A Robust calibration/validation protocol of a hydrological model using hidden Markov states

Etienne Guilpart [1], Vahid Espanmanesh [1], Amaury Tilmant [1], and François Anctil [1]

[1]Département de génie civil et de génie des eaux, Université Laval, Québec, Canada

Correspondence: E. Guilpart (etienne.guilpart@gmail.com)

Abstract. The impacts of climate and land-use changes make the stationary assumption in hydrology obsolete. Moreover, there is still considerable uncertainty regarding the future evolution of the Earth's climate and the extent of the alteration of flow regimes. In that context, it is crucial to assess the performance of a hydrologic model over a wide range of climates and their corresponding hydrologic conditions. We propose a calibration/validation protocol based on the differential split-sample test and numerous, contrasted, climate sequences identified through a Hidden Markov Model (HMM) classification. The proposed protocol is tested on the Senegal River in West Africa. Results show that when the time series of river discharges does not exhibit a clear climate trend, or when it has multiple change points, classical rupture tests are useless and HMM classification is a viable alternative as long as the climate sub-sequences are long enough.

1 Introduction

According to some authors, humanity has entered a new geological Epoch, the Anthropocene, characterized by rapid environmental changes due to human activities (Falkenmark et al., 2019). Among those activities, the massive release of carbon dioxide since the industrial revolution is expected to lead to global warming, which in turn will affect the hydrological cycle (Gleeson et al., 2020). In the past, water engineers were able to design and operate water infrastructure based on the assumption that the climate was stationary, and hence that time series of recorded hydrologic variables such as precipitation and river discharge were representative of future hydrologic conditions (Bernier, 1977; Payrastre, 2003; Naghettini, 2017). Now that the climate is changing, this assumption of stationarity is considered obsolete or even "dead" according to Milly et al. (2008). To deal with this issue, water planners and managers have devoted significant efforts to the development of new decision analytic frameworks that explicitly capture the uncertainties attached to climate change and its impacts on water resources (Brown and Wilby, 2012; Prudhomme et al., 2010).

There are essentially two categories of decision-analytic frameworks : top-down versus bottom-up. The first relies on the sequential coupling of models: GCM models are run to project future precipitations and temperatures which are then downscaled and used as inputs to hydrologic models whose outputs are then processed by water systems models (Peel and Blöschl, 2011). This is consistent with the traditional "predict-then-act" decision-making paradigm (Weaver et al., 2013). The second category

rather seeks to identify robust solutions, i.e. solutions that will perform relatively well across a wide range of hydrologic conditions (Lempert et al., 2006). In terms of decision-making paradigm, the idea here is to "minimize regret".

30   Despite their differences, both frameworks rely at some point on a hydrological model to transform the climate forcings into streamflows. The hydrological model can be stochastic (Borgomeo et al., 2014; Poff et al., 2016), distributed or conceptual (Fortin et al., 2007; Ludwig et al., 2009). When the model is conceptual, its performances must be assessed over contrasting climatic periods because it should be able to perform well over contrasted hydro-climatic conditions (Klemes, 1986). For that purpose, the differential split-sample test principle of Klemes (1986) suggests dividing the whole period into independent periods with different stationary climate features. The hydrological model is then calibrated on a specific climate period and
35 validated on other(s). Also, the technique used to sub-divide a time series is key when calibrating/validating a hydrological model (Thirel et al., 2015a,b; Stephens et al., 2019; Huang et al., 2020).

     Several methods have then been proposed to detect shifts in climate regimes (see e.g. Liu et al. (2016) for a review of detection methods), including the Mann-Kendall test (Mann, 1945; Kendall, 1948) and the Pettitt test (Pettitt, 1979). Those tests can
40 be used to detect trends or ruptures in streamflow observations (Drogue, 2013; Diop et al., 2017; Ali et al., 2019). However, they can only make the distinction between two periods, before and after the change point, and are therefore unable to handle more complex climate sequences with multiple change points. In certain regions, for example, time series of river discharges are characterized by low-frequency shifts, and hence multiple change points, indicating that the underlying hydrological processes are influenced by low-frequency climate signals such as El Nino Southern Oscillation (Bracken et al., 2014; Nalley et al., 2019).
45

     Hidden Markov Models (HMMs) can be used to identify a succession of subsequences in a time series (Rabiner, 1989). Rather than focusing on shifts in the mean of a process, HMMs estimate shifts in the state of a process (Whiting et al., 2004). In other words, a HMM labels the observations according to their state, which ultimately leads to a new time series with states alongside the original time series. If the latter is a time series of river discharges, then the HMM will generate a new time series
50 of climate states. In hydrology, HMMs are typically used to analyze time series exhibiting a regime-like behaviour characterized by long-term persistence (Akintug and Rasmussen, 2005; Whiting et al., 2004; Turner and Galelli, 2016).

     In this article, we propose a calibration/validation protocol integrating a classification obtained by HMM that can handle complex hydrologic sequences. The goal is to improve the robustness of the calibration/validation of a hydrological model, which is a prerequisite to climate change impact assessment. This is illustrated using the Senegal River Basin (SRB) as a case
55 study.

     The paper is organized as follows. We first describe the case study in section 2, and then the selected lumped hydrological model (section 3). The proposed calibration/validation protocol and its application are presented in the section 4. In the next
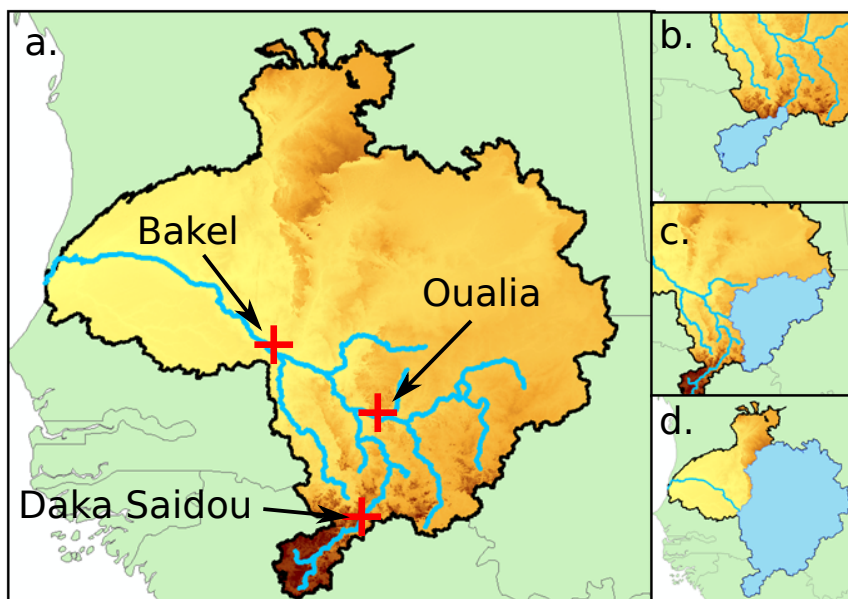
Figure 1. The SRB and its sub-basins boundaries. Red crosses represent sub-basin outlets (a), while sub-basin superficies are shaded in blue (b. Daka-Saidou, c. Oualia, d. Bakel).

60   section 5, the results are described and discussed. Finally, concluding remarks are given.

2   The Senegal River Basin and its sub-basins

The Senegal River drains a basin shared by four countries in West Africa : Guinea, Mali, Mauritania, and Senegal. There are three main tributaries: (i) the Bafing River contributing to ∼ 50% of the Senegal flows, (2) the Bakoye River (∼ 15%), and (iii)

65   the Faleme River (35%). Flowing northward on 500 km, the Bafing River collects precipitation on the Fouta Djallon, a high plateau considered as the water tower of West Africa. After merging with the Bakoye, the Senegal River runs north-west on 200km before the confluence with the Faleme River at Bakel, the last major tributary. After Bakel, the river meanders over 800 km through the floodplain and then discharges into the Atlantic Ocean.

