# Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change?

Pierre Brigode, Ludovic Oudin, Charles Perrin

# Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change?

Pierre Brigode[1], Ludovic Oudin[1], Charles Perrin[2]

[1] UPMC Univ. Paris 06, UMR 7619 Sisyphe, Case 105, 4 Place Jussieu, F-75005 Paris, France

[2] Hydrosystems Research Unit (HBAN), Irstea, 1, rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony Cedex, France

Corresponding author. Email: pierre.brigode@upmc.fr

**Abstract**

This paper investigates the uncertainty of hydrological predictions due to rainfall-runoff model parameters in the context of climate change impact studies. Two sources of uncertainty were considered: (i) the dependence of the optimal parameter set on the climate characteristics of the calibration period and (ii) the use of several posterior parameter sets over a given calibration period. The first source of uncertainty often refers to the lack of model robustness, while the second one refers to parameter uncertainty estimation based on Bayesian inference. Two rainfall-runoff models were tested on 89 catchments in northern and central France. The two sources of uncertainty were assessed in the past observed period and in future climate conditions. The results show that, given the evaluation approach followed here, the lack of robustness was the major source of variability in streamflow projections in future climate conditions for the two models tested. The hydrological projections generated by an ensemble of posterior parameter sets are close to those associated with the optimal set. Therefore, it seems that greater effort should be invested in improving the robustness of models for climate change impact studies, especially by developing more suitable model structures and proposing calibration procedures that increase their robustness.

**Keywords:** Climate change, rainfall-runoff modelling, hydrological model calibration, uncertainty, robustness.

# 1 INTRODUCTION

## *1.1 Hydrological projections under climate change and their associated uncertainties*

The impacts of climate change on catchment behaviour have been extensively investigated over the last few decades (see e.g. for Europe Arnell, 1999a, Arnell, 1999b; for Australia Vaze et al., 2011 and Vaze and Teng, 2011). Quantitatively assessing the uncertainties associated with hydrological projections is a difficult task, even if qualitatively it is now recognised that these uncertainties are considerable. They stem from the methods used to generate climate projections as well as from hydrological modelling. Moreover, the relative importance of the various uncertainty sources is not easy to assess. Wilby and Harris (2006) proposed a framework to assess the relative weights of the sources of uncertainty in future low flows for the River Thames. They consider that uncertainty sources should be ranked in decreasing order as follows: Global Circulation Models (GCMs) > (empirical) downscaling method > hydrological model structure > hydrological model parameters > emission scenario. However, this conclusion was obtained using only two rainfall-runoff models applied to a single catchment. Wilby (2005) noted that depending on the rainfall-runoff model used (and possibly the catchment studied), the uncertainties associated with hydrological modelling may predominate. More recently, Chen et al. (2011) showed on a Canadian catchment that the choices of GCMs and downscaling techniques are the greatest uncertainty sources in hydrological projection estimations, followed by emission scenarios and hydrological model structures, and last hydrological model parameter estimation. On several southeastern Australian catchments, Teng et al. (2012) also showed that uncertainties stemming from fifteen GCM outputs are much greater than the uncertainties stemming from five hydrological models. Focusing on future hydrological trends in the UK, Arnell (2011) showed that "uncertainty in response between climate model patterns is considerably greater than the range due to uncertainty in hydrological model parameterization." These results show that the

uncertainties generated by the hydrological modelling step, though generally lower than that generated by the climate modelling step, can be significant in some cases and should not be ignored in climate change impact studies.

The common sources of uncertainty in hydrological modelling in stationary conditions (in terms of climate conditions and/or physical characteristics) include errors in model structure, problems in the calibration procedure, and errors in the data used for calibration. In non-stationary conditions, as in climate change studies, additional uncertainties may come from parameter instability due to the possible changes in the physical catchment characteristics and in the dominant processes. In both cases, model structure errors and the identification of model parameters can generally be considered as the two main sources of uncertainty in the hydrological modelling step. Several methods exist for studying uncertainties due to model structure (see e.g. Refsgaard et al., 2006). In climate change impact studies, the errors stemming from the model structure are usually assessed using several rainfall-runoff models and quantifying the range of their outputs (Booij, 2005; Wilby, 2005; Wilby and Harris, 2006; Jiang et al., 2007). The problem of parameter identification has been widely investigated and many methods to quantify the associated uncertainty have been proposed (see e.g. Matott et al., 2009, for a review). In a recent study, Bastola et al. (2011) attempted to quantify these two hydrological uncertainty sources (model structure and parameter sets) in a climate change context using a multi-model approach combining multiple emission scenarios and GCMs, four conceptual rainfall-runoff models and two parameter uncertainty evaluation methods (Generalized Likelihood Uncertainty Estimation and Bayesian Model Averaging). The authors concluded that "the role of hydrological model uncertainty is remarkably high and should therefore be routinely considered in impact studies." Note that the type of hydrological model used (physically-based or conceptual, lumped or distributed, etc.) may also be considered as an uncertain choice, in both stationary and non-stationary conditions. It is often

considered that the physical basis of process descriptions is indispensable to maintain the predictive power of hydrological models in a changing climate (see e.g. Ludwig et al., 2009). A few studies have covered this issue by considering both conceptual and physically-based models (see e.g. Poulin et al., 2011). The RExHySS project (Ducharne et al., 2009; Ducharne et al., 2011) addressed this issue on the Seine River basin (France) by considering seven hydrological models, including distributed (predominantly) physically-based models, semi-distributed physically-based models and lumped conceptual models. Interestingly, the results showed that the conceptualisation of the models was not the main source of variability in hydrological projections among the model simulations since large differences were found between models with similar conceptualisations.

### *1.2 Can model parameter instability be a major source of uncertainty?*

Other studies have investigated the dependence of the model parameters on the characteristics of the record period used for calibration. In climate change impact studies, the record period used to calibrate the model differs from the projected period. Since rainfall-runoff model parameters must be calibrated using the available data sets, they will partially account for the errors contained in these data (see e.g. Yapo et al., 1996; Oudin et al., 2006a) and/or their specific climate characteristics (see e.g. Gan and Burges, 1990). This is a well-known issue for conceptual rainfall-runoff models but physically-based models are also affected by this problem (see e.g. Rosero et al., 2010). Model parameters are the integrators of the data's information content. Different time periods used for calibration may provide quite different optimum parameter sets, depending on whether the period is dry or wet, for example, thus providing an estimation of parameter uncertainty with respect to their lack of robustness. Here Beven (1993) states that "it is easy to show that if the same model is 'optimised' on two different periods of record, two different optimum parameter sets will be produced. Extension to multiple calibration periods, if the data were available, would yield multiple optimum

parameter sets. The resulting parameter distributions would reflect the uncertainty in the parameter estimates and the interaction between the individual parameters." As stressed by Gan and Burges (1990), this obviously "should be heeded by modelers who use calibrated conceptual models to explore hydrologic consequences of climate change." As a consequence, and without clear guidelines on how the model should be calibrated for climate change impact studies, most hydrologists calibrate their models with all the available data (e.g. Vaze and Teng, 2011) or with the longest observed period they consider representative of the current hydro-climatology conditions (e.g. Poulin et al., 2011), generally considering a priori that "the longer the calibration period, the more robust the parameter set."

One way to evaluate the capacity of models to represent the hydrological behaviour of a catchment in a changing climate is to apply the differential split-sample test, introduced by Klemeš (1986). In this testing scheme, two contrasted periods are identified in the available record and the split-sample test is performed using these two periods. If the model is intended to simulate streamflow under wetter climate conditions, then it should be calibrated on a dry period selected in the available record and validated on a wet period. Conversely, if it is intended to simulate flows under drier climatic conditions, the reverse should be done. The model should demonstrate its ability to perform well in these contrasted conditions. Despite the simplicity of the test, relatively few authors have followed the differential split-sample test in the past (see e.g. Jakeman et al., 1993; Refsgaard and Knudsen, 1996; Donnelly-Makowecki and Moore, 1999; Seibert, 2003; Wilby, 2005). More recently, Merz et al. (2011) applied the test to a large set of 273 catchments in Austria and found that the parameters of the hydrological model controlling snow and soil moisture processes were significantly related to the climatic conditions of the calibration period. Consequently, the performance of the model was particularly affected if the calibration and the validation periods differed substantially. Vaze et al. (2010) also applied the differential split-sample test to 61 catchments

159 in southeast Australia with four conceptual hydrological models. They found that the
160 performance of these models was relatively dependent on the climatic conditions of the
161 calibration period. On 216 southeastern Australian catchments, Coron et al. (2012)
162 highlighted the same lack of robustness of three hydrological models tested in climatic
163 conditions different from those used for parameter calibration. Vaze et al. (2010) therefore
164 suggest that it would be wiser to calibrate model parameters on a portion of the record with
165 conditions similar to those of the future period to simulate. This idea was also put forward by
166 de Vos et al. (2010), who proposed dynamically re-calibrating model parameters for each
167 temporal cluster by finding analogous periods in the historical record. Following similar
168 motivations, Luo et al. (2011) showed that more consistent model predictions on specific
169 hydrological years are obtained if a selection of calibration periods is performed, and Singh et
170 al. (2011) used adjusted parameter values depending on the aridity of the catchment
171 considered for improving model prediction. These methodologies could be applied for
172 simulating future hydrological conditions, but unfortunately, as stated by Prudhomme and
173 Davies (2009), long records that include climatic conditions similar to what could be expected
174 in the future are lacking. This makes it difficult to identify a set of parameters specific to such
175 future conditions. Note that in some regions, climate changes have occurred in the past and it
176 is therefore possible to objectively assess the potential of hydrological models to cope with
177 changing climate. This is the case for Central and Western Africa, affected by a marked
178 reduction in rainfall and runoff from the year 1970 onwards. Using different models on
179 different catchments in this region, Niel et al. (2003) and Le Lay et al. (2007) showed no
180 evidence that non-stationarity in climate would incur model parameter instability.

181 ### *1.3 Scope of the paper*

182 This paper intends to investigate the uncertainty of hydrological predictions for the future
183 climate. To this aim, we followed Klemeš's differential split-sample test and assessed the

corresponding variability of the simulated hydrological impacts of projected climate when considering alternatively (i) the dependence of the optimal parameter set on the calibration period characteristics and (ii) an ensemble of posterior parameter sets obtained on a given calibration period. Each source of uncertainty was already studied in the context of changing climate, but their relative importance has not been assessed so far. Besides, most studies focusing on the parameter's dependency on climate conditions did not assess the consequences of choosing various calibration strategies on future hydrological projections. Here we will attempt to assess the long-term effects of these two sources of uncertainty in future conditions.

## 2 DATA AND MODELS

### *2.1 Catchment set*

A set of 89 catchments located in northern and central France was used, namely the Somme River at Abbeville, 22 sub-catchments of the Loire River basin and 66 sub-catchments of the Seine River basin (see Figure 1). Catchment area ranges from 32 to 109,930 km², runoff yield ranges from 0.11 to 0.69 and the aridity index (here defined as the ratio of mean annual Penman (1948) potential evapotranspiration to mean annual rainfall) ranges from 0.64 to 1.39. Compared to the Seine basin sub-catchments, the sub-catchments of the Loire River basin add diversity in terms of physiographic characteristics (with generally larger areas, higher elevations, and a different geological context), and hydro-climatic characteristics (with generally higher runoff yields). None of the catchments studied is strongly influenced by upstream dams.

*FIGURE 1: Location and distribution of various characteristics of the 89 catchments used. The boxplots show the 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles (67 is the number of catchments in the Seine and Somme basins, 22 in the Loire basin).*

### *2.2 Hydro-meteorological data*

The hydrological models tested here require only daily time series of potential evapotranspiration (PE) and rainfall (P) as input data. We used climate data from the SAFRAN meteorological reanalysis (Quintana-Segui et al., 2008; Vidal et al., 2010), which provides daily series of Penman PE and P from 1970 to 2007 at a mesoscale (on an 8 × 8 km grid). These data were aggregated for each catchment in order to estimate mean areal inputs. Besides daily streamflow (Q), time series were used to calibrate the models and assess their performance.

Since it is not within the scope of this paper to discuss the uncertainties related to climate projections, the outputs of a single general circulation model (GFDL CM2.1) driven by the A1B emissions scenario (IPCC, 2007) were chosen as climate projections. These outputs were regionalised using a statistical downscaling method based on weather types (Boé et al., 2006), producing a database at the same spatial resolution as the SAFRAN database (8 × 8 km). Three time slices with continuous daily series of PE and P were used in this study:

- 1980–2000, referred to as "present time" and noted PT hereafter;
- 2045–2065, referred to as "mid-century" and noted MC hereafter;
- 2080–2100, referred to as "end-of-century" and noted EC hereafter.

