Transferring global uncertainty estimates from gauged to ungauged catchments

EDITOR DECISION: PUBLISH SUBJECT TO MINOR REVISIONS (EDITOR REVIEW) (09 MAR 2015) BY ROSS WOODS

Comments to the Author:

Thank you for your revisions. One of the referees accepts the revised version as suitable for publication. However, the other referee requests additional minor revisions, and makes a clear case for why they are needed. Please respond to the additional suggestions for changes to the manuscript.

Dear Editor,

We would like to thank you and the two reviewers for the overall positive feedbacks on the modified article.

In the following pages, we respond to the additional suggestions for changes to the manuscript provided by one reviewer.

In addition, we have included the corrections provided by a copy-editing professional.

REVIEWER #2 (R2)

I have had the opportunity to re-review the manuscript. While I believe that the revisions have helped to make the manuscript much easier to understand, there are still sections of the text that require additional editing to clarify the approaches. Therefore, I do not feel the publication can be accepted without these revisions.

We thank Reviewer R2 for his positive feedbacks about the modifications, and his constructive suggestions. We understand that there are some modifications that can improve further the clarity of the paper. We respond to his comments below.

1) Figure 1 is very helpful in understanding the proposed approach; however, additional editing is required. Some of the text in the green circles is not viewable. The ranges of flows for each flow group are not shown; the reader only sees that there are 10 groups. The ranges of flow for each group should be shown either in the figure or explained in the text. Step 3 should explain what the white dots represent in the plot - are these the uncertainty bounds? In step 4, more information is needed as to how do the flow groups get translated to the hydrogrpah?

Thank you for the suggestions. We modified accordingly the Figure to make it more readable. We provide a better explanation of the ranges of flow for each group in the legend of the Figure and also in the text. The white dots represent the multiplicative coefficients that are used to obtain the uncertainty bounds. We add this precision in the legend of the Figure. More information about the step 4 is also given in the text.

2) The description of the method within the text is still quite unclear. As currently presented, there is not enough detail for the reader to repeat this experiment using only the information presented about this method. For example, there is still not enough information as to how the 10 groups were chosen other than "created 10 groups of relative errors according to the magnitude of the simulated flows." The flow ranges for the 10 groups are not described so the reader has no way to apply this approach as described. Given that numerous reviewers had pointed out that additional justification is needed to define how the groups are chosen, the manuscript still has not adequately addressed this point. Further, in steps 3b and 4b, explain what is meant by a multiplicative coefficient and how exactly it is used in this process.

We understand the importance of reproducibility and we really want to make the method understandable so that the reader can test it. Therefore, we modified the description of the method to make every step clearer. The flow ranges are obtained by creating 10 groups equally populated: we use the deciles of the distribution of the simulated discharges to cut the flow range into 10 groups. A multiplicative coefficient is obtained when we work with relative values. Since we use relative errors, i.e., simulated discharge divided by observed discharge, the quantiles of the relative errors are multiplicative coefficients that relate simulated to observed discharge.

3) The Q5 and Q95 notation needs clarification. On pg. 11, sentence 1, Q5 and Q95 are used to describe the width of the interval; then Q95 and Q5 are described as being the 5th and 95th percentiles of the flow duration curve. Which is meant here? This needs clarification. This confusion is also in the conclusions section, which states (on top of p. 17), "We evaluated the approach over a large set of 907 catchments by assessing three expected quantiles of uncertainty estimates…" Are the Q5 and Q95 being referenced here or are there other quantiles that were used?

It seems that there is some misunderstanding, since in the conclusions section we refer to "three expected qualities", not "quantiles". Regarding the Q5 and Q95 notation, we believe that the description in the text and in the Figure 2 is adequate: we repeat several times that Q5 and Q95 are the 5^{th} and 95^{th} percentiles of the flow duration curve, and we gave the definition of the average width: AW = Q95 – Q5.

4) Section 5.3 is not adequate to justify the use of 10 flow groups. There are differences between the use of 10 and 1 group in figure 6 but not a dramatic one; therefore, perhaps it is also useful to try a few other groups (n=3,5,7) as well. Again, in practice, one would not be able to know how to apply this method from only these results.

We agree with the reviewer that defining the number of flow groups is a sensitive aspect of the method. As noticed by the reviewer, the change is not so large, but significant enough to consider it as an important aspect of the approach. However, adding an intermediate number of groups would not provide a clearer answer. Our experience is that the number of flow groups should be determined by considering at least two aspects: the first one is related to the shape of the scatter plots between relative errors and simulated flows. Even when using relative errors, the heteroscedasticity of errors may be low in some cases and larger in others. Besides, the simulation objectives should be also considered: the quantification of uncertainty for extremes values may require a larger number of groups than for intermediate ones. Obviously, the number of points (i.e. the number of time steps available) will also be a constraint to keep statistically significant sub-groups. Therefore, it is difficult to draw general conclusions on the number of groups to choose. The number of ten groups proposed here is a compromise between performance and objective.

We added this discussion in section 5.3 and acknowledged these aspects again in the conclusion. We think that providing more detailed guidance on this aspect will require much more research, which is out of the scope of this article.

Manuscript prepared for Hydrol. Earth Syst. Sci. Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 28 March 2015

Transferring global 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

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

Abstract

Predicting streamflow hydrographs in ungauged catchments is ~~a challengingissue~~challenging, and accompanying the estimates with realistic uncertainty bounds is an even more complex task. In this paper, we present a method to transfer global uncertainty estimates from gauged to ungauged catchments and we test it over a set of 907 catchments located in France, using two rainfall–runoff models. We evaluate the quality of the uncertainty estimates based on three expected qualities: reliability, sharpness, and overall skill. The robustness of the method to the availability of information on gauged catchments was also evaluated using a hydrometrical desert approach. Our results show that the method ~~holds interesting~~ presents advantageous perspectives, providing ~~in a majority of cases~~ reliable and sharp uncertainty bounds at ungauged locations in a majority of cases.

