Journal of Hydrology 411 (2011) 66–76

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

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

A downward structural sensitivity analysis of hydrological models to improve low-flow simulation

Raji Pushpalatha [a], Charles Perrin [a,*], Nicolas Le Moine [b], Thibault Mathevet [c], Vazken Andréassian [a]

[a] Cemagref, UR HBAN, 1, rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony Cedex, France
[b] Université Pierre et Marie Curie, UMR 7619 Sisyphe, 4 place Jussieu, 75252 Paris Cedex 05, France
[c] EDF–DTG, 21, avenue de l'Europe, BP 41, 38040 Grenoble Cedex 09, France

ARTICLE INFO

Article history:
Received 5 May 2011
Received in revised form 18 August 2011
Accepted 23 September 2011
Available online 29 September 2011

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Ezio Todini, Associate Editor

Keywords:
Low flows
Simulation
Lumped model
Model efficiency
Uncertainty
Downward approach

SUMMARY

Better simulation and earlier prediction of river low flows are needed for improved water management. Here, a top–down structural analysis to improve a hydrological model in a low-flow simulation perspective is presented. Starting from a simple but efficient rainfall–runoff model (GR5J), we analyse the sensitivity of low-flow simulations to progressive modifications of the model's structure. These modifications correspond to the introduction of more complex routing schemes and/or the addition of simple representations of groundwater–surface water exchanges. In these tests, we wished to improve low-flow simulation while avoiding performance losses in high-flow conditions, i.e. keeping a general model.

In a typical downward modelling perspective, over 60 versions of the model were tested on a large set of French catchments corresponding to various low-flow conditions, and performance was evaluated using criteria emphasising errors in low-flow conditions. The results indicate that several best performing structures yielded quite similar levels of efficiency. The addition of a new flow component to the routing part of the model yielded the most significant improvement. In spite of the close performance of several model structures, we conclude by proposing a modified model version of GR5J with a single additional parameter.

1. Introduction

1.1. Low flows and rainfall–runoff models

The occurrence of low flows is perhaps less spectacular than high flows, but low-flow consequences can be as costly, because they correspond to crucial periods in the functioning of both ecological and water management systems. For example, the cost of damage caused by the drought events in the years 1988–1989 in the United States was approximately US$40 billion, whereas the cost of the 1993 flood event was US$18–20 billion (Demuth, 2005). Thus, we consider that the simulation and advanced prediction of river low flows is an important challenge to improve low-flow management, both in the present climate and under the projected climate changes, which may well result in an increase in the occurrence of low-flow events (see e.g. Boé et al., 2009; Feyen and Dankers, 2009).

While a variety of lumped rainfall–runoff models are available to simulate streamflow irrespective of the flow conditions (see e.g. Singh and Frevert, 2002a,b), only a limited number of modelling studies focus on low-flow simulation. This study aims at identifying a generic model structure for improved low-flow simulation. Note that given the complexity of hydrological processes and the specificities of each catchment, some modellers have argued that model structures should be catchment-specific (e.g. Fenicia et al., 2008). However, we believe that before identifying catchment-specific models, the best possible general model that would include the representation of most of the dominant processes at work on catchments should be identified. This is the approach followed in this paper.

1.2. Specificities of the downward approach

To identify the general model structures that represent catchment behaviour, we followed a downward approach: a lumped representation of the catchment was used, in which only the main features of catchment hydrological behaviour are represented. This means that we did not attempt to build an explicit physical representation of the system but instead attempted to find the building blocks of the model that maximised modelling efficiency. The tests reported herein can be considered a structural sensitivity analysis. Some studies highlight the usefulness of sensitivity analysis for the improvement of hydrological models (see e.g. Andréassian et al., 2001; Oudin et al., 2006b; Tang et al., 2007; Bahremand and De Smedt, 2008; Ruelland et al., 2008). Other studies used sensitivity

* Corresponding author.
E-mail address: charles.perrin@cemagref.fr (C. Perrin).

0022-1694/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2011.09.034

analysis to better understand model behaviour with respect to inputs such as precipitation and potential evapotranspiration (Oudin et al., 2005a,b; Xu et al., 2006; Meselhe et al., 2009). Here we will focus on the sensitivity of low-flow simulation to the change in the components of the model structure responsible for low-flow simulation.

The main objective of this article is to analyse the extent to which a downward sensitivity analysis can help identify ways to improve low-flow simulation, while keeping the hydrological coherence in simulating the other parts of the flow regime. The downward search starts from a robust and parsimonious model structure. Then we will analyse how sensitive low-flow simulations are to the formulation of the model structure. This is done in trial-and-error mode, by testing many alternative model structures on a large set of catchments representing various physical and hydrometeorological conditions. The best candidate towards which our search converged is finally assessed in comparison with other model structures available in the literature.

1.3. A brief overview of low-flow modelling studies

The number of catchment modelling studies focusing on low-flow simulation using hydrological models is quite limited. One of the major problems with low-flow simulation is to account for surface water–groundwater interactions. During low-flow periods, water exchanges occur through the stream bed: the river may be fed by groundwater or, conversely, it may leak to feed the aquifer. Therefore, groundwater significantly influences low flows. A few studies that investigated these issues can be mentioned here. Fleckenstein et al. (2006) clearly mentioned the river–aquifer interactions and the significance of groundwater contribution during low-flow periods. Herron and Croke (2009) noted the improvement of lumped model predictions with the incorporation of groundwater exchange functions. The conclusions by Anderson et al. (2004) and Hughes (2004) also suggest that the model simulation efficiency can be improved by the addition of functions which represent the interaction between channel and aquifer flows. This is clearly shown in the study by Le Moine et al. (2007), who tested several options to account for inter-catchment groundwater flows using two rainfall–runoff models. Their results indicate that explicitly accounting for these groundwater fluxes significantly improves modelling efficiency.

Along with groundwater exchange functions, additional stores in the routing module can also enhance model performance, especially in the case of delayed flows (Wagener et al., 2004; Mathevet, 2005). Lang (2007) and Lang et al. (2008) analysed the performance of lumped models with respect to the addition of routing stores (to account for different water pathways underground) in an existing structure. Their study showed that some improvement can be achieved in the low-flow simulation, although they conclude that further work would be needed to improve lumped models for low-flow simulation. In a recent study, Kim et al. (2011) used the IHACRES-3S (3 Storage) model to evaluate the low-flow simulation together with the integration of base flow. The results showed a slight improvement in the model's performance, but they concluded that further studies are needed to obtain better low-flow simulation results. Last, Staudinger et al. (2011) analysed the sensitivity of recession simulation to various storage configurations on a snow dominated catchment in Norway within the FUSE framework. They conclude that the structural sensitivity is different in the winter and summer seasons, but that tests on a larger set of catchments are needed to get more general conclusions.

1.4. Scope of the paper

This article presents the end result of a long downward sensitivity analysis process that led to proposing an improved version of the GR4J catchment model (Perrin et al., 2003). Although our aim was to improve low-flow simulation specifically, we intended to find a generic solution, i.e. one that would improve low-flow representation without affecting the representation of high flows. This study builds on the previous studies by Mathevet (2005) and Le Moine (2008) who have already conducted tests to modify the existing model structures to improve modelling efficiency on a wide variety of catchments.

The next section discusses the data set and testing methodology. Then the results are presented and discussed before the concluding remarks.

2. Data set, models and methodology

This section presents the data set, models and testing methodology used for the analysis.

