Anonymous Referee #1

Received and published: 29 July 2018

This paper investigates statistical dependency between extreme river discharge and coastal water level in Rhine river basin. While the authors mainly followed a methodology established by van den Hurk et al. 2015 and Klerk et al. 2015, this study provides a unique contribution in that they used a large set of ensemble model simulation results, not just observations. I think the authors have conducted substantial amount of work and critically analyzed their results, the paper is well written for readers to easily follow, and the findings are scientifically new and interesting. Therefore, I recommend this paper to go through minor revisions before publication. Minor comments are listed below.

We are grateful to the reviewer for the thorough review of our manuscript and useful suggestions, which improved the quality of the manuscript significantly. We provide a point-by-point clarification and response to the reviewer's comments below. For clarity, the reviewer's comments are given in red color, the responses are given in plain black text and the modifications in the manuscript are in blue italics. The manuscript will be modified accordingly.

P 2, L 10: Underestimation of what?

This is a typographic error, and the sentence should be completed as "Ignoring the dependencies may lead to severe over or underestimation of the flood risk". It will be corrected in the revised manuscript.

P2, L10 "Ignoring the dependencies may lead to severe over or under estimation of the flood risk".

P 3, L 3 and after: Please use n-dash (–) not hyphen (-) to indicate certain range of

This will be corrected throughout the manuscript.

P 4, L 5: Add the full name before the abbreviation for TWL.

This will be added.

P 5, L 21: What is E-OBS?

E-OBS are the observed gridded daily precipitation data sets available for European region. Since the data user agreement requires it to be abbreviated as "E-OBS", we use the same acronym and have described this dataset previously in P 3, L30 in original manuscript.

P 6, L 3 and after: The3 in the unit m3/s should be superscript.

We will correct this.

P 6, L4 and L 11: Both 'modeled' and 'modelled' are used throughout the paper, so use either of them.

We will use "modeled" consistently throughout the revised manuscript.

P 6, L 20, 25–26 and 31–32: I agree with your rationale to use two hydrological models to assess model uncertainty as mentioned in P 4, L 8. However, as introduced here, SPHY is strongly biased in reproducing high discharges and HVB performs much better than SPHY. In the supplementary figure 4S, it is shown that SPHY's performance was better than HVB, but given that this paper's objective is to see the dependence between extreme values, it does not support the reason to use SPHY. I am not sure why the authors use such different models in terms of model types (i.e., SPHY is a conceptual model while HVB is a semi-distributed model) and the model physics (written in P 4, L30–31) for comparison.

We acknowledge the reviewer's concern regarding the underestimation of the extremes in the SPHY model. The rationale to use two hydrological models is to include the uncertainty from using different hydrological models. The large bias in SPHY was a clear motivation to include an evaluation of the HBV model as well. However, later recalibration of the SPHY model has led do a clearly improved performance of this model. All SPHY results are replaced and modified conclusions have been included accordingly. The modification to the main results and conclusions are described in more detail in the reply to comments below.

Other aspects and major conclusions remain unchanged. We still point at the need to analyse flood wave length and timing in detail before analyzing the joint flooding risk. We show that flood wave timing is of prime importance while assessing the coastal flood risk. Neither of the models used in this study have perfect performance. The recalibrated version of SPHY outperforms HBV in the representation of the mean annual cycle and daily biases were lower in SPHY. Although HBV performs fairly well in representing the extremes, the flood wave timing was not perfect. The results of two hydrological models with different performance allows evaluation of the impact of these bias on the correlation characteristics, and gives an indication of the contribution of model bias to uncertainty of this joint correlation. Since the results presented in this study are based on quantile thresholds relative to the respective dataset, the biases in the model results do impact the findings concerning the statistical relation between water level and river discharge.

P 7, L 23–24: Why does SPHY have multiple maxima?

The phrasing "multiple maxima" is bit misleading. We intended to say that the maxima is broader in SPHY and is not always a well-defined peak as in the observations. This is the motivation to use the onset of flood wave rather than maxima. However, later recalibration of the SPHY model for the revised manuscript has led do a clearly improved performance of this model. With the improved simulation of extreme flows, more well-defined flood peaks are achieved in the SPHY model simulation. Rather than comparing the onset of discharge peaks, we directly compare the flood wave peak timing. The modification to the main results and conclusions are described in more detail in the reply to comments below.

P 8, L 3–4: 'The broad shape of the distribution of both HBV and SPHY reflects the complex interaction of the climatic and hydrologic processes.' This sentence is too concise to understand the meaning. Why can you say that the broad distribution reflects the complexity in climatology-hydrology interactions? It would be helpful if you can add some more explanations.

