<u>Response to reviews from round 2</u>

Editor Decision: Publish subject to minor revisions (Editor review) (22 Nov 2014) by Prof. Nunzio **Romano**

Comments to the Author:

Dear Authors, The referees and I appreciated your efforts made in the revision of your work. However, as you can see, Ref.#2 required some additional changes/revisions. I have also evaluated this referee's request and I think it is worth in order to further improve your paper. Therefore, I am sending back the paper for additional minor revisions. On a separate sheet, please highlight the changes/revisions and whether or not you agree with the additional changes required or suggested.

We thank the editor for his appreciation. The concerns of ref.#2 are addressed below.

Ref.#1

For final publication, the manuscript should be
accepted as is

The authors spend considerable efforts to reply on the criticism of the referees at the first round of the review. From my side they succeeded in addressing their concerns and the manuscript is ready to be published in HESS.

We thank Ref.#1 for acknowledging our efforts to address the criticism of referees during the discussion stage. This exercise was a very positive experience, and greatly helped in improving the paper focus and readability.

Rev#2
Reconsider after major revisions
I would like to review the revised paper

We thank Ref.#2 for his thorough assessment of the revised version of this paper, which helped in improving it by pointing out the remaining errors and lack of clarity.

There are perhaps four priorities for revision, in no particular order.
1) Justification of serial vs parallel reservoir structure in the model, and discussion of the implications throughout the manuscript.

The information and discussion about the choice of model structures can be found in Hrachowitz et al. (2014). We feel it is out of the scope of the current paper. In fact, in this paper, we are not trying to find the best model for the study catchment, but to discuss the relevance of using hysteretic index to select models. This was probably unclear in the initial version of the paper, but we feel that after addressing the first round of reviews, it has now become quite clear in the current version.

2) Perhaps adding data to determine if there is hysteresis in the storage in different zones (i.e., storage in the riparian vs hillslope zones). This really will show if the model is capturing the correct internal mechanisms.

This is a good idea. However, we do not think we should add this to the present paper, which is already rather long. Our focus was "methodological", and not some much about the actual mechanisms of the study catchment. As mentioned in the discussion, we are currently working on designing better hysteretic indexes, and objective functions, that would be used in the calibration procedure itself, and not only for a posteriori assessment. We added this interesting perspective in the discussion (p 29 line 6):

"Among the possible combination of variables, objective functions based on the relative dynamics of storage in different spatial locations, such as riparian versus hillslope, might provide new insights about the catchment internal processes. We suggest that such..."

3) Ensure consistency with the literature in how storage-streamflow is plotted and on which axis.

We agree with the referee that adopting a standard for discharge storage relationships would be useful. In this paper, we decided to plot stream discharge in the X-axis, mainly for 2 reasons : a) because, of course, this is what is done routinely when looking at hysteretic relationships between stream discharge and stream water composition (solutes, suspended sediments…) and our proxy of HI was inspired from such studies, and 2) because the few studies presenting the discharge-storage studies that we refer to in our discussion (e.g. Ali et al., 2011; Gabrielli et al., 2012) are also plotting stream discharge in the X-axis.

4) Always provide data to substantiate statements.

We think that this concern was mostly raised because of confusion between the results and statements presented in the present paper and results and statements coming from the previous paper of Hrachowitz et al. (2014). Most of this confusion was addressed during the previous round of reviews and we hope the current modified version is absolutely clear in this now.

In this paper the authors analyze annual storage – discharge hysteresis loops observed in a French headwater catchment. They do so to determine if one of four hydrological models can reproduce this behaviour. Testing the ability of a model to do so, the authors argue, is indicative of how well the model may represent the internal system dynamics of the watershed.

The authors found that a more thorough representation of the variety of catchment storages (i.e., upper hillslope, lower hillslope and riparian) best simulated the storage – discharge hysteresis loops. I might argue that many model structures will reproduce storage – discharge hysteresis loops. What would be interesting to know is how the model can reproduce the state variables. Presumably there is a curve or loop observed when comparing hillslope and riparian storages. This would be very interesting to see, and it and some discussion around it would make a nice addition to the paper.