2.1. Data set

A set of 1000 catchments spread over France was used to test the model's generalisability (Andréassian et al., 2006). The database was built by Le Moine (2008). Continuous series of precipitation (P) and potential evapotranspiration (PE) were available for the 1970–2006 time period, providing a good variability of meteorological conditions, with quite severe drought periods (e.g. the years 1976, 1989–1991, 2003 and 2005). Meteorological data come from the SAFRAN reanalysis of Météo–France (Quintana-Segui et al., 2008; Vidal et al., 2010). Daily potential evapotranspiration was estimated using the formulation proposed by Oudin et al. (2005a) based on temperature and extra-terrestrial radiation. Streamflow (Q) data were extracted from the national HYDRO database. The length of the available flow record varies from one catchment to another, but at least 20 years of data were available on each selected catchment within the 1970–2006 period.

The variability of mean streamflow values can be expressed as a function of precipitation and PE. Fig. 1 plots the runoff coefficient (Q/P) as a function of the aridity index (P/PE) (see Mouelhi, 2003; Le Moine et al., 2007). It illustrates the variability of hydro-climatic conditions in the test catchments. As explained in detail by Le Moine et al. (2007), there are many catchments in this data set for which water losses are greater than PE (points lying below the line $y = 1 - 1/x$), which is an indication of leaky catchments. There are also catchments for which flow is greater than rainfall (points above the line $y = 1$), which mainly correspond to

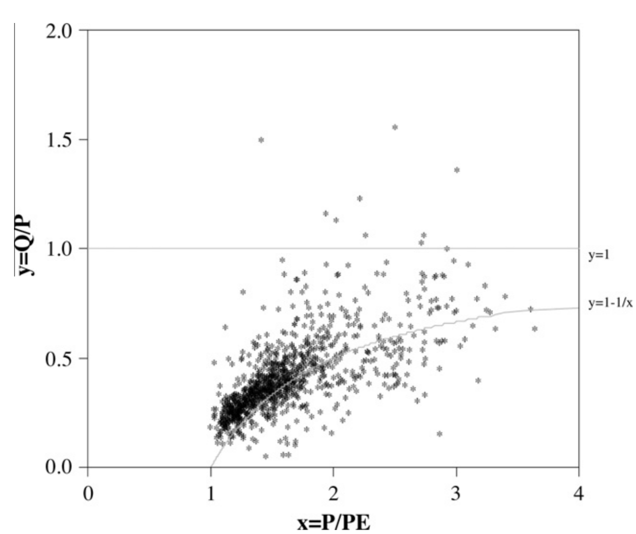

Fig. 1. Daily mean Q/P vs P/PE values for the 1000 catchments in the data set (P – rainfall; Q – streamflow; PE – potential evapotranspiration).

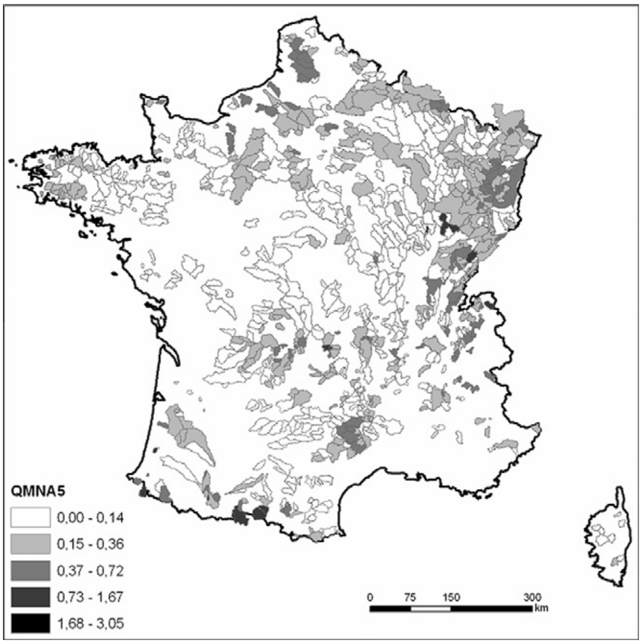

Fig. 2. Catchment location and illustration of the 5-year minimum monthly flow values.

catchments with karstic influences, i.e. those fed by inter-catchment groundwater flows from surrounding areas. Even though these catchments may prove more difficult to model, they were not discarded from the data set, as advocated by Andréassian et al. (2010).

Catchments with flow regulation structures (such as dams) were excluded from this data set. However, low flows may still be influenced by water withdrawals on some catchments: data on these influences were not available for this study. Given the size of our data set, the quality of flow data retrieved from the HYDRO database was trusted and not further checked in this study.

Catchment locations are shown in Fig. 2, along with the value of the minimum monthly flow of 5-year return period (called QMNA5 in France). QMNA5 is highly variable in this data set. It is influenced by various catchment characteristics, such as soil type, vegetation cover, geology and climatic conditions.

2.2. Tested models

The starting point of the present study was the GR4J rainfall–runoff model (Perrin et al., 2003), a lumped four-parameter model (see diagram in Fig. 3). It was already tested in various conditions with good results compared to other model structures. The water balance function that controls water balance in the GR4J model structure consists of a soil moisture accounting (SMA) reservoir (level S) and a conceptual water exchange function (F), expressed as:

$$F = X2 \cdot \left(\frac{R}{X3}\right)^{3.5} \tag{1}$$

in which $X2$ (mm) is the "groundwater" exchange coefficient and R and $X3$ (mm) are the water level and the capacity of the routing store, respectively. $X2$ can be positive or negative, meaning that the water exchange function can simulate imports or exports of water with the underground (i.e. connections with deep aquifers or surrounding catchments). Note that $X3$ is also used to parameterize the outflow from the routing store, which limits the interactions that would unavoidably exist between $X2$ and $X3$ if Eq. (1) was used alone. The routing part of the structure consists in two flow components routed by two unit hydrographs and a non-linear store. The latter is mainly responsible for low-flow simulations, along with leakage (percolation) from the SMA store. The groundwater exchange term F is added to the two flow components of the routing module.

Mathevet (2005) tested several modified versions of this model, especially by increasing the complexity of the routing part of the model and adding stores in parallel to the existing one. His tests, made at the hourly time step, showed limited sensitivity of model results, but the criteria he used focused more on high flows.

Following this work, Le Moine (2008) investigated the interactions between surface and groundwater and evaluated several modifications of the GR4J model to better account for these exchanges. These included different water exchange functions and the addition of a new store representing long-term memory. He proposed a five-parameter version of the model (GR5J) in which the groundwater exchange function has been modified (Fig. 3) to:

$$F = X2 \cdot \left(\frac{R}{X3} - X5\right) \tag{2}$$

where $X5$ is a dimensionless threshold parameter. It allows a change in the direction of the groundwater exchange within the year depending on the water level R in the routing store compared to this threshold. This model has shown significant performance improvement over the GR4J model, especially in low-flow conditions. It can be noted that the time-varying term F is only a very crude way to simulate groundwater–surface water connections. $X5$ can be seen as the external, quasi-stationary potential of the groundwater system and F is a "restoring flux" acting like a spring device with constant $X2$. Usually, $X2$ is negative: the more $R/X3$ departs from $X5$, the more intense the flux is, which tends to restore its value to $X5$.

Based on these previous results, the GR5J model's structure was used as a benchmark in our tests. In the subsequent sensitivity analysis, we will evaluate the extent to which modifications of the components used in the model to simulate low flows have an impact on model performance.

2.3. Model testing and assessment

The split sample testing scheme proposed by Klemes (1986) was used to evaluate model performance. For each catchment, the period where rainfall, PE and flow data were available (at most 1970–2006) was split into two halves (P1 and P2) of similar length, alternatively used for model calibration and validation. It means that for each catchment and each tested model, two calibrations and two validations were systematically performed. The first year of each test period was used for model warm-up. To avoid initialisation problems for catchments with long-term memories, five years of warm-up were considered in addition to the 1-year warm-up period: they were either the five years of observed data preceding the test period when available, or a mean year repeated five times otherwise. In this study, only performance in validation was considered to evaluate models.

