Potsdam, October 31, 2014

Dear Prof. Ralf Merz,

Thank you for your comments and decision regarding our submitted article. Together with this letter, we have uploaded a revised version of our manuscript.

We completely agree that the issue of hydrological model parameter stability under non-stationary conditions needs further discussion. We therefore have added a paragraph to chapter 4.5. where we discuss the problem in detail and also sketch out opportunities for identifying parameter sets which are suitable for simulations under future conditions.

We have also clarified the intention of this paper, which is to consider projected seasonal changes and the uncertainties underlying those projections in a sample of catchments which already show some tendency for a mixed flood generation regime. The intention has not been to develop a regional analysis which covers all possible flood regimes and climatic conditions in Norway, as the likely changes in other areas and types of flood regimes are much better understood, based on previous investigation. In the revised manuscript we have also added a section in the conclusions in which we point out two opportunities for achieving more regional conclusions.

At the end of this letter, please find our point-by-point response to the comments of the three reviewers.

Please also note that we have slightly changed the title of our manuscript to stress the focus of the paper: "Climate change impacts on the seasonality and generation processes of floods – projections and uncertainties for catchments with mixed snowmelt/rainfall regimes". We hope that you will agree with our proposed change.

On behalf of all co-authors,

Klaus Vormoor

Point-by-point response to referee comments and modifications to the manuscript

We wish to thank all three referees again for providing valuable suggestions that have helped to improve the quality, clarity and visual appearance of our manuscript. In the following, we address each of your points and delineate the changes we have made in response to these points in our revised manuscript. For convenience, the referee comments are repeated in *gray-italic*, our answers are written in black. Where we refer to our Author's Comments given in the interactive discussion, we use the abbreviation AC.

Referee #1

SPECIFIC COMMENTS:

The one major point where more insight would have been desirable is the model's predictive skill under changed conditions – e. g. by applying a differential split-sample test (see Andréassian et al., 2011; Klemeš, 1986; Refsgaard and Henriksen, 2004) –, although the problem of time (in)stability of model parameters (Merz et al., 2011) is mentioned, and the uncertainties of the hydrological parameter sets are analysed. If not by extending the study with specific modelling experiments for the reference period 1961–1990, the topic should at least be addressed with a brief discussion.

We agree that this issue needs further discussion. Therefore, we have added a paragraph to the end of section 4.5., which discusses the problem of transferring hydrological model parameters under non-stationary conditions and we sketch out opportunities for identifying parameter sets which are most suitable for simulations under a future climate. The discussion of this issue is further elaborated as suggested in our AC.

With reference to Gudmundsson et al. (2012), the authors consider non-parametric methods as most suitable for bias correction of precipitation (P6281, L26 ff.). In the following, they use these methods also for temperature correction – can they state something about the respective suitability?

We tested the performance of three different versions of quantile mapping and found non-parametric methods to also perform best for temperature correction. We have added a note to our revised manuscript as given in our AC.

The approach to flood generating processes (FGP) is rather straightforward, but appropriate. Still, I suggest putting the approach into the context of more detailed methods, in particular the one described by Merz and Blöschl (2003).

We agree and have included this in our revised manuscript.

Since the extraction of extreme events is based on a Peak-Over-Threshold (POT) approach, did the authors consider POT-specific skill scores (e. g., Lamb, 1999; Viviroli et al., 2009) to evaluate their model?

Referring to our more detailed response given in our AC, we refrain from introducing and discussing these skill scores in our revised manuscript.

How was the catchment-specific normal flood duration (P6284, L 12) determined, i.e. how did the authors define beginning and end of a flood event? Does "normal" flood duration refer to "average" flood duration over all POT events sampled?

We have extended that particular paragraph in section 3.5. to clarify the method we used to determine the "normal flood duration".

I recommend adding a note on the recent study by Köplin et al. (2014) which treats a very similar topic.

We now refer to that study in our revised discussion in chapter 4.4.

