We have now made our proposed revisions based on R1 and Massimiliano’s comments, which we feel has resulted in a much-improved manuscript. All of the revisions in response to the referee comments are detailed below in red. All page, line, section, figure and table references refer to those in the mark-up document. To ensure transparency, we have kept all of the text from our final response document to R1 submitted on 20th December 2018, which includes the referee comments in black and our responses in blue. Additionally, we have added all referee comments made by Massimiliano and our responses for which we proposed revisions (which were originally in separate PDF) to this document using the same colour scheme.

Before detailing specific revisions, we feel it would be useful to highlight several of the major modifications to the manuscript which have been undertaken in light of the review comments and are referred to on multiple occasions in the specific referee comments in the text that follows. These include:

1. Evaluation of signature changes and uncertainties using 25-year time-slices

All results have now been re-analysed using longer 25-year time-slices instead of the original 10-year periods. As such, all figures, tables and corresponding text in the results and discussion sections have been revised accordingly. While the overall picture painted by the new set of results is very similar to that from the original results, substantial revisions to the results and discussion sections were still required and these are listed in more detail under specific referee comments.

2. Separation of RCPs

Both referees highlighted the need to separate out the RCPs when presenting and discussing future climate impacts on ice and snow coverage and river flow regime. Accordingly, several figures have been re-designed to accommodate both RCPs using a consistent colour-scheme to aid the readership. Further details of this are given under specific referee comments. The corresponding text in the results and discussion sections has also been revised to include these new details.

3. Extension of discussion section

Both referees noted the need for a more thorough discussion of limitations of the methods adopted. In response to this we have included an additional Limitations section in the discussion. Additionally, R1 asked for further discussion of our results in the context of previous studies and accordingly, substantial revisions to the initial discussion section have also been made to address this. Further details of these revisions are given under specific referee comments below.

We look forward to the response from the editor at their earliest convenience.

Anonymous referee #1

We thank anonymous referee #1 (R1) for taking their time to read our manuscript and provide us with a very thorough and fair set of comments and recommendations. We are largely in agreement with the recommendations and have therefore proposed some substantial revisions that we feel will result in a much-improved manuscript. We have considered our response to each comment carefully, especially to those that require clarification or are critical for which we provide more detailed responses with clear justifications. We have listed all referee comments in black followed by our responses in blue. Any proposed revisions are highlighted with bold text. We did notice that R1
Dear editor, dear authors,

The authors present a climate impact study including uncertainties from different sources of uncertainties. They included an ensemble of RCPs and climate models as well as different model structures into the model chain. Additionally they used many objectives for calibration and for impact assessment (even if the terminus “multi-objective” or “many-objective” was never used in the presented paper). Although the methods of the study need to be more elaborated, the paper presents a comprehensive work. On larger spatial scales multi-impact-model applications have become very common. The authors try to implement such an approach by choosing different, commonly applied process presentations for the snow/ice processes and for runoff generation processes. For example, runoff generation is represented by two different linear storage methods. However, one limitation with regard to the choice of process representations is that they are, like in the mentioned example, quite similar looking at the variety of process representations possible. In general, the paper is quite long and has a lot of figures. The authors should try to limit word count and balance the length of the different chapters (the discussion is too short compared to other chapters).

There are two points here which are not explicitly raised the remainder of the review and so we will address these before proceeding to the detailed comments.

i) “In general, the paper is quite long and has a lot of figures. The authors should try to limit word count and balance the length of the different chapters (the discussion is too short compared to other chapters).”

As R1 appreciates, this manuscript “presents a comprehensive” piece of work and accordingly we took great care in choosing the level of detail to ensure reproducibility and to achieve the two principal aims of the study outlined in the introduction. While 14 figures and two tables in the main text is perhaps more than average, we don’t feel it is excessive, particularly for a study as comprehensive as this. Indeed, each figure has a clear purpose and demonstrates something new to the reader that cannot be deduced from any of the other figures or from the text. For similar reasons, we don’t feel the word count is excessive. Note, nearly 60% of the wording is contained within the methods and discussion sections and we are in agreement with R1 that there is scope to elaborate further on some aspects of the methodology and to extend the discussion. Even so, on revising the manuscript, we will make every attempt to limit the word count where possible.

ii) “…one limitation with regard to the choice of process representations is that they are, like in the mentioned example, quite similar looking at the variety of process representations possible.”

We appreciate that there are many other runoff-routing models that have been used in glaciated catchments ranging from linear reservoir models to physically-based hydraulic models that simulate discrete flow pathways through the glacier (e.g. Arnold et al. 1998). We adopted the concept of linear reservoirs because they have been so widely used in other glacier-hydrology studies (Hock and Jansson, 2005), particularly in climate change impact assessments. Their popularity can be partly explained by their simplicity. They employ few parameters, making them ideal for data-scarce mountain catchments. Additionally, the concept of linear reservoirs lends itself to structural modifications (e.g. through implementing multiple reservoirs in series or parallel) making them...
extremely versatile across different glaciated settings. Therefore, an investigation into the uncertainty stemming from using linear-reservoirs is particularly relevant to the field of glaciated catchment hydrology. As we state in the manuscript (P10L7), we selected the two models implemented in this study based on results from our previous study (Mackay et al., 2018) where we experimented with three different levels of linear-reservoir complexity. The most complex structure which implemented separate reservoirs for snow, firn, ice and soil actually inhibited model efficiency and therefore we could not justify including it in this study. Indeed, the fact that relatively simple (and more similar) structures were best-suited for this particular catchment is perhaps not surprising. As we note (P30L14), “previous investigations have shown that Virkisjökull has a well developed conduit drainage system that routes runoff efficiently year-round (Phillips et al., 2014; Flett et al., 2017).” This is contrary to many other temperate glaciers which exhibit transient water storage behaviour due to the periodic (seasonal) activation of efficient englacial/subglacial drainage pathways (Jansson, Hock and Schneider, 2003). For these catchments, the use of more complex model runoff-routing model structures might also be justifiable. Even so, we appreciate that if a model interrogation study like that of Mackay et al. (2018) were undertaken which evaluated other runoff-routing models, it could yield additional ‘suitable’ model structures to be included in a study like this. This of course is beyond the scope of this study, but is an important point to raise in the discussion and will be done so in the revised manuscript.

We have included this point as part of the new Limitations section of the discussion (P37L29) and use it to highlight the fact that the “total projection uncertainties reported are not indicative of the true uncertainty” (P39L16) partly because “the representation of uncertainty in the five components evaluated in this study are not exhaustive” (P39L21). We then use the melt and runoff-routing model components to explain this point as the structures implemented “represent a sub-set of a much larger population of available models” (P39L26) and we note more complex, physically-based model structures such as the glacier hydrology model proposed by Arnold et al. (1998) as an example of these.

I have four main points for the authors to consider.

1. Shortcomings in stepwise calibration: The authors calibrate their model in two steps. First, they used different model structures and objectives for calibrating ice/snow processes (TIM) and second they applied different model structures and objectives for calibrating runoff generation (ROC). By calibrating in a first step only different TIM structures, the authors assume that the runoff generation processes will not interact with the TIM processes, which might be true. But I see a problem in the second step of calibration where they use only one TIM representation to further calibrate their model for runoff generation signatures. By this, the calibration is relying only on one set of TIM parameters and its water flux characteristics. I doubt that the selected parameter sets will represent the best possible parameter sets for runoff generation processes. The assumption that the selected 24 different TIM representations will not interact with a particular ROC is in my eyes not correct. I suggest to do a ROC calibration based on all unique TIM representations.

To be clear, the assertion that “only one TIM representation [is used] to further calibrate their model for runoff generation signatures” is not correct. As we state (P13L25), “a single [ROR] parameter set was selected based on its mean performance across the 24 TIM compositions”. In other words, we used all 24 TIM parameter-structure combinations (compositions) in combination with all potential ROR compositions (5000 parameter sets × 2 structures) and selected the most efficient ROR models based on their overall (mean) performance across the 24 TIM compositions. The reason we selected the ROR models in this way was because the second aim of the manuscript necessitated it (P4L18). Namely, “to use ANOVA to quantify the relative influence of the five model chain components to
projection uncertainty across the different characteristics of river flow regime”. More specifically, in order to apply ANOVA and separate out the uncertainties from the TIM and ROR model chain components (factors), we needed an ensemble of model chain runs that included every combination of each factor level so that equations 2-7 could be solved. If we had selected the best ROR compositions separately for each of the 24 TIM compositions, we would not have been able to do this. We emphasise this point in the manuscript (P13L26) by stating that this ROR model selection process “…was done to satisfy the ANOVA requirements so that the TIM and ROR composition uncertainty could be analysed separately”.

As R1 rightly points out, a limitation of this stepwise calibration approach is that it does not fully take into account interactions between the TIM and ROR model components given that they are not calibrated simultaneously. Indeed, we sympathise with R1’s doubts that because of this “the selected parameter sets will [not] represent the best possible parameter sets for runoff generation processes”. We would however, emphasise that the model evaluation (Figure 4) demonstrates that the models are able to capture the majority of the signatures within their observation uncertainty bounds. Furthermore, the ANOVA assessment indicates that for future projections at least, interactions between the TIM and ROR models are negligible except for two of the 25 signatures evaluated (Figure 14). Even so, we agree that it is important to emphasise the assumptions made by adopting this calibration approach and we will this highlight this as a potential limitation in the discussion section of the revised manuscript.

