Comptes rendus - Geoscience 351 (2019) 477–486

Contents lists available at ScienceDirect

Comptes rendus - Geoscience

www.journals.elsevier.com/comptes-rendus-geoscience

Hydrology, Environment

Comparison of hydrological models for use in climate change studies: A test on 241 catchments in West and Central Africa

Kiswendsida Samiratou Ouermi [a, *], Jean-Emmanuel Paturel [b], Julien Adounpke [a], Agnidé Emmanuel Lawin [a], Bi Tié Albert Goula [c], Ernest Amoussou [d]

[a] Laboratoire d'hydrologie appliquée, Université d'Abomey-Calavi, Benin
[b] HSM, IRD, CNRS, Université de Montpellier, Montpellier, France
[c] Université Nangui Abrogoua, Côte d'Ivoire
[d] Faculté des lettres, des arts et des sciences humaines, Université d'Abomey-Calavi, Benin

A R T I C L E I N F O

Article history:
Received 4 April 2019
Received in revised form 9 July 2019
Accepted 21 August 2019
Available online 15 November 2019
Handled by François Chabaux

Keywords:
Rainfall-runoff modelling
Climate change
Model calibration
Model robustness
West and Central Africa

A B S T R A C T

The present work reports on the study and the comparison of the performance of three "Génie rural" (GR) and two Water Balance (WB) models. The calibration and robustness performances are analysed in the light of the hydro-climatic conditions. The study shows that the GR models are much more efficient and robust than the WB models. The behaviour of calibration performance as a function of hydroclimatic variables varies according to the model and goodness-of-fit criteria (GOFC). The GR models are more robust in terms of NSE(Q) and NSE($\sqrt{Q}$) and the WB models in terms of KGE. For more robustness of the models, it is better to transfer the parameters to wetter periods or periods with a lower Potential EvapoTranspiration (PET) than the calibration period. For a loss of robustness of less than 20% for GR and 30% for WB, the variation between calibration and rain validation/PET periods must be around ±15%/±1.5%.

1. Introduction

Africa has been experiencing continuous global warming over the past 50–100 years. A decrease in annual rainfall and an increase in extreme weather events have been observed over the past 30–40 years in West and Central Africa (Deonarain, 2014; IPCC 2013; IPCC 2014; Niang et al., 2014; Panthou et al., 2014; Taylor et al., 2017). However, due to its poverty, the region is struggling to cope with this situation. Indeed, to survive in the changing climate context, the western and central African countries must build infrastructures for the sustainable management of their natural resources, especially water resources. To ensure the sustainability of these infrastructures, their dimensions must be adapted to the current conditions. Given that most of this region's infrastructures were designed with standards established around the 1960s (Rodier, 1964), an update is required for those structures to stand the changing climate conditions.

One of the tools used to obtain "reliable" hydrological data for the design of structures is the use of hydrological models. The main objective of hydrological modelling is to reproduce flows with as minimal error as possible. A good model is assumed to be insensitive to changes in watershed conditions. Seiller et al. (2012) define the robustness of a hydrological model as its degree of insensitivity to climatic and/or environmental conditions. The more insensitive the model is to climatic and/or environmental conditions, the more robust it is. A robust model is therefore capable of reproducing flows over a different period than that of the

* Corresponding author. 688, avenue Professeur-Joseph-Ki-Zerbo, 01 BP 182, Ouagadougou 01, Burkina Faso.
E-mail address: samkisw@yahoo.fr (K.S. Ouermi).

https://doi.org/10.1016/j.crte.2019.08.001
1631-0713/

calibration period with a performance as close as possible to the calibration one. Unfortunately, "robust" models do not exist. The parameters of the models depend on their setting period, even if the characteristics of the flow in the studied catchment remain constant over time (Thirel et al., 2015a). This could be explained by the fact that the models are simplified representations of natural phenomena, however complex these may be (Thirel et al., 2015b; Wang et al., 2018). Because of that, the limits of hydrological models should be evaluated before any use. In West and Central Africa, models are not sufficiently tested due to lack of or limited data. Very few studies have been conducted on the robustness of models in West and Central African catchments (Dezetter et al., 2008; Ouermi et al., 2015). They find that it is preferable to transfer the parameters from drier to wetter periods. Other authors such as Coron et al., 2012; Dakhlaoui et al., 2017; Vaze et al., 2011 and Wilby (2005) have tested the robustness of hydrological models in other parts of the world. Working on Australian (Coron et al., 2012; Vaze et al., 2011) and northern Tunisian (Dakhlaoui et al., 2017) catchments, these authors found it preferable to transfer the parameters to wetter and colder periods. However, in contrast, Wilby (2005), working on five British catchments, recommended to transfer the parameters from wetter to dryer conditions. Do these results allow us to infer that the behaviour of the models is different according to climate zones: tropical climate vs. temperate climate? Further analysis would be required.

The current work consists in studying several hydrological models and comparing their robustness on many catchments of West and Central Africa, seeing how the climate conditions impact model calibration, the robustness of the models in the area of study, and the transferability of the parameters. It should be noted that this study has a larger scope in terms of number of basins studied and space than the other studies conducted on the robustness of the models.

In this article, we will study the behaviour of the different calibration models according to the climatic characteristics of the calibration period, study their robustness in our study area, and finally set the climatic and hydrological regime limits for parameter transfers.

2. Study area and catchment sets

2.1. Study area

West and Central Africa are subject to various hydrological regimes that make the region very unique.

The hydrological regime of a river at a given point is the average behaviour of its flows at this point during one hydrological cycle (the form of its hydrograph). In West and Central Africa, two families of hydrological regimes exist (Rodier, 1964): tropical and equatorial regimes. The tropical regime is characterized by one season of high flows and one season of low flows. The equatorial one is characterized by two seasons of high flows and two seasons of low flows. Depending in particular on the duration of each of these seasons and the extent of the high or low flows, we can distinguish desertic, subdesertic, Sahelian, transitional, and pure tropical hydrological regimes, and boreal transitional, austral transitional, and pure equatorial hydrological regimes. Fig. 1a provides an overview of the spatial location of these different hydrological regimes derived from Rodier (1964) in the study area.

The study focuses on 241 catchments in West and Central Africa (Fig. 1b); these catchments encompass the main rivers of the region and the Sahelian regime, pure tropical, transitional, and Dahomean transitional tropical regimes, and boreal transitional and pure equatorial hydrological regimes.

2.2. Data

Monthly runoff chronological series come from SIEREM database managed by HydroSciences Montpellier (http://www.hydrosciences.fr/sierem/; Boyer et al., 2006). The data were checked for errors.

Monthly rainfall (P) and potential evapotranspiration (PET) continuous time series come from CRU TS 4.00 grids (Harris, 2017). The data are observational products. The spatial resolution of the grids is 0.5×0.5 degree and the data cover the period from 1901 to 2015.

