1. Overall results

A first overview of the results is provided in Table 5 which lists the absolute E_V values, computed as median over the 49 study catchments. We chose the example of 8 measurements with a duration of 2 h each to represent a reasonable effort for a field campaign. The best absolute result, with E_V = 0.72, was achieved with the combined hydrographs of five measurement-oriented and five similarity-oriented donors ($field_{01-05} \otimes NNBR_{01-05}$) and the measurements taking place in spring or summer. All other approaches that combine hydrographs performed only slightly worse, reaching E_V values of 0.70 in spring and summer, and, for the weighted combination of $NNBR_{01-05}$, of 0.71 in summer. The best absolute scores are thus close to the results of a calibration, which achieves a median E_V of 0.73 over the same set of study catchments. Also, the extended regionalisation that combines Nearest Neighbour, Kriging and regression approaches (median E_V of 0.71) was surpassed.

4.2. Improvements by strategy

To examine the efficacy of the individual strategies in more detail, we will look at the improvements possible with each of them below. Note that the reference efficiencies are different for each strategy (see Table 4 and Section 3.3.2).

When using only the one regionalised hydrograph with closest agreement to these measurements ($field_{01}$) (Fig. 4, top), at least 8 measurements were necessary to significantly improve model efficiencies, although results were always better than just using an arbitrary donor catchment (i.e. n_m = 0 measurements). Gauging during spring or summer led to the highest improvements, whereas gauging during fall and especially winter did not lead to noticeable improvements even with many measurements. Gauging during the spring and summer half year led to even better results than seasonal or full-year measurements in some cases. Although measurements of longer duration generally led to somewhat higher efficiencies, the duration of the measurements had less influence on the results than the season during which these measurements were performed. A large number of measurements was necessary to surpass the efficiency of the extended regionalisation scheme by Viviroli et al. (2009c) (see Section 2.2), namely 64 measurements in spring or summer, or 32 measurements over the entire year. The efficiency of a calibrated model cannot be achieved, even with a quasi-continuous sampling of 256 measurements during three months.

Table 5
Efficiencies (E_V score) for all combinations of field-based ($field_i$) and similarity-based ($NNBR_i$) hydrographs with 8 measurements of 2 h duration, computed as the median of the results in 49 study catchments.

Hydrograph evaluated	Efficiency (E_V)				
	Winter	Spring	Summer	Fall	Entire year
$field_{01}$	0.64	0.67	0.66	0.64	0.68
$field_{01-05}$	0.68	0.70	0.70	0.66	0.70
$field_{01-05}$, weighted	0.68	0.70	0.71	0.68	0.72
$field_{01} \otimes NNBR_{01}$	0.68	0.70	0.70	0.69	0.70
$field_{01-05} \otimes NNBR_{01-05}$	0.70	0.72	0.72	0.71	0.72

The combination of the five hydrographs with closest agreement to our field measurements ($field_{01-05}$) (Fig. 4, middle) required more field measurements than $field_{01}$ to improve model performance. However, by weighting the hydrographs according to their agreement with gauged data ($field_{01-05}$, weighted) (Fig. 4, bottom), significant improvements were achieved with 16 or more measurements during the spring season or at least 4 measurements during one year. This is comparable to the effort necessary when using the $field_{01}$ hydrograph, with the important difference that the improvements achieved are clearly smaller.

Combining field measurements and Nearest Neighbours regionalisation ($field_{01} \otimes NNBR_{01}$) yielded significant improvements with four measurements during any season, or two distributed across the entire year (Fig. 5, top). An influence of the gauging season was barely noted for up to 32 measurements, where the results of the extended regionalisation were clearly surpassed in all cases evaluated. Using five measurement-oriented and five similarity -oriented hydrographs ($field_{01-05} \otimes NNBR_{01-05}$) instead of one only led to very similar results (Fig. 5, bottom). The most notable difference is that the influence of the season is discernible again, i.e. that measuring in spring or summer is more favourable than measuring during fall or winter.

4.3. Analysis by regime type

For insights into the importance of seasonal catchment behaviour, results from $field_{01}$ are presented for three regime type groups separately.

Fig. 6 (top) shows the results for the basins of glacial, glacionival and nival regime type. Whereas at least 8 measurements were found to be necessary to achieve a significant improvement in the entire set of study catchments on average (Fig. 4, top), 2 measurements during spring or summer already led to a significantly improved efficiency for these regimes. Marked improvements were achieved with more measurements during spring or summer, although performing more than 32 measurements did not further improve the results. The duration of the individual measurements made little difference to the results. Analyses with an assumed measurement period of 6 months showed that distributing the measurements across the spring and summer half-year leads to results superior to those achieved with measurements in spring or summer alone, most notably if 4 measurements were performed. An exception was the case of one single measurement, where best results were achieved in summer. For 8 or more measurements, spring and summer half-year measurements led to results equal to those of measurements during the entire year.

The nivopluvial regime types (Fig. 6, middle) exhibit a quite different behaviour: First of all, the influence of the season has diminished strongly, although measuring during spring or summer was still more favourable. It is especially noteworthy that spring and summer measurements yielded better results than all-year measurements. Also in contrast to the glacial, glacionival and nival regimes, fall measurements led to equally poor results as winter measurements. Apart from that, many more measurements were generally necessary than in the glacial-glacionival-nival group to achieve an improvement. Also, long measurements led to higher improvements than short measurements.

The pluvial regime types (Fig. 6, bottom), finally, did not show any clear preference for a gauging season. More than that, seasonal measurements were inferior to distributing the same number of measurements over the entire year. In comparison to the nivopluvial regimes, improvements were achieved with fewer measurements, although not as few as for the glacial, glacionival and nival regimes. Also here, long measurements led to higher improvements than short measurements.

