We thank the reviewers for their detailed feedback that helped us to improve the manuscript.
In the following, we provide a short overview of the main changes of the manuscript:

(1) Motivated by the comments of both reviewers we use now the correlation between soil moisture and runoff and between soil moisture and evapotranspiration as a new measures of coupling between these quantities. We agree that this is more comprehensible and that it makes the results easier to interpret.

(2) The Orth et al. 2013 paper that introduces and describes the simple water balance model is now accepted (with minor revisions that will not change the model). As it is not published yet, we upload it as additional material to provide the editor and the reviewers with detailed information on the model.
During the revision process of that paper, the structure of the model has changed. That is, we now distinguish between runoff and streamflow to account for sub-surface transport of water to streambeds and for the transport of surface water to the stream gauge site. This also implies that precipitation can only modify soil moisture at the day it falls and not later, as in the former model version. Furthermore we do not account for interception explicitly anymore as it turned out to have no significant impact. Despite these changes in the model structure the computed memories are mostly similar to the results of the previous version, also our main findings are not impacted.
The performance of the model calibration with the new version changed for some catchments; in most cases it improved, but by a different extent. As a consequence, the selection of catchments has changed. As described in Section 3, we now consider 100 catchments with the highest correlations between the modeled and observed runoff.

(3) We restructured the manuscript to make the results section more comprehensible and to better convey our key findings. The descriptions of the computations of the slopes and the persistence time scales have been moved to the methods section. Furthermore we introduced a subsection on forcing memories and variabilities in the results section.

Reviewer #1:

This is an interesting paper that contributes to the analysis of soil moisture memory in the climate system. The persistence time scale computations at the end are very interesting, particularly the identification of asymmetrical behavior in dry and wet anomalies. I found the discussion of the “coupling diagnostic”, however, confusing and non-intuitive; I think that part of the analysis should be refined or replaced.

A1: We share this important concern of the reviewer. As described in the beginning, we replaced the coupling diagnostics with the correlations between soil moisture and both, runoff and ET. This is described in the manuscript in the newly introduced Section 2.5 (lines 154-167). These correlations are easier to interpret, for instance in terms of the direction or the strength of the respective coupling.

Major comments:
1) The definitions of the coupling strengths in equation 4 and 6 are not intuitive, at least to me. Why are these useful constructs? Why is the correlation between evaporation and radiation subtracted in equation 6, and how is the coupling strength computed with equation 4 affected by the fact that w_n and P_n are strongly correlated with each other? What does it mean for the coupling strength to be about -1 rather than 0 in Figure 5 or 8? The equations are provided without explanation, and substantial explanation is needed. It's not obvious, for example, why the authors didn't try to isolate the impacts of different contributions to memory through, for example, a multiple regression analysis.
Intuitively, it seems like the raw correlations between runoff or evaporation and soil moisture should be the most relevant things to consider. If the authors are trying to show in equation 6 that the variance of R_n reduces the evaporation soil moisture memory connection, they should note that this reduction is already effectively accounted for in the term rho(E_n,w_n) by itself. For me, Figure 8 has great but unmet potential; it would be much more informative if the y-axis was the evaporation or runoff memory or even the correlations between evaporation or runoff and soil moisture.

A2: As described in answer A1 we use now the "raw" correlations as measures of coupling, as also suggested by the reviewer. Correspondingly, the y-axis in Figure 9 now shows the correlations as new measures of the coupling strengths.

2) Given that the runoff consists of contributions from past precipitation events (equation 2), some quantitative statement is needed regarding how the timescale parameter tau is imprinted onto the calculated runoff memories. Analytically, you would think that if you knew tau exactly, you would know precisely what part of the runoff memory is associated with that tau alone (as opposed to the part of the memory associated with precipitation or soil moisture memory). The knowledge that tau affects the runoff memory diminishes the illustrated link between runoff and soil moisture memory; the reader will want to know precisely by how much. Statements such as that on p. 12121, lines 10-12, seem to imply that the contribution of tau to the runoff memory is negligible.

A3: With the new model formulation the runoff occurs instantaneously in response to a precipitation event and is not governed by tau. However, the streamflow occuring in response to runoff does depend on tau.
The reviewer is correct that memory in streamflow may therefore result not only from soil moisture memory but also from the tau parameter. Even if this may

locally weaken the link between the two memories, the latter is nontheless overall the main control of streamflow memory as illustrated in Figure 6.
We added this argument to the manuscript in lines 328-331.