1 Introduction

1.1 Predicting streamflow in ungauged catchments with uncertainty estimates

Predicting the entire runoff hydrograph in ungauged catchments is a ~~challenging issue~~ challenge 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~~ to 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~~ summary 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~~ has yet been

reached on how to adequately quantify uncertainty. In practice, various methodologies are currently used.

For gauged catchments, the methodologies include Bayesian approaches ~~(see e.g., the review of Liu and Gupta, 2007)~~ (see e.g., the review by Liu and Gupta, 2007) , informal methods related to the GLUE framework (Beven and Freer, 2001), multi-model approaches (Duan et al., 2007; Velazquez et al., 2010) and other global 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; Kapangaziwiri et al., 2012). It is a two-step procedure. The first step consists in estimating with uncertainty various summary metrics of the hydrograph, 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 the ~~regionalized~~ regionalised metrics obtained during the first step. A ~~bayesian~~ Bayesian approach can also be used (Bulygina et al., 2011, 2012). The parameter sets are given a relative weight depending on the proximity of their summary metrics compared to ~~regionalized~~ regionalised metrics and depending on a priori information. The reader can refer to Wagener and Montanari (2011) for a comprehensive description of both formal and informal methods belonging to this framework.

One difficulty of the ~~above mentioned~~ 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~~ therefore it remains difficult to use them in practice. In addition, ~~solely relying~~ relying solely on an ensemble of model parameter sets to quantify total predictive uncertainty is often ~~not sufficient~~ insufficient when imperfect rainfall–runoff models are used.

A pragmatic alternative consists in addressing ~~separately~~ the parameter estimation and the global uncertainty estimation issues separately. It has been argued by several authors (Montanari and Brath, 2004; Andréassian et al., 2007; Ewen and O'Donnell, 2012) that a posteriori ~~characterization~~ characterisation of modelling errors of a "best" or "optimal" simulation can yield valid uncertainty bounds at gauged locations. In earlier studies, the terms ~~of total uncertainty, global uncertainty or~~ "total uncertainty", "global uncertainty" and "post-processing" approach have been used interchangeably to refer to this approach. The various sources of uncertainty are indeed lumped into ~~an~~ a unique error term with the goal to ~~estimate~~ estimating uncertainty bounds for model outputs.

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's 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*, under the condition that the ~~increase of~~ increased uncertainty introduced by the regionalisation step compared to the calibration step is adequately taken into account.

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

1.2 Scope of the paper

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

To our knowledge, a framework based on residual errors and global 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 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~~ fast-responding Mediterranean catchments with intense precipitation as well as larger, ~~groundwater-dominated~~ 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 two models use a soil moisture accounting procedure as well as routing stores. However, ~~their~~ they differ markedly in the formulation of their functions. While the GR4J model uses two non-linear stores and a unit-hydrograph, the ~~TOPM~~ TOPMO model uses a linear and an exponential ~~stores~~store, and a pure time delay.

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~~ optimisation procedure has proved to be efficient in past applications for the conceptual models used here. As an objective function, we used the Nash and Sutcliffe (1970) criterion computed on root square transformed flows (NSVQ). 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 are available for calibrating parameter sets at ungauged locations. ~~Thus~~Therefore, 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 NSVQ 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~~ The proximity of the catchments to the gauged ~~ones~~ catchments was quantified by the distances between ~~catchments~~ catchment 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~~ four and seven catchments for GR4J and TOPMO, respectively, following the work reported by Oudin et al. (2008).

3 Proposed approach: transfer of relative errors by flow groups

3.1 Description of the method

Transferring calibrated model parameters from gauged catchments to ungauged ~~catchment~~ catchments is a ~~well established~~ 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~~ ideas underlying the proposed approach ~~is~~ are to (i) ~~to~~ treat each donor as if it was ungauged (simulating flow ~~though~~ through the above described regionalisation approach), (ii) ~~characterize~~ characterise the empirical distribution of relative errors (understood as the ratio between observed and simulated flows, i.e. considering a multiplicative model error) for each of these donors ~~,~~ and (iii) transfer global uncertainty estimates to the ungauged catchment.

The methodology used to transfer global uncertainty estimates can be described by the following steps, illustrated ~~by~~ in Fig. 1:

1. Selection of catchments
   Here we consider a target ~~ungaged~~ ungauged catchment (TUC). This catchment has n neighbouring gauged catchments, called NGC_1, $\text{NGC}_2, \ldots, \text{NGC}_n$, which will be considered as donors for the TUC. For the ith catchment NGC_i, ~~there are~~ one can also select n neighbouring ~~catchment~~ catchments with the notation: NGC_{i1}, $\text{NGC}_{i2}, \ldots, \text{NGC}_{in}$, which can be considered as donors for NGC_i. Obviously, the TUC catchment would be excluded from this set of ~~second order~~ second-order donor catchments.

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_{ij} (with j between 1 and n).

   b. Compute the relative errors of discharge reconstitution (i.e. the ratio between observed and simulated discharges) by comparing simulated and observed discharge series for catchment NGC_i, and create 10 groups of relative errors according to the magnitude of the simulated discharge. To ensure that each group contains the same number of points, we split the simulated discharge ~~variable is cut into quantile groups~~range into 10 sub-groups of equal size, using the deciles

of the simulated discharge distribution. Using several flow groups allows taking into account the possible variability of model ~~errors~~ error characteristics.

