Dear Editor,

let me first of all thank you for providing us extra time to reply to the reviews, which we found most helpful for improving our manuscript.

According to the suggestions of the three anonymous referees the manuscript “Exploring the interplay between state, structure and runoff behaviour of lower mesoscale catchments” was subject major revisions. The most important aspect in this respect was that we splitted the study and deferred its contents into two separate manuscripts: The first part deals with seasonal runoff formation and the water balance, the second part with runoff generation at the event scale. To this end we finished the revision of the first part. A revised version on event scale runoff formation signatures will be submitted separately. In line with our responses to the reviewers we revised our manuscript as follows:

- To clarify the (new) scope of the article we changed the title into: “Unraveling abiotic and biotic controls on the seasonal water balance using data-driven dimensionless diagnostics”.
- Hypothesis and research questions were revised to clarify the overall goal of the manuscript.
- The manuscript was re-structured and several chapters were entirely re-written. This includes particularly abstract, introduction, and discussion and conclusions. The results on seasonal runoff formation are in essence the same but were complemented with additional findings to explain the proposed method in more detail. Compared to the first submission the manuscript was shortened by 15 pages.
- A key aspect we address in our paper is the derivation of dimensionless fluxes which ensure the comparability of the signatures in space and time. We clarify this point in abstract and conclusions and provide details on two different approaches in section 2.2.

All comments raised by the referees and by the editor were considered. In the following we provide brief revision notes to all major aspects.

##

Editor Comments

–

1. ...the units of most of the quantities that goes into these indices are not reported. Using common sense for the units of 1. AC and eFC are usually dimensionless, whereas P, E and Q are usually in mm/time, meaning that dS* is not dimensionless.

Authors reply: in the revised manuscript all units are reported. We furthermore explain that AC and eFC, both provided in (mm/dm), are multiplied by root zone depth (dm) and hence converted to (mm).

–

2. Is there a clear advantage in using scaled quantities, such as the ones proposed, over using not

scaled quantities? Would it be possible to show it? Do scaled quantities lead to better fits than not scaled quantities?

Authors reply: We employ dimensionless diagnostics to ensure the comparability of the signatures in space and time - not to obtain better fits. Comparability is vital for intercomparsion studies. We explain this in more detail in the revised version of the manuscript.

–

3. Clarify the difference between dimensionless and normalized

Authors reply: We apologize for not being precise. In the revised version we only use the term "dimensionless" and avoid any use of "normalized".

–

4. State and structure are terms that are usually applied in the context of models. Here the authors apply these terms in the context of catchments, i.e. of natural systems. The meaning of this terms needs to be clarified at the beginning of the paper.

Authors reply: We clarify this issue in the introduction (P3 L15).

–

5. The objectives are a key part of the paper. I think they are not very clear. What does “feasible” mean in objective 1? Please elaborate and clarify objective 2. Objective 3 would benefit from a clarification of the terminology, e.g. why catchment structure does not include ecology. Results and conclusions should demonstrate how the objectives are reached (i.e. not leave it to the reader to find out).

Authors reply: We reformulated the objective and adapted the research questions according to the new scope of the paper.

–

6. The paper talks about “model based estimates of evapotranspiration” without explaining how these estimates are obtained (some details are given, but much later in the paper). Restructure to clarify and motivate this choice.

Authors reply: The method section was restructured. Information on the site and the available data is now introduced before we detail on the applied methods.

–

7. The catchments are classified based on physiography, but the conclusions say nothing about insights on catchment behaviour. Which catchment characteristics control hydrologic response at that scale?

Authors reply: The conclusions were revised and include several statements concerning the insights we obtained on catchment behaviour. Among others, these include:

- P18 L27: … temperature data prove to be good predictors for CR_S, independent from the physiographic and climatological conditions represented by our data set.