3) On a related note, the discussion in the first paragraph of section 4.3.2 is a little confusing regarding P_n. It seems like P_n* should have two sources of memory: that associated with the prescribed timescale tau, and that associated with true memory in the precipitation forcing. I think that the text is implying that the latter is negligible, but it doesn't have to be. Some separation of the timescale and background forcing contributions to P_n* is needed; it would be nice to state unambiguously the impact of forcing memory on soil moisture memory.*

A4: This confusing paragraph has been removed.
The manuscript now states in lines 313-315 that forcing memories are negligible on a daily time scale whereas they increase towards a monthly time scale. Therefore they impact the streamflow and ET memory on that time scale but nevertheless these are still mostly governed by soil moisture memory and the respective coupling strength, as stated in the manuscript in lines 343-346.

A further investigation of the impact of the forcing memory on soil moisture memory is not within the scope of our study; it is illustrated and described for example in Orth and Seneviratne 2012, JGR.

4) The dots in Figure 4 (and in the other Europe figures) are difficult to read; I don't have a suggestion for fixing this, other than to make the dots much larger and accept that there will be overlap. The differences between the dry and wet cases in Figure 10 (as discussed in the 2nd paragraph of section 4.5) are very difficult to pick out.

A5: Already with the current size of the dots in Figure 5 (former Figure 4) there is some overlap between catchments, especially in Central Europe. We changed the color of the political borders to gray to improve the readability of this figure. We expect further improved readability from the portrait format that this figure will have in the final paper. Therefore we did not adapt the figure to landscape format which would be more suitable for the current version of the manuscript.
The differences between persistences following dry and wet anomalies in Figure 11 (former Figure 10) are indeed not very clear but present. But instead of a shortcoming in the visualization of the results we consider this a result of Figure 11. It supports the conclusion that memory/persistence is associated with both dry and wet anomalies and that the memories are (roughly) similar under extreme conditions regardless which kind of extreme we consider. We adapted the manuscript to state this more clearly in lines 447-448.

5) More importantly, though, it seems like the runoff memory maps (and mean duration maps) should be supplemented with maps generated from the actual runoff observations, so that the features pointed to are demonstrated to be more *than just a reflection of model assumptions. This will lend substantial credibility to the results. One thing would be especially nice: can the authors demonstrate with observations alone that persistence time scales for runoff are lower for the dry cases, as suggested by Figure 9?*

A6: As suggested by the reviewer, we included an investigation of persistence in observed streamflow in Figure 11. It shows similar results as for the modeled streamflow. However, even if this seems to underline the good performance of the model, it is not very surprising as the model is calibrated with observed streamflow as stated in lines 463-465.

6. Overall, the paper would benefit from further discussion on why evaporation and runoff memory is important. Runoff memory is presumably important for water resources, but how about evaporation memory? The evaporation memory of relevance for, say, weather forecasting is probably that for E/Rnet rather than E.

A7: To underline the importance of streamflow and especially evapotranspiration memory we adapted the introduction in lines 48-51:
"In case of streamflow, this question is of high importance in relation with flood prediction and water resource management. An evapotranspiration memory has implications on the exchange of water between the land and the atmosphere and moreover on surface temperature as evapotranspiration is (negatively) related with sensible heat flux."

Minor comments:
6. Figure 1 should show (perhaps with tiny dots) the locations of the catchments generally not considered (though used in Figure 2).

A8: Changed accordingly.

7. Figure 3a: The total water holding capacity of the soil should be stated in the caption, so that people can interpret the values on the x-axis.

A9: Changed accordingly.

8. Section 4.3, introduction. Here the reader should be reminded that the memories plotted fall out of a calibrated model rather than from direct observations of soil moisture, evaporation, and runoff (though see comment 4 above regarding runoff). Another such reminder is appropriate on p. 12122, line 17.

A10: We adapted the manuscript accordingly in lines 285 and 432.

9. Section 4.3.1, first paragraph. "The 1-month-lag memories are higher: : :" This is probably also because the latter part of month n is very close to the first part of *month n+1, and this proximity finds its way into the calculations with the monthly averages.*

A11: It may be the case that the proximity between the end of month n and the beginning of month n+1 tends to increase the estimated memory, but on the other hand the monthly averages we consider also include values from up to 60 days apart (beginning of month n and end of month n+1). Therefore there should be no systematic difference between daily and monthly memories resulting from the calculation method.

10. p. 12117, line 14. This is confusing; the memory of cumulative weighted precipitation is not equivalent to a "forcing memory"; it is some combination of a true forcing memory and a fitted recession timescale (see comment 3 above). The same confusion regarding "forcing memory" finds its way into the "summing up" paragraph at the top of p. 12118.

A12: We thank the reviewer for this comment; to prevent this confusion we removed the first statement about the role of the forcing memory in shaping the resulting streamflow and ET memories and added a reference to the newly introduced Section 4.3.2 that deals with the forcing memory in detail in line 335. We also adapted the concluding paragraph and state now that the forcing memory only has an impact on streamflow and ET memories on longer time scales (line 354).