3. Computation of the multiplicative coefficients applicable to simulated discharge

   a. Put together all the relative errors from the donors NGC_{ij} (with j between 1 and n) according to the group they belong to, i.e. ~~for a group k,~~ all relative errors of groups k of the n donors are assembled ~~.~~ into a master group k. This is done for k between 1 and 10.

   b. Compute the empirical quantiles of the relative error distribution within each ~~group. Each~~ master group k (with k between 1 and 10). Since relative errors were computed (i.e. ratio of simulated to observed discharge values), each quantile of relative ~~error~~ errors can be considered a multiplicative coefficient applicable to the simulated discharge. These multiplicative coefficients will be used to obtain the prediction bounds.

4. Computation of the uncertainty bounds for the target ~~ungaged~~ ungauged catchment TUC

   a. Apply the parameter regionalisation scheme to obtain a simulated discharge time series for the target ~~ungaged~~ ungauged catchment TUC using the parameter sets of the n neighbouring gauged catchments NGC_1, $\text{NGC}_2, \ldots, \text{NGC}_n$.

   b. Multiply the simulated discharge by the set of multiplicative coefficients obtained at Step 3b to obtain the uncertainty bounds. The coefficients calculated for the group k are used when the simulated discharge belongs to the group k.

Note that we based our approach on multiplicative errors and not on additive errors because using multiplicative coefficients ~~yield~~ yields prediction bounds for discharge that are always positive, whereas this might not always be the case with additive errors.

Finally, we mention that the choice to use 10 groups reflects a trade-off between the number of points available to obtain reasonable estimates of empirical quantiles computed for

each group and an adequate treatment of the variability of the characteristics of errors with the magnitude of simulated discharge. A larger (lower) number of groups could obviously be used if more (~~less~~fewer) data are available (see discussion in section 5.3) or based on the analysis of the statistical properties of errors.

3.2 Why ~~donors~~ should donors be considered as ~~ungaged~~ungauged?

The first step deserves a brief explanation. The choice to treat donors as ~~ungaged~~ ungauged is related to the well-known fact that the performance of rainfall–runoff models ~~decrease~~ decreases when they are applied at ~~ungaged~~ ungauged locations with a regionalisation scheme, compared to ~~the case where~~ when local data are available for parameter estimation. The quadratic efficiency criterion used here is the C2M (Mathevet et al., 2006), a bounded version of the Nash and Sutcliffe (1970) efficiency (NSE) criterion. The criterion is ~~solely based~~ based solely on the simulated discharges of the deterministic rainfall–runoff and is completely independent of the application of the uncertainty method. The equations are:

$$\text{C2M} = \frac{\text{NSE}}{2 - \text{NSE}} \tag{1}$$

$$\text{NSE} = 1 - \frac{\sum_{t=1}^{T} \left(Q_t^{obs} - Q_t^{sim}\right)^2}{\sum_{t=1}^{T} \left(Q_t^{obs} - \mu_o\right)^2} \tag{2}$$

where T is the total number of time-steps, Q_t^{obs} and Q_t^{sim} are the observed and simulated discharge~~respectively~~, respectively, at time-step t, and μ_o is the mean of the observed discharges. ~~The advantage of this bounded version is to avoid~~ This bounded version has the advantage of avoiding large negative values which are difficult to interpret.

Figure 3 illustrates the general ~~decrease of performance~~ performance decrease for both models on our catchment set when a regionalisation scheme is used instead of a parameter estimation based on local data. As a consequence~~we should expect~~, predictive uncertainty at ungauged locations we should expect to be larger than predictive uncertainty at gauged

~~location~~locations, i.e. ~~,~~ when the rainfall–runoff model is calibrated with observed discharge data. That is why ~~it is necessary to consider donors as ungaged~~donors must be considered as ungauged. We will come back to this important point in Section 5.

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. A general introduction to probabilistic evaluation can be found in Gneiting et al. (2007) and Wilks (2011), and in Franz and Hogue (2011) for a more hydrological perspective.

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~~ characterised 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. The average width (AW) of the uncertainty bounds is widely used to quantify sharpness~~,~~ :

$$\mathrm{AW} = \frac{1}{T}\sum_{t=1}^{T}\left(Q_t^u - Q_t^l\right) \tag{3}$$

where Q_t^l and Q_t^u are~~respectively~~ , respectively, the lower and upper bounds of the prediction interval $[Q_t^l, Q_t^u]$ at time-step t.

To ease comparison between catchments, we used the width of the 90 % interval [Q5, Q95],

$$\mathrm{AW}^{\mathrm{clim}} = \mathrm{Q95} - \mathrm{Q5} \tag{4}$$

where Q5 and Q95 are the 5th and 95th percentiles of the flow duration curve. This value ~~characterizes~~ characterises the natural variability of the flows for a given catchment and

has the same unit as the average width of the uncertainty bounds. It can be viewed as the average width of the uncertainty bounds of a climatological prediction, where the uncertainty bounds are constant ~~in~~ over time and defined by the interval [Q5, Q95]. A graphical illustration is given in Fig.2.