As with the previous revision we have included additional text in the Limitations section of the discussion which discusses the limitations of the step-wise calibration procedure as outlined in the response above. We emphasise (P38L33 onwards) the “possible model deficiencies brought about by the two-step GHM calibration procedure” and that this was “necessary so that the main effects (Eq. 4) and interaction terms (Eq. 6) for both components could be calculated separately”, but that this approach “neglects any interactions between the TIM and ROR models”.

2. Longer periods for impact assessment: In climate impact studies it is recommended to take at least 30 years of data for the reference period and a 30 years long future period for analysis. The authors take only 25 years. This limitation should be mentioned in the discussion of methods. Later on they even present their results on climate change impacts on the basis of 10 years slices. This is not acceptable. It is likely that e.g. the results on different uncertainty sources is biased and that the uncertainty coming from the “climate” is overestimated.

We certainly agree with R1 that when analysing climate projection time-series for temporal changes in climate, the use of longer time-slices help to smooth out year-to-year climate variability and therefore any derived shifts in their statistical properties are likely to be more robust. Indeed, we can reassure R1 that when defining our reference and future climate periods we made them as long as possible. 25 years was selected because our observed climate data only extends back to 1981 and the EURO-CORDEX historic simulations only run up to the end of 2005. We were therefore limited to this 25-year period to represent our reference climate data period.

We disagree that using 25 years is a significant limitation of the study. Climate impact studies for Iceland have typically used a shorter, 20-year (1981-2000) reference climate period (Aðalgeirsdóttir et al., 2006, 2011; Guðmundsson et al., 2009). The 1981-2000 reference period has also been used to evaluate hydrological impacts in other glaciated regions (e.g. Liu et al., 2013) and in the recently published state-of-the art UK Climate Projections (UKCP18 – see https://www.metoffice.gov.uk/research/collaboration/ukcp) which evaluate changes in means and quantiles of rainfall and temperature like our study.
We sympathise with R1’s reservations about using 10-year time-slices to evaluate changes and uncertainties in the hydrological projections. We could have used 25-year time-slices to maintain consistency with the climate data and to bolster the robustness of any derived changes and uncertainties. We chose 10-year time-slices to provide a more refined indication of changes in the signatures and their sources of uncertainty over time, but on reflection, we agree that by doing so we may have introduced biases in inter-annual variability. We will therefore re-analyse the results using a new set of 25-year long time-slices in accordance with the driving climate data. For this, 1991-2015 will be used as the reference period (note, we do not have ice-coverage data prior to 1988 and so can’t use the 1981-2005 period used for the climate) and a set of future periods centred on the 2030s (2023-2047), 2040s (2033-2057), 2050s (2043-2067), 2060s (2053-2077), 2070s (2063-2087) and 2080s (2073-2097) will be used.

All results have now been re-analysed as described above. In total, six figures required modifying in light of these revisions including Figures 10, 11, 12, 13, 14 and D1. Table 2 was also updated and the text in results sections 3.4 – 3.7 and the discussion was revised to accommodate these changes.

3. Separating impacts for different RCPs: The authors present quite often a mean over both selected RCPs in the result section and partly in the discussion section. I suggest to show the results for both RCPs separately (or only one) as the results strongly depend on the warming level.

The decision to combine the RCPs was made for two reasons. Firstly, as the referee has noted, the manuscript outlines a comprehensive piece of work and we were hesitant to include the additional text and/or figures to present results from the different RCPs separately. The majority of hydrological impact studies that use multiple emission scenarios display their respective projections separately as these are often the main/only source of uncertainty investigated. In our study, we have included five different sources of uncertainty and it is clear that the emission scenario dominates snow and ice projection uncertainty (Figure 10). We therefore agree that it would be beneficial to separate out the projections in Figure 7 by RCP and update the text in the results section accordingly in the revised manuscript. For the hydrological projections, the emission scenario plays an important (although rarely dominant) role in projection uncertainty (see Figures 13 and C1). The decision to prioritise the separation of the RCPs over say the GCM-RCMs for the hydrological projections (which dominate projection uncertainty more frequently) is therefore questionable. However, it would allow for much easier comparison with previous studies, and so we will separate the RCPs for Figures 11 and 12 also in the revised manuscript.

We have modified Figures 7, 11 and 12 and the corresponding text in the manuscript so that the RCP4.5 and RCP8.5 scenarios are shown and described separately. For consistency, we have adopted the same colour scheme as originally used in Figure 5 (RCP4.5 = blue and RCP8.5 = yellow).

Given the additional graphical information added to Figures 11 and 12, several modifications to the original design were made to make them as interpretable to the reader as possible. Firstly, we have removed the graded colour bands used to represent the different confidence intervals, as these were difficult to interpret when using two colour schemes overlaid on one another (we wanted to keep everything on the same axis to allow for easier comparison between the two RCPs). Instead, we have opted for a single colour band to represent the 10th and 90th percentiles so that ≥90% confidence in direction of change can be determined easily. We have also utilised markers on the ensemble mean lines to indicate when a smaller ≥75% level of confidence is achieved which we deemed important given that we frequently refer to this level of confidence in the text. Finally, in Figure 11 we moved the ensemble mean monthly runoff volume lines to insets in each subplot to accommodate both RCPs.
After making these revisions, we also noticed that the ‘Mean extent’ simulation in Figure 8 is itself taken from the aggregated RCP4.5 and RCP8.5 ensemble. Furthermore, given the purpose of Figure 8 is to provide an indication of the range of ice extent simulations, it provides little added value to the reader and was not discussed in detail in the results section. For these reasons, we have removed this part of the figure in the revised manuscript and we now also explicitly state the RCPs used to force each of these simulations in the figure caption.

4. Extend discussion: The authors discuss the consequences of their findings e.g. with regard to threats for infrastructure and changes in water availability. What I miss is the link of their findings to other studies and actual debates in science. The study would also benefit from a critical discussion of advantages (novelty) and limitations of applied methods.

We appreciate that section 4.1 of the discussion is weighted towards impact assessment although we do feel relating the projections to downstream impacts forms and important part of any climate change impact experiment. We agree that there is scope to expand certain aspects of the discussion. In particular, in the revised manuscript we will discuss how our results compare to those of other glacio-hydrological uncertainty analyses (as outlined P3L20-30). We will also make it clearer to the reader what the novelties of our approach are (as outlined P4L7-12) and exactly what they have shown us. Finally, we will include additional text to discuss potential limitations of our approach. Several of these limitations have already been outlined in this response and several more are detailed in the text that follows.

We have made substantial revisions to the discussion section. For brevity, we will not list every revision here, but will instead highlight some of the main revisions undertaken for the revised manuscript.

Firstly we have added additional text to discuss how some of our findings compare to those of other studies. For example, we found general agreement with the projected changes in flow seasonality (P33L17) and high flow magnitudes (P33L33), but we highlight the fact that other studies have projected the reductions in low flow magnitude (while we project an increase), an artefact of the prevailing arid climate system (P37L34).

We have done the same to compare our uncertainty analysis with other studies. For example, we discuss the relative role of the emission scenario and the climate model in contributing to projection uncertainty where our results contrast some other studies (P36L9) and we discuss reasoning for these differences (P36L19).

We have also added some text to the new limitations section which emphasises the fact that findings from this study will not necessarily transfer across to other catchments, both in terms of the projected changes in river flow regime (P37L31) (some of which has been moved from section 4.1) and the contribution of different uncertainty sources (P38L10) and we have used examples from past studies to demonstrate both.

We have also put more emphasis on the novelties our work as outlined in the introduction. In particular, we emphasise that past studies “implemented multiple [hydrological] model codes and therefore cannot make any definite conclusion about the source of the projection uncertainties” (P37L5) and that in this study we have “demonstrated that by using a single, but flexible model code it is possible to separate out the sources of projection uncertainties down to the process level.” Furthermore, we highlight that such insights are beneficial because they can be used to “help prioritise those aspects of the GHM that require: i) additional refinement (e.g. through model
development); and ii) adequate representation of their uncertainty to improve projection robustness.”

We also highlight the novel signature-based analysis that we have undertaken (P3L15) and how this type of analysis can be used to “help prioritise uncertainty sources based on the characteristic of flow one is interested in.”

Page 2, Line 21: Is “GHM” a common abbreviation for this type of models? This abbreviation is often used for “Global Hydrological Model”

We decided to abbreviate glacio-hydrological model for this manuscript because we use the term 55 times. While we appreciate that GHM is also used to abbreviate Global Hydrological Model, we don’t see that as being particularly confusing for a reader of this manuscript given the focus is on a glaciated catchment-based study and the only mention of global-scale hydrological modelling is written as such (P4L1). We also make it clear to the reader at the beginning of the introduction (P2L21) that we are using GHM to refer to glacio-hydrological model. For these reasons we would like to keep the GHM abbreviation in the revised manuscript.

Page 2, Line 30: Upper Yellow is written with capital letters, but upper Indus not

This will be corrected in the revised manuscript.

Noted and corrected.

Page 2, Line 31: put “river” before the citation

This will be corrected in the revised manuscript.

Done.

Page 2, Line 31-33: Why have you decided to include another reference for Q90 that for Q10? Some of the previous mentioned studies (e.g. Vetter 2015) have also analyzed Q90.

The choice of citations for high and low flow projections in glaciated basins was made based on those that robustly showed changes in these variables (i.e. with confidence). On re-reading this paragraph, we appreciate that we did not make this clear and instead it reads as though we’re saying only those studies have attempted to project changes in high/low flows. Accordingly, we will re-write this to make this clear.

We’ve now included the word ‘robust’ in the first sentence of this paragraph so that it’s clear why we’re referring to these studies (P3L3):

“Some impact studies show robust changes in the magnitude of the highest and lowest river flows…”

We also feel it is important to highlight that projections in low flow magnitudes for glaciated river basins have in general been less conclusive, and we will include the Vetter study as an example of this.