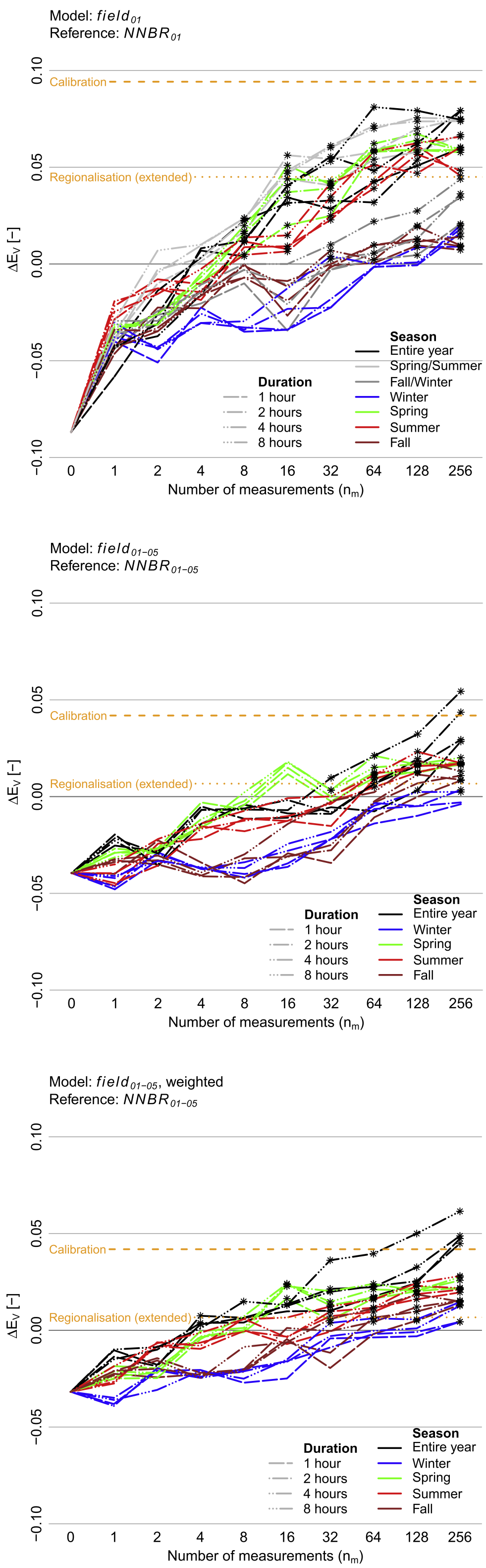

Fig. 4. Results for using the parameter set yielding results closest to those of the measurements (*field*$_{01}$, top), the five parameter sets yielding results closest to those of the measurements (*field*$_{01–05}$, middle), and the five parameter sets yielding results closest to those of the measurements, with weighting (*field*$_{01–05}$, weighted, bottom). Values are relative to the results of using the parameter set of the hydrologically most similar catchment (*NNBR*$_{01}$, top), and the sets of the five hydrologically most similar catchments (*NNBR*$_{01–05}$, middle and bottom). The result of picking arbitrary donor catchments is indicated at $n_m = 0$. Significant improvements (Wilcoxon rank-sum test, two-sided, $\alpha = 0.05$) are denoted with a star. For reference values of E_V see Table 4 and Section 3.3.2.

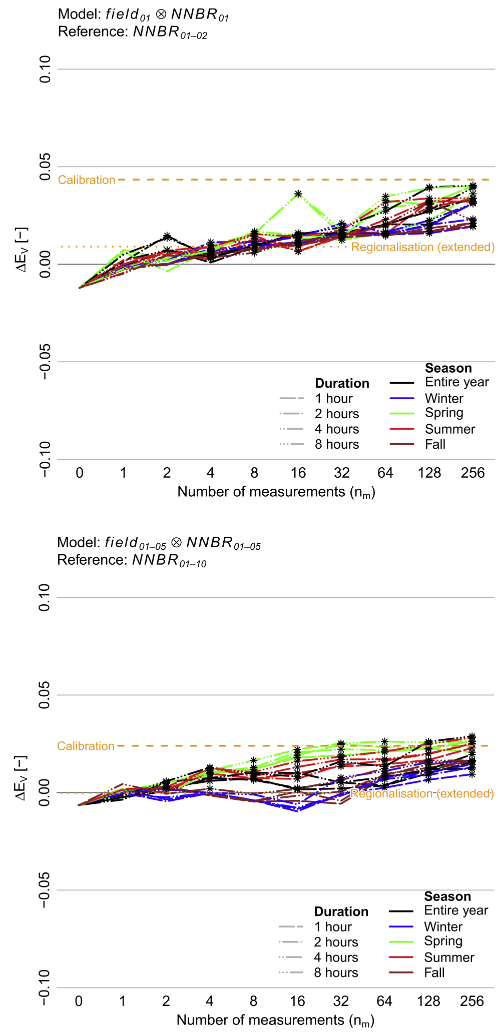

Fig. 5. Results for combining simulations with the parameter set yielding results closest to those of the measurements with those from the hydrologically most similar catchment (*field*$_{01}$ ⊗ *NNBR*$_{01}$, top), and combining simulations with the five parameter sets yielding results closest to those of the measurements with those from the five hydrologically most similar catchments (*field*$_{01–05}$ ⊗ *NNBR*$_{01–05}$, bottom). Values are relative to results of using the parameter sets of the two hydrologically most similar catchments (*NNBR*$_{01–02}$, top), and the sets of the ten hydrologically most similar catchments (*NNBR*$_{01–10}$, bottom). The result of picking arbitrary donor catchments is indicated at $n_m = 0$. Significant improvements (Wilcoxon rank-sum test, two-sided, $\alpha = 0.05$) are denoted with a star. For reference values of E_V see Table 4 and Section 3.3.2.