70   The basin is located in the Soudano-Guinean zone, which is yearly influenced by the monsoon, a rainy season from April to October (Lahtela, 2003; Bodian, 2011). A consequence of the monsoon is a strong north/south precipitation gradient, ranging from 1900mm/y in the south to 100mm/y in the north (Bader et al., 2014; Bodian et al., 2015). In addition, precipitations present strong annual and inter-annual historical variabilities (Faye et al., 2015), with a wet episode (1950s-1970s) and a dry episode (1970s-1990s). With this historical climatic variability, as well as a strong spatial heterogeneity of its hydroclimatic

75   components, the SRB is an interesting case study to analyze the robustness of hydrological models; that is, their ability to

Sub-basins	River	Area (km^2)	Isohyets ranging (mm/y)	Outlet coordinates
Daka Saidou	Bafing	15 897	1500-2000	11,96°N; 10,63°W
Oualia	Bakoye	102 611	500-1500	13,61°N; 10,38°W
Bakel	Senegal	393 754	2000-400	14,91°N; 12,47°W

Table 1. List of the SRB sub-basins. Superficies have been calculated with the GRASS-3.4 model and 1arc sec SRTM elevation data. Indicative isohyets ranging are extracted from Faye et al. (2015).

perform well under contrasted hydrologic conditions.

To take advantage of the hydroclimatic specificities of the SRB and its heterogeneity, we have divided the SRB into three sub-basins (Figure 1.b,c,d and Table 1). This allows us to demonstrate the potential of the proposed protocol based on an HMM classification on basins with contrasting hydrologic characteristics. Sub-basins have been delimited using the GRASS-3.4 software, and the Shuttle Radar Topography Mission (SRTM) 1arc sec elevation data set.

3 The selected hydrological model

We have selected GR2M (Mouelhi, 2003), a monthly time-step conceptual hydrological model, because (1) it has been used in the SRB with satisfactory results (Ardoin-Bardin, 2004; Ardoin-Bardin et al., 2005; Holding, 2012; Bodian et al., 2015, 2016), and (2) the simulated flows will be processed by a hydro-economic model of the SRB working on a monthly time step (Tilmant et al., 2020).

GR2M has only two parameters: X1 and X2 controlling the production and the transfer functions respectively (Figure 2). We use the GR2M version included in the environment "airGR", developed by Coron et al. (2017). GR2M calibration/validation phase requires three time-series: (i) a time series of monthly precipitations (P) in the basin, (ii) a time series of monthly potential evapotranspirations (PET), and (iii) a time series of monthly river discharges (q) at the outlet.

4 The calibration/validation protocol

The calibration/validation protocol follows 3 steps: (1) selecting the inputs, (2) choosing the adequate objective function, and (3) identifying the climate subsequences. The optimization algorithm used for the calibration phase comes from Michel (1991). This paper's contribution lies in step 3 and how HMM can handle complex hydrologic sequences to ultimately assess the ro-
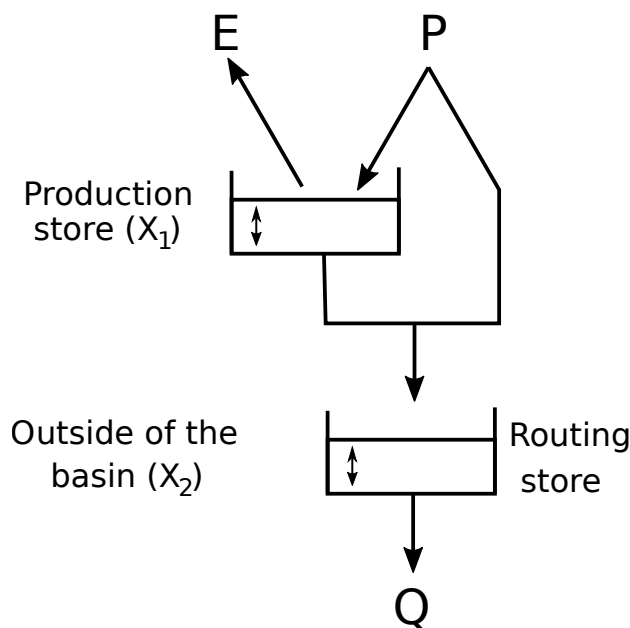
Figure 2. Scheme of the hydrological GR2M model.

bustness of a calibrated hydrological model.

4.1 Hydrological data

Some authors (Paturel et al., 1995; Hossain et al., 2004; Huard and Mailhot, 2006; Kavetski et al., 2006; Huard and Mailhot, 2008) have pointed out that selecting the most accurate hydrological and meteorological inputs can significantly reduce the Bayesian error during the calibration/validation of a hydrological model. Based on a comparison with meteorological observations compiled by SIEREM, and details given by Bader et al. (2014) about hydrological data, we have selected the following dataset: (1) Time series of precipitations were extracted from HSM-SIEREM dataset, stretching from 1940 to 1998(Boyer et al., 2006); (2) PET time series comes from the Climate Research Unite CRU (Harris et al., 2020), and covers a period from 1901 to 2018; (3) Monthly river discharge at sub-basin outlets are naturalized flows extracted from Bader et al. (2014) for the 1903-2012 period. Based on the above datasets, we will analyse the period 1940-01 / 1998-12 (59 years), which is denoted by "the full historical record" ($T^{1940-1998}$) in the remaining of this paper.

4.2 Choosing adequate objective functions

Selecting an objective function to calibrate a conceptual hydrological model is one of the main concerns of the hydrological community (Garcia et al., 2017; Krause et al., 2005; Madsen, 2003). Here, we have selected two objective functions:(1) the Nash Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970), and the (2) Kling-Gupta Efficiency criterion (KGE) (Gupta et al.,

115   2009). The former is a popular criterion and since it mainly focuses on high flows, it is particularly relevant for rivers where much of the annual discharge is generated during the high flow season, which is the case in the SRB. The latter allows for a multi-objective calibration that considers more components than just the errors; that is, correlation, bias and variability.

Mathematically, the NSE and KGE formulations can be written as:

$$NSE = 1 - \frac{\sum_{t=1}^{n}(q_t^{obs} - q_t^{sim})^2}{\sum_{t=1}^{n}(q_t^{obs} - \mu^{obs})^2} \tag{1}$$

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

with q^{obs} is the observed flow at time t, q^{sim} is the simulated flow at time t, μ^{obs} the mean of observed flows; β the ratio between the mean simulated flow and the mean observed flow value $\beta = \mu^{sim}/\mu^{obs}$; α the ratio between the standard deviation of simulated flows and the standard deviation of observed flows $\alpha = \sigma^{sim}/\sigma^{obs}$; and

$$r = \frac{\sum_{t=1}^{n}(q_t^{obs} - \mu^{obs})(q_t^{sim} - \mu^{sim})}{\sqrt{(\sum_{t=1}^{n}(q_t^{obs} - \mu^{obs})^2) * (\sum_{t=1}^{n}(q_t^{sim} - \mu^{sim})^2)}} \tag{3}$$

4.3 Identifying the climate subsequences

The calibration/validation protocol relies on the differential split-sample test proposed by Klemes (1986). The main idea is to split a period into a number of sub-periods with different hydro-climatic features. This amounts to detecting shifts in climate regimes.

130   To achieve this, we use two methods:

1. In the first method, we apply the non-parametric trend Pettitt test (section 4.3.1) to divide T into two climate sub-sequences: (1) a single dry sub-sequence noted $T_{Pettitt.dry}$, and (2) a single wet sub-sequence noted $T_{Pettitt.wet}$.

2. The second method relies on a HMM to river discharge data according to a fixed number of climate states. We apply two versions of the HMM classification:

135   – The 2-states HMM classification, in which each year is labelled as "dry" or "wet". Thus, T is divided into numerous dry or wet subsequences. All years labelled as dry are gathered under the annotation $T_{2HMM.dry}$, and all years labelled as wet under the annotation $T_{2HMM.wet}$.

– Similarly, the 3-states HMM classification, in which years are labelled as "dry", "normal" or "wet". All years labelled as dry, normal and wet are gathered under the annotations $T_{3HMM.dry}$, $T_{3HMM.nor}$, and $T_{3HMM.wet}$
140   respectively.

