Hydrol. Earth Syst. Sci. Discuss., 11, 6273-6309, 2014

Climate change impacts on the seasonality and generation processes of floods in catchments with mixed snowmelt/rainfall regimes: projections and uncertainties

K. Vormoor, D. Lawrence, M. Heistermann, A. Bronstert

Final response to referee #1 Daniel Viviroli:

We want to thank Daniel Viviroli for his valuable comments on our manuscript. Please find below our replies referring to each of his points. For convenience, the comments by the referee are repeated in *gray-italic*. Text designated for inclusion in a revised manuscript is given in blue:

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.*

  Indeed, split sampling tests would deliver some details on the predictive skill of the hydrological model under changed conditions – as far as considerably different conditions can be detected in the observation data. We argue, however, that it is good practice to use long calibration time periods to ensure that a large variety of relevant hydrometeorological conditions are covered (Merz et al., 2009), although we are aware that other studies showing problems with the transfer of hydrological model parameters in time (Brigode et al., 2013; Merz et al., 2011). We agree that this issue needs some further discussion in a revised manuscript. Extending the study with detailed modeling experiments would probably overload the paper. We will, however, clarify why we used such a long calibration period (30 (24; Kråkfoss) years), and we will address the possibilities for testing the predictive skill of the hydrological model under changed climatic conditions. Thus the following modifications are suggested to be added in a revised manuscript:

  P6283, L11 ff: The HBV model was calibrated for each catchment using daily-averaged discharge data. Excepting Kråkfoss, where observed data are only available since 1966, the entire reference period (1961-1990) was used for model calibration. Using such a long calibration period increases the chance that all relevant processes are covered (Merz et al., 2009).

  P6292, L25 ff: It further stresses the need for alternative calibration approaches which improve the robustness of the hydrological model simulations such that they are able to adapt changes in the dominant runoff generation mechanisms under future conditions. For example, split sampling tests (Klemeš, 1986) could be performed for calibration- and validation sub-periods showing differing phases of flood seasonality and/or FGPs in the observation data. That way, parameter sets could be identified which are specialized for runoff simulations under future hydroclimatological conditions. This presupposes, however, that relevant changes can already be detected in the observation data.

- *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?*

This is a good point, which motivated us to compare the performance of EQM with parametric-, and distribution based methods for the adjustment of temperature in the six catchments studied. Generally, it should be noted that different local adjustment methods (both change factor based and bias correction methods) lead to very similar results for the adjustment of temperature (e.g. Räisänen & Räty, 2012). Still, these two authors also found methods based on quantile mapping performing relatively best compared change factor methods and other bias correction methods. Below, we shortly illustrate a quick evaluation of the adjustment for temperature using different types of quantile mappring: (i) EQM (as in our study), (ii) a distribution derived transformation, and (iii) a parametric transformation (Simple Scaling). We calculated the mean absolute error (MAE) between the observation data and the adjusted RCM data for the reference period (1961-1990) to evaluate the overall performance of the different methods (similar as for precipitation in Gudmundsson et al. (2012)):