P3L10 of the revised manuscript now reads:

“The projected trends in $Q_{90}$ for the upper Yellow river basin Vetter et al. (2015) were inconclusive as they showed an even spread of positive and negative trends under the warmest climate scenarios.”

Finally, we noted that we could have included the Wijngaard et al. (2017) study, which clearly shows shifts in low flow quantiles, and as such, we will include this in the revised manuscript.
In the revised manuscript (P3L9), it now reads:

“For the Hindu-Kush, Wijngaard et al. (2017) found the opposite impact with an increase in the magnitude of low flow events.”

Page 3, line 1: “:” could be replaced with “which are”

Agreed. This will be corrected in the revised manuscript.

Corrected

Page 3, line 27: change “21st century river flows” to “21st century river flow projections”

Agreed. This will be corrected in the revised manuscript.

Corrected

Page 3, line 28: instead of “emission scenario (ES)”, I suggest to use “Representative Concentration Pathway (RCP)”

We use the term emission scenario (ES) to broadly refer to all types of emission scenarios of which RCPs are one type. For this reason, we decided not to adopt the term RCP as the Jobst et al. (2018) study uses the older SRES emission scenarios. We also clarify that the ESs we use in this study are of type RCP (P6L23).

Page 3, line 35: instead of “large basin-scale” use “regional-scale”

Agreed. This will be corrected in the revised manuscript.

Corrected

Page 4, line 3-5: I don’t think, that all of the hydrological models used in the mentioned studies match the terminology “Computational glacio-hydrological models”. I suggest to use “impact model” instead.

Yes, we agree that the hydrological models used in these studies are not strictly glacio-hydrological models. The reason we used this term was to thread together glacio-hydrological modelling and uncertainty analysis in the introduction but we appreciate that by doing so this statement is somewhat misleading. Accordingly, we will change the term “GHM” (which is used four times in this paragraph) to “hydrological model” in the revised manuscript.

Done

Page 4, line 15: remove “future”

This will be corrected in the revised manuscript.

Corrected

Page 4, line 31: Isn’t it 1988-2011 as mentioned in table A1?

Yes it is. This will be corrected in the revised manuscript.

Corrected

Page 5, line 4: Can you include altitude and/or station name of the “terminus”?

Yes, we will add the AWS1 station name to the revised manuscript.
Page 6, line 1-2: instead of “than to” use “compared”. Where does this numbers come from?

Agreed, “than to” will be changed to “compared” in the revised manuscript.

The numbers come from the gridded precipitation product detailed in section 2.1.1 (P6L11). We will include the Nawri et al. (2017) reference in the revised manuscript to make this clear.

Page 6, line 8: give the installation time for AWS1!

AWS1 and AWS4 were installed in 2009 and 2011 respectively and this will be included in the revised manuscript.

Page 6, line 5-13: Please indicate all the climatic data which was used in your study. I think AWS1 was installed between 2009 and 2011. So there must be additional climate data.

Yes, we appreciate we have been too brief on the climate data used and simply refer the reader to the Mackay et al. (2018) study for further information. Accordingly, we will detail all of the climate data in the revised manuscript.

We’ve now provided some additional detail on the climate data (P7L1) including the source of the temperature and incident solar radiation data and more specifics about how they were sourced from 1981. We also still refer the reader to the Mackay et al. (2018) study if they require further details.

If only AWS1 was used, please remove AWS3 and AWS4 from Fig.1.

No, both AWS1 and AWS4 were used as will be detailed in the revised manuscript. We will remove AWS3 from Figure 1 in the revised manuscript.

Why are the uncertainties coming e.g. from precipitation measurements/bias correction/downscaling not included in your study, as the authors did for other objectives (like runoff, snow coverage)?

This is an excellent question that we did not address adequately in the discussion, but one we have thought about greatly. Undoubtedly, uncertainties in the driving observation precipitation data, whether from observation error (e.g. at the gauge) or post-processing errors (e.g. bias-correction of the gridded precipitation data) could contribute significantly to projection uncertainty. Indeed, we thought carefully about incorporating these uncertainties into our experiment, but decided against it for two reasons:

1) It would have required us to define a statistical model of the precipitation uncertainties. As noted by R1, we did this for other variables including river discharge, for which we used a method which has been applied in a variety of river basin settings around the world (McMillan and Westerberg, 2015). This was possible because we had the information and
data required to compute these uncertainties. Models of precipitation uncertainty have been used in the past e.g. see Blazkova and Beven (2009) who use a precipitation uncertainty model in a hydrological model selection approach. However, the structure and parameterisation of such a model would have been difficult to define in our study, given that we have almost no information on rain gauge errors. We also have very little information on biases in the grided ICRA precipitation dataset (particularly at high elevations where the gauge network is sparse). Accordingly, we felt that any attempt to incorporate these uncertainties into our study would be difficult to justify and could potentially introduce additional biases that would be detrimental to the robustness of the projections.

2) Of course, 1) would not have precluded us from evaluating uncertainties in the quantile-mapping approach used to bias-correct the ICRA dataset. For example, one could imagine undertaking some type of bootstrapping experiment to evaluate the uncertainties in these models. However, this would have required us to include an additional (sixth) factor in our ANOVA experiments. By doing this, we would have made our ensemble size n-times larger, where n is the number of precipitation time-series. With 94,080 simulations, we were already at very limit of what was computationally feasible (even running on an HPC) and therefore we decided against introducing this as an additional source of uncertainty.

Nevertheless, the fact that we are not accounting for potentially very high uncertainties in the historical precipitation data is a limitation of this work and, as such, we will include this as an additional discussion point in the revised manuscript.

The revised manuscript now includes a discussion of the historical precipitation data in the new Limitations section of the discussion (P39L5). Firstly, we highlight the “lack of available precipitation data at higher elevations” and how this makes precipitation data higher up in the catchment less reliable. We also refer back to the Mackay et al. (2018) study who postulate that known biases in the precipitation data may contribute to “the model's inability to capture snow coverage observations higher up in the catchment and river discharge signatures relating to the timing of flows”. We then go on to emphasise that by not including any representation of uncertainties in the historic precipitation data, “the total projection uncertainties reported are not indicative of the 'true' uncertainty” (P39L15). At this point, we felt that it was important to further expand on other potential sources of uncertainty that have not been included in this study, including, for example, additional model structures (as noted above).

Page 6, line 14: Replace “to 2100” by “until 2100”.

This will be corrected in the revised manuscript.

Corrected

Page 6, line 16: GCM means General Circulation Model and not Global Circulation Model.

This will be corrected in the revised manuscript.

Corrected

Page 6, line 27: Change “the RCP2.6 ES” to “RCP 2.6”

Yes, ES will be removed, but note, RCP and 2.6 should not have a space between them and so this will not been modified as suggested.

Corrected
Page 6, line 21-30: I recommend to make a table where all GCM-RCM combinations are shown which were used in this study. This table could be put to the appendix.

Yes, we agree this would be a useful addition to the manuscript. **We will include this in the revised manuscript.**

We have now included an additional ‘EURO-CORDEX models’ section in the appendix (P40L20) which includes a table of the GCMs and RCMs along with their institutions and the combinations of GCMs and RCMs.

Fig.1: The 0.11° EURO-Cordex grid could be indicated in the figure.

**Yes we’re happy to add the CORDEX grids to the manuscript (note not all RCMs use the same grid) and will do so in the revised manuscript.**

After experimenting with adding the grids to Figure 1, we decided it was not possible to do this without the figure appearing cluttered. Accordingly, we’ve added the grids as a separate figure (Figure A1) in the new appendix section.

Page 6, line 31-32: It is common to use at least 30 years, when the climate between two different data sets (here measured and simulated) is compared. Otherwise climate variability is becoming more dominant. If possible, please use 30 years, otherwise this limitation should be clearly mentioned.

We’ve addressed this in the above comments.

Page 6, line 31 - Page 7, line 6: The authors try to analyze the skills of the RCMs by comparing them to the observed climate, but they do it in a very subjective (visual) way, without any statistics or a clear selection criteria. The authors state “Overall, the GCMRCM performance is good . . .” without any proof or definition, what “good” means. The Authors should explain, why they have not defined any levels of acceptability, as they did for the other inputs.

Yes, we appreciate that on trying to keep the word count down, we provided very little information on the climate model skill, and we agree that Figure 2 is not entirely informative of the potential weaknesses of the different GCM-RCMs. In fact, it is important to state that the purpose of including the climate model skill in the manuscript was not to present any kind of formal climate model selection strategy. Rather, we were trying to be transparent about the potential weaknesses/biases in the GCM-RCMs that were used to drive the projections. On reflection, a table of statistics would have been much more informative. Accordingly, for the revised manuscript we will include a table of statistics for each GCM-RCM showing seasonal average means of each climate variable as well as lower and upper percentiles (e.g. 1 and 99 – just 99 for precipitation) to indicate these biases.

We’ve now added these statistics to the manuscript (Figure 2) and an overview of the main characteristics of these are given in the text (P8L1).

We avoided employing a limits of acceptability approach given the highly non-linear nature of climate model projections and the difficulty in evaluating historic skill and relating it to future projection skill. We did however decide to remove one of the GCM-RCMs from the analysis entirely given it’s relatively poor fit to observation data, particularly for the temperature data. More specifically, we found on comparing monthly average temperatures (Figure R1a below) that the [CNRM-CM5]-[ALADIN53] GCM-RCM shows anonymously large negative biases, particularly during the winter months of the year. A comparison of the ECDFs for the winter months also shows these biases (Figure R1b, c and d). We also calculated the RMSE of the monthly ECDFs for each GCM-RCM
and as we stated in the manuscript the [CNRM-CM5]-[ALADIN53] “was consistently poor across all three climate variables”. More specifically, when ranked according to their RMSE scores, the [CNRM-CM5]-[ALADIN53] GCM-RCM ranked 14, 13 and 15 out of 15.