4.3.1 The non-parametric trend Pettitt's test:

For an given variable (q, referring to inflows for example) , the Pettitt test is defined as follow (Pettitt, 1979):

$$U_{t,T} = \sum_{i=1}^{t} \sum_{j=t+1}^{T} sng(q_i - q_j) \tag{4}$$

$$K_T = max(U_{t,T}) \tag{5}$$

145    $$p \approx 2 * exp(\frac{-6K_T^2}{T^3 + T^2}) \tag{6}$$

Since they are considered as an integrative signal of the whole basin hydro-climatic conditions, Pettitt test has been carried out on the mean annual flows. K_T gives the year of the change-point if the test is significant ($p \leq 0,05$)(Pettitt, 1979).

4.3.2 The Hidden Markov Model:

150   Hidden Markov Model is a class of probabilistic model that can be used to label the observations (Rabiner, 1989). The motivation for adopting this type of model in hydrology is that the climate regime can be represented by a state variable that can take only a limited number of values (e.g. dry/wet for 2 states; dry/normal/wet for 3 states). In other words, in parallel to the time series of historical river discharges, there exists another time series with discrete climate states. Denote $\{q_1, q_2, ..., q_T\}$ the time series of annual flows and $\{\Phi_1, \Phi_2, ..., \Phi_T\}$ the time series of climate states which can only take N possible values (Figure 3).
155

The state variable is unobserved and is accordingly referred to as a hidden variable. The hidden climate state Φ_t is modelled as a N state Markov chain (i.e the probability of a particular state depends only on the previous state) described by a transition probability matrix M with elements M_{ij}:

160    $$M_{ij} = M(\Phi_t = j | \Phi_{t-1} = i) \tag{7}$$

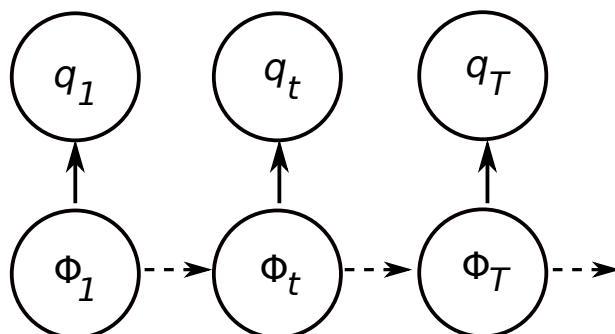With

$$t = 2, ..., T \tag{8}$$

Figure 3. Schematic graph of Hidden Markov Modelling.

$$i, j = 1, ..., N \tag{9}$$

where M_{ij} is the conditional probability of transitioning from a hidden climate state i to a hidden climate state j.

165

The observed variable q_t is assumed to have been drawn from a probability distribution whose parameters are conditional upon the distinct state at time t such that, when Φ_t is known, the distribution of q_t depends only on the current state Φ_t and not on previous states or observations.

170 $$M(q_t|q_{t-1}, ...q_1, \Phi_t, ..., \Phi_1) = M(q_t|\Phi_t) \tag{10}$$

A HMM is described by (1) the parameters of Gaussian distributions i.e., mean $\mu = (\mu_1, \mu_2, ..., \mu_N)$ and standard deviation $\sigma = (\sigma_1, \sigma_2, ..., \sigma_N)$ associated with N states, (2) the N×N matrix of transition probabilities M, and (3) the initial distribution of the Markov chain δ. Consequently, the set of parameters to be estimated is $\theta = \{\mu, \sigma, M, \delta\}$.

175 Fitting a HMM to the observed sequence (here the time series of annual flows), requires evaluating the likelihood of observing that sequence, as calculated under a N-state HMM (see Appendix A for more details). In this study, we use the Expectation-Maximization (EM) algorithm, which is an iterative method for finding the maximum-likelihood estimate of the parameters of an underlying distribution when some of the data are missing. In the context of HMM, the EM algorithm is known as the Baum-Welch algorithm (Welch, 2003) and the hidden climate states are treated as missing data (Bilmes, 1998; 180 Zucchini et al., 2017).

The EM algorithm consists of two main phases: an expectation phase called "E step", followed by a maximization phase called "M step". Given the current estimate of the HMM parameters θ, the following steps are repeated until acceptable convergence is achieved: The "E step" phase of the algorithm computes the expected value of unobserved data (i.e hidden climate states) using the current estimate of the parameters and the observed data. The "M step" phase of the algorithm then provides a new estimate of the parameters by using the data from the "E step" phase as if they were actually measured data. These parameters are then used to calculate the distribution of unobserved data in the next "E step" phase of the algorithm. The resulting values of θ is then the stationary point of the likelihood of the observed data (Please refer to Appendix B for more details).

Given the observation sequence, we want to determine the sequence of hidden climate states $\{\Phi_1, \Phi_2, ..., \Phi_T\}$ that has most likely (under the fitted HMM) given rise to the time series of annual river discharges. In the literature, this is known as the decoding procedure. In this study we use the Viterbi algorithm (Viterbi, 1967) to unfold the sequence of hidden climate states (called the Viterbi path). This, in turn, enables us to divide the whole period into numerous climate sub-sequences.

4.4 Application of the calibration/validation protocol

Recall that the identification of change-points is done on the time series of annual flows while the hydrological model simulates monthly river discharges.

When applying the calibration/validation protocol described above, seven cases arise as shown in Figure 4. Indeed:

- If relevant, the Pettitt test offers two calibration/ validation possibilities: calibration on $T_{pettitt.dry}$ and validation on $T_{pettitt.wet}$, and vice versa.

- The 2-states HMM classification offers two possibilities too: calibration on $T_{2HMM.dry}$ and validation on $T_{2HMM.wet}$, and the opposite.

- Similarly, the 3-states HMM classification leads to three possibilities: calibration on $T_{3HMM.dry}$ and validation on $T_{3HMM.wet} + T_{3HMM.wet}$, and corollaries.

Note that $T_{pettitt.wet}$ and $T_{pettitt.dry}$ are both subsequences made of contiguous years as the original time series are spit in two. In that case, the temporal persistence found in the original time series is very much preserved. However, for $T_{2HMM.wet}$ and $T_{2HMM.dry}$, the situation is different since they are made of numerous, non-contiguous, "wet" or "dry" sub-sequences respectively. This is also true for $T_{3HMM.wet}, T_{3HMM.nor}$ and $T_{3HMM.dry}$. To ensure the continuity of the internal time series of the model, the computation of the objective function (KGE or NSE) during the calibration phase or the validation phase was carried out only on the indices included in the selected climate state.