2.3. Catchments sets

Fig. 2 showed the diversified hydroclimatic characteristics of the catchments under investigation, notably the precipitation, the PET, and the specific annual runoff. Fig. 2 also reveals the period during which the hydrometric stations function, and the percentage of hydrological data gaps by catchment.

The time period covered for the study is from 1950 to 1990 due to the limited services given by the national hydrological services and the difficulties to get recent hydroclimatic data of African regions.

Due to limitation of data, it is admitted an acceptable percentage of gaps in runoff, according to the hydrological regime of the catchments of interest. For tropical catchments, especially the Sahelian and the pure tropical catchments, the acceptable percentage of gaps has been set at 50%. The gaps recorded are generally in periods of low water levels where the flow is most often perennial. For equatorial catchments, the acceptable gap is set at 30% because these rivers have a continuous flow.

Most of the selected catchments range in size from 200 to 1000 km^2. They are mostly unregulated with no major storage or irrigation schemes.

3. Modelling methodology and analysis method

3.1. Hydrological models

Five monthly lumped rainfall-runoff models (RR) with continuous reservoir types were used in this study. In spite of their parsimony (only a few free parameters), they showed a good level of efficiency in previous studies (Ouermi et al., 2015; Paturel, 2014; Paturel et al., 1997) and correspond to several representations of the rainfall–runoff transformation. Within a model, the versions differ due to the number of parameters to calibrate:

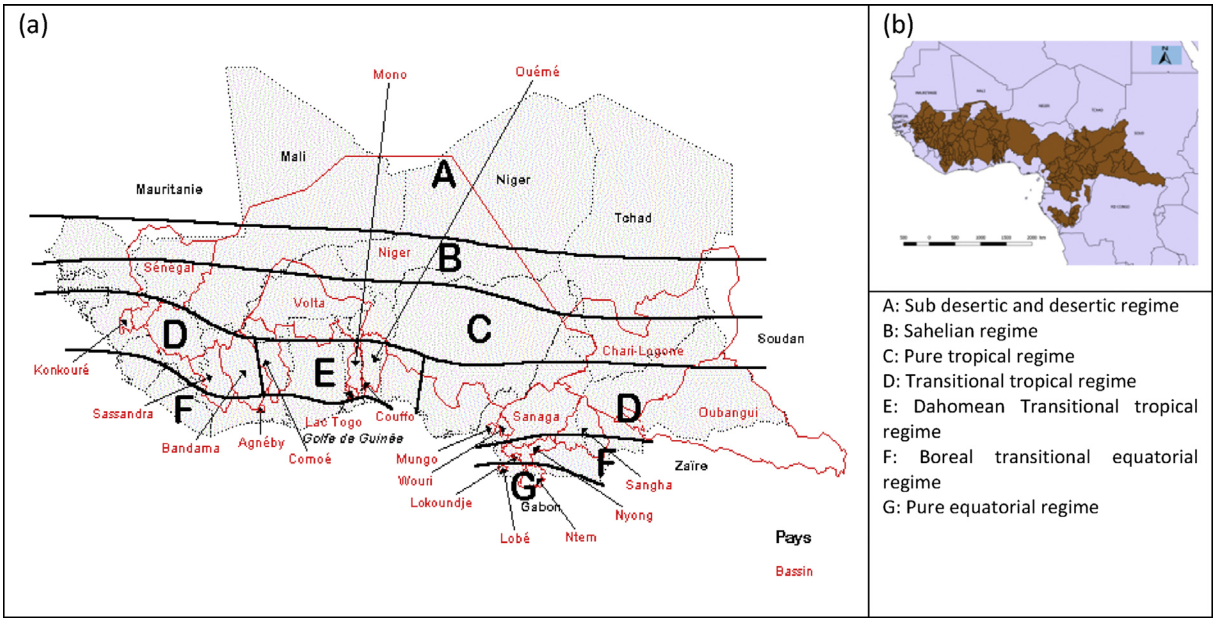

Fig. 1. (a) Hydrological regimes zone based on Rodier (1964) and main rivers catchments of West and Central Africa; (b) 241 studied catchments.

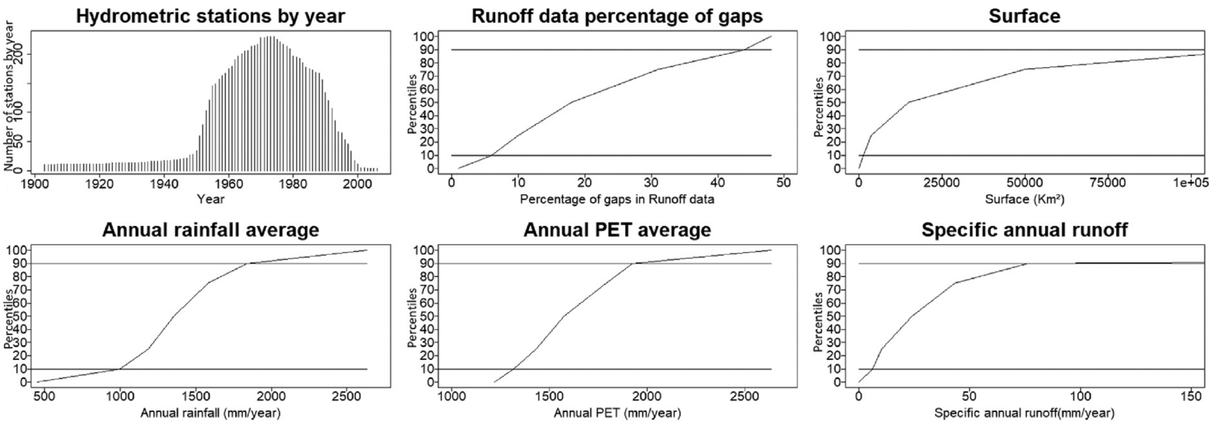

Fig. 2. Catchment characteristics on the entire set of 241 catchments.

- Two versions of the Water Balance model (Conway, 1997) with two or three parameters (WB2 or WB3), which is a combination of the Thornthwaite water balance approach (Thornthwaite and Mather, 1957) and the large scale hydrological modelling approach used by Vörösmarty et al. (1989) and Vörösmarty and Moore (1991);
- Two versions of Makhlouf's model (Makhlouf and Michel, 1994) with two or three parameters (MK2 or MK3);
- One version of Mouelhi's model (Mouelhi, 2003; Mouelhi et al., 2006) with two parameters (MO).

MK and MO models are monthly GR models (https://webgr.irstea.fr) with two reservoirs, a "ground reservoir" and a "transfer reservoir". WB models are single reservoir models.