The parameters were calibrated using a mean square model error calculated on root squared transformed flows as the objective function. This was found by Oudin et al. (2006a) to be a good compromise between high and low flows for model calibration. Here, as we focus on low flows, we could have chosen an objective function putting more weight on low flows. However, it would have been to the detriment of the simulation of high flows. So we preferred to keep this objective function to obtain a general model. This did not prevent us from assessing the model (in validation) over a wider range of criteria.

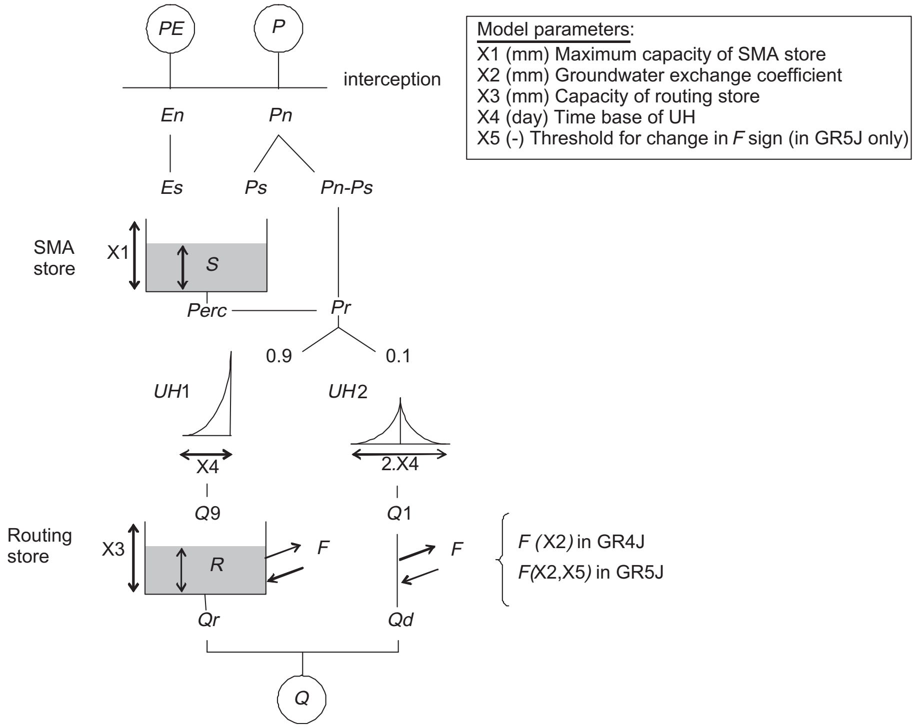

Fig. 3. Structure of the GR4J and GR5J models.

Several criteria based on the Nash and Sutcliffe (1970) efficiency index (*NSE*) were used to evaluate model performance in validation. *NSE* is given by:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^{n}(Q_{obs,i} - \overline{Q_{obs}})^2} = 1 - \frac{E}{E_0} \tag{3}$$

where n is the number of time steps, $Q_{obs,i}$ and $Q_{sim,i}$ are the observed and simulated flows, respectively, at time step i. $\overline{Q_{obs}}$ is the mean of the observed flows over the selected period. E and E_0 are the mean squared error and the variance of observed flows respectively. The *NSE* index takes values over the range $]-\infty; 1]$, 1 indicating perfect simulation and 0 indicating a simulation equivalent to a constant flow equal to the mean observed flow.

As *NSE* has no lower bound, a bounded formulation of *NSE* was preferred (here noted NSE^*) to avoid the influence of strongly negative values while calculating the mean of the model performance over the test catchments (see Mathevet et al., 2006 for more details). NSE^* is derived from *NSE* using the following relationship:

$$NSE^* = \frac{1 - E/E_0}{1 + E/E_0} \tag{4}$$

NSE^* values vary over the range $]-1; 1]$. When $NSE = 1$ (i.e. $E = 0$), $NSE^* = 1$, and when $NSE = 0$ (i.e. $E = E_0$), $NSE^* = 0$, hence the interpretation of the two criteria is similar. Note that for $NSE > 0$, NSE^* values will be lower than *NSE* values and the reverse for $NSE < 0$.

The criterion on natural flows (NSE^*_Q) was used to check simulation consistency in high-flow conditions. The efficiency criteria calculated on logarithm transformed flows (NSE^*_{lnQ}) and inverse transformed flows (NSE^*_{iQ}) were used to put more weight on low-flow simulation. These prior transformations on flows are of the Box–Cox type. Pushpalatha et al. (submitted for publication) analyse the effect of such power transformations on *NSE* efficiency criteria and investigate which transformation seems more relevant to evaluate the efficiency in low-flow conditions. They found that an inverse transformation puts more weight on the 20% of lowest flows on average.

The overall performance of the tested models was computed over the 1000-catchment set, either using the mean value or distribution of performance criteria obtained in validation (i.e. a total of 2×1000 values).

Performance differences with the reference models were also quantified using the relative efficiency index initially suggested by Nash and Sutcliffe (1970) and more recently advocated by Seibert (2001) and Lerat (2009). This is a generalised form of the *NSE* criterion. It compares the performance of the tested model with respect to the performance of a benchmark model structure, by the following equation:

$$RE_{(sim/bench)} = 1 - \frac{\sum_{i=1}^{n}(Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^{n}(Q_{obs,i} - Q_{bench,i})^2} = 1 - \frac{E}{E_1} \tag{5}$$

where $Q_{bench,i}$ is the flow simulated by the benchmark, at time step i, and E_1 is the mean squared error of the benchmark model. Here GR5J was used as the benchmark in all the test cases. Like *NSE*, *RE* can be written under a bounded form (RE^*) using the same transformation as in Eq. (4) (here substituting E_0 by E_1 in Eq. (4)) and can be calculated on transformed flows, depending on the range of flows targeted in the analysis.

2.4. Structural sensitivity analysis

A sensitivity analysis of low-flow simulation to the formulation of the model structure was performed. We systematically evaluated various modifications of the GR5J model. Since it is difficult to detail all the tests, the following sections present the two main types of modifications that were performed, namely modifications of the groundwater exchange function and the routing component.

2.4.1. Sensitivity to changes in the groundwater exchange function

Groundwater (GW) is the main source for river flows during prolonged dry periods. Hence the recharge and release of groundwater is one of the important processes to consider for simulating low flows. Some authors considered only the flow towards the stream using a specific groundwater reservoir (Davison and van der Kamp, 2008), but in the present analysis, we considered an exchange function that can account for both recharge and discharge from the groundwater reservoir, as in Le Moine (2008). During the course of this research, we evaluated the sensitivity of low-flow simulation to various formulations of the existing GW exchange functions.

2.4.2. Sensitivity to the addition of new stores

Stores and the empirical rules governing the transfer of water between them are the main components of rainfall–runoff models. Because of the complexity of the rainfall–runoff transformation, additional stores may improve model performance (Wagener et al., 2004; Mathevet, 2005). For low flows, this may provide additional components corresponding to different flow pathways.

In the GR5J model, a single routing store exists (*R*). A percolation from the soil moisture store also feeds flows during low-flow periods. As suggested by Mathevet (2005) and Le Moine (2008), we considered the parallel addition of new stores to the initial store, with various options to split effective rainfall into the different flow components. We also tested the serial addition of stores, as proposed by Lang (2007).

3. Results and discussion

In the following, we present the main results and discuss how sensitive low-flow simulations are to model formulation, following the modifications presented above. The selected versions of GR5J and their formulations are briefly presented in Table 1. As the number of modifications is almost infinite, we chose to present only a few of them to answer a number of simple questions that may arise when discussing the model's structure. Although these questions are sometimes interrelated, they are presented in sequence for the sake of clarity.

