



Calibration of a rainfall–runoff model at regional scale by optimising river discharge statistics: Performance analysis for the average/low flow regime

Laura Lombardi, Elena Toth*, Attilio Castellarin, Alberto Montanari, Armando Brath

Department DICAM, University of Bologna, Viale del Risorgimento 2, Bologna I-40136, Italy

ARTICLE INFO

Article history:

Available online 20 July 2011

Keywords:

Rainfall–runoff
Runoff variability
Multi-objective optimisation
Regional calibration
Low flows

ABSTRACT

Traditional procedures for rainfall–runoff model calibration are generally based on the fit of individual values of simulated and observed hydrographs. We use here an alternative option that is carried out by matching, in the optimisation process, a set of streamflow statistics. Such an approach has the significant advantage to enable also a straightforward regional calibration of model parameters, based on the regionalisation of the selected statistics. The minimisation of the set of objective functions is carried out by using the AMALGAM algorithm, leading to the identification of behavioural parameter sets. The procedure is applied to a set of river basins located in central Italy: the basins are treated alternatively as gauged and ungauged and, as a term of comparison, the results obtained with a traditional time-domain calibration are also presented. With respect to previous applications of analogous procedures, we investigate here the identification of the target statistics depending on the purposes of the application, and in particular when the focus is on the reproduction of the low-flows. The results show that a suitable choice of the statistics to be optimised leads to interesting results in real world case studies as far as the reproduction of the different flow regimes is concerned.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The procedures traditionally adopted for estimating the parameters of conceptual rainfall–runoff models are based on the fit of individual values of simulated and observed hydrographs, trying to minimise the sum of the simulation errors corresponding to each time step. We use here an alternative approach that is carried out by matching, in the optimisation process, quantitative signatures of the river flow regime (Montanari and Toth, 2007; Schaeffli and Zehe, 2009), and in particular a set of statistics of the river flow (see Castiglioni et al., 2010).

Such an approach has the additional, significant advantage to enable a straightforward regional calibration of model parameters, based on the regionalisation of the selected statistics. Regional calibration of rainfall–runoff (R–R) models is increasingly studied by hydrologists. Indeed, R–R models are more and more used for solving technical problems and very often one has to calibrate model parameters in conditions of data scarcity. One possible solution to the above problem is regional calibration, which is based on the idea of optimising model performances in the simulation of regional information. While many contributions previously focused

on the possibility to directly estimate R–R model parameters with regional relationships (see, for instance, Hundedcha et al., 2008; Zhang et al., 2008; Yadav et al., 2007; Pokhrel et al., 2008; Wagener and Wheeler, 2006; Parajka et al., 2005; Bárdossy, 2007), a number of recent contributions have focused on the idea of seeking the model structure and parameters which provide behavioural simulations of river flow statistics derived through regional studies (see, for instance, Castiglioni et al., 2010).

An obvious advantage of this approach lies in the abundance of studies reported in the literature, which proposes several examples of application of regional approaches for estimating various hydrological variables, such as catchment geo-climatic attributes (see, for instance, in the Apennine region, Brath et al., 2002, 2003) and river flow statistics (Vogel and Kroll, 1992; Stedinger et al., 1993; Hosking and Wallis, 1997; Flood Estimation Handbook, 1999; Pandey and Nguyen, 1999; Furey and Gupta, 2000; Smakhtin, 2001; Castellarin et al., 2004; Laaha and Blöschl, 2006).

To calibrate a R–R model by optimising river flows statistics one needs the availability of long series of input data for the model itself (typically rainfall data). Then, the model is run by using trial values for its parameters and river flows are simulated. These simulated data allow one to estimate river flow statistics that are then compared with those obtained from the observed streamflow data (not necessarily over the same period in which the input data are available, see Montanari and Toth, 2007) or through regionalisation.

* Corresponding author. Tel.: +39 051 2093364.

E-mail addresses: laura.lombardi@unibo.it (L. Lombardi), elena.toth@unibo.it (E. Toth), attilio.castellarin@unibo.it (A. Castellarin), alberto.montanari@unibo.it (A. Montanari), armando.brath@unibo.it (A. Brath).

One key issue is the selection of the statistics to optimise. To address this question, we refer to Gupta et al. (2009) who analysed the properties of estimators based on the minimisation of quadratic goodness of fit indexes. In particular, they showed that minimisation of the model mean square error (least squares calibration) is equivalent to matching mean value, μ , and standard deviation, σ , while maximising cross correlation, of observed and simulated river flow series. Under certain hypotheses, it can be shown that maximising cross correlation can be substituted by matching the lag one autocorrelation coefficients, ρ_1 (for more details see Montanari and Toth (2007) and the discussion in Montanari (in preparation)). This is the reason why Castiglioni et al. (2010) adopted a calibration strategy based on matching the above statistics μ , σ and ρ_1 .

However, when one needs to focus on different river flow regimes, for instance low flows, one may need a different selection of statistics to be optimised. In fact, low flows are often poorly reproduced even when using traditional calibration procedures, while they are of primary importance when dealing with water resources management problems. As above said, in fact, a quadratic objective function is approximated with the approach described above, and such a function tends to emphasise the higher errors, generally corresponding to high flows: it is therefore advisable to include low flow statistics too, to ensure that model performances are evaluated over a wide spectrum of river flow regimes. In addition, in the presence of uncertainty (see e.g. Montanari, 2007) it may be desirable to increase the number of the statistics to be optimised.

For the above reasons, in this study we also include, among the statistics to be optimised, the river flow that is equalled or exceeded for 355 days per year on average, Q_{355} . Q_{355} is commonly adopted in Italy for characterising the low-flow regime, and it is highly correlated to other low-flow indexes used around the world. Several studies proposed methods for regionalising Q_{355} (Castiglioni et al., 2009) and therefore such type of information seems to be worth exploiting. We analyse here the differences in the model performances, with particular focus on low flows.

The above calibration procedure is applied to 5 catchments located in a fairly hydrologically homogeneous region of central Italy. The results show that matching the Q_{355} statistic can provide a valuable contribution to model calibration for simulating the overall flow regimes and the runoff variability for low to average streamflows.