Given the anonymously high biases in temperature and the importance of temperature for driving hydrological change in the catchment (both in terms of melt rate and the proportion of precipitation falling as rainfall), coupled with the fact that the model was relatively poor across all three climate variables, it was deemed appropriate to remove this model from the ensemble. In the revised manuscript we will include additional text to explain our decision in removing this GCM-RCM from the ensemble.

We’ve now included the reasoning behind removing the [CNRM-CM5]-[ALADIN53] GCM-RCM from the ensemble, as part of the new Appendix A and we refer to this in the main text (P8L1). We’ve also included a figure (A2) that has been adapted from R1 below so that it shows the large bias in winter temperature, but also shows the relative ranking of the RMSE scores noted above.

Figure R1: Comparison of observed and simulated (GCM-RCMs) monthly mean temperatures for the recent past (1981-2005)

Fig.2: It is hard to read this figure, especially regarding precipitation. The lines are too close. I guess this is because the extreme precipitation events from climate models are also included. Maybe the 99% quantile is sufficient? Also the ECDF curve for precipitation should start somewhere above the X-axes cross section, indicating the number of days without precipitation.

Yes, we have proposed using a table of statistics to replace Figure 2.

Fig.3: Instead of writing " based on the projections from 1 of the 14 GCM-RCMs with the RCP8.5 ES" you can give the exact name of the GCM/RCP combination which you refer to.

Agreed. This will be corrected in the revised manuscript.

We’ve changed this in the revised manuscript so that it reads:

“...based on the RCP8.5 projections using the CNRM-CM5 GCM and CCLM4-8-17 RCM.”

Page 7, line 19-20: It is common in impact studies to use at least 30 years for reference period and scenario period, respectively.

We’ve addressed this in the comments above.
Page 7, line 25-28: Why don’t you calculate deltas as 25-years (or 30-years) moving window? Then you would not need any interpolation.

The reason we settled on using a linear model for perturbing the deltas over time was because, to our knowledge at least, the study of Farinotti et al. (2012) is the only study that directly addresses the need for transient deltas for the purpose of dynamic glacier evolution modelling for hydrological impact assessment. Given the successful application in this study to a number of catchments of similar size and setting (Swiss Alps), we decided it would be most appropriate and justifiable to adopt this approach also. We recognise though, that we could have chosen a different approach (such as a moving average window). This is therefore a potential source of uncertainty that we have not accounted for in our projections and in our revised manuscript we will include this as an additional limitation in the discussion.

We have now made this point explicitly in the discussion by stating that “other aspects of the downscaling procedure could have also been modified and perturbed (e.g. replacing the linear interpolation of change factors with a moving average model)” (P39L25).

Page 9, line 5-8: You could put the information about the number of sub-samplings directly to step 5. There are two possible (re)sampling schemes: with or without replacement. Why have you chosen sampling without replacement?

On reflection we agree that this would sit better as part of step five and we will edit the manuscript accordingly as part of the revisions.

Done

To be clear, we are sampling with replacement not without replacement as suggested. We will make this clear in step five of the revised manuscript.

Done

Page 9, line 12: What is the resolution of you model domain?

It’s 50 m. We will add this to the revised manuscript.

Done

Page 10, line 3: Are these new methods to calculate soil infiltration and evapotranspiration? Can you name the methods used to calculate infiltration and evapotranspiration? I can’t find/access the given reference Griffiths et al., (2006) in the internet.

No, they are not new methods. The soil moisture model is a sub-model of the UK Centre for Ecology and Hydrology’s Continuous Estimation of River Flows (CERF) regional hydrological model. It’s based on the well-established FAO56 soil moisture accounting procedure (Allen et al., 1998) and has shown to compare favourably to physically-based models at the field scale where interception losses are small (Sorensen et al., 2014). It has been used in the past for national scale hydrological river and groundwater level projections under climate change (Prudhomme et al., 2013; Collet et al., 2018), catchment scale groundwater level modelling (Mackay, Jackson and Wang, 2014; Mackay et al., 2015; Jackson et al., 2016) and distributed recharge modelling (Mansour et al., 2018). We appreciate that the reader would benefit from knowing that the model has been used extensively in the past and so we will include some of the above citations in the revised manuscript to indicate this.

We’ve now included the above citations and it now reads (P10L8): “A soil infiltration and evapotranspiration model (Griffiths et al., 2008) based on the well-established FAO56 soil moisture
accounting 5 procedure (Allen et al., 1998) solves the water balance for model nodes with no ice or snow coverage. This model has been applied extensively (Mackay et al., 2014, 2015; Jackson et al., 2016; Mansour et al., 2018) and has shown to compare favourably to physically-based models at the field scale where interception losses are small (Sorensen et al., 2014).

It appears that the document we cited is no longer freely available to access online, but an identical document (albeit published by the Environment Agency for England and Wales two years later) can be found here: https://www.gov.uk/government/publications/continuous-estimation-of-river-flows. We will update this reference in the revised manuscript.

Done

Page 10, line 5: Given reference Ponce (2014) is gray literature. Chapter is also not given.

We chose to reference the most recent edition of this book which can only be accessed online. However, we appreciate R1’s preference for formally published literature and on reading the first edition published by Prentice-Hall in 1989, the section on linear reservoir models is identical. Accordingly, we will reference the published first edition copy in the revised manuscript with page numbers.

This has now been included in the revised manuscript (P11L1).

Page 10, line 11: Please be more precise here. What kind of observed data was used? How long are the observational periods.

This information is given in appendix A and referred to in the subsequent “GHM calibration” section. However, we appreciate that the reader may wish to know more about these data on the first mention of “observation data of ice melt and snow coverage”. Accordingly, in the revised manuscript we will included a reference to this appendix at the end of this sentence.

Added to the revised manuscript (P11L27).

In addition, we also intend to add some extra information to appendix A including the months used to define the seasonal snow coverage signatures, the elevation ranges of the lower, middle and upper sections of the glacier-free basin area and the years for which the MOD10A1 MODIS data cover.

Added to the manuscript (P42L6 onwards).

Page 10, line 31: “The majority of signatures were selected from past studies . . .” please give a reference for this statement.

Yes, we did not include these references as they are given in Mackay et al. (2018), but we will include them in the revised manuscript.

These have now been included in the revised manuscript (P12L14).

Page 12, line 5: Data description used to create 12 snow and ice signatures is missing

This can be found in the appendix as referenced in the same sentence (P12L6).

Page 12, line 28: Why always 5000 runs? E.g. for TIM1 you have 2 parameters, but for TIM3 5 parameters. Do you think that this makes a difference?
Undoubtedly, the larger parameter space of TIM_{3} means the calibrated models could be ‘less optimal’ than the TIM_{1} models, which has three fewer calibration parameters. As with any Monte Carlo calibration procedure, we tried to balance the density of parameter sampling with our available computational resources (note while the GHM is relatively efficient, it still requires considerable computation time given its distributed structure and hourly timestep). Indeed, we adopted the Sobol parameter sampling strategy to ensure we sampled the parameter space as evenly as possible given these computational limitations. We would emphasise that the calibrated models did show a good fit to the majority of hydrological signatures (Figure 4) and accordingly we are happy that the calibration procedure worked. However, we recognise, that given more computational resources, we would have undertaken a denser parameter sampling, particularly for the more complex model structures. We appreciate that we have not addressed this in the manuscript and will therefore include under-sampling of the parameter space as a potential limitation in the discussion of the revised manuscript.

We have included this as part of the new Limitations section in the discussion. We first state that there are “potential deficiencies in the calibrated GHMs” (P38L22). We then explicitly note one of these which is that “given the distributed structure of the GHM and the fact that it runs on an hourly time-step, running the GHM over multiple years required considerable computation time which limited the number of runs that could be undertaken in the Monte-Carlo calibration procedure” (P38L25). We then recognise that “particularly for the more complex model structures which employ more calibration parameters, a denser parameter sampling could help to find more efficient model parameterisations.”

Page 13, line 5: “Accordingly, only those TIM compositions that captured this signature within the LOA were considered and the rest were discarded.” I have the feeling that this is a very strong assumption. What would be the consequence of having a perfect/better DEM and as a result reducing observational uncertainties? You would probably need to discard most of your simulations.

We agree that if we were to have perfect observations (although of course in reality this is impossible), you would need a perfect model to capture those observations within their uncertainty bounds. Such a result may lead one to obtain no acceptable models leading them to reconsider their model structure and/or parameterisation. They may of course also decide that such LOA are simply impossible to match given limitations/uncertainties associated with boundary conditions/model resolution etc. and in such cases one may choose to relax their LOA (e.g. see Blazkova and Beven, 2009).

We’re not entirely sure what the “strong assumption” alluded to in the above comment refers to? Arbitrary acceptability thresholds (e.g. Nash-Sutcliffe > 0.6 for river flows) are routinely used in hydrological studies to discard large numbers of “unacceptable” or “non-behavioural” models. We would deem the assumptions underlying these approaches much more suspect given they provide little information on what characteristics of model behaviour/processes the behavioural models capture and little reasoning as to why an NSE of 0.6 is good enough. We instead have used multiple model evaluation metrics (signatures) which relate to different aspects of model behaviour and, by quantifying their uncertainty we can define an acceptable model as one that captures the observations within their uncertainty bounds. This also allows us to make more robust statements about model acceptability and (crucially) model deficiencies (e.g. Figure 4d).