3.1. Can we design an improved model for low-flow simulation?

3.1.1. Can the existing groundwater exchange term in GR5J be improved?

We evaluated the sensitivity of low-flow simulation to various formulations of the groundwater exchange function. Starting from the GR5J model, several model versions that differ only by their groundwater exchange formulation were tested. In Table 1, three examples of modifications are provided:

– in M1, we gave seasonal dynamics to the exchanges by making them dependent on the SMA store and not on the routing store;
– in M2, we applied the splitting coefficient of flow components (0.1/0.9, as in Fig. 3) to the exchanges;
– in M3, we applied the formulation proposed by Nascimento (1995), i.e. making the exchanges a function of the level of the two stores, depending on the direction of the exchanges.

Fig. 4 shows the distribution of the performance of the selected versions, indicating significant sensitivity of the model's results to this function, which corroborates the findings of Le Moine et al. (2007). The existing exchange function in the base model appears to provide the best performance. This is in agreement with the results of Le Moine (2008) who had selected this function as the best performing among several other options.

3.1.2. Should the volumetric splitting between flow components be adapted to each catchment?

In GR4J and GR5J, 90% of the total effective rainfall is routed by the non-linear store (see Fig. 3). This volumetric proportion is fixed in the model for any catchment since Edijatno et al. (1999) showed that optimising it did not significantly improve the mean results. One factor to be considered when adding new stores to the initial store is the splitting coefficient of effective rainfall (*SC* in Fig. 5) between the stores. Mathevet (2005) and Le Moine (2008) conducted trials to divide effective rainfall between the existing store and an additional store. Their results tend to confirm that it is difficult to consider *SC* a free parameter. Here we simultaneously tested two model versions with two stores, one in which *SC* (version M4 in Table 1) was optimised on each catchment and the other in which *SC* was set at 0.4 (version M5).

The parameter analysis in M4 shows that *SC* values are very sensitive to the calibration conditions. The *SC* values obtained on the two test periods (P1 and P2) shows that this parameter is poorly defined and it will be difficult to relate it to catchment characteristics. The limited difference in model efficiency between the M4 and M5 versions (see Table 2) shows that *SC* can be set without significant efficiency loss. In the upcoming sections, we test versions considering only fixed splitting coefficients.

3.1.3. Should a new serial or parallel store be added?

Existing models propose a variety of conceptualisations for flow routing, using serial and/or parallel stores. Jakeman et al. (1990) discussed this issue in the IHACRES model, in which the routing

Table 1
Modified versions of the GR5J model and their main characteristics.

Model version	Characteristics of the groundwater exchange function		Characteristics of the additional routing stores					Number of routing stores	Number of free parameters
	Eq. (2)	Others	Power-2 store	Power-5 store	Exponential store	Added in parallel	Added in series		
M1	✓	Exchange dependent on SMA store						1	5
M2	✓	Splitting coefficient applied to F						1	5
M3		Formulation of Nascimento (1995)						1	4
M4	✓			✓		✓		2	7
M5	✓			✓		✓		2	6
M6	✓			✓			✓	2	6
M7	✓		✓			✓		2	6
M8	✓				✓	✓		2	6
M9	✓			✓	✓	✓		3	7
M10	✓				✓ ✓	✓		3	7
M11	✓		✓		✓	✓		3	7
M12	✓				✓	✓		2	6

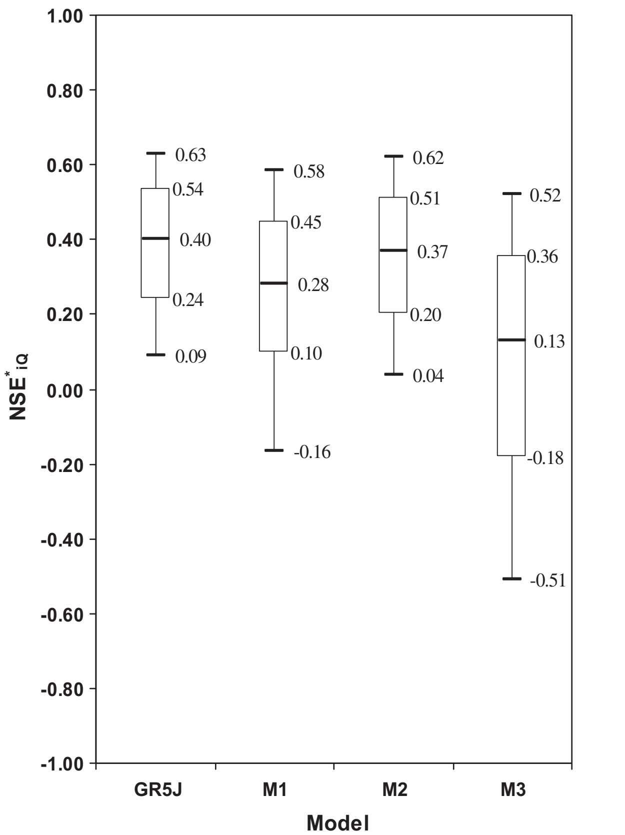

Fig. 4. Box plots of NSE^{*}_{iQ} values obtained in validation over the catchment set by GR5J and three model versions with modified groundwater exchange functions (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles).

module is made of linear stores. This model structure can be adapted to obtain several serial or parallel stores. Despite this flexibility, the authors indicate that in most cases, having two parallel stores is the most efficient configuration.

Here we analysed the sensitivity of the model's performance to the arrangement of routing stores, be they added to the existing parallel or serial stores. Two versions were tested, in which a new parallel store similar to the existing one was added (version M5, see Fig. 5) or a new serial store (version M6). Table 3 shows the mean performance of these two versions. The M5 version reaches higher efficiency values than M6. We also tried to add one more parallel routing store to obtain a third routed flow component (versions M9–M11 in Table 1). The results presented in Table 4 indicate that the improvements for low-flow simulation are not significant, which means that this additional complexity is not warranted by the data.

This confirms that the best compromise on average is to have two parallel stores. The series arrangement did not prove to be an efficient option. Therefore, following Jakeman et al. (1990), we suggest that the complexity of the routing part of the model should be increased by considering two independent flow components. This is a solution that provides more varied flow dynamics.

3.1.4. Does the formulation of the routing stores matter?

Here we tried to identify the best formulation of routing stores, i.e. the solution for which the model shows higher efficiency values. There is a variety of possible formulations of routing stores, ranging from linear to non-linear stores, e.g. power law or exponential stores (see Michel et al., 2003, for a good formulation of this store). In previous studies (see Edijatno and Michel, 1989; Edijatno et al., 1999), a power-5 non-linear routing store was identified as the most efficient. When adding a new parallel store, another formulation may be interesting to introduce a variety of behaviours in the flow components.