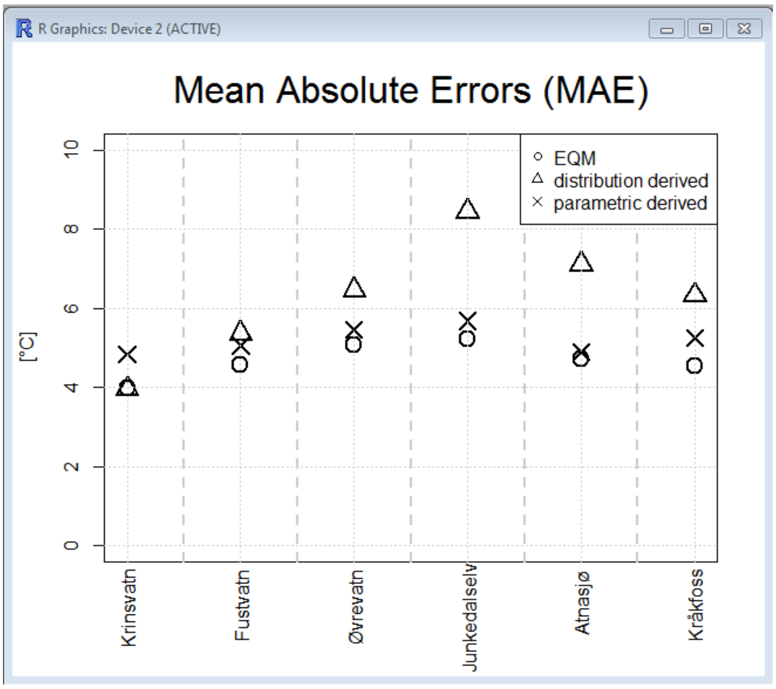

Figure 1: The mean absolute errors (MAE) between observed temperature and bias-corrected RCM-temperature simulations for the reference period (1961-1990). Bias-corrections was performed by three different different versions of quantile mapping.

The MAEs for the RCM temperature data adjusted by EQM are lowest for all catchmetns considered, which indicate that EQM performs best compared to the other quantile mapping methods considered. We will add a note about that observation in section 3.3.1.:

For Norway, Gudmundsson et al. (2012) found that non-parametric transfer methods (as EQM) performed best for the bias correction of precipitation compared to parametric and distribution derived transformations. For temperature, we found the same ranking though the differences are not as large as for precipitation.

- *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 suggest adding a paragraph to section 3.5.2. 'Changes in FGPs':

Note, that there exist more sophisticated approaches to classify flood process types, as combining various process indicators (e.g., flood timing, storm duration, rainfall depth, snowmelt, catchment states) suggested by Merz and Blöschl (2003). The classification proposed here, however, is very easy to apply and straightforward for our purpose since the obliged runoff components can be simply inferred from the output of the HBV model.

- *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?*

Thank you for this suggestion. We checked the POT specific skill scores presented in Viviroli et al. (2009) and found that not all of them are throughout appropriate for our purposes. Especially, the skill scores which are sensitive to the timing and extent of flood events (in terms of days not seasons) will not serve for our purposes since these are not the most important features that our model aims to cover. However, we calculated the skill score which estimates the sum of absolute errors in the POT series. The values confirm our validation results as shown already in the manuscript. Given the length of our manuscript, we would rather refrain from introducing and discussing the suggested skill scores.

- *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?*

This issue was recognized by all three referees. We, therefore, suggest adding a paragraph to section 3.5. explaining on how 'normal duration' was determined:

The normal flood duration has been derived for the six catchments considered by a simple experiment using the HBV model: each catchment was artificially drained to baseflow conditions before twice the amount of annual rainfall was added to completely saturate the catchment again. Concentration and recession time to baseflow was estimated from the resulting hydrographs; concentration and recession time together give 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 will refer to Köplin et al. (2014) in our revised discussion on the relationship between changes in flood seasonality and its underlying causes.

MINOR COMMENTS

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

We will modify the abstract in a revised manuscript so that it contains this information. We suggest extending line 4ff.: Using a multi-model/multi-parameter approach to simulate daily discharge for a reference (1961-1990) and future (2071-2099) period, we analysed the projected changes in flood seasonality and its underlying generation processes in six catchments with mixed snowmelt/rainfall regime in Norway.

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

P6276, L25: *Readability: ... related to changes in the magnitude vs. *changes* in the frequency of events?* We will accept that suggestion.

P6278, L16: *Mention the time period also in the main text, not only in the Figure.* We will do that.

P6278, L24: *The main text discusses *mean* elevation, Table 1 however lists *median* elevation.* Thank you. We will adjust the main text to be in line with Table 1.

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, we will correct that.