Comparing the two values AW and AW^{clim} leads to the following dimensionless criterion called the average with index (AWI):

$$AWI = 1 - \frac{AW}{AW^{clim}} \tag{5}$$

It is positive if the uncertainty obtained by ~~the application of the~~ applying the rainfall–runoff model and the methodology presented here is reduced compared to the climatology, 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 the width of an uncertainty bound and the position of the observed value compared to the uncertainty bound. The scoring rule of the interval score at time-step t is defined as:

$$S_t = \begin{cases} \left(Q_t^u - Q_t^l\right) & \text{if } Q_t^l \leq Q_t^{obs} \leq Q_t^u \\ \left(Q_t^u - Q_t^l\right) + \frac{2}{1-\beta}\left(Q_t^l - Q_t^{obs}\right) & \text{if } Q_t^{obs} < Q_t^l \\ \left(Q_t^u - Q_t^l\right) + \frac{2}{1-\beta}\left(Q_t^{obs} - Q_t^u\right) & \text{if } Q_t^{obs} > Q_t^u \end{cases} \tag{6}$$

where Q_t^{obs} is the ~~observed value~~ value observed at time-step t and β is equal to 90 % since a 90 % interval is sought here. See Fig.2 for an illustration of how S is computed.

IS is the average value of S_t over the time series:

$$IS = \frac{1}{T}\sum_{t=1}^{T} S_t \tag{7}$$

To ease comparison between catchments and evaluate the skill of the prediction bounds, we used the 90 % interval [Q5, Q95] as a benchmark, ~~similarly~~ similar to what we did for the sharpness index. The climatological prediction gives uncertainty bounds that are constant in time and defined by the interval [Q5, Q95], where Q5 and Q95 are the 5th and 95th percentiles of the flow duration curve. Thus we computed the interval skill score:

$$\mathrm{ISS} = 1 - \frac{\mathrm{IS}}{\mathrm{IS}^{\mathrm{clim}}} \tag{8}$$

where $\mathrm{IS}^{\mathrm{clim}}$ is the interval score obtained with the 90 % interval [Q5, Q95]. Using skill scores is a very common approach in probabilistic forecasting. ~~It allows to obtain dimensionless scores , similarly to the~~ Dimensionless scores can thus be obtained, in much the same way as the computation of the well-known Nash and Sutcliffe (1970) efficiency (NSE) criterion for assessing deterministic performance.

The interval skill score ~~ISS~~ (ISS) is positive when the prediction bounds are more skilful than climatology, and negative otherwise. The best value that can be achieved is ~~egal~~ equal to 1.

5 Results and discussion

5.1 Reliability, sharpness and overall skill

Figure 4 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~~ summarise 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 a majority of cases ~~,~~ able to reflect the magnitude of errors when predicting runoff hydrographs in ungauged catchments, even though it is clear that the perfect value of 90 % is not reached in most cases.

The CR values fall below 70 % for around 14 % of the catchments with GR4J ~~,~~ and 13 % with TOPMO, which indicates cases where the proposed approach yields predictive bounds that are clearly too narrow or biased (i.e. ~~,~~ not well centred on the observations). Note that we did not find ~~in the literature any guidance about~~ any guidance on how to properly evaluate the CR values in the literature. The results presented here may be used as a benchmark to comparatively assess the ranges of values that would be obtained in future studies.

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 could consider here uncertainty reduction values between 30 and 80 % as quite satisfactory, even though we ~~recognize~~ recognise that this statement is arbitrary since there ~~is~~ are no widely agreed values to base our analysis on.

Note that negative values are seen for 7 % of the catchments with both GR4J and TOPMO, which indicates cases where the approach yields prediction intervals whose average width is larger than the width of the [Q5, Q95] interval (Q5 and Q95 are the 5th and 95th percentiles of the flow duration curve).

5.1.3 Overall skill

Finally, Fig. 4c shows that the predictive skill for both models is positive for most catchments. 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?

As mentioned earlier, 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). 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~~ Similar to Fig. 4, Fig. 5 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 for both models, 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, since it reflects the ability of the uncertainty method to adequately reflect the magnitude of errors we might expect at locations for which prediction is done. Therefore, the benefit of treating the donor catchments as ungauged clearly appears in Fig. 5a, illustrating the theoretical argument presented in the methodological section.

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 global uncertainty estimates, but this time using only one group instead of 10.

Figure 6 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 increases when going from 10 groups to a single group, whereas both sharpness and skill decrease. It appears that improvements in terms of reliability are not sufficient to compensate for decreases in terms of sharpness when overall skill is considered. This can be understood by the fact that considering a single group instead of a few groups widens the uncertainty bounds on average, since the errors are generally heteroscedastic.

~~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 for the decrease in terms of uncertainty reduction and skill. We definitely prefer to maintain different~~ Obviously, although it appears that a single group is not enough to account for the variability of properties of relative errors, 10 groups may not provide significant performance gains and a compromise may be sought. The visual inspection of scatter plots between relative errors and simulated discharge reveals that the shapes can be very different between catchments, hence potentially requiring different numbers of groups. Besides, the simulation objectives, e.g. simulating intermediate or extreme flows, may also be considered when choosing the number of flow groups. Hence it appears that the number of groups may need further trial-and-error tests in specific applications to obtain the best compromise.

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

Although our tests reveal that the number of groups is a sensitive setting of the method, further research would be needed to evaluate whether different numbers of groups can be advised for specific objectives or conditions.

5.4 How does the performance 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 the characteristics of the uncertainty bounds at ungauged locations, the three criteria used to ~~characterize~~ characterise the uncertainty bounds are plotted in Fig. 7 as a function of a quadratic efficiency criterion for the 907 catchments, the C2M defined in Eq. 1.

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~~ greater 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 an ungauged location because there are cases where poor performance at neighbouring 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 global uncertainty estimates ~~performs well too~~also performs well.

