Exploring the relationship between hydroclimatic stationarity and rainfall-runoff model parameter stability: A case study in West Africa

Matthieu Le Lay, Sylvie Galle, Georges-Marie Saulnier, Isabelle Braud

To cite this version:

Matthieu Le Lay, Sylvie Galle, Georges-Marie Saulnier, Isabelle Braud. Exploring the relationship between hydroclimatic stationarity and rainfall-runoff model parameter stability: A case study in West Africa. Water Resources Research, American Geophysical Union, 2007, 43, pp.W07420. 10.1029/2006WR005257 . insu-00386863

HAL Id: insu-00386863

https://hal-insu.archives-ouvertes.fr/insu-00386863

Submitted on 5 Mar 2021

WATER RESOURCES RESEARCH, VOL. 43, W07420, doi:10.1029/2006WR005257, 2007

Exploring the relationship between hydroclimatic stationarity and rainfall-runoff model parameter stability: A case study in West Africa

M. Le Lay,[1] S. Galle,[1] G. M. Saulnier,[1] and I. Braud[2]

Received 19 June 2006; revised 8 March 2007; accepted 30 March 2007; published 13 July 2007.

[1] Forecasting the hydrological impacts of climatic and anthropic changes assumes that the evolution of model parameters under changing conditions can be predicted. Hence it is necessary to study the relationship between hydroclimatic variability and model parameter values. In this paper, we explore this issue by implementing a daily lumped hydrological model (GR4J, Perrin et al. (2003)) on the Upper Ouémé watershed (10,050 km^2, Benin). West Africa was subjected to changing climatic and hydrological conditions during the second half of the last century, and changes in the water balance can be evidenced on this watershed. Contrasted periods are extracted from the available 1954–2002 data set, so that hydrological and pluviometric extreme periods can be defined. First, the magnitude of changes in model parameter values under changing conditions are analyzed, using a resampling method (first approach) and within an equifinality context (second approach). It is shown that significant changes in the rainfall-runoff relationship do not induce significant changes in the model's parameter values. A third original approach analyzes the signature of hydroclimatic variability in model performance. Hence a test is defined that uses interannual model efficiency variances to measure performance homogeneity and a resampling test to statistically characterize the calculated results. This test demonstrates the hydrological relevance of the calibrated parameter sets because the more stationary the rainfall-runoff relationship, the more homogeneous the model's performance.

Citation: Le Lay, M., S. Galle, G. M. Saulnier, and I. Braud (2007), Exploring the relationship between hydroclimatic stationarity and rainfall-runoff model parameter stability: A case study in West Africa, *Water Resour. Res.*, *43*, W07420, doi:10.1029/2006WR005257.

1. Introduction

[2] Changes in water resources are an increasing matter of concern because of potential consequences of climate nonstationarity and anthropic pressure. In such conditions, the watershed hydrological response is likely to change, and the sustainability of water supplies is questioned, especially in dry areas. In West Africa for instance, the long-lasting drought experienced since the 1970s has had dramatic consequences on water resources and food security. Therefore forecasting the hydrological impacts of climatic and land-use changes has received increasing attention in the last few years. However, it must be recognized that hydrologists are unable to provide satisfactory responses to this question because of uncertainties remaining in the use of hydrological models in nonstationary conditions.

[3] In this context, the detection of watershed behavior changes once they have occurred may be considered an intermediate objective for a better understanding of hydrological system responses to changing conditions. Existing methods to meet this challenge are based either on paired watershed approaches (see the work of *Brown et al.* [2005] for a recent review) or on rainfall-runoff models [e.g., *Lorup et al.*, 1998; *Schreider et al.*, 2002; *Andréassian et al.*, 2003]. Indeed, once it has been calibrated, a model can be considered as a virtual control watershed because it simulates stationary hydrological behavior. Changes occurring on the watersheds studied can therefore be analyzed through statistical trend techniques on simulation residuals.

[4] A complementary issue concerns the relationship between hydrological changes and model parameter changes as the capacity to forecast the hydrological impacts of future changes depend on this relationship. Although this issue is essential, there is a clear lack of research in this area. To the knowledge of the authors, the study of *Niel et al.* [2003] is the main investigation to date. These authors calibrated a rainfall-runoff model on several West African watersheds, before and after the occurrence of the long-lasting rainfall deficit in the region. The optimal parameter values for each calibration period, associated with a confidence region, were compared. From the results, *Niel et al.* [2003] concluded that nonstationarity in rainfall or runoff series

---

[1] Laboratoire d'étude des Transferts en Hydrologie et Environnement, Université de Grenoble, [CNRS, INPG, IRD, UJF], Grenoble, France.

[2] Cemagref, Lyon, France.

0043-1397/07/2006WR005257

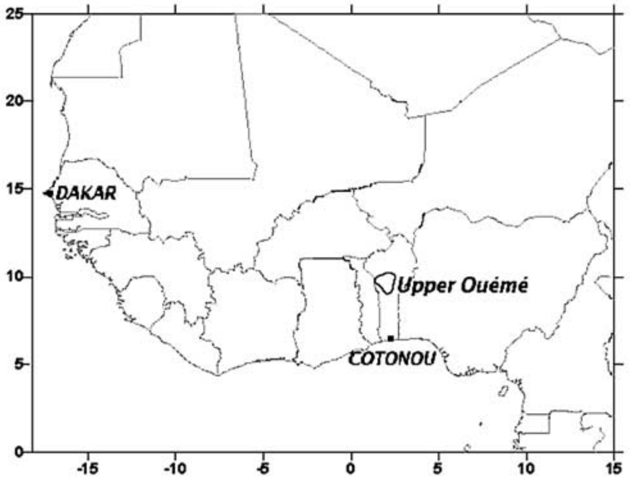

Figure 1. Location of the Upper Ouémé watershed in West Africa.

does not imply nonstability of the model parameters and therefore does not imply variability in the hydrological behavior of watersheds. However, this conclusion relies on the hypothesis that "parameter stability can be translated into hydrologic stability." Aware of the weakness of this hypothesis, they finally conclude that "to judge the relevance of the proposed approach, basins characterized by significant known changes, in land use for instance, could be used to estimate the influence of these changes on the parameter variations." Indeed, the wide use of spatially lumped models and calibration to estimate model parameter values implies that the meaning of parameter values is difficult to define. As such values are the result of an inverse problem that is often considered as ill posed [e.g., *Sorooshian and Arfi*, 1983; *Sorooshian and Gupta*, 1983; *Troutman*, 1983; *Beven*, 1993], they compensate all kinds of errors in the modeling process. It is observed that parameter values depend substantially on the forcing data set, in particular when climatic conditions display a great variability [*Gan and Burges*, 1990].

[5] Therefore this paper aims to extend the work of *Niel et al.* [2003] and explores the relationship between hydroclimatic stationarity and rainfall-runoff model parameter stability. For this purpose, we chose to study a watershed in West Africa, where change in the water balance before and after the long-lasting drought of the 1970s can be evidenced. This case study, as well as the rainfall-runoff model, are presented in section 2. Section 3 describes the methodologies proposed to address the issue of rainfall-runoff model parameter stability under nonstationary hydroclimatic conditions. In section 4, we present and discuss the results obtained from the case study.

