Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Hydrology 279 (2003) 275–289
Journal of Hydrology
www.elsevier.com/locate/jhydrol

# Improvement of a parsimonious model for streamflow simulation

Charles Perrin*, Claude Michel, Vazken Andréassian

*Cemagre, Water Quality and Hydrology Research Unit, Parc de Tourvoie BP 44, 92163 Antony cedex, France*

Received 11 October 2002; revised 15 May 2003; accepted 27 May 2003

**Abstract**

Hydrologists have been struggling over the past decades to improve rainfall-runoff models. As a consequence, models proposed 20–30 years ago still keep evolving as progress is made in the understanding of catchment hydrological behaviour. Here we present the GR4J model, a daily lumped rainfall-runoff model which is the result of a continuous improvement process over the last 15 years. The article provides the mathematical formulation of a new four-parameter version of the model. Model performance is assessed on a large sample of catchments: compared to other rainfall-runoff models, the GR4J performance is among the best ones. It also gives better results than the previous three-parameter model version, especially in the simulation of low flows. The tests indicate that a four-parameter structure corresponds to the maximum level of complexity that could be afforded in the model. Adding more free parameters did not bring significant improvements. The gain in model robustness with this new version should enhance the confidence in the practical use of this simple model for water engineering and resource management. The discussion underlines the potential limits introduced in the modelling process when one relies on a priori concepts in building a model structure and it stresses the value of large catchment samples to assess models.


*Keywords:* Rainfall-runoff modelling; Model complexity; Model improvement; Parsimony; Model robustness; GR4J model

## 1. Introduction

Achieving better streamflow simulations is an objective common to most hydrologists involved in rainfall-runoff modelling. Today, potential gains in model performance become all the more valuable as models are increasingly used for water resource engineering applications. As operational requirements are often found to be best served by lumped soil moisture accounting (SMA) models, this article will focus on this type of models.

A twofold problem appears when attempts are made to build efficient rainfall-runoff models or to improve existing ones. First, a structure—i.e. an ensemble of mathematical functions and devices—must be chosen to reflect the hydrological behaviour at the catchment scale. This requires appropriate criteria by which the expected model qualities can be judged. However, it is difficult to set a definite structure for a given model, mainly because today, no widely accepted general hydrological theory is available to simulate catchment behaviour. Second, one must find an adequate level of complexity for the proposed model structure—i.e. the number of free parameters—that will guarantee optimal performance. Here again, it is difficult to find

* Corresponding author. Tel.: +33-1-40-96-60-86; fax: +33-1-40-96-61-99.
*E-mail address:* charles.perrin@cemagref.fr (C. Perrin).

0022-1694/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0022-1694(03)00225-7

the appropriate number of degrees of freedom: too few will prevent the model from being sufficiently flexible, too many will lead to problems of parameter definition and model robustness. Structure and complexity are linked and must be considered together while the model is being developed. Generally, more mathematical functions in the model means more parameters.

Several approaches can be followed to improve streamflow simulations, depending on what the modeller perceives as the main source of uncertainty in the modelling process: inputs, determination of model parameter values, model structure. A possible way to improve flow simulation is to use the existing model structures as a starting point and then try to modify them. Today, most models are the result of a continuous development process: for example Sugawara (1995) explains how the Tank model was progressively developed, after the end of World War II, from a single reservoir built as an analogue computer to the present version with its more complex structure made of a series of linear stores. Several successive modifications were also introduced to the structure of other models such as the IHACRES (Ye et al., 1997), HBV (Lindström et al., 1997), SMAR (Tan and O'Connor, 1996), TOPMODEL (Beven, 1986) and Xinanjiang (Jayawardena and Zhou, 2000). In most cases, the proposed modifications aimed to improve model sub-processes, but the gains in model performance were not always as large as expected.

In a previous study, Perrin et al. (2001a) assessed 19 existing daily rainfall-runoff models and their results suggested that (1) the definition of the structure of the model plays a key role in its reliability, (2) a complexity of three to five parameters in the model is sufficient to obtain satisfactory performances at a daily time-step, (3) comparative exercises can open ways to improve models starting from very simple structures and making them progressively more complex. The next stage in this research was to try to improve streamflow modelling efficiency. A good candidate for such an improvement process was the three-parameter GR3J model proposed by Edijatno et al. (1999). It was the simplest model tested by Perrin et al. (2001a) and had proved to be a good basis in terms of efficiency.

The objectives of this article are (i) to give an exact description of the improved version of the GR3J model, here called GR4J, by providing a detailed layout of the mathematical model structure (Section 2), (ii) to present the results of model applications as compared with those of other models and demonstrate the gains in model performance and reliability over the former model version (Section 3), and (iii) to discuss the key aspects of the model development and improvement process (Section 4).

## 2. GR4J: a daily four-parameter rainfall-runoff model

The GR4J model (which stands for modèle du *G*énie *R*ural à *4* paramètres *J*ournalier) is a daily lumped four-parameter rainfall-runoff model. It belongs to the family of soil moisture accounting models. The GR4J model is the last modified version of the GR3J model originally proposed by Edijatno and Michel (1989) and then successively improved by Nascimento (1995) and Edijatno et al. (1999).

### 2.1. The choice of a spatial resolution

Though many models adopt a spatially distributed approach today, a lumped one was chosen for the GR4J model for three main reasons. First, although we have a fairly good understanding of some processes occurring in nature such as interception by the vegetation canopy, infiltration in homogeneous soils, etc. we are still unclear about the most important subsoil processes and, moreover, we do not know how these processes act at the catchment scale. Up to now, in spite of the great many scientific publications, nobody has been able to predict the response of any natural land-surface to a given rainfall event. Therefore, we consider that a sensible first move is to try to determine how a catchment works as a whole.

A second reason is that each of the building blocks of any distributed model is itself a lumped rainfall-runoff model. Before gathering an ensemble of elementary models, it is worthwhile to improve the devices that will be used as constituents.

Last, the practical superiority of distributed or semi-distributed approaches over lumped ones for streamflow simulation has not been clearly demonstrated yet (see e.g. Loague and Freeze, 1985; Michaud and Sorooshian, 1994; Refsgaard and Knudsen, 1996; Loumagne et al., 1999; Zhang et al., 2003 cited by Boyle et al., 2001)

### 2.2. Model description

Here we give a summary description of the GR4J model, as proposed by Perrin (2000). The bases of the discrete equations presented below can be found in the previous references dealing with the former versions of the model. To avoid confusion in the sequence and nature of model operations, the corresponding FORTRAN code is available on the *Cemagref* Web site.[1] Fig. 1 shows a diagram of the model.

In the following, for calculations at a given time step, we note $P$ the rainfall depth and $E$ the potential evapotranspiration (PE) estimate that are inputs to the model. $P$ is an estimate of the areal catchment rainfall that can be computed by any interpolation method from available raingauges. $E$ can be a long-term average value, which means that the same PE series is repeated every year.

All water quantities (input, output, internal variables) are expressed in mm, by dividing water volumes by catchment area, when necessary. All the operations described below are relative to a given time step and correspond to a discrete model formulation (obtained after integration of the continuous formulation over the time step).