This scenario was tested on the Seine and the Somme basins within the RExHySS project (Ducharne et al., 2009, Ducharne et al., 2011) and on the Loire River basin within the ICC-Hydroqual project (Moatar et al., 2010). For the Seine and the Somme basins, the downscaled

projection simulates an increase in mean annual air temperature of 1.8°C by MC and 3.1°C by EC, a decrease in mean annual precipitation of 5% by MC and 10% by EC (with an increase of winter precipitation and a decrease of summer precipitation) and an increase in potential evapotranspiration of 16% by MC and 26% by EC. These predictions are close to the mean trends estimated with up to 14 climate projections used in the RExHySS and ICC-Hydroqual projects, making the scenario used in this study an in-between scenario.

### *2.3 Rainfall-runoff models*

Two daily continuous lumped rainfall-runoff models were used to avoid providing model-specific conclusions:

- The GR4J rainfall-runoff model, an efficient and parsimonious (four free parameters) model described in detail by Perrin et al. (2003);
- The TOPMO model (six free parameters), inspired by TOPMODEL (Beven and Kirkby, 1979; Michel et al., 2003), already tested on large data sets. This lumped model is quite different from GR4J but has comparable performance (see e.g. Oudin et al., 2006b). Here the distribution of the topographic index is parameterised and optimised, and not calculated from a digital elevation model. This was found to have only a limited impact on model efficiency, as shown by Franchini et al. (1996) and this eases the application of the model when it is tested on a large set of catchments.

Note that we did not explicitly investigate the uncertainties stemming from hydrological model structures. This was analysed e.g. by Seiller et al. (2012) using a multi-model approach in a climate change perspective.

### *2.4 Model parameterisation*

The optimisation algorithm used to calibrate parameter values is the Differential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt et al., 2009). DREAM optimises model

parameters on a given period and additionally infers the posterior probability distribution of model parameter values through Markov Chain Monte Carlo (MCMC) sampling.

As an objective function, we used the formal Generalized Likelihood function described by Schoups and Vrugt (2010), which considers correlated, heteroscedastic, and non-Gaussian errors (noted GL hereafter). The results in validation are analysed with the Nash and Sutcliffe (1970) efficiency criterion, which is still widely used in modelling studies. It was computed on root square transformed flows (noted NSEsq hereafter), which makes it possible to assess model efficiency for both high and low flows (Oudin et al. 2006b).

# 3 METHODOLOGY FOR INVESTIGATING PARAMETER UNCERTAINTY IN A CHANGING CLIMATE

## *3.1 General methodology*

The building blocks of the method originate from the differential split-sample test recommended by Klemeš (1986) and the methodology followed by Wilby (2005). The parameter uncertainty associated with the changing climate is characterised by the variability of the parameters across calibration sub-periods with varying hydroclimatic characteristics. The methodology is carried out in three steps (see Figure 2):

- Step 1: identification of test periods.
- Step 2: model parameter calibration and identification of posterior parameter sets.
- Step 3: model validation and simulation, and parameter uncertainty quantification.

These three steps are further detailed hereafter.

*FIGURE 2: Illustration of the three-step methodology used for investigating parameter uncertainty in a changing climate.*

Note that a similar methodology was used in the RheinBlick project (Görgen et al., 2010) for quantifying uncertainties due to the parameters of hydrological models in a climate change perspective on the Rhine basin.

### *3.2 Step 1: Identification of climatically contrasted sub-periods*

For each catchment, four climatically contrasted 3-year sub-periods were identified in the available record: a wet sub-period, two dry ones and an intermediate one. The driest sub-period will be used as the validation period (hereafter noted dry validation sub-periods) and the three others will be used as calibration sub-periods. This choice was made because the selected climate projection indicates that future conditions will be drier and warmer on the test basins. The aridity index (here defined as the ratio of mean Penman potential evapotranspiration to mean precipitation) was used to characterise the climatic specificity of each sub-period: the wet sub-period corresponds to the three contiguous hydrological years with the lowest aridity index (here a hydrological year starts on September 1 and ends on August 31). The choice of this index is rather arbitrary and may influence the results obtained hereafter. However, since it is solely based on climate characteristics, this makes it possible to assess the climatic specificity of the chosen sub-periods compared to the projected future climate. Obviously, it was not possible to use a criterion based on observed streamflows, as done by Seibert (2003) on observed data, since future flow observations by definition do not exist.

The choice of the length of the record sub-period to consider is not straightforward since it is based on a trade-off between two opposite expectations: (i) the longer the sub-periods, the more robust the set of parameters should be and (ii) the shorter the sub-periods, the more climatically contrasted sub-periods can be found in the record period. A review of the literature (see e.g. the review proposed by Perrin et al. (2007)) shows that there is no clear

consensus on the minimum length of calibration period for rainfall-runoff models, which is probably attributable to the specificity of the catchments and models used in those studies. Specifically for the two parsimonious models used in this paper, Anctil et al. (2004) obtained good GR4J performance with 3- to 5-year calibration periods and Perrin et al. (2007) showed that the calibration of the GR4J and TOPMO models with the equivalent of only 1 year of data can provide acceptable performance. Thus, it seems that 3-year periods can yield acceptable parameter sets. Those relatively short sub-periods allow representing significantly contrasted climatic conditions. Interestingly, Figure 3 shows that the contrast between the aridity indexes of the different calibration sub-periods is similar to the contrast between the aridity indexes of the observed record and future climate projection. However, it should be noted here that the climate projection simulates systematically drier conditions than the dry validation sub-periods. This means that whatever the selected calibration sub-period, the model is applied in extrapolation in future climate conditions. Note that the aridity index does not reflect seasonal variability of precipitation and potential evapotranspiration: two sub-periods with similar values of the aridity index may be quite different in terms of climate seasonality. This means that seasonal indexes would be useful to consider as additional criteria for period selection if seasonal contrasts were under study.

*FIGURE 3: Comparison of Aridity Index (AI) values for the different calibration and validation sub-periods considered and for the three time slices (PT, MC, EC) for the 89 catchments.*

### *3.3 Step 2: Model calibrations on the specific periods*

For each catchment, the two hydrological models were calibrated using the three climatically contrasted sub-periods (i.e. the wet, mean and dry sub-periods) and the whole record period

329 (except the dry validation sub-periods, which are used for model validation in step 3). A 1-
330 year warm-up period was considered for each simulation.

331 The DREAM algorithm was used to infer the most likely parameter set and its underlying
332 posterior probability distribution. We selected for each calibration run (i) the optimal
333 parameter set defined as the parameter set maximising the GL objective function and (ii) an
334 ensemble of 2000 posterior parameter sets representing the posterior probability distribution
335 of parameter sets. For each catchment and each calibration period, we checked that the
336 DREAM algorithm converged to the stationary distribution representing the model's posterior
337 distribution by analysing the Gelman-Rubin convergence statistics.

338 Note that additional model calibrations were also performed on the dry validation sub-periods.
339 The corresponding calibration performance was used as a reference to evaluate the
340 performance of models validated on these dry validation sub-periods after calibration on other
341 periods.

342 ***3.4 Step 3: Model simulations with different parameter sets***

343 At this stage of the methodology, four optimal parameter sets corresponding to the four
344 calibration periods and an ensemble of 2000 posterior parameter sets identified throughout the
345 record period (except the dry validation sub-periods) are available for each catchment.

346 All these parameter sets were used for each catchment to simulate the streamflow time series
347 over the dry validation sub-periods (illustrated as grey hydrographs in the first line of the third
348 step of Figure 2) and on the three time slices (PT, MC and EC) (illustrated as grey envelops
349 on the flow duration curves plotted in the second line of the third step of Figure 2). Three
350 typical streamflow characteristics were analysed:

351 - The 95th flow exceedance percentile of the flow duration curve, $Q_{95}$ (mm/day),
352 describing low flows;

353 - The mean annual streamflow, $Q_{MA}$ (mm/y), indicating the overall water availability;

- The 5th flow percentile, $Q_{05}$ (mm/day), describing high flows.

For each catchment and each model, the four ensembles of parameter sets were tested first on the dry validation sub-periods. We analysed the dependence of model performance on the climatic specificity of the calibration period. Furthermore, the biases between the observed and the simulated streamflow characteristics were assessed. Second, the variability of the future streamflow simulations obtained using various calibration conditions was analysed for each future time slice. To differentiate the impacts stemming from the specificity of the calibration period from those associated with the “classical" parameter uncertainty approach based on Bayesian inference on the whole record period, the results are presented step by step hereafter.

# 4 RESULTS

## *4.1 Calibration performance results*

In this section, the general calibration performance of the two models are analysed. Figure 4 presents the distributions of the GL function evaluations and the distributions of the Nash-Sutcliffe efficiencies computed on root square transformed flows (NSEsq) obtained by the GR4J and TOPMO models on the catchment set. The distributions were obtained with (i) the calibration efficiencies over the whole record without the dry validation sub-periods obtained with optimal parameter sets, i.e. 89 values for each model (white boxplots, noted OPT) and (ii) calibration performance over the same record periods obtained with the 2000 posterior parameter sets identified for each of the 89 catchments, i.e. 178,000 values for each model (grey boxplots, noted POS). The distributions of the GL objective function values highlight that optimal parameter sets present similar general calibration efficiency to the calibration efficiency obtained using the populations of posterior parameter sets. Considering populations of posterior parameter sets thus adds a limited variability in terms of calibration performance over the whole record periods (without dry validation sub-periods) for the two models. For

GR4J, the distributions of the NSEsq efficiencies similarly show that the optimal parameter identified with the DREAM algorithm and with the GL objective function have similar general performance as the posterior parameter sets. The performance losses when considering populations of posterior parameter sets instead of optimal parameter sets are more significant for TOPMO than for GR4J, with median NSEsq moving from 0.86 with optimal parameter sets to 0.84 with posterior parameter sets. Calibration performance results for the other calibration sub-periods (wet, mean and dry 3-year calibration sub-periods) show the same general tendencies (not shown here). The difference between the two models might stem from the number of free parameters, higher for TOPMO (six free parameters) than for GR4J (four free parameters). Thus, the calibrated parameter values of TOPMO may show greater sensitivity to the choice of the objective function. Finally, note that the general performance of both models is quite reasonable, with half of the catchments studied presenting calibration performance obtained with optimal parameter sets on the whole record periods (without the dry validation sub-periods) greater than 0.85 for the two models.

*FIGURE 4: Distributions of the GL objective function values (top) and of the NSEsq values (bottom) of the two models illustrating (i) calibration performance over the whole record periods without the dry validation sub-periods obtained with optimal parameter sets (white boxplots, noted OPT) and (ii) calibration performance over the whole record periods obtained with posterior parameter sets (grey boxplots, noted POS). Results are shown for GR4J (left) and TOPMO (right). The boxplots show the 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles.*

### *4.2 Sensitivity to the climate characteristics of the calibration period*

In this section, the model outputs are analysed considering different calibration periods. First the efficiency of the model on the dry validation sub-periods is discussed in terms of NSEsq and simulation of standard streamflow characteristics. Second, we analyse the resulting spread of the simulated streamflows for the future time slices.

- **Efficiency on dry validation sub-periods:**

Figure 5 shows the distributions of the NSEsq values on the catchment set obtained by the two models in (i) calibration over the dry validation sub-periods and (ii) validation over the dry validation sub-periods using the other four calibration sub-periods considered (wet, mean, dry, and whole record periods). Models calibrated over different sub-periods generally encountered similar difficulties simulating flows on the dry validation sub-periods since the validation efficiencies are clearly reduced compared to the calibration efficiencies on this sub-period. The differences between the four calibration strategies (over a wet, a mean, a dry sub-period or a long period) are limited but, for both models, using the wettest sub-periods for calibration appears particularly detrimental to simulating the dry validation sub-periods. GR4J and TOPMO obtained marginally better validation results using dry and mean conditions for calibration, respectively. Interestingly, calibrating the models on the whole record period (except the dry validation sub-periods, resulting in 20 years of record on average) does not warrant a particularly robust estimation of optimal parameter sets, since the validation efficiencies are generally similar to those obtained with 3-year calibration periods. This is not consistent with the wide-spread idea that the longer the calibration period, the more robust the parameter set.

These results corroborate the previous findings of Vaze et al. (2010), Merz et al. (2011) and Coron et al. (2012) obtained with different catchments and models, emphasising the lack of

robustness of conceptual rainfall-runoff models when the climatic settings between calibration and validation periods are different.