2. Description of the Case Study: Data and Model

2.1. Data

[6] Over the last 50 years, West Africa has been subjected to significant rainfall variability, characterized both by large interannual fluctuations and by periods of long-lasting drought (see, for instance, the works of *Nicholson and Palao* [1993] or *Le Barbé et al.* [2002]). The analysis of the annual rainfall series displayed a statistical break around 1970 in most West Africa regions [*Hubert et al.*, 1989; *Paturel et al.*, 1997], defining a wet (1950–1970) and a dry (1971–1990) period. On the seasonal scale, these changes are associated with changes in the rainfall regime, as shown by *Le Barbé et al.* [2002], *Lebel et al.* [2003a, 2003b], and *Le Lay and Galle* [2005]. During the 1990s, the annual rainfall depths were more contrasted, and drought persistence is now being questioned [*L'Hôte et al.*, 2003; *Ozer et al.*, 2003].

[7] This study focuses on the Upper Ouémé watershed, covering 10,050 km^2 in Benin (1.5–2.5°E, 9–10°N; Figure 1). It forms part of one of the African Monsoon Multidisciplinary Analysis international program windows, on which atmospheric and continental interactions are being investigated [*Lebel et al.*, 2003a, 2003b]. The hydrology of the Upper Ouémé catchment is also investigated within the IMPETUS project www.impetus.unikoeln.de). Several modeling studies have been carried out [e.g., *Bormann and Diekkrüger*, 2003; *Bormann et al.*, 2005; *Varado et al.*, 2006]. Situated within the Sudanian climatic regime, this area is characterized by a single rainy season, with an average rainfall of 1200 mm spread between April and October. The streamflows are intermittent, with river discharge occurring between the end of June and January. Daily series of rainfall and discharge were collected throughout the period 1954–2002, excluding years 1971, 1980, and 1989 for which data are missing (see the work of *Le Lay and Galle* [2005] for details on data processing). Mean daily rainfall values on the watershed were obtained by kriging from 12 raingauges, using a semivariogram derived from the dense network (40 raingauges) available since 1998. The high hydroclimatic variability of the region studied is presented in Figure 2, through rainfall and runoff indexes. Although rainfall changes that occurred during the 1970s and 1980s are less significant than those that occurred in the Sahel [*Tapsoba et al.*, 2004], the rainfall index clearly underlines the existence of a long-term rainfall deficit over this period. Meteorological data are not available for the entire period; daily Penman-Monteith potential evapotranspiration (PE) calculated over the 1997–2002 period was used to estimate a long-term average of mean daily potential evapotranspiration on the watershed (annual amount is

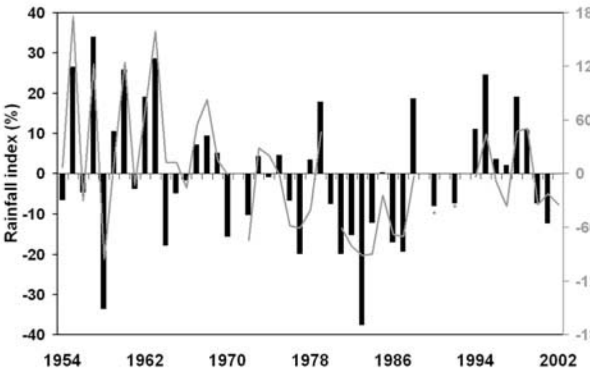

Figure 2. Rainfall and runoff index on the Upper Ouémé watershed.

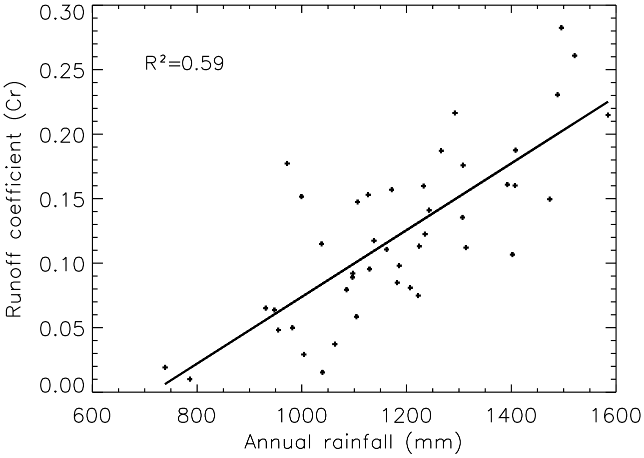

Figure 3. Correlation between annual runoff coefficient and annual rainfall on the Upper Ouémé watershed from 1954 to 2002.

approximately 1500 mm). Although the use of such a climatological PE is frequent in rainfall-runoff modeling, the context of hydroclimatic variability makes this assumption questionable. In particular, neglected interannual PE variability may have impacts on model results. These aspects are discussed latter in section 4.4.

2.2. Stationarity Analysis

[8] In conjunction with the climatic nonstationarity described above, the Upper Ouémé watershed is experiencing a substantial growth in anthropic pressure because of the significant increase in population [*Deichmann*, 1996] and land-use changes in the region. According to the *Food and Agricultural Organization* [2003], an annual deforestation rate of 2.3% (1990–2000) is observed on the investigated area. Therefore changes in the hydrological behavior of the watershed are likely to occur. Many different variables can be considered to characterize changes in hydrological behavior at the watershed scale. As PE variability cannot be documented, this study focuses on the rainfall-runoff relationship; consequently, evapotranspiration changes due to land cover changes or PE changes are not distinguished. Then, the annual runoff coefficient $C_r = R / P$ (with R and P the annual runoff and rainfall depth, respectively) is used to define the annual hydrological yield of the watershed.

[9] Because of the changes in rainfall conditions observed on the 1954–2002 period, C_r is closely related to the rainfall depth (R^2 = 59%), as shown in Figure 3. Therefore it is necessary to decorrelate the rainfall-runoff relationship descriptor from the climatic conditions. Residuals of the C_r variable, defined as follows, were considered to evidence the nonstationarity of the rainfall-runoff relationship:

$$Res(C_r) = C_r - C_r^*$$

where C_r* is the linear estimator of C_r, such that $C_r* = a*P + b$ (with a and b as linear regression coefficients). The choice of such a simple model for the annual yield coefficient stems from our limited knowledge of the water balance, which prevents the use of finer models [e.g., *Milly*, 1994]. Figure 4 shows Res(C_r) for the 1954–2002 period. The annual hydrological yield of the watershed displays a significant decrease, with a clear shift around 1970. This feature may be statistically characterized, using break detection methods (see the work of *Kundzewicz and Robson* [2004] for a review and caveats). Among the possible tests, the Hubert segmentation procedure [*Hubert et al.*, 1989] was applied on Res(C_r). Thanks to a specific algorithm, this technique provides one or several break dates (or possibly none) which separate contiguous segments whose means are significantly different in terms of the Scheffé test [*Scheffé*, 1959]. The results display only one significant break of the rainfall-runoff relationship, with a maximum probability between 1970 and 1975.