4.4. Range of results

The above results always related to the median of our 49 study basins, and thus to the improvements that can be expected in at least 50% of the cases. In a practical application, however, we are also interested in a single basin and a single year, and thus in the range of the results. Therefore, the interquartile and whisker ranges for Figs. 4–6 are provided in Figs. 7–9. For clarity, the ranges of the individual measurement durations (1, 2, 4 and 8 h) have been averaged. Note that the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box (Murrell, 2005). Also note that the scale of the ordinate is different from that of the original figures.

In general, the range of the results is considerable, at least in relation to the improvements possible. The range is widest when only the single regionalised hydrograph with closest agreement to the measurements (*field*$_{01}$) is used, and it is narrowest for the

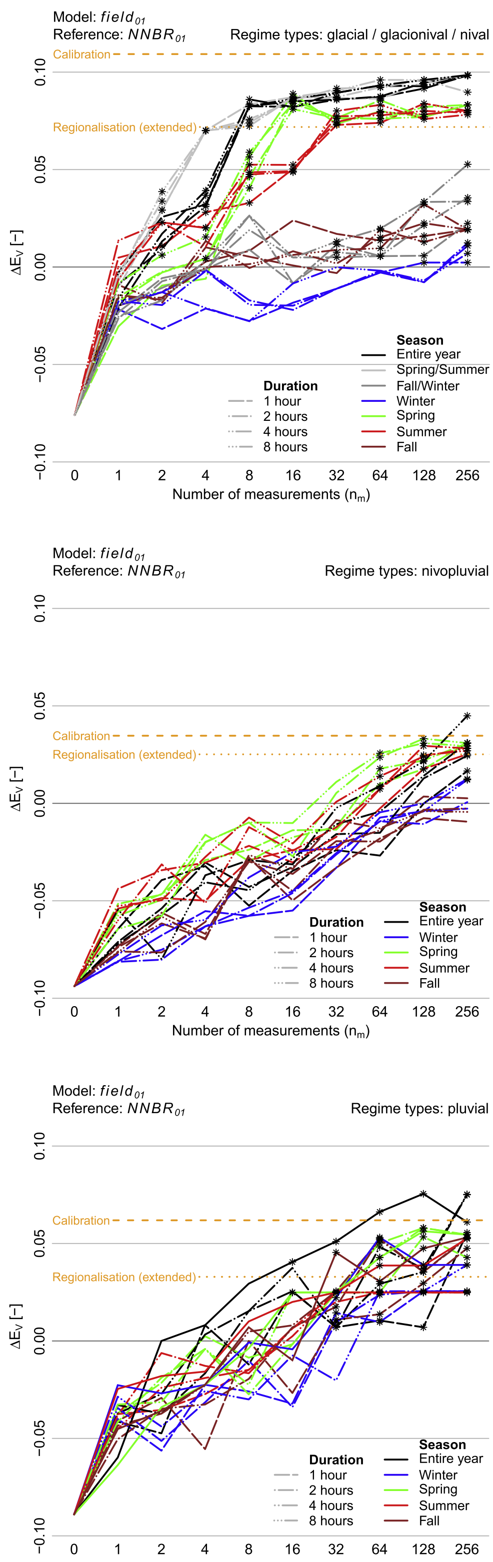

Fig. 6. Results for using the parameter set yielding results closest to those of the measurements (*field*$_{01}$, similar to Fig. 4, top), grouped by regime types. Values are relative to results of using the parameter set of the hydrologically most similar catchment (*NNBR*$_{01}$). The result of picking arbitrary donor catchments is indicated at $n_m = 0$. Significant improvements (Wilcoxon rank-sum test, two-sided, $\alpha = 0.05$) are denoted with a star. For reference values of E_V see Table 4 and Section 3.3.2.

combination of five measurement-oriented and five similarity-oriented hydrographs ($field_{01-05} \otimes NNBR_{01-05}$).

Also, the range does not narrow down systematically when more measurements are done, although it rises simultaneously with the median results, which we have already seen improve at significant levels. An exception is found in catchments with glacial, glacionival and nival regimes (Fig. 9, top), where the range decreases slightly for spring and summer measurements and increases for fall and winter measurements.

5. Discussion

It is apparent from the median results (Figs. 4–6) that measurements in fall or winter generally yield inferior results when compared to measurements in spring or summer or over the course of an entire year. The explanation for this must be sought in the information contained in the field data at different times of the year, and the relevance of this information for identifying well-suited parameter sets (or rather, in our case, the corresponding donor catchments). This hypothesis is corroborated by our analysis by regime type: For catchments with significant influence of snow and ice melt, it proved more effective to gauge during spring or summer when the dominance of these processes is high. Although the knowledge of fall and winter time runoff might lead to simulations with improved low-flow behaviour, the much more active parts of the hydrograph in spring (dominated by snow melt) and summer (dominated by ice melt) will not benefit from this information. Consequently, an exclusive focus on inactive parts of the hydrograph will lead to lower efficiencies in the overall simulation.

In the particular case of the glacial, glacionival and nival regime types, our findings agree well with those for glacierised catchments of Konz and Seibert (2010), who showed that even a few runoff measurements led to improved simulations, provided that the measurements are taken during the melting season. Also in agreement with our results, they noted that additional measurements during the low flow period are necessary, however, to achieve further improvements.

