Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-596-AC4, 2021

Hydrology and Earth System Sciences
Discussions

Interactive comment on "A Robust calibration/validation protocol of a hydrological model using hidden Markov states" *by* Etienne Guilpart et al.

Etienne Guilpart et al.

etienne.guilpart@gmail.com

Received and published: 17 March 2021

We highly appreciate your time and effort in conducting a thorough review of our manuscript. Please find below our responses (your comments are in bold, our responses in normal font, and proposed changes to the manuscript in italic).

The study applies the differential split-sample test to evaluate the robustness of hydrological model parameters using the HMM to identify sub-sequences of different hydroclimatic conditions. The approach is tested over three sub-basins

of the Senegal River employing the GR2M model; the 2-states and 3-states HMM classifications are compared with the non parametric Pettitt's test which allows to detect a single change point. Authors state that results show that HMM can be a viable classification option for long time series exhibiting multiple change points.

There are several points that deserve further insights and that are not sufficiently explored in the study.

(1) The title of the paper hints at a protocol of calibration/validation of hydrological models that in my opinion is still substantially based on the differential split sample test proposed by Klemes (1986). The study rather focuses on the comparison of different approaches to identify sub-periods characterized by different climatic conditions. The 3steps protocol outlined at sec. 4 is not novel and should at least include calibration and validation steps. I suggest rephrasing the title.

We propose the following title :

Combining split-sample testing and Hidden Markov Modeling to assess the robustness of hydrological models.

(2) I agree with remark in RC1 about the necessity of a more detailed analysis to support the climate classification sequence identified by the HMM in terms of changes in meteorological data or land use.

We agree that river discharges are the result of hydrological processes taking place upstream and are influenced by changes in precipitation, land use, etc. Anthropogenic changes may indeed alter the flow regime and hence influence an HMM-derived classification which would no longer rely solely on natural factors. However, that anthropogenic influence can be removed so that the analysis is done on naturalized flows, something fairly standard in time series analysis but also in process-based hydrological modelling.

The overall goal of this paper is to highlight the relevance of combining HMMs classifications with the differential split-sample test to assess the robustness of a hydrological model to be used in climate change studies, i.e. to generate hydrologic projections from contrasted climate ones. So, the emphasis is on changes in physical processes rather than human-induced ones. The Senegal River basin is only used as a case study to illustrate the proposed method to test the robustness of a hydrological model. For that river basin, most of the runoff and headwaters of two of the three sub-basins (Daka Saidou, Oualia) are located in the Fouta Djallon, a sparsely populated plateau where vegetation cover is relatively stable (Descroix et al, 2020), anthropogenic impacts on runoff seem to be negligible (Faty, 2017) or not even mentioned in the updated Senegal River monography (Bader et al, 2014) and in the River Basin Master Plan (OMVS, 2011). The areas mainly concerned with massive land-use conversions are located downstream of Bakel, a region not considered in our analysis. For the third sub-basin, river discharges at the outlet were naturalized by Bader et al. (2014) after removing the influence of the Manantali dam on the flow regime.

In this study, the term "robustness" refers to the ability of the hydrological model to perform well under contrasted hydro-climatic conditions. This definition is coherent with the so-called robust decision-making framework that is often advocated to handle the deep uncertainty attached to climate change (Lempert et al., 2006). In other

words, we do not seek the "best" model (with the best fit) but a model that performs reasonably well under different conditions.

Also, we propose replacing the second to last paragraph in the introduction with:

In this article, we combine a classification obtained by an HMM with the differential split-sample testing framework. The goal is to improve the robustness of the calibration/validation of a hydrological model, which is a prerequisite to climate change impact assessment. The term “robustness” refers to the ability of the hydrological model to perform well under contrasted hydro-climatic conditions. This definition is coherent with the so-called robust decision-making framework that is often advocated to handle the deep uncertainty attached to climate change (Lempert et al., 2006). This is illustrated using the Senegal River Basin (SRB) as a case study. Headwaters in the SRB are still largely natural areas (Descroix et al, 2020; Faty, 2017) and the flow regime in the upper part of the basin exhibits regime-shifting behavior with departures from the inter-annual average over extended periods of time (Faye et al. (2015); Paturel et al. (2004); Dacosta et al. (2002). These characteristics makes the SRB an interesting case study to illustrate the split-sample testing framework with hydrologic sequences identified from an HMM.

We propose replacing the abstract with:

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. In this paper, *numerous, contrasted, climate sequences identified by a Hidden Markov Model (HMM) are used in a differential split-sample testing framework to assess the robustness of a hydrologic model. The split-sample test based on an HMM classification is implemented on the time series of monthly river discharges in the upper Senegal River Basin in West Africa, a region characterized by the presence of low-frequency climate signals. A comparison with the results obtained using classical rupture tests shows that the diversity of hydrologic sequences identified using the HMM can help assessing the robustness of the hydrologic model.*

Section 5 only focuses on the 1940-1998 period, which displays clear subsequences for each sub-basin. To further illustrate the relevance of the HMM classification, we investigated two shorter 26-years periods (1945-1971 and 1972-1998) which were put in the Supplementary Materials section. We first decided to put those results in that section because we thought that they would somehow distract the reader from the main objective of the paper. Since both reviewers find such an analysis relevant, we will bring it back to the main text. This will give us the opportunity to discuss in more detail the importance of the length of the period.