[10] It is beyond the scope of this paper to explain these hydrological changes. However, the following several reasons may be suggested: (1) the observed changes in rainfall regimes [*Le Lay and Galle*, 2005], which are likely to modify the runoff generation processes; (2) an increase in air temperature, inducing a PE increase; and (3) land-cover changes, due to rainfall deficit and/or growth of anthropic pressure. It is worth noting that other watersheds of sub-Sahelian West Africa have experienced the same hydrological yield decrease since the 1970s [*Mahé et al.*, 2005]. Actually, the Upper Ouémé watershed offers a great data set of nonstationary hydroclimatic conditions and provides an excellent opportunity to explore the relationship between hydroclimatic variability and model parameter instability.

2.3. GR4J Rainfall-Runoff Model

[11] The methodologies described in this article can be implemented with any conceptual rainfall-runoff model (here "conceptual" means that it requires calibration). Obviously, the capacity to detect hydrological changes will depend to a large extent on the efficiency and the robustness of the model used. Furthermore, model complexity supported by daily rainfall-runoff data is very limited [e.g., *Jakeman and Hornberger*, 1993]. It is therefore recommended to avoid overparametrized modeling structures, as they tend to lack robustness.

[12] For this study, we used the GR4J model [*Perrin et al.*, 2003]. This is a reliable lumped model, which operates on a

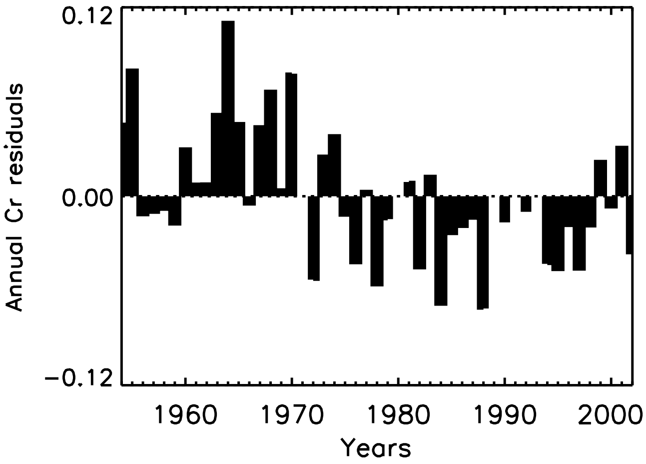

Figure 4. Residuals of runoff coefficients on the Upper Ouémé watershed from 1954 to 2002.

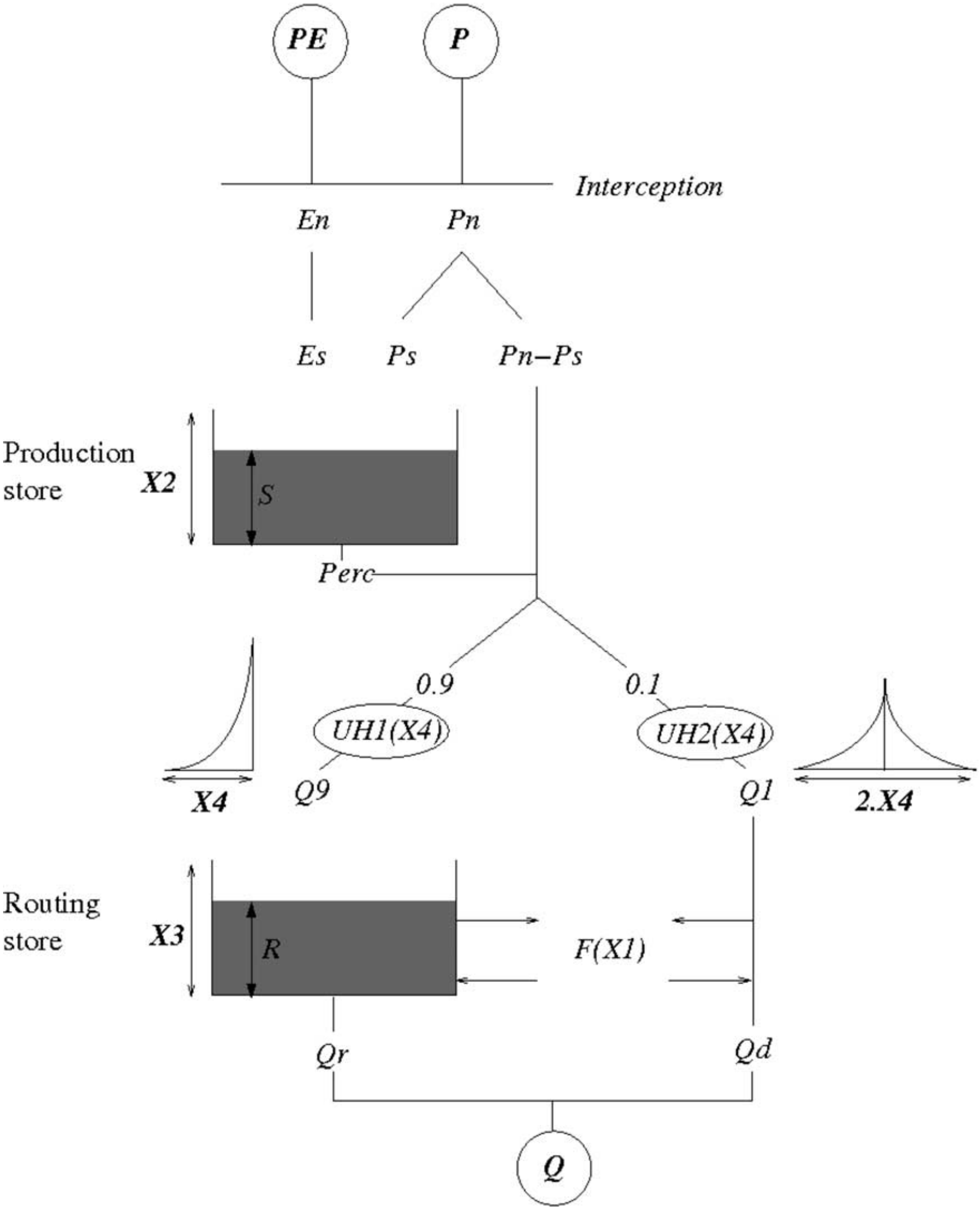

Figure 5. Diagram of the GR4J rainfall-runoff model (after *Perrin et al.*, 2003): P and PE are the forcing data; X_1, X_2, X_3, and X_4 are the four free parameters; internal state variables and fixed parameters are also represented.

daily basis. Figure 5 shows a synthetic diagram of the model. Its structure is composed of two stores and four free parameters, which account for water balance (groundwater exchange coefficient, X_1; maximum capacity of the production store, X_2) and water transfer (maximum capacity of the nonlinear routing store, X_3; unit hydrograph time base, X_4). The input variables are areal daily rainfall and potential evapotranspiration. After subtraction of water losses (evapotranspiration and interception), the remaining water amount is transferred by a nonlinear routing store and a linear routing with unit hydrographs. The total streamflow is the addition of quick and slow flow components. A detailed description of the model is beyond the scope of this paper; the reader can refer to the work of *Perrin et al.* [2003] for a complete discussion. However, it has to be noticed that the GR4J model was developed along empirical lines, questioning every part of the structure. This approach follows the methodology encouraged by *Nash and Sutcliffe* [1970], who were “prepared to accept additional parts and hence greater difficulty in determining parametric values only if

Table 1. Results of the Split-Sample Test on the 1998–2002 Period

	1998–1999 (Calibration)	2000–2002 (Validation)
E	0.89	0.84
R^2	0.89	0.88
B	−0.002	0.09

increased versatility of the model makes it much more likely to obtain a good fit between observed and computed output.''