1.1. Description of the calibration procedure

We applied a multi-objective calibration procedure (to simultaneously match the selected statistics) therefore identifying the full Pareto set of the non-dominated parameter vectors $\{\theta\}$ for the R–R model (Yapo et al., 1998), which enables one to analyse a wider range of behavioural simulations in the presence of uncertainty in each of the above statistics. In this way the user might also choose the fitting accuracy for mean value, variability and persistence properties, depending on the scopes of the problem at hand and the specific behaviours of the case study.

The multiobjective optimisation was carried out by using a recently developed multimethod evolutionary search, named AMALGAM (Vrugt and Robinson, 2007). A Multi ALgorithm Genetically Adaptive Method (AMALGAM) runs simultaneously, for population evolution, a set of different optimisation methods (namely NSGA-II, Differential Evolution, Adaptive Metropolis Search and Particle Swarm Optimisation), resulting in a combination of the respective strengths by adaptively updating the weights of these individual methods based on their reproductive success (for more details see Vrugt and Robinson (2007) and references therein). This ensures a fast, reliable and computationally efficient solution to

multiobjective optimisation problems. The self-adaptive search properties of AMALGAM make it able to quickly adjust to the specific peculiarities and difficulties of each search problem and, in addition, they reduce the need for tuning of the algorithmic parameters.

In detail, we optimise with AMALGAM the following objective functions:

$$\phi_1(\theta) = \left[\frac{\mu[Q_s(\theta)] - \mu(Q)}{\mu(Q)} \right]^2 \quad (1)$$

$$\phi_2(\theta) = \left[\frac{\sigma[Q_s(\theta)] - \sigma(Q)}{\sigma(Q)} \right]^2 \quad (2)$$

$$\phi_3(\theta) = \left[\frac{\rho_1[Q_s(\theta)] - \rho_1(Q)}{\rho_1(Q)} \right]^2 \quad (3)$$

$$\phi_4(\theta) = \left[\frac{Q_{355}[Q_s(\theta)] - Q_{355}(Q)}{Q_{355}(Q)} \right]^2 \quad (4)$$

where $\mu[Q_s(\theta)]$, $\sigma[Q_s(\theta)]$, $\rho_1[Q_s(\theta)]$ and $Q_{355}[Q_s(\theta)]$ are the mean, standard deviation, lag-one autocorrelation coefficient and Q_{355} of the simulated runoff series, and $\mu(Q)$, $\sigma(Q)$, $\rho_1(Q)$ and $Q_{355}(Q)$ are the corresponding empirical values. In the ungauged analysis, the observed statistics will be replaced by those obtained with regional relationships (see Section 3.1).

2. Rainfall-runoff model

The calibration method outlined above is in principle applicable to any R–R model. However, in view of the application to the ungauged case it is advisable to restrain parameter uncertainty by using a parsimonious model. In this study we use HYMOD, a five-parameter lumped and conceptual model that was proposed by Boyle (2000) and recently used by Wagener et al. (2001), Vrugt et al. (2003) and Montanari and Toth (2007) among others. This model consists of a relatively simple rainfall excess model that is connected to two series of linear reservoirs. In detail, three identical reservoirs in series are used to simulate the quick response and a single reservoir is adopted for the slow response. HYMOD input data are precipitation and evapotranspiration time series in addition to catchment area. The five parameters of the model are the maximum water storage capacity of the soil over the catchment, C [mm], a parameter representing the degree of spatial variability of the water storage capacity itself over the catchment, b [–], the fraction of rainfall excess that flows downstream as quick response, a [–], and the fractions of quick, R_q [–], and slow, R_s [–], reservoirs that empty at each time step.

3. Study area

The study area is a wide geographical region of northern central Italy including 52 catchments for which at least five years of daily series of streamflow records are available (Fig. 1). The daily streamflow series for the set of study catchments are not significantly affected by water storages and withdrawals, and the study region covers a wide spectrum of hydrologic conditions. The study region includes scarcely urbanized, mountainous and hilly, as well as pervious and practically impervious basins, with the upper part covered by pastures and broad leaved woods (see also Castellarin et al., 2004, Table 1 on p. 958 and Section 6.3).

Simultaneous daily time series of streamflow, precipitation and evapotranspiration with continuous record lengths varying from two to four years are available for a subset of 5 catchments (Candigliano River at Acqualagna, Metauro River at Barco di Bellaguardia, Esino River at Moie, Potenza River at Cannuciaro, and Tenna River at Amandola). Table 1 illustrates the main hydrological features of these catchments in detail by reporting the values of the

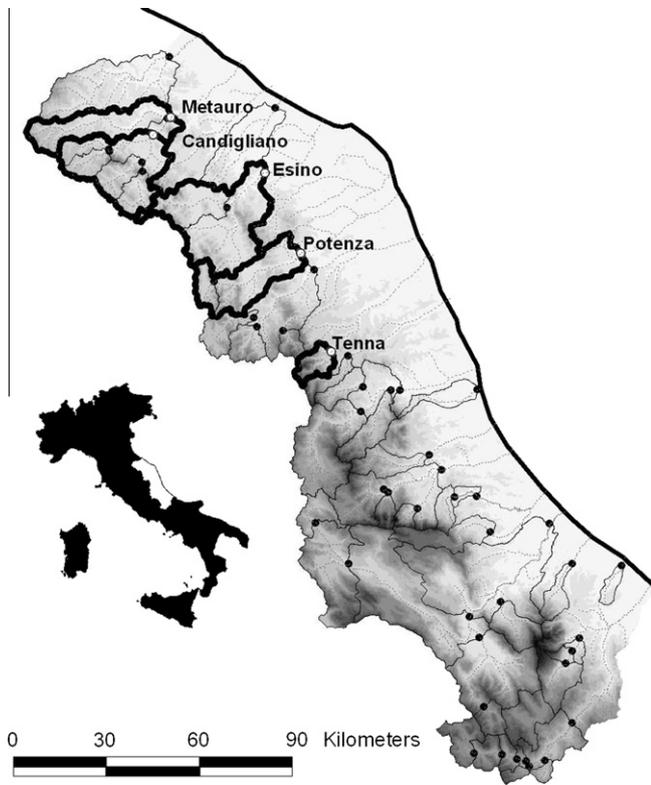


Fig. 1. Study area: 52 basins utilized in the regional analysis (thin black lines) and corresponding streamgauges (black dots); 5 study catchments where the calibration procedure was tested in ungauged mode (thick black lines) and corresponding streamgauges (white dots).