(3) Although HMM allows for a finer labelling of each year, NSE and KGE coefficients in Table 3 do not show higher performance in fitting observed flows. Classification based on Pettitt's test provides comparable or better results, particularly for the Daka Saidou and Oualia basins. An interesting insight could be given on which is the most convenient case to apply to reach better performance (Case 1 vs Case 2; Case 3 vs Case 4;: : :.).

We would like to insist on the fact that we do not necessarily want to achieve the highest calibration/validation values. Rather, the goal is to assess the robustness, i.e. the ability of the hydrological model to perform well under contrasted hydro-climatic

conditions. Checking that a calibration on a dry state is in general better than a calibration on a wet state (or opposite) can indeed be investigated with the proposed approach but it is not the main objective of the paper. The table below nevertheless presents that comparison. It shows NSE/KGE values for calibration and validation for the seven cases.

		Pettitt Test		2-states HMM		3-states HMM			
Phase	Sub-sequence	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Catchment
Calibration	Dry	-	0.935/0.967	0.923/0.961	-	0.927/0.963	-	-	Daka Saidou
	Normal	-	-	-	-	-	0.925/0.962	-	
	Wet	0936/0.967	-	-	0.942/0.971	-	-	0.942/0.971	
Validation	Dry	0.932/0.940	-	-	0.921/0.943	-	0.913/0.887	0.919/0.937	
	Normal	-	-	-	-	0.913/0.892	-	0.921/0.911	
	Wet	-	0.933/0.953	0.939/0.948	-	0.934/0.952	0.938/0.920	-	
Calibration	Dry	-	0.803/0.775	0.775/0.864	-	0.738/0.819	-	-	Oualia
	Normal	-	-	-	-	-	0.772/0.868	-	
	Wet	0.886/0.941	-	-	0.888/0.937	-	-	0.888/0.937	
Validation	Dry	0.727/0.803	-	-	0.709/0.796	-	0.717/0.786	0.579/0.753	
	Normal	-	-	-	-	0.762/0.816	-	0.714/0.795	
	Wet	-	0.836/0.845	0.832/0.865	-	0.804/0.814	0.837/0.853	-	
Calibration	Dry	-	0.858/0.919	0.776/0.883	-	0.805/0.899	-	-	Bakel
	Normal	-	-	-	-	-	0.827/0.909	-	
	Wet	0.900/09.41	-	-	0.899/0.949	-	-	0.913/0.943	
Validation	Dry	0.693/0.533	-	-	0.715/0.678	-	0.766/0.745	0.633/0.478	
	Normal	-	-	-	-	0.782/0.734	-	0.755/0.692	
	Wet	-	0.644/0.470	0.804/0.679	-	0.622/0.436	0.810/0.684	-	

We can see that no case systematically stands out from the others. When calibrated over a dry sequence, validation over the wet sequence generally displays a lower criterion value. The opposite is also true. Consequently, we are not able to state that a calibration on a specific subsequence is more desirable than another one.

We will replicate this discussion for sections dealing with shorter periods (now in Supplementary materials).

(4) The case study and the results do not allow a generalization of the approach

and probably it is not the ideal test case to show the advantages of the proposed method mostly because there is a marked rupture in streamflow observations.

This paper shows how an HMM classification can be used within the differential split-sample testing framework to assess the robustness of hydrological models under shifting flow regimes. The Senegal River, for which the presence of low-frequency climate signals has been discussed in the literature (Ardoin-Bardin, 2004; Bodian et al, 2014, Descroix et al, 2015), is used as a case study. The results are therefore specific to the case study and cannot be generalized. However, the idea of using HMM to identify contrasted subsequences of streamflows is relevant for other rivers exhibiting regime shifting behavior. Since an HMM classification has the potential to identify more than two subsequences, the method automatically offers more flexibility to assess the robustness of a hydrological model. See our responses to your second comment for proposed changes in the manuscript.

(5) The content of the final paragraph at sect. 5.1 is not clear and differences (if any) with the traditional differential split sample test should be better outlined. Moreover, conclusions not always refer to by presented analysis and results showed in the study and do not seem to support statement in the abstract.

Both the split-sample and the differential split-sample tests belong to the calibration/validation approach proposed by Klemes (1986). Although both tests require splitting the time series, the split-sample test is applied to stationary time series whereas the differential split sample is test is used with non-stationary ones. With the split sample test, splitting the time series does not require specific modeling techniques and is mostly left to the modeler's judgment. With the differential split sample test, however, we need a technique to detect change-points in hydrological series. In this paper, change-points are detected after carrying out a HMM-based classification

of the data, which allows for the identification of multiple subsequences instead of only two with the traditional Pettit test. We have reformulated the title and rewritten the abstract to better explain the scope of the paper, which is about combining an HMM classification of hydrological series with the differential split sample test to help assessing the robustness of hydrological models.