We added explanation for this statement in the revised manuscript: *This wide distribution is a result of multiple drivers of flood rather than a single flood generation mechanism. The climatic mechanism includes persistent synoptic weather conditions favoring a very extreme event or episodes of moderate precipitation events resulting in a multiple day extreme event or extreme positive temperature anomaly causing a quick melt of snow in the catchment (Gaál et al., 2012; Nied et al., 2014; Prudhomme and Genevier, 2011). The hydrological processes such as antecedent soil moisture conditions, snow and ice storage in the catchment, rain on snow mechanisms and antecedent ground water level play an important role in defining the magnitude and length of the flood wave (Merz and Blöschl, 2003, 2008). Further, the superimposition of flood waves from different tributaries of the river also contributes towards the increased length and magnitude of the flood wave. Moreover, coincidence of any of the extremes from the climatic and hydrological processes results in amplification of the flooding magnitude and extent.*

P 8, L 8: Looking at Figure 4, the half of the data was located in the range -1-+1 in case of HVB, which does not seem so broad a distribution. P 9, L 3–4: Again, the SPHY results are strongly affected by the underestimation of river discharge. I am not sure whether the use of such a poorly biased model can provide meaningful indications.

We acknowledge the reviewer's concern regarding the large biases in reproducing the absolute extremes in the SPHY model. However, after recalibration the results have changed strongly as indicated above.

Since, the reviewers have some concern over the SPHY model results, we decided to couple an advanced kinematic routing scheme to the model which improves the representation of the flood wave characteristics, and allows a better quantification of the role of model uncertainty. We use the PCR-GLOWB2 kinematic wave routing scheme (Sutanudjaja et al., 2018). The higher quantiles flow has significantly improved as compared the SPHY with simple routing scheme (Figure R1 and Table 1.)

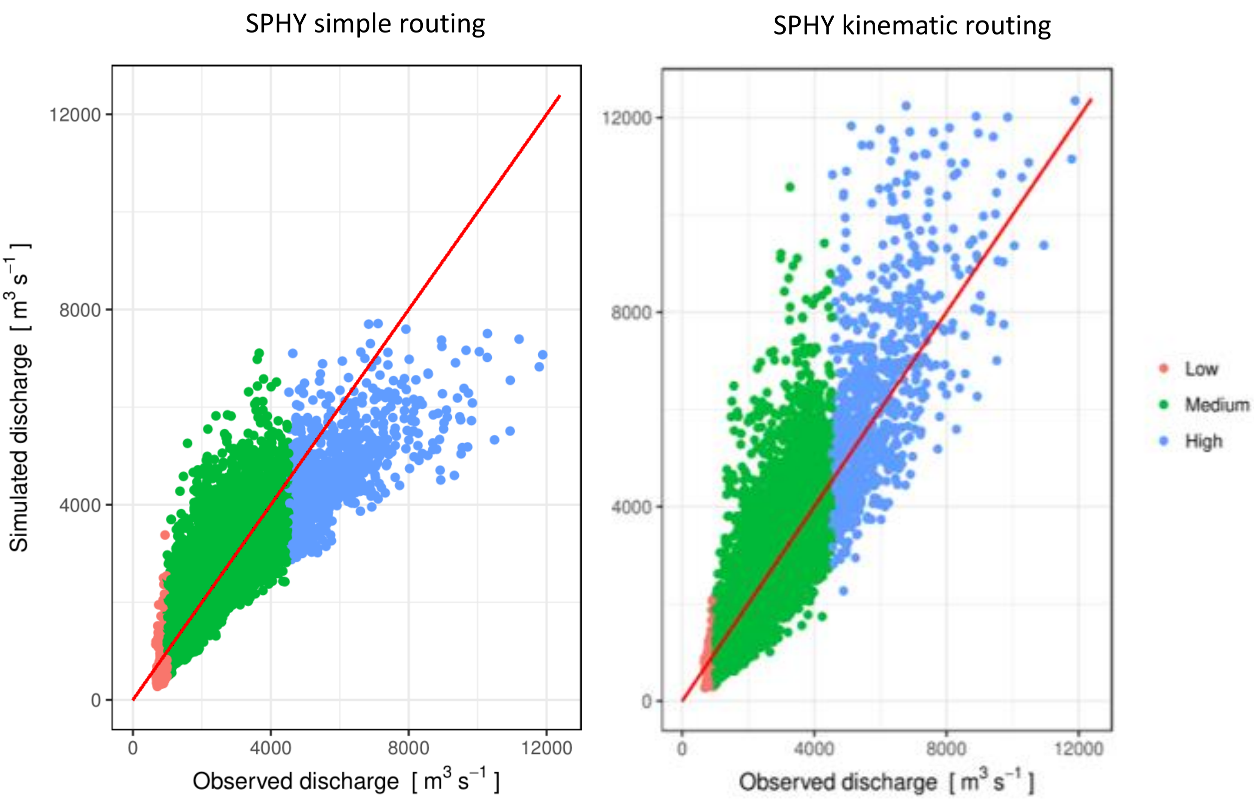

Figure R1. Observed versus SPHY modeled daily discharge at Lobith for the period between 1951 to 2000 for (left) the original simple routing scheme, and (right) with a kinematic routing scheme from PCR-GLOWB-2. Colors indicate three ranges based on observed percentiles: "Low" (<5%, red), "Medium" (5-95%, green) and "High" (>95%). The solid red line represents the 1:1 slope.