The first version of the MK model has four parameters. According to Makhlouf and Michel (1994), two of these parameters can be fixed in specific climatic conditions. They suggested that, in other climatic conditions, it is preferred that all four parameters be calibrated. After various uses of the MK model, Ouédraogo (2001) and Lubès-Niel et al., 2003 found that, for most of the African catchments studied, the parameter linked to direct flow is zero and the parameter linked to soil condition characteristics (A) can be assimilated to the Water Holding Capacity (WHC) of the soil, data mapped by the FAO (1995). Thus, it is proposed in this study that the third parameter, γ in WB and A in MK, be replaced by WHC.

Even if MK and MO are GR models, their structure is a little bit different: the MO model assumes that there is conceptually an exchange of water with outside parts of catchment, a factor which is able to adjust the rainfall inputs to provide higher NSE values in calibration; not in the MK model.

We refer to Supplementary Material 1 for a further overview of the characteristics of the five tested models.

3.2. Model calibration: optimization method and goodness-of-fit criteria

The parameters are calibrated using the method of Rosenbrock followed by the simplex of Nelder and Mead (Servat and Dezetter, 1988). For this, we use three goodness-of-fit criteria (GOFC), ranging from $-\infty$ to 1:

- $NSE(Q)$ Criterion (Nash and Sutcliffe, 1970),

$$NSE(Q) = 1 - \frac{\sum (Q_{sim} - Q_{obs})^2}{\sum Q_{obs} - (\overline{Q_{obs}})^2} \tag{1}$$

- $NSE(\sqrt{Q})$ Criterion (Nash and Sutcliffe, 1970),

$$NSE\left(\sqrt{Q}\right) = 1 - \frac{\sum\left(\sqrt{Q_{sim}} - \sqrt{Q_{obs}}\right)^2}{\sum\left(\sqrt{Q_{obs}} - \overline{\sqrt{Q_{obs}}}\right)^2} \tag{2}$$

- KGE Criterion (Gupta et al., 2009),

$$KGE = 1 - \sqrt{(\mathrm{r}-1)^2 + (\alpha-1)^2 + (\beta-1)} \tag{3}$$

with $\beta = \frac{Q_{obs}}{Q_{obs}}$, $\frac{\alpha^2 = \frac{\sigma^2_{sim}}{\sigma^2_{obs}}, r = \frac{1}{n}\sum_1^n (Q_{obs} - \overline{Q_{obs}})(Q_{sim} - \overline{Q_{sim}})}{\sigma_{obs}\sigma_{sim}}$

where Q_{sim} and Q_{obs} are the simulated and observed flows, σ_{sim} and σ_{obs} are respectively the standard deviation of the simulated and observed flows.

The *NSE* criterion (Q) is the most widely used one for evaluating the performance of hydrological models because of its simplicity. However, it gives more importance to high flows than other parts of hydrograph. Its NSE variant ($\sqrt{Q}$) gives equal consideration to all parts of the hydrograph.

The *KGE* index is a result of decomposition of $NSE(Q)$ and represents a compromise between three evaluation criteria, correlation coefficient, bias error and standard deviation ratio (Dakhlaoui et al., 2017).

3.3. Generalized split-sample test

The differential split-sample test (DSST) proposed by Klemes (1986) is the common typical testing procedure to investigate the parameter dependency on climate and the related consequences on model efficiency. This is a specific case of the split-sample test (SST), where calibration and validation periods are chosen according to their climatic differences. Finding limits to the DSST, Coron et al. (2012) proposed a generalized methodology to evaluate the validity of RR models for use under non-stationary climatic conditions: the Generalised-Split-Sample Test (GSST). According to the authors, GSST overcomes the limits to provide comparable results under various conditions and over a wide range of parameter transfer conditions, thus resulting in more robust conclusions on parameter transferability.

The method follows three steps

- numerous subperiods are created by a sliding window moving by one year according to one climatic characteristic (e.g., mean rainfall or PET for the catchment);
- a hydrological model is calibrated on each subperiod using previously selected quality criteria; this provides one parameter set per period;
- for each calibration subperiod, the optimized parameter set is used to perform all the possible validation tests on independent subperiods. Validation subperiods overlapping with the calibration ones are not considered to ensure strict independence of calibration and validation conditions.

Nevertheless, Coron et al. (2012) note that their method is limited because it does not always make it possible to attribute the change in model behaviour to a specific climate or environmental change experienced in a catchment.

In the present study, the length of the chosen subperiod is five years for calibration and validation. Subperiods of five years permit to maximize the climatic differences between two periods and to test models in much contrasted conditions.

However, the number of sub-periods is different from one catchment to another because of the available runoff data for the considered catchment. Given the available runoff data, 7634 calibration runs were possible in this study.

3.4. Robustness

The goodness-of-fit criterion in the ‘model simulation’ (using an optimized parameter set from model calibration versus a different subperiod) is then compared to the model calibration results to quantify the variation in GOFC. This relative variation measures the transferability of the model to climate-contrasted periods or the robustness of the model. It is formulated as:

$$\mathrm{Robustness} = \frac{\mathrm{Performance}_{receiver} - \mathrm{Performance}_{donor}}{\mathrm{Performance}_{donor}} \tag{4}$$

The “donor” period is the calibration period and the “receiver” period is the validation period. So Performance$_{donor}$ is the calibration performance and Performance$_{receiver}$ is the validation performance.

Some authors, such as Seiller et al. (2012) and Dakhlaoui et al., 2017, criticized this way of expressing model robustness. They found that it underestimates the robustness of the models by exaggerating the performance losses of the models. They found other methods of measuring robustness. Dakhlaoui et al., 2017 proposed:

$$\text{Robustness}_{\text{Dakhlaoui}} = \frac{\text{Performance}_{\text{val}_{\text{receiver}}} - \text{Performance}_{\text{cal}_{\text{receiver}}}}{\text{Performance}_{\text{cal}_{\text{receiver}}}} \quad (5)$$

$\text{Performance}_{\text{val\_receiver}}$ is the validation performance of the model with parameters set to a period other than the receiving period and $\text{Performance}_{\text{cal}_{\text{receiver}}}$ is the calibration performance of the model with parameters set to the receiving period.

4. Results and conclusion

In the present study, the calibration performance and robustness of the models have been analysed only for parameter sets where at least one calibration performance criterion was equal or higher than 60 in terms of GOFC. This condition permits to assume that the model can be considered "good" at a given catchment (Hounpke, 2016). Therefore, only a part of the 7634 possible calibration runs were analysed. Moreover, some runs of calibration failed and were not taken into account. The failure of model calibration on a catchment has not been studied. Also if all calibrations had been successful, the possible number of calibration-validation runs would be 297 731 but, given the conditions set above, the maximum possible number of these calibration-validation runs considered for a model is 292 738.

