Response to referee comment Anonymous Referee #2

We would like to thank Anonymous Referee #2 for taking their time and effort to read our manuscript. The constructive comments will be used to improve our manuscript. Below, we will respond to the comments made by Anonymous Referee #2 (original comments in black, our response in blue).

The presented study investigates the differences in hydrological response of 5 mesoscale Swiss catchments to seasonal extremes of temperature and precipitation at two different spatial model resolutions. Therefore, the authors apply the SPHY model at hyperresolution (500m) and at a resolution of 40km, which respresents a lumped model approach. To identify hydrological response they analyse seasonally aggregated runoff and evapotranspiration and conclude that hydrological response simulated with the hyperresolution set up can be more extreme than with the lumped set up. Investigating the effects of model resolution on hydrological response is of scientific significance and is within the scope of HESS. Due to the following specific somments I suggest major revisions before publishing the manuscript.

Thank you for the constructive comments, which will improve and clarify our manuscript.

P1 L4f: “For four seasonal extremes representing flood and drought/heatwave conditions we investigate the simulated response at both model resolutions.” – What are the four seasonal extremes or do you mean you investigate extremes in four seasons?

Based on temperature and precipitation anomalies, we have selected four extreme seasons; winter of 1995, spring of 2007, summer of 2003 and autumn of 2002. Due to the extreme precipitation and temperature values during these seasons, the discharge is showing extreme values as well.

P1 L1-L12: In the abstract you mention hydrological responses. An on P1 L8f mention “two simulated responses” – It needs to be stated in the abstract what these responses are.

We understand the confusion, and will clarify this in the revised version. With the “two simulated responses” we mean the response simulated with the high resolution model on one hand, and the response simulated with the low resolution model on the other hand.

P1 L1-L12: In the abstract you mention the terms hyperresolution, low resolution, corse resolution and distributed models and compare them to each other. If you use these terms, you need to explain them or provide a scale for the specific resolutions. Why do you compare low resolution model response to distributed model response (P1 L9f)? In P1 L2 you introduce a distributed model. Did you apply a lumped model, too?

We will add the spatial resolutions in meters to the abstract, and replace “distributed model” (P1 L10) with “high resolution model” to avoid confusion. We are using the same hydrological model, where the high resolution model contains around 1500 cells, and the low resolution model consists of only one cell (hence the terms distributed and lumped).

P1 L1-L12: Wording and sentences of the abstract need to be revised since they are often distracting.

We will reduce the number of different terms in order to be avoid confusion and be more consistent with the terminology.

P2 L1f: “We note the recent trend in hydrological modeling, which convenes coarse global modeling with local high resolution modeling: hyper resolution modeling (Wood et al., 2011; Bierkens, 2015; Bierkens et al., 2015).” – Please be more precise. The sentence is misleading.

We want to state that hyperresolution modeling is becoming more popular, yet due to limitations in data and/or computational power, most studies are still ran on coarse resolutions. We will adapt this paragraph so our message is more clear.

P2 L17f: What are the normal, new and old situations? Please define them.

We refer here to impact or climate studies, where the current situation is often taken as baseline for comparisons.

P2 L20f: Please rephrase the sentence.

We agree that this sentence is confusing and will adapt it to become better understandable.

P3 L1: What are “complex” basins?

We define complex basins as basins with a lot of land cover and/or elevation variation.

P1 L13 – P3 L16 section Introduction: The introduction is rather an overview of the shortcomings of coarse scale global simulations with regard to their validity for local hydrological processes, of the effects landscape and land cover heterogeneity has on hydrological processes and of the issue of spatial resolution of hydrological models. However, it is not written what your study is about, what your goals are, what you plan to do and how you want to achieve your goals. This needs to be included and the introduction needs to be thoroughly revised.

We will add a paragraph stating the goals and methods of this study.

P3 L6f: “With more than one-sixth. . .” – This sentence does not fit into this section, where is the connection to the content of the section

With this sentence we wanted to stress the importance of snow and glaciers for the water supply. We agree that this sentence is currently not connect to the paragraph, which we will fix in the revised version.

P3 L3 – P3 L16? P3 L4, L23: Scale is not the right word.

Thanks for spotting this error, we agree that scale is not the right word, and will replace it with “pixel size”.

Section 2 “Methods, model and data” should be divided into subsections for better readability.

We agree that the section is currently rather long, and that adding subsections will improve readability.

P3 L29 – P4 L14: The model is described in detail in the cited literature. Please focus on the parameters, routines, etc., which are necessary for your study.