Table 1. Performance index for HBV and SPHY model on a daily time scale. The low, med and high represent the statistics for Q<Q5th, Q5th < Q < Q95th and Q > Q95th quantile of the observed flow whereas, all, represents the overall flow series.

	HBV				SPHY (Kinematic routing)			
Objective function	**low**	**Med**	**high**	**all**	**low**	**med**	**high**	**all**
R2	0.52	0.87	0.65	0.91	0.19	0.65	0.34	0.77
PBIAS (%)	-18.3	-10.6	-7.3	-10.3	-20	-0.1	6.7	0.3
RMSE(m3/s)	180	359	1045	415	300	605	1732	695
NSE	-4.9	0.79	0.26	0.87	-5.26	0.59	0.2	0.69
Volumetric Efficiency	0.82	0.87	0.85	0.87	0.72	0.81	0.79	0.81

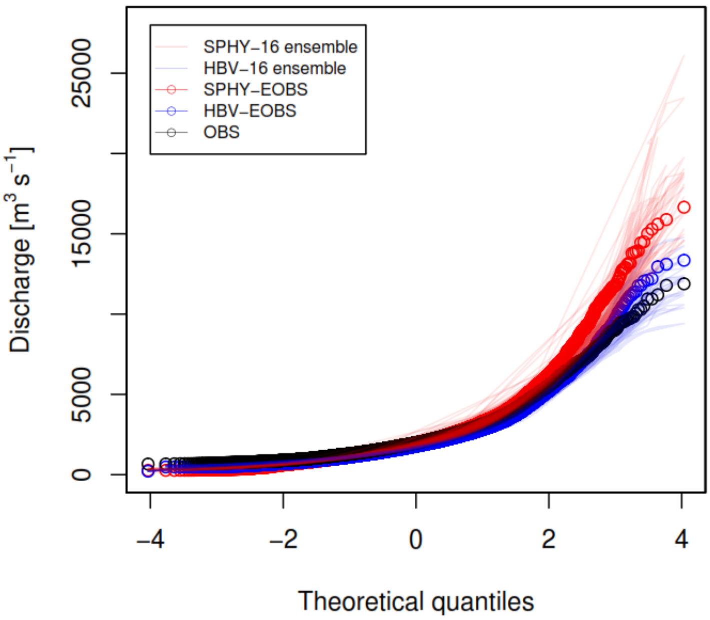

Figure R2: Normal quantile plot for HBV (blue), SPHY (red) and Observation (black). On the horizontal axis, the distributions are centered and scaled (divided by the standard deviation). The light blue & red lines represent 16 ensemble members for HBV & SPHY.

With the improved simulation of extreme flows, more well-defined flood peaks are observed in the model simulation. Rather than comparing the onset of discharge peaks, we directly compare the flood wave peak timing. We found that with the new routing scheme the flood wave travel time has significantly improved in SPHY, even outperforming the HBV model (Figure R3).

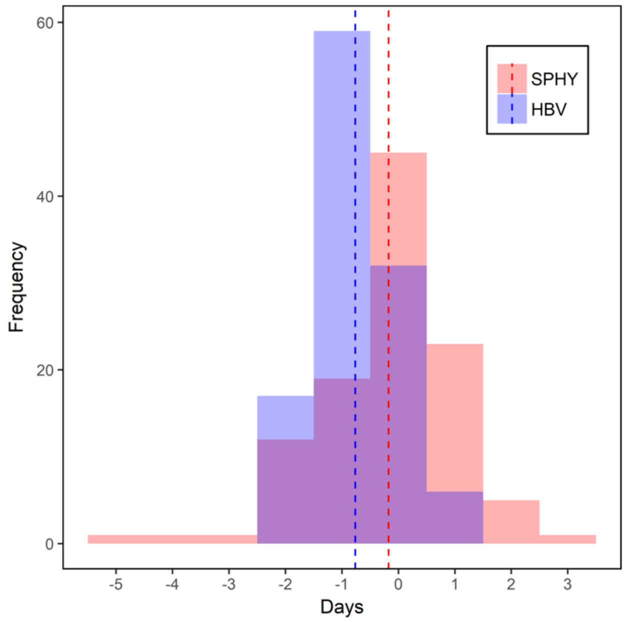

Figure R3. Comparison of the distribution of the timing of the discharge wave peak in SPHY with the observations for 50 years.

With the model routing improvement, the surge and discharge composite plots generated using the climate model ensemble are also improved, as shown in Figure R4 This suggest the underestimation of the extreme flows in SPHY are mainly due to routing scheme used and not due to poor calibration of the physical processes. With improved timings of the flood wave, the model uncertainty can be reduced further. A clear dependence at higher quantile for the range of lags can be seen in SPHY model with kinematic routing which was not evident in the SPHY model with simple routing. Improvement in the model timing does not change our previous claim that probability for finding a co-occurrence of extreme river discharge at Lobith and storm surge conditions at Hoek van Holland are up to four times higher (Figure R6). Though there are some minor differences in the figures, the main conclusion remains unchanged.