4.1. Analysis of the calibration performances of the models

An overview of calibration performance was provided to evaluate the quality of the sets of reference parameters and to check that the models perform reasonably well in calibration. The calibration performance in terms of GOFC is summarised as boxplots according to models (Fig. 3) and according to criteria (Fig. 4). The horizontal line in the box is the median of GOFC values. The upper and lower envelopes show the 75th and the 25th percentile values and the upper and lower whiskers show the 95th and 5th percentile values, respectively.

An overall analysis of the calibration performance of the models showed that the GR models are more efficient than the WB models. The median and the upper whisker of the plots are lower for WB models than for the GR models. The number of runs with at least 60 for GOFC is higher for GR models than for WB models. The concept of GR is more applicable to the study area investigated, which is probably due to the structure of GR models, which have one routing reservoir instead of none for WB models.

The MK models are slightly more performant than the MO model in calibration. This is confirmed by the number of runs, which is higher for MK models than for the MO model.

The results of calibration performance are more distinct for NSE($\sqrt{Q}$) than for the two others GOFC because of the lower whisker, which is higher than for others. We must note that in the case of KGE, the upper whisker is slightly similar to the two other GOFCs.

The different versions of the MK and WB models with two or three parameters are almost equivalent in calibration, despite the different numbers of parameters to calibrate whatever the performance criteria may be. It seems that assimilating the soil reservoir parameter to a water holding capacity of the soil is appropriate. It must be due to

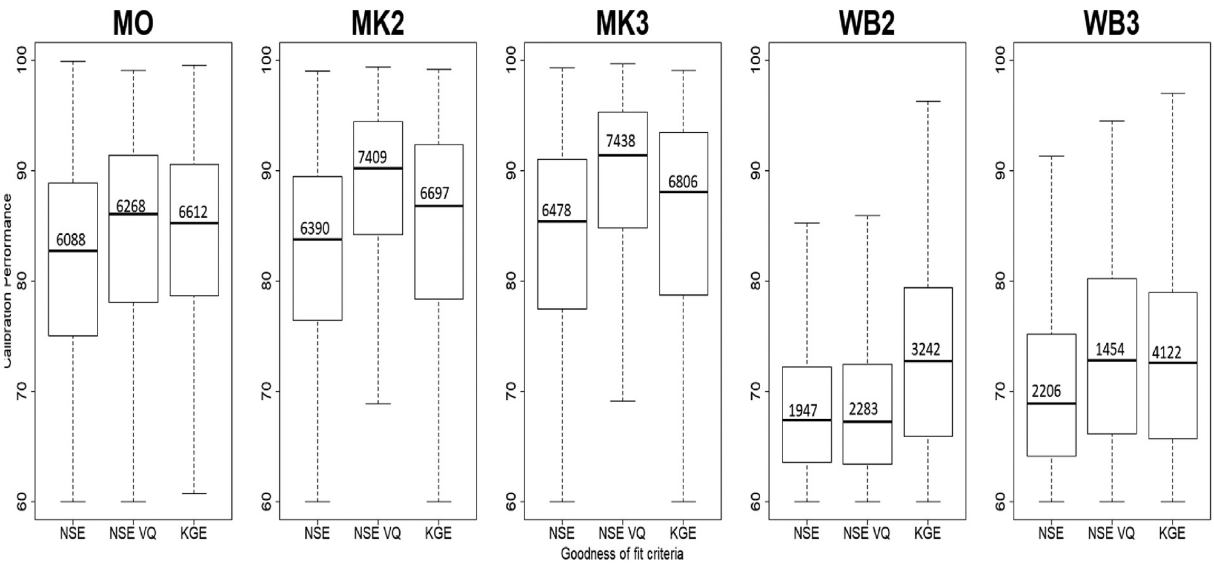

Fig. 3. Summary of goodness-of-fit criteria values across the 241 catchments for the calibration of the five rainfall–runoff models for the modelling periods according to the goodness-of-fit criteria. The number of runs with at least 60 as performance is indicated on the median line of the boxplots.

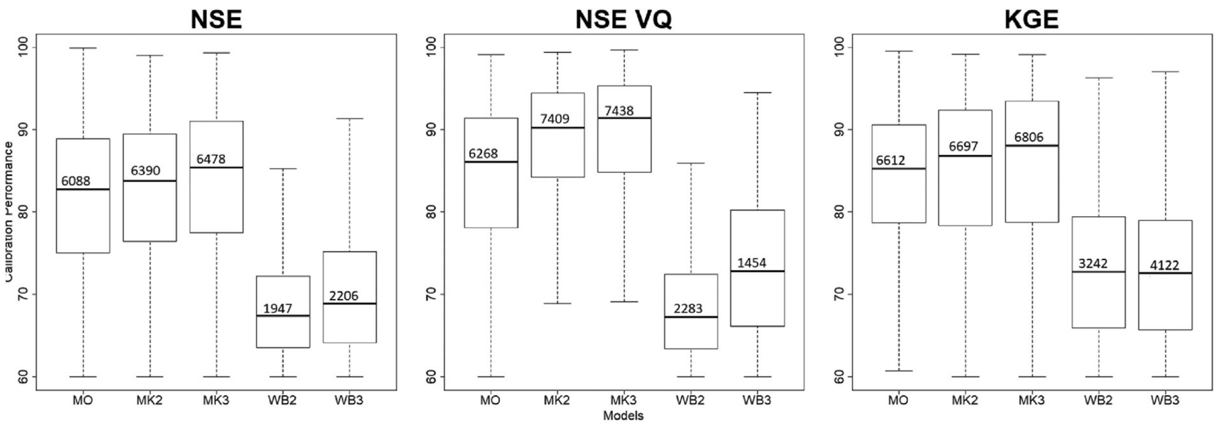

Fig. 4. Summary of goodness-of-fit criteria values across the 241 catchments for the calibration of the five rainfall–runoff models for the modelling periods according to the models. The number of runs with at least 60 as performance is on the median line of the boxplots.

the fact that the two models are not very sensitive to this parameter.

More specifically, the analysis of the calibration performance of the models shows also behaviour differences between GOFC components. Fig. 4 shows that GR models are the most efficient in terms of NSE($\sqrt{Q}$) and WB models in terms of KGE (confirmed by number of runs). Does that mean that the structure of a model and GOFC are linked? This could be partly explained by the fact that models are generally implemented using a particular GOFC.

Because of highest values for NSE($\sqrt{Q}$), the assumption can be made that the structure of GR models is the most capable of reproducing the entire hydrograph than reproducing the highest flows. With higher KGE coefficients compared to NSE, the structure of WB models gives a better representation of the variability of the flows than the flows themselves.