*FIGURE 5: Distributions of the NSEsq values obtained by the two models illustrating (i) calibration performance over the dry validation sub-periods (black boxplots) and (ii) validation performance over the dry validation sub-periods using the other four calibration sub-periods considered (wet, mean, dry, and whole record without the dry validation sub-period illustrated, respectively, with blue, green, red and white boxplots). Results are shown for GR4J (left) and TOPMO (right). The boxplots show the 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles.*

Figure 6 summarises the results of the models' sensitivity to the climatic specificity of the calibration period on the observed dry validation sub-periods. This figure is organised as a table with two columns and three rows: each column represents a hydrological model (left: GR4J; right: TOPMO) and each row represents a specific characteristic of the simulated flow series (from top to bottom: $Q_{95}$, $Q_{MA}$ and $Q_{05}$). For each model and for each streamflow characteristic, the plot on the left shows the observed versus simulated value for each catchment, each dot representing the mean of simulated values obtained with the four optimal parameter sets and each bar representing the range of simulated values when using the four optimal parameter sets. Ideally, all range bars should be centred on the 1:1 line, meaning that streamflow simulated by parameter sets originating from different calibration sub-periods are all equal to the observed streamflow. The boxplots on the right represent the distributions of the relative errors on the flow characteristic on the dry validation sub-periods over the 89 catchments when considering the four calibration sub-periods. These relative errors were estimated as the ratio of the difference between simulated and observed flow characteristics to

451 the observed flow characteristics. Ideally, all the boxplots should be centred on a null value of
452 the bias between the observed and the simulated streamflow.

453 The first main result is that the two rainfall-runoff models present similar overall efficiency in
454 simulating the flow characteristics on the dry validation sub-periods (graphs on the left). This
455 efficiency is rather limited since the median absolute bias is greater than 0.1 for both models.
456 Even for the estimation of mean annual flow ($Q_{MA}$), none of the four calibration strategies
457 yields a median absolute bias lower than 0.1. The second main result is that the impact of the
458 climatic specificity of the calibration sub-periods on the modelled flow characteristics is not
459 straightforward (graphs on the right). For GR4J, it seems that the 3-year dry calibration sub-
460 periods provide the least biased estimations of the three streamflow characteristics of the dry
461 validation sub-periods. Using wet and mean 3-year calibration sub-periods tends to yield
462 overestimated flow simulations on the dry validation sub-periods. Conversely, TOPMO tends
463 to underestimate flows of the dry validation sub-periods. The mean 3-year calibration sub-
464 periods seems to provide less biased estimation of the streamflow characteristics on the dry
465 validation sub-periods. Finally, using 3-year calibration sub-periods (dry ones for GR4J and
466 mean ones for TOPMO) yields less biased predictions than when considering the whole
467 available records for calibration for both hydrological models and for the three streamflow
468 characteristics studied here, which corroborates the validation performance illustrated in
469 Figure 5. Note, however, that all calibration periods produce highly biased predictions and
470 that the differences between the calibration strategies are relatively limited compared to these
471 biases.

472

473 *FIGURE 6: Sensitivity of the simulated flow characteristics (from top to bottom: $Q_{95}$, $Q_{MA}$*
474 *and $Q_{05}$) on the dry validation sub-periods after calibration on climatically specific periods*
475 *(wet, mean, dry, total record) (left column: GR4J; right column: TOPMO). The Q-Q plots*

*show the observed versus simulated value for each catchment, each dot representing the mean of values simulated with the four optimal parameter sets and each bar representing the range of simulated values when using the four optimal parameter sets. The boxplots on the right represent the distributions of the relative errors on the flow characteristic on the dry validation sub-periods over the 89 catchments when considering the four calibration periods. The boxplots are constructed with the 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles.*

- **Impacts on simulated flow evolutions:**

Then we assessed the change in future simulated streamflow characteristics when considering the variability stemming from the climatic specificity of the calibration periods. Figure 7 shows the models' outputs on future time slices when using the sets of model parameters obtained on the four different calibration periods considered so far. The range of streamflow characteristics simulated for future time slices (mid-century (MC) and end-of-century (EC)) with the four parameter sets are plotted against those simulated for the present time slice (PT).

In the following, we assume that a model simulates a significant hydrological change on a catchment if the range bar is completely above or below the 1:1 line, meaning that the change can be considered beyond the variability generated by the climate specificity of the calibration periods considered. For example, a decreasing trend of $Q_{05}$ between PT and MC is assumed for a particular catchment if a model calibrated over the four different sub-periods simulates four $Q_{05}$ values lower in MC than in PT.

Considering the range across centres, the two models suggest a rather similar decreasing trend for the values of the three streamflow characteristics from PT to EC. This trend is not observed for the MC time slice. The $Q_{05}$ streamflow characteristic (high flows) increases for this time slice, before decreasing more sharply by the end of the century. Some catchments show particularly large range bars. An analysis of these catchments (not shown here) indicates

501 that the calibration performance is particularly poor for at least one calibration sub-period.
502 This shows that a model with poor performance in current conditions will add substantial
503 uncertainty to future predictions.

504

505 *FIGURE 7: Comparison of the simulations of three streamflow characteristics (from top to*
506 *bottom: $Q_{95}$, $Q_{MA}$ and $Q_{05}$) obtained on the present time slice (PT) and future time slices (MC*
507 *and EC) under projected climate conditions with the two hydrological models (left: GR4J;*
508 *right: TOPMO). The range bars represent, for each catchment, the range of estimated values*
509 *with the four optimal parameter sets corresponding to the four calibration periods.*

510

511 The sensitivity of the two models to the climatic specificity of the calibration periods is of the
512 same magnitude for all three time slices considered, meaning that the sensitivity to the
513 calibration periods is relatively stable in the future time slices. Figure 8 synthesises the results
514 of these trends (e.g. a decreasing trend of $Q_{05}$ between PT and MC is assumed for a particular
515 catchment when a model calibrated over the four different sub-periods simulates four $Q_{05}$
516 values lower in MC than in PT), showing the proportion of catchments where hydrological
517 trends between present (PT) and future (MC and EC) time slices have been simulated
518 considering different calibration sub-periods for the two hydrological models. It also
519 compares the information given by a hydrological model calibrated over a long period (here
520 the entire available record without the dry validation sub-periods) and the information given
521 using the four different calibration periods. These results confirm the previously obtained
522 general trend of a decrease in the three streamflow characteristic values from PT to EC, with a
523 particular increasing trend for high flows ($Q_{05}$) from PT to MC. Nevertheless, when
524 considering the four different calibration periods, a number of catchments show no clear
525 trends for the MC time slice, which attenuates the general trends highlighted when using the

whole record periods as the only calibration periods. Last, differences between the two models can be observed: when considering only the whole record for calibration, GR4J seems to simulate a more regionally homogeneous decrease in low- to medium-flow characteristics ($Q_{95}$ and $Q_{MA}$), since the percentage of catchments with a decrease in flow is larger than for TOPMO. Considering the four calibration periods, the two models yield more homogeneous simulations for the catchment set.

*FIGURE 8: Proportions of catchments showing (or not) hydrological trends between present (PT) and future (MC and EC) time slices considering different calibration sub-periods for the two hydrological models: white highlights a clear decrease, black highlights a clear increase and grey highlights no clear trend.*

### *4.3 Sensitivity to the use of a posterior ensemble of parameter sets*

In this section, the model outputs are analysed considering 2000 posterior parameter sets obtained on the whole record period for each catchment and for each model. First, we discuss the efficiency of these ensembles of posterior parameter sets on the dry validation sub-periods in terms of NSEsq and simulation of the three streamflow characteristics ($Q_{95}$, $Q_{MA}$ and $Q_{05}$). Second, we analyse the resulting variability of the simulated streamflow characteristics for the future climate conditions.

- **Efficiency on dry validation sub-periods:**

Figure 9 shows the distribution of NSEsq values obtained by the two models illustrating (i) the calibration performance of the optimal parameter sets over the dry validation sub-periods (i.e. 89 NSEsq values for each model) and (ii) the validation performance over the dry validation sub-periods using optimal parameter sets (i.e. 89 NSEsq values for each model) and posterior parameter sets (i.e. 178,000 NSEsq values for each model) identified on the whole

551 record periods without the dry validation sub-periods. For GR4J, the validation performance
552 obtained by posterior parameter sets is very similar to that produced by the individual optimal
553 parameter sets presented in Figure 5, meaning that for the catchments studied, the DREAM
554 algorithm produces posterior parameter sets yielding efficiency close to the value obtained by
555 optimal parameter sets over the dry validation sub-periods. The NSEsq performance
556 distributions obtained with optimal and posterior parameter sets are not similar for TOPMO:
557 the optimal parameter sets appear to be less efficient than the posterior parameter sets in terms
558 of NSEsq validation performance. This means that rather different optima exist when using
559 the GL function and a likelihood function based on a standard least squares errors scheme
560 (NSEsq here). Nevertheless, differences between optimal parameter set performance and
561 posterior parameter set performance are less significant in the validation step than in the
562 calibration step, as shown in Figure 4. It should be remembered that the NSEsq was not used
563 as an objective function.

564

565 *FIGURE 9: Distribution of NSEsq values obtained by the two models illustrating (i)*
566 *calibration performance of the optimal parameter sets over the dry-validation sub-periods*
567 *(black "OPT" boxplots) and (ii) validation performance over the dry validation sub-periods*
568 *using optimal (white "OPT" boxplots) and posterior (grey "POS" boxplots) parameter sets*
569 *identified on the whole record periods without the dry validation sub-periods. Results are*
570 *shown for GR4J (left) and TOPMO (right). The boxplots show the 0.10, 0.25, 0.50, 0.75 and*
571 *0.90.*

572

573 Figure 10 shows the variability of the models' outputs on the dry validation sub-periods when
574 considering the ensembles of posterior parameter sets instead of the single optimal parameter
575 set. This figure is organised like Figure 6. The dots represent the means of flow characteristics

on a catchment simulated with the 2000 posterior parameter sets and the bars represent the range of simulated flow characteristics on a catchment when considering its 2000 posterior parameter sets. The boxplots synthesise the distributions of the relative errors on the flow characteristics simulated by the models with the posterior parameter sets identified.

The first major result is that considering posterior parameter sets does not significantly increase the variability of the simulated flows. The biases between observed flows and flows simulated by the ensembles of posterior parameter sets (grey boxplots) are close to those obtained with the ensembles of optimal parameter sets (white boxplots). Moreover, this variability is very limited compared to the variability observed when considering the four climate-specific parameter sets (see Figure 6). Here, the hydrological responses associated with 2000 posterior parameter sets are similar to those associated with optimal parameter sets. The flow characteristics obtained by TOPMO present generally greater uncertainty than those obtained by GR4J. These predictive uncertainty values are again probably due to the larger number of free parameters for TOPMO. Note, however, that the predictive uncertainty for TOPMO is often consistent with the observed biases for the validation sub-periods since the estimation range often encompasses the observed flow value.

*FIGURE 10: Sensitivity of the simulated flow characteristics (from top to bottom: $Q_{95}$, $Q_{MA}$ and $Q_{05}$) on the dry validation sub-periods using the 2000 posterior parameter sets determined on the whole record periods without the dry validation sub-periods for the two hydrological models (left: GR4J; right: TOPMO). The Q-Q plots show the observed versus simulated value for each catchment, each dot representing the mean of simulated values when using the 2000 posterior parameter sets and each bar representing the range of simulated values when using the 2000 posterior parameter sets. The boxplots on the right represent the distributions of the relative errors on the flow characteristic on the dry validation sub-periods*

*over the 89 catchments when considering the 2000 posterior parameter sets. The boxplots are constructed with the 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles.*

- **Impacts on simulated flow evolutions:**

Figure 11 synthesises the results on the evolution of flows when considering the posterior parameter sets obtained throughout the whole record periods. This figure is organised like Figure 7: for each catchment, a range cross quantifies the variability in the estimation of flow characteristics for a time slice simulated by the posterior parameter sets obtained on the whole record periods. The results are very different from those obtained when considering only the individual optimal parameter sets for each of the four calibration periods (Figure 7). The variability of simulated flows considering 2000 posterior parameter sets for each catchment is much lower than the variability considering four climate-specific parameter sets for each catchment. Nevertheless, the variability in TOPMO outputs considering posterior parameter sets is higher than GR4J's variability.

*FIGURE 11: Comparison of the simulations of three streamflow characteristics (from top to bottom: $Q_{95}$, $Q_{MA}$ and $Q_{05}$) obtained on the present time slice (PT) and future time slices (MC and EC) under projected climate conditions with the two hydrological models (left: GR4J; right: TOPMO). For each catchment, the range bars represent the range of estimated values with the 2000 posterior parameter sets obtained over the whole record period.*

Figure 12 illustrates the proportion of catchments showing (or not showing) clear changes when considering the ensemble simulations obtained with the posterior parameter sets. The additional consideration of the ensembles of 2000 posterior parameter sets yields a slight increase in the number of catchments for which no clear trend is observed, particularly

between the MC and the PT. Nevertheless, the future trends are similar to those observed without taking into account the ensembles of posterior parameter sets, i.e. when using only the optimal parameter sets. There is a sharp decrease in all streamflow characteristics by the EC and a slight but significant increase in the high-flow characteristic for the MC.