*Determination of net rainfall and PE.* The first operation is the subtraction of $E$ from $P$ to determine either a net rainfall $P_n$ or a net evapotranspiration capacity $E_n$. In GR4J, this operation is computed as if there were an interception storage of zero capacity. $P_n$ and $E_n$ are computed with the following equations:

$$\text{If } P \geq E, \quad \text{then } P_n = P - E \text{ and } E_n = 0 \tag{1}$$

$$\text{otherwise } P_n = 0 \quad \text{and} \quad E_n = E - P \tag{2}$$

*Production (SMA) store.* In case $P_n$ is not zero, a part $P_s$ of $P_n$ fills the production store. It is determined as a function of the level $S$ in

[1] http://www.antony.cemagref.fr/webqhan/projets%20themes/Hydrologie/Code%20fortran.htm.

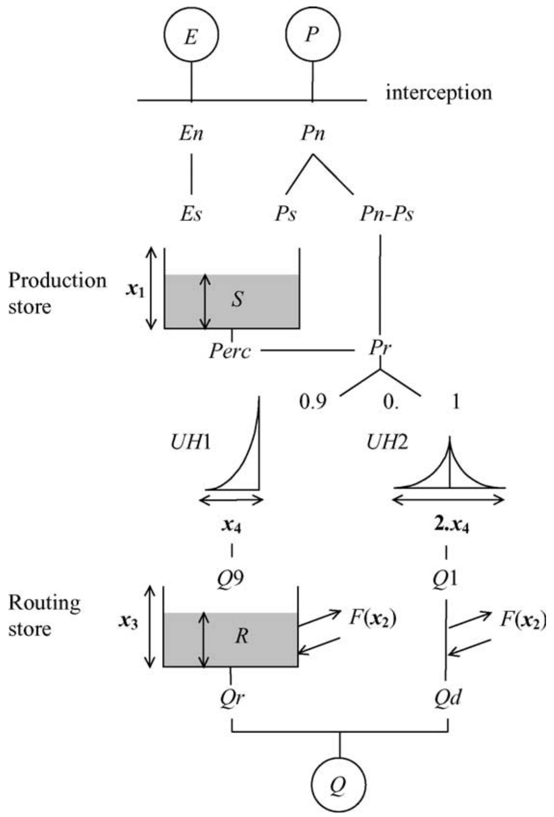


Fig. 1. Diagram of the GR4J rainfall-runoff model.

the store by:

$$P_s = \frac{x_1\left(1 - \left(\frac{S}{x_1}\right)^2\right)\tanh\left(\frac{P_n}{x_1}\right)}{1 + \frac{S}{x_1}\tanh\left(\frac{P_n}{x_1}\right)} \tag{3}$$

where $x_1$ (mm) is the maximum capacity of the SMA store. Eqs. (3) and (4) below result from the integration over the time step of the differential equations that have a parabolic form with terms in $(S/x_1)^2$, as detailed by Edijatno and Michel (1989).

In the other case, when $E_n$ is not zero, an actual evaporation rate is determined as a function of the level in the production store to calculate the quantity $E_s$ of water that will evaporate from the store. It is obtained by:

$$E_s = \frac{S\left(2 - \frac{S}{x_1}\right)\tanh\left(\frac{E_n}{x_1}\right)}{1 + \left(1 - \frac{S}{x_1}\right)\tanh\left(\frac{E_n}{x_1}\right)} \tag{4}$$

The water content in the production store is then updated with:

$$S = S - E_s + P_s \tag{5}$$

Note that $S$ can never exceed $x_1$. A representation of the rating curves obtained with Eqs. (3) and (4) is shown in Fig. 2.

A percolation leakage Perc from the production store is then calculated as a power function of the reservoir content:

$$\text{Perc} = S\left\{1 - \left[1 + \left(\frac{4}{9}\frac{S}{x_1}\right)^4\right]^{-1/4}\right\} \tag{6}$$

Perc is always lower than $S$. The reservoir content becomes:

$$S = S - \text{Perc} \tag{7}$$

The introduction of a percolation function from the production store constitutes one of the main differences with the previous GR3J version proposed by Edijatno et al. (1999). Such a percolation (or infiltration) function from a production store exists in many conceptual models, for example in the model proposed by Georgakakos and Baumer (1996).

The percolation function in Eq. (6) occurs as if it originated from a store with a maximum capacity of $9/4x_1$. Given the power law of the mathematical formulation, this means that the percolation does not contribute much to the streamflow and is interesting mainly for low flow simulation.

*Linear routing with unit hydrographs*. The total quantity $P_r$ of water that reaches the routing functions is given by:

$$P_r = \text{Perc} + (P_n - P_s) \tag{8}$$

$P_r$ is divided into two flow components according to a fixed split: 90% of $P_r$ is routed by a unit hydrograph UH1 and then a non-linear routing store, and the remaining 10% of $P_r$ are routed by a single unit hydrograph UH2. With UH1 and UH2, one can simulate the time lag between the rainfall event and the resulting streamflow peak. Their ordinates are used in the model to spread effective rainfall over several successive time steps. Both unit hydrographs depend on the same time parameter $x_4$ expressed in days. However, UH1 has a time base of $x_4$ days whereas UH2 has a time base of $2x_4$ days. $x_4$ can take real values and is greater than 0.5 days.

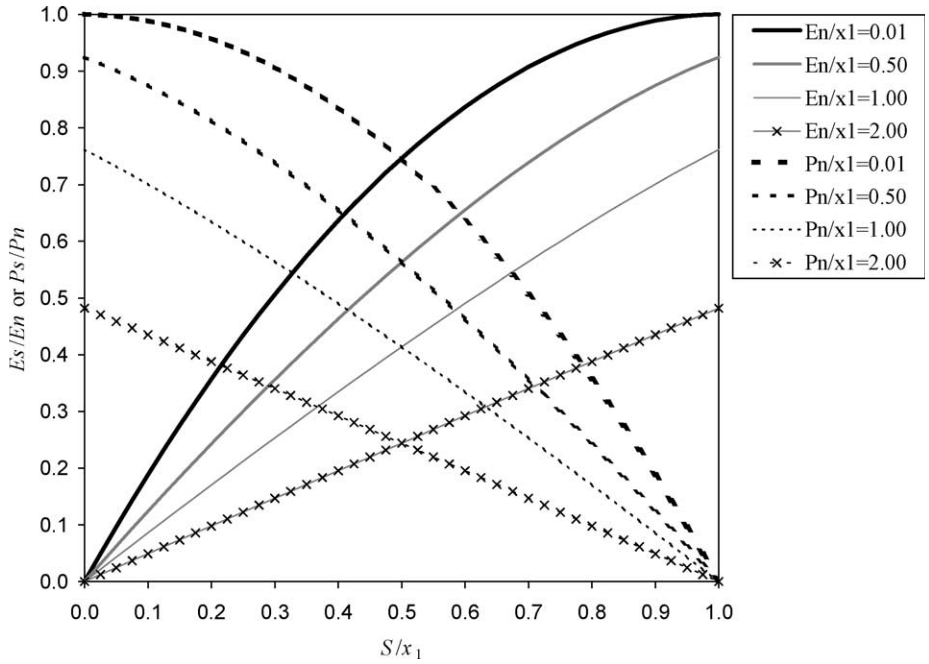


Fig. 2. Illustration of the behaviour of the production functions ($E_s/E_n$ : solid line; $P_s/P_n$ : dashed line) as a function of storage rate $S/x_1$ for different values of $E_n/x_1$ or $P_n/x_1$.