MINOR COMMENTS:

P6274, Abstract: *I suggest adding the number of catchments studied and the daily time-step used.* Considered.

P6274, L12: *...in flood regimes *result*...* Thank you.

P6276, L25: *Readability: ... related to changes in the magnitude vs. *changes* in the frequency of events?* Accepted

P6278, L16: *Mention the time period also in the main text, not only in the Figure.* Considered.

P6278, L24: *The main text discusses *mean* elevation, Table 1 however lists *median* elevation.* Thank you. We corrected that.

P6280, L15: *Maybe mention here already why the two time periods are almost (but not completely) identical.*

There is, to our knowledge, no particular reason why some of the RCMs of ENSEMBLES are only run up to 2099 instead of 2100. To be consistent in our study, we therefore applied the period 2071-2099 for all RCMs considered (as already stated in the same section).

P6288, L07: *... least *pronounced*...* Thank you.

P6304, *Figure 2: (1) The ordinate's point of origin is not 0. This is perfectly OK, but I would mention it (either in the main text or in the caption) as it makes the differences between the various series appear larger. (2)My interpretation of the NSEw value indicated here is that refers to series (i), and that it refers to the entire series (and not to the POT values which constitute the main content of the Figure). Consider clarifying this.*

Both points are clarified in a revised figure caption for Figure 2. Please also note that we updated the NSE_w value for Kråkfoss (from 0.77 to 0.87) since this was wrongly reported in our discussion paper.

P6306, Figure 4: *The vertical gridlines could be improved to aid the figure's interpretation, i. e. for easier comparison of the number of events within each group (box width). Also, since visually interpreting box width via square-root as number of events is not straightforward, I suggest adding* *a scale for box width.* Ibid.: *Point of origin for ordinate not 0: see above.* ibid.: *How was the maximum value of the ordinate determined? It does not seem to correspond to the maximum values displayed (i. e., the whiskers).*

We have modified Figure 4 and its figure caption according to our suggestion in our AC.

P6307, Figure 5: *Like for Figure 4, I suggest adding a scale for linking the pie diameter to the total number of events, as this is visually not straightforward.*

We agree that Figure 5 was not completely straightforward and have modified the Figure according to our suggestion in our AC.

Referee #2

One important point that is not stressed in the paper, but somewhat relevant for the whole study, is the fact that you assume stationary conditions for the whole reference period for which you calibrate your hydrological model. …would be interesting to see whether changes in seasonality can already be observed in the measured data. If so, this may influence your modelling results for future time periods as model parameters may not be stationary for the reference period

We agree and have added a paragraph to section 4.5. with a detailed discussion on that issue. Unfortunately, changes in seasonality cannot be discerned for all catchments in the observed data.

Page 6280, lines 16-17: *Why did you choose the 2071-2099 as future period? Did you also have a look at changes in nearer future?* We have added a comment on this as pointed out in our AC.

Page 6284, line11-12: *How did you determine the "normal flood duration" for the catchments?* We extended this paragraph in section 3.5. to clarify the method we used to determine of the "normal flood duration".

Page 6286, lines 15-24 and Figure 2: *In the Figure you specify that for certain simulations you only apply one best-fit HBV parameter set? Why? And why is this not discussed in the related text?* We have added a sentence to p.6288, line 17 as suggested in our AC.

Page 6287, line 5: *Can you specify any reason for the lower performance? Are certain processes not represented well with the model?* We have added a short note to chapter 4.1. that this due to a systematic underestimation of flood peaks.

Page 6287, lines 7-14 and Page 6288, lines 8-9: *The performance of the HBV model regarding POT events for the validation period (Figure 2) shows a low performance for the catchments Fustvatn and Junkerdalselv which is mainly relevant for the assessment of changes in flood magnitudes for the future period, while regarding the representation for flood seasonalities the performance of the model for the Krakfoss catchment is rather low (Figure 3) which is important for the assessment of changes in flood seasonalities (You could stress this aspect more clearly in the paper). Can you comment on why e.g. model performance in the Krakfoss catchment is low regarding seasonality and high regarding flood magnitudes?*

We have made the changes in chapter 4.2. and 4.3. as suggested in our AC.