As we state in the manuscript (P13L3), “Given the high degree of glaciation in the study catchment, and its recent rapid retreat, an initial emphasis of the calibration was put on the model’s ability to capture the long term glacier volume change signature.”
In other words, given its recent retreat, the glacier has the potential to retreat much further over the 21st century and this could significantly alter the hydrological regime of the catchment (our projection metrics). Given the sensitivity of the projections to the projected glacier retreat (note summer runoff from ice melt was approximately double that of rainfall in the 1990s – Figure 11), we feel we are justified in our decision to use the ice volume change signature to prune the initial set Monte-Carlo runs.

Page 13, line 10: “. . . best captures the 11 remaining . . .” should be 10 instead

Agreed, we will correct this in the revised manuscript.

Done.

Page 14: The whole chapter 2.5.3 is about results and not about methods.

We agree, this chapter would sit much better in the results section. We will move this chapter to the beginning of the results section in the revised manuscript.

Done (P18L22).

Page 14, line 23: “use model runs” instead of “model chain runs”

Agreed. This will be corrected in the revised manuscript.

Done. We also noted the same terminology was used later on (P16L19). This has also been amended.

Page 14 line 27- Page 16 line 15: This paragraph should be shortened and transferred to the section on model description.

Agreed. This will be corrected in the revised manuscript.

In response to this suggestion we’ve made a number of modifications to the methodology section:

- Shortened the aforementioned text and moved it to section 2.3.1
- We’ve also included sub-headings in 2.3 to make it clearer to the reader
- We’ve now got rid of section 2.6 (which was almost exclusively associated with the aforementioned text) and combined with the ‘ANOVA uncertainty analysis’ section (P15L7)

Page 16, line 19-20: It is common to take at least time slices of 30 years for climate impact assessment. By using only 10 years, you will increase the influence of climate variability” and consequently you are overestimating ES uncertainty.

We’ve addressed this in the previous responses.

I also ask myself whether you have put the duplicated TIM and ROR parameterizations into the ANOVA?

No, we did not include duplicates into the ANOVA given that we are sub-sampling all factors down to two levels and wanted to avoid generating sums of squares equal to zero. We appreciate we should have stated this in manuscript given that it could influence the results shown. Given that we will need to run the ANOVA analysis again, we are happy to experiment with including the duplicates in the sub-sampling and assess for convergence of results according to P16L18-19.

We have included duplicates for the new results.
How many unique parameterizations do you have for a given TIM structure and a given ROR structure, respectively?

As we state (P14L2), there are “24 unique TIM compositions were obtained from this calibration stage made up of eight unique parameterisations of each of the three TIM structures.” Also, we state (P15L5) that “14 unique ROR compositions were selected made up of seven unique parameterisations of the ROR1 and ROR2 structures, giving a total of 24×14 = 336 unique GHM compositions”.

Page 16, line 23: Please check your calculation. I get 1×10^8 (102.850.020)

Yes, in our defence the statement and calculation is correct i.e. there are ~ 4×10^8 unique combinations of factors levels when sub-sampled down to two. However, a large number of these would include level repetitions for at least one factor and indeed, we did not sub-sample these in our study making this statement inconsistent with the applied methods. We will amend this in the revised manuscript.

Accordingly, in the revised manuscript (P18L15) we have revised the statement and calculation so that it specifically refers to unique combinations of factors levels without repetitions:

“However, given that there are >10^8 unique combinations of factor levels when sub-sampled down to two (and discarding factor level repetitions), it would have been …”

Page 18, line 3: I suggest to write “[...] consistently show an increase relative to the reference period [. . .]” instead of “consistently predict an increase relative to the recent past

While we appreciate the referee’s preference for using “reference period” rather than “recent past”, we feel that we have made it clear to the reader exactly what we mean by “recent past” in the methodology section. Also, we use the term “reference period” to refer to the changes in hydrological signatures which must have a different reference period to the climate (see previous responses). Accordingly, to avoid confusion we would prefer to keep the term “recent past”. We do however agree that the word “predict” should be replaced with “show” and this modification will be made in the revised manuscript.

Done

Page 18, line 4: “The largest increases are predicted [. . .]” Please always use project and not predict for climate projections.

This will be corrected in the revised manuscript along with the two other occurrences of the word “predict” (P20L1 and P24L4).

Done.

Page 18, line 7-8: Please present climate impacts separate for each RCP. Averaging over both RCPs is not very common.

Agreed. We’ll amend these statements in the revised manuscript.

On re-reading this it’s not obvious that we’re quoting the seasonal range rather than the annual range which would indeed be more appropriate. We have now edited the description of the projections for temperature, incident solar radiation, and precipitation (P19L12 onwards) so that they are reported as changes in the annual mean and for each RCP separately.
Page 18, line 14: correct “Figure 5l”

We’re referring figure 5l represents changes in precipitation for the autumn months.

Page 18, line 15: change “The sign of change” into “The direction of change”

This will be corrected in the revised manuscript along with the eight other identified occurrences of this phrase (P1L11, P22L8, P24L10, P25L3, P25L7, P29L34, P30L21, P31L23).

Done

Page 18, line 28: Maybe I missed that, but what exactly is “annual snow coverage”?

We mean the “annual mean watershed snow coverage” which we realise we did not make clear to the reader. We will revise this wording along with the wording in the caption of figure 7 to make this clear.

Done

In addition, we noted that we express the snow coverage signatures as the proportion of glacier-free basin area covered by snow in Table A1 rather than the total coverage in km² as in Figure 7. To avoid confusing the reader, we will add a sentence to the end of the caption for this table in the revised manuscript which reads: “Note, snow coverage is expressed as a proportion of the glacier-free basin area.”

Done

Fig. 7: Why are the snow coverage confidence bands only positive compared to the mean? Does snow coverage also include ice, or does it only count as snow coverage if there is snow on the glacier?

We appreciate it looks as though the confidence bounds are always higher than the mean (note the mean is not actually shown in Figure 7), but the confidence bounds do in fact extend below the mean. What you’re seeing is the tail-end of the ensemble above the mean. The majority of simulations reside near the lower bound (i.e. the distribution of simulations is skewed).

Page 20, line 12: improve English: “The maximum coverage simulations show higher than . . .”

We’re not entirely sure what the issue is here. We state (P20L10) that “Figures 9a, b and c show the climate projection time-series that produced the minimum (dotted lines) and maximum (dashed lines) snow (blue lines) and ice (red lines) coverage by 2100.” We then refer to these as the minimum coverage and maximum coverage simulations. We therefore cannot identify a reason to change this phrasing.

Fig 5. instead of “recent past “ use reference period

Please see above comment.

Page 20, line 1: “projects” instead of “predicts”

Yes, this will be modified as noted in the previous responses.

Done

Line 20, line 1-10: Again: I guess these results are highly dependent on the warming levels and it would be good to either choose one RCP or to analyze both separately.
We will analyse both separately.

Done. See above.

In addition to the proposed revisions above, we also noticed that Figure 4 does not have the a,b,c and d labelling on it. These labels will be added in the revised manuscript.

Done.

Fig. 11: Could you please present the results in mm and not in km³, because mm is more intuitive and easier to compare to other studies. Is snow included in rainfall? In previous plots you named it total precipitation.

Yes, we can change the units to mm. No rainfall only includes rainfall otherwise we would have named it total precipitation. Note, as described in the figure caption, we’re showing the runoff components, hence why we’re showing rainfall and not total precipitation here.

Done.

Referee #2 (Massimiliano Zappa)

We thank Massimiliano for taking his time to provide us with thorough, fair and helpful comments and recommendations. We also very much appreciate the positive comments throughout regarding the strength of the methods employed for which we spent considerable time choosing and implementing. We are in agreement with almost all of the recommended revisions and we feel that our proposed revisions to the manuscript will result in a much-improved paper. We also noted two key similarities with the recommendations from anonymous reviewer #1 (R1). In particular, the need to separate out the RCPs and the need to include a more thorough discussion of potential limitations of the approaches adopted in this study. We have highlighted any overlaps with R1s comments in our responses and in some cases refer back to detailed responses to R1s comments to avoid repeating ourselves.

P1L14: Maybe be more explicit here and give some 2-3 examples of findings

Agreed. In the revised manuscript we will include some examples of findings.

We have now included several examples of how the projection uncertainty contributed by different model chain components varies across the different signatures (P1L16). Specifically, we write, “For example, the numerical climate model is the dominant source of uncertainty for projections of high-magnitude, quick-release flows, while the runoff-routing model is most important for signatures related to low-magnitude, slow-release flows. The emission scenario dominates mean monthly flow projection uncertainty, but during the transition from the cold to melt season (April and May) the snow/ice melt model contributes up to 23% of projection uncertainty.”


Indeed, an example of glacier-runoff compensating runoff deficit during the anomalously dry summer 2003 heatwave would be an excellent addition to this section of the introduction. We will incorporate this into the revised manuscript.

We’ve now included this reference after (P2L12) and use it to further expand on the compensation effect described in the previous sentence which states that “the so-called ‘compensation effect’ has
been widely observed in the northern hemisphere, whereby partially-glaciated catchments demonstrate reduced intra-annual flow variability (Fountain and Tangborn, 1985; Chen and Ohmura, 1990).” We then use the above study to note that “…the compensation of runoff from melt inputs can actually serve to increase mean runoff during anomalously dry heat wave events (Zappa and Kan, 2007).”

P2L21: the NCM acronym is new to me, we always use GCM/RCM.