For the nivopluvial regime types – where the influence of melt processes is smaller, but still important –, we noted that spring and summer measurements yield better results than all-year measurements. This could be due to the superposition of snowmelt and rainfall influences, making it difficult to capture both with the same number of measurements distributed across the entire year. Also in contrast to the glacial, glacionival and nival regimes, fall measurements led to equally poor results as winter measurements, because there are no ice melt processes to be captured in the transition from late summer to early fall months. The necessity of both more and longer measurements for achieving improvements is explained by the behaviour of rainfall (pluvial influence) being more erratic than that of melt processes (glacial, glacionival and nival influences).

In the case of the pluvial regime types, no single season seems to contain more important hydrological information than any other, which is in full agreement with the premise of the pluvial regime, namely rainfall being the sole dominant influence across the entire year. In contrast to the more regular snow and ice melt processes, the temporal distribution of rainfall may even show notable variations between and within years, which limits the effectiveness of measurements during one season only. In this case, it might be favourable to perform measurements during

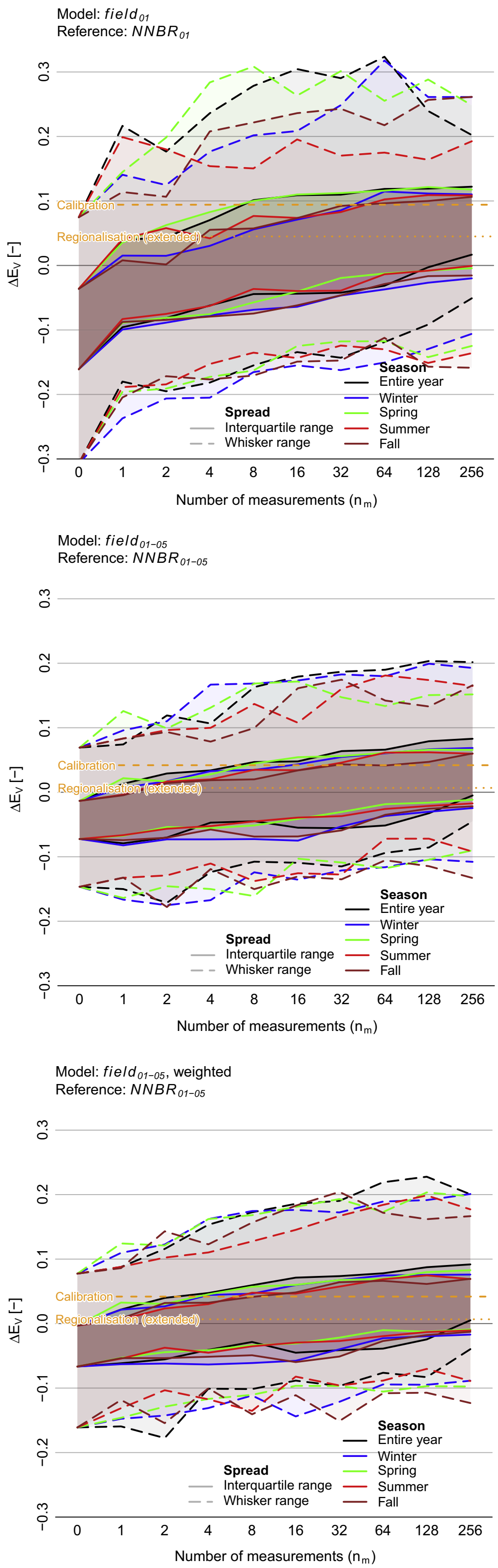

Fig. 7. Results for using the parameter set yielding results closest to those of the measurements ($field_{01}$, top), the five parameter sets yielding results closest to those of the measurements ($field_{01-05}$, middle), and the five parameter sets yielding results closest to those of the measurements, with weighting ($field_{01-05}$, weighted, bottom). Values are relative to results of using the parameter set of the hydrologically most similar catchment ($NNBR_{01}$, top), and the sets of the five hydrologically most similar catchments ($NNBR_{01-05}$, middle and bottom). The result of picking arbitrary donor catchments is indicated at $n_m = 0$. For reference values of E_V see Table 4 and Section 3.3.2. See also Fig. 4.