Based on this, we change section 3.1.1 Basic metrics and distribution, 3.1.3 Timing of the peak, 5. Discussion, table 1, and all the figures 1-8 accordingly in the main manuscript.

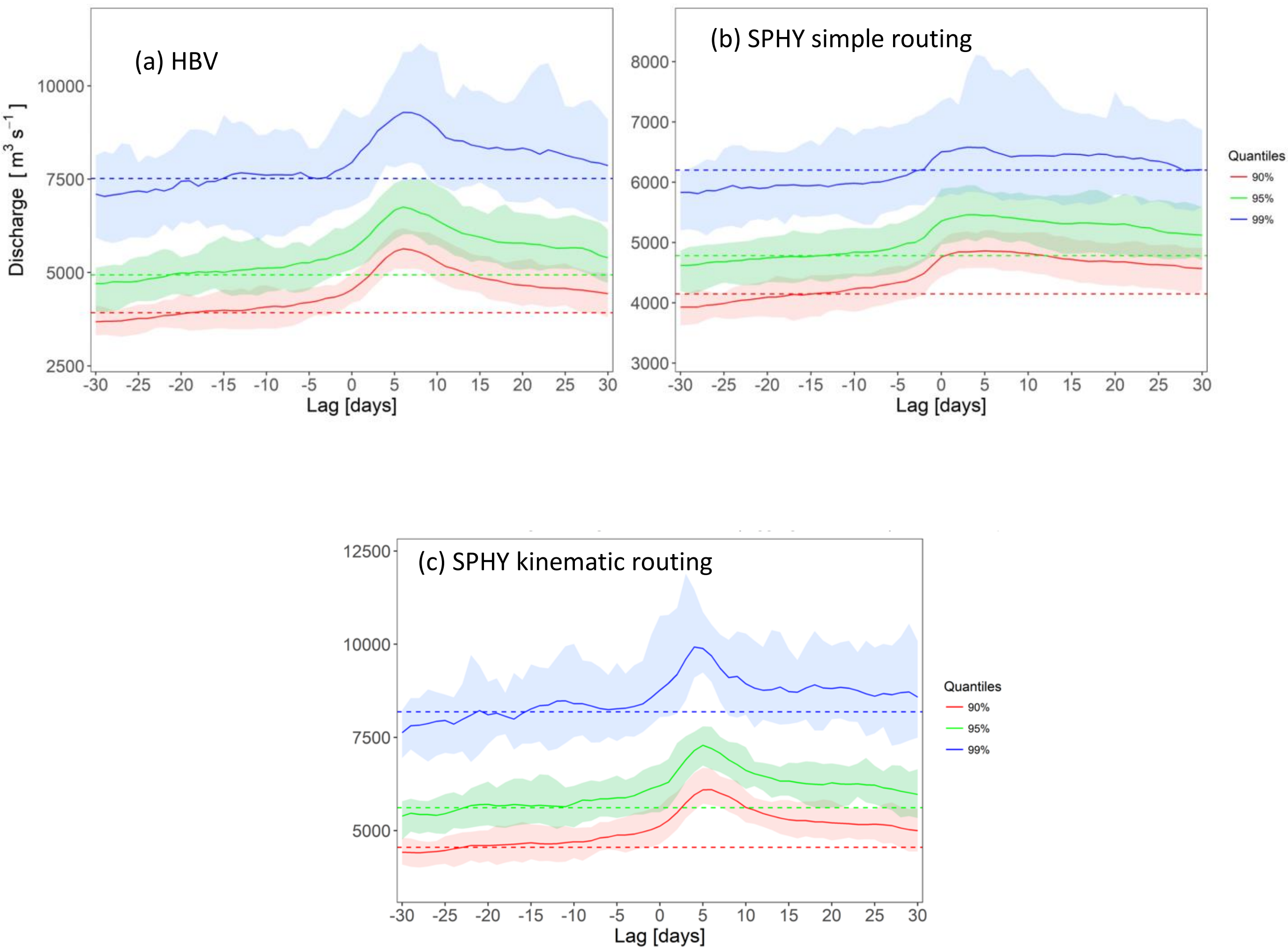

Figure R4. Mean temporal evolution of the 90th (red), 95th (green) and 99th (blue) quantile of discharge at Lobith for total water level events exceeding the 90th percentile at HvH in WAQUA as modeled by (a) HBV, (b) SPHY with simple routing and (c) SPHY with kinematic routing. The lag in discharge at Lobith is relative to the peak in total sea water level at HvH. Negative & positive lag days indicate that discharge peak occurs before & after the day of the high sea water event, respectively. The dashed lines are the unconditional discharge quantiles, i.e. discharge quantiles independent of water level; solid lines are the ensemble mean of the conditional quantiles. The shaded area represents 16 different lines for each ensemble and we only took the 5th and 95th percentile of those 16 lines to show spread of 16 ensemble member.