*FIGURE 12: Proportion of catchments showing (or not showing) hydrological trends between present (PT) and future (MC and EC) time slices considering (or not considering) posterior parameter sets for the two hydrological models: white highlights a clear decrease, black highlights a clear increase and grey highlights no clear trend.*

## 5 DISCUSSION AND CONCLUSION

This paper attempted to investigate the uncertainty of hydrological predictions for the future climate when considering either (i) the dependence of the optimal parameter set on calibration period specificity or (ii) the use of several posterior parameter sets over a given calibration period. The first aspect often refers to the robustness of model parameters, while the second often refers to parameter uncertainty estimation based on Bayesian inference.

The two conceptual hydrological models tested here were sensitive to the use of climatically contrasted calibration sub-periods. This sensitivity was highlighted by a wide range of possible simulated streamflows for both the dry observed validation sub-periods and the future climate time slices. Even if general future changes can be observed when considering four optimal parameter sets (obtained with the calibration on three sub-periods and the whole record periods except the dry validation sub-periods) for each catchment, the proportion of catchments showing clear changes is much lower than when considering a unique parameter set (obtained by calibration on the whole record periods except the dry validation sub-periods). However, the impact of the calibration period climate specificity on the simulated

streamflows is not straightforward since for a majority of the catchments studied, using a wet calibration sub-period for a dry validation sub-period does not systematically generate a larger bias between observed and simulated flows than when using a dry calibration sub-period. Moreover, considering long periods for model calibration does not generate more robust simulation than using 3-year sub-periods, which is not consistent with the common belief that "the longer the calibration period, the more robust the parameter set". Since the use of two different hydrological models did not provide equivalent results, the relation between the model considered and the impact of the climatic specificity of the calibration period on calibration and validation performance should be further investigated.

Concerning the "classical" parameter uncertainty assessment followed in this study, it seems that the prediction bounds obtained from the ensembles of posterior parameter sets are considerably thinner than what would be expected, especially for the GR4J model. Nevertheless, it is important to note that these results are dependent to some extent on the method used (the DREAM algorithm (Vrugt et al., 2009) and the GL objective function (Schoups and Vrugt, 2010)), the catchments studied and the models considered. It appeared that DREAM provided posterior parameter sets that were close to the optimal ones in terms of Nash-Sutcliffe validation efficiency over the dry validation sub-periods. Other methods to quantify parameter uncertainty could produce posterior parameter sets with greater differences than the optimal ones and thus yield larger uncertainty bounds. Considering the ensembles of 2000 posterior parameter sets yields a slight increase in the number of catchments for which no clear trend is observed, especially for TOPMO. The results obtained by the two conceptual models were found to be relatively consistent. The main differences were the larger uncertainty bounds observed for TOPMO. This is probably attributable to the larger number of degrees of freedom of TOPMO, which has six free parameters, compared to the four free parameters of GR4J. TOPMO's calibrated parameters are thus likely to depend

676 more on the choice of the calibration period and the objective function used during the
677 optimisation process. Still, further research is needed to confirm these hypotheses.

678 Our results show that, given the evaluation approach followed here, model robustness was the
679 major source of variability in streamflow projections in future climate conditions. They
680 corroborate the previous findings of Vaze et al. (2010), Merz et al. (2011) and Coron et al.
681 (2012) obtained with different catchment sets and models, emphasising the lack of robustness
682 of conceptual rainfall-runoff models when the climatic context between calibration and
683 validation periods are different. Note that for these three studies, long-term regional non-
684 stationarities were observed on the catchments studied: southeastern Australian catchments
685 suffered from long drought periods while Austrian catchments experienced a significant
686 increase in temperature over the last few decades, generating a shift in the hydrological
687 regimes, particularly for snow-affected catchments. These situations allow testing the
688 hydrological models on long as well as significantly different sub-periods in terms of climatic
689 conditions. Even if these actual non-stationarities were not observed everywhere, it seems
690 possible to test the sensitivity of models' calibration on climatically contrasted sub-periods.

691 Thus, from these results, it seems difficult to provide general guidelines for calibrating
692 hydrological models for climate change studies. The robustness issue should be investigated
693 more thoroughly, by proposing and testing calibration procedures that increase this
694 robustness. For example, Coron et al. (2012) proposed the Generalized Split Sample Test
695 procedure, which aims at testing all possible combinations of calibration-validation periods
696 and thus studying the capability of the tested model to be used in different climatic contexts.
697 Other tests could be performed, inspired by the methodology defined in this work.

698 This study also stresses that hydrological models do not efficiently reproduce streamflow
699 characteristics, even if the NSEsq coefficient estimated after calibration is quite high. The
700 median bias obtained for mean annual flow was generally greater than 10%. This is a

considerable limitation for the use of hydrological models to simulate extreme high or low flows in a changing climate. To cope with this notable failure, one could suggest using multi-objective calibration procedures and/or adapting the objective function to the estimated flow characteristic.

## 6 ACKNOWLEDGMENTS

All hydro-meteorological data were provided by two research programs that addressed the potential impacts of climate change: the RexHySS project (Ducharne et al., 2009; Ducharne et al. 2011) on the Seine and the Somme basins and the ICC-Hydroqual (Moatar et al., 2010) project on the Loire basin.

The authors thank the reviewers who provided constructive comments on an earlier version of the manuscript, which helped clarify the text. Among them, Jasper Vrugt is thanked for providing the codes to implement the optimisation approach he advised. Finally, François Bourgin (IRSTEA) is also acknowledged for his comments and suggestions.

## 7 REFERENCES


Anctil, F., Perrin, C. and Andréassian, V. 2004. Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models. Environmental Modelling & Software 19, n°. 4: 357-368. doi: 10.1016/S1364-8152(03)00135-X.

Arnell, N.W., 1999a. The effect of climate change on hydrological regimes in Europe: a continental perspective. Global Environmental Change 9, 5–23. doi: 10.1016/S0959-3780(98)00015-6.

Arnell, N.W., 1999b. Climate change and global water resources. Global Environmental Change 9, 31–49. doi: 10.1016/S0959-3780(99)00017-5.

Arnell, N.W., 2011. Uncertainty in the relationship between climate forcing and hydrological response in UK catchments. Hydrology and Earth System Sciences. 15(3), 897–912. doi: 10.5194/hess-15-897-2011.

Bastola, S., Murphy, C., Sweeney, J., 2011. The role of hydrological modelling uncertainties in climate change impact assessments of Irish river catchments. Advances in Water Resources, 34(5), 562–576. doi: 10.1016/j.advwatres.2011.01.008.

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

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in Water Resources 16, 41–51. doi: 10.1016/0309-1708(93)90028-E.

Boé, J., Terray, L., Habets, F., Martin, E., 2006. A simple statistical–dynamical downscaling scheme based on weather types and conditional resampling. Journal of Geophysical Research. 111(D23), D23106. doi: 200610.1029/2005JD006889.

Booij, M.J., 2005. Impact of climate change on river flooding assessed with different spatial model resolutions. Journal of Hydrology 303, 176–198. doi: 10.1016/j.jhydrol.2004.07.013.

Chen, J., Brissette, F. P., Poulin, A. and Leconte, R. (2011), Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed, Water Resources. Research., 47, W12509. doi: 201110.1029/2011WR010602.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., and Hendrickx, F. 2012. Crash Testing Hydrological Models in Contrasted Climate Conditions: An Experiment on 216 Australian Catchments. Water Resources Research. doi: 10.1029/2011WR011721.

Donnelly–Makowecki, L.M., Moore, R.D., 1999. Hierarchical testing of three rainfall–runoff models in small forested catchments. Journal of Hydrology 219, 136–152. doi : 10.1016/S0022-1694(99)00056-6.

Ducharne, A., Habets, F., Déqué, M., Evaux, L., Hachour, A., Lepaillier, A., Lepelletier, T., Martin, E., Oudin, L., Pagé, C., Ribstein, P., Sauquet, E., Thiéry, D., Terray, L., Viennot, P., Boé, J., Bourqui, M., Crespi, O., Gascoin, S., Rieu, J., 2009. Impact du changement climatique sur les Ressources en eau et les Extrêmes Hydrologiques dans les bassins de la Seine et la Somme. Rapport final du projet RExHySS, Programme GICC, 62 pp (available at www.sisyphe.upmc.fr/~agnes/rexhyss/, in French).

Ducharne, A., Sauquet, E., Habets, F., Deque, M., Gascoin, S., Hachour, A., Martin, E., Oudin, L., Page, C., Terray, L., Thiery, D., Viennot, P, 2011. Evolution potentielle du

régime des crues de la Seine sous changement climatique. La Houille Blanche 1. 51-57. doi: 10.1051/lhb/2011006.

Franchini, M., Wendling, J., Obled, C., Todini, E., 1996. Physical interpretation and sensitivity analysis of the TOPMODEL. Journal of Hydrology 175, 293–338. doi : 10.1016/S0022-1694(96)80015-1.

Gan, T.Y., Burges, S.J., 1990. An assessment of a conceptual rainfall–runoff model's ability to represent the dynamics of small hypothetical catchments, 2: hydrologic responses for normal and extreme rainfall. Water Resources Research 26, 1605–1619.

Görgen, K., Beersma, J., Brahmer, G., Buiteveld, H., Carambia, M., de Keizer, O., Krahe, P., Nilson, E., Lammersen, R., Perrin, C. and Volken, D., 2010. Assessment of Climate Change Impacts on Discharge in the Rhine River Basin: Results of the RheinBlick2050 Project, CHR report, I–23, 229 pp., Lelystad, ISBN 978–90–70980–35–1.

IPCC, W.G.I., 2007. Climate change 2007: the physical science basis, 4th Assessment Report. Genève.

Jakeman, A.J., Chen, T.H., Post, D.A., Hornberger, G.M., Littlewood, I.G., Whitehead, P.G., 1993. Assessing uncertainties in hydrological response to climate at large scale. Macroscale modelling of the hydrosphere 214, 37–47.

Jiang, T., Chen, Y.D., Xu, C., Chen, X., Chen, X., Singh, V.P., 2007. Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang Basin, South China. Journal of Hydrology 336, 316–333. doi: 10.1016/j.jhydrol.2004.07.013.

Klemeš, V., 1986. Operational testing of hydrological simulation models. Hydrological Sciences Journal 31(1), 13-24. doi: 10.1080/02626668609491024.

Le Lay, M., Galle, S., Saulnier, G.M., Braud, I., 2007. Exploring the relationship between hydroclimatic stationarity and rainfall–runoff model parameter stability: A case study in West Africa. Water Resources Research 43, W07420. doi: 10.1029/2006WR005257.

Ludwig, R., I. May, R. Turcotte, L. Vescovi, M. Braun, J. F. Cyr, L. G. Fortin, et al., 2009. The role of hydrological model complexity and uncertainty in climate change impact assessment. Advances in Geosciences 21 (2009): 63–71. doi: 10.5194/adgeo-21-63-2009.

Luo, J., Wang E., Shen S., Zheng H., Zhang H., 2011. Effects of conditional parameterization on performance of rainfall-runoff model regarding hydrologic non-stationarity. Hydrological Processes, doi: 10.1002/hyp.8420.

Matott, L.S., Babendreier, J.E., Purucker, S.T., 2009. Evaluating uncertainty in integrated environmental models: A review of concepts and tools. Water Resources. Research. 45, W06421. doi: 200910.1029/2008WR007301.

Merz, R., Parajka, J., Blöschl, G., 2011. Time stability of catchment model parameters: Implications for climate impact analyses. Water Resources. Research. 47, W02531. doi : 201110.1029/2010WR009505.

Michel, C., Perrin, C., Andreassian, V., 2003. The exponential store: a correct formulation for rainfall–runoff modelling. Hydrological Sciences Journal 48(1), 109–124. doi: 10.1623/hysj.48.1.109.43484.

Moatar, F., Ducharne, A., Thiéry, D., Bustillo, V., Sauquet, E., Vidal, J.–P., 2010. La Loire à l'épreuve du changement climatique. Geosciences 12, 78–87.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I–A discussion of principles. Journal of Hydrology 10(3), 282–290. doi: 10.1016/0022-1694(70)90255-6.

Niel, H., Paturel, J.E., Servat, E., 2003. Study of parameter stability of a lumped hydrologic model in a context of climatic variability. Journal of Hydrology 278, 213–230. doi: 10.1016/S0022-1694(03)00158-6.

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., Michel, C., 2006a. Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. Journal of Hydrology 320, 62–83. doi: 10.1016/j.jhydrol.2005.07.016.

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., Michel, C. 2006b. Dynamic averaging of rainfall-runoff model simulations from complementary model parametrizations. Water Resources Research 42. doi: 200610.1029/2005WR004636.

Penman, H.L., 1948. Natural Evaporation from Open Water, Bare Soil and Grass. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 193, 120–145.

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. Journal of Hydrology 279(1-4), 275–289. doi: 10.1016/S0022-1694(03)00225-7.

Perrin, C., Oudin, L., Andreassian, V., Rojas–Serna, C., Michel, C., Mathevet, T., 2007. Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models. Hydrological Sciences Journal 52(1), 131. doi: 10.1623/hysj.52.1.131.

