Hydrol. Earth Syst. Sci. Discuss., 11, 8039–8066, 2014
www.hydrol-earth-syst-sci-discuss.net/11/8039/2014/
doi:10.5194/hessd-11-8039-2014

Hydrology and Earth System Sciences
Open Access
Discussions

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

Transferring model uncertainty estimates from gauged to ungauged catchments

F. Bourgin[1], **V. Andréassian**[1], **C. Perrin**[1], **and L. Oudin**[2]

[1]Irstea, UR HBAN, 1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony Cedex, France
[2]UPMC Univ Paris 06, UMR 7619 Metis, Case 105, 4 place Jussieu, 75005 Paris, France

Received: 18 June 2014 – Accepted: 29 June 2014 – Published: 14 July 2014

Correspondence to: F. Bourgin (francois.bourgin@irstea.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Predicting streamflow hydrographs in ungauged catchments is a challenging issue, and accompanying the estimates with realistic uncertainty bounds is an even more complex task. In this paper, we present a method to transfer model uncertainty estimates from gauged to ungauged catchments and we test it over a set of 907 catchments located in France. We evaluate the quality of the uncertainty estimates based on three expected qualities: reliability, sharpness, and overall skill. Our results show that the method holds interesting perspectives, providing in most cases reliable and sharp uncertainty bounds at ungauged locations.

1 Introduction

1.1 Predicting streamflow in ungauged catchments with uncertainty estimates

Predicting the entire runoff hydrograph in ungauged catchments is a challenging issue that has attracted much attention during the last decade. In this context, the use of suitable conceptual rainfall–runoff models has proved to be useful, and because traditional calibration approaches based on observed discharge data cannot be applied in ungauged catchments, other approaches are required. Various methods have been proposed for the estimation of rainfall–runoff model parameters in ungauged catchments, as reported by the recent synthesis of the Prediction in Ungauged Basins (PUB) decade (Blöschl et al., 2013; Hrachowitz et al., 2013; Parajka et al., 2013).

The estimation of predictive uncertainty is deemed good practice in any environmental modelling activity (Refsgaard et al., 2007). In hydrological modelling, the topic has been widely discussed for years, and there is still no general agreement about how to adequately quantify uncertainty. In practice, various methodologies are currently used.

For gauged catchments, the methodologies include Bayesian calibration and prediction approaches (see e.g., the review of Liu and Gupta, 2007), informal methods related

HESSD
11, 8039–8066, 2014

to the GLUE framework (Beven and Freer, 2001), multi-model approaches (Duan et al., 2007; Velazquez et al., 2010) and other total uncertainty quantification methods (Montanari and Brath, 2004; Solomatine and Shrestha, 2009; Weerts et al., 2011; Ewen and O'Donnell, 2012). A comprehensive review of the topic can be found in Matott et al. (2009) and Montanari (2011).

While many methods have been proposed for gauged catchments, only a few have been proposed for the estimation of predictive uncertainty on ungauged catchments. McIntyre et al. (2005) presented a GLUE-type approach consisting of transferring ensembles of parameter sets obtained on donor (gauged) catchments to target (ungauged) catchments. More recently, a framework based on constrained parameter sets was applied in several studies (Yadav et al., 2007; Zhang et al., 2008; Winsemius et al., 2009; Bulygina et al., 2011, 2012; Kapangaziwiri et al., 2012). It is a two-step procedure. The first step consists in estimating with uncertainty various summary metrics of the hydrographs, also called "signatures" of the catchments, or gathering other "soft" or "hard" information at the target ungauged catchment. The second step is the selection of an ensemble of model parameter sets. "Acceptable" or "behavioural" parameter sets are those that yield sufficiently close simulated summary metrics compared to regionalized metrics. The reader can refer to Wagener and Montanari (2011) for a comprehensive description of this framework.

One difficulty of the above mentioned approaches lies in the interpretation of the uncertainty bounds obtained from the parameter ensemble predictions. As noted by McIntyre et al. (2005) and Winsemius et al. (2009), the uncertainty bounds cannot easily be interpreted as confidence intervals, and thus it remains difficult to use them in practice. In addition, solely relying on an ensemble of model parameter sets to quantify total predictive uncertainty is often not sufficient when imperfect rainfall–runoff models are used.

A pragmatic alternative consists in addressing separately the parameter estimation and the uncertainty estimation issues. It has been argued by several authors (Montanari and Brath, 2004; Andréassian et al., 2007; Ewen and O'Donnell, 2012) that a posteriori characterization of modelling errors of a "best" or "optimal" simulation can yield valid uncertainty bounds at gauged locations. As stated by Solomatine and Shrestha (2009),

> The historical model residuals (errors) between the model prediction $\hat{y}$ and the observed data y are the best available quantitative indicators of the discrepancy between the model and the real-world system or process, and they provide valuable information that can be used to assess the predictive uncertainty.

Similarly, one could argue that the model residuals between the model prediction and the observed data at *neighbouring gauged locations* are, perhaps, the best available indicators of the discrepancy between the model and the real-world system at *the target ungauged location*.

The only attempt we are aware of to apply a total uncertainty estimation approach at ungauged location is the one presented by Roscoe et al. (2012). They quantified uncertainty for river stage prediction at ungauged locations by first interpolating the residual errors at ungauged locations, and then applying quantile regression to these errors.

1.2 Scope of the paper

The aim of this paper is to provide an estimation of the total uncertainty affecting runoff prediction at ungauged locations when a rainfall–runoff model and a regionalisation scheme is used.