In addition, even though the KGE is based on a decomposition of the NSE, the corresponding scores cannot be directly compared. Therefore, we will discuss the results obtained with NSE and KGE separately. During all calibration phases, the
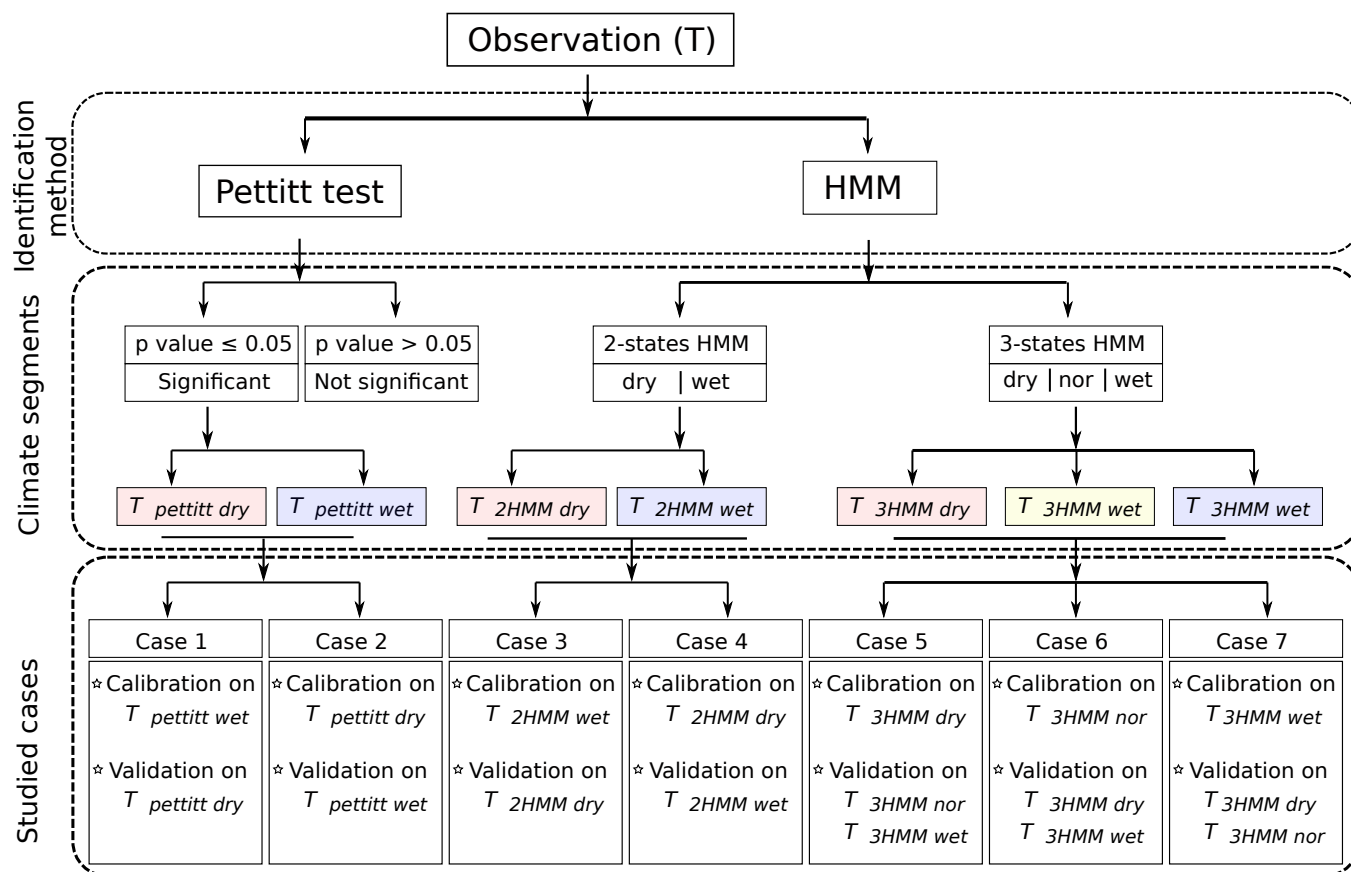
Figure 4. Application of our protocol to identify climate sequences in a given period T. Seven cases for the calibration/validation phase are obtained.

215 first year is considered as warming-up period and not considered.

5 Results and discussion

5.1 Complex climate sequences and HMM classifications

The results of the division of the full historical record $T^{1940-1998}$ are displayed in Table 2 and Figure 5a., while the corre-

220 sponding parameters are given in Table 2.

The full historical record $T^{1940-1998}$

	Basins	p value	Year break		
Pettitt test	Daka Saidou	1.10-6	1970		
	Oualia	8.10-8	1971		
	Bakel	2.10-6	1971		
	Basins	μ	σ	δ	M
2-states-HMM	Daka Saidou	183.8; 300.2	31.3; 53.8	1,0	$\begin{bmatrix} 0.96 & 0.04 \\ 0.04 & 0.96 \end{bmatrix}$
	Oualia	178.3; 65.1	32.1; 45.9	1,0	$\begin{bmatrix} 0.966 & 0.034 \\ 0.038 & 0.962 \end{bmatrix}$
	Bakel	433.9;855.1	134.8;210.0	1,0	$\begin{bmatrix} 0.963 & 0.037 \\ 0.041 & 0.959 \end{bmatrix}$
3-states-HMM	Daka Saidou	300.7; 206;162.5	22.8; 22.5;54.1	0,1,0	$\begin{bmatrix} 0.934 & 0.066 & 0 \\ 0.071 & 0.81 & 0.061 \\ 0.037 & 0 & 0.963 \end{bmatrix}$
	Oualia	178.4;87.8;37.8	11.4;25.3;45.8	0,1,0	$\begin{bmatrix} 0.785 & 0.215 & 0 \\ 0.129 & 0.81 & 0.061 \\ 0.037 & 0 & 0.963 \end{bmatrix}$
	Bakel	363.39; 553.32; 925.3	90.2; 149.0; 179.5	0,1,0	$\begin{bmatrix} 0.927 & 0.073 & 0 \\ 0.617 & 0.853 & 0.850 \\ 0 & 0.087 & 0.913 \end{bmatrix}$

Table 2. Pettitt test results and Hidden Markov Model parameters (N=2 and N=3) for Daka Saidou, Oualia, and Bakel sub-basins, on the full historical record $T^{1940-1998}$.

For the three sub-basins, the Pettitt test is significant and shows a rupture in 1970 or 1971 (Figure 5a. red vertical line). The 2-state HMM classification provides similar results with nearly aligned climate sub-sequences for all sub-basins. This is also true for the 3-states HMM classification.

225

With the 2-states or 3-states HMM classification, the states are clearly distinct, which enables us to divide the time series into numerous climate sub-sequences. The values close to one on the diagonal indicate that when the climate is in a particular state, it will likely remain in that state in the next time period (year). With a 2-states-HMM classification (Figure 5a. blue line), the dry is dominant in all sub-basins, whereas with a 3-states-HMM classification (Figure 5b. orange line), the wet state

230 is dominant in both Daka Saidou and Oualia sub-basins. This is due to the fact that some years that were labelled as "dry" with the 2-states HMM classification are now recognized as "normal" when using the 3-states HMM classification. Similarly, several years move from "wet" to "normal" with the 3-states HMM classification. However, since the number of years that are switching from dry to normal is larger than those that are switching from wet to normal, the result is a dominant wet state.

235 We note that Pettitt change-point is aligned with the 2 or 3-states-HMM transitions in 1970 (Daka Saidou) or in 1971 (Oualia and Bakel). The length of climate sub-sequences are given in Table 3.
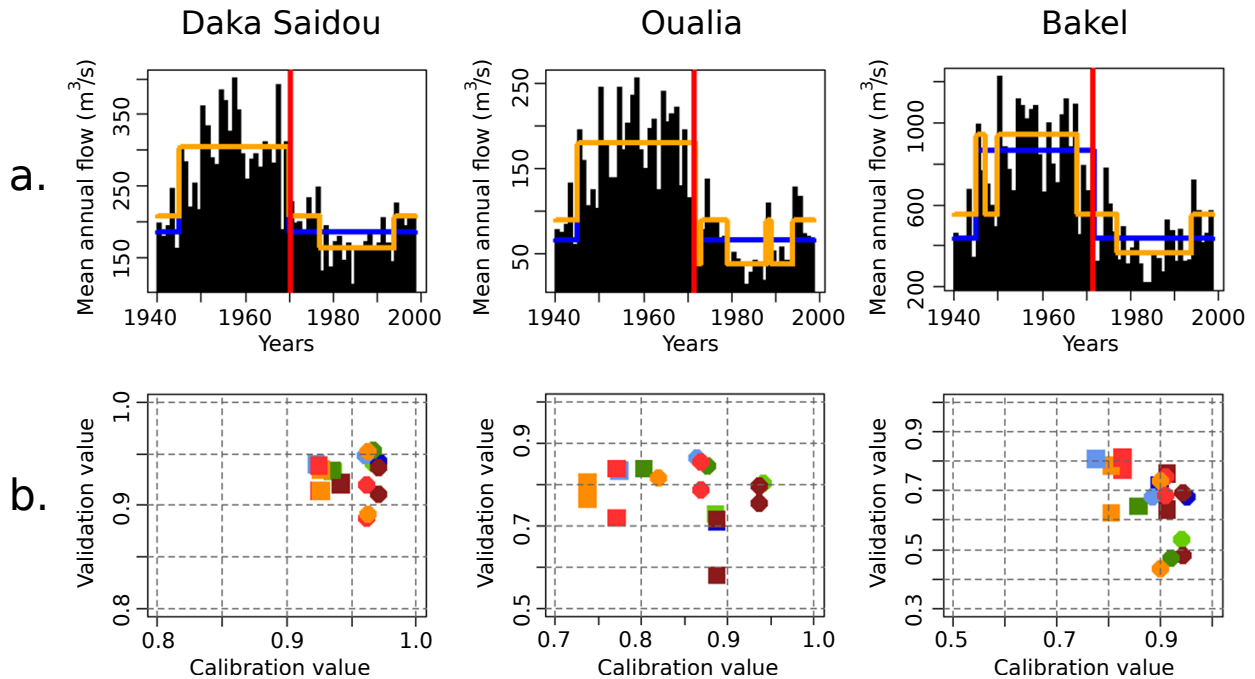
Figure 5. a. Years classifications of $T^{1940-1998}$ according to the Pettitt test (vertical red lines), 2-states-HMM (in blue) and 3-states-HMM (in orange); b. Scatter-plot of NSE (squares) and KGE (dots) calibration/validation values. Light green refers to the case 1 (Pettitt test, calibration on $T_{pettitt.wet}$ and validation on $T_{pettitt.dry}$); dark green to the case 2(Pettitt test, calibration on $T_{pettitt.dry}$ and validation on $T_{pettitt.wet}$); Light blue to the case 3 (2-states-HMM, calibration on $T_{2HMM.dry}$ and validation on $T_{2HMM.wet}$.), and dark blue to the opposite (case 4); Orange to the case 5 (3-states-HMM, calibration on $T_{3HMM.dry}$ and validation on $T_{3HMM.nor}$ and $T_{3HMM.wet}$); red to the case 6, and dark red to the seventh case.