5.5 How does the method perform in data-sparse conditions?

The results presented so far were obtained with a dense network of gauging stations. To investigate the impact of the network density on our results, we applied a demanding test called the hydrometrical desert. It consists in excluding potential donors that are closer to

the target ~~ungaged~~ ungauged catchment than a given threshold. For example, a threshold distance of 100 km means that the closest donor catchment must be at least 100 km ~~far from the ungaged~~ from the ungauged target catchment. This test results in a notable decrease of deterministic performance, as shown in Table 2, where the mean of the C2M efficiency criterion over the 907 catchments is reported ~~,~~ for both models. Note that this is a more demanding test than a decrease of network density, because catchments ~~keeps the possibility to still have~~ retain the possibility of still having close donors.

Figure ~~6~~ 8 shows the distributions of the three criteria obtained by applying the hydrometrical desert with threshold values of 10, 20, 50, 100 and 200 km, respectively. A clear decrease appears with increasing distances. While we should expect that the sharpness of the uncertainty bounds decreases because of larger errors, and that this situation leads to a decrease of skill, the results in terms of reliability reveal the limitation of the method. With increasing distances, the method becomes less able to transfer the appropriate magnitude of the larger errors.

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 have proposed a method allowing the transfer of global uncertainty estimates from gauged to ungauged catchments. The method extends the parameter transfer approach to the domain of global uncertainty estimation.

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

Nonetheless, the following limitations ~~of our~~ to the study can be mentioned:

1. Although the approach seems promising on average on the large catchment set we used, it is not able to adequately quantify the predictive uncertainty for some catchments and it failed in some cases.

2. The method might not perform well in in regions with sparser gauging networks than the one used here, as revealed by the application of a demanding test called the hydrometrical desert.

3. We only tested the 90 % prediction intervals, whereas the method could be applied to obtain other prediction intervals. We made this choice to keep the article as simple as possible, but further work could be done in that direction.

4. We also noted that the number of flow groups used in the approach may be a sensitive setting of the method, and further research would be needed to provide more detailed guidance on this point depending on the structure of the model errors and the modelling objectives.

It is worth stressing that although we used a transfer based on spatial proximity, the approach presented in this article is not only independent of the rainfall-runoff model but also of the regionalisation scheme used to obtain deterministic prediction at ungauged locations. Any other similarity measure could be a basis for transferring residual errors, including physical-based similarity measures. Accordingly, the regionalisation settings, including the output averaging option, could be adapted if deemed more appropriate.

Since we believe that uncertainty quantification has to be considered in any modelling study, further work should be devoted to the search for similarity measures that ~~do~~ not only perform well in allowing the transfer of parameter sets from donor to target catchments, but also allow transferring modelling error characteristics.

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. The authors also thank the Editor Ross Woods, who reviewed the manuscript, and the five reviewers, including Alberto Viglione and Denis Hughes, for their constructive comments, which helped improve the manuscript.

References

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

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

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

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.

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 Resources Research, 47, W02 503, doi:10.1029/2010wr009240, 2011.

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 Resources Research, 48, W06 519, doi:10.1029/2011wr011207, 2012.

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

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

Franz, K. J. and Hogue, T. S.: Evaluating uncertainty estimates in hydrologic models: borrowing measures from the forecast verification community, Hydrology and Earth System Sciences, 15, 3367–3382, doi:10.5194/hess-15-3367-2011, 2011.

Gneiting, T. and Raftery, A. E.: Strictly proper scoring rules, prediction, and estimation, Journal of the American Statistical Association, 102, 359–378, doi:10.1198/016214506000001437, 2007.

Gneiting, T., Balabdaoui, F., and Raftery, A. E.: Probabilistic forecasts, calibration and sharpness, Journal of the Royal Statistical Society Series B-Statistical Methodology, 69, 243–268, doi:10.1111/j.1467-9868.2007.00587.x, 2007.

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, Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 58, 1198–1255, doi:10.1080/02626667.2013.803183, 2013.

Kapangaziwiri, E., Hughes, D. A., and Wagener, T.: Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa, Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 57, 1000–1019, doi:10.1080/02626667.2012.690881, 2012.

Liu, Y. Q. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resources Research, 43, W07 401, doi:10.1029/2006wr005756, 2007.

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.

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, pp. 211–219, 2006.

Matott, L. S., Babendreier, J. E., and Purucker, S. T.: Evaluating uncertainty in integrated environmental models: A review of concepts and tools, Water Resources Research, 45, W06 421, doi:10.1029/2008wr007301, 2009.

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, Water Resources Research, 41, W12 434, doi:10.1029/2005wr004289, 2005.

Michel, C., Perrin, C., and Andreassian, V.: The exponential store: a correct formulation for rainfall-runoff modelling, Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 48, 109–124, doi:10.1623/hysj.48.1.109.43484, 2003.

Montanari, A.: Uncertainty of Hydrological Predictions, in: Treatise on Water Science, edited by Peter, W., pp. 459–478, Elsevier, Oxford, 2011.

Montanari, A. and Brath, A.: A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, Water Resources Research, 40, W01 106, doi:10.1029/2003wr002540, 2004.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I - A discussion of principles, Journal of Hydrology, 10, 282–290, 1970.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andreassian, 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, Journal of Hydrology, 303, 290–306, doi:10.1016/j.jhydrol.2004.08.026, 2005.

Oudin, L., Andreassian, V., Mathevet, T., Perrin, C., and Michel, C.: Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations, Water Resources Research, 42, W07 410, doi:10.1029/2005wr004636, 2006.

Oudin, L., Andreassian, 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 Resources Research, 44, W03 413, doi:10.1029/2007wr006240, 2008.

Oudin, L., Kay, A., Andreassian, V., and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, Water Resources Research, 46, W11 558, doi:10.1029/2009wr008887, 2010.