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.*

We will add the following information to the Figure caption:

The NSE_w values given for each catchment represent the goodness-of-fit of the HBV model for the entire series (not only POT events) using the best parameter set identified by the calibration. Note that the ordinate's point is not zero and differs between the single plots. Note that we will also update the NSE_w values in Figure 2 since the NSE_w for Kråkfoss is actually higher (0.87) than given in the current version of the figure (0.77). Sorry for that mistake.

P6306, Figure 4: *(1) 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. (2) Point of origin for ordinate not 0: see above. How was the maximum value of the ordinate determined? (3) It does not seem to correspond to the maximum values displayed (i. e., the whiskers).*

We will modify Figure 4 with respect to your suggestions and clarify the figure caption. The new Figure will have vertical lines supporting the readability of changes in the number of events (width of the boxes). The width of the boxes will no longer be given as square roots of the actual number of events but as scales compared with the largest amount of events (spring/summer vs.

autumn/winter) within each period. The modified version of Figure 4 and the new figure caption will read like this:

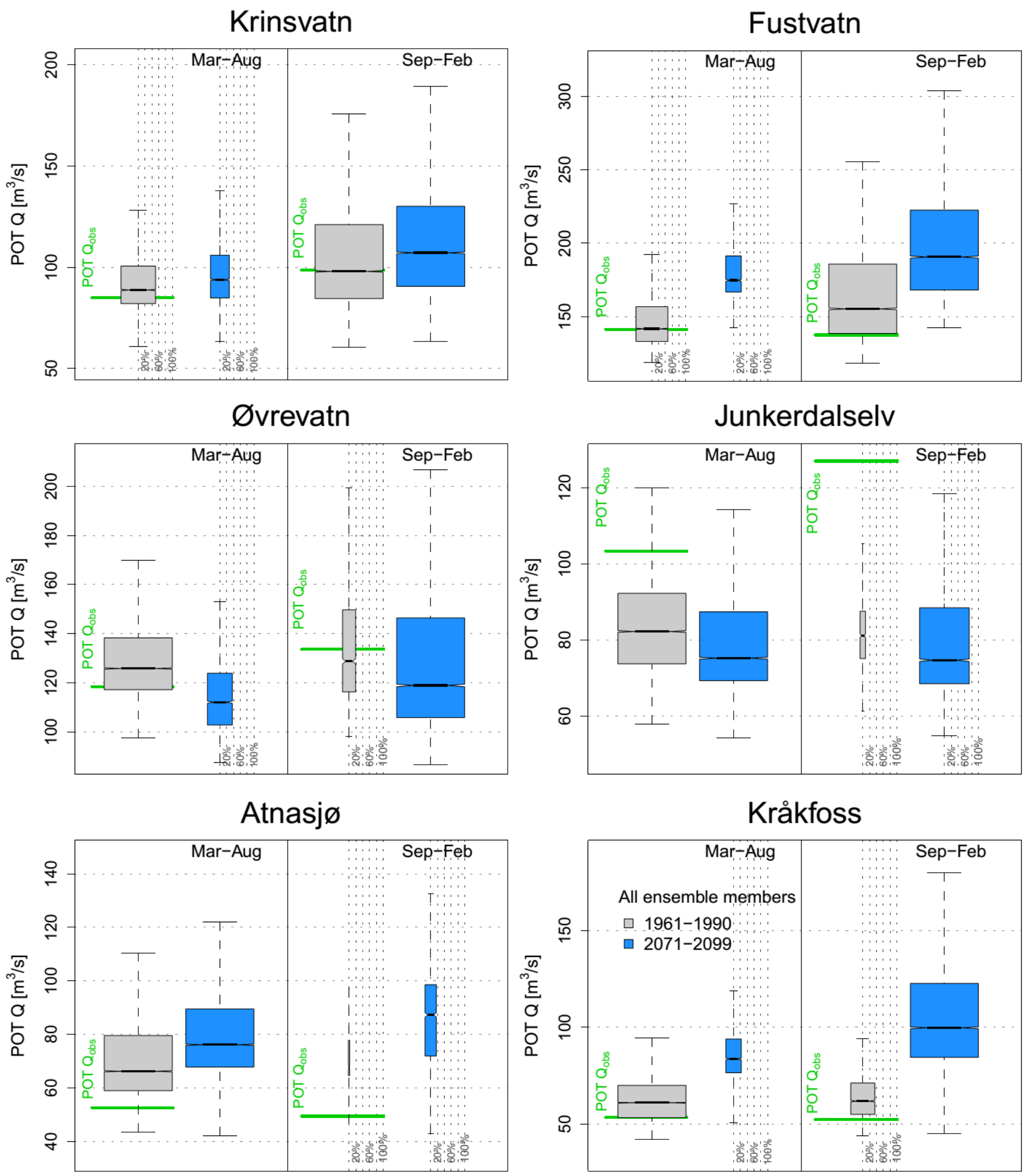

Figure 4. Boxplots showing the median and interquartile magnitudes of the simulated POT events from all ensemble realizations for the reference (grey boxes) and future period (blue boxes), separated with respect to the two basic flood seasons in Norway (spring/summer - left; autumn/winter - right). The whisker-range corresponds to twice the interquartile range. The green bars (POT_{obs}) indicate the median magnitudes of observed POT events. The width of the boxes illustrates the seasonal distribution in the frequency of the POT events: Per catchment and period, the smaller boxes are scaled compared to the larger boxes representing the dominant flood season in terms of flood frequency.

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 the current version of Figure 5 is not completely straightforward with respect to the visualization of the direction of change in the total number of events. We will thus modify the figure in a revised manuscript. However, instead of adding a scale, we will give the total number of events simulated for the reference period, and show the percentage change in the pie for the future period. The change in size of the pies will then underline the directions of change. The modified version of Figure 5 (including modified captions) will look like this:

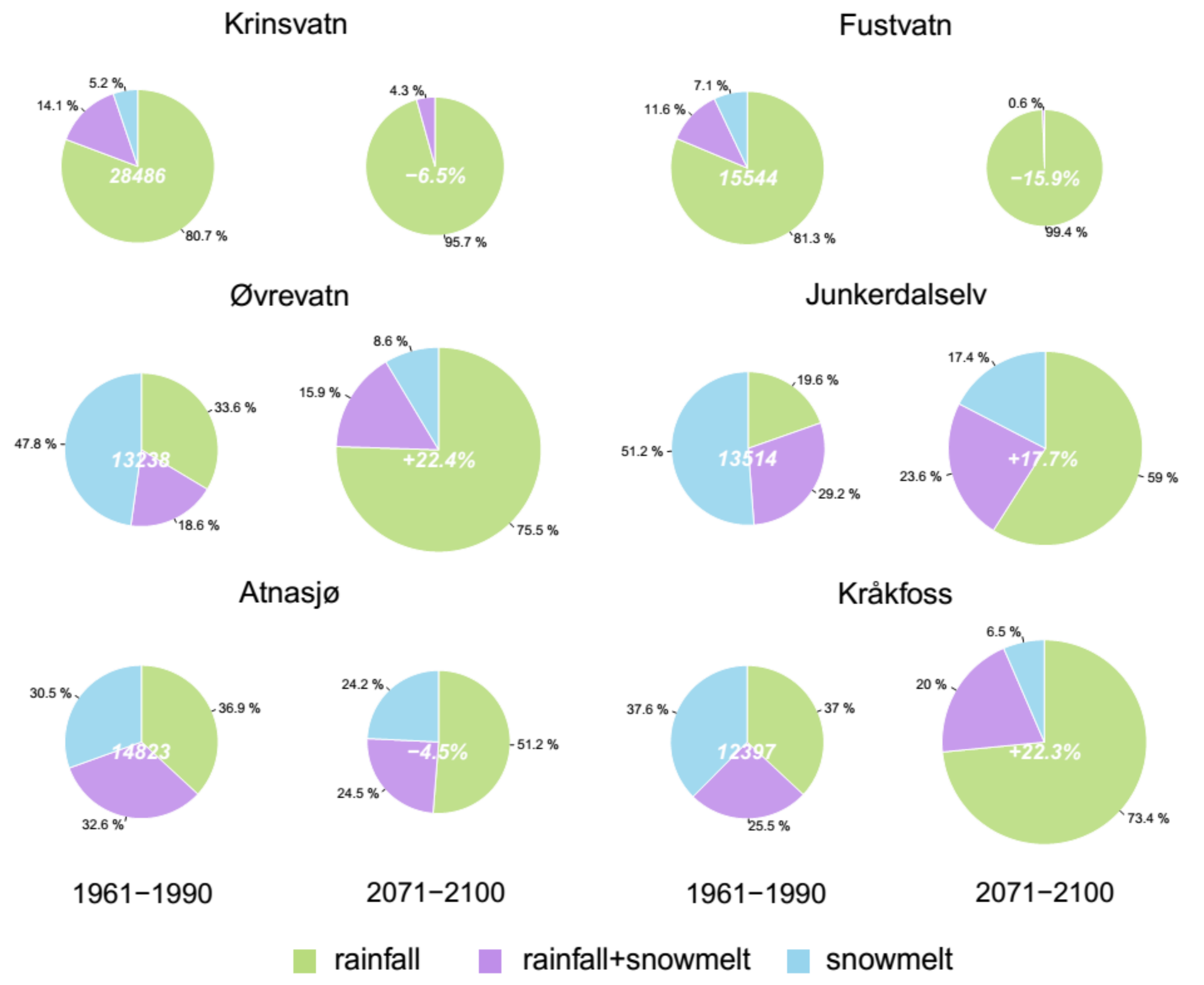

Figure 5. Percentage of POT events according to their FGPs in relation to the total number of events for the reference (left pies) and future period (right pies) derived by all ensemble realizations. The diameter of the pies for the future period indicates the direction of change in the total number of events. Total numbers of events for the reference period and the percentage change in the number of events for the future period are given by the white numbers within the pies.

REFERENCES

Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.-H., and Valéry, A.: Crash tests for a standardized evaluation of hydrological models, Hydrol. Earth Syst. Sc., 13, 1757–1764, 2009.

Brigode, P., Oudin, L. and Perrin, C.: Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change?, J. Hydrol., 476, 410–425, doi:10.1016/j.jhydrol.2012.11.012, 2013.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E. and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods, Hydrol. Earth Syst. Sci., 16(9), 3383–3390, doi:10.5194/hess-16-3383-2012, 2012.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrol. Sci. J., 31(1), 13–24, doi:10.1080/02626668609491024, 1986.