11. Section 4.4.1, first paragraph. There are a lot of different time scales considered in this paper: the runoff recession time scale, the applied lag, and the averaging time scale. In this paragraph I would label the time scale as the "averaging time scale", for clarity.

A13: We adapted the manuscript accordingly in line 365.

12. Figures 9 and 10 are discussed out of order.

A14: Changed accordingly.

Reviewer #2:

Using a simple water balance approach, the authors investigated how soil moisture/storage and soil moisture memory affects both runoff and evapotranspiration memory in >100 catchments across Europe.

The general topic of this paper (catchment storage and its effect on various hydrologic processes) has, especially recently, received a great deal of attention from the hydrologic community and will likely stay topical in the future. While the *authors present some promising results, the paper in its current form raises questions I will discuss below.*

General comments: The authors acknowledge that their "soil moisture," a residual of the water balance, is not actually soil moisture but incorporates all sorts of storages in the catchment. Why not simply call it (catchment) storage? It would be a more accurate term for it in my opinion.

B1: We are thankful for this comment. However, with the new formulation of the model, soil moisture does not incorporate groundwater dynamics (and storage) anymore. Therefore we feel that "soil moisture" is the appropriate term in this case. This is furthermore consistent with the Orth et al. 2013 (JHM) study that introduces the simple water balance model.

Model and model calibration: since the authors are referencing a paper in review for the methods (see technical comments), more details about model structure and especially model calibration would be desirable.

B2: We share this important concern of the reviewer. As mentioned in the beginning the Orth et al. 2013 paper is now accepted and we provide the final version as additional material.

What was the reason to simply use a (Pearson?) correlation coefficient and not one of the established objective functions used in hydrology (e.g. Nash-Sutcliffe)?

B3: We use a Pearson correlation coefficient for this purpose because with the simple model we try to capture the temporal evolution of soil moisture or streamflow rather than their absolute amount, because this is important to represent memory characteristics. We clarify this point in the manuscript in lines 102-104.

What exactly is the accuracy term in Table 1?

B4: This accuracy term denotes the step width that is used for each parameter in the streamflow optimization procedure to calibrate the model, and therefore it reflects the accuracy of each parameter. To clarify this point, we adapted the manuscript in lines 105-106.

Also, the fact that the model actually reaches the upper limits of Cs, alpha, and gamma suggests that the model would exceed those boundaries in order to get a better fit. Unreasonable model parameters are usually an indicator of a model structure that is not suitable for a given watershed or runoff characteristics. And it would be nice to see how the simulated Q actually compares to the measured Q (at least for the case study) since all further analysis depends on the quality of the simulated runoff.

B5: We thank the reviewer for this comment. Additionally to the validation of streamflow (memory) in Section 4.1 we added the observed streamflow evolution to Figure 4 as suggested by the reviewer. It shows that, corresponding to our focus mentioned above in B3, the modeled streamflow reasonably captures the temporal evolution of the observed streamflow.
We share the important concern of the reviewer that model parameters colliding with the respective bounds are a problem. Even with the updated model version the problem is present for one parameter, the maximum ET ratio. However, only in a few catchments this parameter collides with its upper bound, such that this is no general problem. This question is also adressed in the Orth et al. 2013 paper, who conclude that this upper bound is physically justified and that in nature energy input from processes not captured by the simple model, such as warm air advection, may contribute to ET rates that may temporarily exceed the available net radiation, which may consequently lead to this parameter collision.

Also, how sensitive are the model parameters for the calibration, especially Cs, the only somewhat physically interpretable parameter? Simulated Q and subsequent analysis/interpretation highly depends on Cs. A non-sensitive Cs parameter would raise questions about the interpretability and significance of the results. Are the derived Cs values in a meaningful range for the chosen catchments, i.e. is there a way to compare/validate the calibrated Cs values with actual soil measurements like depth and porosity? A simple check could be plotting the Cs values for all catchments to check for consistency in catchments that are spatially close together and could be considered (somewhat) hydrologically similar.

B6: Also this question about the sensitivity of the model performance to particular parameters is addressed in the Orth et al. 2013 paper; it is shown that parameter uncertainties have only a minor impact on the estimated (soil moisture) memory. Motivated by the reviewers suggestion we added Figure 3 to the present manuscript illustrating the fitted water holding capacites. We generally find similar values in nearby catchments. Some few exceptions are probably due to the heterogeneous nature of soil and land cover characteristics.
We added this information to the manuscript in lines 236-241.