same geomorphological and climatic catchment descriptors considered in the regional analysis: drainage area, A (km²); percentage of permeable area, P (%); maximum, mean and minimum elevations, H_{max} , H_{mean} and H_{min} (m above the sea level, m asl); average elevation relative to H_{min} , $\Delta H = H_{mean} - H_{min}$ (m); main channel length, L (km); concentration time, t_c (h); mean annual precipitation, MAP (m); mean annual temperature, TAM (C°). The concentration time, t_c (h), was computed through Giandotti's equation (Giandotti, 1937; Castiglioni et al., 2009). The observation period spans from 1951 to 1978. Mean areal precipitation over the catchment was computed by applying the Thiessen polygons technique (Castellarin et al., 2004). Daily potential evapotranspiration series at catchment scale were derived from the Thornthwaite (Thornthwaite, 1946) equation. The quality of the data was tested and certified by the former National Hydrographic Service of Italy.

3.1. Regional analysis

To optimise the HYMOD parameters in the ungauged conditions experiment, a regional analysis was first developed in order to estimate $\mu(Q)$, $\sigma(Q)$, $\rho_1(Q)$ and $Q_{355}(Q)$ depending on the geomorphological and climatic attributes.

Table 1
Geomorphological and climatic descriptors of the 5 study catchments considered for the rainfall–runoff modelling.

	A (km ²)	P (%)	H_{max} (m asl)	H_{mean} (m asl)	H_{min} (m asl)	ΔH (m)	L (km)	t_c (h)	MAP (m)	TAM (C°)
Candigliano	613.8	26.0	1702	600	183	417	56.3	11.2	1118.4	12.1
Metauro	1043.6	20.0	1702	560	110	450	75.7	14.3	1115.0	12.5
Esino	798.2	47.5	1702	529	96	433	69.9	13.1	1228.2	12.4
Potenza	430.7	57.0	1570	616	168	448	58.4	10.1	1096.8	12.2
Tenna	99.7	71.0	2334	1170	425	745	19.4	3.2	899.9	11.1

Castiglioni et al. (2010) developed regional multiregression models for predicting μ , σ and ρ_1 in the study region. The authors regressed the catchment descriptors of the 52 sites (see Table 1) against the corresponding empirical values of μ , σ and ρ_1 through a multivariate stepwise regression analysis (see Weisberg, 1985). Then, the regional models were validated through a jack-knife cross-validation procedure (see also Castellarin et al., 2004; Castellarin et al., 2007). In detail, one catchment is supposed to be ungauged and therefore the related streamflow data-set is discarded. Then, multiregression equations are recalibrated against the reduced data base. The obtained regional models are then used to estimate the statistics for the hypothesized ungauged site. The procedure is repeated by excluding all stations in turn, thereby obtaining a set of 52 jack-knifed estimates of each statistic in ungauged mode.

Castiglioni et al. (2009) and Castiglioni et al. (2011) also developed for the study area a smooth regional estimator of $Q_{355}(Q)$ through the application of the Physiographic-Space Based Interpolation technique (PSBI, Chokmani and Ouarda, 2004). PSBI spatially interpolates the hydrometric index of interest (in our case Q_{355}) in a two-dimensional space of physiographic and climatic descriptors, the so-called physiographic space. The x and y coordinates of the physiographic space are derived from an adequate set of $n > 1$ geomorphoclimatic catchment descriptors (in our case the descriptors listed in Table 1) through the application of the Principal Component Analysis (PCA). Once defined the physiographic space, any given basin (gauged or ungauged) can be represented as a point in the $x - y$ space. The empirical values of the quantity of interest (i.e., Q_{355}) can be represented along the third dimension z for each gauged catchment, and can then be spatially interpolated by applying a standard interpolation algorithm (e.g., ordinary kriging). The spatial interpolation enables one to represent Q_{355} over the entire portion of the $x - y$ space containing empirical data, and therefore to estimate it at ungauged sites lying within the same portion of the space. A jack-knife resampling procedure was then adopted to validate the proposed regional estimator and showed that PSBI is more accurate than multiregression for predicting Q_{355} in ungauged sites over the study region.

Since we exploit the results of the jack-knife validation exercises performed in Castiglioni et al. (2010) and Castiglioni et al. (2009), we report them in Fig. 2, for the sake of completeness. The reader is referred to these studies for further details.

4. Rainfall–runoff model calibration

4.1. Baseline calibration in the time-domain

In order to evaluate the results of the proposed calibration procedure, we first performed, as a baseline calibration, a traditional calibration in the Time-Domain (TD), by minimising the Root Mean Square Error (RMSE) of the simulated and observed hourly time-series, from now on referred to as TD calibration. This approach, certainly the most widely applied in calibration procedures, aims at fitting the individual values of the observed reference time series and in particular, since the objective function is of the quadratic

type, it is strongly sensitive to high and extreme error values, that typically correspond to high and extreme streamflow values.

4.2. Calibration optimising the observed statistics

The proposed methodology based on the fit of the streamflow statistics is first applied by referring to a gauged case, using the empirical values of the statistics, estimated by using the hydrometric series that were actually observed at the watershed outlets. Different subsets of the four statistics were simultaneously matched in the calibration: as a reference, the match of $\mu(Q)$, $\sigma(Q)$ or $\rho_1(Q)$ was considered, as proposed in Castiglioni et al. (2010): such calibration will be referred to as Mean, Variance, Autocorrelation (MVA). Secondly, the $Q_{355}(Q)$ statistics is added to the three above, implementing the MVA Q_{355} multiobjective calibration. Lastly, in order to give more weight to the goodness of fit of the Q_{355} statistic, the set formed by $\mu(Q)$, $\sigma(Q)$ and $Q_{355}(Q)$ is considered, thus neglecting the match of $\rho_1(Q)$ in the MV Q_{355} calibration. The reason why $\rho_1(Q)$ was eliminated is its sensitivity to high flows, being correlation a quadratic measure, and given the well known different behaviours of correlation for low flows with respect to high flows, which are governed by different hydrological processes. Some preliminary studies confirmed the above conclusion for the study area and the available data sets.