- P19 L11: ...the product of the topographic gradient and the saturated hydraulic conductivity is an important predictor for the average winter runoff coefficients while the two variables alone are not.
- And more general P20 L6: dimensionless double mass curves in combination and ecological temperature index is well suited to unravel biotic and abiotic controls on seasonal runoff formation as long as runoff formation does monotonously increase with storage.

–

8. Can some of the tables be converted to figures?

Authors reply: The revised paper was shortened by 15 pages. Only 7 out of 11 figures and 4 out of 6 tables are included in the revised version of the paper. For this reason we decided not to convert any tables into figures.

##
Reply to anonymous Referee #1

–

1) I cannot find a clear main novel contribution of this paper from either the abstract or the conclusion. This may be challenging to formulate as the paper tries to understand catchment similarity for different time-scales and processes, but I do expect that there are some overarching goals.

Authors reply: Abstract and conclusions were re-written.

–

2) The abstract is a mess: first you state a very broad and poorly defined problem, sub- sequently you state many details of the analysis that are not informative is the problem is still unclear. For example how is the reader supposed to know what you mean by “extensive/additive” and “intensive/non-additive”. After you introduced the methods, your results focus on differences in functioning at different time-scales. Why didn’t you introduce that you are looking at different timescales in your problem statement or methods? Your results do not clearly link to the introduction and methods part of your paper, and read like things you ad-hoc found and listed. You start stating results “Our dimensionless signatures evidently detect. . .” before the reader knows what data this comes from. This seems like an illogical and confusion order to me. Your conclusions feel like a random list of unconnected findings. Do you really need so many words for your abstract? I think you can (and should) be more to the point.

Authors reply: The abstract was re-written. Terms such as “extensive/additive” were deleted.

–

3) The introduction section lacks a logical build-up to a problem statement. Although I acknowledge that the introduction is giving a relevant overview of many challenges we face in

concisely characterizing how catchments function, it reads like a long list of problems, rather than a structured introduction working defining a clear problem and working towards a clear goal.

Authors reply: The introduction was revised according to this suggestion and build-up towards a clear goal. We hope that the new version is more clear.

–

4) Are you sure that “Question 2: Can we detect intensity controlled runoff formation essentially a high frequency process based on low frequency data?” is a clearly defined question for the (ignorant) reader (like me).?

Authors reply: Question 2 was reformulated. A more precise definition of “ intensity controlled” is provided (P8 L32).

–

5) What is the coherence of Q1-Q3 beyond “testing dimensionless measures to discriminate differences in runoff generation”? If there is coherence, please formulate it such that this is clear. If there is no coherence, why would you address these three questions within a single manuscript?

Authors reply: Due to the splitting of the manuscript Q1-Q3 were reformulated. This issue is not relevant any more.

–

6) The section “Conceptual framework and candidate diagnostics” needs to be improved. Although you list requirements of the paper in the different subsection the writing is messy. I think you should be much more to the point in clearly stating the requirements. Problems that I came across and a. “Requirements of functional diagnostics” is mainly an overview of your perception of how the hydrology of different landscapes function. Right now the reader reads one sentence about the “requirements” but are confused by the end of the section 2.1 what the purpose of this section is. b. The requirements of “Normalization of states and re- sponse measures” and ““Coherence and quality of integral storage measures” should be concisely written and to the point. c. I understand that you choose the metrics in section 2.2 based on all the requirements you have set before, but when you introduce them in this section I expect that the rationale of choosing them is clear. Currently it is not. For example “Lastly we use a normalized specific pre-event discharge (Q) av- eraged across the last seven days:” is just an announcement, but leaves it completely unclear where the rationale to choose this metric comes from. The introduction of other metrics suffers from similar issues.

Authors reply: The section “Conceptual framework and candidate diagnostics” was entirely re-structured and re-written.

–

7) Is it logical that you provide the details of your study area only after you have discussed all the metric you use? I presume these metrics are largely based on your perception how the study catchment function. Therefor I think it would be more logical to switch their order.

Authors reply: We switched the order of the sections.