The full historical record ($T^{1940-1998}$) could be seen as a textbook case with a clear climate trend and long climate and hydrological records. In a such situation, classical rupture tests like the Pettitt test are adequate to identify two climate sub-
240 sequences by detecting a single change point. However, 2-states and 3-states HMM classifications allow for a finer labelling of the years. For example, according to the Pettitt test, the years prior to 1945 are considered as wet even though they are classified as either dry or normal years by the 2-states and 3-states HMM classification respectively.

To further highlight the relevance of integrating HMM classifications in a calibration/validation protocol, the protocol has
245 been implemented over two other smaller periods (27 years each), which display more complex climate sequences with no clear climate trend: the period $T^{1945-1971}$ and the period $T^{1972-1998}$. Results are not shown due to a lack of space, but available for consultation in the supplementary material section.

The full historical record $T^{1940-1998}$

Daka Saidou

	Sub-sequence(s)	Pettitt's Test Case 1	Case 2	2-states HMM Case 3	Case 4	3-states HMM Case 5	Case 6	Case 7
Calibration	Dry		0.93/0.97(28y)	0.92/0.96(33y)		0.93/0.96(17y)		
	Normal						0.92/0.96(16y)	
	Wet	0.94/0.97(30y)			0.94/0.97(25y)			0.94/0.97(25y)
Validation	Dry	0.93/0.94(28y)			0.92/0.95(33y)		0.92/0.93(17y)	0.91/0.9(17y)
	Normal					0.92/0.93(16y)		0.92/0.95(16y)
	Wet		0.93/0.95(30y)	0.94/0.95(25y)		0.94/0.96(25y)	0.94/0.97(25y)	

Oualia

	Sub-sequence(s)	Pettitt's Test Case 1	Case 2	2-states HMM Case 3	Case 4	3-states HMM Case 5	Case 6	Case 7
Calibration	Dry		0.80/0.88(27y)	0.77/0.86(31y)		0.74/0.82(15y)		
	Normal						0.77/0.89(16y)	
	Wet	0.89/0.94(31y)			0.89/0.94(27y)			0.89/0.94(27y)
Validation	Dry	0.73/0.80(27y)			0.71/0.8(31y)		0.72/0.79(15y)	0.58/0.75(15y)
	Normal					0.76/0.82(16y)		0.71/0.79(16y)
	Wet		0.84/0.84(31y)	0.83/0.86(27y)		0.8/0.81(27y)	0.84/0.85(27y)	

Bakel

	Sub-sequence(s)	Pettitt's Test Case 1	Case 2	2-states HMM Case 3	Case 4	3-states HMM Case 5	Case 6	Case 7
Calibration	Dry		0.86/0.92(27y)	0.78/0.88(31y)		0.8/0.9(17y)		
	Normal						0.83/0.91(21y)	
	Wet	0.9/0.94(31y)			0.90/0.95(27y)			0.91/0.94(20y)
Validation	Dry	0.69/0.53(27y)			0.71/0.68(31y)		0.77/0.74(17y)	0.63/0.48(17y)
	Normal					0.78/0.73(21y)		0.75/0.69(21y)
	Wet		0.64/0.47(31y)	0.80/0.68(27y)		0.62/0.44(20y)	0.81/0.68(20y)	

Table 3. Table of NSE/KGE calibration and validation scores according to the seven cases for the full historical record $T^{1940-1998}$. The numbers of years used for the calibration or validation are given between brackets.

Depending on the length of period T, the climate sub-sequences can be short. This is the case with $T^{1945-1971}$ and $T^{1972-1998}$, for which climate sub-sequences provided by the 3-states HMM classification could reach a length of 5 years. Thus, the question of the minimum of years required to ensure a reliable calibration or validation raises one more time. No consensus has been reached by the hydrologist community at this point, but a number from two to eight years could be enough depending on the "hydrological events" included (Razavi and Tolson, 2013; Juston et al., 2013; Singh and Bárdossy, 2012).

In addition, we want to underline the following two points: (1) According to our protocol, when a model is calibrated on a specific climatic state, it will not be evaluated (validated) on this same state. To overcome this, the "split-sample test" of Klemes (1986) could be combined with the differential split-sample test, but a longer record could be required. (2) Please recall that the principle of the differential split-sample test could be applied to detect complex climate sequences and to detect massive conversions in land uses when applied on flows.

5.2 Towards an enhancement of calibration and validations scores with HMM classifications?

For Daka Saidou, all seven cases have high scores (≥ 0.9), indicating that the model is robust. However, for Oualia and Bakel, calibration/validation scores are more scattered (Figure 5b.).

265   When comparing the NSE and KGE values for calibration and validation, more several optimal cases are explored (seven cases in this article), which could lead to better model performances than if we had limited ourselves to the 2 cases provided by the Pettitt test.

     For Daka-Saidou, calibration/validation NSE (or KGE) scores are gathered in the same area, showing that the model will
270   have similar performances regardless of the method used to divide the period (Pettitt test, 2-states HMM or 3-states HMM classifications). For Oualia and Bakel, results are more scattered, and the HMM classifications could enhance NSE or KGE scores. This is due to the fact that Daka-Saidou is a small upstream basin with relatively homogeneous precipitation and evapotranspiration (Table 1), whereas Oualia and Bakel are larger and heterogeneous basins.

275 **6   Conclusions**

In this article, we have shown how an HMM can deal with complex climates sequences, and how the resulting classification can be used to develop a robust calibration/validation protocol. The protocol has been implemented in the Senegal River basin using the GR2M model and the historical flow from 1940-1998.

280   The main concluding remarks are:

   – When records display a single point change, a classical rupture trend (as Pettitt test) remains an adequate tool to divide the records into two climate sub-sequences.

   – If the records contain multiple change points, HMM classifications are a good alternative to divide the records into several climate sub-sequences. However, records must be long enough (typically 20-25 years for a 2-states HMM classification,
285   and 30-35 years for a 3-states HMM classification) to (i) have a sufficient number of usual and unusual hydrological events (as mentioned by Singh and Bárdossy (2012)), and (ii) to have a minimum number of years for each climate state.

   – Regardless of the division method used, the range of climate conditions over which the hydrological model can perform depends on the intrinsic variability of the series used during the calibration/validation phase.

   – HMM classifications open up the range of possibilities for calibrate/validate a hydrological model, which can lead to an
290   enhancement of the criterion function (but not necessarily).