To our knowledge, a framework based on residual errors and total uncertainty quantification has not yet been extensively tested in the context of prediction in ungauged catchments. This paper contributes to the search for methods able to provide uncertainty estimates when runoff predictions in ungauged catchments are sought.

2 Data and methods

Our objective is not to develop a new parameter regionalisation approach. Therefore, we purposely chose to use ready-to-use materials and methods and only focus on the uncertainty quantification issue. This study can be considered as a follow-up of the work made by Oudin et al. (2008) on the comparison of regionalisation approaches. We only provide here an overview of the data set, the rainfall–runoff models and the parameter calibration and regionalisation approach, since the details can be found in Oudin et al. (2008).

2.1 Data set

A database of 907 French catchments was used. They represent various hydrological conditions, given the variability in climate, topography, and geology in France. This set includes fast responding Mediterranean catchments with intense precipitation as well as larger, groundwater-dominated catchments. Some characteristics of the data set are given in Table 1. Catchments were selected to have limited snow influence, since no snowmelt module was used in the hydrological modelling. Daily rainfall, runoff, and potential evapotranspiration (PE) data series over the 1995–2005 period were available. Meteorological inputs originate from Météo-France SAFRAN reanalysis (Vidal et al., 2010). PE was estimated using the temperature-based formula proposed by Oudin et al. (2005). Hydrological data were extracted from the HYDRO national archive (www.hydro.eaufrance.fr).

2.2 Rainfall–runoff models

Two daily, continuous lumped rainfall–runoff models were used:

- The GR4J rainfall–runoff model, an efficient and parsimonious daily lumped continuous rainfall–runoff model described by Perrin et al. (2003).

- The TOPMO rainfall–runoff model, inspired by TOPMODEL (Beven and Kirkby, 1979). This version was tested on large data sets and showed performance comparable to that of the GR4J model, while being quite different (Michel et al., 2003; Oudin et al., 2008, 2010).

Using these two models rather than a single one makes it possible to draw more general conclusions.

The GR4J and TOPMO models have four and six free parameters respectively. On gauged catchments, parameter estimation is performed using a local gradient search procedure, applied in combination with a pre-screening of the parameter space (Mathevet, 2005; Perrin et al., 2008). This optimization procedure has proved to be efficient in past applications for the conceptual models used here. As objective function, we used the Nash and Sutcliffe (1970) criterion computed on root square transformed flows. This criterion was shown to yield a good compromise between different objectives (Oudin et al., 2006).

2.3 Regionalisation approach

By definition, no discharge data is available for calibrating parameter sets at ungauged location. Thus, other strategies are needed to estimate the parameters of hydrological models for prediction in ungauged catchments.

Oudin et al. (2008) assessed the relative performance of three classical regionalisation schemes over a set of French catchments: spatial proximity, physical similarity and regression. Several options within each regionalisation approach were tested and compared. Based on their results, the following choices were made here for the regionalisation approach, as they offered the best regionalisation solution:

- Poorly modelled catchments were discarded as potential donors: only catchments with a performance criterion in calibration above 0.7 were used as possible donors.

– The spatial proximity approach was used. It consists of transferring parameter sets from neighbouring catchments to the target ungauged catchment. Proximity of the ungauged catchments to the gauged ones was quantified by the distances between catchments centroids.

– The output averaging option was chosen. It consists of computing the mean of the streamflow simulations obtained on the ungauged catchment with the set of parameters of the donor catchments.

– The number of neighbours was set to 4 and 7 catchments for GR4J and TOPMO respectively.

3 Proposed approach: transfer of relative errors by flow groups

Transferring calibrated model parameters from gauged catchments to ungauged catchment is a well established approach when parameters cannot be inferred from available data. The method presented here extends the parameter transfer approach to the domain of uncertainty estimation.

The main idea underlying the proposed approach is (i) to treat each donor as if it was ungauged (simulating flow though the above described regionalisation approach), (ii) characterize the empirical distribution of relative errors for each of these donors, and (iii) transfer model uncertainty estimates to the ungauged catchment.

The methodology used to transfer model uncertainty estimates can be described by the following steps, illustrated by Figs. 1 to 5:

1. Selection of catchments
Here we consider a target catchment as ungauged, called TUC. This catchment has n neighbouring gauged catchments, called NGC_1, $\mathrm{NGC}_2,\ldots,\mathrm{NGC}_n$. If the NGC_i catchment was now considered ungauged, one could also consider n neighbouring catchments, called NGC_i^1, $\mathrm{NGC}_i^2,\ldots,\mathrm{NGC}_i^n$. Obviously, the TUC catchment would be excluded from this set of second order donor catchments.

HESSD
11, 8039–8066, 2014

2. Application of the parameter regionalisation scheme to the donor catchments NGC_i

   a. Apply the parameter regionalisation scheme to obtain a simulated discharge time series for each NGC_i using neighbours NGC_i^j.

   b. Compute the relative errors of streamflow reconstitution, and create 10 groups of relative errors according to the magnitude of the simulated discharge. The groups are based on the quantiles of the simulated discharges, so that each group is equally populated. The subdivision into flow groups allows accounting for the heteroscedasticity of model errors.

3. Computation of the multiplicative coefficients

   a. Put together the relative errors from the donors according to the group they belong to.

   b. Compute the empirical quantiles of the relative errors distribution within each group. Each quantile of relative error can be considered a multiplicative coefficient. These multiplicative coefficients will be used to obtain the prediction bounds.

4. Computation of the uncertainty bounds for the target catchment TUC

   a. Apply the parameter regionalisation scheme to obtain a simulated discharge time series for TUC using the parameter sets of the neighbouring catchments NGC_i.

   b. Multiply the simulated discharge by the set of multiplicative coefficients obtained at Step 3b to obtain the uncertainty bounds.