–

8) There should be a story in your results, rather than that it is a long list of ad-hoc results you have obtained. For example you begin with the statement that “During low flow conditions, the storage estimators dS and are in most cases linearly independent.” And subsequently you list the associate statistics. The logic in your story is impossible to follow (for me). Listing the how the storage metrics link to flow characteristics is fine, but please make it more read like a story.

Authors reply: We re-structured the results to clarify the storyline. We hope that the revised version is easier to read.

–

9) The paper presents the findings as “generally applicable for meso-scale catchments”. This is an extremely bold statement since only 22 catchments in a very small part of the world are used.

Authors reply: This was a misunderstanding that we clarified in the revised version. In P16 L16 we “conclude that season-specific dDMCs are a well suited fingerprint for characterizing seasonal runoff formation in meso-scale catchments and that dDCMs are suited for intercomparison studies.”

–

10) The paper is uses many words, while I am not against long papers, it does feel that with 30% less words the story can be told too. The current version of the manuscript is not clear enough to allow me to properly evaluate the results obtained in the paper. Therefor I suggest that this paper can only be considered for HESS after rigorous rewriting and restructuring of the paper has been done. I am sure that interesting new findings are available within the results, but at this stage I can not properly evaluate this.

Authors reply: The revised version was shortened by 15 pages.

###

reply to referee #2

Although one of the main focuses is on the normalized dimensionless storage predictors, as primarily described in the method, the manuscript also touches upon other things including the relationship between topography and runoff ratios, temporal sampling frequency, triple mass curves etc. As a result, it is currently very difficult to understand what the authors would like to solve or propose in this single paper and the significance of the manuscript becomes unclear.

Authors reply: For this reason we splitted the manuscript. This study focuses on runoff formation at the seasonal and annual time scale. The methods for the event scale will be presented in a follow-up study.

–

Furthermore, the logic of the evaluation for the three storage predictors must be well defined. Obviously different predictors represent different properties, yet they are all related to the catchment storage. Therefore well understanding of the predictors’ char- acteristics is very important in qualitative way. However, what confuses me is that the authors tend to say the predictor showing the higher correlation to the compared in- dices is the best. Related statements appear many times, for example on L784 in P.24. Please describe clearly your logic on the evaluation of the predictors.

Authors reply: This aspect refers to the event scale measures. In this manuscript we focus on the seasonal time scale. The issues are hence not relevant any more.

–

Secondly, the rationale of the normalization is unclear. The values can be easily converted to be non-dimensional, but it is effective only if you can normalize them in a physically meaning manner. In the current manuscript, all of the result figures show the relationship between the predictors and the compared variables at each catchment and its evaluation basically conducted based on the rank correlation. If this is the purpose, I do not see any necessity on the normalization.

Authors reply: We clarify this aspect and state that dimensionless approaches are required to ensure comparability of signatures in space and time. The derivation of dimensionless quantities is a key aspect of our paper. We discuss this aspect in detail in section 2.2.

–

Moreover, the equation (3) normalizes the average discharge volume divided by soil porosity. What does this mean physically? Also the equation (2) sums up the difference between (P - E) whose total values vary significantly depending on the duration of the summing up. If so, how can you convince it was successfully normalized by the porosity? Why can it be better than the original values? The same concern is applied to the equation (4) also, especially Ks from a soil map can vary significantly by some orders with large uncertainty.

Authors reply: The key to obtain scale invariant dimensionless quantities is to divide a state variable of interest – for instance a force, velocity or length by a characteristic quantity of the system (Blöschl and Sivapalan, 1995). A popular example is the Reynolds number in fluid mechanics which relates inertial forces to viscous forces. It is defined as flow velocity times a characteristic length divided by the dynamic viscosity of the fluid, which is suited to compare turbulence in open channel flow independent of the channel width. The best known example of a dimensionless response measure in hydrology is the runoff coefficient (CR) (-) defined as specific discharge (L/T) over specific rainfall (L/T). The latter is often used as "diagnostic" variable to detect scale invariant differences in generation of runoff volume (Merz et al., 2006, Graeff et al. 2012).