14

Appendix A: Likelihood of Hidden Markov Models

We suppose there is an observation sequence $Q = \{q_1, q_2, ..., q_T\}$ and the associated (unobserved) state variables $\Omega = \{\Phi_1, \Phi_2, ..., \Phi_T\}$ generated by such a model. Given the set of HMM parameters $\theta = \{\mu, \sigma, M, \delta\}$, the joint density of complete data set $Z = (Q, \Omega)$ can be expressed as:

295

$$p(Z|\theta) = p(Q, \Omega|\theta) = p(Q|\Omega, \theta)p(\Omega|\theta) \tag{A1}$$

Assuming the data belonging to each hidden state are characterized by a specific Gaussian probability distribution, the two terms on the right-hand side are:

$$p(Q|\Omega, \theta) = \prod_{t=1}^{T} p(q_t|\mu_{\Phi_t}, \sigma_{\Phi_t}) \tag{A2}$$

$$p(\Omega|\theta) = \delta \prod_{t=1}^{T-1} p((\mu_{\Phi_{t+1}}|\sigma_{\Phi_t})|M) \tag{A3}$$

300   The complete data likelihood function $\zeta(\theta|Z)$ can be calculated as:

$$\zeta(\theta|Z) = \zeta(\theta|Q, \Omega) = p(Q, \Omega|\theta) \tag{A4}$$

For a HMM which has the initial distribution δ and transition probability matrix M for the Markov chain, let us define the probability mass function of Q if the Markov chain is in state i at time t as:

$$p_i(q) = p(Q = q|\Omega = i) \tag{A5}$$

305   With $i = 1, 2, ...N$

The general form of likelihood function is then given by (Zucchini et al., 2017):

$$\zeta = \delta \Gamma(q_1) M \Gamma(q_2) ... M \Gamma(q_T) 1' \tag{A6}$$

where $\Gamma(q)$ is defined as the diagonal matrix with i the diagonal element $p_i(q)$ and $1'$ is N dimensional vector of 1.

Appendix B: HMM Likelihood maximization with EM algorithm

In order to set out the likelihood computation in the form of Baum-Welch algorithm (Welch, 2003), which involves the forward $\alpha(t)$ and backward $\beta(t)$ probabilities, we define $\alpha(t)$ and $\beta(t)$ as:

$$\alpha(t) = \delta\Gamma(q_1)M\Gamma(q_2)...M\Gamma(q_t) = \delta\Gamma(q_1)\prod_{n=2}^{t} M\Gamma(q_n) \tag{B1}$$

and

$$\beta(t) = \delta\Gamma(q_{t+1})M\Gamma(q_{t+2})...M\Gamma(q_t)1' = (\prod_{n=t+1}^{T} M\Gamma(q_n))1' \tag{B2}$$

respectively. More specifically, $\alpha_i(t)$ is the probability of observing the partial sequence $q_1, q_2, ..., q_t$ and ending up in state i at time t, and $\beta_i(t)$ is the probability of observing the remaining sequence. Numerical calculation of $\alpha_i(t)$ and $\beta_i(t)$ is not trivial (Akintug and Rasmussen, 2005). Here we use the method suggested by Durbin et al. (1998) for scaling forward and backward probabilities to overcome this problem. Now let us define $u_j(t)$ and $v_{jk}(t)$ as (Zucchini et al., 2017):

$$u_j(t) = p(\Phi_t = j|Q, \theta) = \frac{\alpha_j(t)\beta_j(t)}{\zeta} \tag{B3}$$

$$v_{jk}(t) = p(\Phi_{t-1} = j, \Phi_t = k|Q) = \alpha_j(t-1)M_{jk}p_k(q_t)\beta_k(t)/\zeta \tag{B4}$$

Where M_{jk} is the probability of transition from hidden climate state j to climate state k, and ζ is the likelihood function. With EM algorithm, we aim to maximize the log-likelihood of the parameters of interest θ, based on complete data (i.e. both the observed data and the hidden climate states). Now let us represent the sequence of climate states (missing data) by the Markov chain by the zero-one random variables. The complete data log-likelihood can be formulated as:

$$\log(\zeta(\theta|Z)) = \sum_{j=1}^{N} u_j(1)\log(\delta_j) + \sum_{j=1}^{N}\sum_{k=1}^{N}(\sum_{t=2}^{T} v_{jk}(t))\log(M_{jk}) + \sum_{j=1}^{N}\sum_{t=1}^{T} u_j(t)\log(p_j(q_t)) \tag{B5}$$

where $u_j(t) = 1$ if and only if $\Phi_t = j(t = 1, 2, ..., T)$, and transition probability $v_{jk}(t) = 1$ if and only if $\Phi_{t-1} = j$ and $\Phi_t = k(t = 2, 3, ..., T)$, N is the number of hidden climate states, δ_j is the initial transition of Markov chain, and $p_j(.)$ is the probability mass function if the Markov chain is in state j at time t. Maximization of the complete data log-likelihood function is performed with the EM algorithm through an iterative process presented in Figure B1.

Author contributions. A. Tilmant suggested the integration of HMM classifications into a calibration/validation protocol. V. Espanmanesh ran the HMM onto flows to provide climate subsequences. E. Guilpart carried out all the calibrations and validations, and write this paper
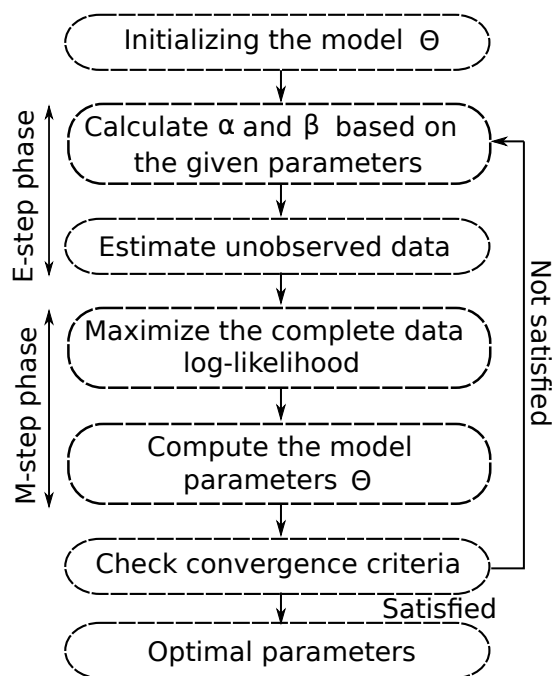
Figure B1. Expectation Maximization algorithm for a HMM parameter estimation.

(excepted the section 4.3.2 which was written by E. Espanmanesh). A. Tilmant supervised the writing. F. Anctil brought his points of view and proofred the paper.

Hydrology and
Earth System
Sciences
Discussions

References

Akintug, B. and Rasmussen, P. F.: A Markov switching model for annual hydrologic time series, Water Resources Research, 41, 1–10, https://doi.org/10.1029/2004WR003605, 2005.

340 Ali, R., Ismael, A., Heryansyah, A., and Nawaz, N.: Long term historic changes in the flow of lesser Zab River, Iraq, Hydrology, 6, 1–12, https://doi.org/10.3390/hydrology6010022, 2019.

Ardoin-Bardin, S.: Variabilité hydroclimatique et impacts sur les ressources en eau de grands bassins hydrographiques en zone soudano-sahélienne, Ph.D. thesis, Université Montpellier II, https://doi.org/10.1038/ni.2208, 2004.

Ardoin-Bardin, S., Dezetter, A., Servat, E., and Mahe, G.: Évaluation des impacts du changement climatique sur les ressources en eau d ’

345 Afrique de l ’ Ouest et Centrale, in: Regional Hydrological Impacts of Climatic Change - Hydroclimatic Variability, April, pp. 194–202, IAHS, Foz de Iguaçu (Brazil), 2005.

Bader, J.-C., Cauchy, S., Duffar, L., and Saura, P.: Monographie hydrologique du fleuve Sénégal. De l'origine des mesures jusqu'en 2011, IRD, Marseille (France), ird editio edn., https://www.documentation.ird.fr/hor/fdi:010065190, 2014.

Bernier, J.: Etude de la stationnarité des séries hydroméléorologiques, La houille blanche, 4, 313–219, 1977.