AMALGAM is used to calibrate the model by minimising the three different chosen sets of objective functions (1)–(4), thus identifying the model parameter sets that generate simulated series characterised by statistics close to those of the observed series. For each set of statistics to be matched (MVA, MVA Q_{355} , MV Q_{355}), using a multiobjective search, we are able to find a Pareto set of solutions and not a single optimal solution. In order to compare the results with the single optimal solution of the TD calibration, we identify, in the Pareto Front, the point that is closest to the origin of the axes (as illustrated in Fig. 3), from now on referred to as MVA-O, MVA Q_{355} -O, MV Q_{355} -O simulations.

4.3. Calibration optimising the regionalised statistics (ungauged case)

In ungauged conditions, AMALGAM is used to fit the statistics of the simulated series to the corresponding ones calculated through the regional regression equations. The algorithm minimises in fact the differences obtained through the different subsets (MVA, MVA Q_{355} , MV Q_{355}) of the objective functions (1)–(4), where $\mu(Q)$, $\sigma(Q)$, $\rho_1(Q)$ and $Q_{355}(Q)$ were replaced by their estimates $\hat{\mu}$, $\hat{\sigma}$, $\hat{\rho}_1$ and \hat{Q}_{355} obtained using the regionalisation procedure presented in Section 3.1 – through a jack-knife resampling procedure.

For each subset of the statistics, having identified the Pareto set of the non-dominated solutions for the study area in ungauged conditions, we were then able to single out the parameter set in the Pareto front that is the closest to the origin of the axes, from now on referred to as MVA-R, MVA Q_{355} -R and MV Q_{355} -R simulations. A comparison of the calibrations in the ungauged case (MVA-R, MVA Q_{355} -R, MV Q_{355} -R) with the calibrations based on the observed statistics (MVA-O, MVA Q_{355} -O, MV Q_{355} -O) will allow a quantification of the additional uncertainty introduced by using regional statistics instead of those obtained from the observed data.

5. Results and discussion

5.1. Performance Indexes

The performance indexes used for evaluating hydrological models are generally of the quadratic type (mean squared error, correlation coefficient, Nash–Sutcliffe efficiency, etc.) but, as above said,

the use of squares implies a greater influence on the index by way of the larger flow values. This is appropriate when, like in many streamflow forecasting applications, the focus is on the ability to reproduce potentially dangerous flood events, but not when the aim of the modelling is the reproduction of low and average flows. If we are interested in all the flow regimes, we should in fact chose a non-quadratic index, as confirmed by the results obtained by Legates and McCabe (1999) and by Krause et al. (2005), who recommend the use of a modification of the index of agreement first introduced by Willmott (Willmott, 1981; Willmott et al., 1985):

$$d_j = 1 - \frac{\sum_{t=1}^N |Q_s(t) - Q(t)|^j}{\sum_{t=1}^N (|Q_s(t) - \mu(Q)| + |Q(t) - \mu(Q)|)^j} \quad (5)$$

where t is the time instant, Q and Q_s are the observed and simulated streamflow, respectively, $\mu(Q)$ is the mean value of Q , N is the total number of simulation steps and j is a weighting exponent.

The index of agreement varies from 0.0 to 1.0, with higher values indicating better agreement between the model and observations. In the original version, j was equal to 2, thus resulting sensitive to extreme values; for increasing the sensitivity for lower values, $j = 1$ may be used, so that the errors are given their appropriate weighting, resulting in a more overall sensitivity measure for the quality of the model results during the entire period (Legates and McCabe, 1999; Krause et al., 2005).

As suggested by Legates and McCabe (1999), a complete assessment of model performance should include, in addition to a relative error measure (like the modified index of agreement d_1), also one absolute error measure (e.g., root mean square error, RMSE, or mean absolute error, MAE), that provides an evaluation of the error in the units of the variable. In the present study, in order to avoid the inflated sensitivity to the larger values, MAE is preferred over RMSE:

$$MAE = \frac{\sum_{t=1}^N |Q_s(t) - Q(t)|}{N} \quad (6)$$

The mean absolute error varies from 0 to $+\infty$, for increasing discrepancy, and gives the same weight to all errors.

In addition to the above performance indexes – that are based on the fit of individual values of the observed reference time series – we add an evaluation of the model performances in terms of a signature of runoff variability (Atkinson et al., 2002; Yanmaz et al., 2008). This approach enables us to “assess how well the model is able to predict the streamflow response of the catchment without the reliance on hydrograph fitting, and moreover, can give us considerable insight into catchment response.” (Atkinson et al., 2002).

In particular, a comparison of observed and simulated flow duration curves is here considered; such comparison is deemed by Pickup (1977) to be “probably the most effective test of model performance”. In fact, an assessment of the model performance on a frequency basis is the most useful in case systematic errors are present, since it tends to remove most of the effect of random data errors (Pickup, 1977).

Flow-duration curves have previously been used either as objective functions in model calibration or as measures of model performance, among the others, by Refsgaard and Knudsen (1996), Houghton-Carr (1999), Yu and Yang (2000), Andersen et al. (2002), Farmer et al. (2003), Butts et al. (2004), Blazkova and Beven (2009) and Westerberg et al. (2010).