In the revised version of the manuscript we explain two versions to obtain dimensionless double mass curves. We show that the kind of scaling strongly influences the diagnostic potential of the signatures.

We are well aware that data on soil porosity are highly uncertain, particular at the catchment scale, but this is what is operationally available in Southern Germany.

##
reply to reviewer 3

1. While I generally enjoyed the paper, I found the organization of section 3 very difficult to follow. What data do you use to compute your statistics? Many different data are mentioned in this section including met data (rel- ative humidity, etc) and modeled data (ET, etc). I am guessing some met data are mentioned only because they were needed for modeling, while other products were used directly to compute statistics. It would be helpful if you directly state which data are used to compute your statistics and which are used for LARSIM. I think you were trying to be inclusive, but the generality here leaves me a little confused. For instance, what data are included in hydro-meteorological time series?

Authors reply: Section 3 was entirely re-written and re-structured.

–

2. Along the same lines, you mention catchment characteristics data in section 3.1, but do not direct readers to Tables 1 and 2, or mention which catchment characteristics you derive or why you derive them. Brevity is good, but I found that without introducing them, when you mentioned derivatives from these datasets in results and discussion I was a little lost.

Authors reply: In the revised version we introduce the tables and describe how the different properties are derived.

–

3. I guess I see the requirement of modeling (to obtain ET) as a potential source of added uncertainty in this analysis. While I see that the modeling results are published in a second paper, and I recognize you do not have to defend these results in this new paper (and as well that ET is an important part of a catchment functioning analysis), I do think it is important that you acknowledge some of the uncertainty that comes from using a derivative of a model in this type of analysis. I think our community does a good job acknowledging potential sources of uncertainty from modeling and problems related to equifinality, but I also think we run the risk of incorrect interpretation when we do not directly include this uncertainty in analyses such as the one presented in this paper. For instance, if your input data to your model was assumed to have an error of +/- 5 or 10%, and this error was propagated you're your ET signal, would this change your conclusions? It might be worth adding a short discussion of potential limitations in such an analysis. I'd be interested to see if you come to similar conclusions if one or two of the sites are located close to an Eddy-Covariance tower, and if those measured values were used in this analysis (recognizing as well that ET from Eddy-Covariance is a derivative data product with error).

Authors reply: Due to the splitting of the manuscript the focus is now on data-driven diagnostics, that means we only use observables. ET is hence not required any more. Accordingly, uncertainties are less important - though of course not negligible. However, to keep the manuscript as concise as possible we decided not to discuss uncertainty issues in our paper.

–

4. In the introduction to the paper and in the discussion, I felt like there were two sources of literature not acknowledged. Again, brevity is important, but this work reads as very similar to the literature on runoff generation and the literature on catchment classification. I noticed the absence

of some seminal/recent papers from both sources that could potentially bolster your introduc- tion and conclusions.

Authors reply: In the revised version of the paper we provide literature on runoff generation and catchment classification.

–

5. I completely understand needing to use abbreviations in results/discussion, but I found it again very difficult to link your acronyms back to their description in Section 2 of the paper. A table linking acronyms, descriptors, and their importance in determining catchment functioning would organize the many different pieces.

Authors reply: Due to the splitting of the manuscript the paper is clearer and we use less abbreviations and acronyms. For this reason we do not provide a table that links acronyms and descriptors.

–

6. You introduce the ideas of additive vs. non-additive in the abstract, but this does not follow throughout the paper. Consider removing, reframing.

Authors reply: These terms have been removed.

7. What makes a dimensionless quantity better than a dimensional quantity? How is what you have framed different from other catchment classification studies? Why should we normalize measures others have used?

Authors reply: Please see our answer to the last question of referee #2 and section 2.2 in the revised version of the paper.