Figure 2: *Add the info on what the NSE_w is in the Figure caption.* Considered. We modified the figure caption and updated the NSE_w value for Kråkfoss since this was wrongly reported in the first version.

Figure 4: *Are the changes shown here all significant or not?* The notches on the boxplots indicate significant changes (with 95 % confidence). We have added this to section 4.3.

Referee #3

GENERAL REMARKS:

It is well known that there will be a shift in flood seasonality due to climate warming, from snow dominating floods to more rain controlling floods in regions with a seasonal snow cover and accordingly, a change in controlling processes (e.g. J. Parajka, 2010). However, as also mentioned in the paper, precipitation is projected to increase in the region as already documented, particularly on the western coast. Thus, it is important to account also for changes in seasonal precipitation when discussing changes in flood seasonality. The paper briefly mentions this aspect, however, it is recommended that is also include a quantitative analysis of changing (seasonal) precipitation and temperature pattern to better distinguish the relative importance of increasing temperature versus changes in precipitation.

We have added a figure and a sub-chapter to the Results & Discussion-section with a quantitative analysis of the projected changes in seasonal precipitation sums and mean temperature after local adjustment for the six study catchments.

It would further have been of interest to assess any trends in the observed period and compare these with future predictions.

We have split the first paragraph of the introduction into two sections elaborating on (i) observed trends in (extreme) streamflow associated with observed changes in the meteorological triggers, and (ii) the projected changes in temperature and precipitation and their implications for the snow regime. The latter is closely connected to the revised discussion on the changes in the FGPs.

The use of the AR4 scenarios rather than the CMIP5, makes the study somewhat outdated (although the main conclusions may not change that much).

Downscaled RCM projections for CMIP5 covering the whole of Norway are now only becoming available (end of 2014) and have yet to be made available for general research purposes. Thus they were not available for the study presented here. As mentioned by the reviewer him-/ herself, it probably would not affect the main conclusions.

The use of only one (conceptual based and calibrated) model in (what is likely) a non-stationary climate should be commented on, and more general, the role of hydrological model uncertainty in climate change impact studies (e.g. Velázquez et al., 2013; Bosshard, et al., 2012).

We agree that it is necessary to comment on the issue of hydrological model parameter stability under non-stationary conditions. Please see the paragraph which we have added to chapter 4.5. where we discuss this in more detail. Regarding the use of only one hydrological model we have made the changes to manuscript as suggested in our AC.

The use of only six catchments and their location. It is noteworthy that the selection does not include a catchment in western Norway, which is specifically mentioned as an area of interest due to high precipitation rates (ref. Introduction). This is also a region where precipitation is projected to increase significantly in the future (and already has).

As we have now clarified in the revised manuscript, the purpose of this study was not to develop climate projections for all types of catchments represented in Norway, as general patterns for rainfall-dominated catchments as well as for snowmelt-dominated catchments are well established in previous work. The focus here is on catchments that already show a tendency towards a mixed regime in the current climate. Thus, we have not included a catchment from western or southern coastal Norway, as these are already dominated by autumn/winter rainfall flood regimes. In addition, we have avoided catchments with a significant glacier component, and have limited the catchment size to a reasonable range, such that hydrological modelling with a daily timestep is appropriate.

The topic of the study lends itself to a regional study and six catchments is a rather low number given the high hydroclimatic variability across Norway. Only with a better coverage can one conclude on regional patterns and trends in flood patterns (in the current as well as future climate), as these can vary considerable locally. This can be achieved either by increasing the number of catchments or by using a gridded dataset for Norway (e.g. data from seNorge.no, which contains both interpolated climate and simulated runoff based on a gridded version of HBV). The current study design is in my opinion not sufficient to conclude on regional patterns in flood seasonality (refer Objective 1). Accordingly (provided that the study is not extended), the conclusions must be revised to be more catchment specific and less general.