Köplin, N., Schädler, B., Viviroli, D. and Weingartner, R.: Seasonality and magnitude of floods in Switzerland under future climate change, Hydrol. Process., 28(4), 2567–2578, doi:10.1002/hyp.9757, 2014.

Merz, R. and Blöschl, G.: A process typology of regional floods, Water Resour. Res., 39(12), 1340, doi:10.1029/2002WR001952, 2003.

Merz, R., Parajka, J. and Blöschl, G.: Scale effects in conceptual hydrological modeling, Water Resour. Res., 45(9), W09405, doi:10.1029/2009WR007872, 2009.

Merz, R., Parajka, J. and Blöschl, G.: Time stability of catchment model parameters: Implications for climate impact analyses, Water Resour. Res., 47(2), W02531, doi:10.1029/2010WR009505, 2011.

Räisänen, J. and Räty, O.: Projections of daily mean temperature variability in the future: cross-validation tests with ENSEMBLES regional climate simulations, Clim. Dyn., 41(5-6), 1553–1568, doi:10.1007/s00382-012-1515-9, 2012.

Viviroli, D., Zappa, M., Schwanbeck, J., Gurtz, J. and Weingartner, R.: Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part I: Modelling framework and calibration results, J. Hydrol., 377(1-2), 191–207, doi:10.1016/j.jhydrol.2009.08.023, 2009.