We fully agree with the referee, looking at different kind of hysteresis, including between state variables, would be very interesting. As we replied above (comment 2) this is the focus of currently ongoing work.

The only other major comment I have is in regards to the orientation of the storage-discharge graphs. Storage – discharge hysteresis curves are being presented quite regularly in recent literature. However, there seems to be no convention on which term is plotted on the ordinate or y-axis. For instance Gabrielli et al (2012) plot water table fluctuations (i.e., storage) on the y-axis, but Ewen and Birkinshaw (2007) plot storage on the x-axis. Presumably it should be streamflow on the y-axis, as the dependent term. The authors plot streamflow on the abscissa or x-axis. This is a crucial point, as it influences the direction of the hysteresis. I suggest the authors double check to

make sure they are interpreting comparable curves. Agreed upon standards for plotting storage-discharge curves would be beneficial to our community as they become more common.

See answer above (major comment 3)

Some specific comments:
Page 2 Line 23 – 29: There are a lot of different ideas in this sentence; perhaps try rewriting it.
We respectfully disagree with the referee. To our view, the sentence (p2 Line 24-28, is it the same?) is conveying one single idea, i.e. tradeoff between loss of information in lumped models vs lack of relevant information for (semi) distributed models:
"While spatial aggregation of storage estimates (e.g. catchment averages) in lumped models may lead to a loss of crucial information and thus to overly-simplistic representations of reality, allowing for explicit incorporation of spatial storage heterogeneity in (semi) distributed models may prove elusive in the presence of data error and the frequent absence of detailed spatial knowledge of the properties of the flow domain".

Page 4 Line 17 – 18: It would be good if the authors were to add a couple of sentences that tie together the thoughts in these two paragraphs. It might improve the structure of the introduction.

We propose to add, before the last paragraph, the following sentence:
"Hysteretic patterns between hydrological variables are potentially good candidates to build such tools."

Page 5 Line 13: There is a reference to Figure 2 and streamflow, but Figure 2 does not show streamflow data.

This is right. We corrected as: (Fig.2 and 3a)

Page 5 Line 27: Perhaps the authors need to be explicit in what the final selection of the study and model period was at this point in the paper.

We agree, we added the following sentence:
"In summary, stream discharge, water table levels and soil moisture were considered for the years 2002-2012, 2002-2012, and 2010-2012 respectively."

Page 6 Line 7: Because of the methodology that includes normalizing storage over the study period, it is important to have some statement about how representative these locations and this time period are. The authors assume that the sample period covers the entire range of absolute storage. That would mean the measurements span record highs and lows. A figure would best illustrate this. If the study period did not span this range, then some statement about the broader applicability of the results needs to be made.

We feel we have extensively discussed the difficulty of evaluating storage at the watershed scale in the introduction (p2 Line 22 to Page 3 Line 16) and the discussion (p19 Lines 9-18). We agree that the range of observed values will affect the normalization. However, we have tested that the normalization is not affecting the absolute value of HI (p8 Lines 11-15).

Page 6 Line 24: I'm sorry, I don't follow, to which years are you referring when you say "almost all the years".

What was meant is that whatever the hydro-climatological conditions in a given year, the observed hysteretic patterns are not likely to change. In our case, only one of the year 2011-2012 would have given different. We modified the sentence as follow:

"For storage-discharge hysteresis at the annual scale, this approach is not sufficient as the same type of hysteretic loop is likely to happen for almost all the years, when a strong seasonality exists and its pattern is repeated across years. This is the case in your study, where seasonality of groundwater level and discharge was showing a strong unimodal pattern for all years, except 2011-2012 which was bimodal (Figure 2 and 3a)."

Page 6 Line 32: The authors are clearly discussing streamflow – chemistry concentration hysteresis loops, and should explicitly say so. The processes driving these types of curves are very different than the ones controlling storage – streamflow curves.

We do not quite understand the referee's concern. We feel it is quite clear that we are discussing how to quantify hysteretic loops, which would be relevant whatever process at stake.

Page 7 Line 25: The sentence that begins "Water storage dynamics ….." perhaps should come earlier in the section.