3. Methods

3.1. Model Implementation

[13] Model calibration is based on a Monte Carlo sampling with a uniform prior distribution. The popular Nash and Sutcliffe efficiency [*Nash and Sutcliffe*, 1970] has been used as the objective function, and may be expressed as:

$$E = 1 - \sigma^2_{\mathrm{mod}}/\sigma^2_{\mathrm{obs}} \tag{1}$$

where σ^2_{mod} and σ^2_{obs} are the model error variance and the observed variance, respectively, for the period under consideration.

[14] For model evaluation, two additional statistics were computed, the determination coefficient (R^2) and a water balance criterion (B) defined as follows:

$$B = \frac{\sum_{i=1}^{n} Q_{i,\mathrm{mod}}}{\sum_{i=1}^{n} Q_{i,\mathrm{obs}}} - 1 \tag{2}$$

where n is the number of days during the simulation period, $Q_{i,\mathrm{obs}}$ and $Q_{i,\mathrm{mod}}$ are the observed and simulated daily flows, respectively. Split sample tests were performed for model evaluation. Examples of simulation results for the 1998–2002 period are shown in Table 1 and in Figure 6. Model performance is quite good, with efficiency values up to 0.8 and bias on the volume of stream discharge under 10%. We therefore assume that the GR4J model correctly captures the watershed behavior.

3.2. Definition of Reference Periods

[15] In order to explore the relation between hydroclimatic stationarity and rainfall-runoff model parameter stability, we defined several reference periods considered as representative of different hydroclimatic conditions. Hence the 1954–2002 period was first restricted to the strongly contrasted 1954–1990 period, from which two pairs of data sets were extracted.

[16] The hydrological nonstationarity outlined in section 2.2 defines two 17-yearlong periods (1954–1970 and 1972–1990, hereafter called high-yielding (HY) and low-yielding (LY) periods) assumed to represent two different states of the watershed. However, we cannot speak of stationary periods, as stationarity would mean that the watershed is experiencing no significant land-use change, no significant long-term climate change, and no climate extremes. This is not the case, even if the changes during these periods are small compared to the changes between the two periods.

[17] The calibration process creates a statistical relation between model parameter values and the data set used to force the model. Thus, although the model's parameter values are expected to represent the hydrological behavior of a watershed, their dependence on climatic forcing conditions has to be questioned [e.g., *Gan and Burges*, 1990]. The rainfall variability makes it possible to define two extreme composite periods, one composed of the 17 wettest years of the 1954–1990 period (hereafter called the wet composite (WC) period) and the other, of the 17 driest years (hereafter called the dry composite (DC) period). Definition of composite periods is based on a fundamental hypothesis, the hydrological independence of each year. The latter is ensured for the case study by the intermittent nature of the water cycle, typical of many arid and tropical regions. Memory effects are then quite low, as suggested by the low rank 1 autocorrelation coefficient of annual runoff coefficients (r = 0.26). This hypothesis may appear too restrictive in the application of these methodologies. However, it is worth noting that numerous tropical or arid watersheds present this feature. Moreover, if long time series are available, one may define independent hydrological periods on most small or medium-sized watersheds, as done by *Andréassian et al.* [2003].

3.3. Applied Methodologies

[18] Different methodologies have been implemented to explore the relationship between hydroclimatic variability and model parameter variability. The two first approaches attempt to use changes in model parameter values as a signature of hydroclimatic variability. They can be resumed as follows: if model parameter values are representative of a

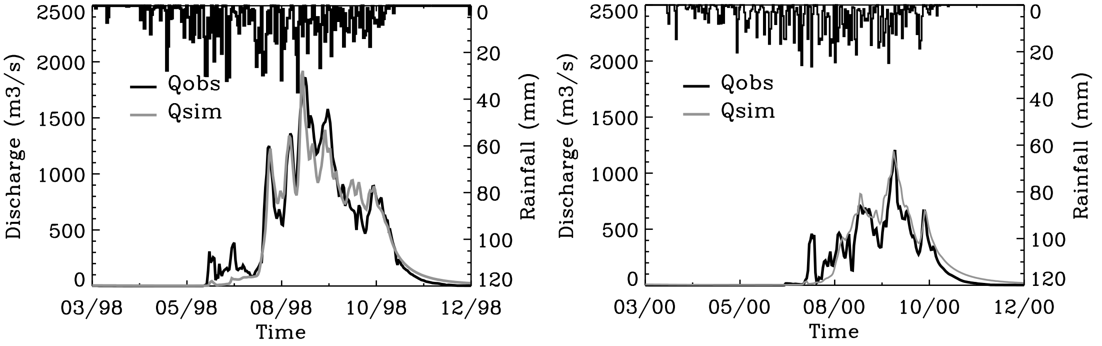

Figure 6. Example of observed and simulated daily hydrographs, for 1998 (left) and 2000 (right). Calibration period: 1998–1999.

watershed behavior, one may think that changes in parameter values will be maximum between two sharply contrasted hydrological periods, such as HY and LY periods. To investigate this hypothesis, changes in model parameters were analyzed using a resampling method (first approach) and the equifinality concept (second approach). The third implemented approach is deeply different and is based on an analysis of the signature of hydroclimatic variability in model performance.

3.3.1. Resampling Test on Optimal Parameter Values

[19] In the first method, changes in model parameters were analyzed using a resampling method. This type of technique is attractive because it is distribution-free and does not require any assumption on the distribution of the data [*Kundzewicz and Robson*, 2004]. The test proposed is based on the following two hypothesis: (1) the absence of a trend in hydrological behavior of the watershed on the 1954–1990 period (hereafter called the null hypothesis H_0); and (2) the hydrological independence of each year, discussed in section 3.2. According to these hypotheses, the chronological order of the observations is not important, and data can be shuffled many times. In what follows, the resampling covers the 34-yearlong 1954–1990 period (excluding years 1971, 1980, and 1989 for which data are missing). One hundred synthetic 17-year periods (P_i) were sampled, and complementary (and therefore independent) 17-year periods (P_i') were inferred. For each subsample, the model was calibrated, leading to new estimated parameter sets. To characterize the changes in optimal parameter values, the distance $D_{i,j}$ between the complementary periods was measured as follows:

$$D_{i,j} = \left|X_j(P_i) - X_j(P_i')\right|$$

where $X_j(P_i)$ is the optimal value of the parameter X_j for the period P_i. The D distances were calculated after each shuffle, so that at the end of the permutation round (i.e., after 100 distance calculations), an empirical distribution of these distances was generated for each parameter. It was then possible to analyze the relative positions of the distance obtained (for each parameter) between the reference periods (HY and LY, WC and DC) within the distribution. If this distance was somewhere in the middle of the distribution, one could conclude that there was no reason to reject the null hypothesis H_0. If this distance was larger than almost all the values of the distribution, we rejected H_0, given that such a value was unlikely with this hypothesis. It should be noted that such a technique relies on a deterministic calibration procedure, as it yields to an optimal parameter set assumed to be representative of watershed behavior during the calibration period.