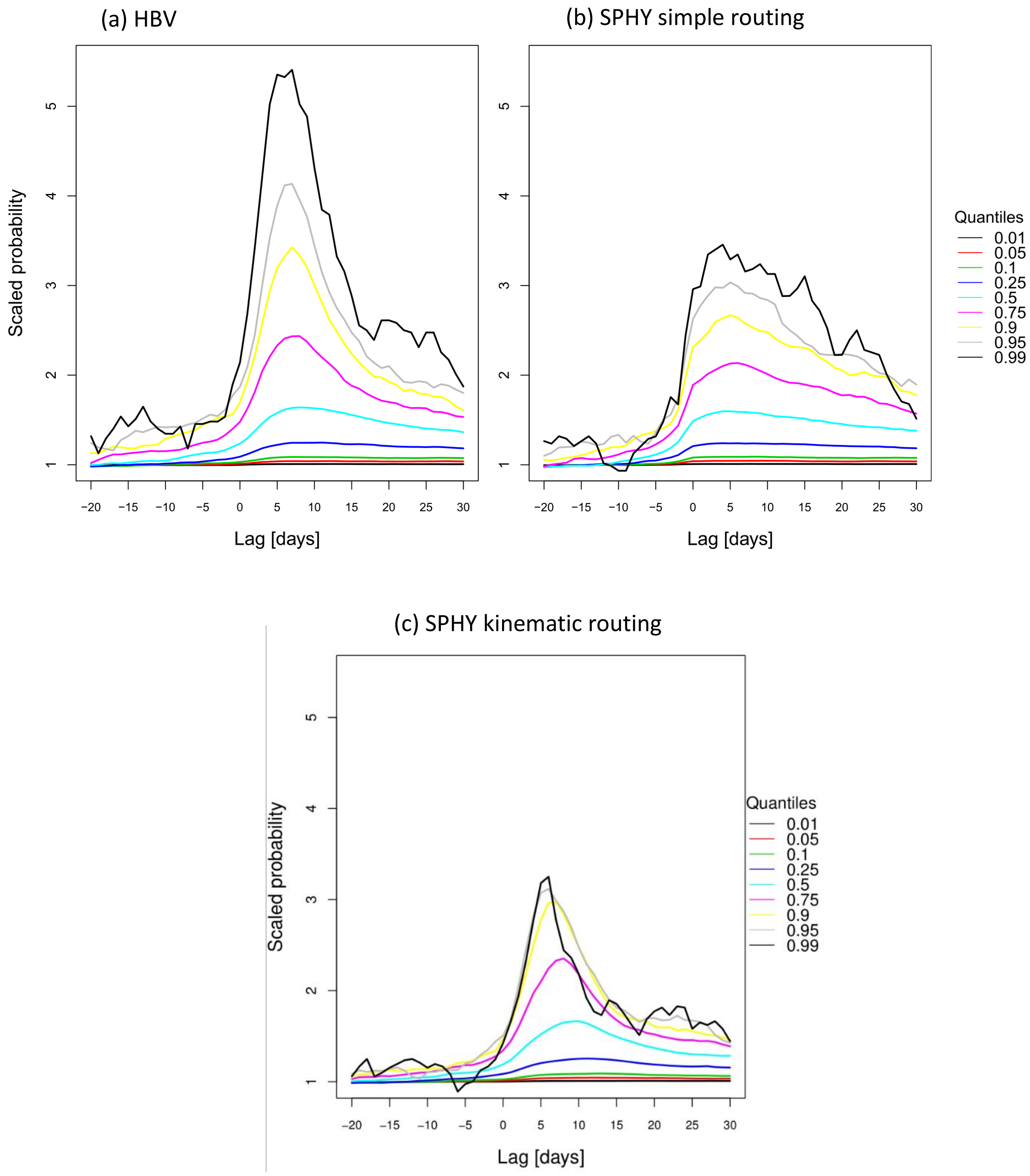

Figure R5. Exceedance probability of river discharge above indicated percentile, conditioned on the 97.5% exceedance of total water level for (a) HBV, (b) SPHY with simple routing and (c) SPHY with kinematic routing. For each discharge percentile, the probability is scaled by the random probability of the event.

P 9, L 12–13: The word ‘tail’ is duplicated in the sentence ‘analyzing the tails of the tail of distributions’.

We rephrased this as follows:

P 9, L 12-13 *‘analyzing the extremes in the tails of the distributions’*

P 9, L 29: What does the width of the bands represent?

The width of the band, for instance the grey, represents the conditional discharge distribution i.e. the discharge distribution between the 50th and 90th quantile only for the cases where the total water level is above 90th quantile. Similarly, the blue color band represents the conditional discharge distribution i.e. the discharge distribution between the 50th and 90th quantile only for the cases where the total water level is above 50th quantile.