We agree that the sentence was confusing. It was rephrased as
"The difference in water storage in the unsaturated and …"

Page 8 Line 9: This is not a paragraph as it does not convey a single complete idea, but could easily be attached to the previous paragraph.
We agree and attached it to the previous paragraph

Page 8 Line 30: Just curious, why add a threshold that controls movement into a deeper groundwater store.

The long-term water balance of the catchment exhibits substantial deficits, indicating the importance of deep infiltration "losses", which is corroborated by the observation that groundwater levels continue to fall even during no-flow conditions. The incorporation of this thresholds permits to mimic precisely this behavior. More detailed information and discussion about the choice of model structures can be found in the cited paper by Hrachowitz et al. (2014).

Page 9 Line 4: Perhaps similarly, I was a bit surprised that the authors built a model that included only parallel reservoirs. Given the description of the system they wish to simulate (i.e., hillslope – riparian – stream) would a serial system not be more appropriate? The authors built their own virtual experiment, so a parallel system is fine, but I wonder if some content as to why a serial system was not selected might be helpful.

A range of serial set-ups was indeed tested for the Hrachowitz et al. (2014) paper, however, not included in that paper as these set-ups could not reproduce several aspects of the system response. We also have to admit that we do not see how a purely serial conceptualization (e.g. Nash cascade) would permit, for example, the conceptualization of the differences in unsaturated/root zone storage capacity between wetlands and hillslopes, which is essentially the non-linear core of these model types as it determines the different timing of when connectivity of the two individual landscape units is established. We agree, however, that it is likely that flows generated on the hillslope will eventually have to pass through wetlands in this first order stream (which may not necessarily be the case in larger catchments). Given the process understading for this catchment and its conceptualization in the model, the hillslope starts to be connected quite some time after connectivity is established in the wetland (due to its limited storage capacity). Thus, on the daily time scale of this study and given the fast response of the wetland, this will hydrologically not be relevant and is not warranted by the data in our search for a parsimonious model set-up. A connection between hillslope and wetland in a combined parallel-serial set-up will however be necessary once biogeochemical transport processes are incorporated in the model, which is investigated in detail in a paper that is currently in preparation. Further information and discussion about the choice of model structures can be found in Hrachowitz et al. (2014).

Page 9 Line 28: I do not see how the information in Table 4 indicates performance.

Table 4 is provided to help the understanding of the "DE" values given in Figure 3 b and has been added following the first round of review as explicitly requested by the reviewers.

Page 10 Line 3: Where are the data that show that the model failed to reproduce flow in wet periods?

Figure 3 a. depicts the observed and simulated flow for a calibration period and an evaluation period which is wetter than the calibration period as showed by the higher values of discharge. This explains partly the difference in global performance of model M1 between calibration and evaluation period (see e.g. greater mean Euclidian distance showed in the Figure 3 b). As said at the beginning of section 2.5 (p8 line 22 in modified version):

"This section aims at summarizing the results of this previous study Hrachowtiz et al. (2014) that are used as a basis for the present work."

In order to make it clearer we also added at the beginning of section 2.6 the following sentence:
"This section is also a summary of the findings of Hrachowitz et al. (2014) that served as a basis for this study, and not results of the current study. "

And we added a reference to the relevant figure at modified p9 L8:
"such as the evaluation period in Fig. 3a"

Page 10 Line 12: This seems to be a pervasive problem with the paper; not providing data or information in figures and tables to back up some of the statements that are being made. I can't glean performance details from Figure 3b.

It is correct that Figure 3 does not give any details of the performance but rather an overview of the results of a detailed study (Hrachowitz et al., 2014) which are used here as a basis, e.g. something published and known. According to the previous review round we all agreed that it was important to avoid redundancy with the previous paper which contains already all detailed illustrations about model performances, for different criteria, signatures and periods.

Page 12 Line 14: In stating that there are "fast responding flow pathways", the authors are interpreting what is going on. What evidence - even from the literature – do they have of this behaviour?

