Detailed response to Anonymous Referee #2

Joost Buitink, Remko Uijlenhoet, and Adriaan J. Teuling

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 manuscript "Evaluating seasonal hydrological extremes in mesoscale pre-Alpine basins at coarse 0.5 deg and fine hyper-resolution" by Buitink et al. provides an interesting excursion into the effects of spatial resolution in hydrological modelling. Two grid resolutions, 500x500 m and 40,000x40,000 m are used in the STAHY model to simulate hydrological processes, and the results are compared in 5 mesoscale basins in the Swiss Alps. The main message is that the coarse resolution model fails to capture the "diverse and contrasting response" from the high resolution model, because topography and land cover are not accurately represented. This is found to be especially true for extremes, where anomalies in climate and their effect on runoff and ET were quantified.

It has to be said that these conclusions are not suprising, in fact are to be expected, and there are numerous studies published in hydrological literature that report the same or similar findings. In this sense, the potential innovation of the paper is rather limited, and would have to be found in the details and/or implications that are specific to the study cathments, climates, model used, etc. To highlight what is really innovative and make the relevance of the paper more clear I suggest to focus in the revision on the following points and questions.

1. What is the real aim of the paper? In the introduction (p2, 10-15) the authors claim that many studies have explored the effect of spatial resolution, but few (none) have explored the "effect of the modelling approach". It should be clarified from the start what is meant by this, because the authors in my opinion also only show the effect of spatial resolution and not modelling approach. They use the same conceptual model, same parameterisations, only the input data differ.

   The aim of this study is to quantify the differences in anomalies as result of a different spatial resolution, using a newly defined metric. We understand that our usage of "modelling approach" can be confusing. With modelling approach we link to the two scales of modelling: global and regional. In general, we identify two common practices related to the scale of modelling: a relatively coarse resolution for global studies, and relatively fine resolution for regional studies. In our study, we aim to compare these two "approaches" while keeping as many factors constant. We understand the confusion, and will clarify this in the revised version of the manuscript. Another innovation of our study is that, as far as we are aware, we are the first to show that it is not uncommon to have both extreme positive and negative flux anomalies occurring simultaneously within a catchment.

2. The studied catchments are all smaller than one coarse pixel in the analysis. This is quite clearly stated in Section 2.4. The input data are resampled to the 500x500 grid and averaged to a 40,000x40,000 m grid for the coarse application. So this sounds to me like comparing a spatially distributed model to a point model, not to a coarse resolution model. This also means that all elevation dependencies in hydrological processes in the point application are gone. Is this correct?

   This is indeed correct, SPHY does only allow for sub-grid variability in glacier cover, but not in elevation via e.g. elevation zones. We are aware that some hydrological models do allow for different elevation zones within a single pixel, yet this still limits the model to output only a single value per pixel. In the "coarse resolution" model setup, each basin is captured within a single pixel, as would happen when simulating this region at 0.5x0.5° resolution. We chose for this configuration to present the difference in a "worst-case scenario" regarding model setup.

3. I would have liked to see at least a table with the values of the key parameters that were calibrated, i.e. something that gives more credibility to the SPHY model application. It seems in Section 2.4 that the model was calibrated for both the spatial application and the point application separately. I assume it is the high resolution application shown in Figure 5. How different were the parameter values? What do the differences (if any) mean for the results of the simulations, e.g. temperature dependencies, etc.

   Thank you for this suggestion. We have added a table with parameter values in Table C1. In some cases, the optimization algorithm reach the parameter boundary. This happens more often with the low resolution model, indicating that SPHY struggled to simulate the hydrograph at these resolutions. Furthermore, we have also updated the calibration and validation figure, so it also includes the discharge simulations with the low resolution model (see Fig. C1).

Table C1. Parameter values resulting from optimization, for both the high and low resolution models.

		rootdepth	tcrit	snowsc	ddfsnow
high resolution	Reuss	461.22	-0.99	0.01	2.11
	Rhone	1719.33	0.06	0.18	2.10
	Inn	2997.40	-0.86	0.01	1.38
	Emme	453.35	-1.00	0.99	3.68
	Thur	402.60	-0.14	0.03	1.05
low resolution	Reuss	402.60	-0.75	0.01	1.01
	Rhone	2997.40	-0.40	0.01	1.01
	Inn	2997.40	-1.00	0.01	1.01
	Emme	1919.75	-0.40	0.70	3.38
	Thur	537.97	0.24	0.19	5.41

4. The results were compared on seasonal anomalies of runoff and ET, summed over the catchment areas. There are no supporting data and plots to actually show how the model performed for spatially distributed variables, beyond Figure 6.