In their discrete form, unit hydrographs UH1 and UH2 have $n$ and $m$ ordinates, respectively, where $n$ and $m$ are the smallest integers exceeding $x_4$ and $2x_4$, respectively. This means that the water is staggered into $n$ unit hydrograph inputs for UH1 and $m$ inputs for UH2. The ordinates of both unit hydrographs are derived from the corresponding $S$-curves (cumulative proportion of the input with time) denoted by SH1 and SH2, respectively. SH1 is defined along time $t$ by:

$$\text{For } t \le 0,\ \text{SH1}(t) = 0 \tag{9}$$

$$\text{For } 0 < t < x_4,\ \text{SH1}(t) = \left(\frac{t}{x_4}\right)^{5/2} \tag{10}$$

$$\text{For } t \ge x_4,\ \text{SH1}(t) = 1 \tag{11}$$

SH2 is similarly defined by:

$$\text{For } t \le 0,\ \text{SH2}(t) = 0 \tag{12}$$

$$\text{For } 0 < t \le x_4,\ \text{SH2}(t) = \frac{1}{2}\left(\frac{t}{x_4}\right)^{5/2} \tag{13}$$

$$\text{For } x_4 < t < 2x_4,\ \text{SH2}(t) = 1 - \frac{1}{2}\left(2 - \frac{t}{x_4}\right)^{5/2} \tag{14}$$

$$\text{For } t \ge 2x_4,\ \text{SH2}(t) = 1 \tag{15}$$

UH1 and UH2 ordinates are then calculated by:

$$\text{UH1}(j) = \text{SH1}(j) - \text{SH1}(j-1) \tag{16}$$

$$\text{UH2}(j) = \text{SH2}(j) - \text{SH2}(j-1) \tag{17}$$

where $j$ is an integer. If $0.5 \le x_4 \le 1$, UH1 has a single ordinate equal to one and UH2 has only two ordinates. Fig. 3 shows an example of unit hydrograph ordinates for $x_4 = 3.8$ days.

*Catchment water exchange*. A groundwater exchange term $F$ that acts on both flow components, is then calculated as:

$$F = x_2\left(\frac{R}{x_3}\right)^{7/2} \tag{18}$$

where $R$ is the level in the routing store, $x_3$ its 'reference' capacity and $x_2$ the water exchange coefficient. $x_2$ can be either positive in case of water imports, negative for water exports or zero when there is no water exchange. The higher the level in the routing store, the larger the exchange. In absolute value, $F$ cannot be greater than $x_2$ : $x_2$ represents the maximum quantity of water that can be added (or released) to (from) each model flow component when the routing store level equals $x_3$.

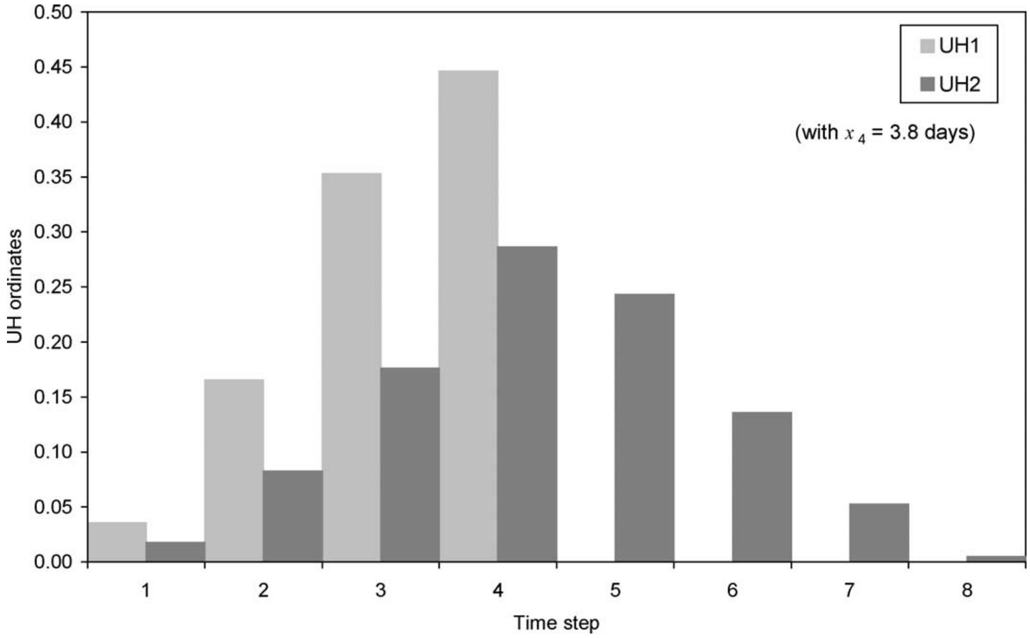


Fig. 3. Example of the ordinates of UH1 and UH2 for parameter $x_4 = 3.8$ days.

*Non-linear routing store*. The level in the routing store is updated by adding the output $Q9$ of UH1 and $F$ as follows:

$$R = \max(0; R + Q9 + F) \quad (19)$$

The outflow $Q_r$ of the reservoir is then calculated as:

$$Q_r = R\left\{1 - \left[1 + \left(\frac{R}{x_3}\right)^4\right]^{-1/4}\right\} \quad (20)$$

$Q_r$ is always lower than $R$, as shown in Fig. 4. The level in the reservoir becomes:

$$R = R - Q_r \quad (21)$$

Note that, although the reservoir can receive a water input greater than the saturation deficit $x_3 - R$ at the beginning of a time step, the level in the reservoir can never exceed the capacity $x_3$ at the end of a time step, as shown in Fig. 4. Therefore, the capacity $x_3$ could be called the 'one day ahead maximum capacity'. This routing store is able to simulate long streamflow recessions, when necessary.

*Total streamflow*. Like the content of the routing store, the output $Q1$ of UH2 is subject to the same water exchange $F$ to give the flow component $Q_d$ as follows:

$$Q_d = \max(0; Q1 + F) \quad (22)$$

Total streamflow $Q$ is finally obtained by:

$$Q = Q_r + Q_d \quad (23)$$

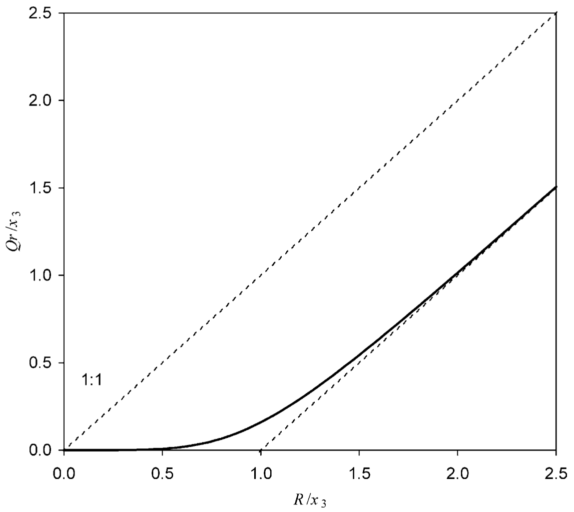


Fig. 4. Illustration of the outflow $Q_r$ from the routing reservoir as a function of the level in the store after the introduction of input $Q9$.