We agree that this is a fine line between observation and interpretation. We would however argue that the activation of fast responding flow paths can unambiguously be seen from the observed hydrograph. More specifically, at points in time when marked breaks in the recession behavior occur, i.e. changes in slope of the recession curve when regarding flow on a log-scale, it is well understood that changes in the flow regime occur and that connectivity of flow paths active on different time scales becomes established/lost. In other words, at moments in time when the slopes are steeper, faster processes "need" to be active.

Page 12 Line 28: If you put this last sentence near the beginning of the paragraph, the ideas would flow better.

We tried to take into account the referee proposition, but couldn't find a way to move that sentence.

Page 12 Line 31: "illustrated" instead of "confirmed" perhaps?

We agree that "confirmed" was unadequate. We replaced it by "reflected"

Page 13 Line 25: Similar to the earlier point about flow pathways, some evidence that shows the reader the flow was generated in the riparian zone would improve the authors' argument.

In principle we strongly agree and we would also be more than glad to have data showing this. Unfortunately, like so often in hydrology, such data are not available in the study catchment. However, again following the perceptual understanding of how the catchment functions, flow must be generated in the riparian zone before it is generated at the hillslope, due to the differences in available storage capacity. Also clearly no proof, the model results of the parallel set-ups lend quite some support to this hypothesis as they perform better than the lumped set-ups, and they therefore cannot be rejected.

Page 13 Line 32: Is this the authors' Figure 7 or Hrachowitz et al.'s?

We added the Figure 7 following the comments from the first round of review and this Figure is original. It is not of course in the manuscript of Hrachowitz et al. (2014) which is not explicitly dealing with hysteresis signatures. Note that almost all the authors of the present manuscript were of course authors of the previous manuscript as well.

Page 14 Line 4: The flow of the paper would be better if the content in this paragraph were integrated into the previous paragraph.

As we understood, the suggestion is to insert the content of p14 L6-10 into the previous paragraph. We feel that the reference cited in these lines relate to the main conclusion of the previous paragraph, (and are consistent with it, but were drawn using other approaches) and then should be cited after. Also, we couldn't find a way to integrate them in the previous paragraph (already long), without affecting the flow of the paper.

Page 14 Line 17: Maybe add "in 2011 – 2012" after "unsaturated zones ….."

We added "**in 2011 – 2012"** as proposed.

Page 14 Line 27: I have to admit, I am not familiar with how to reference data and supplementary material in HESS, but certainly it requires more than the word "Data".

Page 14 line 27 our indication (Figure 9, Data) was for indicating to the reader to focus on the observed values in figure 9 (labelled in the legend as "Data"), and these dots are named "Data" by opposition to model simulation. There is no reference to supplementary material. We rephrased the sentence as:

For the saturated zone, observed values of HI values were negatively correlated with the total annual rainfall for both the hillslope and the riparian zone, with a more negative slope for the hillslope (Figure 9).

And we modified the Figure 9 caption as follow:

Figure 9. Variations of hysteresis Index, observed (Data) and simulated (M1 to M4), versus (…)

Page 15 Line 13: The two sentences that begin after "The Figure 11a …." can both be removed and the beginning of the next paragraph relocated to after "….. stream dynamics."

These sentences were added to introduce the acronym of the various relationships, which are used a lot in the following text and in several Figures. We agree that these two sentences are a bit heavy, nevertheless we feel that without a proper definition of these terms frequently used in the following text and figures, the reader would find difficult to follow the argument.

Page 16 Line 15: Could these errors be due to the difference between a serial and parallel reservoir model structure?

As indicated above, a purely serial model could not adequately reproduce the observed system dynamics, as tested for but not shown in Hrachowitz et al. (2014). The main problem is how to represent the rewetting period where riparian zone is reacting very fast to rain while the upstream hillslope reservoir does not react because unsaturated zone is not filled. As discussed above, a combined parallel-serial set-up will, however, be necessary to model nutrient transport.

Page 17 Line 16: I agree. The authors could explain why they chose to build models M3 and M4 this way.