The analysis of calibration performance displays overall similar trends of calibration performances according to the annual precipitation amount, except for the highest ones (Supplementary Material 2). For the WB3 model, the calibration performance increases with the average annual rainfall whatever the GOFC. For WB2 confronted in terms of NSE criteria, performance increases with annual rainfall and falls around the average rainfall of 2600 mm/year. In terms of KGE, the performance increases with annual rainfall. It has been proven above that WB models better reproduce flow variability. The more variable the climatic conditions are over the calibration period, the more difficult it is for the model to reproduce the flows. Wet catchments often display less variability. This could explain WB3's behaviour with rain. The one of WB2 is probably different because of the replacement of the third parameter by WHC_{max}, a measurable value in the field. For the GR models in whatever NSE criterion, performance increases slightly with annual rainfall, before falling to around 2600 mm/year in rainfall. It is indeed less clear for the MO model. For the KGE criterion, performance drops to around 1000 mm/year in rainfall, from where it rises with rainfall and then falls to around 2600 mm/year. We do not find any explanation for this overall behaviour.

As with annual rainfall, the behaviour of model calibration performance as a function of PET depends on the model and the GOFC (Supplementary Material 3). For the MO model and for all GOFCs, the calibration performance increases with annual PET up to around PET 1800 mm/year and drops sharply to around PET 2200 mm/year. For MK models and in terms of NSE(Q), the same behaviour as that of MO is recorded. In terms of NSE($\sqrt{Q}$) and KGE, there is a slight decrease in performance up to around PET 1200 mm/year, which then rises to 1800 mm/year, from which it falls to 2200 mm/year. The WB models show very irregular behaviour with PET. For WB2, the performance decreases with PET in terms of NSE. In terms of NSE($\sqrt{Q}$) and KGE, it drops to values of 1400 mm/year and 1600 mm/year. It then rises to the value of 1800 mm/year and 2000 mm/year, from which it falls to 2200 mm/year. As for WB3, it seems to be insensitive to PET because a clear relationship does not emerge between these calibration performances and PET. An explanation for these behaviours is hardly found because the previous tests show a very low sensitivity of the studied models with respect to PET. An analysis of the calibration performance in terms of hydrological regimes of the catchment was carried out against the background of a geographical map. No link was found between the proper calibration of flows and the hydrological regime of the considered catchment.

4.2. Analysis of the robustness of the models to climate-contrasted periods

As described earlier, the GSST procedure is used over five year periods for 241 catchments. Fig. 5 (according to the GOFC) shows the performance loss in the GOFC compared to the model calibrations for all of the catchments and all the periods. The cumulative distribution function of performance loss in the GOFC (*y*-axis) is plotted versus the percentage difference in the GOFC (*x*-axis) in the simulation period (receiver period) relative to the calibration period (donor period). The percentage difference is calculated in such a way that the donor period is used as the denominator (Eq. (5)). Negative values on the *x*-axis represent simulation results where

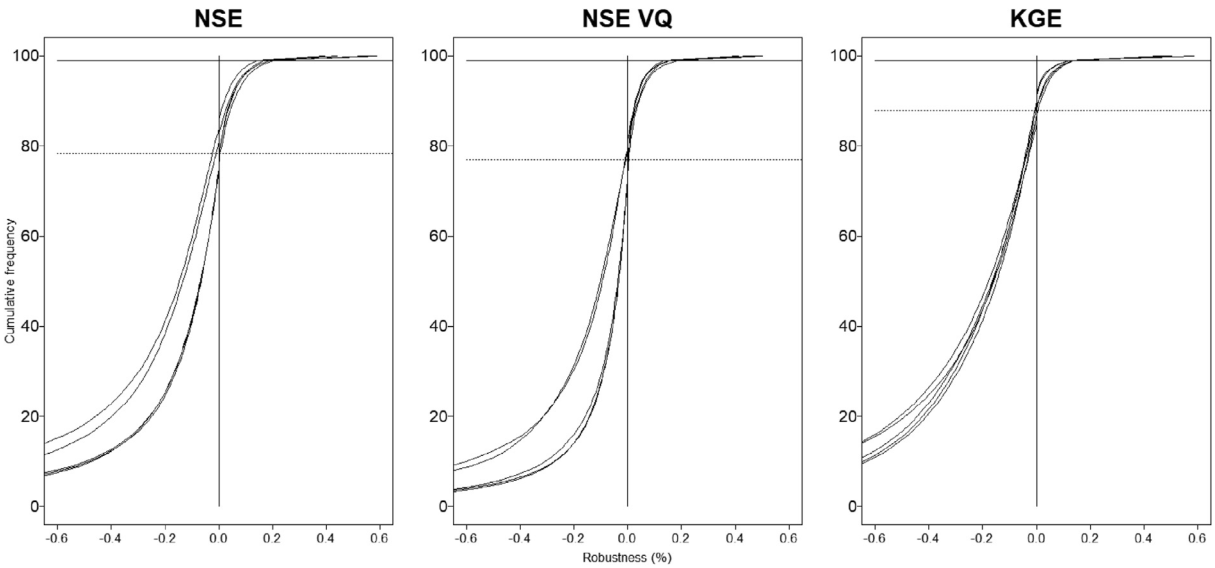

Fig. 5. Cumulative frequency graphs of model robustness.

there is a reduction in GOFC compared to calibrated GOFC values and vice-versa.

Fig. 5 shows an interesting point in about 10% (up to 15–20%) of the simulations; the simulated GOFC values can be greater than the calibrated GOFC values. This means that parameters from a donor period transferred over a receiver period lead to higher GOFC values than over the donor period. This is possible if, between the two periods, donor and receiver, the values of GOFCs in calibration are sufficiently different, and in particular if the GOFC in calibration of the donor period is lower than the GOFC in calibration of the receiver period. For a reason that is not necessarily known (data? punctual change over time in the rain-runoff relationship? etc.), the model does not match well for one period, while it matches better for another one.

An overall analysis of models robustness with respect to the performance criterion shows that the GR models are more robust than the WB models, whatever the GOFC. The order of magnitude of robustness is higher, in terms of NSE($\sqrt{Q}$), then NSE(Q) and finally KGE.

The robustness of GR models varies according to the chosen GOFC. It is almost the same in terms of NSE(Q) and slightly different in terms of NSE($\sqrt{Q}$) and KGE. The MK models have very close robustness which is higher than MO one in terms of NSE or NSE($\sqrt{Q}$). On the other hand, in terms of KGE, the MO model is more robust than MK models.

The comparison of the robustness of the versions of MK and WB models shows that the version with three parameters is slightly more robust than the version with two parameters.

The three graphs in Fig. 5 show a break at the percentile 99% of the cumulative frequency graph for all studied models and GOFC and around 20% of transfers gain in performance. No explanation is found for the catchments for which the robustness is high. Differences exist from one model to another and from one criterion to another. But, in general, models are all robust for Central Africa pure equatorial catchments. Very wet watersheds with less variation in hydroclimatic conditions would be easier to model.