The sample of six study catchments is not sufficient to draw robust regional conclusions nor has this been the intention of the study as mentioned above. We have clarified this in the manuscript. Moreover, we have added a paragraph to the conclusions where we discuss the value of our results in more detail and where we point out two opportunities to gain more robust regional conclusions, if desired.

Objective 3 can only be answered if the role of changing precipitation and temperature patterns are included explicitly (ref. point 1 above).

Please see also our response to your first General Remark. We have included a quantitative analysis of the changes in precipitation and temperature. Note, however, that the FGPs are defined by runoff components, which are simulated by the HBV model. In fact, hydrological modeling is required to highlight the changing role of snow storage and its effects on flood generation. We have added a note on that in section 4.5.

When objective 4 is presented, we have not yet been informed about the different ensemble components. The latter aspect needs to be better introduced, including the design of the modelling strategy. Section 3.1. says what it consists of, but not why this particular design was chosen.

Perhaps it is partly what is said on p.6286, line 10: "identify the fractional uncertainty emerging from different sources within the model chain for three variables…".

In the revised manuscript, we introduce the different ensemble components before presenting objective 4. These changes refer to the suggestion made in our AC.

SPECIFIC COMMENTS

The introduction gives reference to various trend studies (in observations), but not to particular studies on trends in floods, which should be added.

We have modified the introduction according to our suggestion in the AC.

The result of the paper should be discussed in light of similar studies, and not be limited to national (or Nordic) studies. Also pan-European trend studies would be of interest as well as studies from similar regions in other continents (e.g. U.S. and Canada).

Accepted. In the revised manuscript, we refer to comparable findings from a pan-European study, and from studies for different regions in the Alps and North America. Please also note, that we only included studies in regions with similar processes and scales (AC)..

Reference should be made to existing regime classifications for Norway (here only two regimes classes are suggested). Other regime classifications distinguish more classes and could also be used as a starting point for selecting representative catchments.

We have added a section on the classification of hydrological regimes in the Nordic countries (Tollan 1975, Gottschalk et al. 1979) and discuss that in the context of our simplified discretization of flood regimes in section 2.1. as suggested in our AC.

Clarification on the seasonality index, S_D:

i. *Why is the second term in the index included (does it add any information)?*

Please see our AC. We have not modified our manuscript referring to this point.

ii. *The first term describes the ratio between the flood peaks in m-3s-1; does this mean that you sum the POT discharge values?*

Yes, we have clarified this in the revised manuscript.

iii. *Is it valid to use the same two seasons for all catchments given their high variability in hydroclimatic regime (and will they be representative in the future)?*

The variability in seasons is supposed to be larger for the current (reference) climate than in the projected future climate. Since the classification holds for the current situation it is likely that this will also hold for future conditions.

iv. *How will the use of a fixed threshold (here the 98.5 streamflow percentile) influence the selection of events if there is a change in annual precipitation (and thus streamflow) in the future?*

This was not made clear enough in the first version. We have clarified this in the revised manuscript.

v. *How is the normal flood duration defined? Is there a different value for snow generated events as compared to rainfall (different response times)?*

We have extended the paragraph in section 3.5. in order to clarify the method we have used to determine the "normal flood duration" as suggested in our AC. We have not used different methods for snowmelt- vs. rainfall-generated events. The 'normal flood duration' refers to the maximum temporal extent of a flood in a specific catchment independent of its generation process.

vi. *Present and argue for your proposed seasonality index in light of existing definitions (e.g. J. Parajka, 2010).*

Please see our detailed response in the AC. Again, our apologies, that we did not make clear in the earlier version that we actually apply the same seasonality for the analysis of changes in the FGPs. We have significantly modified the method section and the discussion of the corresponding results as suggested in our AC.