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

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

Perrin, C., Andreassian, 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 Resources Research, 44, W08 447, doi:10.1029/2007wr006579, 2008.

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, Environmental Modelling & Software, 22, 1543–1556, doi:10.1016/j.envost.2007.02.004, 2007.

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

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

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

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, International Journal of Climatology, 30, 1627–1644, doi:10.1002/joc.2003, 2010.

Wagener, T. and Montanari, A.: Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, Water Resources Research, 47, W06 301, doi:10.1029/2010wr009469, 2011.

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), Hydrology and Earth System Sciences, 15, 255–265, doi:10.5194/hess-15-255-2011, 2011.

Wilks, D. S.: Statistical methods in the atmospheric sciences, Academic, Oxford, 3rd edn., 2011.

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 Resources Research, 45, W12 422, doi:10.1029/2009wr007706, 2009.

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

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 Resources Research, 44, W00B04, doi:10.1029/2008wr006833, 2008.

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

Table 2. Mean C2M over the 907 catchments of the data set, with calibration (CAL), regionalisation (REGIO), and with the hydrometrical desert (HD) defined by increasing distance 10, 20, 50, 100 and 200 km.

	CAL	REGIO	HD-10	HD-20	HD-50	HD-100	HD-200
GR4J	~~0,67~~ 0.67	~~0,51~~ 0.51	~~0,49~~ 0.49	~~0,46~~ 0.46	~~0,43~~ 0.43	~~0,41~~ 0.41	~~0,35~~ 0.3
~~TOPM~~ TOPMO	~~0,59~~ 0.59	~~0,47~~ 0.47	~~0,46~~ 0.46	~~0,44~~ 0.44	~~0,41~~ 0.41	~~0,39~~ 0.39	~~0,34~~ 0.3

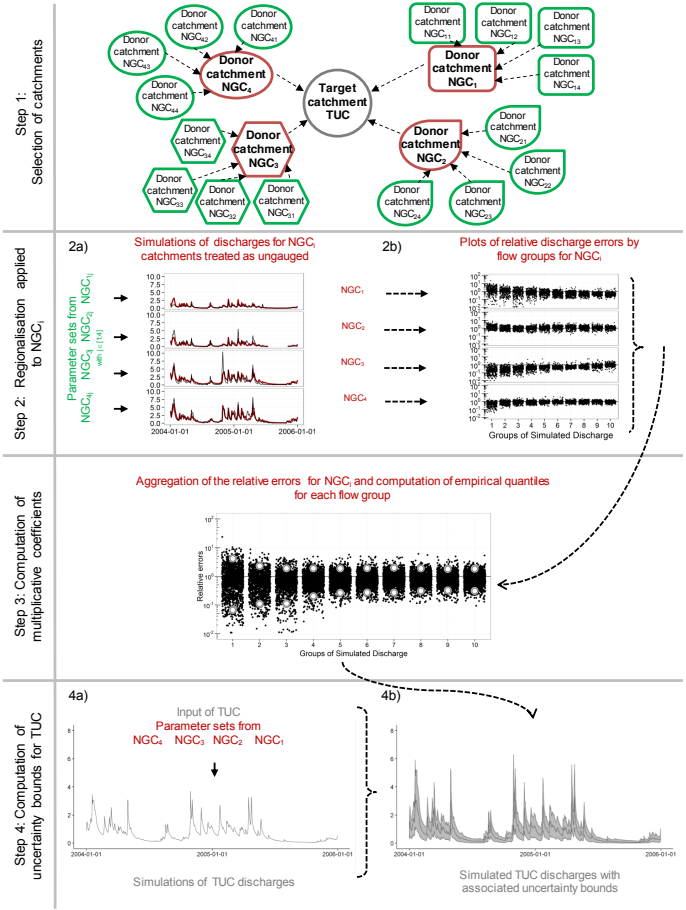
Step 1: Selection of catchments
Donor catchment NGC42
Donor catchment NGC41
Donor catchment NGC43
Donor catchment NGC4
Donor catchment NGC44
Donor catchment NGC34
Donor catchment NGC3
Donor catchment NGC33
Donor catchment NGC32
Donor catchment NGC31
Target catchment TUC
Donor catchment NGC11
Donor catchment NGC12
Donor catchment NGC13
Donor catchment NGC1
Donor catchment NGC14
Donor catchment NGC21
Donor catchment NGC2
Donor catchment NGC22
Donor catchment NGC24
Donor catchment NGC23
Step 2: Regionalisation applied to NGCi
2a)
Simulations of discharges for NGCi catchments treated as ungauged
Parameter sets from NGC1 NGC2 NGC3 NGC4 with i ∈ [1:4]
2004-01-01
2005-01-01
2006-01-01
2b)
Plots of relative discharge errors by flow groups for NGCi
Groups of Simulated Discharge
Step 3: Computation of multiplicative coefficients
Aggregation of the relative errors for NGCi and computation of empirical quantiles for each flow group
Relative errors
Groups of Simulated Discharge
Step 4: Computation of uncertainty bounds for TUC
4a)
Input of TUC
Parameter sets from
NGC4 NGC3 NGC2 NGC1
Simulations of TUC discharges
4b)
Simulated TUC discharges with associated uncertainty bounds

Figure 1. Illustration of the proposed approach, in the case of $n = 4$ donors. Red catchments are first-level donors while green catchments are second-level donors. For Step 2b, the simulated discharge variable (x axis) is split into 10 equal-size groups. In Step 3, white dots represent the values of the upper and lower multiplicative coefficients for each group. See the text for the description of the four steps.