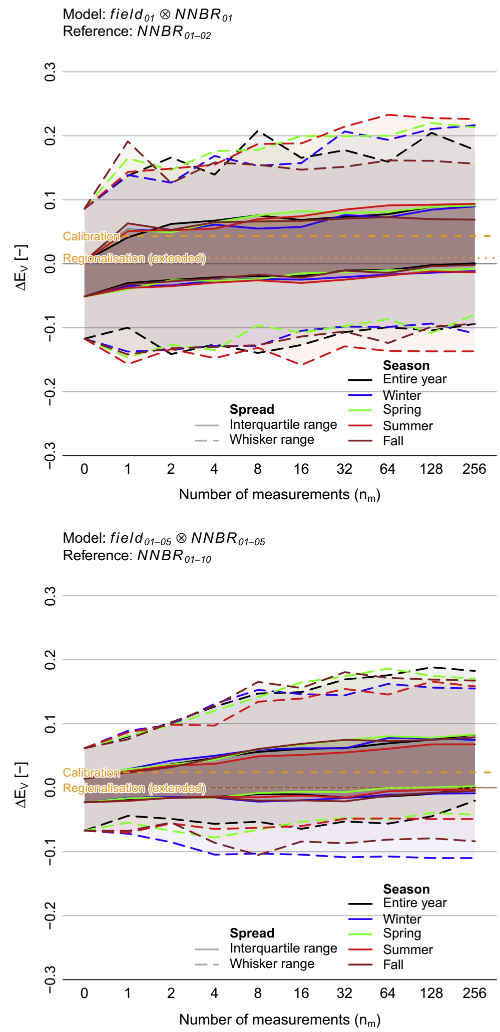

Fig. 8. Results for combining simulations with the parameter set yielding results closest to those of the measurements with those from the hydrologically most similar catchment ($field_{01} \otimes NNBR_{01}$, top), and combining simulations with the five parameter sets yielding results closest to those of the measurements with those from the five hydrologically most similar catchments ($field_{01-05} \otimes NNBR_{01-05}$, bottom). Values are relative to results of using the parameter sets of the two hydrologically most similar catchments ($NNBR_{01-02}$, top), and the sets of the five hydrologically most similar catchments ($NNBR_{01-05}$, bottom). The result of picking arbitrary donor catchments is indicated at $n_m = 0$. For reference values of E_V see Table 4 and Section 3.3.2. See also Fig. 5.

particularly dry or wet periods or otherwise unusual conditions, as discussed for model calibration by Singh and Bárdossy (2012). Analysing such advanced measurement schemes (e.g. based on an Antecedent Precipitation Index) was however not within the scope of the present study that focused on taking field measurements during average flow conditions.

Using only the one regionalised hydrograph with closest agreement to the measurements ($field_{01}$) showed largest gains in efficiency against using the hydrologically most similar Nearest Neighbour ($NNBR_{01}$) when a sufficient number of measurements were done. This points to the value of short measurements where it is not possible to determine hydrological similarity comprehensively. The limitations of using agreement to the measurements only are however shown for combining five corresponding hydrographs ($field_{01-05}$, in particular when no weighting is done), where hydrological similarity alone ($NNBR_{01-05}$) is more effective unless

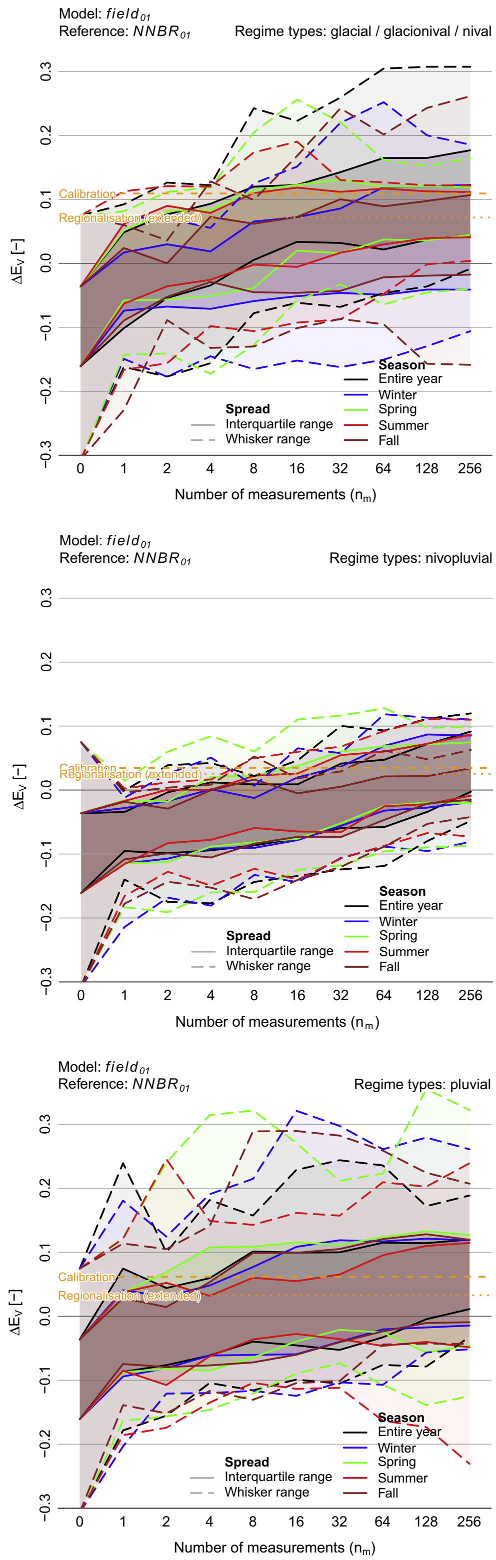

Fig. 9. Results for using the parameter set yielding results closest to those of the measurements (*field*$_{01}$, similar to Fig. 7, top), grouped by regime types. Values are relative to results of using the parameter set of the hydrologically most similar catchment (*NNBR*$_{01}$). The result of picking arbitrary donor catchments is indicated at $n_m = 0$. For reference values of E_V see Table 4 and Section 3.3.2. See also Fig. 6.

32 or more measurements are performed. Combinations of information from the field measurements and information from hydrological similarity ($field_{01} \otimes NNBR_{01}$, $field_{01-05} \otimes NNBR_{01-05}$) lead to smaller absolute gains in efficiency against the Nearest Neighbour benchmarks ($NNBR_{01-02}$ and $NNBR_{01-10}$, *respectively*). The risk of causing losses in efficiency is however much smaller, and it is even possible to attain the efficiency of a calibration. Provided that a sufficient effort is taken, it is thus still possible to improve a sophisticated regionalisation with short measurements.

Occasionally, the combination of many short measurements and regionalisation leads to results higher than those of standard calibration. This is possible because the model was calibrated for each catchment with runoff data from the same five years 1994–1997, whereas the assumed short measurements cover any of the years 1984–2003 and may thus add information to the parameter estimation problem.