*2.3. Model parameters*

In the GR4J model, four parameters have to be optimised:

$x_1$ : maximum capacity of the production store (mm)
$x_2$ : groundwater exchange coefficient (mm)
$x_3$ : one day ahead maximum capacity of the routing store (mm)
$x_4$ : time base of unit hydrograph UH1 (days)

Here, the capacity of the production store is optimised whereas it was a fixed parameter with a value of 330 mm in GR3J. All four parameters are real numbers. $x_1$ and $x_3$ are positive, $x_4$ is greater than 0.5 and $x_2$ can be either positive, zero or negative.

Note that some figures in the model equations may appear as fixed parameter values, e.g. a power 4 in Eqs. (6) and (20), a fixed split 10–90% of effective rainfall, a 2.5 exponent in the computation of the unit hydrographs, a 2.25 coefficient related to the percolation function in Eq. (6). These values were chosen as those yielding the best model results in many different test conditions. They were fixed because leaving them free did not significantly improve (or even degraded) the model results while adding unhelpful complexity to the model structure. The case of the fixed split of effective rainfall may appear surprising: in some other models (e.g. IHACRES), this split is used to identify quick and slow flow components and is determined by a free parameter or even as a function of a level in a model store that varies in time. Here we chose not to introduce a priori ideas on how this split could be made while building the model structure. Indeed, when we tried to let this split free in the model structure, no significant improvement of model efficiency could be obtained, indicating that the usefulness of this additional parameter could not be confirmed by the data. Fixing this coefficient in the model structure did not prevent from obtaining almost as satisfactory results.

Most optimisation algorithms used to calibrate the model parameters require knowledge of an initial parameters set. This initial set may consist of median values obtained on a large variety of catchments (see Table 1). Approximate 80% confidence intervals

Table 1
Values of median model parameters and approximate 80% confidence intervals

| | Median value | 80% Confidence interval |
|---|---|---|
| $x_1$ (mm) | 350 | 100–1200 |
| $x_2$ (mm) | 0 | −5 to 3 |
| $x_3$ (mm) | 90 | 20–300 |
| $x_4$ (days) | 1.7 | 1.1–2.9 |

for the four parameters are provided in Table 1. They were derived from the 0.1 and 0.9 percentiles of the distributions of model parameters obtained over a large sample of catchments. Given the small number of model parameters, simple optimisation algorithms are generally capable of identifying parameter values yielding satisfactory results. The choice of an objective function depends on the objectives of model user. Note that care should be taken to set appropriate initial conditions of the internal state variables in the model to avoid discrepancies at the beginning of the simulation periods. One year can be used for model warm-up at the beginning of each simulation, as done, for example, by Chiew and McMahon (1994).

## 3. Results of model application

In this section, we assess the value of the GR4J model for streamflow modelling.

### *3.1. Comparative assessment of model performances*

The evaluation of model performances in absolute terms is difficult and it is better to use benchmark references as advocated for example by Seibert (2001). One possibility is to evaluate performances in a comparative way by testing several models on the same study cases; one of these models, either very simple and/or generally accepted as being a standard in modelling, might be considered as a reference. A few comparative studies have already been carried out in different climatic contexts (e.g. WMO, 1975; Chiew et al., 1993; Michaud and Sorooshian, 1994; Refsgaard and Knudsen, 1996), to examine the differences between several models or modelling approaches.

Here, the evaluation of the GR4J model performances was carried out in a comparative way according to the approach adopted by Perrin et al. (2001a). Full details of the methodology can be found in their article and only a brief outline of the test conditions is given below:

- test sample of 429 catchments situated in different climate conditions, from semi-arid to temperate and tropical humid. For each catchment, between two and six non-overlapping, one- to eight-year long periods, were selected from the data record;
- split-sample test scheme for an evaluation of model performances in simulation mode, with a total of 3204 simulation tests;
- five criteria (CR1–CR5) were used to judge model efficiency. These criteria are:
  CR1: the classical Nash–Sutcliffe (Nash and Sutcliffe, 1970) criterion,
  CR2: the Nash–Sutcliffe criterion calculated on square root transformed streamflow,
  CR3: the Nash–Sutcliffe criterion calculated on the logarithm transformed streamflow,
  CR4: a criterion of absolute error,
  CR5: a water balance criterion.

All criteria vary in the interval $]-\infty, 1]$, 1 meaning a perfect agreement. Criterion CR2 is also used as an objective function to calibrate models, since it was found to be of general interest, i.e. not too specific to any particular model application: CR2 is a median way between CR1 that puts emphasis on the simulation of flood events and CR3 that puts emphasis on the quality of low flow simulation.

All results presented here were obtained in simulation mode. The performances of the GR4J model were compared to those of the 19 models tested in Perrin et al. (2001a), that are also lumped rainfall-runoff models with various levels of complexity (at most, nine parameters to be optimised). The GR3J model from which the GR4J model was derived is one of these models. A hypothetic model, defined as an envelope curve and gathering the best performance among all models for each catchment (see Perrin et al., 2001a), was also used as an upper bound of model performances (called the Mαω model in the following). Fig. 5 shows the distribution of results

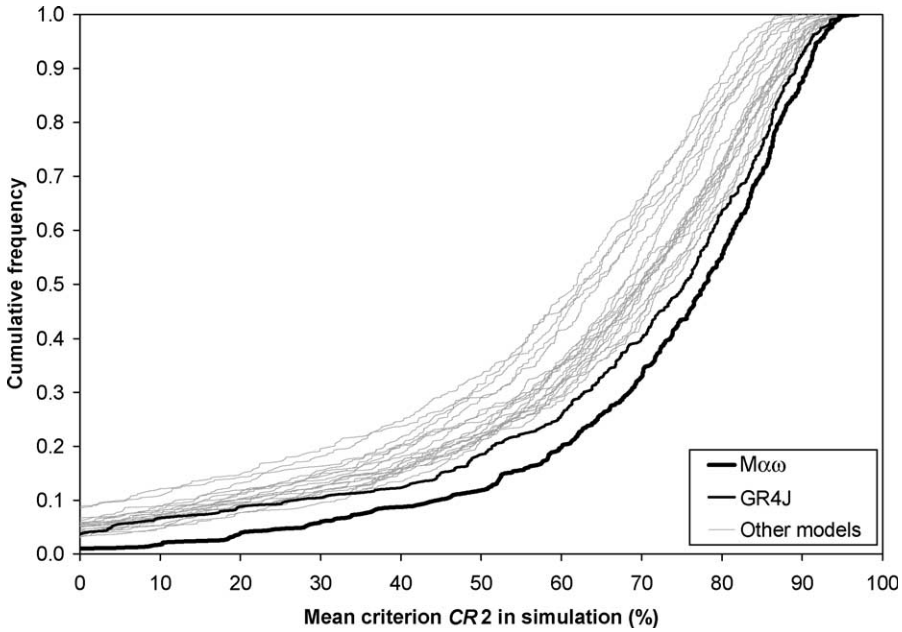


Fig. 5. Distribution of mean performances per catchment obtained in simulation with criterion CR2.

obtained for all models in the case of criterion CR2; the more the curve on the left hand side of the graph, the better the performance of a model. The distributions indicate that the GR4J model performs fairly well in comparison with the 19 other models. The distribution curve of its efficiency is even closer to that of the Mαω model than the other distributions are.