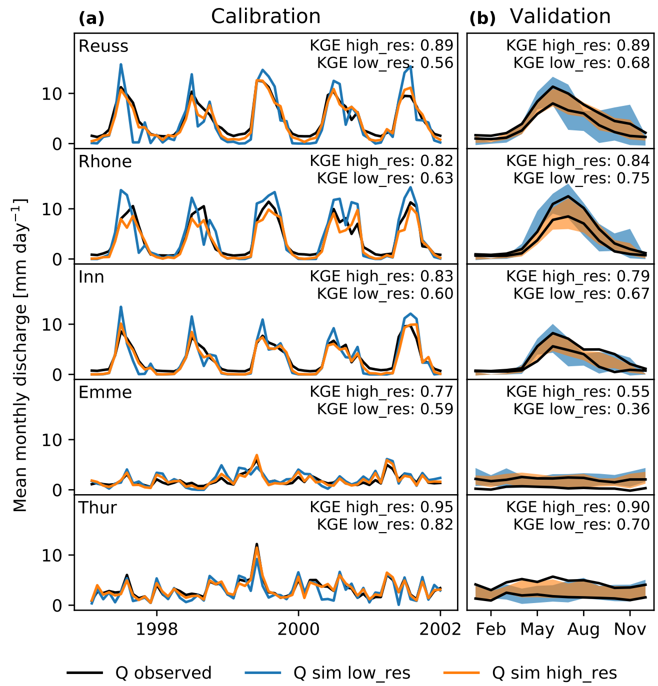

Figure C1. Calibration and validation of SHPY. The model is validated over the entire period, excluding the calibration period. The two black lines, and the shaded areas present the average discharge plus/minus the standard deviation. For Inn, all data after 2003 was excluded, since the observed discharge pattern changed after this period.

For example, snow cover could have easily been compared with data to show how snow accumulation and melt processes are simulated. There is little confidence given to the spatial predictions of the model, on which the entire analysis is based.

This was also stated by the first reviewer, and we added an internal validation of the model by comparing it with observations from the Rietholzbach, a research catchment situated in the Thur basin (Seneviratne et al., 2012). We obtained the evaporation data as measured with a long-term research lysimeter, and discharge data from this catchment. Both discharge and evaporation from the corresponding pixel were extracted from SPHY, to compare with the observations. We calculated the anomalies over the entire simulation period. The comparison between the observed and simulated anomalies can be found in Table C2. This table shows that the simulated anomalies agree well with the direction and magnitude of the observed anomalies. Winter and autumn values for evaporation are gray, since they are not the focus of this study due to the fact that SPHY does not allow for evaporation during snow covered periods. We believe that this table is a valuable addition to the validation of the model.

There is a slight mismatch between the evaporation anomalies during the summer of 2003, yet both describe unusually high values. To further investigate this, we also plotted the evaporation time series in Figure C2. This figure shows that

Table C2. Comparison between anomalies simulated with SPHY and observed in the Rietholzbach, anomalies are based of the entire simulation period.

Event	**Runoff**		**Evaporation**	
	Observed	Simulated	Observed	Simulated
DJF 1995	1.68	2.34	0.86	0.53
MAM 2007	-0.52	-0.27	1.61	0.98
JJA 2003	-2.15	-2.17	1.66	3.61
SON 2002	2.65	2.62	-1.76	-0.26

SPHY is able to accurately simulate the evaporation, yet there are some differences between the two time series. These can be attributed to the scale difference between the lysimeter and a single high resolution SPHY pixel, and the fact that SPHY does not account for all factors influencing evaporation since it uses the temperature-based Hargreaves method. The difference between the two time series influences the mean and standard deviation, and therefore the resulting anomaly values.