When it comes to assessing the various strategies examined in absolute terms, the median improvements achieved are small to moderate, and depending on the individual catchment and year studied, there is scope for both gains as well as losses in model efficiency (Figs. 7–9). It should be noted, however, that the Nearest Neighbour regionalisation reaches high efficiencies already on its own. Finding that improvements are not higher proves that the regionalisation indeed works well and that runoff measurements – in the framework of the present study – only provide limited, albeit still important additional information. This has to be viewed in the context of the data-rich environment in which our regionalisation was developed: It is based upon comprehensive sets of calibrated model parameters and catchment attributes, which may not be available in other parts of the world. The improvements possible with limited measurements might thus be higher in data-scarce regions where the development of an effective regionalisation is not feasible.

6. Conclusions and outlook

A comprehensive set of study catchments was examined regarding the value of field data for improving a Nearest Neighbour regionalisation scheme. Our analyses highlight that runoff data measured in a short field campaign can indeed further improve model performance. The possible improvements depend, to a large extent, on the regime type of the basin considered and the season during which these measurements take place. In basins where processes of snow and ice melt are dominant, higher improvements are achieved when these processes are captured by measurements in spring or summer, and even two measurements can lead to a significant improvement. This highlights the high value of field measurements in mountain areas, which is amplified by the low density of operational measurements at high altitudes (Perks et al., 1996; Viviroli et al., 2011). Even in comparatively data-rich regions, like Switzerland in the present study, a few well-targeted measurements can lead to improvements. Whether it is worthwhile making short measurements in a particular case is ultimately a question of balancing the cost of the measurements against the potential benefits achieved through improved simulations.

When provided with the possibility of distributing the same number of measurements across an entire year instead of a single season, choosing the entire year resulted in slightly better results for purely melt or purely rainfall dominated catchments. When a mixture of these influences is present, however, spring or summer measurements are more favourable.

In comparison to the measurement season and the number of measurements, the duration of each of the individual measurements is less decisive. Taking more than one measurement at a time is thus only advisable if the cost of getting to a site is much larger than the cost of doing a measurement. However, a differentiation by regime type is due also here: Where melt processes are dominant, the measurement duration will not make a noticeable difference in the results. This is because the corresponding hydrographs do not exhibit extreme fluctuations over the day. Rainfall-dominated catchments, in contrast, benefit from longer measurements, because their runoff can change sharply from one hour to the next.

To maximise the impact of the measurements, it proved most successful to combine simulations chosen on the basis of these measurements with simulations selected upon catchment similarity (Nearest Neighbours). With such a strategy, 4 measurements in spring or summer were sufficient to surpass the results of an extended regionalisation that combines hydrological similarity (Nearest Neighbours) with parameter interpolation in space (Kriging) and model parameters derived from catchment attributes (regression). This underscores the high value of actual "field truth" from the catchment under consideration, even if many catchments with calibrated model parameter sets are available as potential donors.

Our analyses focused on models with good overall performance and on measurements during average flow, which is relevant for a wide range of hydrological applications – most specifically for those focusing on overall water balance characteristics. Further research should also focus on peak and low flow and on the importance of unusual (e.g. particularly wet or dry) periods. In this context, we recommend testing additional efficiency criteria that are sensitive on different parts of the hydrograph. From a practical point of view, it should also be examined in future studies to what extent the benefits of short measurements are diminished if the envisaged runoff measurement strategy (such as measuring during average runoff as in the present paper) is not followed exactly.

Acknowledgements

We thank the Federal Office for the Environment (FOEN) and the Federal Office of Meteorology and Climatology (MeteoSwiss) for providing extensive hydrological and meteorological data making this study possible. Two anonymous reviewers as well as the Associate Editor are acknowledged for their constructive comments on a previous version of the manuscript.

References

Addor, N., Jaun, S., Fundel, F., Zappa, M., 2011. An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios. Hydrol. Earth Syst. Sci. 15 (7), 2327–2347.

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R., Seibert, J., 2014. Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments. Water Resour. Res. http://dx.doi.org/10.1002/2014WR015549 (in press).

Ambroise, B., Freer, J., Beven, K.J., 1996. Toward a generalization of the TOPMODEL concepts: topographic indices of hydrological similarity. Water Resour. Res. 32, 2135–2145.

Aschwanden, H., Weingartner, R., 1985. Die Abflussregimes der Schweiz. Publikationen Gewässerkunde 65, Geographisches Institut der Universität Bern.

Archfield, S.A., Vogel, R.M., 2010. Map correlation method. Selection of a reference streamgage to estimate daily streamflow at ungaged catchments. Water Resour. Res. 46 (10), W10513.

Bárdossy, A., 2007. Calibration of hydrological model parameters for ungauged catchments. Hydrol. Earth Syst. Sci. 11, 703–710.

Bergström, S., 1972. Utveckling och tillämpning av en digital avrinningsmodell. [Development and application of a digital runoff model, in Swedish.] SMHI notiser och preliminära rapporter, serie hydrologi 22, Norrköping.

Beven, K.J., 2002. Towards a coherent philosophy for modelling the environment. Proc. Roy. Soc. London, Series A – Math., Phys. Eng. Sci. 458 (2026), 2465–2484.

Bitterli, T., Aviolat, P., Brändli, R., Christe, R., Fracheboud, S., Frey, D., George, M., Matousek, F., Tripet, J.-P., 2004. Groundwater Resources. Hydrological Atlas of Switzerland, Plate 8.6. Federal Office for the Environment, Bern.

Bosshard, T., Kotlarski, S., Zappa, M., Schär, C., 2014. Hydrological climate-impact projections for the Rhine River: GCM–RCM uncertainty and separate temperature and precipitation effects. J. Hydrometeorol. 15 (2), 697–713.

