This study applies the time varying parameter method previously developed by the authors to a Vietnamese catchment and two lumped daily hydrological models. The authors test the suitability of their method to reflect observed land use changes within the catchment as well as the compatibility of the method with different model structures. The manuscript is well written, the results very interesting and I appreciate the author's efforts to present their method in a very clear and concise manner.
That said, I consider the manuscript can still be improved on several aspects.

Please note that all line numbers refer to the submitted document without changes tracked.

Key comments

1) The reader could benefit from more precise explanations on the following points.
The fact that the method is applied to two lumped, conceptual, daily models needs to be stated from the beginning (abstract and introduction) of the article. These are specific methodological choices and could impact the conclusions.
The scope of the paper needs to be more clearly stated by underlining what research gap this study fills (i.e. how your specific contribution will advance understanding) and the novelty of the approach (i.e. what can the time variable parameters method do that existing methods can't when studying the impacts of land use changes).
The perspectives of the study could be better articulated with the paper's scope and better motivated given the outputs of the study. More specifically, the authors propose to apply the time varying parameter method (TVPM) to physically-based models. However, the lines 294-296 state that parameter dimensionality can be an issue and, as acknowledged by the authors, physically-based models are usually less parsimonious than conceptual models. Likewise, the other perspective is to applied the TVPM within a multi-model framework. According to the findings of the analysis, model structure is a key factor in assuring the success of the time varying parameter method: wouldn't it be the same problem to find a single model compatible with the TVPM than to find a compatible multi-model?

2) The temporal scales in the introduction and throughout the manuscript need to be defined more consistently.
Please quantify : L53: "short-term" (one time step ahead/days/week/month?), L54 "dynamic" (daily dynamic/weekly…?), L63 and 71: "real time", L72: "given time", L87: "gradual", L288: "longer time horizons".
The pre-change conditions are different between L206-207 (1973-1979) and Table 1 (1970-1994).
The observed results (Figure 2) are presented between 1970 and 2004 when the modeling results (Figure 3) are presented for the 1975-2004 period.
Likewise why calibrate the models between 1973 and 1979 and not between 1970 and 1994?
It is quite difficult in the present manuscript to gather the different time resolutions.

3) Section 2.2 mixes methods with results.
I would suggest to keep the methodological parts (computation of the base flow index, description of the MASH method and the Mann-Kendall test) as section 2.2 and move the result parts (analysis of

figure 2) as a new section 3.1. It would also be easier for the reader to recall the outputs of the observed changes analysis while moving to the analysis of the time varying parameter method (L367: “as discussed in section 2.2”).
Regarding the computation of the BFI please consider adding the equation as well as the chosen values for the two parameters to the text as it can impact the BFI values.

4) The benchmark used in this study appears quite weak for two reasons.
First, the study is retrospective which means both the benchmark and the TVPM should be based on the whole streamflow record. Secondly, the authors mentioned the use of split sample calibration for retrospective studies in the introduction (lines 47-49), why not choose a benchmark based on split sample calibration? The use of such a benchmark could better highlight the benefits of the TVPM over existing methodologies. In particular, it could supplement the discussion the authors provided on the benefits of updating both parameters and states over updating solely the model parameters.
If changing the benchmark is not feasible, the results analysis and discussion should at least acknowledge that better-performing benchmarks already exist and nuance the relative assessment of the efficacy of the TVPM accordingly.

5) I believe the paper could benefit from a more detailed discussion on two aspects.

Could you please expand the explanation of the observed increase of BFI with regard to the physical processes involved. Indeed as stated by the authors, forest coverage decrease for the benefit of cropland. If this is the case, I would expect an observed decrease of BFI since forests usually favor infiltration while cropland are usually characterized by more compact soils and managed to maximize the use of soil water by crops. Are these newly agricultural soils drained or irrigated? It could result respectively in increased soil infiltration and increased available water without changes in the precipitation signal.