Some of the methodological choices made here will be further discussed in Sect. 5.

HESSD
11, 8039–8066, 2014

4 Quantitative evaluation of uncertainty bounds

We assessed the relevance of the 90 % uncertainty bounds by focusing on three characteristics: reliability, sharpness and overall skill.

Reliability refers to the statistical consistency of the uncertainty estimation with the observation, i.e., a 90 % prediction interval is expected to contain approximately 90 % of the observations if prediction errors are adequately characterized by the uncertainty estimation. To estimate the reliability, we used the coverage ratio (CR) index, computed as the percentage of observations contained in the prediction intervals.

Sharpness refers to the concentration of predictive uncertainty. We used a quantitative index based on the average width of the uncertainty bounds. To ease comparison between catchments, we used the width of the 90 % intervals of historical flows [Q5, Q95], where Q5 and Q95 are the 5th and 95th percentiles of the flow duration curve, as a benchmark. The ratio (R) between these two values provides information about the reduction of uncertainty obtained by the application of the rainfall–runoff and the methodology presented here, compared to the climatology. The value $1 - R$ indicates the percentage of reduction of the average width. We call this criterion the average width index (AWI). It is positive if the average width is reduced, and negative otherwise.

Uncertainty bounds that are as sharp as possible and reasonably reliable are sought: indeed sharp intervals that would consistently miss the target would be misleading, while overly large intervals that would successfully cover the observations at the expense of sharpness would be of limited value for decision making.

To complete the assessment of the prediction bounds, we used the interval score (Gneiting and Raftery, 2007). The interval score (IS) accounts for both reliability and sharpness and provides an overall assessment of the quality of the prediction bounds. The scoring rule of the interval score is defined as

$$S = (u - l) + \frac{2}{1-\beta}(l - q)\mathbf{1}\{q < l\} + \frac{2}{1-\beta}(q - u)\mathbf{1}\{q > u\} \quad (1)$$

HESSD
11, 8039–8066, 2014

where $[l, u]$ is the prediction interval and q is the observed value; $\mathbf{1}\{x < y\}$ is the indicator function, equal to 1 if $x < y$ and 0 otherwise, and β is equal to 90 % since a 90 % interval is sought here. IS is the average value of S over the time series.

To ease comparison between catchments and evaluate the skill of the prediction bounds, we used the unconditional climatology as a benchmark and computed the interval skill score

$$\mathrm{ISS} = 1 - \frac{\mathrm{ISS}}{\mathrm{ISS}^{\mathrm{clim}}} \quad (2)$$

where $\mathrm{ISS}^{\mathrm{clim}}$ is the interval score obtained with the 90 % climatological interval [Q5, Q95] (Q5 and Q95 are the 5th and 95th percentiles of the flow duration curve).

The skill score is positive when the prediction bounds are more skillfull than the climatological interval, and negative otherwise.

5 Results and discussion

5.1 Reliability, sharpness and overall skill

Figure 6 shows the distributions of the three criteria used to evaluate the uncertainty bounds on the 907 catchments. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) are used to synthesize the variety of scores over the 907 catchments of the data set.

5.1.1 Reliability

For both models, half of the catchments (from the lower quartile to the upper quartile) have CR values between 80 and 95 %. The median values are equal to 89 and 90 % for GR4J and TOPMO respectively. Since a value of 90 % is expected for 90 % prediction bounds, these results suggest that the prediction bounds are, in most cases, able to reflect the magnitude of errors when predicting runoff hydrographs in ungauged catchments. The CR values fall below 0.7 for around 14 % of the catchments with GR4J, and 13 % with TOPMO, which indicates cases where the proposed approach yields predictive bounds that might be too narrow or biased (i.e., not well centered on the observations).

5.1.2 Sharpness

Regarding sharpness, it can be seen that for GR4J, half of the catchments (from the lower quartile to the upper quartile) have AWI values between 39 and 67 %, while for TOPMO corresponding values are equal to 38 and 65 %. The median values are equal to 57 and 55 % for GR4J and TOPMO respectively. The higher the AWI values, the lower the predictive uncertainty is. Since it would be utopic to expect that no errors will be made when predicting runoff hydrographs for ungauged catchments, we considered here uncertainty reduction values between 30 and 80 % as quite satisfactory. Note that negative values are seen for 7 % of the catchments with both GR4J and TOPMO, which indicates cases where the approach yield prediction intervals whose average width is larger than the width of the historical [Q5, Q95] interval (Q5 and Q95 are the 5th and 95th percentiles of the flow duration curve).

5.1.3 Overall skill

Finally, Fig. 6c shows that the predictive skill is positive for most catchments (around 92 %) for both models. For both models, half of the catchments (from the lower quartile to the upper quartile) have ISS values between 40 and 70 %. The median values are equal to 61 and 59 % for GR4J and TOPMO respectively. While it might be argued that the unconditional climatology is not a very challenging benchmark, we consider that it is still a positive and reassuring result.

5.2 Do we need to treat the donor catchments as ungauged?

A critical step of the proposed approach is to apply the regionalisation scheme to obtain a simulated discharge time series for each donor catchment (Step 2a). This is done

HESSD
11, 8039–8066, 2014

because we expect that predictive uncertainty at ungauged locations is larger than predictive uncertainty at gauged location, i.e., when the rainfall–runoff is calibrated with observed discharge data. To assess the impact of this methodological choice, we applied the methodology described earlier to transfer uncertainty estimates, but this time the donor catchments are treated as gauged.