The caption of Figure 6. P22 L7 in the original manuscript will be rephrased as:

“The width of the band, for instance the grey, represents the conditional discharge distribution i.e. the discharge distribution between the 50th and 90th quantile only for the cases where the total water level is above 90th quantile. Similarly, the blue color band represents the conditional discharge distribution i.e. the discharge distribution between the 50th and 90th quantile only for the cases where the total water level is above 50th quantile.”

P 10, L 13–15: This analysis is interesting, but could you add some literature to support your reasoning about the hydrological characteristics of the target basin?

We added references to relevant literature and rephrased a part of which will be included in a “Study area” section which is not included in the original manuscript.

:

The Rhine basin covers an area of 185,000 km^2 and runs over 1320 km from its source in the Alps to the North Sea. The streamflow in Rhine is mainly dominated by snowmelt and rainfall-runoff from the Alps during summer for the upper part of Rhine (Viviroli, Daniel; Messerli, 2003). However, for lower parts at Lobith, streamflow is mostly dominated by rainfall resulting in streamflow peaks during winter. A peak-shift in the average annual hydrograph can be observed from summer to winter from the upper Rhine at Basel (50 Km downstream of Untersiggenthal as shown in Figure R6) and lower Rhine at Lobith (Engel, 2001; Photiadou et al., 2011). The snowmelt contribution to the streamflow at Lobith is significant and total annual contribution to streamflow is around 30 percent (Stahl et al., 2016). Area upstream of Basel produces almost 50% of the discharge despite only covering around 20% of the total area of the Rhine catchment (Kwadijk and Deursen, 1999). The flood wave travel time from Basel to Lobith is around 5 days (Hegnauer et al., 2014). Further, the slow melt from snow and glaciers would require additional day or two to reach the Basel.

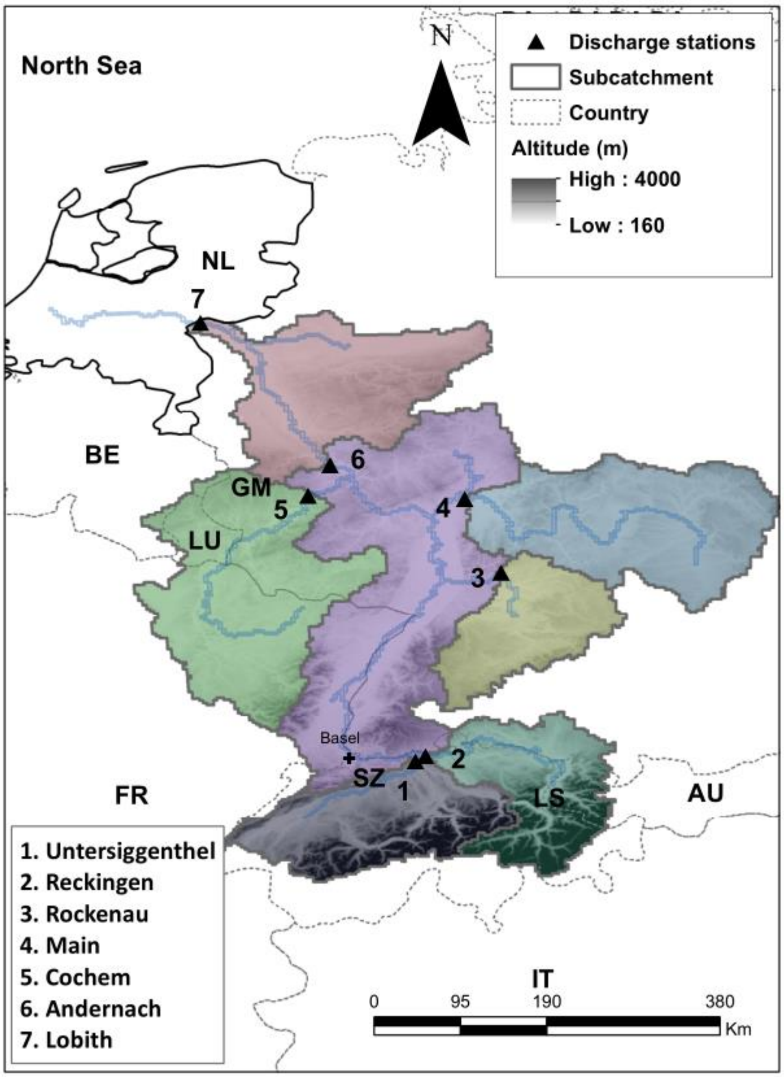

Figure R6: The Rhine basin, with seven sub catchments used for the calibration in the study.

P 10, L 26: 'in which the physical...' maybe 'physics' not 'physical'?

We changed this to 'physics' instead of 'physical'.

P 12, L 18: 'still it's not...' The abbreviation should be avoided.

We agree with the reviewer and 'still it's' would be rephrased as *''still it is''* in the sentence.

P 18, Figure 1 and after: Add the model names to each sub plot.

We added the model names to each sub plot.

P 19, Figure 2: Maybe better to use 'and' instead of '&'.