The classification into three flood generation types is based on the contribution of rain and snow to the runoff. What about rain on snow events; how would these be classified based on the HBV model simulations?

The classification of the FGPs is defined as the percentage contribution of rainfall vs. snowmelt to runoff components during the flood event. Events are classified as 'snowmelt', 'rainfall', and 'rainfall+snowmelt' events. Correspondingly, rain on snow events are implicitly included in this definition. No distinction is made regarding a possible spatial differentiation in FGPs within a given catchment, i.e. if one has snowmelt in the higher reaches of a catchment without rainfall and rainfall in the lower reaches. The moderate size of the catchments considered supports the use of this 'lumped' approach for classifying FGPs.

Combining the result and discussion section can be challenging. Here, the results are discussed under specific headings, which is fine. However, this requires an overall discussion bridging between the different sections (option to add such a section at the end of the combined section).

As stated in our AC, we would prefer to hold our proposed narrative style in a revised manuscript, such that we progress from the detection of changes in flood seasonality to the reasons for these changes and finally to its underlying causes. However, in our revised discussion, of Figure 6 (now Figure 7) we could explain the changes in the previous sections and connect them to one picture.

It is concluded that the relative role of hydrological parameter uncertainty is highest in catchments showing a high change in flood seasonality. Is this not just a result of high model sensitivity to the threshold temperature (snow/rain and melt/no melt), implying a widely different response in runoff to small changes in temperature?

The HBV model parameters which tend to be most sensitive are not the parameters that are directly related to the snow routine of the model (TS – temperature for no melt; TX – temperature for snow, ice) but rather the precipitation and snow correction factors (PKORR, SKORR) (Lawrence and Haddeland, 2011). Moreover, the model is trained in a period with regular snowmelt seasons. All snow related processes should therefore be calibrated well in the relevant catchments.

The abstract needs to better represent details of the study, e.g. number of catchments, multi-model in what sense, what are the ensemble components?

We extended the abstract with this information.

The abstract reads "Changes towards more dominant autumn/winter events correspond to an increasing relevance of rainfall as a flood generating process (FGP) which is most pronounced in those catchments with the largest shifts in flood seasonality. Here, rainfall replaces snowmelt as the dominant FGP". Later it is stated (Section 4.4) "Rainfall becomes the dominant FGP in the future period in all investigated catchments". There is here a need to distinguish the relative contribution of a precipitation increase (rain or snow) vs. a shift in precipitation from snow to rain due to a temperature increase. In other words; what is the role of increasing temperature vs. changes in precipitation patters for the different catchments (should be evaluated on a seasonal basis). Ref. point 1 under General comments.

We argue that we consider the relative role of increasing temperature vs. changes in precipitation patterns through the use of the hydrological model. The results shown in Figure 5 and Figure 6 can be directly linked to projected changes in the temperature and precipitation regime in Norway, which have already been investigated by other authors (e.g. Hanssen-Bauer et al. 2003, 2009; Engen-Skaugen et al. 2007). We have pointed out this linkage more clearly in the discussion of our revised manuscript.

An important observation, although a bit hidden, is given in Section 4.4, p.6290, line 21: "the rainfall-generated POT events tend to occur later in the year". This should be further elaborated and possible reasons discussed.

We have revised that part of section 4.4. and have combined this with a better introduction of the methods and the results shown in Figure 6.

It is argued that the selection of only two classes is chosen to obtain a broad picture of flood seasonality. Why not simply look at changes in the flow regimes, i.e. changes in the month of the highest peak? This would allow you to analyse a more general shift in flood occurrence, not restricted by the choice of a fixed season (temporal as well as spatially).

We were looking for a simple classification of dominant flood seasonality for catchments with mixed snowmelt-rainfall regimes. S_{D} is intuitive (since the two seasons are associated with dominant flood generating processes), it is easy to apply (as well as transferable to other geographical areas, if others wish to apply it), and we therefore believe it is well suited this study. Moreover, we show more general shifts in flood occurrences in the discussion of Figure 6 (section 4.4), where we illustrate the change in the mean annual timing of floods separated by the FGPs.