When comparing these preliminary results on robustness to those on calibration previously obtained, it appears that there is a link between the chosen GOFC and a model. The GR models have better calibration and robustness performances with NSE($\sqrt{Q}$) compared to the WB models that have better performances with KGE.

In terms of rainfall, a link can be established between model robustness (performance loss) and the difference in mean rainfall between calibration and simulation periods (Supplementary Material 4). The behaviour is different depending on the GOFC and the model chosen. Nonetheless, for all models and selected GOFCs, the minimum of performance loss (median) is for a relative change in P of $\pm 5\%$: the receiver period is relatively 5% wetter or drier than the donor period. In the same way, the performance loss in GOFC values is not symmetric for positive and negative relative changes in mean rainfall. The performance loss in GOFC values is generally greater when a model calibrated over a wet period is used to model runoff over a drier period compared to when a model calibrated over a dry period is used to model runoff over a wetter period. For WB models and regardless of the chosen GOFC, as expected, the performance-loosing GOFC values are generally higher for a larger difference between the rainfall in the calibration (donor) and simulation (receiver) periods. For the GR models, it was observed a particular behaviour since there is a slight inflexion point for the positive ΔP (the receiver period is relatively wetter than the donor period), at $\Delta P = +30\%$: the median of the performance losses at $\Delta P = +50\%$ corresponds to a less significant loss of performance than for $\Delta P = +40\%$. This observation should perhaps be put into perspective since the number of runs that could be made corresponding to $\Delta P = +50\%$ is much less important than the number of runs corresponding to $\Delta P = +40\%$. However, this observation is common to all GOFCs, which still gives it some weight. The analysis undertaken in this study does not provide any explanation for these results and more analysis is required to investigate the details of the calibrated parameter values and link them to the model structure.

The range of performance losses during parameters transfers depends on the chosen model and the GOFC. However, it is minimal for a $\Delta P = 0\%$. For NSE(Q), it varies overall as the median. For KGE, the range remains significant, regardless of the observed rainfall change.

Therefore, it appears that in terms of median of performance losses by class of ΔP, but also in terms of range, it is better to calibrate the models on a dry period than on a wet period. The NSE criteria, and in particular the NSE($\sqrt{Q}$) criterion, also allow a higher robustness of the GR models. The GR models have very similar robustness performances.

To have an overview of the level of error obtained when parameters are transferred under similar rainfall

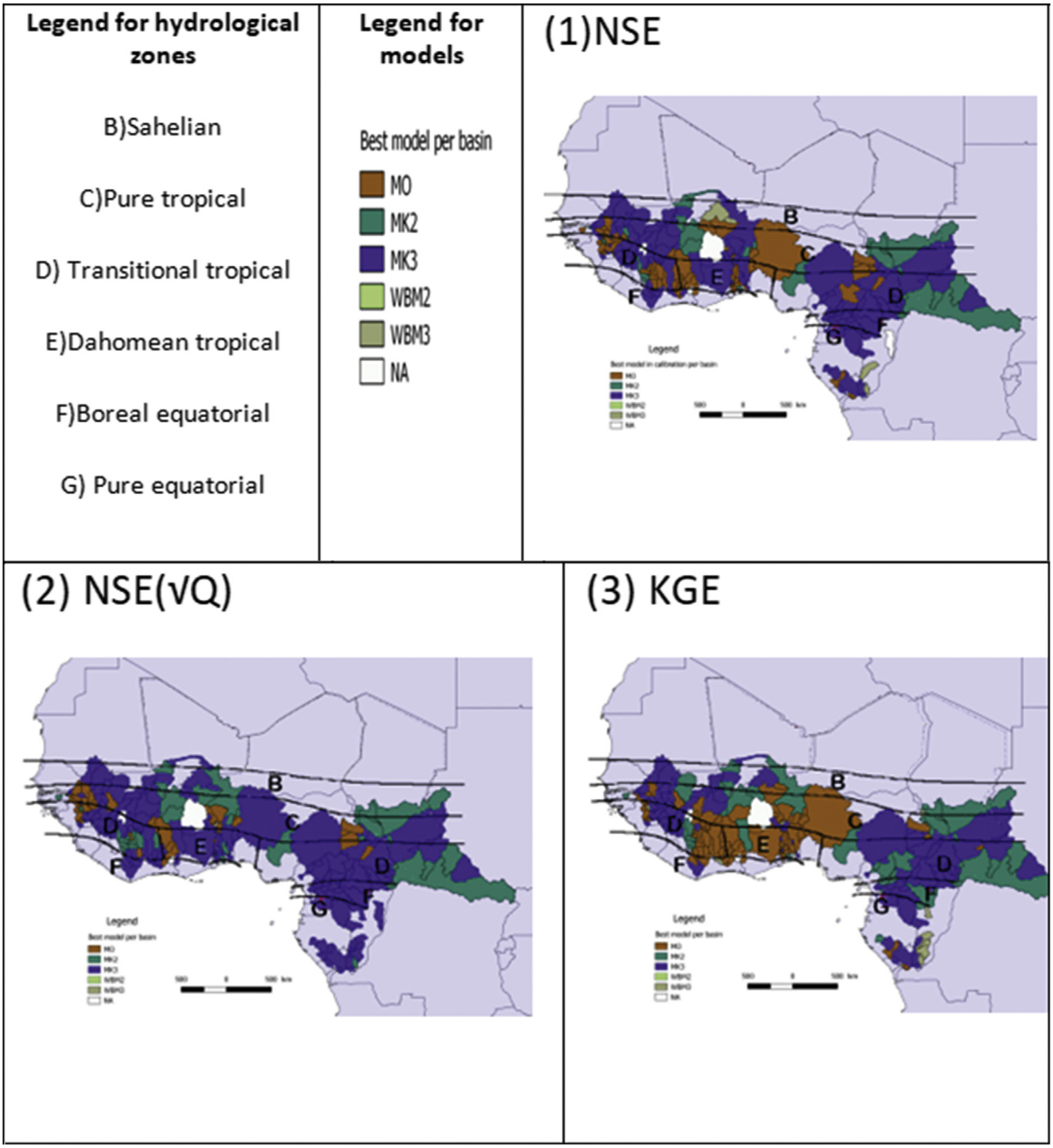

Fig. 6. Best model in robustness according to the hydrological regime.

conditions, averaging the box plots corresponding to −10% and +10% can be done. With NSE($\sqrt{Q}$), the median of performance loss is inferior to 10%; slightly higher with NSE(Q) and around 15–20% with KGE. With NSE($\sqrt{Q}$), the range is then in the order of ±20% on either side of the median. Calibrating GR models with NSE($\sqrt{Q}$) on a dry period and applying them on a wetter period will lead in 75% of cases to a loss of performance that will be inferior to 15–20% at most. This can be considered acceptable if the calibration is considered of good quality.