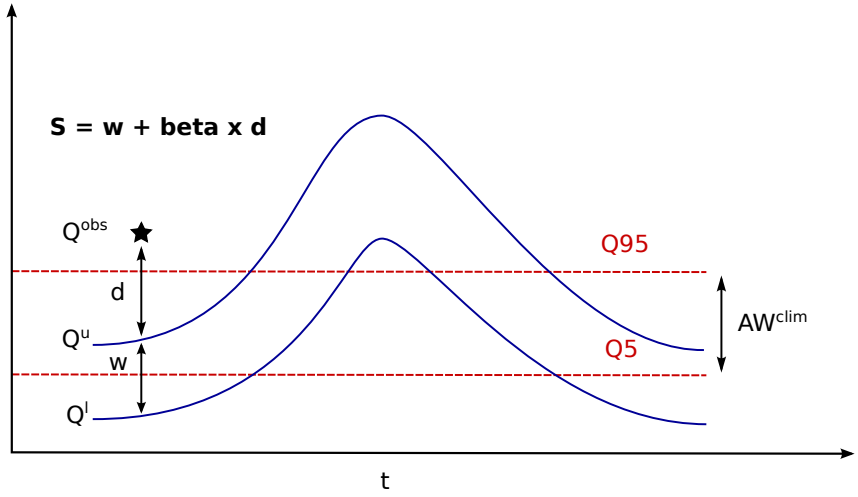

Figure 2. Illustration of the evaluation of the uncertainty bounds. Q5 and Q95 are the 5th and 95th percentiles of the flow duration curve. S is the interval score defined at one time-step for the situation where the observed value is above the upper limit of the uncertainty bound. See the text for further details.

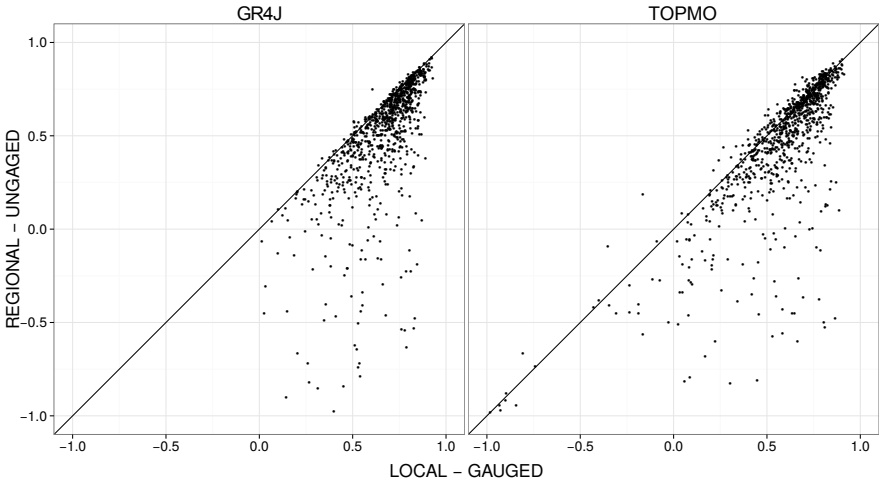

Figure 3. Impact of the regionalisation scheme on deterministic performance, as quantified by the bounded C2M efficiency criterion. Note that in a very few cases, the performance obtained with the regionalisation scheme is better than the performance obtained with calibration. This is possible because of the output averaging option used by the regionalisation scheme.

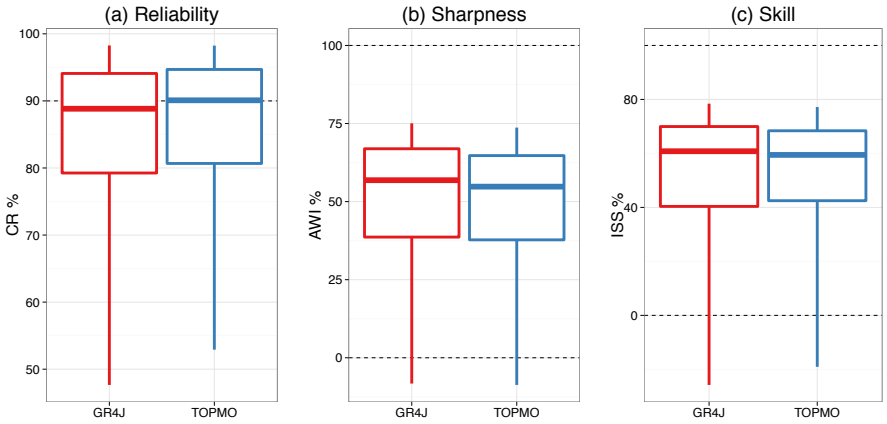

Figure 4. Distributions of the three performance criteria. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) ~~synthesize~~ summarise the variety of scores over the 907 catchments of the data set.

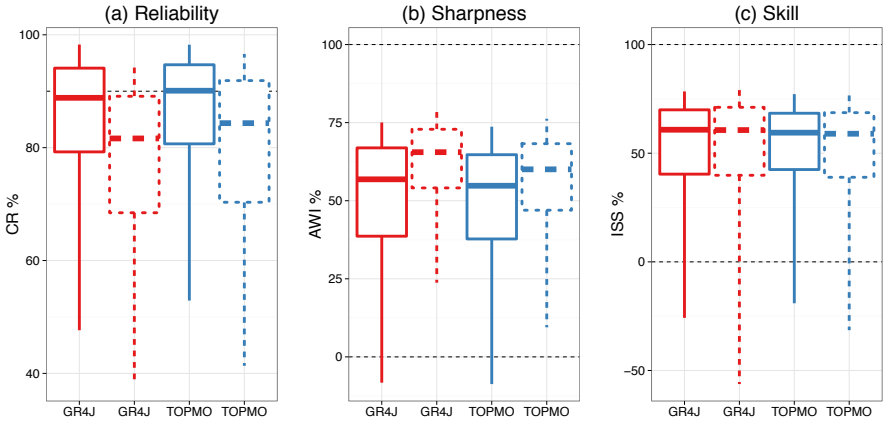

Figure 5. Distributions of the three performance criteria, obtained in two cases, (i) when the donor catchments are treated as ungauged (~~continuous~~ solid lines) and (ii) when the donor catchments are treated as gauged (dashed lines). Boxplots (5th, 25th, 50th, 75th and 95th percentiles) ~~synthesize~~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