3.3.2. Analysis of Marginal Posterior Parameter Distributions

[20] In the second approach, changes in model parameters were analyzed within an equifinality context, to implicitly consider dependence between the model's structure, parameter values, and forcing data in model calibration. Even in parsimonious rainfall-runoff models such as GR4J, compensation effects preclude the interpretation of individual parameter values and the whole parameter set must be considered instead. In a Bayesian context, implementing the Regional Sensitivity Analysis from the work of *Hornberger and Spear* [1981] leads to marginal posterior distributions of parameter values rather than to one optimal set. These posterior distributions describe the set of acceptable parameters, leading to so-called behavioral solutions, with reference to the measure of model performance. In order to detect the signature of hydroclimatic variability, we hypothesize that behavioral distributions obtained for the HY and LY periods, as well as those obtained for the WC and DC periods, should be significantly different. The significance of this difference was measured using the χ^2 test, which, for a given significance level α, detects whether two empirical distribution functions are both realizations of the same unknown theoretical distribution law. The *p* value synthesized the results of the test; it corresponds to the probability of obtaining, within the hypothesis that a common distribution law exists (the null hypothesis), a χ^2 distance value greater than the distance computed between the two reference periods.

3.3.3. Signature of Hydroclimatic Variability in the Model's Performance

[21] The third approach we considered is based on an original analysis of calibration results. Using a global mathematical criterion during the model calibration may appear to be insufficient as a test of model relevance, as it often reduces the calibration process to a curve fitting. For *Boyle et al.* [2000], the resulting loss of information can even lead to seeing equifinality when there is none, which justifies the use of multicriteria strategies. As stressed by *Wagener et al.* [2003], "it also leads to problems with the identification of those parameters associated with response modes that do not significantly influence the selected objective function." Nevertheless, a finer study of the calibration results may lead to a better understanding of the model's performance. As an illustration, should it be considered that two calibrations on *x* years of data resulting in the same value of any objective function mean two equally good model simulations? Clearly not, since observation of the year-to-year fitting may lead to very different conclusions.

[22] Our hypothesis can be stated as follows: for any model, if calibrated parameter values are representative of the rainfall-runoff relationship, hydrological homogeneity (i.e., no changes in the rainfall-runoff relationship) of a calibration data set should result in a certain homogeneity of the calibrated model's performance. On the contrary, if calibrated parameter values are affected by substantial noise due to the statistical behavior of the model, hydrological homogeneity of a data set will not result in homogeneity of model performance. For this purpose, we used the variance of annual efficiencies as a descriptor of performance homogeneity over a calibration period. Therefore we could use year-to-year calibration results to analyze the model's behavior under changing conditions.

4. Results

4.1. Resampling Test on Optimal Parameter Values

[23] Results of the resampling test defined in section 3.3.1 are shown in Figure 7. We observe a large dispersion in optimal parameter values suggested by the distributions. Furthermore, the distances obtained between the HY (1954–1970) and LY (1972–1990) periods lie somewhere

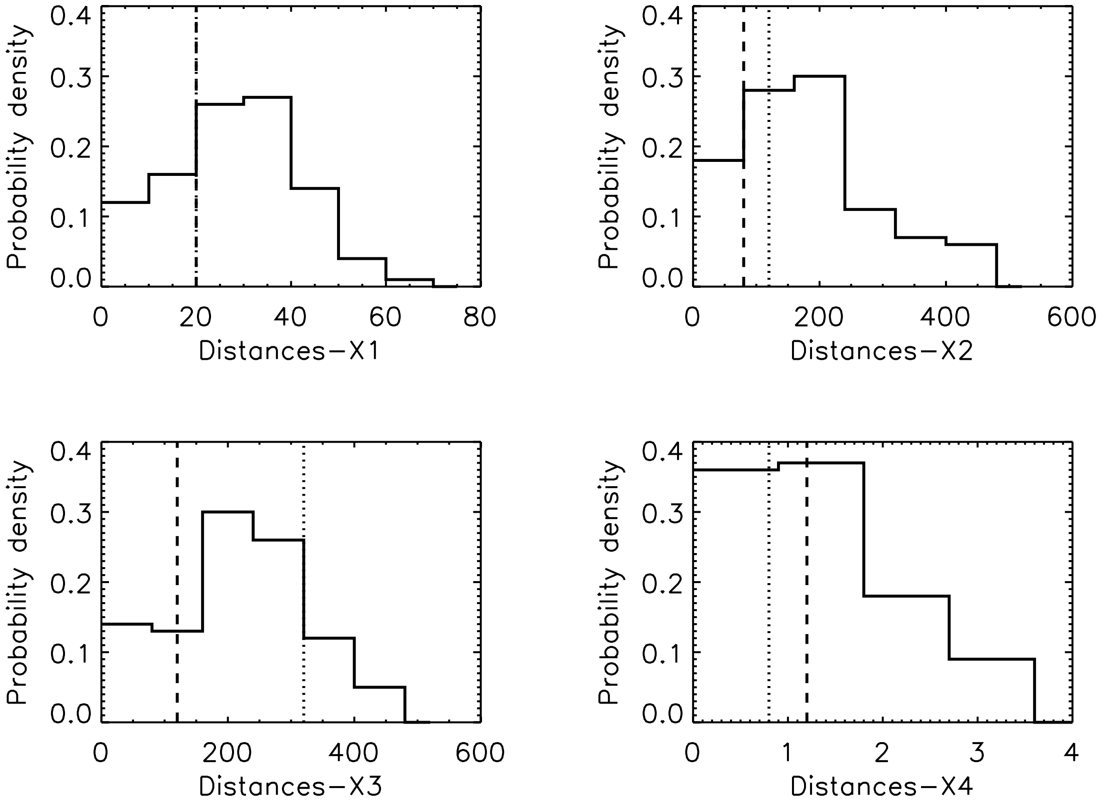

Figure 7. Parameter distance distributions for all four model parameters, from optimal calibrated values, compared to the parameter distances between HY and LY periods (dashed lines), WC and DC periods (dotted lines).

in the middle of the distributions, except perhaps for X_3. The conclusion is that there is no reason to reject H_0. In other words, changes in the rainfall-runoff relationship are not associated with significant changes in calibrated parameter values. As a result, does it mean that the calibration process is primarily driven by forcing conditions? To answer, one may observe distances obtained between the wet and dry (WC and DC) composite periods. Figure 7 shows that these distances are also in the middle of the distribution.