The analysis of the robustness of the model to PET variation shows much more consistent results in terms of models and chosen criteria (Supplementary Material 5). The variations observed in terms of PET are much lower than those of rainfall and are only in the order of a few percent in the study area: −6.5% to +6.5%. However, the magnitude of the performance losses is of the same order of magnitude as that obtained with a change in rainfall. The WB models are much less robust than the GR models. NSE criteria also give better robustness qualities. As for rainfall, the performance loss in GOFC values is not symmetric for positive and negative relative changes in mean rainfall. However, the dissymmetry to change in mean PET is opposite to change in mean rainfall. The loss of performance is lower when the parameters are transferred from a period with higher PET than in the opposite case. The minimum loss of performance corresponds to a ΔPET = 0%. As with rainfall, the NSE($\sqrt{Q}$) leads to the best robustness of GR models. In a range from ΔPET = −1% to +1%, the loss of performance is in 75% of cases inferior to 15% for MK, 20% for MO.

Using Eq. (5) of Dakhlaoui et al., 2017, robustness analysis led to results similar to those obtained with conventional robustness calculation. The remark made by the authors that the conventional method would exaggerate the performance losses is not verified in this case.

Fig. 6 is a set of maps giving a spatial representation of the robustness of the different models according to GOFC. These maps were cross-referenced with those of hydrological regimes to also allow an analysis of the robustness of the models in relation to hydrological regimes. For each GOFC and for each calibration/validation run (all periods combined), the number of times a model is the most robust has been counted, and the one with the highest number has been selected.

Fig. 6 shows that in terms of NSE(Q) and NSE($\sqrt{Q}$), the MK3 model is slightly more robust than MK2s, whatever the hydrological regime of the catchment may be. In terms of KGE, the MO model seems to be most robust for the West African catchments of pure equatorial and equatorial boreal transitional regimes and for transitional tropical regime Dahomean catchments. For other hydrological regimes, the MK2 and MK3 models are equivalent.

5. Conclusion

This study compared the robustness of five monthly conceptual models widely used in West and Central Africa: three GR and two WB models. The investigation was undertaken on 241 catchments covering the main river catchments of the region. The models were tested under three GOFC: NSE(Q), NSE($\sqrt{Q}$) and KGE. For the analysis of the results, only those obtained with parameters giving at least 60 as the calibration performance were considered.

An analysis of calibration performance showed that GR models are more performing than WB models, regardless the GOFC. The performance of MK models is somewhat similar, albeit slightly higher than that of MO models. The behaviour of calibration performance according to climate variables is different from one model to another and from one GOFC to another. No relationship was found between calibration performance and the hydrological regime.

Robustness analysis also showed GR models to be more robust than WB models. The models are most robust in terms of NSE($\sqrt{Q}$). They seem to work better when the same weight is given to all parts of the hydrograph. MK models are the most robust in terms of NSE(Q) and NSE($\sqrt{Q}$). In terms of KGE, the MO model is the most robust. WB is the most robust in terms of KGE than any other GOFC.

The robustness of MK models is very similar independently of the number of calibrated parameters. This confirms the hypothesis emitted by Ouédraogo (2001), who proposed to replace the third parameter of the MK3 model by a Water Holding Capacity mapped by FAO.

The analysis of the robustness of the models according to climatic variations between calibration and validation periods (Rain and PET) showed that it is better to transfer the parameters from a drier and a higher PET period than doing the reverse. A more refined analysis showed that a rain variation of −10%–10% or PET of −1%–1% between the calibration and validation period causes:

- a loss of robustness less than 10% for GR models and 30% for WB models in terms of NSE(Q) and NSE($\sqrt{Q}$);
- a loss of robustness between 15 and 20% for GR models in terms of KGE.

Authors such as Vaze et al. (2011), Coron et al. (2012), who worked on Australian catchments, and Dakhlaoui et al., 2017, who worked on northern Algerian catchments, reached conclusions similar to those in the present study. Coron et al. (2012) also found that the difference between calibration and validation periods should not exceed 10% for better transfer results.

A question then arises of why does the behaviour of models differ from one climate to another? From one model to another? From one catchment to another? From one GOFC to another? Some authors such as Wagener et al. (2001), Wilby (2005), Coron et al. (2012), Thirel et al. (2015a) believe that the degree of transferability of a model is related to the sensitivity of the model to its parameters over the calibration period. According to them, the sensitivity of the parameters depends on the processes that predominate in the flow over the calibration period. The more sensitive the parameter over a period, the more likely it is to be the optimal parameter for other periods. The next step in this study will be on the sensitivity of the models to their parameters. Previous studies (Ouermi et al., 2015) have not shown a clear relationship between the local sensitivity of models to parameters and their degree of transferability. A more detailed study will be conducted on the equifinality of the parameters, the global sensitivity of the models to the parameters, and their relationship with the climatic conditions of the calibration periods.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crte.2019.08.001.

References

Boyer, J.-F., Dieulin, C., Rouché, N., Cres, A., Servat, E., Paturel, J.-E., Mahé, G., 2006. SIEREM: an environmental information system for water resources. Climate Variability and Change—hydrological Impacts. In: Proceedings of the Fifth FRIEND World Conference Held at Havana, Cuba, November 2006, vol. 308. IAHS Publ, pp. 19–25.

Conway, D., 1997. A water balance model of the upper blue Nile in Ethiopia. Hydrol. Sci. J. 42 (2), 265–286. https://doi.org/10.1080/02626669709492024.

Coron, L., Andreassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., Hendrickx, F., 2012. Crash testing hydrological model in constrasted climate conditions: an experiment on 216 Australian catchments. Water Resour. Res. 48 https://doi.org/10.1029/2011WR011721. W05552.

Dakhlaoui, H., Ruelland, D., Tramblay, Y., Bargaoui, Z., 2017. Evaluating the robustness of conceptual rainfall-runoff models under climate variability in northern Tunisia. J. Hydrol. 550, 201–217. https://doi.org/10.1016/j.jhydrol.2017.04.032.

Deonarain, B., 2014. https://350.org/fr/8-impacts-du-changement-climatique-qui-affectent-deja-lafrique/.

Dezetter, A., Girard, S., Paturel, J.-E., Mahé, G., Ardoin-Bardin, S., Servat, E., 2008. Simulation of runoff in West Africa: is there a single data-model combination that produces the best simulation results? J. Hydrol. 354, 203–212.

FAO, 1995. Digital Soil Map of the World and Derived Soil Properties (CDROM). FAO Land and Water Digital Media Series.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377 (1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Harris, I., 2017. https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.00/Release_Notes_CRU_TS4.00.txt.