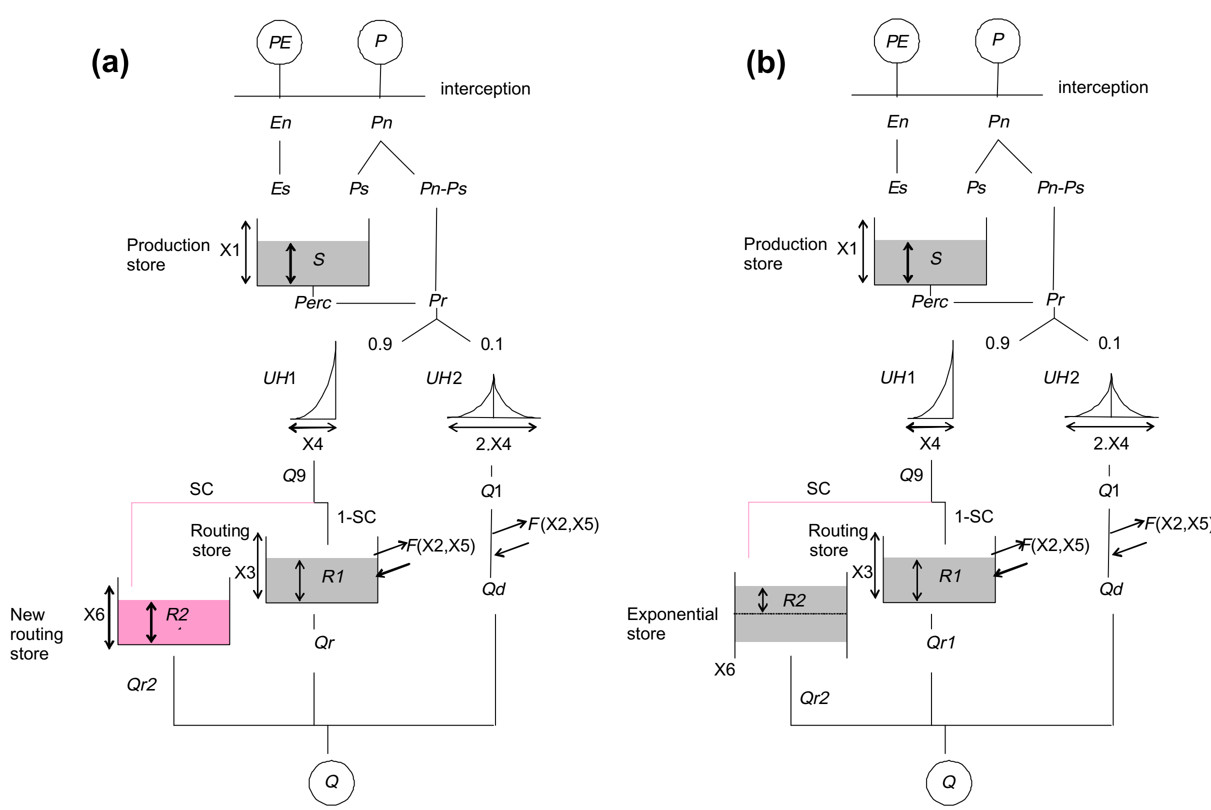

Fig. 5. Schematic representation of modified versions of the GR5J model (M5 (a) and M8 (b)).

Table 2
Mean model performance for versions M4 and M5.

	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}	Number of free parameters
M4	0.637	0.661	0.369	7
M5	0.634	0.659	0.365	6

Table 3
Mean efficiency values for versions M5 and M6.

	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}
M5	0.634	0.659	0.365
M6	0.625	0.641	0.310

Table 4
Mean model performance for versions M8–M11 (multiple routing stores) and mean relative performance RE^* with reference to M8 over the catchment set.

	M8	M9	M10	M11
NSE^*_{iQ}	0.383	0.384	0.385	0.384
RE* (%)	–	0.10	0.16	0.10

Various formulations were tested, among which we give the examples of versions M5, M7 and M8 in Table 1. Fig. 6 shows the corresponding distributions of efficiency values over the catchment set. The percentiles of the distribution of the model version M8 (with an additional exponential store) indicate better performance. The exponential store is known to be an efficient tool to simulate long recession spells (see Michel et al., 2003).

As suggested by Le Moine (2008) we also analysed the performance of M8 by removing the direct flow component (version M12). Indeed, the introduction of a new flow component may make this direct flow component unnecessary. However, the results are slightly lower than version M8 (Table 5), so we chose to keep this direct flow in the model. Note that this direct flow does not require specific free parameters.

Other versions were tested and several gave similar although slightly lower results. Thus, in all our tests, the M8 version was shown to be the most satisfactory and we chose to select it as a good candidate for providing improved low-flow simulation. We will call it GR6J hereafter (daily (J) version of the GR model with six free parameters).

3.2. Comparing the results of GR4J, GR5J and GR6J

This section quantifies the differences in the model's behaviour and performance between the GR4J, GR5J and GR6J versions in greater detail. Since GR5J was shown by Le Moine (2008) to yield better efficiency than GR4J, we mainly focus on the relative performance of GR5J and GR6J.

3.2.1. Relative performance of GR6J

The percentage improvement in the NSE^*_{iQ} values of GR6J are calculated in terms of relative efficiency values (RE^*). The RE^* values of the GR6J model are calculated with reference to the GR5J model. Table 6 shows the average relative performance for the three NSE^* criteria. The RE^* value based on NSE^*_{iQ} in Table 6 indicates a significant improvement in the simulation of low flows without losing efficiency on high flows. The significance of the improvement in performance is evaluated using the Student *T*-test at a 99% confidence level (*T*-values should be above 2.576). Although the differences may not seem large, remember that they were obtained on a large set of catchment, which makes them very significant (see also Mathevet (2005) for further discussion). When looking at the criterion on inverse flows, RE^* is positive on a majority of catchments, which means that the additional store improves this set of catchments.

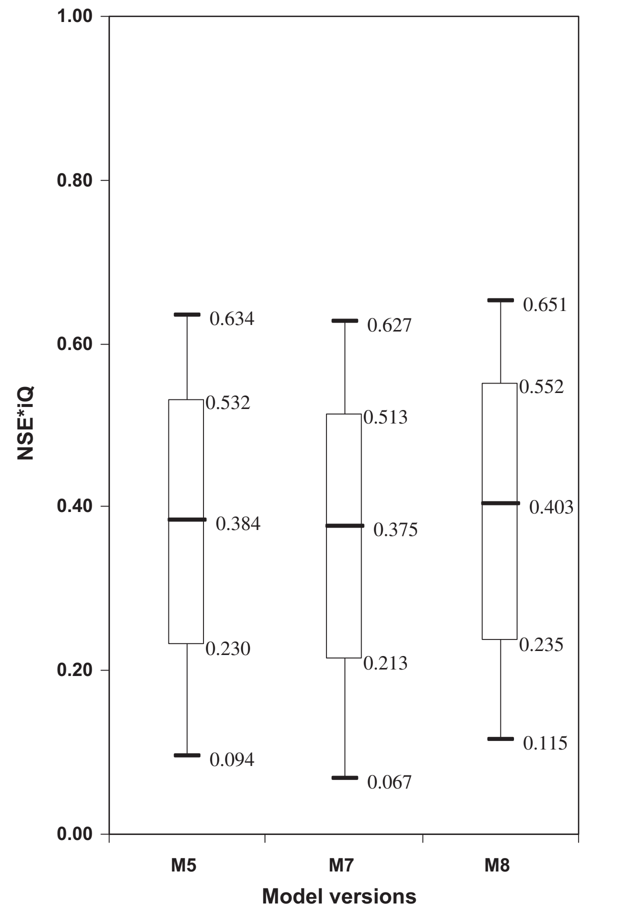

Fig. 6. Box plots of NSE^*_{iQ} values obtained in validation by model versions having different formulations of the additional routing store (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles).

Table 5
Mean efficiency values of M8 vs M12.

	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}
M8	0.634	0.662	0.383
M12	0.631	0.657	0.378

Table 6
Mean performance of GR5J and GR6J and relative performance of GR6J with reference to GR5J over the catchment set for various criteria (criteria on Q and iQ put more emphasis on floods and low flows, respectively), and significance of the improvement using the *T*-test (*T*-values should be greater than 2.576 at a 99% confidence level). The results were obtained in validation after calibration using another objective function.