[24] Finally, it seems impossible to associate model parameter values 276 with rainfall-runoff relationship or climatic forcing conditions. Parameter identification problems may explain these results. The issue is examined in the next section by taking into account equifinality in calibrated parameter sets.

4.2. Analysis of Marginal Posterior Parameter Distributions

[25] From the initial Monte Carlo sampling calibration scheme described in section 3.1, we selected the top 10% of parameter sets performing well as behavioral sets, so as to calculate the posterior parameter distributions. Figure 8 shows the cumulative marginal distributions of behavioral parameter values for the HY (1954–1970) and LY (1972–1990) periods. The production module parameters (X_1 and X_2) are the most sensitive, since they show greater deviation from the original uniform distribution than the transfer module parameters (X_3 and X_4). In particular, the X_4 parameter displays an apparent insensitivity, which may be explained by the smooth hydrological response of this large watershed. Given a chosen significance level $\alpha = 0.01$, parameters X_1, X_2, and X_3 are shown to be significantly different for the two periods, as the probability of obtaining a greater distance within the null hypothesis is lower than 1%. In particular, the X_1 decrease observed is consistent with the decrease in runoff coefficient on the watershed, as it tends to increase groundwater losses.

[26] Figure 9 shows the same distributions for the wet and dry (WC and DC) composite periods. The results are similar to those obtained for the HY and LY periods, as several parameter distributions (here for X_1 and X_3) are significantly different for the two periods. As a result, changes in the rainfall-runoff relationship cannot be associated with significant changes in model parameter posterior distributions. Once again, it is impossible to conclude on the relationship between the hydrological behavior of the watershed and the model parameter values.

4.3. Signature of Hydroclimatic Variability in the Model's Performance

[27] The alternative methodology described in section 3.3.3 was also implemented, using the same resampling approach that the one applied in section 4.1. What differs is the measured statistic, here defined as the interannual efficiency variances and calculated as follows:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(E_i - \overline{E}\right)^2$$

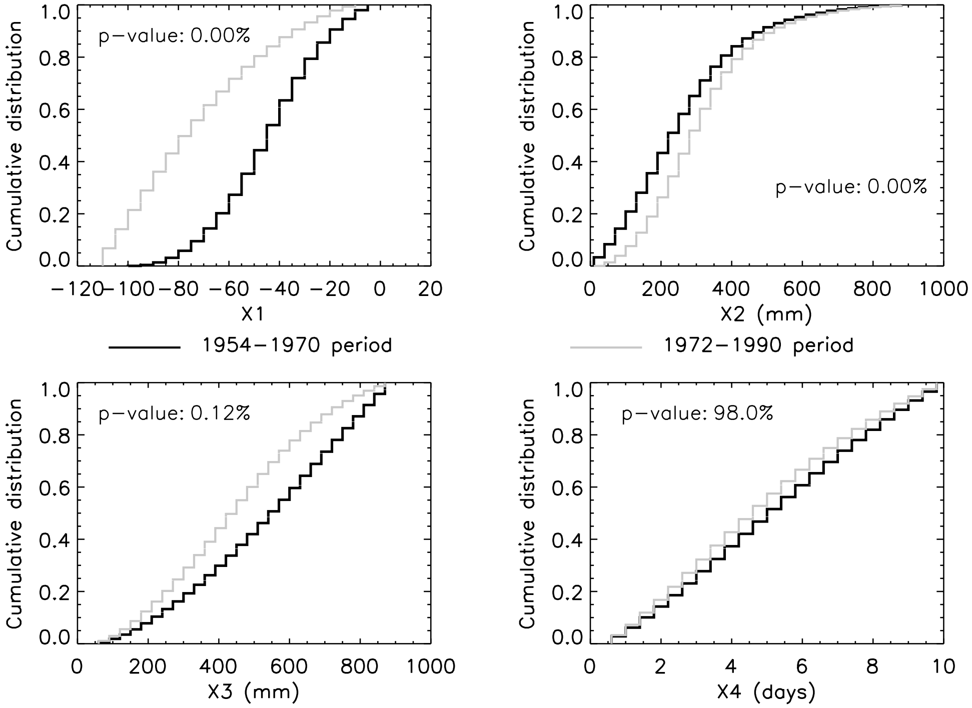

Figure 8. Cumulative marginal likelihood distributions, for all four model parameters, from behavioral parameter sets for 1954–1970 (HY) and 1972–1990 (LY) periods.

where n is the number of years in the calibration period, E_i is the efficiency for the year i, and E is the averaged global efficiency. As the Nash and Sutcliffe efficiency uses the mean flow on the calibration period as the reference model, comparison of efficiency values between different periods is biased. Therefore the efficiency criterion was modified, with the reference model becoming the interannual mean flow (i.e., over the entire 1954–2002 period). This variance is calculated for the 200 calibration periods defined in section 3.3.1, thus generating an empirical distribution.

[28] Figure 10 shows the distribution of efficiency variances. This distribution is compared to the variances

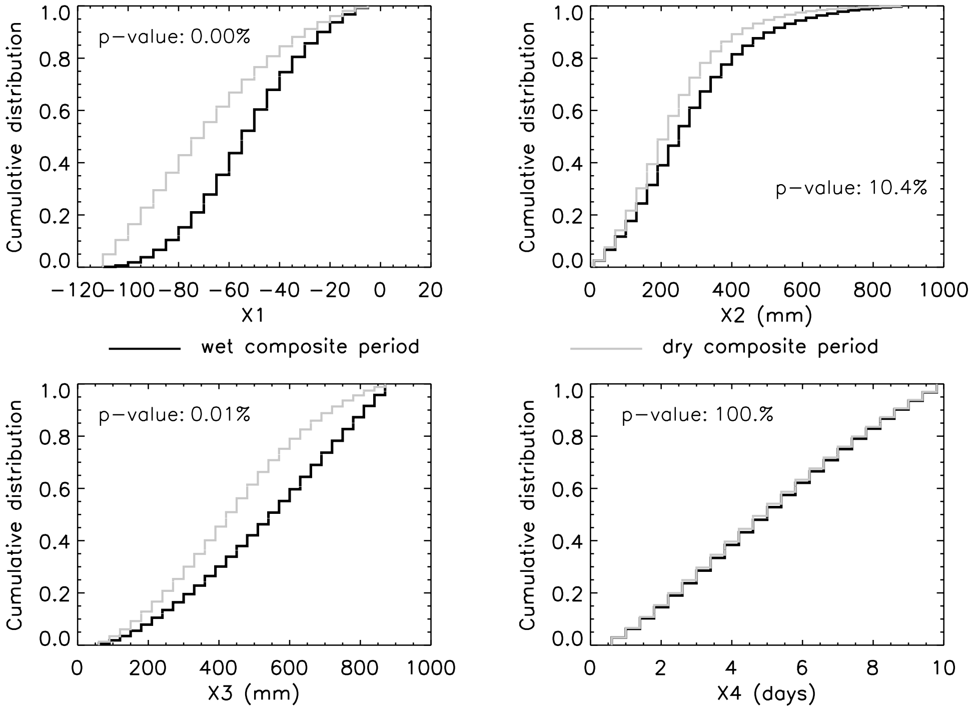

Figure 9. Cumulative marginal likelihood distributions, for all four model parameters, from behavioral parameter sets for the wet (WC) and dry (DC) composite periods.