We propose the following changes in section 5:

This article proposes an HMM-based classification to deal with complex climate sequences and shows how the resulting classification can be used in a differential split sample test to assess the robustness of a hydrological model. A modeling experiment is carried out in the Senegal River basin using the GR2M model and historical flows from 1940-1998.

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, an HMM classification can divide the series into several climate sub-sequences without the need for additional data. However, records must be long enough (typically 20-25 years for a 2-states HMM classification, and 30-35 years for a 3-states HMM classification).

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.

Compared to the Pettit test, however, the HMM classification allows for a finer labelling of the years, therefore better exploiting the intrinsic variability in the series to enrich a differential split sample test.

There are some other minor comments (see below) that I recommend to consider. Table 1: Check isohyets ranging for the Bakel sub-basin. Table 2: can be improved by associating parameters to dry, wet and nor period. Figure 5b: I suggest to use the same axis limits for both validation and calibration performances. Eq. (3): Check equation symbol (*) Eq. (4, 5, 6): several symbols are not introduced or explained in text. Line 205: should be T3HMMnor .

Thanks for you careful reading. We will consider it in our ongoing proofreading step.

Recent examples of differential split-sample validation tests that can be included as references have been reported in:

- **D.F. Motavitaab, R. Chowab, A. Guthkea, W. Nowaka. (2019). The comprehensive differential split-sample test: A stress-test for hydrological model robustness under climate variability, Journal of Hydrology, Volume 573, Pages 501-515**
- **H. Dakhlaouiab, D. Ruellandc, Y.Tramblayd. (2019) A bootstrap-based differential splitsample test to assess the transferability of conceptual rainfall-runoff models under past and future climate variability, Journal of Hydrology, Volume 575, Pages 470-486**

Thanks for your contribution, and we will add them to the Introduction.

References:

Ardoin-Bardin, S. (2004). Variabilité hydroclimatique et impacts sur les ressources en eau de grands bassins hydrographiques en zone soudano-sahélienne. PhD thesis, Université Montpellier II.

Bader, J.-C., Cauchy, S., Duffar, L., and Saura, P. (2014). Monographie hydrologique du Fleuve Sénégal. De l'origine des mesures jusqu'en 2011. IRD, Marseille (France), ird edition.

Bodian, A. (2014). Caractérisation de la variabilité temporelle récente des précipitations annuelles au Sénégal (Afrique de l'Ouest). Physio-Géo, 8(Volume 8) :297-312.

Dacosta, H., Kandia, K. Y., and Malou, R. (2002). La variabilité spatio-temporelle des précipitations au Sénégal depuis un siècle. Regional Hydrology : Bridging Ihe Gap between Research and Practice (Proceedings, (2) :499-506.

Descroix, L., Faty, B., Manga, S. P., Diedhiou, A. B., Lambert, L. A., Soumaré, S., Andrieu, J., Ogilvie, A., Fall, A., Mahé, G., Diallo, F. B. S., Diallo, A., Diallo, K., Albergel, J., Tanimoun, B. A., Amadou, I., Bader, J. C., Barry, A., Bodian, A., Boulvert, Y., Braquet, N., Couture, J. L., Dacosta, H., Dejacquelot, G., Diakité, M., Diallo, K., Gallese, E., Ferry, L., Konaté, L., Nnomo, B. N., Olivry, J. C., Orange, D., Sakho, Y., Sambou, S., and Vandervaere, J. P. (2020). Are the fouta djallon highlands still the water tower of west africa ?, volume 12.

Descroix, L., Diongue Niang, A., Panthou, G., Bodian, A., Sane, Y., Dacosta, H., Malam Abdou, M., Vandervaere, J.-P., and Quantin, G. (2015). Évolution récente de la pluviométrie en Afrique de l'ouest à travers deux régions : la Sénégambie et le Bassin du Niger Moyen. Climatologie, 12(Volume 12) :25-43.

Faye, C., Sow, A. A., and Ndong, J. B. (2015). Étude des sècheresses pluviométriques et hydrologiques en Afrique tropicale : caractérisation et cartographie de la sècheresse par indices dans le haut bassin du euve Sénégal.

Faty, A. (2017). Modélisation hydrologique du haut bassin versant du Fleuve Sénégal dans un contexte de variabilité hydro-climatique : Apport de la télédétection et du modèle Mike SHE. PhD thesis, Université de Cheikh Anta Diop de Dakar.

Klemes, V. (1986). Operational testing of hydrological simulation models. Hydrological Sciences Journal, 31(1) :13-24.

Lempert, R. J., Groves, D. G., Popper, S. W., and Bankes, S. C. (2006). A general, analytic method for generating robust strategies and narrative scenarios. Management Science, 52(4) :514-528.

OMVS (2011). SDAGE - Schéma directeur. Rapport téchnique.

Paturel, J.-E., Ibrehim, B., and L'Aour, A. (2004). Evolution de la pluviométrie annuelle en Afrique de l'Ouest et centrale au XXeme siècle. Sud Sciences et technologies, 13 :40.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-596, 2020.