For a quantitative evaluation of the predicted flow duration curves, different exceedance percentages have been considered

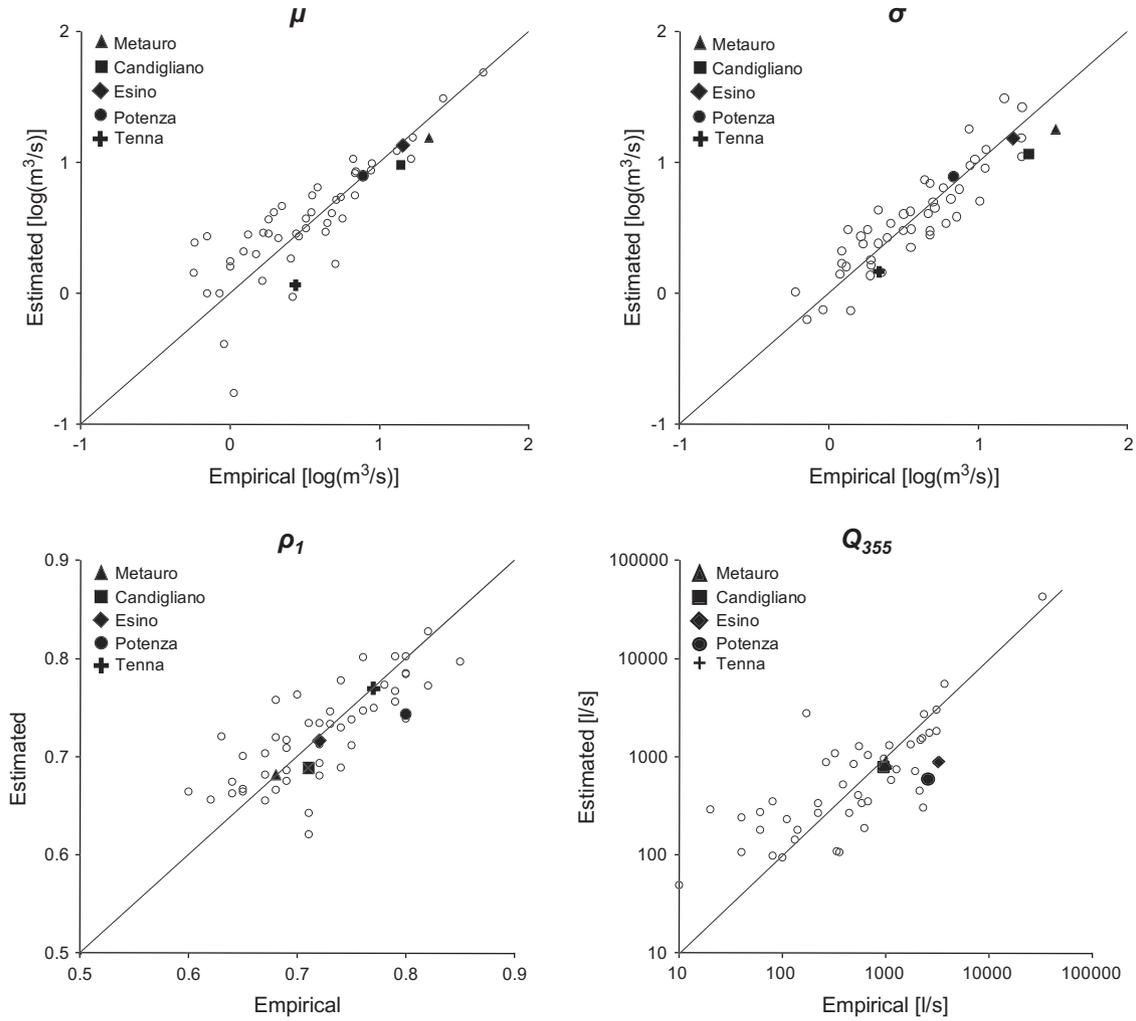


Fig. 2. Empirical values versus cross-validated regional estimates of $\mu(Q)$, $\sigma(Q)$, $\rho_1(Q)$ and Q_{355} (see Castiglioni et al., 2009; Castiglioni et al., 2010); values for the 5 test sites are also highlighted.

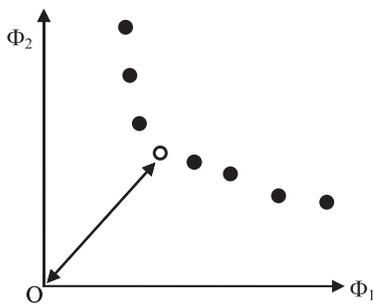


Fig. 3. Illustration of the concept of Pareto front (i.e. the set of non-dominated solutions) for a problem having two criteria (Φ_1 , Φ_2), in the objective space. The white circle is the element of the front having the minimum (Euclidean) distance from the origin of the axes.

and the sum of the absolute values of the differences between observed and simulated quantiles have been computed:

$$ErrFDC = \frac{\sum_{i=1}^K |P_{o,k_i} - P_{s,k_i}|}{K} \quad (7)$$

where P_{o,k_i} and P_{s,k_i} are the discharges of exceedance percentage $k_i(\%)$, in the observed and in the simulated flow duration curve,

respectively, and K is the number of quantiles taken into consideration.

Since the proposed calibration approach has been conceived in order to improve in particular the reproduction of the low and average flow behaviour, the $K = 10$ flow-duration percentiles from 50% to 95% exceedance were chosen to characterise the low flow part of duration curve: $k_i = 50\%, 55\%, \dots, 90\%, 95\%$.

5.2. Discussion of calibration results

Table 2 reports the MAE calculated on the simulated series modelled by using the TD calibration, the calibrations based on the observed statistics (MVA-O, MVA_{Q355}-O, MVQ₃₅₅-O) and the calibrations based on regionalized statistics (MVA-R, MVA_{Q355}-R, MVQ₃₅₅-R).

Analogously, Table 3 shows the modified index of agreement (d_1) and Table 4 the values obtained for the flow duration curve error ($ErrFDC$) of the same simulations.

We will first analyze the results of the calibration based on the match of the statistics of the observed streamflows: MVA-O, MVA_{Q355}-O, MVQ₃₅₅-O, comparing them to the baseline calibration in the time domain.

Tables 2 and 3 show that when considering the agreement of individual values, considering all the flow regimes, the traditional TD calibration allows lower MAE (with the only exception of the

Table 2
MAE obtained with time-domain calibration (TD) and with the fit of observed (-O) and regionalised (-R) statistics, for the test catchments.