Poulin, A., Brissette, F., Leconte, R., Arsenault, R., Malo, J-S., 2011. Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. Journal of Hydrology 409. 626-636. doi: 10.1016/j.jhydrol.2011.08.057.

Prudhomme, C., Davies, H., 2009. Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 2: future climate. Climatic change 93, 197–222. doi: 10.1007/s10584-008-9461-6.

Quintana–Segui, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., Morel, S., 2008. Analysis of near–surface atmospheric variables: Validation of the SAFRAN analysis over France. Journal of Applied Meteorology and Climatology 47, 92–107. doi: 10.1175/2007JAMC1636.1.

Refsgaard, J.C., Knudsen, J., 1996. Operational validation and intercomparison of different types of hydrological models. Water Resources Research 32, 2189–2202. doi : 10.1029/96WR00896.

Refsgaard, J.C., Van der Sluijs, J.P., Brown, J., Van der Keur, P., 2006. A framework for dealing with uncertainty due to model structure error. Advances in Water Resources 29, 1586–1597. doi : 10.1016/j.advwatres.2005.11.013.

Rosero, E., Yang Z.L., Wagener T., Gulden L., Yatheendradas S. Niu G.Y., 2010. Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season. Journal of Geophysical Research 115. doi: 10.1029/2009JD012035.

Seibert, J., 2003. Reliability of model predictions outside calibration conditions. Nordic Hydrology 34, 477–492. doi : 10.2166/nh.2003.028.

Schoups, G, Vrugt J.A., 2010.A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Research 46. doi: 201010.1029/2009WR008933.

Seiller, G., Anctil, F. and Perrin, C., 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. Hydrology and Earth System Sciences, 16(4): 1171-1189. doi: 10.5194/hess-16-1171-2012.

Singh, R., Wagener T., van Werkhoven K., Mann M. E., Crane R., 2011. A trading-space-for-time approach to probabilistic continuous streamflow predictions in a changing

862 climate – accounting for changing watershed behavior. Hydrology and Earth System
863 Sciences 15 3591-3603. doi: 10.5194/hess-15-3591-2011.
864 Teng, J, Vaze, J., Chiew, F H. S., Wang, B. and Perraud, J.M.. 2012. Estimating the Relative
865 Uncertainties Sourced from GCMs and Hydrological Models in Modeling Climate
866 Change Impact on Runoff. Journal of Hydrometeorology 13 (1) 122–139.
867 doi:10.1175/JHM-D-11-058.1.
868 Vaze, J., Post, D.A., Chiew, F.H.S., Perraud, J.M., Viney, N.R., Teng, J., 2010. Climate non–
869 stationarity – Validity of calibrated rainfall–runoff models for use in climate change
870 studies. Journal of Hydrology, 394(3-4), 447-457. doi : 16/j.jhydrol.2010.09.018.
871 Vaze, J., A. Davidson, J. Teng, and G. Podger. 2011. Impact of Climate Change on Water
872 Availability in the Macquarie-Castlereagh River Basin in Australia. Hydrological
873 Processes 25 (16): 2597–2612. doi:10.1002/hyp.8030.
874 Vaze, J., and J. Teng. 2011. Future Climate and Runoff Projections Across New South Wales,
875 Australia: Results and Practical Applications. Hydrological Processes 25 (1): 18–35.
876 doi:10.1002/hyp.7812.
877 Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M. and Soubeyroux, J.-M. (2010), A 50-
878 year high-resolution atmospheric reanalysis over France with the Safran system.
879 International Journal of Climatology, 30: 1627–1644. doi: 10.1002/joc.2003.
880 Vos, N.J. de, Rientjes, T.H.M., Gupta, H.V., 2010. Diagnostic evaluation of conceptual
881 rainfall–runoff models using temporal clustering. Hydrological Processes, 24(20),
882 2840-2850. doi : 10.1002/hyp.7698.
883 Vrugt, J.A., Ter Braak C., Diks C., Robinson B., Hyman J.M., Higdon D., 2009. Accelerating
884 Markov chain Monte Carlo simulation by differential evolution with self-adaptive
885 randomized subspace sampling. International Journal of Nonlinear Sciences and
886 Numerical Simulation 10 (3): 273–290.
887 Wilby, R.L., Harris, I., 2006. A framework for assessing uncertainties in climate change
888 impacts: Low–flow scenarios for the River Thames, UK. Water Resources. Research.
889 42(2), W02419. doi: 200610.1029/2005WR004065
890 Wilby, R.L., 2005. Uncertainty in water resource model parameters used for climate change
891 impact assessment. Hydrological Processes 19(16), 3201–3219. doi:
892 10.1002/hyp.5819.
893 Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall–
894 runoff models: sensitivity to calibration data. Journal of Hydrology 181, 23–48. doi:
895 10.1016/0022-1694(95)02918-4.
896

## 8 FIGURES

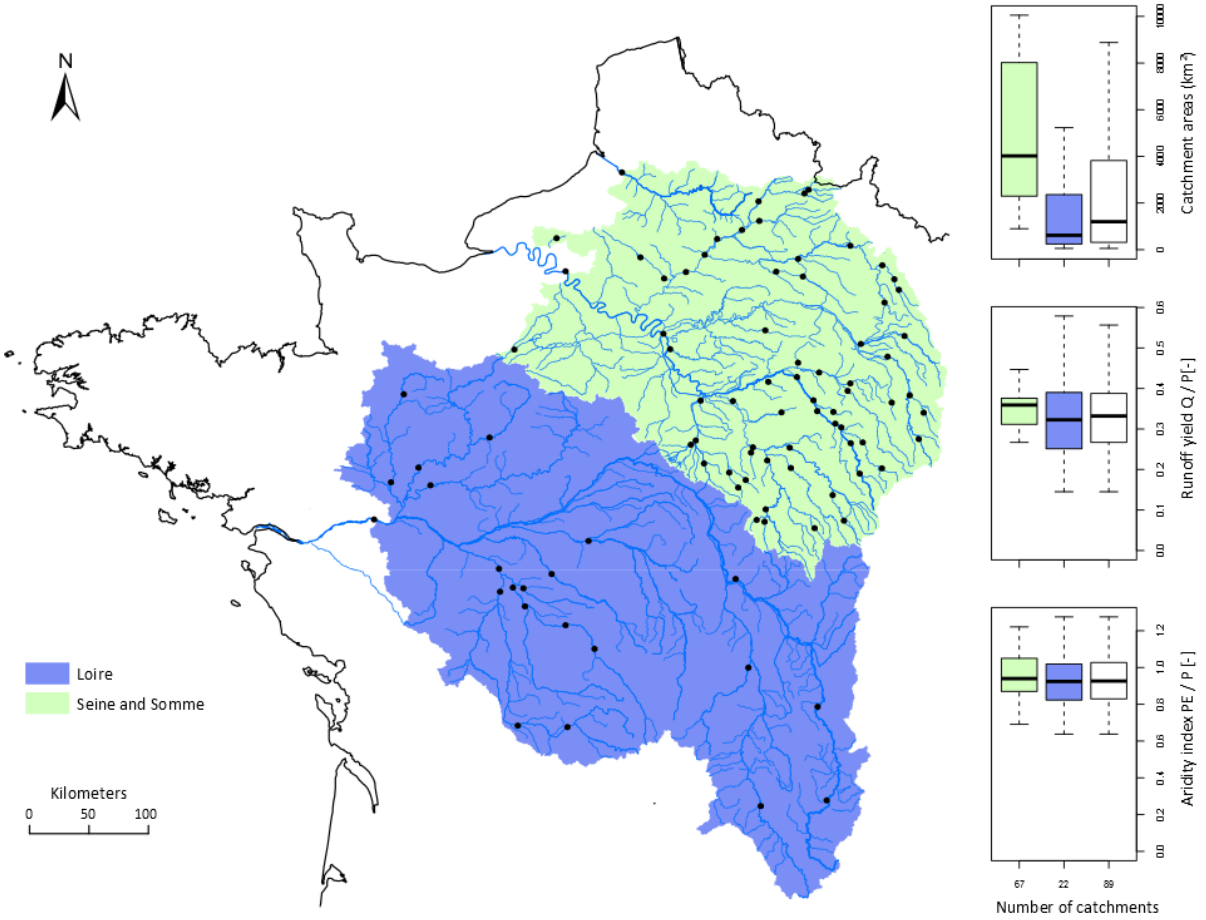


Fig. 1. Location and distribution of various characteristics of the 89 catchments used. The boxplots show the 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles (67 is the number of catchments in the Seine and Somme basins, 22 in the Loire basin).

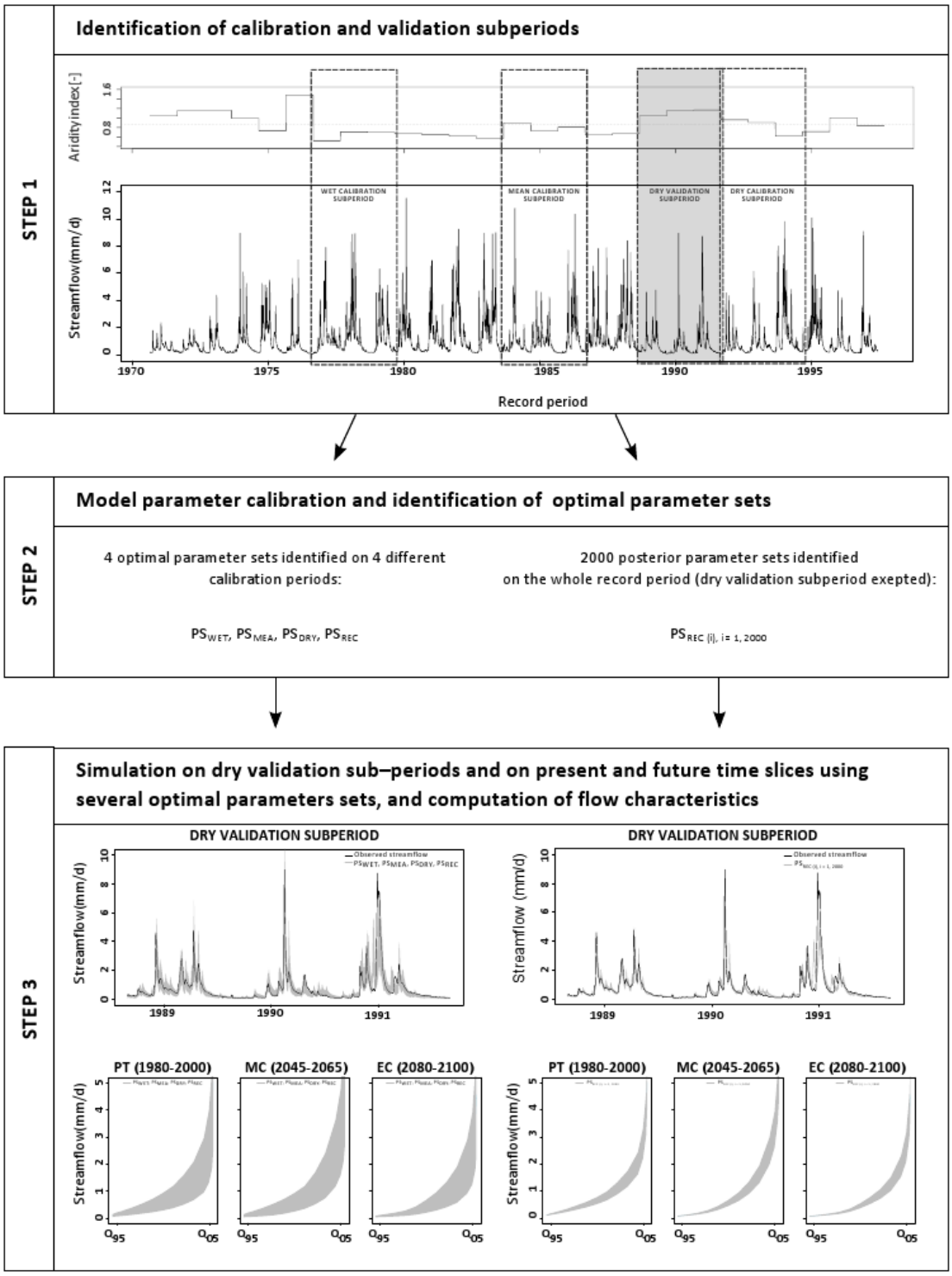


Fig. 2. Illustration of the three-step methodology used for investigating parameter uncertainty in a changing climate.