350 Bilmes, J. A.: A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, Tech. Rep. 510, International Computer Science Institute, Berkeley, https://doi.org/10.1080/0042098032000136147, http://www.leap.ee.iisc.ac.in/sriram/teaching/MLSP{_}18/refs/GMM{_}Bilmes.pdf, 1998.

Bodian, A.: Approche par modélisation pluie - débit de la connaissance régionale de la ressource en eau : Application au haut bassin du fleuve Sénégal, Ph.D. thesis, Université Cheikh Anta Diop de Dakar, http://hydrologie.org/THE/BODIAN.pdf, 2011.

355 Bodian, A., Dezetter, A., and Dacosta, H.: Rainfall-runoff modelling of water resources in the upper Senegal River basin, International Journal of Water Resources Development, 32, 89–101, https://doi.org/10.1080/07900627.2015.1026435, 2015.

Bodian, A., Dezetter, A., Deme, A., and Diop, L.: Hydrological evaluation of TRMM Rainfall over the Upper Senegal River basin, Hydrology, 3, 1–18, https://doi.org/10.3390/hydrology3020015, 2016.

Borgomeo, E., Hall, J. W., Fung, F., Watts, G., Colquhoun, K., and Lambert, C.: Risk-based water resources planning: Incorporating proba-

360 bilistic nonstationary climate uncertainties, Water Resources Research, 50, 6850–6873, https://doi.org/10.1002/2014WR015558, 2014.

Boyer, J. F., Dieulin, C., Rouche, N., Cres, A., Servat, E., Paturel, J. E., and Mahé, G.: SIEREM: An environmental information system for water resources, in: FRIEND World Conference, 308, pp. 19–25, IAHS, Havana (Cuba), 2006.

Bracken, C., Rajagopalan, B., and Zagona, E.: A hidden Markov model combined with climate indices for multidecadal streamflow simulation, Water Resources Research, 50, 7836–7846, https://doi.org/10.1002/2014WR015567, 2014.

365 Brown, C. and Wilby, R. L.: An alternate approach to assessing climate risks, Eos, 93, 401–402, https://doi.org/10.1029/2012EO410001, 2012.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., and Andréassian, V.: The suite of lumped GR hydrological models in an R package, Environmental Modelling and Software, 94, 166–171, https://doi.org/10.1016/j.envsoft.2017.05.002, 2017.

Diop, L., Yaseen, Z. M., Bodian, A., Djaman, K., and Brown, L.: Trend analysis of streamflow with different time scales: a case study

370 of the upper Senegal River, ISH Journal of Hydraulic Engineering, 24, 105–114, https://doi.org/10.1080/09715010.2017.1333045, https://doi.org/10.1080/09715010.2017.1333045, 2017.

Drogue, G.: Etudes hydro-climatologiques régionales . Applications à l ’ évolution du climat et aux écoulements de rivière dans un, Tech. rep., Université Paris-Diderot, Paris, https://tel.archives-ouvertes.fr/tel-00920438/, 2013.

Durbin, R., Eddy, S. R., Krogh, A., and Mitchison, G.: Biological Sequence Analysis, Biological Sequence Analysis, https://doi.org/10.1017/cbo9780511790492, 1998.

Falkenmark, M., Wang-Erlandsson, L., and Rockström, J.: Understanding of water resilience in the Anthropocene, Journal of Hydrology, 2, 100 009, https://doi.org/10.1016/j.hydroa.2018.100009, https://doi.org/10.1016/j.hydroa.2018.100009, 2019.

Faye, C., Diop, E. H. S., and Mbaye, I.: Impacts des changements de climat et des aménagements sur les ressources en eau du fleuve sénégal: Caractérisation et évolution des régimes hydrologiques de sous-bassins versants naturels et aménagés, Belgeo - Revue belge de géographie, 4, 1–25, https://doi.org/10.4000/belgeo.17626, 2015.

Fortin, L. G., Turcotte, R., Pugin, S., Cyr, J. F., and Picard, F.: Impact des changements climatiques sur les plans de gestion des lacs Saint-François et Aylmer au sud du Québec, Canadian Journal of Civil Engineering, 34, 934–945, https://doi.org/10.1139/L07-030, 2007.

Garcia, F., Folton, N., and Oudin, L.: Which objective function to calibrate rainfall–runoff models for low-flow index simulations?, Hydrological Sciences Journal, 62, 1149–1166, https://doi.org/10.1080/02626667.2017.1308511, http://dx.doi.org/10.1080/02626667.2017.1308511, 2017.

Gleeson, T., Wang-Erlandsson, L., Porkka, M., Zipper, S. C., Jaramillo, F., Gerten, D., Fetzer, I., Cornell, S. E., Piemontese, L., Gordon, L. J., Rockström, J., Oki, T., Sivapalan, M., Wada, Y., Brauman, K. A., Flörke, M., Bierkens, M. F. P., Lehner, B., Keys, P., Kummu, M., Wagener, T., Dadson, S., Troy, T. J., Steffen, W., Falkenmark, M., and Famiglietti, J. S.: Illuminating water cycle modifications and Earth system resilience in the Anthropocene Water Resources Research, Water Resources Research, 56, 1–24, https://doi.org/10.1029/2019WR024957, 2020.

Gupta, H. V., Kling', H., Yilmaz, K. K., and Martinez-Baquero, G. F.: Decomposition of the Mean Squared Error & NSE Performance Criteria: Implications for Improving Hydrological Modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, https://www.sciencedirect.com/science/article/pii/S0022169409004843?casa{_}token=DGYSV9oOoK4AAAAA:Qy-hbGCA2zIgzryYETf-l3s9mddklhjC72DLnJtG1mTVN6aJpS8CbjJS-KWs8ZS53wQUQCW8Uyc, 2009.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Scientific data, 7, 109, https://doi.org/10.1038/s41597-020-0453-3, 2020.

Holding, W.: Apport De La Modélisation Pluie-Débit Pour La Connaissance De La Ressource En Eau : Application Au, Climatologie, 9, 109–125, http://lodel.irevues.inist.fr/climatologie/index.php?id=223,https://doi.org/10.4267/climatologie.223, 2012.

Hossain, F., Anagnostou, E. N., Dinku, T., and Borga, M.: Hydrological model sensitivity to parameter and radar rainfall estimation uncertainty, Hydrological Processes, 18, 3277–3291, https://doi.org/10.1002/hyp.5659, 2004.

Huang, S., Shah, H., Naz, B. S., Shrestha, N., Mishra, V., Daggupati, P., Ghimire, U., and Vetter, T.: Impacts of hydrological model calibration on projected hydrological changes under climate change—a multi-model assessment in three large river basins, Climatic Change, https://doi.org/10.1007/s10584-020-02872-6, 2020.

Huard, D. and Mailhot, A.: A Bayesian perspective on input uncertainly in model calibration: Application to hydrological model "abc", Water Resources Research, 42, 1–14, https://doi.org/10.1029/2005WR004661, 2006.

Huard, D. and Mailhot, A.: Calibration of hydrological model GR2M using Bayesian uncertainty analysis, Water Resources Research, 44, 1–19, https://doi.org/10.1029/2007WR005949, 2008.

Juston, J., Seibert, J., and Johansson, P.: Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment, Hydrological Processes, 23, 3093–3109, https://doi.org/10.1002/hyp, http://jamsb.austms.org.au/courses/CSC2408/semester3/resources/ldp/abs-guide.pdf, 2013.

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, Water Resources Research, 42, 1–10, https://doi.org/10.1029/2005WR004376, 2006.

Kendall, M.: Rank correlation methods, Charles Griffin & Co. Ltd., London, 1948.

Klemes, V.: Operational testing of hydrological simulation models, Hydrological Sciences Journal, 31, 13–24, https://doi.org/10.1080/02626668609491024, 1986.

Krause, P., Boyle, D. P., and Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment, Advances in Geosciences, 5, 89–97, https://doi.org/10.5194/adgeo-5-89-2005, 2005.

Lahtela, V.: Managing the Senegal River: National and local development dilemma, International Journal of Water Resources Development, 19, 279–293, https://doi.org/10.1080/0790062032000089365, 2003.