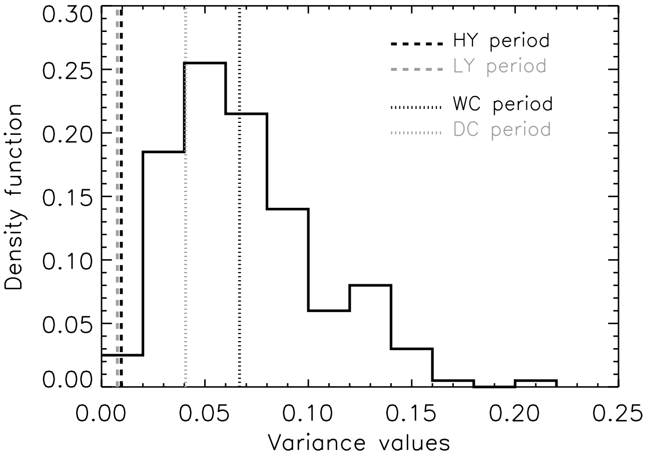

Figure 10. Distribution of year-to-year efficiency variances, compared to variances obtained for HY, LY, WC, and DC periods.

obtained for the two reference data sets. Variances associated with the HY and LY periods, which may be considered as more hydrologically homogeneous than the randomly generated ones, are shown to be the two lowest variances of the distribution. Moreover, one can observe that the variances calculated for WC and DC periods, which do not present any particular hydrological homogeneity, are close to the median of the distribution. This methodology therefore makes it possible to consider calibrated parameter values of the model as mostly representative of a given rainfall-runoff relationship on the watershed.

4.4. Discussion

[29] The first two methods did not provide a conclusion on the relationship between model parameter values and the rainfall-runoff relationship. It can therefore be concluded that the magnitude of changes in parameter values is not a good indicator of changes in watershed behavior, even taking equifinality problems into account. These results outline, first, the statistical relation that calibration processes create between the model's parameter values and the data set used to force the model. They also make it necessary to discuss the use of a climatological PE. At the intra-annual scale, PE variability is mostly seasonal and daily fluctuations have a negligible impact on rainfall-runoff model simulations [*Oudin et al.*, 2005]. However, the neglected interannual variability of PE probably biases the model's parameter values. In particular, production module parameters can compensate an inaccurate PE estimation. We therefore regret the lack of long-term PE data, but two comments must be made. First, the variability in annual PE, such as calculated on the contrasted 1997–2002 period, is quite limited (between 1410 and 1580 mm, i.e., a maximum relative variation of approximately 11%). We therefore cannot suggest only this reason to explain the results obtained. Secondly, PE is always, and by far, the worst estimated forcing variable. It is also rarely available on long-term periods. Efficient methodologies must therefore be developed to deal with this lack of data.

[30] The third method, on the basis of the analysis of year-to-year calibration results, led to more definitive conclusions, as the HY and LY periods were found to exhibit the lowest variances. In this case, one can conclude that calibrated model parameters are relevant from a hydrological point of view. Consequently, year-to-year efficiency variance can be used to investigate the hydrological homogeneity of any given period. Hence the test has been applied to two other periods, 1954–1990 and 1992–2002. The 1954–1990 period constitutes a validation test for the proposed method, as its hydrological heterogeneity has been evidenced. On the contrary, we have no preconceived idea of the hydrological homogeneity of the 1992–2002 period. Results are shown in Figure 11. First, the large variance calculated for the 1954–1990 period gives further confidence in the method proposed. Secondly, the rainfall-runoff relationship appears to be very stable on the 1992–2002 period, given the very strong homogeneity of the model's performance.

5. Conclusions

[31] The applicability of rainfall-runoff models under changing conditions is an issue of increasing interest. Indeed, it appears to be a necessary condition for dealing with the impact of climate and land-use changes on water sustainability and for predicting the response of ungauged basins. In this context, the relationship between hydroclimatic stationarity and rainfall-runoff model parameter stability has to be questioned. Hence this study proposed an exploration of rainfall-runoff model behavior under changing hydroclimatic conditions.

[32] The Upper Ouémé watershed, in Benin, was chosen as the case study, because it is a good example to understand and predict changes in water resources in West Africa. This watershed was shown to have experienced a high hydrological yield decrease during the 1970s and the 1980s. First, the signature of these hydrological changes in model parameter sets was investigated. Two methods were applied, on the basis of an analysis of changes in optimal parameter values and on changes in marginal posterior distributions. The results showed that the magnitude of changes in the model's parameter values is not a suitable measure of changes in watershed behavior.

[33] An alternative and original methodology was also proposed. If the model's parameter values are relevant, that

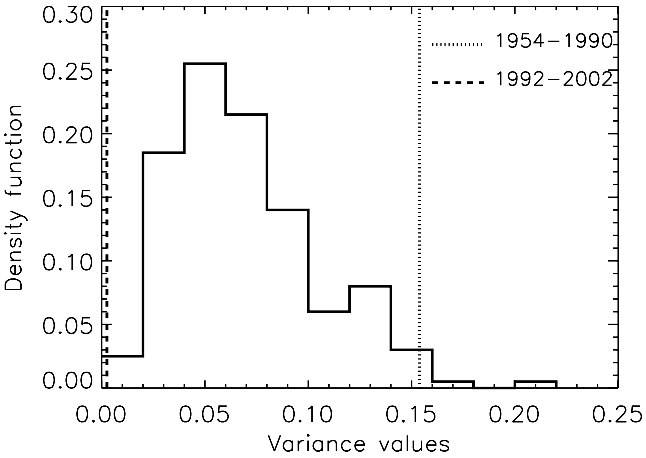

Figure 11. Distribution of year-to-year efficiency variances, compared to variances obtained for 1954–1990 and 1992–2002 periods.

is, they are representative of watershed's behavior, we suggest that for any calibration period, a relative stability in the rainfall-runoff relationship will be associated with a relative stability in the model's performance over the period considered. The test applied used interannual model efficiency variances to measure performance homogeneity and a resampling test to statistically characterize the calculated results. Applying this test showed that calibrated parameter values have hydrological relevance. We therefore feel that the approach proposed is promising, as it can be used to evaluate a modeling ensemble (i.e., a model structure and a calibration strategy), before considering scenarios of changes in parameter values. When hydrological relevance of calibrated parameter values is thus verified, this test can also be used to study stationarity of a watershed's behavior. To apply this methodology, resampling techniques must be usable on data, which implies that independent periods can be identified. However, this restriction is not limiting if long periods of data are available.

[34] **Acknowledgments.** These researches are part of the AMMA international program and were partly funded by the French ECCO/PNRH program. The authors wish to thank Vazken Andréassian and Charles Perrin for their expert assistance on the GR4J model. The constructive comments of the three reviewers were appreciated and considerably helped to improve the paper.