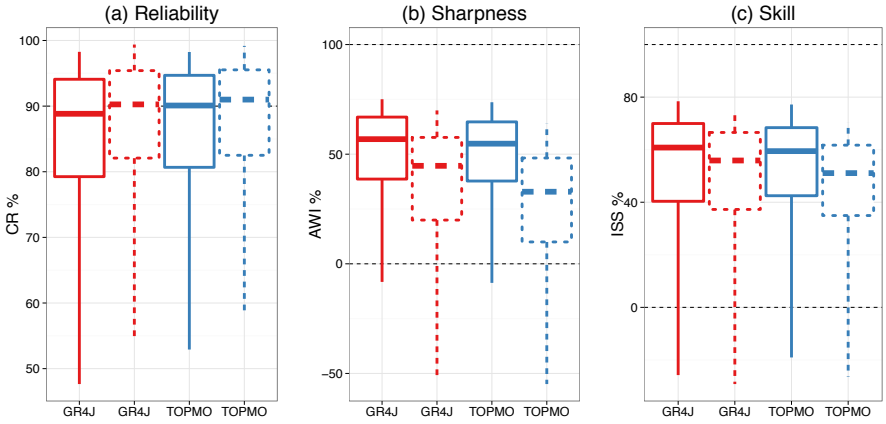

Figure 6. Distributions of the three performance criteria, obtained in two cases, (i) when 10 groups of ~~relatives~~ relative errors are used (~~continuous~~ solid lines) and (ii) when only one group is used (dashed lines). Boxplots (5th, 25th, 50th, 75th and 95th percentiles) ~~synthesize~~ summarise the variety of scores over the 907 catchments of the data set.

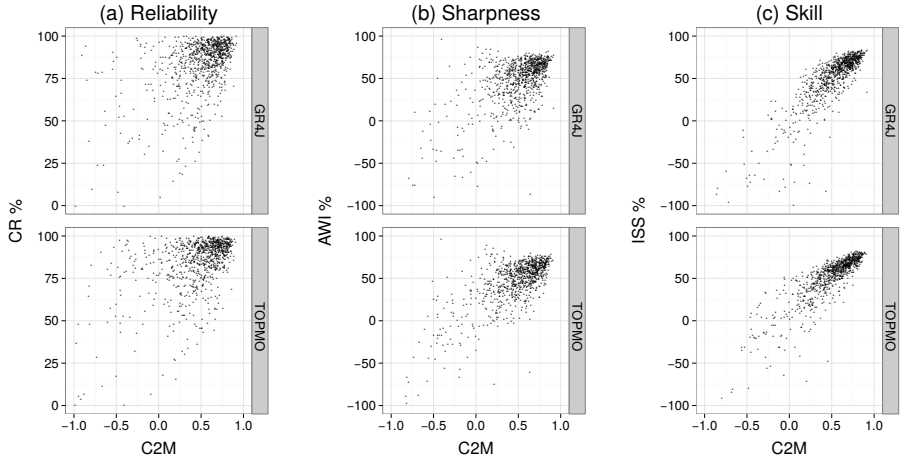

Figure 7. Impact of deterministic performance, as quantified by the bounded C2M quadratic criterion, on the three performance criteria for the 907 catchments. Note that for ~~easing~~ easier visualisation, the lower limits of the AWI **(b)** and ISS **(c)** values are set to $-100\,\%$ but lower ~~values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

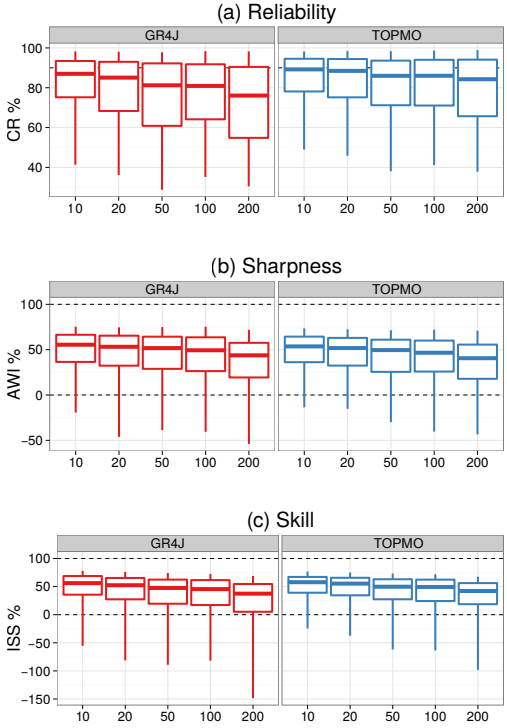

Figure 8. Impact of the hydrometrical desert on the distributions of the three performance criteria. Potential ~~donors~~ donor catchments are not retained as donors when their distance to the target catchment is below 10, 20, 50, 100 and 200 km. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) ~~synthesize~~ summarise the variety of scores over the 907 catchments of the data set.