Burn, D.H., 1997. Catchment similarity for regional flood frequency analysis using seasonality measures. J. Hydrol. 202, 212–230.

Drogue, G.P., Plasse, J., 2014. How can a few streamflow measurements help to predict daily hydrographs at almost ungauged sites? Hydrol. Sci. J. 59 (12), 2126–2142.

Fundel, F., Zappa, M., 2011. Hydrological ensemble forecasting in mesoscale catchments: sensitivity to initial conditions and value of reforecasts. Water Resour. Res. 47 (9), W09520.

Fundel, F., Jörg-Hess, S., Zappa, M., 2013. Monthly hydrometeorological ensemble prediction of streamflow droughts and corresponding drought indices. Hydrol. Earth Syst. Sci. 17 (1), 395–407.

Garen, D.C., Marks, D., 2001. Spatial fields of meteorological input data including forest canopy corrections for an energy budget snow simulation model. In: Dolman, A.J., Hall, A.J., Kavvas, M.L., Oki, T., Pomeroy, J.W., (Eds.), Soil-Vegetation-Atmosphere Transfer Schemes and Large-Scale Hydrological Models. Proceedings of an International Symposium Held during the Sixth Scientific Assembly of the International Association of Hydrological Sciences (IAHS) at Maastricht, the Netherlands, from 18 to 27 July 2001. IAHS Publication 270, Wallingford, Oxfordshire, pp. 349–353.

Geiger, H., Röthlisberger, G., Stehli, A., Zeller, J., 1992. Extreme point rainfall of varying duration and return period 1901–1970. Hydrological Atlas of Switzerland, Plate 2.4. Federal Office for the Environment, Bern.

Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. Hydrol. 228, 113–129.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martínez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377 (1–2), 80–91.

Gurtz, J., Baltensweiler, A., Lang, H., 1999. Spatially distributed hydrotope-based modeling of evapotranspiration and runoff in mountainous basins. Hydrol. Process. 13, 2751–2768.

Hollander, M., Wolfe, D.A., 1999. Nonparametric Statistical Methods, second ed. Wiley, New York.

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB) – a review. Hydrol. Sci. J. 58 (6), 1198–1255.

Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.

Jensen, H., Lang, H., Rinderknecht, J., 1997. Extreme Point Rainfall of Varying Duration and Return Period 1901–1970 (updated map). Hydrological Atlas of Switzerland, Plate 2.4[2]. Federal Office for the Environment, Bern.

Jörg-Hess, S., Kempf, S.B., Fundel, F., Zappa, M., 2014. The benefit of climatological and calibrated reforecast data for simulating hydrological droughts in Switzerland. Meteorol. Appl. 22 (3), 444–458.

Juston, J., Seibert, J., Johansson, P.-O., 2009. Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment. Hydrol. Process. 23, 3093–3109.

Khu, S.-T., Werner, M.G.F., 2003. Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling. Hydrol. Earth Syst. Sci. 7 (5), 680–692.

Kokkonen, T.S., Jakeman, A.J., Young, P.C., Koivusalo, H.J., 2003. Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. Hydrol. Process. 17 (11), 2219–2238.

Kölla, E., 1987. Estimating flood peaks from small rural catchments in Switzerland. J. Hydrol. 95, 203–225.

Konz, M., Seibert, J., 2010. On the value of glacier mass balances for hydrological model calibration. J. Hydrol. 385, 238–246.

Köplin, N., Viviroli, D., Schädler, B., Weingartner, R., 2010. How does climate change affect mesoscale catchments in Switzerland? – a framework for a comprehensive assessment. Adv. Geosci. 27, 111–119.

Köplin, N., Schädler, B., Viviroli, D., Weingartner, R., 2012. Relating climate change signals and physiographic catchment properties to clustered hydrological response types. Hydrol. Earth Syst. Sci. 16 (7), 2267–2283.

Köplin, N., Schädler, B., Viviroli, D., Weingartner, R., 2013. The importance of glacier and forest change in hydrological climate-impact studies. Hydrol. Earth Syst. Sci. 17 (2), 619–635.

Köplin, N., Schädler, B., Viviroli, D., Weingartner, R., 2014a. Seasonality and magnitude of floods in Switzerland under future climate change. Hydrol. Process. 28 (4), 2567–2578.

Köplin, N., Rößler, O., Schädler, B., Weingartner, R., 2014b. Robust estimates of climate-induced hydrological change in a temperate mountainous region. Climatic Change 122 (1–2), 171–184.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97.

Legates, D.R., McCabe Jr., Gregory J., 1999. Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35 (1), 233–241.

Liechti, K., Panziera, L., Germann, U., Zappa, M., 2013. The potential of radar-based ensemble forecasts for flash-flood early warning in the southern Swiss Alps. Hydrol. Earth Syst. Sci. 17 (10), 3853–3869.

Lindström, G., 1997. A simple automatic calibration routine for the HBV model. Nord. Hydrol. 28 (3), 153–168.

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.