Model	GR5J	GR6J	RE^* (%)
NSE^*_Q	0.629	0.634	0.83
NSE^*_{lnQ}	0.648	0.662	2.45
NSE^*_{iQ}	0.346	0.383	4.26

3.2.2. Illustration of the model's results

It is always difficult to select representative examples when working on a large catchment set. However, we wished to illustrate the model's results on a few case studies, by providing simulated hydrographs. We selected three catchments (see Table 7) with different hydro-climatic conditions and considered their streamflow values for a period of 1 year. We chose the year 2003, which was one of the driest years over the past decade in France. Fig. 7 shows the observed flow series and the flow series simulated by the two models, GR5J and GR6J, for the three catchments. Note that the graphs use logarithmic scales to emphasise differences in low flows. In catchment A and B, the performance of the GR6J model is significantly better than GR5J's performance on the very low-flow. The GR5J model tends to underestimate these flows, especially in the case of catchment B. In catchment J, the two models give similar results and also similar dynamics in low-flow conditions, which indicates that the introduction of the new store is neutral on this catchment.

Table 7
Characteristics of sample catchments.

Catchment characteristics	A	B	J
Gauging station	Custines	Saint Michel	Drennec
River	Moselle River	Meuse River	Aber Wrac'h River
Catchment code	A7010610	B2220010	J3205710
Mean rainfall (mm/year)	1109	954	1087
Mean streamflow (mm/year)	530	378	593
Mean potential evapotranspiration (mm/year)	614	619	643
Catchment area (km^2)	6830	2540	24

3.2.3. Parameter stability and identifiability

Fig. 8 compares the stability of parameters of the GR5J and GR6J models obtained on the two calibration periods (P1 and P2). In general, there is a quite good agreement between periods, with the parameters showing good stability, which is a desirable property. However, the threshold values for groundwater exchange

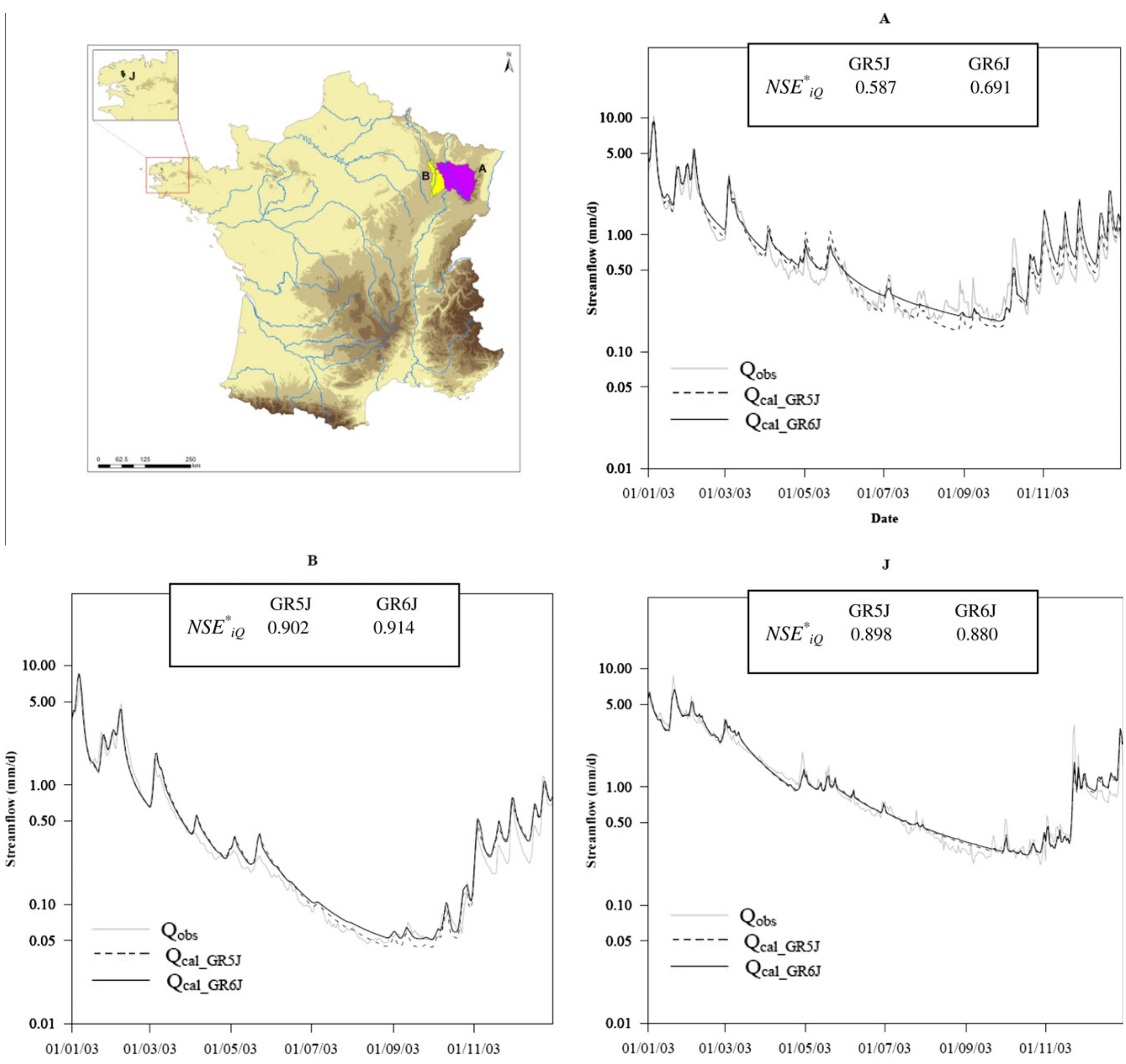

Fig. 7. Location of sample catchments and illustration of their hydrographs simulated by the GR5J and GR6J models, with corresponding NSE^*_{iQ} efficiency values.

(X5) and the sixth parameter of the GR6J model change significantly for a number of catchments, which may be due to a lower identifiability of these parameters for these catchments. The scatter seems a bit lower with the GR6J model for the X2 and X5 parameters. Interestingly, the reverse is observed for the capacity of the routing store (X3), for which the spread seems greater in