We adopted the suggestion in P19 Figure 2 caption.

P 19, Figure 3: What do the dotted lines in the right figure represent? Mean values?

The dotted line represents the median values (50^{th} quantile) as mentioned in the caption of Figure 3 as 'The dash vertical lines show the mean of the flood waves for SPHY (red), OBS (black) and HBV (blue)'. The mean will be replaced by median and term "**HBV**" is duplicated twice instead one of them should be “OBS”. We also noticed that the black vertical dash line is missing from the figure and this will be added.

P 22, Figure 6: In the scatter plots of the left figure, the blue dots represent events exceeding 99th quantile, but on the other hand, the black and blue lines in the left and right figures represent 95th and 50th quantiles, so the large/small relationship between gray and blue colors is inverse within the same figure. This is very confusing! In the right figures, what the triangular and the bands represent?

The colors of the figures do not correspond in Figure 6 (a) and 6 (b) and we agree that this leads to confusion in perceiving the figures. Therefore, we change the coloring in the revised manuscript as displayed below:

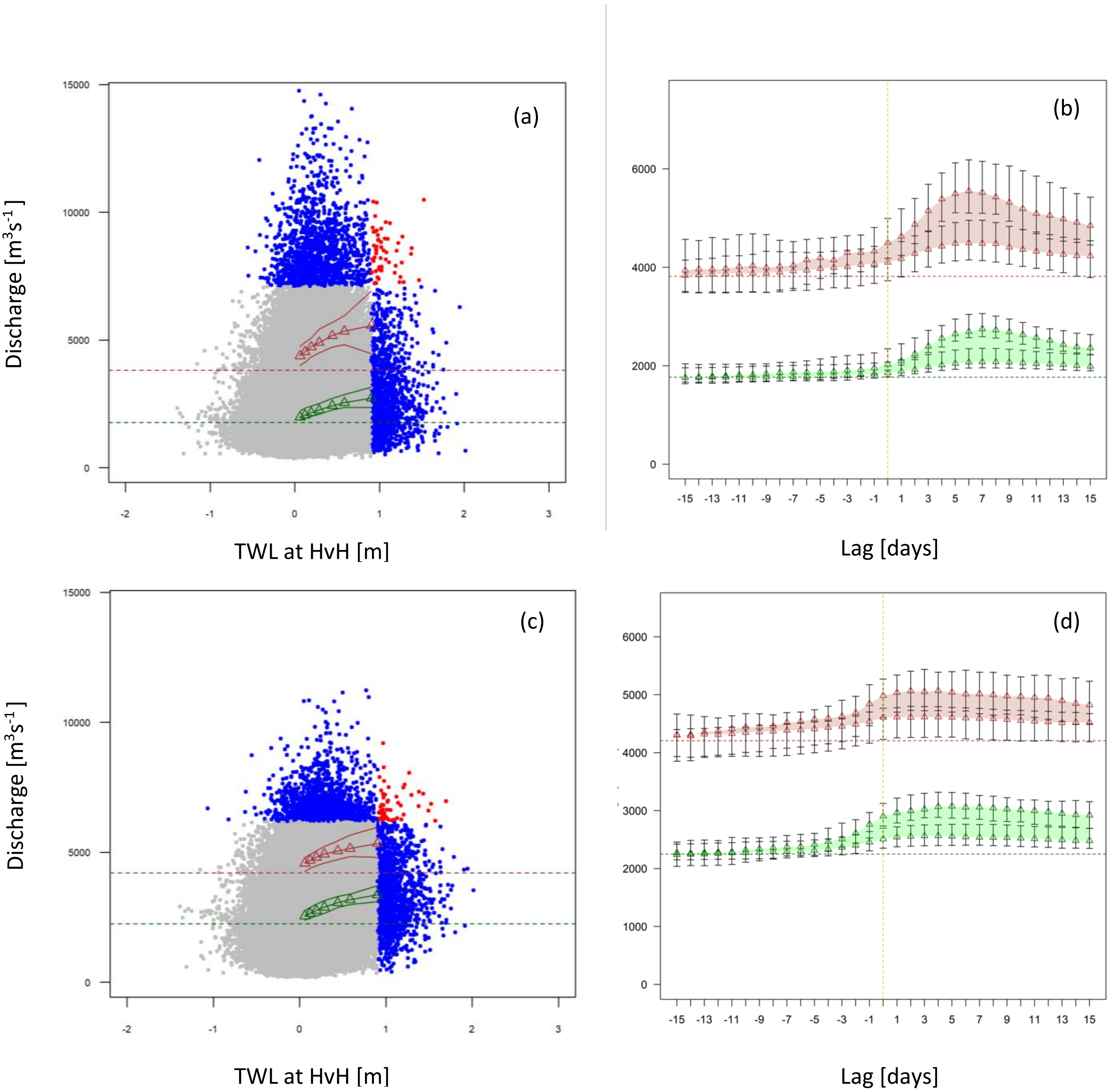

Figure 6: Left: Scatter plot of coastal water levels and discharge for a lag of three days ((a) HBV and (c) SPHY) and 16 ensemble members. Events exceeding the 99^{th} quantile of either of the variables are marked in blue. Events exceeding the 99^{th} quantile of both variables are marked in red. The triangles (green/brown) represent the ensemble mean of the conditional discharge (50^{th} and 90^{th}). The green solid lines represent the spread of ensemble i.e. 5^{th} and 95^{th} quantiles of the conditional discharge (50^{th} quantile). Similarly, brown solid lines represent the 5^{th} and 95^{th} quantiles of the conditional discharge (90^{th} quantile). Right: Conditional discharge plot for 50^{th} and 95^{th} quantile of surge ((b) HBV and (d) SPHY). The green band, represents the conditional discharge distribution i.e. the discharge distribution between the 50^{th} and 90^{th} quantile only for the cases where the total water level is above 50^{th} quantile. Similarly, the brown band represents the conditional discharge distribution i.e. the discharge distribution between the 50^{th} and 90^{th} quantile only for the cases where the total water level is above 90^{th} quantile. The upper and lower green triangle represent the mean of the 50^{th} and 90^{th} quantile of discharge conditioned on 50th quantile of TWL. Similarly, the upper and lower brown triangle represent the mean of the 50th and 90th quantile of discharge conditioned on 90th quantile of TWL. The yellow dash line indicates the time lag zero. The error bar on triangle represent the confidence interval estimates of the mean (5th and 95th quantiles) from 16 ensemble members