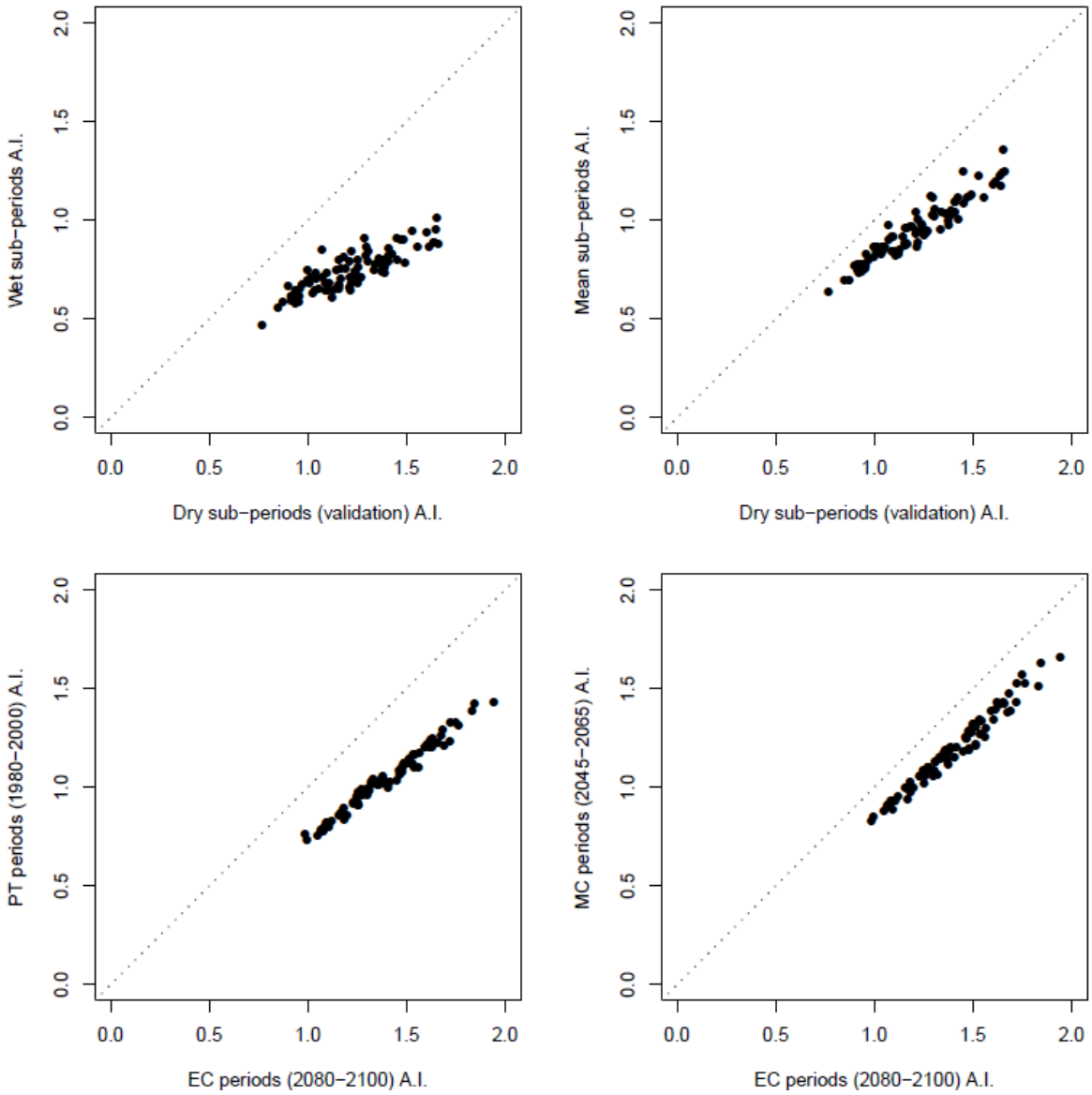


Fig. 3. Comparison of Aridity Index (AI) values for the different calibration and validation sub-periods considered and for the three time slices (PT, MC, EC) for the 89 catchments.

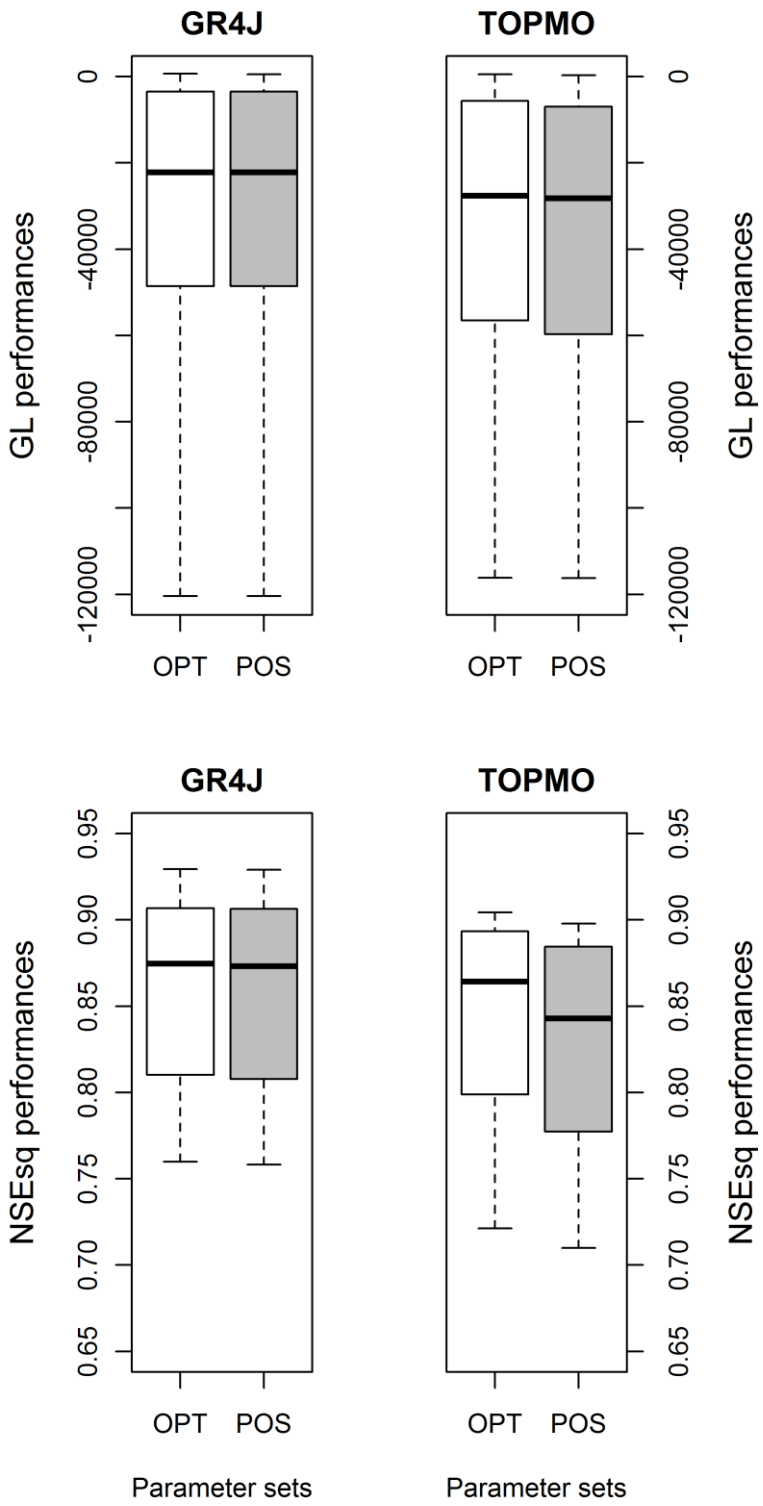


Fig. 4. Distributions of the GL objective function values (top) and of the NSEsq values (bottom) of the two models illustrating (i) calibration performance over the whole record periods without the dry validation sub-periods obtained with optimal parameter sets (white boxplots, noted OPT) and (ii) calibration performance over the whole record periods obtained with posterior parameter sets (grey boxplots, noted POS). Results are shown for GR4J (left) and TOPMO (right). The boxplots show the 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles.

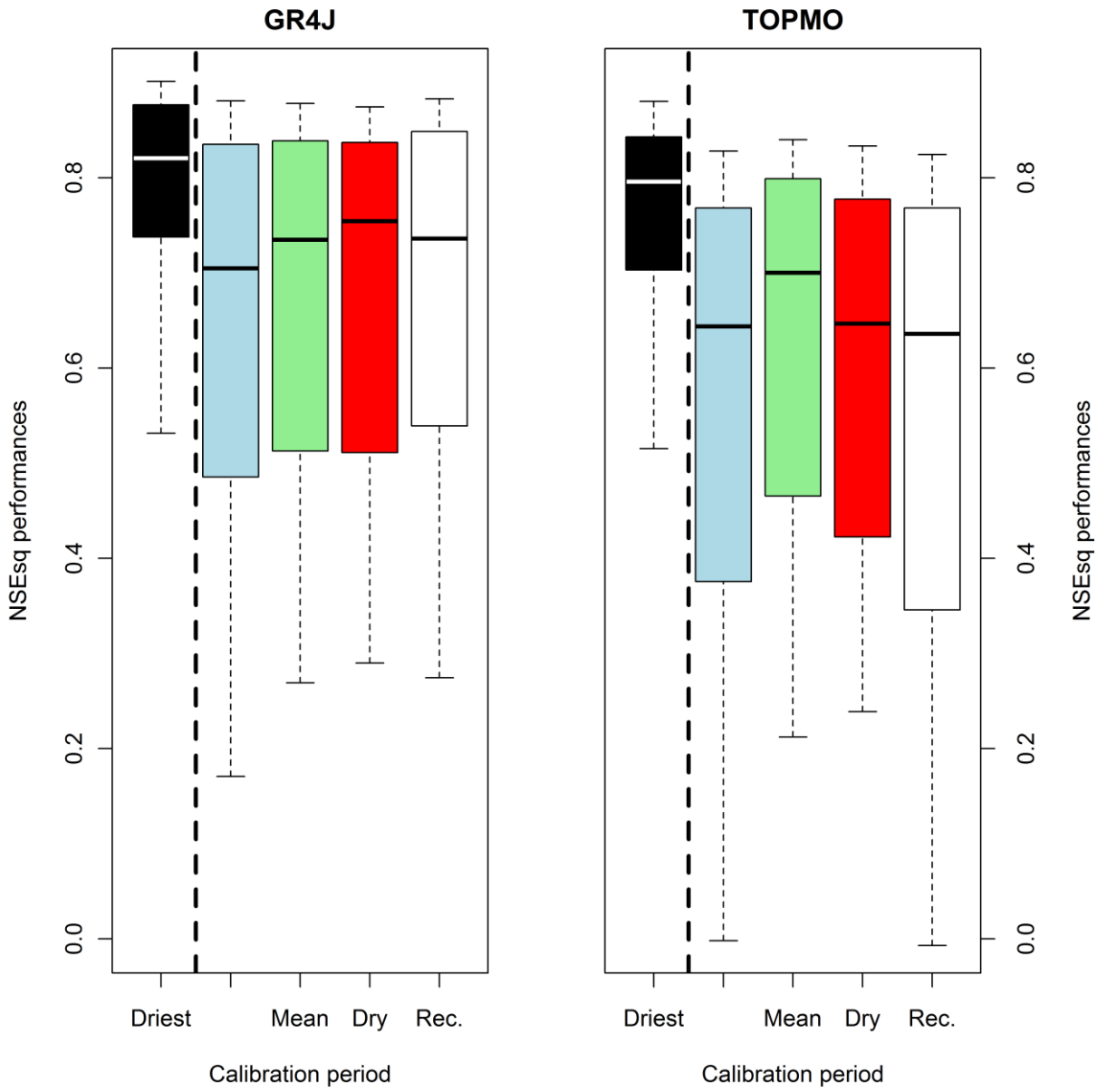


Fig. 5. Distributions of the NSEsq values obtained by the two models illustrating (i) calibration performance over the dry validation sub-periods (black boxplots) and (ii) validation performance over the dry validation sub-periods using the other four calibration sub-periods considered (wet, mean, dry, and whole record without the dry validation sub-period illustrated, respectively, with blue, green, red and white boxplots). Results are shown for GR4J (left) and TOPMO (right). The boxplots show the 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles.