References

Andréassian, V., E. Parent, and C. Michel (2003), A distribution-free test to detect gradual changes in watershed behavior, *Water Resour. Res.*, *39*(9), 1252, doi:10.1029/2003WR002081.

Beven, K. J. (1993), Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, *16*, 41–51.

Bormann, H., and B. Diekkrüger (2003), Possibilities and limitations of regional hydrological models applied within an environmental change study in Benin (West Africa), *Phys. Chem. Earth*, *28*, 1323–1332.

Bormann, H., T. Fass, S. Giertz, B. Junge, B. Diekkrüger, B. Reichert, and A. Skowronek (2005), From local hydrological process analysis to regional hydrological model application in Benin: Concept, results and perspectives, *Phys. Chem. Earth*, *30*, 347–356.

Boyle, D., H. V. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrologic models: Combining the strengths of manual and automatics methods, *Water Resour. Res.*, *36*(12), 3663–3674.

Brown, A., L. Zhang, T. McMahon, A. Western, and R. Vertessy (2005), A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation, *J. Hydrol.*, 1–34.

Deichmann, U. (1996), Africa population database, National Center for Geographic Information and Analysis, United Nations Environment Programme, World Resources Institute, Internet, www.grid2.cr.usgs.gov.

Food and Agricultural Organization (2003), The state of the world's forests, Food and Agriculture Organisation of the United Nation, Available from: http://www.fao.org/DOCREP/005/Y7581E/y7581e00.htm.

Gan, T. Y., and S. Burges (1990), An assessment of a conceptual rainfall-runoff model's ability to represent the dynamics of small hypothetical catchments: 2. hydrologic responses for normal and extreme rainfall, *Water Resour. Res.*, *26*(7), 1605–1619, doi:10.1029/90WR00146.

Hornberger, G., and R. Spear (1981), An approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, *12*, 7–18.

Hubert, P., J. Carbonnel, and A. Chaouche (1989), Segmentation des séries hydrométéorologiques—Application des séries de précipitations et de débits de l'afrique de l'ouest, *J. Hydrol.*, *110*, 349–367.

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

Kundzewicz, Z., and A. Robson (2004), Change detection in hydrological records—A review of the methodology, *Hydrol. Sci. J.*, *49*(1), 7–19.

Le Barbé, L., T. Lebel, and D. Tapsoba (2002), Rainfall variability in West Africa during the years 1950–1990, *J. Clim.*, *15*(2), 187–202.

Le Lay, M., and S. Galle (2005), Variabilités interannuelle et intra-saisonni'ere des pluies aux échelles hydrologiques. La mousson ouest africaine en climat soudanien, *Hydrol. Sci. J.*, *50*(3), 509–524.

Lebel, T., A. Diedhiou, and H. Laurent (2003a), Seasonal cycle and interannual variability of the Sahelian rainfall at hydrological scales, *J. Geophys. Res.*, *108*(D8), 8389, doi:10.1029/2001JD001580.

Lebel, T., J. L. Redelsperger, and C. Thorncroft (2003b), African Monsoon Multidisciplinary (AMMA): An international research project and field campaign, *CLIVAR Exchanges*, *27*(8), 52–54.

L'Hôte, Y., G. Mahé, and S. Bonaventure (2003), The 1990s rainfall in the sahel: The third driest decade since the beginning of the century, *Hydrol. Sci. J.*, *48*(3), 493–496.

Lorup, J., J. Refsgaard, and D. Mazvimavi (1998), Assessing the effect of land use change on catchment runoff by combined use of statistical tests and hydrological modelling: Case studies from Zimbabwe, *J. Hydrol.*, *205*, 147–163.

Mahé, G., J. E. Paturel, E. Servat, D. Conway, and A. Dezetter (2005), The impact of land use change on soil water holding capacity and river flow modelling in the Nakambe River, *J. Hydrol.*, *300*, 33–43.

Milly, P. C. D. (1994), Climate, soil water storage, and the average annual water balance, *Water Resour. Res.*, *30*(7), 2143–2156.

Nash, J., and J. Sutcliffe (1970), River flow forecasting through conceptual models. Part I—A discussion of principles, *J. Hydrol.*, *10*, 282–290.

Nicholson, S., and I. Palao (1993), A re-evaluation of rainfall variability in the Sahel: Part 1. Characteristics of rainfall fluctuations, *Int. J. Climatol.*, *13*, 371–389.

Niel, H., J. Paturel, and E. Servat (2003), Study of parameter stability of a lumped hydrologic model in a context of climatic variability, *J. Hydrol.*, *278*, 213–230.

Oudin, L., C. Michel, and F. Anctil (2005), 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*, 275–289.

Ozer, P., M. Erpicum, G. Demarée, and M. Vandiepenbeeck (2003), The sahelian drought may have ended during the 1990s, *Hydrol. Sci. J.*, *48*(3), 489–492.

Paturel, J., E. Servat, B. Kouamé, H. Lubs, M. Ouedraogo, and J. Masson (1997), Climatic variability in humid africa along the Gulf of Guinea. Part 2: An integrated regional approach, *J. Hydrol.*, *191*, 1–15.

Perrin, C., C. Michel, and V. Andréassian (2003), Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, *279*, 275–289.

Scheffé, M. (1959), *The Analysis of Variance*, 477 pp., John Wiley, Hoboken, N. J.

Schreider, S., A. Jakeman, R. Letcher, R. Nathan, B. Neal, and S. Beavis (2002), Detecting changes in streamflow response to changes in non-climatic catchment conditions: Farm dam development in the murray-darling basin, Australia, *J. Hydrol.*, *262*, 84–98.

Sorooshian, S., and F. Arfi (1983), Response surface parameter sensitivity analysis methods for postcalibration studies, *Water Resour. Res.*, *18*(5), 1531–1538.

Sorooshian, S., and V. K. Gupta (1983), Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness, *Water Resour. Res.*, *19*(1), 260–268.

Tapsoba, D., M. Haché, L. Perreault, and B. Bobée (2004), Bayesian rainfall variability analysis in West Africa along cross sections in space-time grid boxes, *J. Clim.*, *17*, 1069–1082.

Troutman, B. (1983), Runoff prediction errors and bias in parameter estimation induced by spatial variability of precipitation, *Water Resour. Res.*, *19*(3), 791–810.

Varado, N., I. Braud, S. Galle, M. Le Lay, L. Sguis, B. Kamagate, and C. Depraetere (2006), Multi-criteria assessment of the representative elementary watershed approach on the donga catchment (benin) using a downward approach of model complexity, *Hydrol. Earth Syst. Sci.*, *10*(3), 427–442.

Wagener, T., N. McIntyre, M. Lees, H. Wheater, and H. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Processes*, *17*, 455–476.

I. Braud, Cemagref, Lyon, France.

S. Galle, M. Le Lay, and G. M. Saulnier, LTHE, Université de Grenoble, [CNRS, INPG, IRD, UJF], BP 3091, Grenoble cédex, France. (matthieu.lelay@hmg.inpg.fr)