### *3.2. Performances improvement from GR3J to GR4J*

Table 2 shows the mean performances on the 429 catchments by both GR3J and GR4J models according to the five criteria. To quantify the improvement obtained with the new GR4J model, we used the $r^2$ statistics proposed by Nash and Sutcliffe (1970) and used, for example, by Senbeta et al. (1999). It is given by:

$$r^2 = \frac{R_2 - R_1}{1 - R_1} \quad (24)$$

where $R_1$ and $R_2$ are the efficiencies obtained by the original and modified model versions, respectively. This statistic compares the improvement achieved when going from model version #1 to #2 to the maximum possible improvement that might be achieved if one could build a perfect model (efficiency equal to 1). Initially proposed for the Nash–Sutcliffe criterion, this statistic can be applied to the five proposed evaluation criteria since they all vary within the same interval $]-\infty, 1]$. Senbeta et al. (1999) consider in their study that a value of $r^2$ greater than 10% indicates a significant improvement of model performance in the case of the Nash–Sutcliffe criterion. However, this threshold of significance is usually chosen subjectively and depends largely on modeller experience and on the chosen performance criterion.

The values of $r^2$ in Table 2 show that model performances are improved for all criteria. This improvement is more significant for the three

Table 2
Mean performances of the GR3J and GR4J models and estimation of performance improvement

| | GR3J | GR4J | $r^2$ (%) |
|---|---|---|---|
| CR1 (%) | 47.0 | 51.0 | 7.5 |
| CR2 (%) | 58.6 | 61.9 | 7.8 |
| CR3 (%) | 52.6 | 57.5 | 10.3 |
| CR4 (%) | 50.0 | 52.2 | 4.4 |
| CR5 (%) | 78.4 | 79.0 | 2.9 |

first criteria, especially for CR3, which means that the proposed model version provides better simulations of low flows. This improvement is related to the introduction of the percolation function into the model and the calibration of a fourth parameter (SMA reservoir capacity). For CR4 and CR5, the improvement is less significant. The low improvement obtained for the balance criterion CR5 is not surprising since conceptual rainfall-runoff models generally manage to get quite a good water balance. Here we have considered that the results of Table 2 were satisfactory enough to accept the modifications of the GR3J model and justify the calibration of a fourth parameter in the model.

### *3.3. Comparison with previous studies of model improvement*

As a comparison, we calculated in Table 3, the $r^2$ statistics from the mean simulation results in the studies by Tan and O'Connor (1996), Lindström et al. (1997) and Senbeta et al. (1999), that all aimed at improving some models. The first authors proposed a modified version, SMARY, of the SMAR model and quantified the gain on four catchments in Tanzania, Japan, Australia and China. The second authors gave a new version, HBV-96, of the HBV model that they tested on seven catchments in Sweden. The last authors proposed two new versions of the PDISC model that were evaluated on six catchments from different locations in Bangladesh, China, Nepal and Australia. All three studies used the Nash–Sutcliffe criterion as efficiency index.

At first sight, the results of the above studies appear more satisfactory than ours. Note, however, that they were obtained on a limited number of tests, whereas the test of the new GR4J model was carried out on a sample of 429 catchments. If we look at the performances obtained on the different catchments with GR3J and GR4J models, we can notice that the range of $r^2$ statistics obtained in simulation is quite large. The distributions of $r^2$ statistics for the 429 catchments with CR1 and CR3 criteria are shown in Fig. 6 (we used the mean performance obtained on each catchment). First it can be seen that the GR4J version does not produce improvements in all cases. For the CR1 criterion, that puts emphasis on floods, performances are improved ($r^2 > 0$) in 68% of the tests and significantly so ($r^2 \geq 10\%$) in 34%. In contrast, they are significantly debased ($r^2 \leq -10\%$) in 7% of the tests. For the CR3 criterion, that puts more emphasis on low flows, performances are improved in 71% and significantly so in 47% of the tests, but are significantly debased for only 9% of the tests. These results indicate that in some small sub-samples of catchments, the improvement of the results may be more significant than that shown in Table 2 for the 429 catchments.

## 4. Discussion of model development

The development of the GR4J model is based on some fundamental assumptions regarding the choice of a modelling approach, the assessment of the model structure and the determination of the maximum level of model complexity allowed. These aspects are discussed in the following sections.

### *4.1. Methodology for model development and improvement*

The GR4J model and its former versions have been developed mostly along empirical lines. Empiricism is understood here in the sense of a modelling process

Table 3
Values of $r^2$ statistics calculated with the mean results (Nash–Sutcliffe criterion CR1) obtained in simulation mode in three different studies

| Authors | Number of tests | Mean efficiency of ref. model (%) | Mean efficiency of version #1 (%) | Mean efficiency of version #2 (%) | $r^2$ (#1) (%) | $r^2$ (#2) (%) |
|---|---|---|---|---|---|---|
| Tan and O'Connor (1996) | 4 | 49.8 | 62.7 | | 25.6 | |
| Lindström et al. (1997) | 7 | 83.6 | 87.4 | | 23.2 | |
| Senbeta et al. (1999) | 6 | 78.0 | 80.8 | 80.6 | 12.7 | 11.8 |

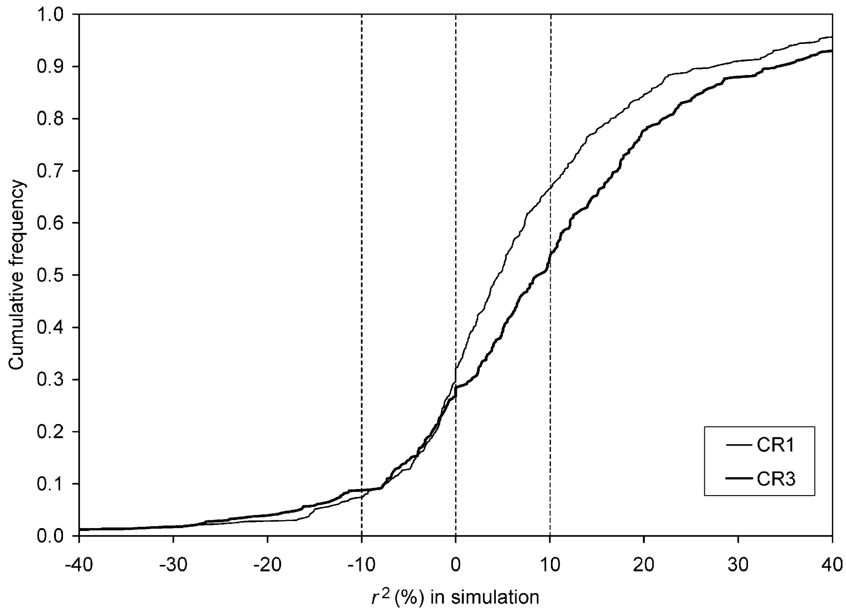


Fig. 6. Distribution of $r^2$ statistics (relative measure of performance modification) computed with the Nash–Sutcliffe criterion values obtained on the sample of 429 catchments when going from the GR3J to the GR4J model.

that combines correct mathematical operators to yield the best estimate of the observed output of the actual rainfall-runoff transformation. GR4J could be included among the hybrid metric-conceptual models as defined by Wheater et al. (1993) or Young (2001).

This mode of development differs from most other rainfall-runoff models whose construction relies on some prior perceptions of the processes dominating the rainfall-runoff transformation (see among others Jayawardena and Zhou, 2000; Joukainen, 2000).