Similarly to Fig. 6, Fig. 7 shows the distributions of the three criteria obtained in the two cases: whether or not the donor catchments are treated as ungauged. We can observe for both models a drop in reliability, whereas sharpness increases. This is because the relative errors are smaller when the donor catchments are treated as gauged, yielding narrower but less reliable prediction bounds for the target catchment. Interestingly, this results in skill scores that are quite similar: improvements in terms of sharpness compensate decreases in terms of reliability.

Note that reliability is generally considered as a prevailing characteristic over sharpness. Therefore, the benefit of treating the donor catchments as ungauged clearly appears in Fig. 7a.

5.3 Do we need to use groups of relative errors?

Another critical step of the proposed approach is to use 10 groups of relative errors. The groups are defined according to the magnitude of the simulated discharge (Step 2b). This was done to take into account the fact that the characteristics of errors usually change according to the magnitude of the simulated discharge. To assess the impact of this methodological choice, we again applied the methodology described earlier to transfer model uncertainty estimates, but this time only one group is used instead of 10.

Figure 8 shows the distributions of the three criteria obtained in the following two cases: whether 10 groups or only one group of relative errors are used. For both models, reliability slightly increase, whereas both sharpness and skill decrease. It appears that improvements in terms of reliability are not sufficient to compensate decreases in terms of sharpness when overall skill is considered.

HESSD
11, 8039–8066, 2014

While it could be argued that using only one group is the preferable option because of the slight improvement in terms of reliability, in our opinion, the improvement is not sufficiently important to compensate the decrease in terms of uncertainty reduction and skill. We definitely prefer to maintain different flow groups.

5.4 How do the performances of the rainfall–runoff models relate to the characteristics of uncertainty bounds?

To gain insights into the possible relationships between the performance of the deterministic rainfall–runoff simulations and characteristics of the uncertainty bounds at ungauged locations, the three criteria used to characterize the uncertainty bounds are plotted in Fig. 9 as function of a quadratic efficiency criteria for the 907 catchments. The quadratic efficiency criterion is the C2M (Mathevet et al., 2006), a bounded version of the Nash and Sutcliffe (1970) efficiency (NSE) criterion. The equations are:

$$\mathrm{C2M} = \frac{\mathrm{NSE}}{2 - \mathrm{NSE}} \tag{3}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{n} \left(Q_{\mathrm{o},t} - Q_{\mathrm{s},t}\right)^2}{\sum_{t=1}^{n} \left(Q_{\mathrm{o},t} - \mu_{\mathrm{o}}\right)^2} \tag{4}$$

where n is the total number of time-steps, $Q_{\mathrm{s},t}$ is the simulated discharge at time-step t, $Q_{\mathrm{o},t}$ is the observed discharge at time-step t, and μ_{o} is the mean of the observed discharges. The advantage of this bounded version is avoiding large negative values difficult to plot.

A trend appears between deterministic performance and characteristics of the prediction bounds at ungauged locations, for the two rainfall–runoff models. The reliability index exhibits larger variability compared to the sharpness index, and the stronger link is seen for the skill score. Reliability is relatively less affected by the poor deterministic performance of the simulation at ungauged location because there are cases where poor performance at neighbour locations leads (though the transfer of relative errors) to wide prediction bounds that are able to cover the observed values. We can also observe that skill scores and C2M scores are strongly related, which indicates that when the transfer of model parameters performs well, the transfer of model uncertainty estimates performs well too.

6 Conclusions

Runoff hydrograph prediction in ungauged catchments is notoriously difficult, and attempting to estimate the predictive uncertainty in that context is a further challenge. We proposed a method allowing the transfer of model uncertainty estimates from gauged to ungauged catchments. The method extends the parameter transfer approach to the domain of uncertainty estimation.

We evaluated the approach over a large set of 907 catchments by assessing three expected qualities of uncertainty estimates, reliability, sharpness and overall skill. Our results demonstrate that the method is generally able to reflect model errors at ungauged locations and provide reasonable reliability. We applied two different rainfall-runoff models (GR4J and TOPMO) to ensure that the presented results are not model-specific.

Although we used a transfer based on spatial proximity, the approach is independent of the regionalisation scheme used to obtain deterministic prediction at ungauged locations, and any other similarity measure could be a basis for transferring residual errors.

Last, we would like to stress that the results presented in this study are expressed in terms of dimensionless measures. As such, they can provide a basis for future comparisons when prediction in ungauged catchments with uncertainty estimates is performed.

Acknowledgements. The authors thank Météo-France for providing the meteorological data and Banque HYDRO for the hydrological data. The financial support of SCHAPI to the first author is also gratefully acknowledged.

HESSD
11, 8039–8066, 2014

References

Andréassian, V., Lerat, J., Loumagne, C., Mathevet, T., Michel, C., Oudin, L., and Perrin, C.: What is really undermining hydrologic science today?, Hydrol. Process., 21, 2819–2822, doi:10.1002/hyp.6854, 2007. 8041

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol., 249, 11–29, 2001. 8041

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrol. Sci. Bull., 24, 43–69, doi:10.1080/02626667909491834, 1979. 8044

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H.: Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales, Cambridge University Press, 2013. 8040

Bulygina, N., McIntyre, N., and Wheater, H.: Bayesian conditioning of a rainfall–runoff model for predicting flows in ungauged catchments and under land use changes, Water Resour. Res., 47, W02503, doi:10.1029/2010wr009240, 2011. 8041

Bulygina, N., Ballard, C., McIntyre, N., O'Donnell, G., and Wheater, H.: Integrating different types of information into hydrological model parameter estimation: application to ungauged catchments and land use scenario analysis, Water Resour. Res., 48, W06519, doi:10.1029/2011wr011207, 2012. 8041