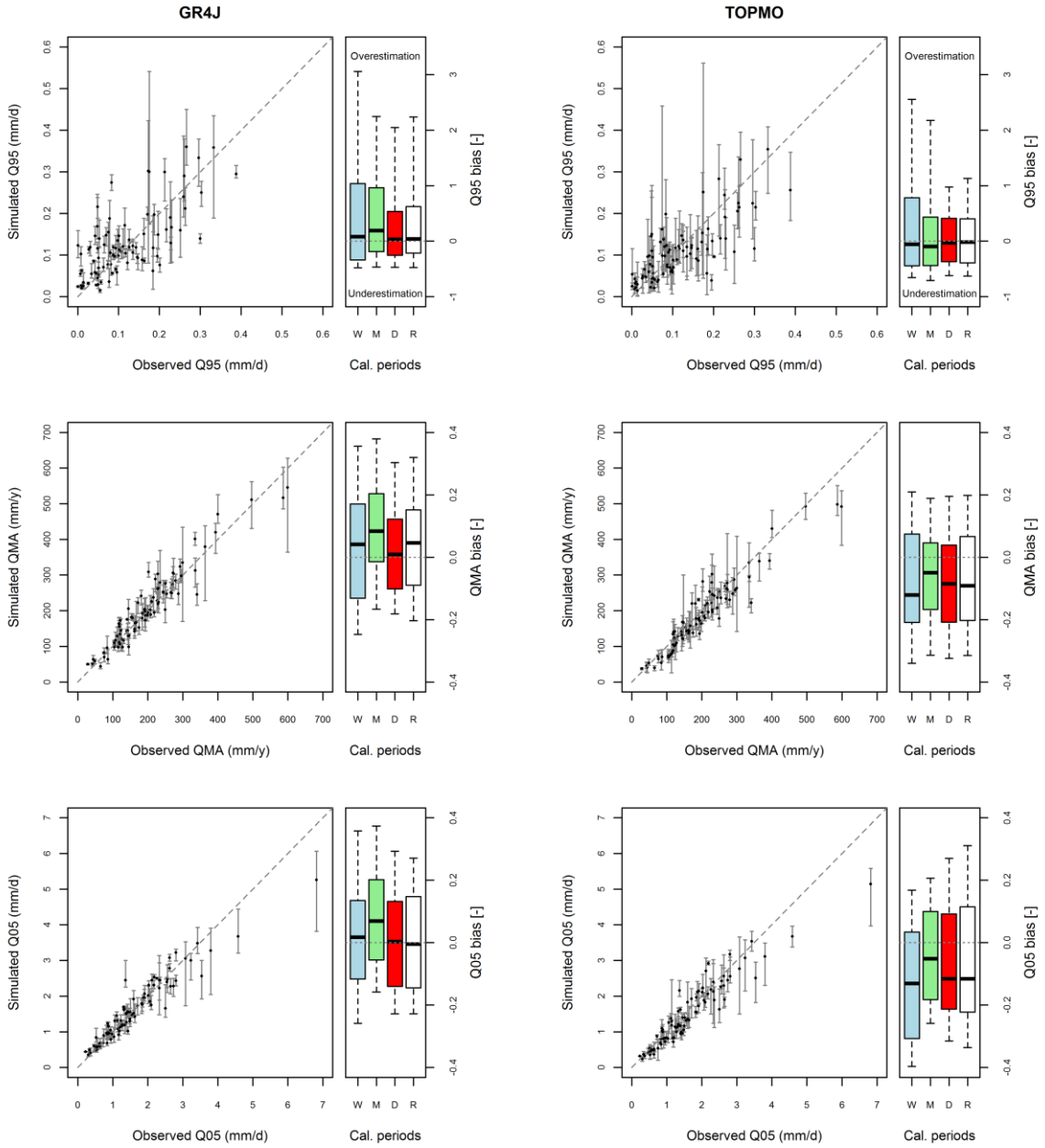


Fig. 6. Sensitivity of the simulated flow characteristics (from top to bottom: Q95, QMA and Q05) on the dry validation sub-periods after calibration on climatically specific periods (wet, mean, dry, total record) (left column: GR4J; right column: TOPMO). The Q-Q plots show the observed versus simulated value for each catchment, each dot representing the mean of values simulated with the four optimal parameter sets and each bar representing the range of simulated values when using the four optimal parameter sets. The boxplots on the right represent the distributions of the relative errors on the flow characteristic on the dry validation sub-periods over the 89 catchments when considering the four calibration periods. The boxplots are constructed with the 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles.

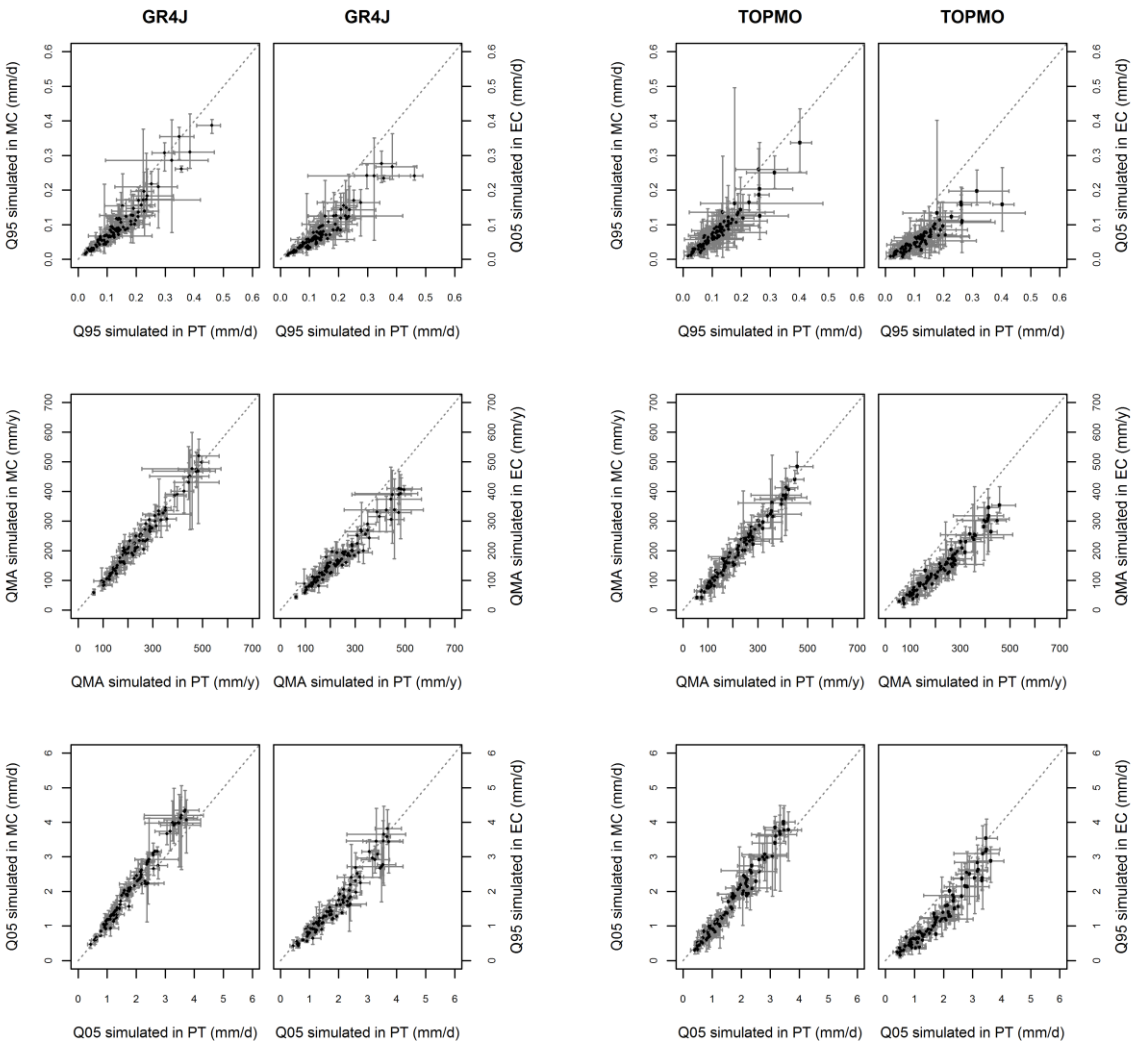


Fig. 7. Comparison of the simulations of three streamflow characteristics (from top to bottom: Q95, QMA and Q05) obtained on the present time slice (PT) and future time slices (MC and EC) under projected climate conditions with the two hydrological models (left: GR4J; right: TOPMO). The range bars represent, for each catchment, the range of estimated values with the four optimal parameter sets corresponding to the four calibration periods.

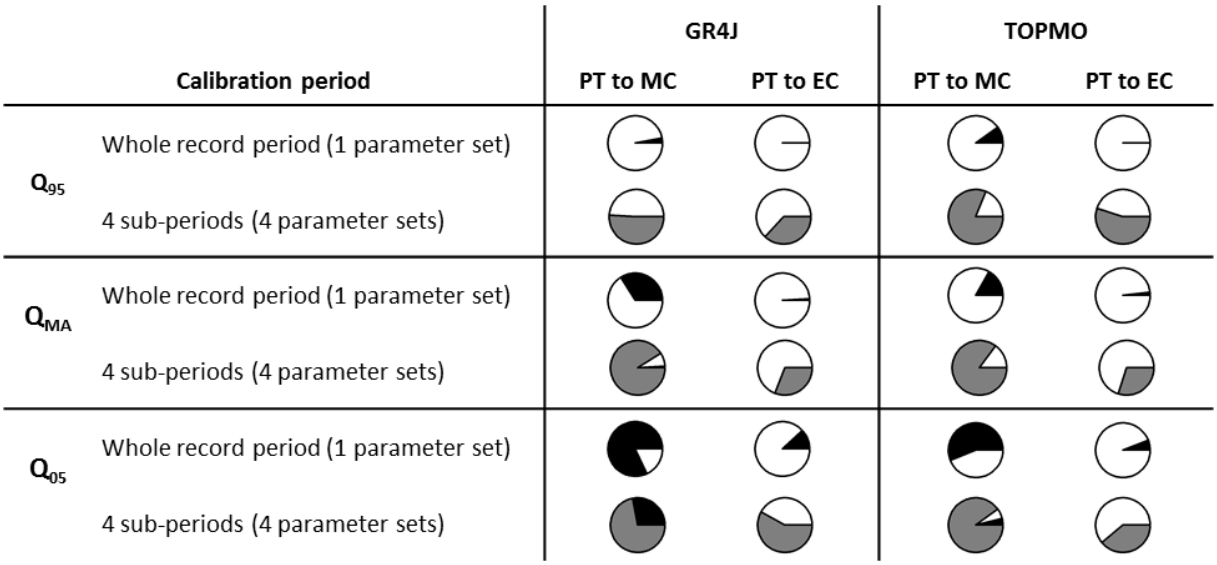


Fig. 8. Proportions of catchments showing (or not) hydrological trends between present (PT) and future (MC and EC) time slices considering different calibration sub-periods for the two hydrological models: white highlights a clear decrease, black highlights a clear increase and grey highlights no clear trend.

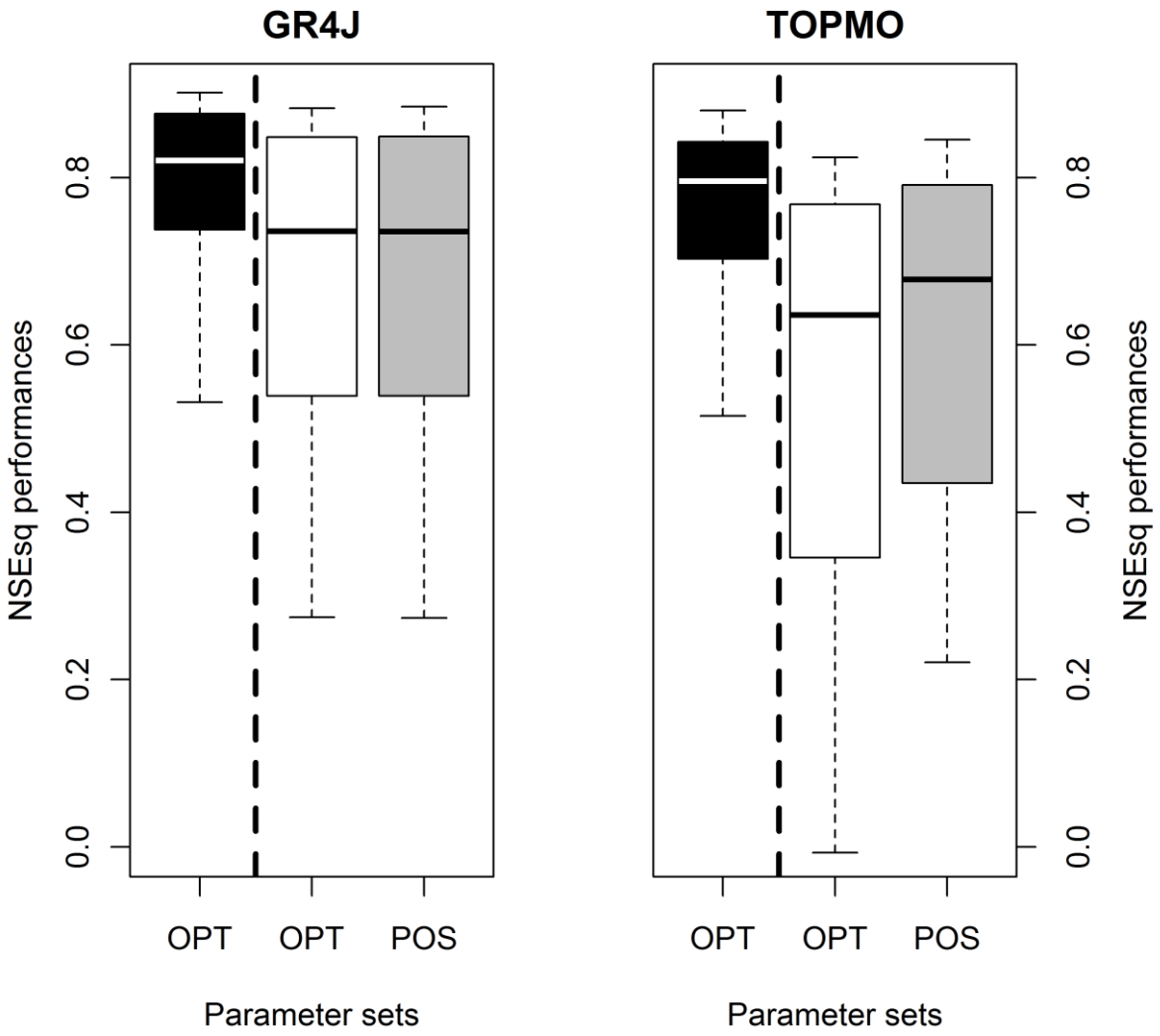


Fig. 9. Distribution of NSEsq values obtained by the two models illustrating (i) calibration performance of the optimal parameter sets over the dry-validation subperiods (black ‘‘OPT’’ boxplots) and (ii) validation performance over the dry validation sub-periods using optimal (white ‘‘OPT’’ boxplots) and posterior (grey ‘‘POS’’ boxplots) parameter sets identified on the whole record periods without the dry validation sub-periods. Results are shown for GR4J (left) and TOPMO (right). The boxplots show the 0.10, 0.25, 0.50, 0.75 and 0.90.