Yes, in fact we think we can improve this by using the term numerical climate model (or just climate model) in full. We can then introduce the GCM and RCM abbreviations when discussing the Ragettli et al. (2013) and Huss et al. (2014) studies (P3L20 and P3L24 respectively) so that we can then refer to the GCM-RCM-DS-GHM model chain later on (P4L13). We will revise the manuscript in this way.

We now introduce the terms GCM and RCM on P3L26 which reads, “Projections of river flow regime are inherently uncertain due to assumptions made about the formulation, parameterisation and boundary conditions of the underlying GHM (Ragettli et al., 2013; Huss et al., 2014; Jobst et al., 2018) and climate model, be that a general circulation model (GCM), or combined GCM and regional climate model (GCM-RCM)...”

P2L24: Maybe you can elaborate more on this

Yes on reflection we could expand on the concept of "peak water". We will do so in the revised manuscript.

We’ve added to this sentence now by making it clear exactly what we mean by peak water. The sentence now reads: “Decadal changes in runoff are inevitable over the coming century (e.g. Bliss et al., 2014; Lutz et al., 2014; Shea and Immerzeel, 2016) where enhanced melt will result in increased river discharge to a point in time termed ‘peak water’ after which the continued loss of snow and ice will result in an overall decrease in river flow.” We then tie this into the Huss and Hock (2018) study already included in the original manuscript by saying: “It has been shown that many basins, particularly those with small glaciers, have already reached peak water and face a future of dwindling water supply (Huss and Hock, 2018).”

P3L8: I think you mean Sweden :-) !!!

Yes we meant Sweden. We’ll correct this in the revised manuscript.

Done.

P4L4: Highly?

We agree, "highly" would be more suitable. We will change this for the revised manuscript.

Done.

P4L18: The aim is "to quantify ..." not to "use ANOVA". No need to declare ANOVA here.

We agree. We’ll take out the reference to ANOVA in the revised manuscript.

We’ve now removed this and moved it to P5L6 instead so that it reads “…allowing for uncertainty stemming from these to be localised using ANOVA”.

P6L1: In general or really within this small watershed?
We mean within the region of Oraefajokull. **We will make this clearer in the revised manuscript.**

In the original manuscript, we described a lateral precipitation gradient as being “double the annual precipitation falling to the east of Òræfajökull (3500 mm yr\(^{-1}\)) than to the west (1500 mm yr\(^{-1}\))”. We appreciate the difficulty the reader would have in visualising/interpreting this and so to make the west-east precipitation gradient easier for the reader to interpret, we now refer to total precipitation (taken from the Nawri et al. (2017) data) at two locations visible in Figure 1 - at the summit of Òræfajökull (to the east) and at the catchment outlet to the west (PSL21). It now reads:

“There is a significant lateral precipitation gradient due to the prevailing north easterly winds and orographic effects with more than five-times the annual precipitation falling at the Öræfajökull summit (~8000 mm yr\(^{-1}\)) compared to lower down at the catchment outlet to the west (~1500 mm yr\(^{-1}\)) (Nawri et al., 2017).”

P9L9: Very clear section with adequate referencing to prior work. As your The Cryosphere paper is open access, there is no need to add figures here, but for any reader, it is useful here to have a look at Figure 2 in https://www.the-cryosphere.net/12/2175/2018/tc-12-2175-2018.pdf

We agree, although we should point out that Figure 2 specifically details the runoff-routing component of the model and so we will refer to it below as suggested.

We’ve now included a reference to Figure 2 (P11L2).

P11L11: Did you re-calibrate with respect to Mackay et al (2018, TC)? If yes, why?

Yes we recalibrated them and are happy to clarify why. In the TC paper we calibrated the TIM and ROR model parameters simultaneously so that we could evaluate the model’s ability to capture the ice melt, snow coverage and river discharge signatures simultaneously. In this paper, because we wanted to evaluate the projection uncertainty from the TIM and ROR model components separately using ANOVA, we needed a calibration approach that would give us a set of TIM models and a set of ROR models that we could mix and match with each other so that every possible combination was included in the model chain ensemble. We emphasise this point in the manuscript (P13L26) by stating that this model selection process "...was done to satisfy the ANOVA requirements so that the TIM and ROR composition uncertainty could be analysed separately". Note that R1 highlighted this as a potential limitation of the model skill (a point that we acknowledge) and as such we intend to include this as a discussion point in the revised manuscript.

See previous responses.

Figure 4: Put a,b,c,d where appropriate on the artwork

Agreed, we'll include these in the revised manuscript.

Done.

P18: Here you start mixing RCP4.5 and RCP8.5. Only in Figure 5 you have separate analyses. I think you should not make it throughout the paper, but show in Figures separate colors for RCP4.5 and RCP8.5 trajectories

We agree, and in fact, the same suggestion was made by R1 and we will revise Figures 7, 8, 11 and 12 to reflect these suggestions.
See above responses.

P18L28: Don't you need to declare here a moment of each year that you take to evaluate snow coverage? or is this the maximum snow coverage for each winter?

Yes, R1 also touched on this. To be clear we mean the "annual mean watershed snow coverage". **We will modify this in the revised manuscript.**

Done (P22L8) – also see above responses.

Figure 7 caption: At which time of the year? Or it is at least one time per year?

As noted above, this is the annual mean watershed snow coverage. **We will modify this in the revised manuscript.**

Done – also see above responses.

Figure 8 caption: Role of different RCPs?

Yes, we appreciate that it would be useful to see the difference between the two RCPs in this plot and that the mean projection is somewhat meaningless here as it’s averaged over the two RCPs. Accordingly, **in the revised manuscript we will include the minimum and maximum extents for both RCPs separately.**

We’ve now included RCP4.5 and RCP8.5 in the figure caption.

Figure 10: Nice plot. Please detail the labels in the caption. This agreed well with Bosshard et al.

Agreed. **We will include caption label descriptions in the revised manuscript.**

Done.

Figure 11 caption: declare better how the decadal changes are shown (next each to other per month).

After reading this again we agree that it is not clear how the decadal changes relate to the lines on the plots. **We will modify this to make it clear to the reader in the revised manuscript by changing:**

"Each plot includes projected decadal changes (2020s to 2090s) relative to the 1990s with projection confidence intervals (blue bands) and ensemble mean projections (yellow dotted lines) and ensemble mean monthly runoff volume averaged over the 1990s (black solid line) and 2090s (black dash line)"

to:

"For each month, the trajectory of the decadal (2020s to 2090s) ensemble mean change (yellow dotted lines) and projection confidence intervals (blue bands) are shown. Additionally, the ensemble mean monthly runoff volumes averaged over the 1990s (black solid line) and 2090s (black dash line) are shown."
We’ve modified this caption, although it reads slightly differently from the above proposed revisions given that we are now using different 21\textsuperscript{st} century time-slices and have had to modify the design of this figure to include the separate RCPs. Accordingly, it now reads:

“Projections of monthly mean runoff components including rainfall (a), snowmelt (b), ice melt (c) and evapotranspiration (d) for RCP4.5 (blue) and RCP8.5 (yellow). For each month, the trajectory of the ensemble mean change over the 21st century time-slices (2030s to 2080s) relative to the reference period (1991-2015) is shown by the solid coloured lines. These lines are marked for each time-slice where there is \geq 75\% confidence in the direction of change. They are bounded by the 10th and 90th percentiles of the projections (bands). Inset in each plot are ensemble mean monthly runoff volumes averaged over the reference period (black solid line) and 2080s (dashed lines).”

Figure 12 caption: Typo

This will be corrected in the revised manuscript

Done. In fact, we’ve modified the caption of Figure 12 so that it is more closely aligned to that of Figure 11.

Figure 13: I would not mix the decades here and go for two distinct decades, e.g. 50s and 90s.

On reflection, we agree that mixing decades takes away from the temporal variability in the effect sizes that we show are important in Figure C1. Accordingly, we will revise this plot so that it shows the results from two separate decades.

We’ve now included two sub-figures in Figure 13 which show the effect sizes calculated for the time slices centred on the 2030s at the start of the 21\textsuperscript{st} century and the 2080s at the end of the 2080s. Because of this, we have also made considerable revisions to section 3.7 in the results (P29L28 onwards) to describe the findings from these two figures.

Discussion: You should add here a section on limitations: - short time period of series used for calibration - only a specific catchment

Yes we completely agree with this and we have highlighted a number of limitations in our responses to R1 including limitations of calibration approach, driving climate data and the delta-change approach. Accordingly, these limitations along with the short time-series used for calibration will be discussed in the revised manuscript.

Yes, we’ve included the point about the short time-series in the new limitations section of the discussion (P38L30).

With regards to focussing on a specific catchment, as we noted in a previous response, we feel we have made it clear that these projections are specific to this region in the discussion (see P30L8 - P30L19) and as such we don’t feel the manuscript would benefit from additional emphasis of this point.