Duan, Q., Ajami, N. K., Gao, X., and Sorooshian, S.: Multi-model ensemble hydrologic prediction using Bayesian model averaging, Adv. Water Resour., 30, 1371–1386, doi:10.1016/j.advwatres.2006.11.014, 2007. 8041

Ewen, J. and O'Donnell, G.: Prediction intervals for rainfall–runoff models: raw error method and split-sample validation, Hydrol. Res., 43, 637–648, doi:10.2166/nh.2012.038, 2012. 8041

Gneiting, T. and Raftery, A. E.: Strictly proper scoring rules, prediction, and estimation, J. Am. Stat. Assoc., 102, 359–378, doi:10.1198/016214506000001437, 2007. 8047

Hrachowitz, M., Savenije, H. H. G., Bloeschl, 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., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)a review, Hydrolog. Sci. J., 58, 1198–1255, doi:10.1080/02626667.2013.803183, 2013. 8040

Kapangaziwiri, E., Hughes, D. A., and Wagener, T.: Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa, Hydrolog. Sci. J., 57, 1000–1019, doi:10.1080/02626667.2012.690881, 2012. 8041

Liu, Y. Q. and Gupta, H. V.: Uncertainty in hydrologic modeling: toward an integrated data assimilation framework, Water Resour. Res., 43, W07401, doi:10.1029/2006wr005756, 2007. 8040

Mathevet, T.: Quels modèles pluie-débit globaux au pas de temps horaire ? Développements empiriques et comparaison de modèles sur un large échantillon de bassins versants, Ph. D. thesis, Paris, 2005. 8044

Mathevet, T., Michel, C., Andréassian, V., and Perrin, C.: A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins, IAHS-AISH Publication, 211–219, Wallingford, 2006. 8051

Matott, L. S., Babendreier, J. E., and Purucker, S. T.: Evaluating uncertainty in integrated environmental models: a review of concepts and tools, Water Resour. Res., 45, W06421, doi:10.1029/2008wr007301, 2009. 8041

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, Water Resour. Res., 41, W12434, doi:10.1029/2005wr004289, 2005. 8041

Michel, C., Perrin, C., and Andréassian, V.: The exponential store: a correct formulation for rainfall–runoff modelling, Hydrolog. Sci. J., 48, 109–124, doi:10.1623/hysj.48.1.109.43484, 2003. 8044

Montanari, A.: Uncertainty of hydrological predictions, in: Treatise on Water Science, edited by: Peter, W., Elsevier, Oxford, 459–478, 2011. 8041

Montanari, A. and Brath, A.: A stochastic approach for assessing the uncertainty of rainfall–runoff simulations, Water Resour. Res., 40, W01106, doi:10.1029/2003wr002540, 2004. 8041

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models, Part I – A discussion of principles, J. Hydrol., 10, 282–290, 1970. 8044, 8051

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2 – Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, J. Hydrol., 303, 290–306, doi:10.1016/j.jhydrol.2004.08.026, 2005. 8043

HESSD

11, 8039–8066, 2014

Transferring model uncertainty estimates from gauged to ungauged catchments

F. Bourgin et al.

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., and Michel, C.: Dynamic averaging of rainfall–runoff model simulations from complementary model parameterizations, Water Resour. Res., 42, W07410, doi:10.1029/2005wr004636, 2006. 8044

Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged catchments: a comparison of regionalization approaches based on 913 French catchments, Water Resour. Res., 44, W03413, doi:10.1029/2007wr006240, 2008. 8043, 8044

Oudin, L., Kay, A., Andréassian, V., and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, Water Resour. Res., 46, W11558, doi:10.1029/2009wr008887, 2010. 8044

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies, Hydrol. Earth Syst. Sci., 17, 1783–1795, doi:10.5194/hess-17-1783-2013, 2013. 8040

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, J. Hydrol., 279, 275–289, doi:10.1016/s0022-1694(03)00225-7, 2003. 8043

Perrin, C., Andréassian, V., Serna, C. R., Mathevet, T., and Le Moine, N.: Discrete parameterization of hydrological models: evaluating the use of parameter sets libraries over 900 catchments, Water Resour. Res., 44, W08447, doi:10.1029/2007wr006579, 2008. 8044

Refsgaard, J. C., van der Sluijs, J. P., Hojberg, A. L., and Vanrolleghem, P. A.: Uncertainty in the environmental modelling process – a framework and guidance, Environ. Modell. Softw., 22, 1543–1556, doi:10.1016/j.envsoft.2007.02.004, 2007. 8040

Roscoe, K. L., Weerts, A. H., and Schroevers, M.: Estimation of the uncertainty in water level forecasts at ungauged river locations using quantile regression, Int. J. River Basin Manage., 10, 383–394, doi:10.1080/15715124.2012.740483, 2012. 8042

Solomatine, D. P. and Shrestha, D. L.: A novel method to estimate model uncertainty using machine learning techniques, Water Resour. Res., 45, W00B11, doi:10.1029/2008wr006839, 2009. 8041, 8042

Velázquez, J. A., Anctil, F., and Perrin, C.: Performance and reliability of multimodel hydrological ensemble simulations based on seventeen lumped models and a thousand catchments, Hydrol. Earth Syst. Sci., 14, 2303–2317, doi:10.5194/hess-14-2303-2010, 2010. 8041

Vidal, J.-P., Martin, E., Franchisteguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, Int. J. Climatol., 30, 1627–1644, doi:10.1002/joc.2003, 2010. 8043