Lempert, R. J., Groves, D. G., Popper, S. W., and Bankes, S. C.: A general, analytic method for generating robust strategies and narrative scenarios, Management Science, 52, 514–528, https://doi.org/10.1287/mnsc.1050.0472, 2006.

Liu, Q., Wan, S., and Gu, B.: A Review of the Detection Methods for Climate Regime Shifts, Discrete Dynamics in Nature and Society, 2016, 1–10, https://doi.org/10.1155/2016/3536183, 2016.

Ludwig, R., May, I., Turcotte, R., Vescovi, L., Braun, M., Cyr, J. F., Fortin, L. G., Chaumont, D., Biner, S., Chartier, I., Caya, D., and Mauser, W.: The role of hydrological model complexity and uncertainty in climate change impact assessment, Advances in Geosciences, 21, 63–71, https://doi.org/10.5194/adgeo-21-63-2009, 2009.

Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, Advances in Water Resources, 26, 205–216, https://doi.org/10.1016/S0309-1708(02)00092-1, 2003.

Mann, H.: Non parametric tests against trend, Econometrica, 13, 245–259, https://doi.org/10.2307/1907187, 1945.

Michel, C.: Michel, C. (1991), Hydrologie appliquée aux petits bassins ruraux, Hydrology handbook (in French), Cemagref, Antony, France., Antony, 1991.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Stationarity Is Dead: Whither Water Management?, Science, 319, 573–574, https://doi.org/10.1126/science.1151915, http://www.sciencemag.org/content/319/5863/573.short, 2008.

Mouelhi, S.: Vers une chaîne cohérente de modèles pluie-débit conceptuels globaux aux pas de temps pluriannuel, annuel, mensuel et journalier, Ph.D. thesis, Université Paris VI, Ecole des Mines de Paris, http://www.cemagref.fr/webgr/Download/Rapports{_}et{_}theses/2003-MOUELHI-THESE.pdf, 2003.

Naghettini, M.: Fundamentals of Statistical Hydrology, Springer, Cham, https://doi.org/10.1007/978-3-319-43561-9, http://link.springer.com/10.1007/978-3-319-43561-9, 2017.

Nalley, D., Adamowski, J., Biswas, A., Gharabaghi, B., and Hu, W.: A multiscale and multivariate analysis of precipitation and streamflow variability in relation to ENSO, NAO and PDO, Journal of Hydrology, 574, 288–307, https://doi.org/10.1016/j.jhydrol.2019.04.024, 2019.

Nash, J. and Sutcliffe, J.: Nash and Sutcliffe - 1970 - River flow forecasting though conceptual models Part 1 - A discussion of principles, Journal of Hydrology, 10, 282–290, 1970.

Paturel, J. E., Servat, E., and Vassiliadis, A.: Sensitivity of conceptual rainfall-runoff algorithms to errors in input data - case of the GR2M model, Journal of Hydrology, 168, 111–125, https://doi.org/10.1016/0022-1694(94)02654-T, 1995.

Payrastre, O.: Utilité de l'information historique pour l'étude du risque de crues, in: 14èmes Journées Scientifiques de l'Environnement : l'Eau, la Ville, la Vie,, edited by Thévenot, D. R., Créteil (France), https://hal.archives-ouvertes.fr/hal-00203088, 2003.

Peel, M. C. and Blöschl, G.: Hydrological modelling in a changing world, Progress in Physical Geography, 35, 249–261, https://doi.org/10.1177/0309133311402550, 2011.

450 Pettitt, A.: A non-parametric approach to the change-point problem, Appl. Statist., 28, 126–135, https://doi.org/10.2307/2346729, 1979.

Poff, N. L., Brown, C. M., Grantham, T. E., Matthews, J. H., Palmer, M. A., Spence, C. M., Wilby, R. L., Haasnoot, M., Mendoza, G. F., Dominique, K. C., and Baeza, A.: Sustainable water management under future uncertainty with eco-engineering decision scaling, Nature Climate Change, 6, 25–34, https://doi.org/10.1038/nclimate2765, 2016.

Prudhomme, C., Wilby, R. L., Crooks, S., Kay, A. L., and Reynard, N. S.: Scenario-neutral approach to climate change impact studies:
455 Application to flood risk, Journal of Hydrology, 390, 198–209, https://doi.org/10.1016/j.jhydrol.2010.06.043, http://dx.doi.org/10.1016/j.jhydrol.2010.06.043, 2010.

Rabiner, L. R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE, 77, 257–286, https://doi.org/10.1109/5.18626, 1989.

Razavi, S. and Tolson, B. A.: An efficient framework for hydrologic model calibration on long data periods, Water Resources Research, 49,
460 8418–8431, https://doi.org/10.1002/2012WR013442, 2013.

Singh, S. K. and Bárdossy, A.: Calibration of hydrological models on hydrologically unusual events, Advances in Water Resources, 38, 81–91, https://doi.org/10.1016/j.advwatres.2011.12.006, http://dx.doi.org/10.1016/j.advwatres.2011.12.006, 2012.

Stephens, C. M., Marshall, L. A., and Johnson, F. M.: Investigating strategies to improve hydrologic model performance in a changing climate, Journal of Hydrology, 579, 124 219, https://doi.org/10.1016/j.jhydrol.2019.124219, https://doi.org/10.1016/j.jhydrol.2019.124219, 2019.

465 Thirel, G., Andréassian, V., and Perrin, C.: De la nécessité de tester les modèles hydrologiques sous des conditions changeantes, Hydrological Sciences Journal, 60, 1165–1173, https://doi.org/10.1080/02626667.2015.1050027, http://dx.doi.org/10.1080/02626667.2015.1050027, 2015a.

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., and Vaze, J.: Hydrologie sous changement: un protocole d'évaluation pour
470 examiner comment les modèles hydrologiques s'accommodent des bassins changeants, Hydrological Sciences Journal, 60, 1184–1199, https://doi.org/10.1080/02626667.2014.967248, http://dx.doi.org/10.1080/02626667.2014.967248, 2015b.

Tilmant, A., Pina, J., Salman, M., Casarotto, C., Ledbi, F., and Pek, E.: Probabilistic trade-off assessment between competing and vulnerable water users - The case of the Senegal River basin, Journal of Hydrology, 587, https://doi.org/10.1016/j.jhydrol.2020.124915, https://www.sciencedirect.com/science/article/pii/S0022169420303759?casa{_}token=
475 U5HbM86r8aEAAAAA:mRMK2xZGvlDdQxuN2McRyndrxkIjNzodfOJiuMPMgNEKgDG9flHbcu5-P6ruap5VbOdVC3xGSv0, 2020.

Turner, S. and Galelli, S.: Regime-shifting streamflow processes: Implications for water supply reservoir operations, Water Resources Research, 52, 3984–4002, https://doi.org/10.1002/ 2015WR017913, 2016.

Viterbi, A. J.: Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm, IEEE Transactions on Information Theory, 13, 260–269, https://doi.org/10.1109/TIT.1967.1054010, 1967.

480 Weaver, C. P., Lempert, R. J., Brown, C., Hall, J. A., Revell, D., and Sarewitz, D.: Improving the contribution of climate model information to decision making: The value and demands of robust decision frameworks, Wiley Interdisciplinary Reviews: Climate Change, 4, 39–60, https://doi.org/10.1002/wcc.202, 2013.

Welch, L. R.: Hidden Markov models and the Baum-Welch algorithm, IEEE Information Theory Society Newsletter, 53, 9–13, 2003.

Whiting, J., Lambert, M., Metcalfe, A., and Kuczera, G.: Development of non-homogeneous and hierarchical Hidden Markov models for

485   modelling monthly rainfall and streamflow time series, Proceedings of the 2004 World Water and Environmetal Resources Congress: Critical Transitions in Water and Environmetal Resources Management, pp. 1588–1597, https://doi.org/10.1061/40737(2004)212, 2004.

Zucchini, W., MacDonald, I. L., and Langrock, R.: Hidden Markov Models for Time Series: An Introduction Using R, vol. 80, https://doi.org/10.18637/jss.v080.b01, 2017.