Margot, A., Schädler, B., Sigg, R., Weingartner, R., 1992. Influence on rivers by water power stations (⩾300 kW) and the lake control. Hydrological Atlas of Switzerland, Plate 5.3. Federal Office for the Environment, Bern.
MeteoSwiss [Federal Office of Meteorology and Climatology], 2010. SwissMetNet: an Observation Network for the Future, Zürich.
MeteoSwiss [Federal Office of Meteorology and Climatology], 2013. Time Series of Meteorological Variables, Zürich.
Meyer, R., Schädler, B., Viviroli, D., Weingartner, R., 2011. The role of base flow to estimate climate change impacts on low flow in the Swiss Midlands – determination, analysis and assessment. Hydrol. Wasserbewirtsch. 55 (5), 244–257.
Murrell, P., 2005. R Graphics. CRC Press, Boca Raton, FL.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models: Part I – a discussion of principles. J. Hydrol. 10, 282–290.
Oudin, L., Andréassian, V., Perrin, C., Michel, C., Le Moine, N., 2008. Spatial proximity, physical similarity and ungaged catchments: a comparison based on 913 French catchments. Water Resour. Res. 44, W03413.
Patil, S., Stieglitz, M., 2012. Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catchment. Hydrol. Earth Syst. Sci. 16 (2), 551–562.
Perks, A., Winkler, T., Stewart, B. (Eds.), 1996. The Adequacy of Hydrological Networks: A Global Assessment. World Meteorological Organisation (WMO), Geneva.
Perrin, C., Oudin, L., Andréassian, 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 (1), 131–151.
Perrin, C., Andréassian, V., Rojas Serna, C., Mathevet, T., Le Moine, N., 2008. Discrete parameterization of hydrological models: evaluating the use of parameter sets libraries over 900 catchments. Water Resour. Res. 44, W08447.
Romang, H., Zappa, M., Hilker, N., Gerber, M., Dufour, F., Frede, V., Bérod, D., Oplatka, M., Hegg, C., Rhyner, J., 2011. IFKIS-Hydro: an early warning and information system for floods and debris flows. Nat. Hazards 56 (2), 509–527.
Ross, B.B., Contractor, D.N., Shanholtz, V.O., 1979. A finite-element model of overland and channel flow for assessing the hydrologic impact of land-use change. J. Hydrol. 41 (1–2), 11–30.
Schaefli, B., Gupta, H.V., 2007. Do Nash values have value? Hydrol. Process. 21 (15), 2075–2080.
Schattan, P., Zappa, M., Lischke, H., Bernhard, L., Thürig, E., Diekkrüger, B., 2013. An approach for transient consideration of forest change in hydrological impact studies. In: Boegh, E., Blyth, E., Hannah, D.M., Hisdal, H., Kunstmann, H., Su, B., Yilmaz, K.K. (Eds.), Climate and Land Surface Changes in Hydrology. IAHS, Wallingford, pp. 311–319.
Seibert, J., Beven, K.J., 2009. Gauging the ungauged basin: how many discharge measurements are needed? Hydrol. Earth Syst. Sci. 13, 883–892.
Seibert, J., McDonnell, J.J., 2013. Gauging the ungauged basin: the relative value of soft and hard data. J. Hydrol. Eng., A4014004
SFSO [Swiss Federal Statistical Office], 2003. GEOSTAT Database Products. Licence G158000315, ©SFSO, Neuchâtel.
Singh, S.K., Bárdossy, A., 2012. Calibration of hydrological models on hydrologically unusual events. Adv. Water Resour. 38, 81–91.
Viviroli, D., 2007. Ein prozessorientiertes Modellsystem zur Ermittlung seltener Hochwasserabflüsse für ungemessene Einzugsgebiete der Schweiz. Geographica Bernensia G77. Institute of Geography, University of Bern, Bern, ISBN 978-3-90583500-7.
Viviroli, D., Weingartner, R., 2011. Comprehensive flood hydrological assessment of ungauged catchments in the Swiss part of the River Rhine basin by means of process-oriented modelling. Hydrol. Wasserbewirtsch. 55 (5), 258–272.
Viviroli, D., Gurtz, J., Zappa, M., 2007. The Hydrological Modelling System PREVAH. Geographica Bernensia P40. Institute of Geography, University of Bern, ISBN 978-3-905835-01-0.
Viviroli, D., Zappa, M., Gurtz, J., Weingartner, R., 2009a. An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. Environ. Model. Softw. 24 (10), 1209–1222.
Viviroli, D., Zappa, M., Schwanbeck, J., Gurtz, J., Weingartner, R., 2009b. Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part I: modelling framework and calibration results. J. Hydrol. 377 (1–2), 191–207.
Viviroli, D., Mittelbach, H., Gurtz, J., Weingartner, R., 2009c. Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part II: parameter regionalisation and flood estimation results. J. Hydrol. 377 (1–2), 208–225.
Viviroli, D., Archer, D.R., Buytaert, W., Fowler, H.J., Greenwood, G.B., Hamlet, A.F., Huang, Y., Koboltschnig, G., Litaor, M.I., López-Moreno, J.I., Lorentz, S., Schädler, B., Schreier, H., Schwaiger, K., Vuille, M., Woods, R., 2011. Climate change and mountain water resources: overview and recommendations for research, management and policy. Hydrol. Earth Syst. Sci. 15 (2), 471–504.
Weingartner, R., Aschwanden, H., 1992. Discharge regime – the basis for the estimation of average flows. Hydrological Atlas of Switzerland, Plate 5.2. Federal Office for the Environment, Bern.
Young, A.R., 2006. Stream flow simulation within UK ungauged catchments using a daily rainfall–runoff model. J. Hydrol. 320, 155–172.
Zappa, M., Kan, C., 2007. Extreme heat and runoff extremes in the Swiss Alps. Nat. Hazards Earth Sys. Sci. 7, 375–389.
Zappa, M., Bernhard, L., Spirig, C., Pfaundler, M., Stahl, K., Kruse, S., Seidl, I., Stähli, M., 2014. A prototype platform for water resources monitoring and early recognition of critical droughts in Switzerland. In: Castellarin, A., Ceola, S., Toth, E., Montanari, A., (Eds.), Evolving Water Resources Systems. Understanding, Predicting and Managing Water–Society Interactions. Proceedings of ICWRS2014, Bologna, Italy, June 2014. IAHS, Wallingford, pp. 492–498.