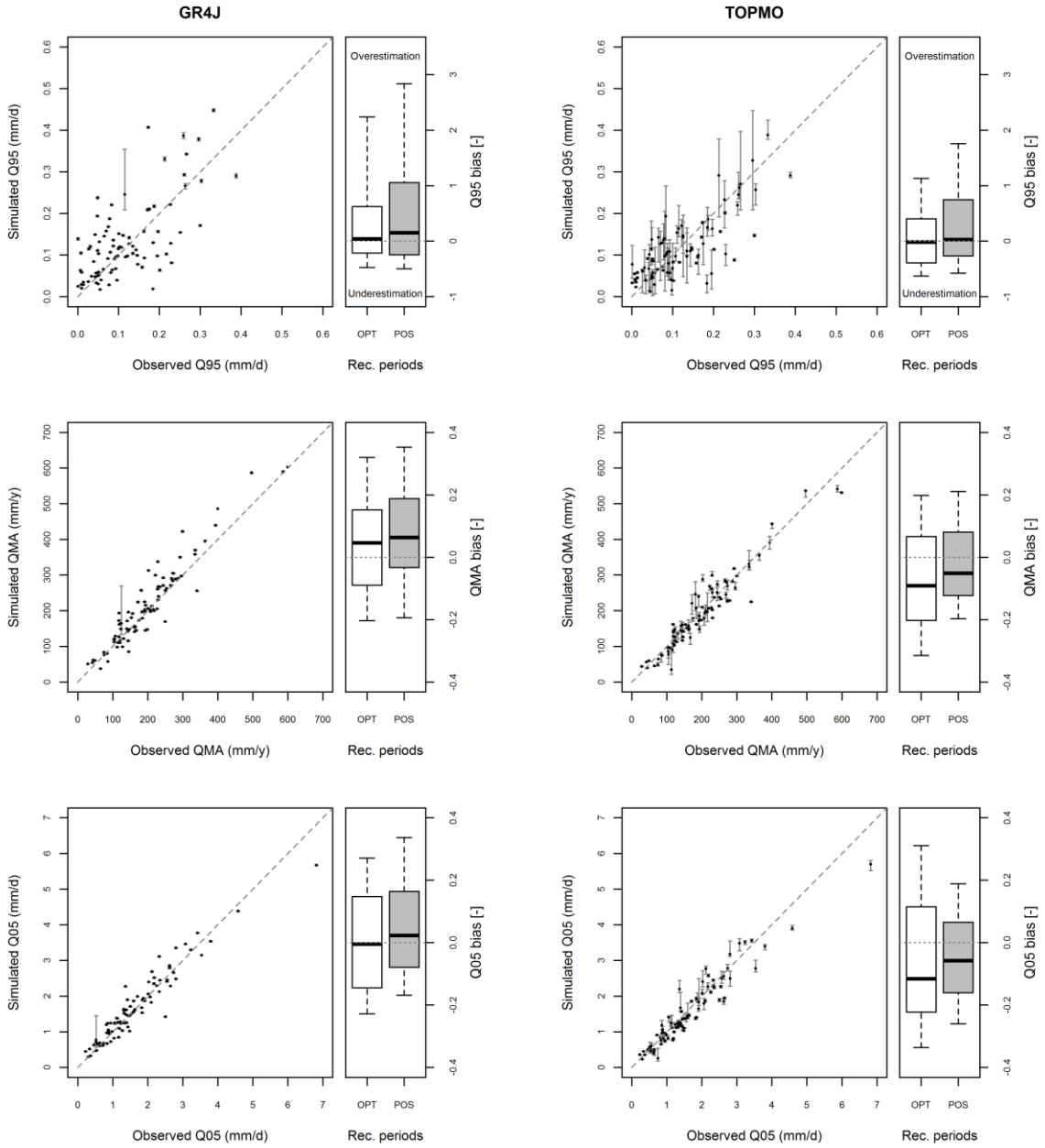


Fig. 10. Sensitivity of the simulated flow characteristics (from top to bottom: Q95, QMA and Q05) on the dry validation sub-periods using the 2000 posterior parameter sets determined on the whole record periods without the dry validation sub-periods for the two hydrological models (left: GR4J; right: TOPMO). The Q–Q plots show the observed versus simulated value for each catchment, each dot representing the mean of simulated values when using the 2000 posterior parameter sets and each bar representing the range of simulated values when using the 2000 posterior parameter sets. The boxplots on the right represent the distributions of the relative errors on the flow characteristic on the dry validation sub-periods over the 89 catchments when considering the 2000 posterior parameter sets. The boxplots are constructed with the 0.10, 0.25, 0.50, 0.75 and 0.90 percentiles.

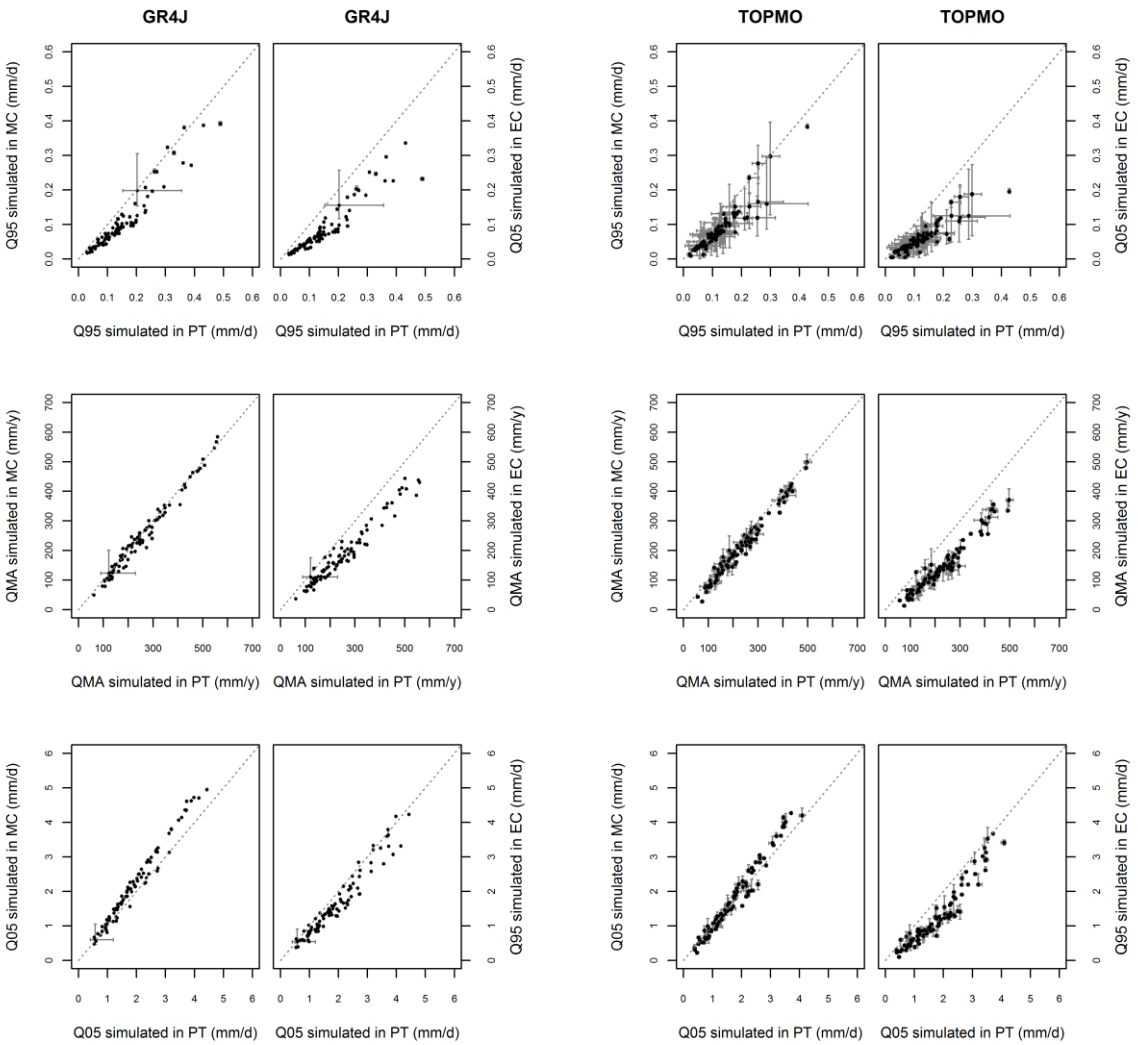


Fig. 11. Comparison of the simulations of three streamflow characteristics (from top to bottom: Q95, QMA and Q05) obtained on the present time slice (PT) and future time slices (MC and EC) under projected climate conditions with the two hydrological models (left: GR4J; right: TOPMO). For each catchment, the range bars represent the range of estimated values with the 2000 posterior parameter sets obtained over the whole record period.

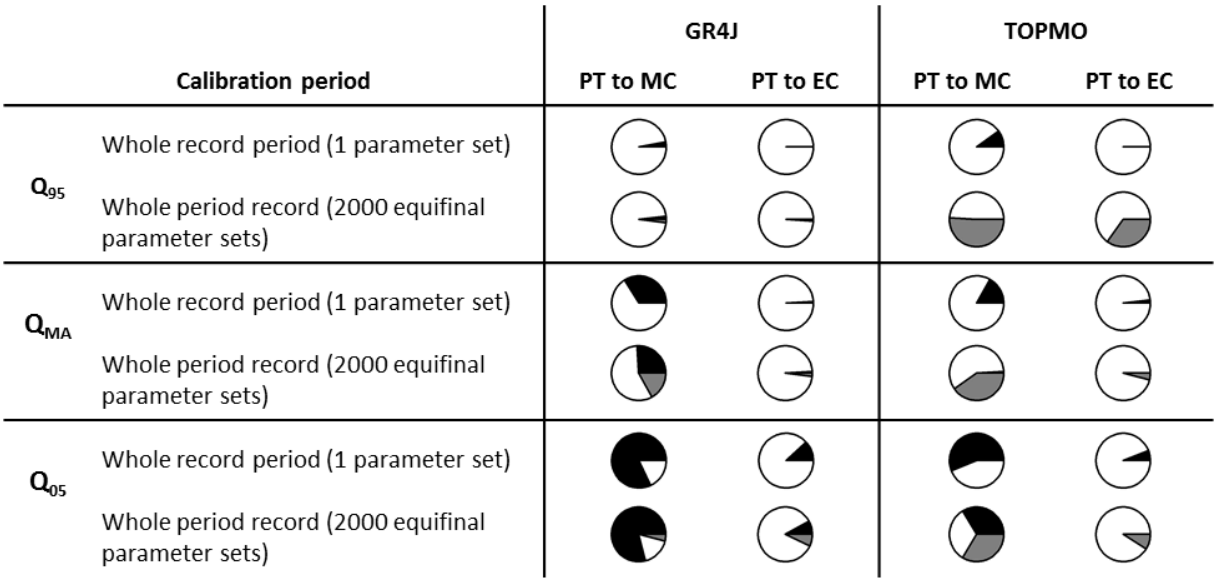


Fig. 12. Proportion of catchments showing (or not showing) hydrological trends between present (PT) and future (MC and EC) time slices considering (or not considering) posterior parameter sets for the two hydrological models: white highlights a clear decrease, black highlights a clear increase and grey highlights no clear trend.