*What do Equations 4 and 6 actually mean? It looks like can theoretically range from -2 to +2. What do those values mean for the coupling strength? Isn't there some equifinality in the results? For example, you can get (Qn, wn)=0 for an infinite number of (Qn, wn) and (Qn, P*n) combinations. Would the interpretation of be the same, no matter what combination it is derived with? How meaningful is it to look at correlations between Q and wn, and E and wn when wn is actually used to calculate both Q and E?*

B7: We share this important concern of the reviewer regarding the coupling diagnostics and replaced them with simple correlations between soil moisture and both, runoff and ET, as introduced in the new Section 2.5. Concerning the latter question of the reviewer, we think it is meaningful to compute these correlations. Because with our coupling diagnostics we assess the impact of soil moisture on runoff and ET **in comparison to** the impact of the atmospheric forcing. This is possible because both, soil moisture and the atmospheric forcing (precipitation for runoff and net radiation for ET) are used in the computation of runoff and ET.

Figure 3a: Shouldn't the fitted lines have a y-maximum of 1 since the ratios are bounded by P and Rn, respectively?

B8: The reviewer is correct, we changed the figure accordingly.

How do you define "significance" in the results section? (e.g. 12116/1)

B9: As we did not perform statistical tests on this point, we removed this term from the manuscript.

I would consider restructuring section 4. The results seem to contain methods as well, e.g. the computation of the persistence time scales in section 4.5, or the slopes of the normalized runoff and ET functions in 4.4.2. Maybe break it up into distinct "Results" and "Discussion" sections and keep methods restricted to the actual Methods section? In general, I have found section 4 to be somewhat confusing, and it's possible that some of the key findings are unintentionally obscured by the mildly confusing structure of the entire section.

B10: We thank the reviewer for this remark and addressed it as mentioned in the beginning. As suggested by the reviewer we moved the descriptions of the computations of the persistence time scales and slopes to the methods section. Furthermore we introduced the new section 4.3.2 about forcing variabilities and memories to better convey our key findings in the other subsections of the results section.

What is the reason for the lack of a clear spatial pattern in the soil moisture anomalies in Figure 10? I would assume that the extent of dry anomalies exceeds the local scale and affects much greater areas. This is not evident from Figure 10.

B11: The reason for the weak large-scale pattern in Figure 10 is probably that soil moisture reacts differently to dry anomalies depending on the catchment characteristics and on the initial soil moisture at the beginning of the dry anomaly. As reported in Orth and Seneviratne 2012, soil moisture memory depends to a large extent on soil and vegetation characteristics, and these are known to be rather heterogeneous even between spatially close catchments. We clarified this in the manuscript in lines 435-436.

Technical comments:
12105/20, 12106/5, 12107/20: Referencing a paper that is in review (Orth et al., 2012) is less than ideal (especially in the methods section).

B12: See answer B2.

Fig 5: Consistency with labels. I'd suggest using (En, wn) instead of (ET, Soil moisture). The same goes for runoff. The axes labels don't have units. And I would also suggest renaming the axes labels to (Qn, Wn+lag) etc, but that is just a personal preference.

B13: Changed accordingly.

Koster and Mahanama (2012) reference title wrong? Actual title: Land surface controls on Hydroclimatic means and Variability?

B14: We displayed their preliminary title in our references and now updated the manuscript accordingly.

The nice maps (especially Figures 4 and 10) may be a little too small in their current form if the paper is printed. It's difficult to actually recognize what is going on.

B15: We trust that the portrait format of the final paper allows Figure 5 (former Figure 4) to be larger with consequently improved readability. Therefore we did not adapt the figure to landscape format which would only be more suitable for the current version of the paper. We adapted and enlarged Figure 10 after removing the ET persistence maps and moving the color bar to the bottom. Furthermore, in both Figures we draw the political boundaries in gray to improve the readability of the catchment points.

Figures 9 and 10 should be switched as Figure 10 is discussed first.

B16: Changed accordingly.

Final remark: The paper deals with an interesting topic and shows some nice results. However, referencing a paper in review in the methods section is problematic, especially when writing that the model was validated in that particular paper (12112/10). Because of that, I suggest either greatly expand the methods section and be more elaborate on the model validation or even hold off a publication until the Orth et al., 2012 paper has been published, as this appears to be a crucial paper since the current model structure is apparently being discussed there. Also, like I suggested/mentioned above, section 4 is slightly confusing (especially the "coupling" part), and I personally would prefer a clearer distinction between methods, results, and discussion.

B17: We thank the reviewer again for the comments and suggestions. The two main comments mentioned in this final remark were adressed as mentioned in answers B2 and B10.

Lastly I would suggest having a revised manuscript checked by an English editor, as parts of it are somewhat difficult to understand. To some extent the confusion from section 4 may be a result of the somewhat confusing language as well.

B18: To clarify the manuscript, we adapted it at various points (shown in red in the revised manuscript), especially in the results section.