MAE (m ³ /s)	Candigliano	Metauro	Esino	Potenza	Tenna
TD	3.3	7.0	7.0	1.8	0.7
MVA – O	3.8	7.7	8.4	2.4	1.1
MVA – R	4.1	10.0	7.5	3.1	1.3
MVAQ ₃₅₅ – O	3.2	7.8	8.9	2.6	1.1
MVAQ ₃₅₅ – R	3.7	7.8	7.3	3.1	1.2
MVQ ₃₅₅ – O	4.5	8.6	9.2	2.4	1.0
MVQ ₃₅₅ – R	3.3	11.3	8.9	3.0	1.2

Table 3
Index of agreement obtained with time-domain calibration (TD) and with the fit of observed (-O) and regionalised (-R) statistics, for the test catchments.

d_1	Candigliano	Metauro	Esino	Potenza	Tenna
TD	0.79	0.80	0.71	0.81	0.76
MVA – O	0.76	0.78	0.67	0.73	0.69
MVA – R	0.73	0.67	0.68	0.67	0.55
MVAQ ₃₅₅ – O	0.80	0.78	0.65	0.70	0.67
MVAQ ₃₅₅ – R	0.77	0.77	0.70	0.69	0.57
MVQ ₃₅₅ – O	0.73	0.76	0.63	0.73	0.69
MVQ ₃₅₅ – R	0.79	0.67	0.66	0.70	0.56

Table 4
Flow duration curve error obtained with time-domain calibration (TD) and with the fit of observed (-O) and regionalised (-R) statistics, for the test catchments.

ErrFDC (m ³ /s)	Candigliano	Metauro	Esino	Potenza	Tenna
TD	1.1	1.1	1.9	0.5	0.1
MVA – O	1.2	0.8	1.8	0.2	0.3
MVA – R	1.2	2.2	2.3	0.3	0.3
MVAQ ₃₅₅ – O	0.7	0.3	0.8	0.3	0.1
MVAQ ₃₅₅ – R	0.8	0.6	2.9	0.6	0.2
MVQ ₃₅₅ – O	0.5	0.3	0.3	0.1	0.1
MVQ ₃₅₅ – R	1.3	1.1	3.3	0.7	0.2

Candigliano) and higher d_1 than the calibrations based on the match of the statistics. Such result was expectable, given the nature of the TD calibration, that is based exactly on the minimisation of this type of individual discrepancy. Comparing the calibrations based on the fit of the observed statistics, the two tables demonstrate that, among the subsets of the four statistics, it is necessary to include the match of the autocorrelation coefficient, since the calibration based on the fit of mean, variance and Q_{355} alone (MVQ₃₅₅-O) leads to the highest MAE (with the only exception of the MAE values for the Tenna river, that are all very close) and to the lowest index of agreement. It is therefore shown the importance of an adequate reproduction of the lag one autocorrelation coefficient when aiming at a good minimisation of the individual errors considering all the flow regimes, thus reaching performances closer to that of the squared error minimisation in the time domain. In fact, as stated above, the minimisation of the model mean square error, may be considered, under certain hypotheses, equivalent to matching mean value, variance and lag one autocorrelation coefficients, following Gupta et al. (2009).

On the other hand, when considering an assessment of the model performance on the low/average river regime, the ErrFDC (Table 4) indicates that the (MVQ₃₅₅-O) calibration is always the best performing, followed by the (MVAQ₃₅₅-O) and, at a greater distance, by the MVA-O and the TD calibrations. It follows that including in the calibration procedure the match of the Q_{355} statistic seems to be crucial for a satisfactory fit of the flow duration curve, allowing a good reproduction of the low to average flow quantiles and outperforming the time domain calibration. It is also confirmed that matching autocorrelation may lead to a poorer fit of the low-flows regime.

When considering the calibrations based on the regionalized statistics (-R calibrations), there is an overall deterioration of all the performance indexes in comparison to the use of the observed statistics. This result was expected, in view of the additional uncertainty introduced by regional statistics with respect to those obtained from the observed data.

In particular, the regional simulations on the Metauro and Tenna watersheds highlight a clear effect of underprediction due to a poor reproduction of the mean and/or of the variance values. In order to illustrate this effect, Fig. 4 shows the scatterplots of observed river flows versus the simulations for the Metauro river (other scatterplots are not presented for sake of brevity). It may be noticed a strong underestimation in the regionalized calibrations for the Metauro river, whose regionalized mean and especially variance are sensibly lower than the observed ones.

As far as the Esino river basin is concerned, Table 4 highlights a remarkable deterioration of the ErrFDC when the calibration is based on the match of the regionalised Q_{355} . This is due to the inadequate reproduction of such statistic: the regionalised value is in fact much less than the empirical one (0.88 vs. 3.23 m³/s), thus causing a strong deterioration of the lowest quantiles of the MVAQ₃₅₅-R and MVQ₃₅₅-R calibrations.

For the Candigliano and the Potenza river basins, that have the highest discrepancy between the observed and regionalized lag one autocorrelation coefficients (as may be seen in Fig. 2), the best MAE and d_1 are obtained by the calibration that does not include such statistic (MVQ₃₅₅-R). This result confirms, once again, the importance of a good reproduction of the lag one correlation coefficient when aiming at a good minimisation of the individual errors.