Hounpke, J., 2016. Assessing the Climate and Land Use Changes Impact on Flood Hazard in Ouémé River Basin, Benin (West Africa). PhD thesis, 169 p.

IPCC, 2013. In: Stocker, T.F., Qin, D., Plattner, G.-K., MTignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 p.

IPCC, 2014. In: Pachauri, R.K., Meyer, L.A. (Eds.), Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team. IPCC, Geneva, Switzerland, 151 p.

Klemes, V., 1986. Operational testing hydrological of hydrological simulations models. Hydrol. Sci. J. 31, 1. https://doi.org/10.1080/02626668609491024.

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

Makhlouf, Z., Michel, C., 1994. A two-parameter monthly water balance model for French watersheds. J. Hydrol. 162, 299–318.

Mouelhi, S., 2003. Vers une chaîne cohérente de modèles pluie-débit conceptuels globaux aux pas de temps pluriannuel, annuel, mensuel et journalier. ENGREF, Cemagref Antony, France, p. 323. Thèse de doctorat.

Mouelhi, S., Michel, C., Perrin, C., Andréassian, V., 2006. Stepwise development of a two-parameter monthly water balance model. J. Hydrol. 318 (1–4), 200–214. https://doi.org/10.1016/j.jhydrol.2005.06.014.

Nash, J.E., Sutcliffe, J.V., 1970. Riverflow forecasting through conceptual models. J. Hydrol. 10, 282–290.

Niang, I., Ruppel, O.C., Abdrabo, M.A., Essel, C., Lennard, J., Padgham, J., Urquhart, P., 2014. Africa. In: Barros, V.R., Field, C.B., Dokken, D.J., Mastrandrea, M.D., Mach, K.J., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., Girma, B., Kissel, E.S., Levy, A.N., MacCracken, S., Mastrandrea, P.R., White, L.L. (Eds.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1199–1265.

Ouédraogo, M., 2001. Contribution à l'étude de l'impact de la variabilité climatique sur les ressources en eau en Afrique de l'Ouest. Analyse des conséquences d'une sécheresse persistante : normes hydrologiques et modélisation régionale. Ph.D. thesis. Université Montpellier-2, France, 257 p.

Ouermi, S., Paturel, J.-E., Karambiri, H., 2015. Transposabilité temporelle des paramètres de modèles hydrologiques dans un contexte de changement climatique en Afrique de l'Ouest. Hydrol. Sci. J. https://doi.org/10.1080/02626667.2015.1072275.

Paturel, J.-E., 2014. Exercice de scénarisation hydrologique en Afrique de l'Ouest – Bassin du Bani. Hydrol. Sci. J. 59 (6), 1135–1153. https://doi.org/10.1080/02626667.2013.834340.

Panthou, G., Vischel, T., Lebel, T., 2014. Short communication: recent trends in the regime of extreme rainfall in the central Sahel. Int. J. Climatol. 34, 3998–4006. https://doi.org/10.1002/joc.3984.

Paturel, J.-E., Servat, E., Lubes, H., Kouame, B., Ouedraogo, M., Masson, J.M., 1997. Climatic variability in humid Africa along the Gulf of Guinea – Part 2: An integrated regional approach. J. Hydrol. 191, 16–36.

Rodier, J., 1964. Régimes hydrologiques de l'Afrique noire à l'ouest du Congo. ORSTOM, Paris, 163 p.

Seiller, G., Anctil, F., Perrin, C., 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. Hydrol. Earth Syst. Sci. (4), 1171–1189. https://doi.org/10.5194/hess-16-1171-2012.

Servat, E., Dezetter, A., 1988. Modélisation globale de la relation pluie-débit: des outils au service de l'évaluation des ressources en eau. Hydrol. Cont. 3 (2), 117–129. ISSN 0246–1528.

Thirel, G., Andréassian, V., Perrin, C., Audouy, J.-N., Berthet, L., Edwards, P., Folton, N., Furusho, C., Kuentz, A., Lerat, J., Lindström, G., Martin, E., Mathevet, T., Merz, R., Parajka, J., Ruelland, D., Vaze, J., 2015a. Hydrology under change: an evaluation protocol to investigate how hydrological models deal with changing catchments. Hydrol. Sci. J. 60 (7–8), 1184–1199. https://doi.org/10.1080/02626667.2014.967248.

Taylor, C.M., Belušić, D., Guichard, F., Parker, D.J., Vischel, T., Bock, O., Harris, P.P., Janicot, S., Klein, C., Panthou, J., 2017. Frequency of extreme Sahelian storms tripled since 1982 in satellite observations. Nature 544 (7651), 475–478.

Thirel, G., Andréassian, V., Perrin, C., 2015b. On the need to test hydrological models under changing conditions. Hydrol. Sci. J. 60 (7–8), 1165–1173. https://doi.org/10.1080/02626667.2015.1050027.

Thornthwaite, C.W., Mather, J.R., 1957. Instructions and Tables for Computing Potential Evapotranspiration and the Water Balance: Centerton, N.J., Laboratory of Climatology. Publications in Climatology. Drexel Institute of Technology, Laboratory of Climatology, Centerton, NJ, USA, pp. 185–311.

Vaze, J., Chiew, F.H.S., Perraud, J.M., Viney, N., Post, D., Teng, J., Wang, B., Lerat, J., Goswami, M., 2011. Rainfall-runoff modelling across Southeast Australia: datasets, models and results. Aust. J. Water Resour. 14.

Vörösmarty, C.J., Moore III, B., 1991. Modeling basin-scale hydrology in support of physical climate and global biogeochemical studies: an example using the Zambezi River. Surv. Geophys. 12 (1–3), 271–311. https://doi.org/10.1007/BF01903422.

Vörösmarty, C.J., Moore III, B., Grace, A.L., Gildea, M.P., Melillo, J.M., Peterson, B.J., Rastetter, E.B., Steudler, P.A., 1989. Continental scale models of water balance and fluvial transport: an application to South America. Glob. Biogeochem. Cycles 3, 241–265. https://doi.org/10.1029/GB003i003p00241.

Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V., Sorooshian, S., 2001. A framework for development and application of hydrological models. Hydrol. Earth Syst. Sci. 5 (1), 13–26.

Wang, S., Ancell, B.C., Huang, G.H., Baetz, B.W., 2018. Improving robustness of hydrologic ensemble predictions through probabilistic pre- and post-processing in sequential data assimilation. Water Resour. Res. 54, 2129–2151. https://doi.org/10.1002/2018WR022546.

Wilby, R.L., 2005. Uncertainty in water resource model parameters used for climate change impact assessment. Hydrol. Process. 19, 3201–3219. https://doi.org/10.1002/hyp.5819.