References:


Future evolution and uncertainty of river flow regime change in a deglaciating river basin

Jonathan D Mackay1,2, Nicholas E Barrand1, David M Hannah1, Stefan Krause1, Christopher R Jackson2, Jez Everest3, Guðfinna Aðalgeirsdóttir4, and Andrew R Black5

1School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK
2British Geological Survey, Environmental Science Centre, Keyworth, Nottingham, NG12 5GG, UK
3British Geological Survey, Lyell Centre, Research Avenue South, Edinburgh, EH14 4AS, UK
4Institute of Earth Sciences, University of Iceland, 101 Reykjavík, Iceland
5Geography and Environmental Science, University of Dundee, Dundee, DD1 4HN, UK

Correspondence: Jonathan D Mackay (joncka@bgs.ac.uk)

Abstract. The flow regime of glacier-fed rivers are sensitive to climate change due to strong climate-cryosphere-hydrosphere interactions. Previous modelling studies have focussed on projecting projected changes in annual and seasonal flow magnitude, but neglect other changes in river flow regime that could also have socio-economic and environmental impacts. This study employs a more comprehensive, signature-based analysis of climate change impacts on the river flow regime for the deglaciating Virkisá river basin in southern Iceland. 25 metrics (signatures) are derived from 21st century projections of river flow time-series to evaluate changes in different characteristics (magnitude, timing and variability) of river flow regime over sub-daily to decadal timescales. The projections are produced by a model chain that links numerical models of climate and glacio-hydrology. Five components of the model chain are perturbed to represent their uncertainty, including the emission scenario, numerical climate model, downscaling procedure, snow/ice melt model and runoff-routing model are perturbed to propagate their uncertainties through to the river discharge projections. The signature-based analysis indicates that glacier-fed rivers will exhibit changes in. The results show that the magnitude, timing and variability of glacier-fed river flows over a range of timescales will change in response to climate change. For most signatures there is high confidence in the sign-direction of change, but the magnitude of change is uncertain and varies substantially across the different signatures. A decomposition of the projection uncertainties using analysis of variance (ANOVA) shows that all five perturbed model chain components contribute to projection uncertainty, but their relative contributions vary across the signatures. For example, the numerical climate model is the dominant source of uncertainty for projections of high-magnitude, quick-release flows, while the runoff-routing model is most important for signatures related to low-magnitude, slow-release flows. The emission scenario dominates mean monthly flow projection uncertainty, but during the transition from the cold to melt season (April and May) the snow/ice melt model contributes up to 23% of projection uncertainty. Signature-based decompositions of projection uncertainty can be used to better design impact studies to provide more robust projections.
1 Introduction

Mountain watersheds have been referred to as the world’s water towers (Viviroli and Weingartner, 2004; Viviroli et al., 2007), partly because they receive large quantities of precipitation relative to adjacent lowlands, but also because they regulate runoff through the seasonal accumulation and melt of snow and ice. The presence of snow and ice profoundly affects characteristics of downstream river flow regime including flow magnitude, timing and variability over a range of timescales (Jansson et al., 2003; Mankin et al., 2015). This is partly due to the periodic (diurnal and seasonal) variations and longer-term (decadal) trends in melt water inputs brought about by fluctuations in glaciological mass balance. In addition, the dynamic water storage and realease properties of snow and ice (runoff-routing) control downstream river flow response to runoff over hourly to seasonal timescales (Willis, 2005). As such, glaciated basins exhibit river flow regimes that differ from their non-glaciated equivalents. For example, the so-called ‘compensation effect’ has been widely observed in the northern hemisphere, whereby partially-glaciated catchments demonstrate reduced intra-annual flow variability (Fountain and Tangborn, 1985; Chen and Ohmura, 1990). Indeed, the compensation of runoff from melt inputs can actually serve to increase mean runoff during anomalously dry heat wave events (Zappa and Kan, 2007).

Mountain glaciers are retreating at unprecedented rates (Zemp et al., 2015) while snow coverage is receding (Vaughan et al., 2013) resulting in observable changes to downstream river flows (Luce and Holden, 2009; Singh et al., 2016; Hernández-Henríquez et al., 2017; Matti et al., 2017). With near-surface air temperature projected to rise over the coming decades (Collins et al., 2013) future changes in river flow regimes in response to cryosphere change could have wide-ranging socio-economic and environmental impacts. Long-term reductions in melt water inputs will disrupt the supply of water available for irrigation (Nolin et al., 2010; McDowell and Hess, 2012; Carey et al., 2014; Baraer et al., 2015). Increased inter-annual and intra-annual flow variability will threaten infrastructure projects such as hydroelectric power stations (Laghari., 2013; Gaudard et al., 2014; Carvajal et al., 2017). The loss of the runoff-regulating effects of snow and ice could result in more frequent short-term very high flows putting downstream populations and infrastructure at risk (Laghari., 2013; Stoffel et al., 2016). Changes in flow magnitude and variability from annual to sub-daily timescales will threaten the sustainability of some of the world’s most pristine freshwater ecosystems (Bunn and Arthington, 2002; Naiman et al., 2008; Beamer et al., 2017). Therefore, it is of paramount importance to make reliable projections of changes in downstream river flow regimes from glaciated watersheds so that future impacts can be adapted to and mitigated.

Computational glacio-hydrological models (GHM) driven by numerical climate model (NCM) projections allow us to assess how future river flow regimes will change in glaciated river basins. Past studies have focussed on projecting changes in decadal, annual and seasonal variations in runoff magnitude. Decadal changes in runoff are inevitable over the coming century (e.g. Bliss et al., 2014; Lutz et al., 2014; Shea and Immerzeel, 2016) with many basins already having reached ‘peak water’ and facing where enhanced melt will result in increased river discharge to a point in time termed ‘peak water’ after which the continued loss of snow and ice will result in an overall decrease in river flow. It has been shown that many basins, particularly those with small glaciers, have already reached peak water and face a future of dwindling water supply (Huss and Hock, 2018).
Seasonal flow magnitudes are also projected to change as melt cycles evolve and watersheds shift from glacial-nival to pluvial runoff regimes (Kobierska et al., 2013; Duethmann et al., 2016; Ragettli et al., 2016; Garee et al., 2017).

Some studies have projected impact studies show robust changes in the magnitude of the highest and lowest river flows (Wijngaard et al., 2017) including Wijngaard et al. (2017) who projected an increase in the magnitude of the 10% exceedance flow (Q_{10}) for river basins across the Hindu-Kush-Himalayan region. Similar patterns of change have been shown for the Rhine (Bosshard et al., 2013), upper Indus (Lutz et al., 2016) and Upper Yellow (Vetter et al., 2015) river basins upper Yellow river basin (Vetter et al., 2015) show high flow magnitudes will increase. Stewart et al. (2015) projected a decrease in low flow magnitude (Q_{90}) for the snow-covered Sierra Nevada and Upper Colorado river basins due to shifts in the snowmelt season and changes in precipitation type from snow to rain. For the Hindu-Kush, Wijngaard et al. (2017) found the opposite impact with an increase in the magnitude of low flow events. The projected trends in Q_{90} for the upper Yellow river basin by (Vetter et al., 2015) were inconclusive as they showed an even spread of positive and negative trends under the warmest climate scenarios.

Of course, one could go beyond projecting changes in seasonal to decadal mean flow magnitudes and quantiles of the flow duration curve (FDC). A branch of streamflow analysis that has been widely adopted in hydrology is the calculation of river flow ‘signatures’ which are metrics derived from river discharge time-series that represent different characteristics of river flow over specific time scales. These may include mean flows and FDC quantiles as well as metrics to quantify the variability (e.g. coefficient of variation), timing (e.g. peak flow month) and flashiness (e.g. autocorrelation) of flows. Signatures have been used in the past to analyse catchment runoff behaviour and similarity (Yadav et al., 2007; Ali et al., 2012). Furthermore, their ability to localise specific aspects of runoff behaviour make them ideal diagnostic evaluation metrics for model hypothesis testing (Euser et al., 2013; Coxon et al., 2014; Hrachowitz et al., 2014) and calibration (Hingray et al., 2010; Shafii and Tolson, 2015; Kelleher et al., 2017; Schaeffli, 2016). They also offer an opportunity to evaluate past (Sawicz et al., 2014) and future (Casper et al., 2012) river flow regime change. For example, Teutschbein et al. (2015) projected changes in 14 different river flow signatures for 14 snow-covered catchments in Switzerland. They projected changes in Switzerland and showed daily to annual river flow magnitude, timing and variability highlighting the breadth of river flow regime sensitivity were all sensitive to climate change. An analysis like this is yet to be undertaken for any glaciated river basins.

Projections of river flow regime are inherently uncertain due to assumptions made about the formulation, parameterisation and boundary conditions of the underlying NCMs (Giorgi et al., 2009) and GHMs (Ragettli et al., 2013; Huss et al., 2014; Jobst et al., 2018), GHM (Ragettli et al., 2013; Huss et al., 2014; Jobst et al., 2018) and climate model, be that a general circulation model (GCM), or combined GCM and regional climate model (GCM-RCM) (Giorgi et al., 2009). Uncertainties may also be introduced by intermediary steps employed to link the two sets of models together such as downscaling (DS)the NCM output. It is important to quantify the propagation of uncertainties from all sources in the model chain as this provides a basis for assigning more robust levels of confidence to river flow projections. Additionally, one can assess the relative contributions of model chain components to the total projection uncertainty, providing empirical evidence for future research needs (e.g. Meresa and Romanowicz, 2017). Ensemble-based experiments have been used in the past to provide this understanding. Here, different components of the model chain are perturbed, typically using a ‘one at a time’ (OAT) approach where the spread in projections
for each perturbed component is evaluated. Ragettli et al. (2013) perturbed three components of a model chain applied to the Hunza River Basin, northern Pakistan including the NCM-structure-GCM, statistical DS model and parameterisation of the GHM. They showed that all three sources contributed to annual runoff projection uncertainty, but for the heavily glaciated sub-regions of the catchment, the GHM parameter uncertainty exceeded the effect of other sources. Huss et al. (2014) investigated uncertainty in seasonal river flow projections over the 21st century for the Findelengletscher catchment, Switzerland by modifying the NCM-structure-GCM-RCM, GHM melt model structure and initial ice volume boundary condition. Of these, they found that the NCM-structure-GCM-RCM and initial ice volume were most important while the melt model structure was of secondary importance. Jobst et al. (2018) investigated uncertainties in 21st century river flows at flow projections for the Clutha river basin, New Zealand. They evaluated contributions from emission scenario (ES), NCM-structure-GCM-RCM, statistical DS approach and melt model structure. Similarly to Huss et al. (2014), they found that NCM-structure-uncertainty in the choice of GCM-RCM dominated total projection uncertainty.