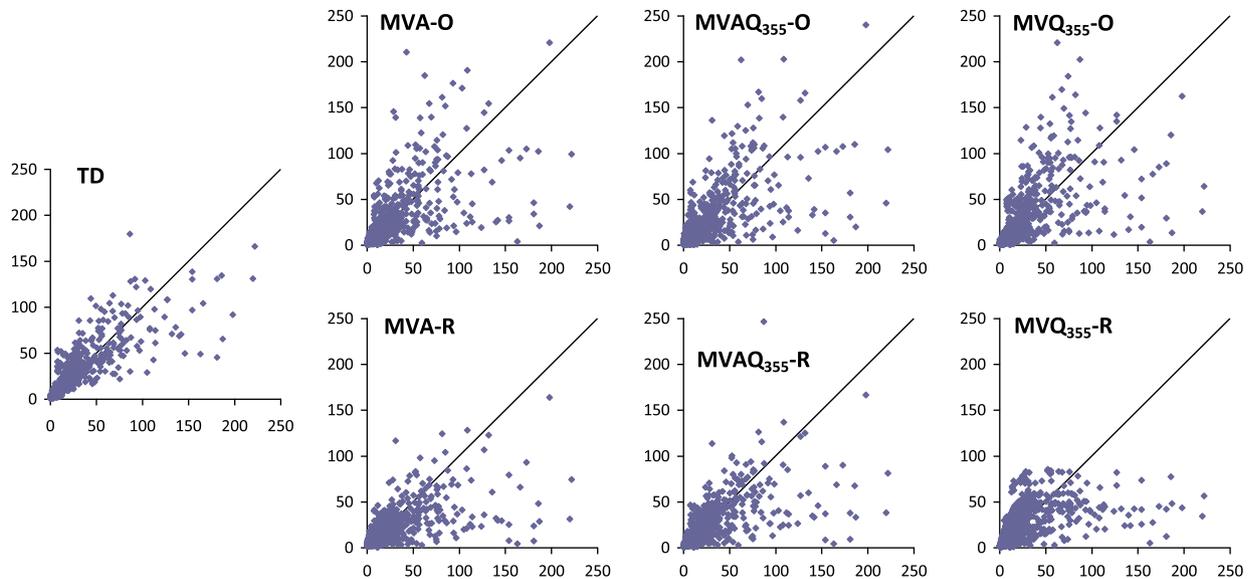


Fig. 4. Metauro River at Barco di Bellaguardia. Scatterplot of observed versus simulated river flows obtained with time-domain calibration (TD) and with the fit of observed (-O) and regionalised (-R) statistics.

6. Conclusions

The idea proposed in this paper is to implement a calibration procedure for rainfall–runoff models based on the fit of quantitative signatures of the river flow regime, which may provide interesting perspectives also in ungauged conditions. The approach maximises, through the AMALGAM multiobjective calibration procedure, the match between different streamflow statistics simulated by the HYMOD model and the corresponding empirical estimates; in ungauged conditions, the empirical values are replaced by regional estimates of the selected statistics.

With respect to previous applications of analogous procedures, we have analysed herein the potential improvement allowed by identifying the target statistics as a function of the hydrological application, and in particular when the focus is on the reproduction of the low-flows.

The procedure is applied to a set of river basins located in central Italy: the basins are treated alternatively as gauged and ungauged and, as a term of comparison, the results obtained with a traditional time-domain calibration are also presented.

The analysis shows that a suitable choice of the statistics to be optimised leads to interesting results in real world case studies as far as the reproduction of the different flow regimes is concerned. In particular, if simulation of the low to average flow regime is the main target, low flow indexes provide an information certainly worth exploiting, which also lend itself to regionalisation. The type of information to be considered in the model parameterisation should therefore be carefully selected in view of the practical needs required for the considered application.

The regional calibration procedure is potentially able to convey useful information, although in some cases the presence of relevant uncertainty makes it unlikely that the regional information is enough to calibrate a rainfall–runoff model with the reliability that is required in real world applications. However, we believe that regional information might be in any case useful to constrain the value of model parameters in ungauged basins.

The approach proposed here may be of help while calibrating a model, in view of the idea that the integration of different information is the way forward to reduce the feasible space for the model parameters. We believe that integrating different types of soft information, for instance by following the approach recently

proposed by Winsemius et al. (in press), and analysing different modelling competencies and skills is an important step for process understanding and hydrological modelling, both in gauged and in ungauged conditions.

Acknowledgements

The authors are grateful to Jasper Vrugt for providing the software for the application of AMALGAM. The useful comments of two anonymous referees are also gratefully acknowledged. The work presented here has been carried out in the framework of the activity of the Working Group at the University of Bologna of the Prediction in Ungauged Basins (PUB) initiative of the International Association of Hydrological Sciences. The study has been partially supported by the Italian Government through its national grants to the programmes on “Relations between hydrological processes, climate, and physical attributes of the landscape at the regional and basin scales” and “Advanced techniques for estimating the magnitude and forecasting extreme hydrological events, with uncertainty analysis”.

References

- Andersen, J., Dybkjaer, G., Jensen, K.H., Refsgaard, J.C., Rasmussen, K., 2002. Use of remotely sensed precipitation and leaf area index in a distributed hydrological model. *Journal of Hydrology* 264, 34–50.
- Atkinson, S.E., Woods, R.A., Sivapalan, M., 2002. Climate and landscape controls on water balance model complexity over changing timescales. *Water Resources Research* 38 (12), 1314.
- Bárdossy, A., 2007. Calibration of hydrological model parameters for ungauged catchments. *Hydrology and Earth System Sciences* 11, 703–710.
- Blazkova, S., Beven, K., 2009. A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resources Research* 45, W00B16.
- Boyle, D.P., 2000. Multicriteria Calibration of Hydrological Models, Ph.D. Dissertation. Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tempe.
- Brath, A., Castellarin, A., Montanari, A., 2002. Assessing the effects of land-use changes on annual average gross erosion. *Hydrology and Earth System Sciences* 6, 255–265.
- Brath, A., Castellarin, A., Montanari, A., 2003. Assessing the reliability of regional depth-duration-frequency equations for gaged and ungauged sites. *Water Resources Research* 39. doi:10.1029/2003WR002399.
- Butts, M.B., Payne, J.T., Kristensen, M., Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology* 298 (1–4), 242–266.

- Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., Brath, A., 2004. Regional flow-duration curves: reliability for ungauged basins. *Advances in Water Resources* 27 (10), 953–965.
- Castellarin, A., Camorani, G., Brath, A., 2007. Predicting annual and long-term flow-duration curves in ungauged basins. *Advances in Water Resources* 30 (4), 937–953.
- Castiglioni, S., Castellarin, A., Montanari, A., 2009. Prediction of low-flow indices in ungauged basins through physiographical space-based interpolation. *Journal of Hydrology* 378, 272–280.
- Castiglioni, S., Lombardi, L., Toth, E., Castellarin, A., Montanari, A., 2010. Calibration of rainfall-runoff models in ungauged basins: a regional maximum likelihood approach. *Advances in Water Resources* 33, 1235–1242. doi:10.1016/j.advwatres.2010.04.009.
- Castiglioni, S., Castellarin, A., Montanari, A., Skøien, J.O., Laaha, G., Blöschl, G., 2011. Smooth regional estimation of low-flow indices: physiographical space based interpolation and top-kriging. *Hydrology and Earth System Sciences* 15, 715–727.
- Chokmani, K., Ouada, T.B.M.J., 2004. Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. *Water Resources Research* 40, W12514. doi:10.1029/2003WR002983.
- Farmer, D., Sivapalan, M., Jothityangkoon, C., 2003. Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: downward approach to water balance analysis. *Water Resources Research* 39 (2), 1035.
- Flood Estimation Handbook (FEH), 1999. Institute of Hydrology, Wallingford, UK.
- Furey, P.R., Gupta, V.A., 2000. Space-time variability of low streamflows in river networks. *Water Resources Research* 36 (9), 2679–2690.
- Giandotti, M., 1937. *Idrologia*, Florence (IT): Barbera Edizioni.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* 377, 80–91.
- Houghton-Carr, H.A., 1999. Assessment criteria for simple conceptual daily rainfall-runoff models. *Hydrological Sciences Journal* 44, 237–261.
- Hosking, J.R.M., Wallis, J.R., 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge, UK.
- Hundecha, Y., Ouada, T.B.M.J., Bárdossy, A., 2008. Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the spatial structures of the parameters within a canonical physiographic-climatic space. *Water Resources Research* 44, W01427. doi:10.1029/2006WR005439.
- Krause, P., Boyle, D.P., Båse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* 5, 8997.
- Laaha, G., Blöschl, G., 2006. Seasonality indices for regionalizing low flows. *Hydrological Processes* 20, 3851–3878.
- Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research* 35 (1), 233–241.
- Montanari, A., Toth, E., 2007. Calibration of hydrological models in the spectral domain: an opportunity for scarcely gauged basins? *Water Resources Research* 43, W05434. doi:10.1029/2006WR005184.
- Montanari, A., 2007. What do we mean by uncertainty? The need for a consistent wording about uncertainty assessment in hydrology. *Hydrological Processes* 21, 841–845.
- Pandey, G.R., Nguyen, V.T.V., 1999. A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology* 225 (1–2), 92–101.
- Parajka, J., Merz, R., Blöschl, G., 2005. A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences* 9, 157–171.
- Pickup, G., 1977. Potential and limitations of rainfall-runoff models for prediction on ungauged catchments: a case study from the Papua New Guinea Highlands. *Journal of Hydrology (NZ)* 16 (1), 87–102.
- Pokhrel, P., Gupta, H.V., Wagener, T., 2008. A spatial regularization approach to parameter estimation for a distributed watershed model. *Water Resources Research* 44, W12419. doi:10.1029/2007WR006615.
- Refsgaard, J.C., Knudsen, J., 1996. Operational validation and intercomparison of different types of hydrological models. *Water Resources Research* 32, 2189–2202.
- Schaeffli, B., Zehe, E., 2009. Hydrological model performance and parameter estimation in the wavelet-domain. *Hydrology and Earth System Sciences* 13, 1921–1936.
- Smakhtin, V.U., 2001. Low flow in hydrology: a review. *Journal of Hydrology* 240, 147–186.
- Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1993. Frequency analysis of extreme events. In: Maidment, D.R. (Ed.), *Handbook of Hydrology*. McGraw-Hill Inc., New York, NY, USA, Chapter 18.
- Thornthwaite, C.W., 1946. An approach toward a rational classification of climate. *Transactions, American Geophysical Union* 27 (1).
- Vogel, R.M., Kroll, C.N., 1992. Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics. *Water Resources Research* 28 (9), 2451–2458.
- Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., Shuffled, A., 2003. Complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* 39, 1201. doi:10.1029/2002WR001642.
- Vrugt, J.A., Robinson, B.A., 2007. Improved evolutionary optimization from genetically adaptive multimethod search. *Proceedings of the National Academy of Sciences of the United States of America* 104, 708–711.
- Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V., Sorooshian, S., 2001. A framework for development and application of hydrological models. *Hydrology and Earth System Sciences* 5, 13–26.
- Wagener, T., Wheater, H.S., 2006. Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. *Journal of Hydrology* 320, 132–154.
- Weisberg, S., 1985. *Applied Linear Regression*, second ed. John Wiley, New York.
- Westerberg, I.K., Guerrero, J.L., Younger, P.M., Beven, K.J., Seibert, J., Halldin, S., Freer, J., Xu, C.Y., 2010. Calibration of hydrological models using flow-duration curves. *Hydrology and Earth System Sciences Discussions* 7, 9467–9522.
- Willmott, C.J., 1981. On the validation of models. *Physical Geography* 2, 184–194.
- Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., ODonnell, J., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research* 90, 8995–9005.
- Winsemius, H.C., Schaeffli, B., Montanari, A., Savenije, H.H.G., in press. On the calibration of hydrological models in ungauged basins: a framework for integrating hard and soft hydrological information. *Water Resources Research*, doi:10.1029/2009WR007706.
- Yadav, M., Wagener, T., Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources* 30 (8), 1756–1774.
- Yanmaz, K.K., Gupta, H.V., Wagener, T., 2008. A process-based diagnostic approach to model evaluation: application to the NWS distributed hydrologic model. *Water Resources Research* 44, W09417.
- Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. *Journal of Hydrology* 204, 83–97.
- Yu, P.S., Yang, T.C., 2000. Using synthetic flow duration curves for rainfall-runoff model calibration at ungauged sites. *Hydrological Processes* 14, 117–133.
- Zhang, Z., Wagener, T., Reed, P., Bhushan, R., 2008. Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. *Water Resources Research* 44, W00B04. doi:10.1029/2008WR006833.