Detailed justifications are given in Hrachowitz et al. (2014), where it is also explicitly stated that the best performing model of that study (here: M4) does not have to be seen as the end point in the model development process, but as the current state of knowledge for the study catchment. Further refinements are subject to currently ongoing research, which will, of course, be partly based on the results and interpretations reported here, i.e. a potentially different formulation of the riparian zone.

Page 19 Line 15: While the content of this paragraph is interesting, I think it takes away from the focus of the paper. I suggest removing it.

In our point of view the perspective in terms of water chemistry in really important as it has been shown that taking into account only water will lead to miss important processes (e.g. Hrachowitz et al., 2013 a) therefore we prefer to keep this paragraph.

Figure 1: The legend makes it difficult to discern some of the details of the map (e.g., soil moisture sensors vs. buildings). Where in France is this, near Quimper? Perhaps the map of France could be made larger. I think you have the space. Where is the climate station?

The figure 1 has been modified to improve its readability and we modified the figure caption in order to provide the requested information: "Study site in West Brittany (square near Quimper) and location of the monitoring equipments. The weather station is located 500 m north of the catchment."

Figure 2: Maybe state in the caption that the two profiles are sB1 and sB2.

The minimum and maximum are not refereeing to sB1 and sB2, as explained we integrated the two profiles together and their different depths to compute this normalized unsaturated zone storage variable.

Figure 6: In the middle panel – c) I believe – streamflow goes up faster than saturated storage. So where is the streamflow coming from? What does the streamflow vs. riparian storage curve look like? In order to glean from which storages the streamflow is originating (hillslope unsaturated zone, saturated zone, riparian, surface, etc.), all these curves should be provided. This will help explain which runoff processes and sources are predominant at which times.

This is largely due to the limited storage capacity in the riparian zone: most of the additional precipitation cannot be stored there and is directly routed to the stream. In contrast, the "saturated" storage on hillslope is still very low because most of the incoming precipitation will first, to a large extent, be stored in the unsaturated/root zone to overcome the storage deficits that accumulated at the end of the summer. Although percolation to recharge the groundwater also slowly starts to pick up, this will only become effective after a while. The detailed interpretation of the stream flow vs. riparian storage relationship is explained all along the previous section (similar to what is illustrated in Figure 4 a and c). There are already a lot of figures and for the sake of clarity we had to make some compromise. The origin of the stream flow is precisely the purpose of the section 3.1.3 Interpretation.

Figure 7: In the third panel from the left, does the hillslope cross section not imply the conceptual model should be include serial reservoirs? Also, how is HUS going up as HSS goes up? Does this imply that the increase in unsaturated storage does not offset the loss of thickness in the vadose zone as the water table rises? Is this realistic?

Given that the study catchment consists of a first-order channel, it is of course likely that water generated on the hillslope sooner or later needs to pass through the riparian zone. However, as the riparian zone is likely to be fully connected with no more additional storage capacity at the time the hillslope starts being active, the flows from the hillslope are likely to pass the riparian zone without any significant delays on the daily time scale, in particular also due to the comparatively high storage coefficent of the riparian reservoir (KR; see Table 3). We thus do not think that a serial set-up is necessary nor warranted by the available data here.

We do not see where the reviewer refers to when saying that HUS goes up as HSS is going up, as this does largely not happen in the model (Fig.7). Only once HUS is converging towards a maximum the rates of increase of HSS significantly increase (Fig.7, column 3, row 2). However, due to the deep groundwater table in the study catchment it is true that increased storage in HSS is not offset by a loss of thickness of HUS. This is owed to two linked reasons: (1) in the model, HUS is not the entire unsaturated zone, but rather only its dynamic part (i.e. the root zone), which can experience substantial storage deficits that cannot occur to that level below the root zone. In other words, in the root zone the available storage capacity is between wilting point and field capacity, while below that the amount of water stored, on average, is close to field capacity without further storage capacity against gravity. The unsaturated zone below the root zone can thus be considered as a mere flow transfer zone that is only introducing a time lag to recharge but is otherwise hydrologically irrelevant at the spatial and temporal scale of this application. (2) The groundwater tables are comparatively deep on the hillslopes, thereby not interfering with the root zone depth. Thus, yes, we consider this quite realistic in the study catchment.