HESSD

11, 8039–8066, 2014

Transferring model uncertainty estimates from gauged to ungauged catchments

F. Bourgin et al.

Wagener, T. and Montanari, A.: Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, Water Resour. Res., 47, W06301, doi:10.1029/2010wr009469, 2011. 8041

Weerts, A. H., Winsemius, H. C., and Verkade, J. S.: Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales), Hydrol. Earth Syst. Sci., 15, 255–265, doi:10.5194/hess-15-255-2011, 2011. 8041

Winsemius, H. C., Schaefli, B., Montanari, A., and Savenije, H. H. G.: On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, Water Resour. Res., 45, W12422, doi:10.1029/2009wr007706, 2009. 8041

Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Adv. Water Resour., 30, 1756–1774, doi:10.1016/j.advwatres.2007.01.005, 2007. 8041

Zhang, Z., Wagener, T., Reed, P., and Bhushan, R.: Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization, Water Resour. Res., 44, W00B04, doi:10.1029/2008wr006833, 2008. 8041

HESSD
11, 8039–8066, 2014

Transferring model uncertainty estimates from gauged to ungauged catchments

F. Bourgin et al.

Table 1. Characteristics of the 907 catchments. P – precipitation, PE – potential evapotranspiration, Q – discharge.

	Percentiles				
	0.05	0.25	0.50	0.75	0.95
Catchment area (km^2)	27	73	149	356	1788
Mean annual precipitation ($mm\,yr^{-1}$)	753	853	978	1176	1665
Mean annual potential evapotranspiration ($mm\,yr^{-1}$)	549	631	659	700	772
Mean annual runoff ($mm\,yr^{-1}$)	133	233	344	526	1041
Q/P ratio	0.17	0.27	0.34	0.45	0.68
P/PE ratio	1.06	1.25	1.47	1.83	2.9
Median elevation (m)	76	149	314	645	1183

HESSD
11, 8039–8066, 2014

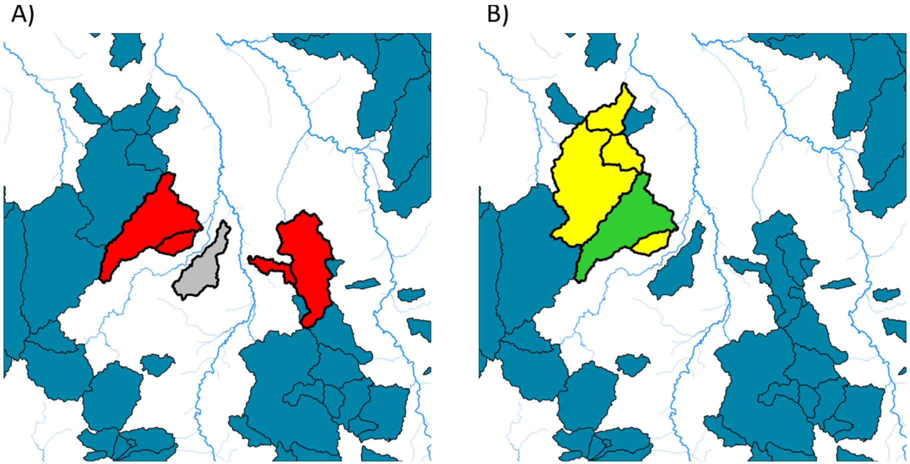

Figure 1. Illustration of the proposed approach – Step 1: in **(A)**, a target catchment (grey) is considered as ungauged; this catchment has *n* neighbouring gauged catchments (red). In **(B)**, if one of the neighbouring catchment is now considered ungauged (green), we also consider *n* neighbouring catchments (yellow). Note that the target catchment is excluded from this set of second order donor catchments.

HESSD
11, 8039–8066, 2014

Transferring model uncertainty estimates from gauged to ungauged catchments

F. Bourgin et al.

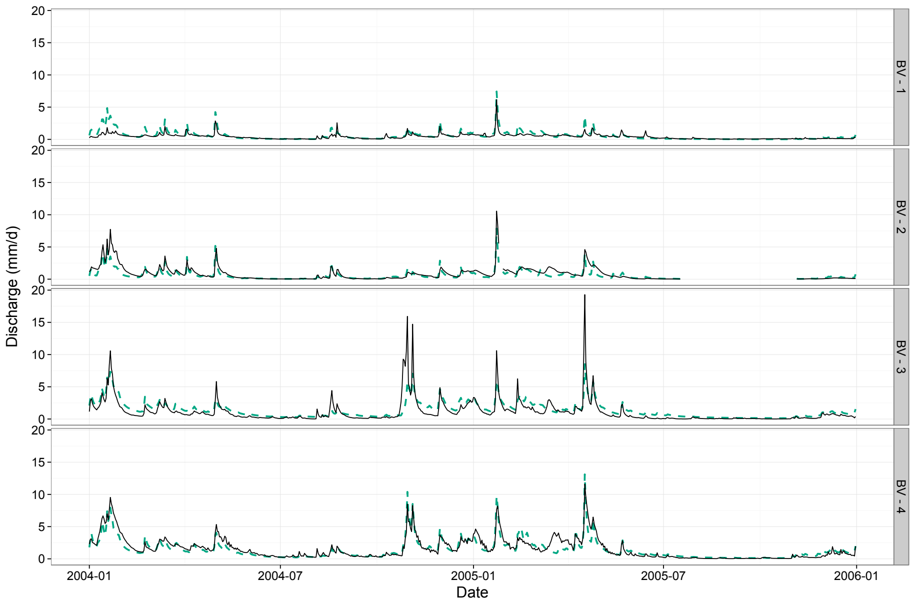

Figure 2. Illustration of the proposed approach – Step 2a: simulated (green, dashed) and observed (black) discharge time series for four donor catchments treated as ungauged, i.e., in which model parameters must be estimated from a regionalisation approach.