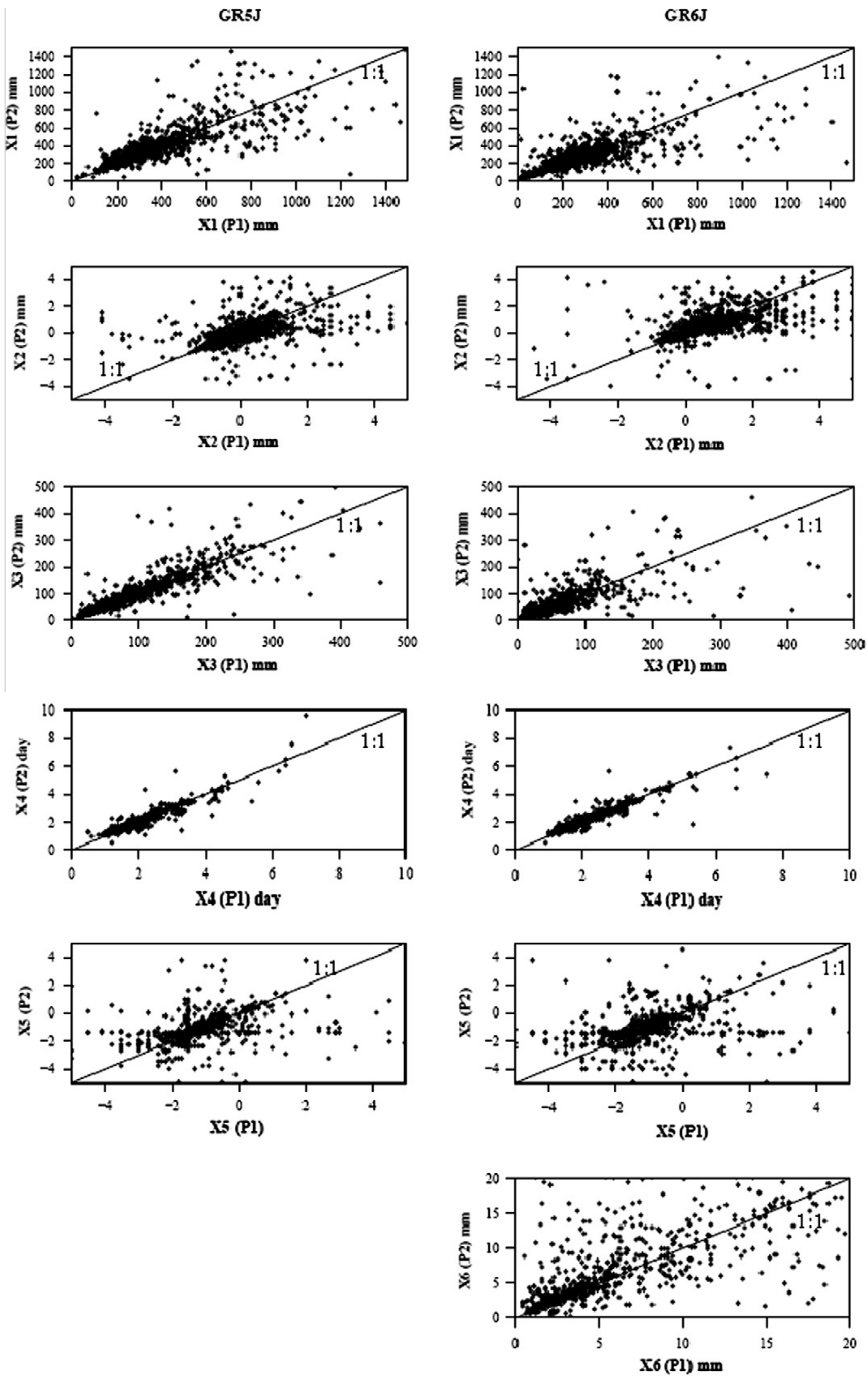

Fig. 8. Comparison of the parameter values obtained on the two calibration periods P1 and P2 for the GR5J and GR6J models.

Table 8
Average efficiency values for five lumped models compared to GR5J and GR6J for various criteria (criteria on Q and iQ put more emphasis on floods and low flows, respectively). Results were obtained in validation after calibration using another objective function.

Model acronym	Reference describing the original version	Number of free parameters	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}
HBV0	Bergström and Forsman (1973)	9	0.546	0.559	0.156
IHAC	Jakeman et al. (1990)	6	0.528	0.556	0.196
MOHY	Fortin and Turcotte (2007)	7	0.493	0.554	0.229
MORD	Garçon (1999)	6	0.603	0.616	0.302
TOPM	Beven and Kirby (1979)	8	0.574	0.584	0.216
GR4J	Perrin et al. (2003)	4	0.621	0.617	0.230
GR5J	Le Moine (2008)	5	0.629	0.648	0.346
GR6J		6	0.634	0.662	0.383

the case of GR6J. This means that the introduction of the new routing store impacted the rest of the routing module, especially the initial routing store. The additional complexity in the model seems to be at the cost of a lower identifiability for some components. Note that in spite of the precautions taken for model initialisation, parameter optimisation may still have been hampered by unsuitable initial conditions on some groundwater dominated catchment, as discussed by Le Moine (2008).

3.2.4. GR6J vs existing models

To finalise the comparative assessment, the proposed GR6J version of the model was compared to independent lumped models. The selection of models is shown in Table 8, along with model's mean performance for three criteria. Note that, to be able to apply the models in exactly the same conditions (i.e. the same data, calibration procedure and testing scheme), we had to recode the models and sometimes slightly modify them. For this comparison, it was important to rely on model structures that were representative of various conceptualisations of low-flow simulation. More details on the modifications made are given by Perrin et al. (2003) and Mathevet (2005).

The average performance values calculated over the entire catchment set indicate that GR6J is suitable to simulate low flows on this data set, since it compares favourably well with the other models. While showing significant gains in low flows, it still remains efficient in high flows. This indicates that GR6J is a good candidate for various hydrological modelling applications, i.e. it is a generic model for end-users interested in the advantages of one-size-fits-all models (which we acknowledge may not be a generally shared opinion, see e.g. Savenije, 2009) or a good starting skeleton for hydrologists aiming at customised solutions.

4. Conclusions

Improving the low-flow simulation ability without impacting the high-flow simulation ability: this was one of our objectives in this study. We chose to proceed by trial and error, as recommended by Nash and Sutcliffe (1970) and Michel et al. (2006). Working on a large set of catchments proved to be a good way to prevent undue complexity in the proposed modified versions of the model. Here we started from the simple GR5J model and tested a number of modified versions, some having higher performance values compared to the initial model structures. The model's performance was not equally sensitive to all the tested modifications. Among the modifications that proved the most robust, the addition of an exponential routing store, in parallel to the existing routing store in the GR5J model, showed improvement in low flows on average, still remaining efficient in high-flow conditions. The complexity added by this modification (an additional free parameter) seems to be warranted by the model's results, as well as by the comparison to other existing models. In spite of this improvement, it is not possible to say that we improved the physical realism of this model, since the initial intention was not to explicitly represent the physical mechanisms. Instead this study focused more on identifying the main features of the rainfall–runoff transformation at the catchment scale and improving the model's predictive power. The improved model version provides a better representation of the catchment's hydrological behaviour.

Last, let us note that the level of performance in low-flow conditions seems to remain quite low. This may be for several reasons, including structural model errors, data quality or artificial influences. Nonetheless, it shows that specific research should be continued to improve the efficiency of hydrological models for low-flow simulation, a domain that was probably overlooked in the past.

Acknowledgements

We are very grateful to Météo–France for providing meteorological data and SCHAPI for providing flow data. We also wish to thank Domaine d'Intérêt Majeur Agrosciences, territoires, écologie, alimentation (DIM ASTREA) of Région Ile–de–France and Office National de l'Eau et des Milieux Aquatiques (ONEMA) for their financial support to conduct this study.

References

Anderson, M.L., Chen, Z.Q., Kavvas, M.L., 2004. Modeling low flows on the Cosumnes River. J. Hydrol. Eng. 9 (2), 126–134.
Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I., Lavabre, J., 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. J. Hydrol. 250, 206–223.
Andréassian, V., Hall, A., Chahinian, N., Schaake, J., 2006. Introduction and synthesis: why should hydrologists work on a large number of basin data sets? Large sample experiments for hydrological model parameterisation: results of the model parameter experiment – MOPEX. IAHS Publ. 307, 1–5.
Andréassian, V., Perrin, C., Parent, E., Bardossy, A., 2010. Editorial – The court of miracles of hydrology: can failure stories contribute to hydrological science? Hydrol. Sci. J. 55 (6), 849–856. doi:10.1080/02626667.2010.506050.
Bahremand, A., De Smedt, F., 2008. Distributed hydrological modeling and sensitivity analysis in Torysa Watershed, Slovakia. Water Resour. Manage 22, 393–408.
Bergström, S., Forsman, A., 1973. Development of a conceptual deterministic rainfall–runoff model. Nord. Hydrol. 4, 147–170.
Beven, K., Kirby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. J. Hydrol. Sci. 24 (1), 43–69.
Boé, J., Terray, L., Martin, E., Habets, F., 2009. Projected changes in components of the hydrological cycle in French river basins during the 21st century. Water Resour. Res. 45, W08426.
Davison, B., van der Kamp, G., 2008. Low-flows in deterministic modelling: a brief review. Can. Water Resour. Assoc. 33 (2), 181–194.
Demuth, S., 2005. Low flows and droughts – a European prospective. IAHS Newsletter 82, March 2005, 7–8.
Edijatno, Michel, C., 1989. Un modèle pluie-débit journalier à trois paramètres. La Houille Blanche (2), 113–121.
Edijatno, Nascimento, N.O., Yang, X., Makhlouf, Z., Michel, C., 1999. GR3J: a daily watershed model with three free parameters. Hydrol. Sci. J. 44 (2), 263–277. doi:10.1080/02626669909492221.
Fenicia, F., Savenije, H.H.G., Matgen, P., Pfister, L., 2008. Understanding catchment behavior through 664 stepwise model concept improvement. Water Resour. Res. 44, W01402. doi:10.1029/2006WR005563.
Feyen, L., Dankers, R., 2009. Impact of global warming on streamflow drought in Europe. J. Geophys. Res. 114, D17116.
Fleckenstein, J.H., Niswonger, R.G., Fogg, G.E., 2006. River–aquifer interactions, geologic heterogeneity, and low flow management. Ground Water 44 (6), 837–852.
Fortin, V., Turcotte, R., 2007. Le modéle hydrologique MOHYSE. Internal note, CEHQ, Québec.
Garçon, R., 1999. Modéle global Pluie-Débit pour la prévision et la prédétermination des crues (Lumped rainfall–runoff model for flood forecasting and design flood estimation, in French). La Houille Blanche 7 (8), 88–95.