The empirical approach follows the methodology encouraged by Nash and Sutcliffe (1970), who advised starting from very simple models and searching for the best ways to improve their efficiency by testing several model modifications and keeping the most satisfactory one. These authors say that they “are prepared to accept additional parts and hence greater difficulty in determining parametric values only if increased versatility of the model makes it much more likely to obtain a good fit between observed and computed output”. This modelling process is also partly advocated by Jakeman et al. (1994) with the system identification approach that begins with simple assumptions and evaluates refinements by confrontation to observations. We agree with several authors, e.g. Klemeš (1982), who think that the “striving for improvement of an empirical model in the absence of additional information tends to be scientifically sterile and to have an extremely low benefit/cost ratio from the point of view of applications”. Here, this additional information is provided by the large spectrum of catchments used to check the relevance of model improvements. The usefulness of this trial and error approach is also acknowledged by Bergström (1991) and Lindström et al. (1997) who used it in the development of the widely applied HBV model.

In the case of the GR4J model, efforts have been made to avoid introducing misconceptions due to the apparently trivial task of applying well-known hydraulic theories with unknown boundary conditions. In developing the GR4J model, we asked such a question: what could be done if we tried to mimic catchment behaviour using a single reservoir (see Michel, 1983)? A large number of very simple formulations were tested, in turn, to try to mimic reality. In this study, we tested a total of 235 modified versions of the GR3J model before selecting the most satisfactory one presented at the beginning of this article.

A good example of empirical model modification is the water exchange function that was introduced in the GR3J model structure by Nascimento and Michel (1992). Conceptually, this function corresponds to the fact that most catchments are not isolated bodies, ideally underlain by an impervious substratum. Initially the function was introduced in the model as an ad hoc solution to solve the problem of ephemeral catchments. The proposed function eventually proved very valuable for establishing a satisfactory water balance on all types of catchments. Conversely, removing it from the model leads to significant performance losses.

### *4.2. Use of large catchment samples for empirical model development*

The empirical approach followed to develop GR4J relies on large hydrological data sets that are used as the only referees to accept or reject model specifications. The test sample must include a large number of catchments (typically a few hundred) with various climate conditions, which is possible thanks to the high computing power quite easily available today.

Furthermore, with a large sample of catchments, it is possible to assess the versatility of models: a model is considered more versatile (and reliable) if it performs well on a wide range of climate and catchment conditions. This is explicitly or implicitly acknowledged by the many hydrological studies that assessed models in very different conditions (see among others WMO, 1975; Chiew and McMahon, 1994; Chiew et al., 2002). A model that is able to perform well for very different climate conditions can be considered as a good safeguard against model failure due to natural climate variability. For example, in the spring and summer of 1976 in France, there was a six-month period of drought with almost zero rainfall, making climate conditions similar to those found in tropical zones. We believe that a model that can perform satisfactorily in semi-arid conditions is less likely to fail in 1976 in France than a model that does not behave well in such conditions.

Last, large test samples are the only way to make sure that improvements are significant. We join the point of view of Andersson (1992) who mentions that “a certain change of model structure can improve the model performance on some basins whereas it is unchanged or deteriorated for other basins. (…) It is therefore important to test the new model for a large set of basins (…) before drawing conclusions of a general model improvement”.

### *4.3. Determining the appropriate level of complexity*

Several studies on rainfall-runoff modelling have discussed the influence of complexity (understood here as the number of optimised parameters) on model efficiency (see e.g. Jakeman and Hornberger, 1993; Kokkonen and Jakeman, 2001; Perrin et al., 2001a). It has been demonstrated that the increase in the number of model parameters could lead to overparameterisation and ill-conditioned problems during optimisation. Perrin et al. (2001a) showed that a small number of parameters (three to five), is sufficient to produce satisfactory results of streamflow simulations at a daily time step provided that the model structure is accurately built.

Bergström (1991) states that “optimal complexity is the point we should try to reach for a specific problem and it is not difficult to find examples of too complex models which have not contributed significantly to the solution of the problem”. Kokkonen and Jakeman (2001) add that “the determination of the appropriate level of model complexity is far from straightforward”. Here, for the modification of the GR3J model, the adequate level of complexity was the one that allowed the best performance in simulation mode. It was determined by testing many different model versions with different numbers of parameters. In total, 235 different model structures derived from the simple GR3J model were tested, in turn, on the 429 catchments and assessed with the set of performance criteria. The number of free parameters in these model versions was changed either by freeing or fixing some parameters in the mathematical functions or by adding or removing some parts of the model. The level of complexity of the tested versions was ranging between zero (no calibrated parameters, i.e. a single model for all catchments) and six optimised parameters.

Fig. 7 shows the performances obtained in simulation mode on the test catchments for criteria CR1, CR3 and CR5 (similar results were obtained for