Provide some more context to evaluate the results on the model structure impact.
On Figure 3 please ensure that all parameters and states are represented, at least those involved in the TVPM. For example the b parameter (HyMOD) is primarily impacted by the TVPM but not presented in the model scheme so that the reader cannot understand how it is used by the model. Be more specific in the legend of Figure 3: for example, on Fig 3b there is a q_b in the legend but none in the scheme, it is also unclear whether sowat, stw1, Sq1… are the store names or the store content (i.e. the state variable to be updated)? On Figure 5, there is a k_b parameter which is not displayed on Figure 3. If possible, please display parameters using one color and states using another color to help the reader understand model structure quickly.
For the HBV model, perc and β are the two most heavily impacted by TVPM but are also the two most sensitive. I do not find surprising that TVPM would preferably adjust sensitive parameters but a discussion of the relation between model sensitivity and effects of TVPM is missing. To this aim, it would also be very interesting to have the results of the sensitivity analysis for the HyMOD model.
Which lead to the following point.
Can the authors elaborate on lines 399-401: “The annual runoff and annual direct runoff are severely under-estimated in the post-change period by the TVP-HyMOD, whilst the Annual Baseflow Index has an increasing trend of magnitude far greater than observed (Figure 7c).”? As stated by the authors (l191-192), the three cascading tanks represent quick flows while slow flow is represented by the S_s store. In Figure 5 the mean alpha parameter is inferior to 0,5 in the post-change period, meaning more flow is routed through the slow flow store, hence the increase of BFI in Figure 7c. My understanding of

these results is that it is easier for the model to adjust its response (simulated streamflow) by modifying the S_s store behavior than to adjust the quick flow response. This could be due to: (i) a high model sensitivity towards ks (especially when alpha is low and b high) and/or (ii) incompatibility between cascading tanks (need of multiple time steps to have an impact on streamflow) and data assimilation frameworks (Markov chain). If this is indeed the case, I would argue that based on their results, the authors should make some concrete recommendations on which type of model structure is compatible with TVPM (parallel tanks, high sensitivity for all parameters, low parameter cross correlation...)

With the above key review comments I have the following minor comments.

Minor comments

Line 64-67: “It can also...an assessment.” The link with the above paragraph is not obvious at this point of the introduction
Line 72: “given time”, do you mean in forecasting mode?
Lines 74-76: please rephrase “the time scale of the observation frequency”
Lines 75-77: Regarding the applications of the method for 1): please clarify the advantages of the approach compared to existing split sample calibration procedures you mentioned (l 48-49), 2) and 3): seam out of the paper scope since the method/results do not include a part on forecasts. Please justify more clearly the use of the method for forecasting. Regarding 3) is on-line water resource water management on the same time scale as the time varying parameter method?
Line 103: Is the efficiency of the method dependent on catchment size? Please specify in the text.
Line 109: Please specify to which dates you are referring
Line 134: Could you explain the reason behind using two different data sets to assess land use? Are the two datasets equally reliable? Please specify in the text
Line 143: Can you describe the variation of altitude within the catchment, as it can help understand the uncertainties associated with the meteorological forcing.
Line 158: Please insert the BFI equation and specify the chosen values for the two parameters
Line 182: Please specify that the daily time step is used
Lines 205-206: Did you use both algorithms on each model or the SCE was used to calibrate HBV and BEA for HyMOD (or reversed)? If a different algorithm was used to calibrate the models, please include the importance of the calibration procedure in the discussion of your results
Line 210: Can you explain why these streamflow threshold values were retained?
Line 247 (eq1): Please name $\mathbf{m}_t$
Line 254: Is $\mathbf{m}_{max}$ the same as the “allowable rate of change” in tables 4 and 5? If yes please unify the notations. Could you also specify how $\mathbf{m}_{max}$ is set (experience with the model, external data…)?
Line 295: Is a large number of parameters a limit to the application of the method? If yes, please acknowledge it in the text
Line 296: Could you briefly explain the Sobol method?
Line 361: Please refer to Figure 4
Lines 375-376: Is the problem the difference between dry and wet seasons or catchment size and heterogeneity? Please clarify.
Lines 379-380: “increased difficulty in accurately modeling the hydrologic response (even in pre-change conditions)”: does this mean bad calibration for both models? Please clarify
Line 412: Can the extreme updated values be prevented with smaller allowable change values?
Line 451: “the time varying parameter method”

Line 463: “(i.e. model equation)” could maybe be moved to the beginning of the article to help the reader
Line 464: Is HyMOD unsuited or the association of the time varying method with the HyMOD structure proves inefficient?
Line 466: “unknown future”: please rephrase since (i) data assimilation cannot be performed without streamflow measurements (“unknown”) and (ii) the “future” has not been explored in this study
References: The formatting of the doi appears different between the citations.
Line 644: Table 1: please add the mean observed BFI values in the Hydro-Meteorological Properties since it is a key variable in your study