It is mentioned that the HBV snow and melting module has a semidistributed structure. More details are here needed as the formulation of the snow routine is vital for the study, e.g. what is the spatial resolution of the elevation zones, how is the climate input interpolated to different elevation zones, how is snow melt calculated?

Considered. We give more details on the structure and input data of/to the 'Nordic' version of the HBV model in a revised manuscript.

Is the RCM downscaled to the scale of the catchment area or to a gridded structure? and how is the climate input distributed to the different elevation zones? More details needed.

We give more details on the input data and on the HBV model with respect to the hydroclimatic inputs to the elevation zones as suggested in our AC.

The reference to 'equifinality' should be deleted as I cannot see that the work specifically addresses this aspect; instead focus should be on parameter uncertainty only.

We have deleted 'equifinality'.

The last paragraph of Section 3.5.2 is not clear. What is the 'flood duration time of the core event' and what implication does it have that the duration is extended by adding 'the catchment specific recession time'.

We agree that this needs further explanation and have extended the paragraph as suggested in our AC.

Section 4.3 is important, but the approach (changes in magnitudes vs. the frequency of events) has not been well introduced in the Method section.

We agree and have added a note on that to section 3.5.1. referring to our AC.

Figure 2: comment also on the spread, not only on the median.

Accepted. We have made the suggested extension of that paragraph as given in our AC.

Figure 4: add the observations to the seasonal plot.

Please note that the distributions of the observed POT events (though not divided by seasons) are already given in Figure 2. We give the median of the observed POT events in Figure 4 (green bars). We think that adding the distributions of observed POT events to Figure 4 would overload the plot. Please also note that modified the plot for a better readability referring to a comment by referee one.

Figure 5: Is this result based on an average across the model ensemble for all 25 parameter sets?

No, the pie charts represent the total number of events by the entire ensemble. We clarified this in the figure caption. A modified version of this figure will also show the total number of observations and the percentage change of these for the future period.

Figure 6 needs a better introduction (hard to read and not well explained). Difficult to understand the text that follows (p.6291, line12-20), and this section needs revision.

We completely agree that a better introduction and explanation of this figure is needed. Also the methods which were used to derive the results shown in Figure 6 will need a detailed introduction. Therefore, in addition to adding a paragraph to the method section, we revised the entire part of section 4.4. where we discuss the results shown in Figure 6. Since Figure 6 summarizes many aspects which already have been shown in the previous sub-chapters, this allows us to give a summarizing discussion as recommended by an earlier Comment.

TECHNICAL CORRECTIONS

P.6275 (line 21). The reference by Lawrence and Hisdal (2011) cite change in flood frequency, then refers to flood magnitudes; please clarify.

Considered, we have clarified this as suggested in our AC.

P.6227, line 17: rewrite as i. reads like snowmelt in inland and northernmost Norway causes high flow s during spring and summer in the whole of Norway (similar for ii.).

Thank you, we rewrote the sentences referring to your suggestion.

P.6282, line 6: 'this approach performs remarkable well'; provide details of what performs well and where.

We have added the information that the approach performs remarkable well 'without seasonal subsampling'.

Overall use comma more (particular to distinguish between the use of 'that' and 'which').

We have re-checked the use of comma throughout the text.

Suggest to replace the word 'mismatch' when discussing model performance with something more informative, e.g. underestimation, …

Considered. We have clarified the direction of the mismatches in the main text.

P.6287, line 21: Sentence starting: "For Fustvatn", is this the correct catchment here?

Yes, this is the correct catchment. The information provided here refers to the three boxes on the right.

P.6293, line 4: replace "different regions" with "six catchments representing different …".

Considered, we have rephrased this paragraph as suggested in our AC. Generally, we have modified the entire text to be more catchment specific and less general in our conclusions.