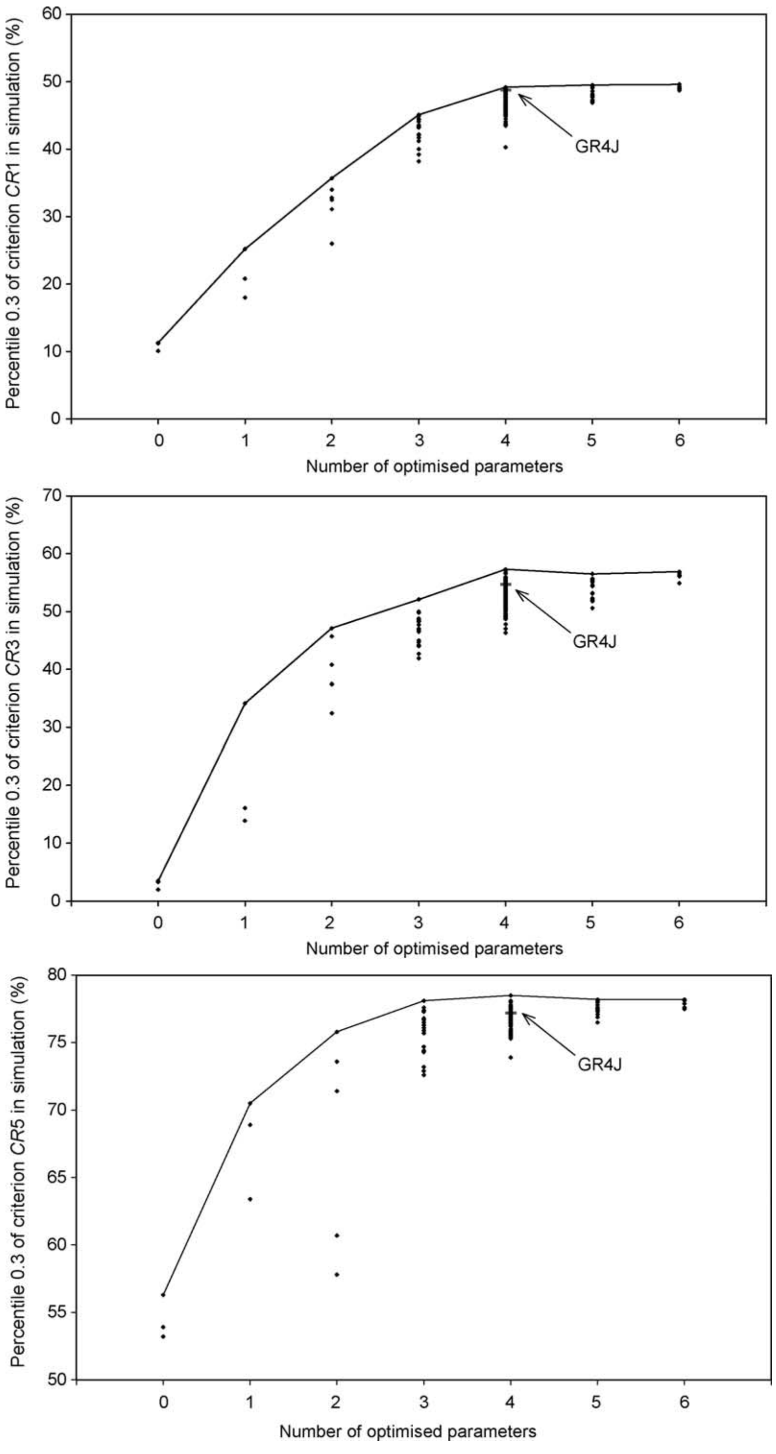


Fig. 7. Percentiles 0.3 of the distributions of results in simulation for the 235 model structures with 0–6 optimised parameters (the curve joins the best performances).

the other criteria). We used the percentiles 0.3 of the distribution of results because mean performances are less representative when model complexity decreases due to negative efficiency values. The graphs indicate that the performances increase rapidly when the number of parameters increases from zero to three. A fourth parameter still brings some improvement, albeit small in the case of the balance criteria (CR5). Performances then reach a plateau when a fifth and a sixth parameters are added and show no significant gain. The maximum performance even decreases slightly in the case of criteria CR3 and CR5. Hence, at a daily time-step, four parameters seem to be the most reliable level of complexity, confirming that a parsimonious model is sufficient to simulate catchment behaviour, as also argued by Kokkonen and Jakeman (2002) among others.

Note, however, that among the tested four-parameter structures, the proposed GR4J model is not the best performing one according to all the performance criteria, as shown by the arrows in Fig. 7. The choice of this particular structure is, in fact, a compromise between several modelling qualities highlighted by the five evaluation criteria.

## 5. Conclusion

In this article, we presented an improved version of the daily lumped rainfall-runoff GR4J model. This model was developed to get the best average results on a large sample of catchments, as a guarantee of model robustness. The model structure is concise with a small number of parameters. The test of the model showed that, compared to other models of the same type, its results are statistically among the best. The improvement made in model efficiency is significant in comparison with the previous model version, especially with respect to low-flow simulation. These results suggests that improvements can be achieved when the model structure remains simple.

The GR4J model was developed along empirical lines, questioning every part of the structure. Only four free parameters were kept in the model. This parsimony was not a specific objective of our research but an interesting by-product of the stringent modelling approach adopted here.

During our research, we found that the choice of the model structure was incredibly difficult due to the lack of knowledge of how the catchment behaves as a whole, which may be a 'handicap of empirical modelling' as remarked by Klemeš (1982). Several functions appeared equivalent in terms of model efficiency and the final choice often leaned towards the simplest formulations. It is likely that the model version proposed here will evolve in the future, to further improve its reliability. As progress is made, a better definition of the model structure may be reached. Complementary ways will have to be explored to make other significant advances. The complementarity between models could also define new ways of model improvement, exploring model structure with different scales of temporal lumping (see Mouelhi, 2003).

The improved reliability of models like GR4J heightens confidence in their use within a range of hydrological applications where water engineers are looking for reliable tools. The GR4J model, coupled in some cases with a stochastic rainfall model, has already been applied to medium or extreme flood assessment (Perrin and Michel, 2002), reservoir design, reservoir management (Yang et al., 1991, 1995), long-term drought forecasting (Perrin et al., 2001b), short-term flood forecasting (Yang and Michel, 2000). It has been argued that models of this type cannot be used for other applications such as assessing the impacts of land-use and climate change, or water quality management, on the grounds that they lack physical foundation. Although it is acknowledged that conceptual lumped rainfall-runoff models are far from being able to tackle satisfactorily the formidable problem of assessing the consequences of climate change, models like the GR4J seem to be the best suited to detecting changes in a basin behaviour (Andréassian, 2002), as opposed to physically based models that incorporate pre-arranged answers. When we know enough about the consequences of already identified phenomena, it will be possible to establish correlative rules that would enable engineers to predict the result of the most commonly observed land-use changes. Last, regarding water quality problems, it seems wise to start with a simple model that produces good streamflow predictions, since pollutants and suspended-matter fluxes are closely related to flow. Attempts have already been made to combine simple models and nitrate transport modules, e.g. by Ma et al. (1990) and Van Herpe (2000) on the basis of the GR3J model and TOPMODEL, respectively. We believe that a fruitful way forward would be to develop empirically, bit by bit, the additive components needed to produce other outputs such as chemical-substance concentrations. Such an approach is likely to be more effective in the long-term than hoping that all problems will be solved simultaneously once a perfect physically based model has been developed.

## Acknowledgements

The authors thank the two anonymous reviewers for their suggestions which helped to improve the text. ENGEES (Ecole Nationale du Génie de l'Eau et de l'Environnement de Strasbourg) is thanked for its support of this study. The authors thank researchers and instutions who provided data sets for model testing: Dr Francis H.S. Chiew at the Department of Civil and Environmental Engineering of the University of Melbourne, Australia, for data sets of the Australian catchments; Dr Eric Servat at the Institut de Recherche pour le Développement (formerly ORSTOM) in Montpellier, France, for data sets of catchments in the Ivory Coast; Jane L. Thurman from the Water Data Center at the US Department of Agriculture, Beltsville, United States, for data sets in the US (ARS data base); Dr Nilo Oliveira Nascimento at the University of Minas Gerais, Belo Horizonte, Brazil for data of Brazilian catchments; Dr John Schaake and the MOPEX team (see Web site www.nws.noaa.gov/oh/mopex) for data sets of American catchments; the HYDRO database of the French Ministry for the Environment and Météo France for hydrological and meteorological data sets of the French catchments.

## References


Andersson, L., 1992. Improvements of runoff models. What way to go? Nordic Hydrology 23, 315–332.

Andréassian, V., 2002. Impact de l'évolution du couvert forestier sur le comportement hydrologique des bassins versants. PhD Thesis, Université Pierre et Marie Curie Paris VI, Cemagref, France, 276 pp.

Bergström, S., 1991. Principles and confidence in hydrological modelling. Nordic Hydrology 22, 123–136.

Beven, K., 1986. Hillslope runoff processes and flood frequency characteristics. In: Abrahams, A.D., (Ed.), Hillslope Processes, pp. 187–202.

Boyle, D.P., Gupta, H.V., Sorooshian, S., Koren, V., Zhang, Z., Smith, M., 2001. Toward improved streamflow forecasts: value of semidistributed modeling. Water Resources Research 37 (11), 2749–2759.

Chiew, F., McMahon, T., 1994. Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian catchments. Journal of Hydrology 153, 383–416.

Chiew, F.H.S., Stewardson, M.J., McMahon, T.A., 1993. Comparison of six rainfall-runoff modelling approaches. Journal of Hydrology 147, 1–36.

Chiew, F.H.S, Peel, M.C., Western, A.W., 2002. Application and testing of the simple rainfall-runoff model SIMHYD. In: Singh, V.P., Frevert, D.K. (Eds.), Mathematical Models of Small Watershed Hydrology and Applications, Water Resources Publications, Highlands Ranch, pp. 335–367.