Herron, N., Croke, B., 2009. Including the influence of groundwater exchanges in a lumped rainfall–runoff model. Math. Comput. Simul. 79, 2689–2700.
Hughes, D.A., 2004. Incorporating groundwater recharge and discharge functions into an existing monthly rainfall–runoff model. J. Hydrol. Sci. 49 (2), 297–311.
Jakeman, A.J., Littlewood, I.G., Whitehead, P.G., 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. J. Hydrol. 117, 275–300.
Kim, H.S., Croke, B.F.W., Jakeman, A.J., Chiew, F.H.S., 2011. An assessment of modelling capacity to identify the impacts of climate variability on catchment hydrology. Math. Comput. Simul. 81, 1419–1429.
Klemes, V., 1986. Operational testing of hydrological simulation models. J. Hydrol. Sci. 31, 13–24.
Lang, C., 2007. Etiages et tarissements: vers quelles modélisations? L'approche conceptuelle et l'analyse statistique en réponse à la diversité spatiale des écoulements en étiage des cours d'eau de l'Est français. Ph.D. thesis. Univerité Paul Verlaine (Metz), 292 pp.
Lang, C., Gille, E., Francois, D., Drogue, G., 2008. Improvement of a lumped rainfall–runoff structure and calibration procedure for predicting daily low flow discharges. J. Hydrol. Hydromech. 56 (1), 59–71.
Le Moine, N., 2008. Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performance et du réalisme des modéles pluie-débit? Ph.D. thesis. Université Pierre et Marie Curie (Paris), Cemagref (Antony), 324 pp.
Le Moine, N., Andréassian, V., Perrin, C., Michel, C., 2007. How can rainfall–runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments. Water Resour. Res. 43, W06428.
Lerat, J., 2009. Quels apports hydrologiques pour les modèles hydrauliques? Vers un modèle intégré de simulation des crues. Ph.D. thesis. Université Pierre et Marie Curie (Paris), Cemagref (Antony, Montpellier), 300 pp.
Mathevet, T., 2005. Quels modéles pluie-debit globaux au pas de temps horaire? Développements empiriques et comparison de modéles sur un large échantillon de bassins versants. Ph.D. thesis. ENGREF (Paris), Cemagref (Antony), France, 463 pp.
Mathevet, T., Michel, C., Andréassian, V., Perrin, C., 2006. A bounded version of the Nash–Sutcliffe criterion for better model assessment on large sets of basins, large sample basin experiments for hydrological model parameterization: results of the model parameter experiment – MOPEX. IAHS Publ. 307, 211–219.
Meselhe, E.A., Habib, E.H., Oche, O.C., Gautam, S., 2009. Sensitivity of conceptual and physically based hydrologic models to temporal and spatial rainfall sampling. J. Hydrol. Eng. 14 (7), 711–720.
Michel, C., Perrin, C., Andreassian, V., 2003. The exponential store: a correct formulation for rainfall-runoff modelling. J. Hydrol. Sci. 48 (1), 109–124.
Michel, C., Perrin, C., Andréassian, V., Oudin, L., Mathevet, T., 2006. Has basin scale modelling advanced far beyond empiricism? IAHS Publ. 307, 108–116.
Mouelhi, S., 2003. Vers une chaîne cohérente de modèles pluie-débit conceptuels globaux aux pas de temps pluriannuel, annuel, mensuel et journalier. Ph.D. thesis. ENGREF, Cemagref Antony, France, 323 pp.
Nascimento, N.O., 1995. Appréciation à l'aide d'un modèle empirique des effets d'action anthropiques sur la relation pluie-débit à l'échelle du bassin versant. Ph.D. thesis. CERGRENE/ENPC, Paris, 550 pp.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I – A discussion of principles. J. Hydrol. 10, 282–290.
Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., Loumagne, C., 2005a. Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 – Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. J. Hydrol. 303 (1–4), 290–306.
Oudin, L., Michel, C., Anctil, F., 2005b. Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 1 – Can rainfall–runoff models effectively handle detailed potential evapotranspiration inputs? J. Hydrol. 303 (1–4), 275–289.
Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., 2006a. Dynamic averaging of rainfall–runoff model simulations from complementary model parameterizations. Water Resour. Res. 42, W07410.
Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., Michel, C., 2006b. Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. J. Hydrol. 320, 62–83. doi:10.1016/j.jhydrol.2005.07.016.
Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279 (1–4), 275–289.
Pushpalatha, R., Perrin, C., Le Moine, N., Andréassian, V., submitted for publication. Efficiency criteria suitable for evaluating low-flow simulations. J. Hydrol.
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. J. Appl. Meteorol. Climatol. 47 (1), 92–107.
Ruelland, D., Ardoin-Bardin, S., Billen, G., Servat, E., 2008. Sensitivity of a lumped and semi-distributed hydrological model to several methods of rainfall interpretation on a large basin in West Africa. J. Hydrol. 361, 96–117.
Savenije, H.H.G., 2009. The art of hydrology. Hydrol. Earth Syst. Sci. 13, 157–161.
Seibert, J., 2001. On the need for benchmarks in hydrological modelling. Hydrol. Process. 15 (6), 1063–1064.
Singh, V.P., Frevert, D.K. (Eds.), 2002a. Mathematical Models of Large Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado, p. 14 pp.
Singh, V.P., Frevert, D.K. (Eds.), 2002b. Mathematical Models of Small Watershed Hydrology and Applications. Water Resources Publications, Highlands Ranch, Colorado, p. 72 pp.
Staudinger, M., Stahl, K., Seibert, J., Clark, M.P., Tallaksen, L.M., 2011. Comparison of hydrological model structures based on recession and low flow simulations. Hydrol. Earth Syst. Sci. Discuss. 8, 6833–6866. doi:10.5194/hessd-8-6833-2011.
Tang, Y., Reed, P., Wagener, T., van Werkhoven, K., 2007. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. Hydrological Earth System Sciences 11, 793–817.
Vidal, J.-P., Martin, E., Franchisteguy, L., Baillon, M., Soubeyroux, J.-M., 2010. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. Int. J. Climatol. 30, 1627–1644.
Wagener, T., Wheater, H.S., Gupta, H.V., 2004. Rainfall–Runoff Modelling in Gauged and Ungauged Catchments. Imperial College Press, London.
Xu, C., Tunemar, L., Chen, Y.D., Singh, V.P., 2006. Evaluation of seasonal and spatial variations of lumped water balance model sensitivity to precipitation data errors. J. Hydrol. 324, 80–93.