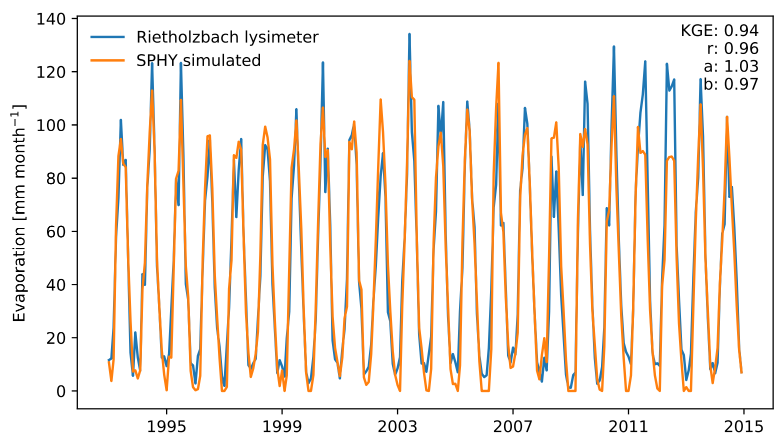

Figure C2. Comparison between observed evaporation in a lysimeter, and the simulated evaporation of the corresponding pixel.

5. A new metric DWD is proposed to show the place of the point model value in the spatial distribution of the spatially resolved values. This is an interesting metric. I appreciate the effort to illustrate its use in Figure 3.

   Thank you for your kind words. Please note that we slightly adapted the equations of this metric, so that the distances are properly weighted. This new formulation only has a limited effect on the results, and does not alter the conclusions. See our detailed response to Davide Zoccatelli for the new equations and the effect on the DWD values.

6. The relation of the seasonal anomalies in ET and runoff to temperature across the basins in Figure 7 is interesting. I have one concern that the results here are probably strongly dependent on the parameterisations and structure of the model. Some of these relations, e.g. between runoff and temperature can be gleaned directly from observations. Will you get the same sensitivities? In Figure 8 the anomalies are plotted for every grid cell as a function of land cover. What about soil depth? Does SPHY assume that soil depth and soil properties are constant in space?

   The simulations are indeed strongly dependent on the model structure and parameterization. However, since we analyzed all results on a three monthly time step, we do not expect the model to be a large factor here, but that differences in forcing are of great importance. SPHY does assume soil depth to be constant in space (the rootzone parameter in Table C1), but soil properties are related to the soil type and are variable in space. We chose to plot the anomalies against land cover since this corresponds to the studies by Theurillat and Guisan (2001) and Jolly et al. (2005). Since land cover and soil properties are related, we do not expect the results to be very different.

7. Overall, I find the physical relations between the results and hydrological processes in Section 3.1 nicely covered, the arguments are logical, especially the elevation effect is coming out strongly. I suggest also looking at the paper by Fatichi et al. (2015) "High resolution distributed analysis of climate and anthropogenic changes on the hydrology of an Alpine catchment" in Journal of Hydrology for another demonstraiton of this effect with a physically-based model.

   Thank you for highlighting this paper. Really interesting to see the strong influence of elevation regardless of climate model used. We will cite this paper in our manuscript as it will be a valuable addition to put our result in perspective.

8. As mentioned above, the resolution effects are less instructive than the explanation of the anomalies. I am not sure what to take out of Figure 9, other than the point model lies within the range of the cells of the distributed model. This was for an extreme year, what about the entire simulation? Probably the results are the same. The message simply seems to be that lumping in space averages hydrological response, which is something very well known. Is there more to it than that? If yes, this has to be brought to the forefront more clearly

   We have included Figure 9 to visualize the DWD results in Table 2. This figure helps to explain why we see some large DWD values in the table, especially for the ET results. During seasons with less or no extreme conditions, the violinplots from the high resolution model are generally less spread out (less variation in anomaly values), also resulting in lower DWD values. We agree that this is currently not clearly explained in the text, and will add this to the revised version of the manuscript.

References

Jolly, W. M., Dobbertin, M., Zimmermann, N. E., and Reichstein, M. (2005). Divergent vegetation growth responses to the 2003 heat wave in the Swiss Alps. *Geophysical Research Letters*, 32(18):L18409.

Seneviratne, S. I., Lehner, I., Gurtz, J., Teuling, A. J., Lang, H., Moser, U., Grebner, D., Menzel, L., Schroff, K., Vitvar, T., and Zappa, M. (2012). Swiss prealpine Rietholzbach research catchment and lysimeter: 32 year time series and 2003 drought event. *Water Resources Research*, 48(6):W06526.

Theurillat, J.-P. and Guisan, A. (2001). Potential Impact of Climate Change on Vegetation in the European Alps: A Review. *Climatic Change*, 50(1-2):77–109.