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. Hydrological Sciences Journal 44 (2), 263–277.

Georgakakos, K.P., Baumer, O.W., 1996. Measurement and utilization of on-site soil moisture data. Journal of Hydrology 184, 131–152.

Jakeman, A.J., Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model ? Water Resources Research 29 (8), 2637–2649.

Jakeman, A.J., Post, D.A., Beck, M.B., 1994. From data and theory to environmental model: the case of rainfall runoff. Environmetrics 5, 297–314.

Jayawardena, A.W., Zhou, M.C., 2000. A modified spatial soil moisture storage capacity distribution curve for the Xinanjiang model. Journal of Hydrology 227, 93–113.

Joukainen, S., 2000. Improving the calculation of potential evapotranspiration of the HBV model: application to the Ounasjoki watershed. In: Nilsson, T., (Ed.), Nordic Hydrological Conference 2000, Uppsala, Sweden, pp. 347–354.

Klemeš, V., 1982. Empirical and causal models in hydrology. In: National Research Council Geophysics Study Commitee, (Ed.), Scientific Basis of Water-Resource Management, National Academy Press, Washington, DC, pp. 95–104.

Kokkonen, T.S., Jakeman, A.J., 2001. A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications. Water Resources Research 37 (9), 2345–2352.

Kokkonen, T.S., Jakeman, A.J., 2002. Structural effects of landscape and land use on streamflow response. In: Beck, M.B., (Ed.), Environmental Foresight and Models: A Manifesto, Elsevier Science, Amsterdam, pp. 303–321.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. Journal of Hydrology 201, 272–288.

Loague, K.M., Freeze, R.A., 1985. A comparison of rainfall-runoff modeling techniques on small upland catchments. Water Resources Research 21 (2), 229–248.

Loumagne, C., Michel, C., Palagos, B., Baudez, J.C., Bartoli, F., 1999. D'une approche globale vers une approche semi-distribuée en modélisation pluie-débit (From a global to a semi-distributed approach in rainfall-runoff modelling). La Houille Blanche (6), 81–88.

Ma, Z.C., Leviandier, T., Ferry, M., 1990. Fitting a conceptual hydrologic model accounting for nitrate loss in an agricultural representative basin. Proceedings of the Conference on Hydrological research basins and the environment. Verslagen en Mededelingen, Commissie voor hydrologish onderzoek TNO no 44, pp. 319–327.

Michaud, J., Sorooshian, S., 1994. Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. Water Resources Research 30 (3), 593–605.

Michel, C., 1983. Que peut-on faire en hydrologie avec un modèle conceptuel à un seul paramètre ? La Houille Blanche (1), 39–44.

Mouelhi, S., 2003. Vers une chaîne cohérente de modèles pluie-débit conceptuels globaux aux pas de temps pluriannuel, annuel, mensuel et journalier. PhD Thesis, ENGREF, Paris/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. PhD Thesis, CERGRENE/ENPC, Paris, France, 550 pp.

Nascimento, N.O., Michel, C., 1992. Some epistemological aspects of the development and use of hydrologic conceptual models. In: Graillot, D., Schilling, W., Einfalt, T. (Eds.), Proceedings of the Fourth Junior Scientist Course Assessment of Modelling Uncertainties and Measurement Error in Hydrology, St-Etienne (France), pp. 245–264.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I—A discussion of principles. Journal of Hydrology 27 (3), 282–290.

Perrin, C., 2000. Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative. PhD Thesis, INPG (Grenoble)/Cemagref (Antony), France, 530 pp.

Perrin, C., Michel, C., 2002. Robustness of two flood estimation methods with data availability. In: Spreafico, M., Weingartner, R., Dobmann, J., Hodel, H., Hunziker, S. (Eds.), Proceedings of the International Conference on Flood Estimation, Berne, Switzerland, 6–8 March 2002, Report of CHR no. II-17, in press.

Perrin, C., Michel, C., Andréassian, V., 2001a. Does a large number of parameters enhance model performance ? Comparative assessment of common catchment model structures on 429 catchments. Journal of Hydrology 242 (3–4), 275–301.

Perrin, C., Michel, C., Andréassian, V., 2001b. Long-term low flow forecasting for French rivers by continuous rainfall-runoff modelling. Meeting of the British Hydrological Society on Continuous River Flow Simulation. BHS Occasional Paper No. 13, Littlewood I.G. (Ed.), Wallingford, UK, 5th July 2001, pp. 21–29.

Refsgaard, J.C., Knudsen, J., 1996. Operational validation and intercomparison of different types of hydrological models. Water Resources Research 32 (7), 2189–2202.

Seibert, J., 2001. On the need for benchmarks in hydrological modelling. Hydrological Processes 15 (6), 1063–1064.

Senbeta, D.A., Shamseldin, A.Y., O'Connor, K.M., 1999. Modification of the probability-distributed interacting storage capacity model. Journal of Hydrology 224, 149–168.

Sugawara, M., 1995. Tank model. In: Singh, V.P., (Ed.), Computer Models of Watershed Hydrology, Water Resources Publications, pp. 165–214, Chapter 7.

Tan, B.Q., O'Connor, K.M., 1996. Application of an empirical infiltration equation in the SMAR conceptual model. Journal of Hydrology 185, 275–295.

Van Herpe, Y., Troch, P.A., De Troch, F.P., 2000. Applicability of a parsimonious nitrate transport model in catchment management. In: V. Elias, I.G. Littlewood (Eds.), Catchment hydrological and biochemical processes in the changing environment, Seventh Conference of the European Network of Experimental and Representative Basins. Technical Documents in Hydrology No. 37. IHP/UNESCO, Liblice (Czech Republic), 22–24 September 1998, pp. 37–46.

Wheater, H.S., Jakeman, A.J., Beven, K.J., 1993. Progress and directions in rainfall-runoff modelling—Chapter 5. In: Jakeman, A.J., Beck, M.B., McAleer, M.J. (Eds.), Modelling Change in Environmental Systems, Wiley, pp. 101–132.

WMO, 1975. Intercomparison of Conceptual Models Used in Operational Hydrological Forecasting, World Meteorological Organization, Geneva, Switzerland.

Yang, X., Michel, C., 2000. Flood forecasting with a watershed model: a new method of parameter updating. Hydrological Sciences Journal 45 (4), 537–546.

Yang, X., Parent, E., Michel, C., Roche, P.A., 1991. Gestion d'un réservoir pour la régularisation des débits. La Houille Blanche (6), 433–440.

Yang, X., Parent, E., Michel, C., Roche, P.A., 1995. Comparison of real-time reservoir-operation techniques. Journal of Water Resources Planning and Management 121 (5), 345–351.

Ye, W., Bates, B.C., Viney, N.R., Sivapalan, M., Jakeman, A.J., 1997. Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments. Water Resources Research 33 (1), 153–166.

Young, P., 2001. Data-based mechanistic modelling and validation of rainfall-flow processes. In: Anderson, M.G., Bates, P.D. (Eds.), Model Validation: Perspectives in Hydrological Science, Wiley, Chichester, pp. 117–161.

Zhang, Z., Koren, V., Smith, M., Reed, S., Wang, D., 2003. Use of NEXRAD data and basin disaggregation to improve continuous hydrograph simulations. Journal of Hydrologic Engineering, in press.