The triangle on the right figure (b and d) represent the mean of conditional discharge on TWL. For instance, the upper and lower brown triangles in figure 6(b) represent the part of the conditional discharge distribution (i.e. region between 50^{th} and 90^{th} quantile) conditioned on 90^{th} quantile of TWL. The band is plotted just to distinguish the area between the upper and lower triangle. Similarly, the error bar on triangle represent the confidence interval estimates of the mean (5^{th} and 95^{th} quantiles) from 16 ensemble members.

References

Engel, M. D. and H.: Flood events in the rhine basin: genesis, influences and mitigation, Nat. Hazards, 23(2–3), 271–290, doi:10.1023/A, 2001.

Gaál, L., Szolgay, J., Kohnová, S., Parajka, J., Merz, R., Viglione, A. and Blöschl, G.: Flood timescales: Understanding the interplay of climate and catchment processes through comparative hydrology, Water Resour. Res., 48(4), 1–21, doi:10.1029/2011WR011509, 2012.

Hegnauer, M., Beersma, J. J., van den Boogaard, H. F. P., Buishand, T. A. and Passchier, R. H.: Generator of Rainfall and Discharge Extremes (GRADE) for the Rhine and Meuse basins. Final report of GRADE 2.0, , 84, 2014.

Kwadijk, J. and Deursen, W. Van: Internationale Kommission für die Hydrologie des Rheingebietes Commission internationale de l'Hydrologie du bassin du Rhin Development and testing of a GIS based water balance model for the Rhine drainage basin., 1999.

Merz, R. and Blöschl, G.: A process typology of regional floods, Water Resour. Res., 39(12), 1–20, doi:10.1029/2002WR001952, 2003.

Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, Water Resour. Res., 44(8), 1–17, doi:10.1029/2007WR006744, 2008.

Nied, M., Pardowitz, T., Nissen, K., Ulbrich, U., Hundecha, Y. and Merz, B.: On the relationship between hydro-meteorological patterns and flood types, J. Hydrol., 519(PD), 3249–3262, doi:10.1016/j.jhydrol.2014.09.089, 2014.

Photiadou, C. S., Weerts, A. H. and M. Van Den Hurk, B. J. J.: Evaluation of two precipitation data sets for the Rhine River using streamflow simulations, Hydrol. Earth Syst. Sci., 15(11), 3355–3366, doi:10.5194/hess-15-3355-2011, 2011.

Prudhomme, C. and Genevier, M.: Can atmospheric circulation be linked to flooding in Europe?, Hydrol. Process., 25(7), 1180–1190, doi:10.1002/hyp.7879, 2011.

Stahl, K., Weiler, M., Kohn, I., Freudiger, D., Seibert, J., Vis, M. and Gerlinger, K.: The snow and glacier melt components of streamflow of the river Rhine and its tributaries considering the influence of climate change, , (I), 146 [online] Available from: www.chr-khr.org/en/publications, 2016.

Sutanudjaja, E. H., Van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., Van Der Ent, R. J., De Graaf, I. E. M., Hoch, J. M., De Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D. and Bierkens, M. F. P.: PCR-GLOBWB 2: A 5 arcmin global hydrological and water resources model, Geosci. Model Dev., 11(6), 2429–2453, doi:10.5194/gmd-11-2429-2018, 2018.

Viviroli, Daniel; Messerli, B.: Assessing the Hydrological Significance of the World's Mountains, Mt. Res. Dev., 23(4), 369–375, doi:10.1659/0276-4741(2003)023, 2003.