Since SPHY is a rather new model, we wanted to give a quick overview of the concept of the model, so that it is easier to understand our results.

P4 L12: “. . .sparse/bare vegetation (referred to as “other”).” Does this mean that rocks is neglegted, or do you neglegt sparse vegetation? Please clarify!

The “other” class contains both rocks and sparse vegetation types; rocky landscapes are not neglected.

P4 L16f: How did you resample input data?

For the high resolution model, we used bilinear interpolation to resample the forcing data from the original 2x2km to the required 500x500m. To resample the forcing data to the 40x40km pixel, we averaged all values within this grid cell.

P4 17f: Do you use measured runoff data for calibration? If so, where is this runoff data measured?

Yes, we did use measured runoff for calibration. The runoff is measured at measurement stations at the outlet of each catchments, which is obtained by the Federal Office for the Environment (FOEN).

P4 L18f: On what monthly discharge do you calibrate? Mean, sum, median, maximum, etc.? Please clarify!

We calibrate on the sum of the monthly discharge. We will also add this to the method section.

P6 Formula 1 and 2: It is not necessary to give the formula for the mean and standard deviation.

We will remove the formulas for the mean and standard deviation.

P6 L11f: “For this study, we are also interested in the difference between lumped and distributed simulations and its implications on the simulated anomalies.” Isn’t that the core of the study? At least the title suggests that. Please make this clear in the introduction section.

This is indeed the core of our study, which we will emphasize in the introduction.

P7 L16f: Do you just propose this metric or do you introduce it? Is there any literature where this metric has been applied or developed?

We propose the DWD as a new metric to quantify the differences between the lumped and distributed model, since we felt like most used metrics were not able to correctly quantify the differences between those two results. As a result, there is not any literature available where this metric has been applied of developed.

P9 L13: You mention the models inability to simulate discharge during low flow periods. Is the model then really suitable for the study, since you focus on extremes? Please explain this.

The bad model performance was during winter period, only for the Alpine basins (Reuss, Rhone and Inn). During these periods, most water is stored as snow or ice in these basins, and the catchments are not really active. The other three seasons generally have higher discharges, where the model is reaching better performance.

P9 L13 – P10 L1: You assume that the weak model performance at low flow periods might be related to the coarse monthly calibration time step. Why don’t you choose a finer time step? In your work, you focus on extremes and you realize that your model has problems in simulating low flows, which you trace back to the coarse calibration time step, but you don’t test another calibration time step. You should definitely check and explain this.

We looked again the model performance during these low flow (winter) period. We now see that it is related to an error in the model. The measured discharge shows a rather constant baseflow during

these periods, which SPHY was not able to simulate. SPHY simulates discharges very close to zero during winter period, since most water is stored as snow or ice in these catchments, and is not contributing to the discharge. Since we aggregate all our results to 3 months, we do not expect that a finer calibration time step will change the results.

P10 L35: What are “interesting relations”? Please explain.

The precipitation – evaporation scatter plot consisted just of random scatter, indicating that there was not a relation between these two factors.

P11 L1: What is the measure for a possible correlation? Please specify and give a measure for correlation and its significance.

We are sorry for this typo, “correlation” is not the right word. We meant to say “… seems related to …”

P15 L1f: “Nevertheless, this figure does not provide enough information to draw a frim conclusion about the performance of the lumped model with respect to the distributed model results.” – Please rephrase. The sentence is partly unclear. Moreover it raises the question why you show this figure if it is not good enough for what you want to show.

This figure depicts the difference between the high and low resolution model. We want to quantify this difference in order to draw conclusions on how the low resolution model compares with the high resolution model. Therefore we have introduced the DWD metric.

P15 Figure 8: Please give a more informative description. What do the numbers 2056 and 2181 mean?

Those number are FOEN station identification codes. We will replace those codes with the catchments names as they are presented throughout this study.

P16 L21f: “All basins are for large parts covered with snow during the winter month” better: Large parts of the basins are covered with snow during the winter month.

Thanks for this suggestion, we will change this sentence in the revised version of our manuscript.

Technical corrections

Entire text: Please check usage of the word "land use". Land use refers to agriculture, industry, buildings and everything where the natural environment is modified by human activities. Land cover might be a better word in most cases since you do not refer to human activities.

Thank you very much for this suggestion. The term land cover is indeed a better word than land use for out study. We will alter this in the revised version of our manuscript.

List of other technical corrections

We thank Anonymous Referee #2 for their sharp eye on spotting typos in our manuscript. We will definitely fix those errors in the next version of our manuscript.