HESSD
11, 8039–8066, 2014

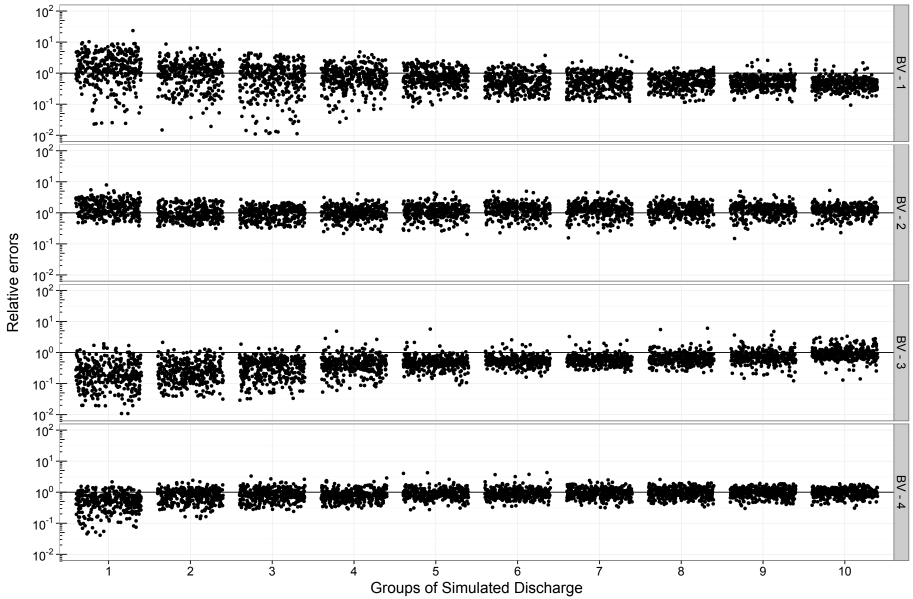

Figure 3. Illustration of the proposed approach – Step 2b: relative errors by flow groups; groups of relative errors are defined according to the magnitude of the simulated discharge.

HESSD
11, 8039–8066, 2014

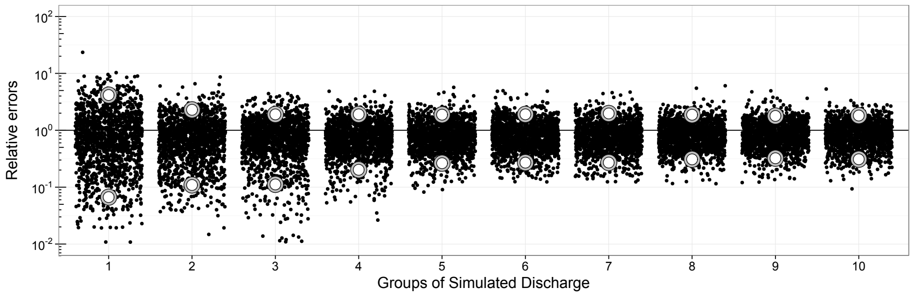

Figure 4. Illustration of the proposed approach – Steps 3a and b: aggregating the relative errors observed at the donors catchments; white dots correspond to the empirical quantiles (5 and 95 %) of the relative errors distribution within each group.

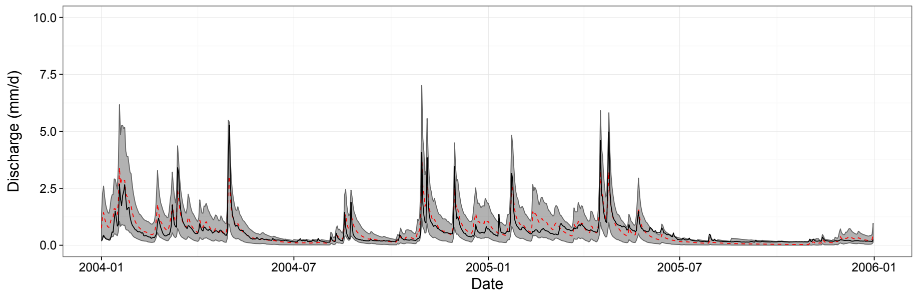

Figure 5. Illustration of the proposed approach – Steps 4a and b: simulated (red, dashed) and observed (black) discharge time series for the ungauged catchments; 90 % uncertainty bounds in grey.

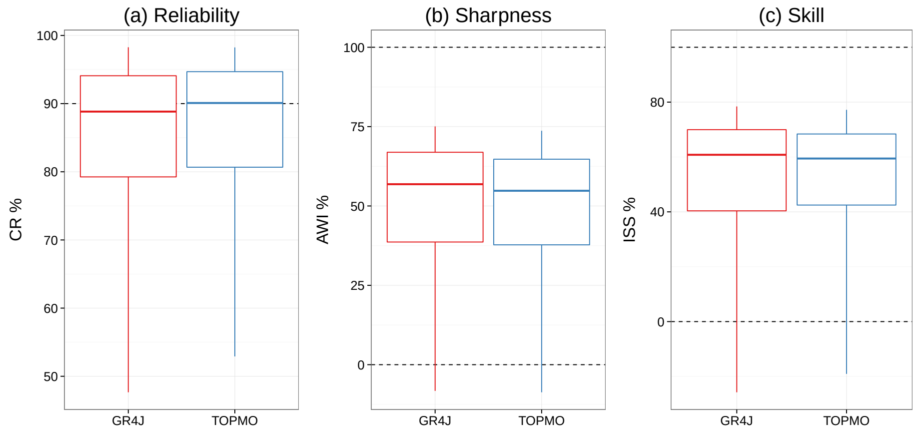

Figure 6. Distributions of the three performance criteria. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 907 catchments of the data set.

HESSD
11, 8039–8066, 2014

Transferring model uncertainty estimates from gauged to ungauged catchments

F. Bourgin et al.

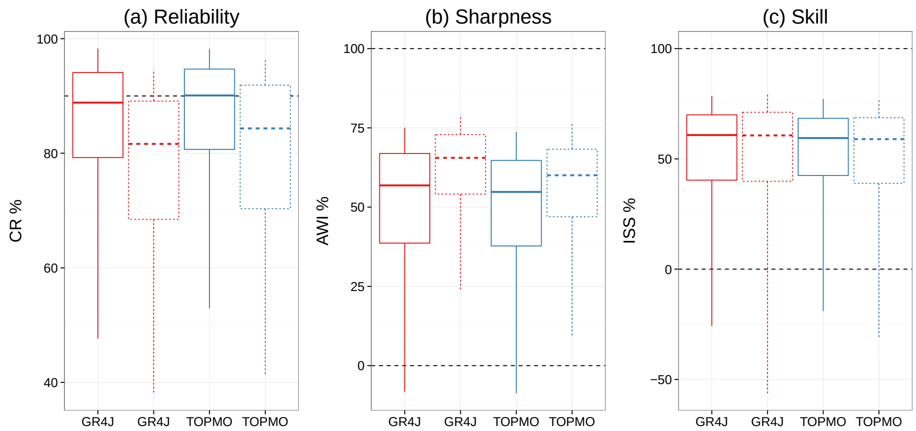

Figure 7. Distributions of the three performance criteria, obtained in two cases, (i) when the donor catchments are treated as ungauged (continous lines) and (ii) when the donor catchments are treated as gauged (dashed lines). Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 907 catchments of the data set.

HESSD
11, 8039–8066, 2014

Transferring model uncertainty estimates from gauged to ungauged catchments

F. Bourgin et al.

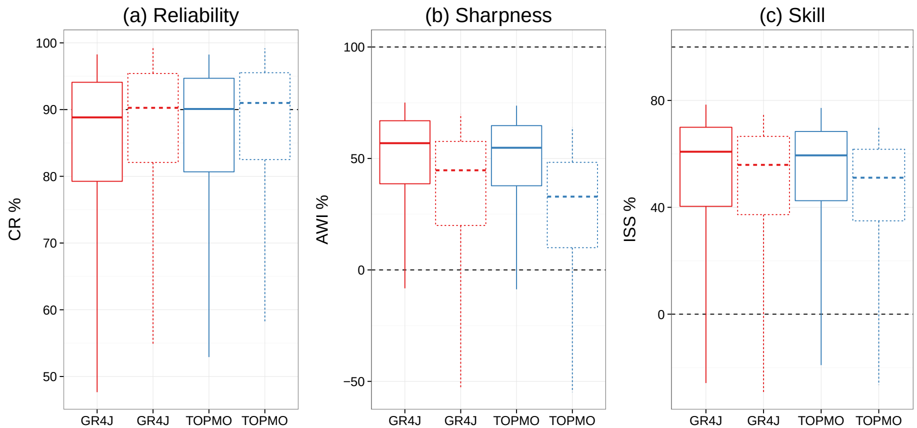

Figure 8. Distributions of the three performance criteria, obtained in two cases, (i) when 10 groups of relatives errors are used (continous lines) and (ii) when only one group is used (dashed lines). Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 907 catchments of the data set.

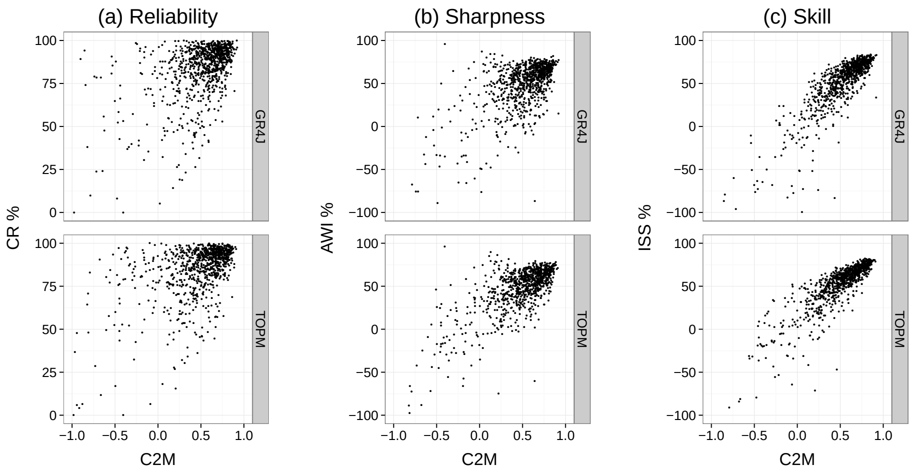

Figure 9. Impact of deterministic performance, as quantified by the bounded C2M quadratic criterion, on the three performance criteria for the 907 catchments. Note that for easing visualisation, the lower limits of AWI **(b)** and ISS **(c)** values are set to −100 % but lower values of AWI are obtained in 7 cases for both models, and lower values are obtained in 18 and 22 cases for GR4J and TOPMO respectively.

HESSD

11, 8039–8066, 2014