The OAT method provides a useful-first-order approximation of the relative contribution of each component to the total projection uncertainty. However, findings are dependent on how the non-perturbed model components are fixed. Furthermore, this approach cannot resolve interactions between model components which may also contribute to projection uncertainty (Pianosi et al., 2016). An alternative approach that addresses these shortcomings is the The Analysis of Variance (ANOVA) statistical method (von Storch and Zwiers, 1999; Tabachnick and Fidell, 2014). ANOVA addresses these shortcomings and has been adopted in a number of recent large basin-scale and global-scale regional and global scale hydrological modelling studies (Bosshard et al., 2013; Addor et al., 2014; Giuntoli et al., 2015; Vetter et al., 2015; Samaniego et al., 2017; Vetter et al., 2017; Yuan et al., 2017) and used to compare uncertainties stemming from ES, NCM, GHM-climate model, hydrological model structure and DS approach. While uncertainties associated with future climate tend to dominate projections of future river flows, glaciated catchment-river flow, glacier-fed river flow projections have shown to be uniquely sensitive to GHM highly sensitive to hydrological model structure (Addor et al., 2014; Giuntoli et al., 2015), particularly in relation to high flows (Vetter et al., 2017). Furthermore, the contribution of projection uncertainty from interactions between model chain components can exceed individual components (Bosshard et al., 2013; Addor et al., 2014; Vetter et al., 2015). Several issues not considered in these studies, however, are yet to be addressed. Firstly, none have investigated a full range of characteristic changes in river flow regime covering decadal to sub-daily timescales. Second, all have incorporated GHM-hydrological model uncertainty using multiple GHM model codes, each with their own unique set of process representations, resolution, timestep and climate interpolation strategies making it difficult to determine which model components contribute most to projection uncertainty. Finally, none included a fully integrated mass-conserving, dynamic glacier evolution model component and therefore could not fully account for atmosphere-cryosphere-hydrosphere atmosphere-cryosphere-hydrosphere feedbacks.

This study uses a NCM-DS-GHM-GCM-RCM-DS-GHM model chain to simulate the impact of 21st century climate change on the downstream river flow regime in the deglaciating Virkísá river basin in southern Iceland. Five different components of the model chain are perturbed to represent uncertainty of future ES, NCM-structure ES, GCM-RCM, statistical DS parameterisation and structure-parameterisation of two primary controls on river flow regime in the GHM: melt and runoff-routing processes. The study has two principal aims: i) to determine how climate change and consequent cryospheric change will impact on
downstream river flow regime over the 21st century; and ii) to use ANOVA to quantify the relative influence of the five model chain components to projection uncertainty across the different characteristics of river flow regime. This study addresses each of the aforementioned gaps in previous work. Firstly, changes in river flow regime are assessed quantitatively using 25 river discharge signatures which define different characteristics of river flow regime over a range of timescales. Second, a single, consistent, GHM code is used that can incorporate different model structures and parameterisations of melt and runoff-routing processes allowing for uncertainty stemming from these to be localised using ANOVA. Finally, a fully integrated mass-conserving, dynamic glacier evolution routine is included in the GHM code.

2 Methodology

2.1 Study site

The Virkisá river basin covers an area of 22 km$^2$ on the western side of the ice-capped Öræfajökull stratovolcano in south-east Iceland (Figure 1) and forms a primary drainage channel for accumulating ice at the mountain summit (∼ 2000 m asl). The glacier flows in a south-westerly direction (average ice surface slope = 0.25) along two distinct glacier arms, Virkisjökull and Falljökull, (hereafter referred to as Virkisjökull) around a bedrock ridge before meeting again at the terminus (∼ 150 m asl). Virkisjökull currently covers ∼ 60% of the river basin area, but has been in a phase of retreat since 1990. Between 1990 and 2011 Virkisjökull lost ∼ 0.3 km$^3$ of ice and retreated ∼ 0.5 km. A small proglacial lake at the terminus forms the headwater of the Virkisá River. The Virkisá flows through an 800 m bedrock-controlled section flanked on either side by push moraines and then over the Skeiðarársandur floodplain typically comprising unconsolidated glacial outwash sediment. The steep-sided valley walls and glacial activity only allow for sporadic development of thin soils with limited vegetation including mosses, grass and shrubs.

The local climate is characterised by cool summers (∼ 10 $^\circ$C on average at the terminus AWS1) and mild winters (∼ 1 $^\circ$C on average at the terminus AWS1) with an average temperature lapse rate of -0.44 $^\circ$C 100 m$^{-1}$ (Flett, 2016). There is a significant lateral precipitation gradient due to the prevailing north-easterly winds and orographic effects with more than double-five-times the annual precipitation falling to the east of at the Óræfajökull summit (∼ 8000 mm yr$^{-1}$) than to the west (compared to lower down at the catchment outlet to the west (∼ 1500 mm yr$^{-1}$), an artefact of the prevailing north-easterly winds and orographic effects (Nawri et al., 2017).

2.2 Climate data

2.2.1 Historical climate

Historical climate data were available from 1981 to 2016 inclusive. A detailed description of these is are provided by Mackay et al. (2018). For brevity, only a summary of these data are provided here. The historical climate data include continuous hourly near-surface air temperature and incident solar radiation from automatic weather station 1 (AWS1) measurements from two automatic weather stations (AWS) in the catchment (Fig. 1c) which were installed in 2009 (AWS1) and 2011 (AWS4).
Figure 1. Location of Virkisá river basin in Iceland with glaciated areas highlighted in grey (a); on Öræfajökull (b); and detailed topographical map of basin including major land surface types and meteorological and stream gauging stations (c).
Temperature data from the nearby Icelandic Meteorological Office Fagurhólsmýri weather station (12 km south of the study site) were used to extend the AWS1 time-series back to 1981 using a linear regression model ($R^2=0.92$) to bias-correct against the AWS data. A seasonally variable hourly lapse rate calculated between AWS1 (156 m asl) and AWS4 (805 m asl) was used to extrapolate near-surface air temperature across the study region and an on-ice temperature correction function (Shea and Moore, 2010) was employed to account for katabatic cooling of air in the glacier valleys. Continuous hourly incident solar radiation data was also available from AWS1. A random resampling strategy that accounted for the dependence between intra-day solar radiation and temperature variability was employed to generate a continuous time-series back to 1981. For precipitation, the 2.5 km gridded hourly total precipitation data produced as part of the ICRA atmospheric reanalysis project were used, which are currently considered the most accurate gridded precipitation product over Iceland (Nawri et al., 2017). These were bias-corrected against the AWS precipitation data using the hourly precipitation measurements from AWS1 using equidistant quantile mapping approach (Li et al., 2010).

### 2.2.2 Regional climate projections

Future climate time-series for until 2100 were constructed using the regional climate projections from the Coordinated Regional Climate Downscaling Experiment (CORDEX) (Giorgi et al., 2009). The CORDEX projections are based on an ensemble of regional climate models (RCMs) driven by global circulation model (GCM) simulations from the latest RCMs driven by GCM projections from the Coupled Model Intercomparison Project (CMIP5) (Taylor et al., 2012). Iceland is covered by the EURO-CORDEX and ARCTIC-CORDEX regional model domains. Following the review by Gosseling (2017), the EURO-CORDEX data were used as these include projections at a higher 0.11° spatial resolution and a larger ensemble of GCM-RCM combinations allowing better exploration of climate model uncertainty.

The 0.11° EURO-CORDEX simulations span the years 1950-2100 with simulations up to 2005 constituting the ‘recent past’ where influences such as atmospheric composition, solar forcing and emissions are imposed based on observations. From 2006, three future ESs or Representative Concentration Pathways (RCPs) were imposed on the models including RCP2.6, RCP4.5 and RCP8.5 which represent an additional radiative forcing by 2100 relative to pre-industrial values of +2.6, +4.5 and +8.5 W m$^{-2}$ respectively. All simulations are available at 3-hourly to 3-monthly resolution, however the 3-hourly simulations were only produced using 4 GCM-RCMs while daily to seasonal simulations were produced using 15. Given the intent of this study to analyse projection uncertainty, it was decided that the daily data were most suitable. The RCP2.6 ES was omitted from the model experiments as only 8 of 15 GCM-RCMs within the CORDEX archive were used with this ES. Furthermore, the probability of achieving the RCP2.6 targets is increasingly unlikely (Sanford et al., 2014; Fyke and Matthews, 2015) and arguably completely infeasible (Mora et al., 2013) given the current global emission trajectory.

GCM-RCM skill was analysed by comparing the empirical distribution functions (ECDFs) of the daily catchment average historical observation climate data and the GCM-RCM simulations for the recent past (1981—2005). While the skill was variable across the 15 GCM-RCMs, one was consistently poor across all three climate variables (GCM:CNRM-CM5, RCM:CNRM-ALADIN53). Accordingly, this GCM-RCM was removed from the climate model ensemble given that it showed an extreme negative winter temperature bias and a consistently low skill when compared to daily observed
Figure 2. Comparison of daily-seasonal catchment-average observed and simulated near surface air temperature (14 GCM-RCMs) monthly ECDFs, incident solar radiation (SW) and total precipitation (P) between 1981 and 2005 for the 14 [GCM]-[RCM] used in this study. The top row shows the observed value and all subsequent rows indicate the GCM-RCM biases. The 1st percentile, mean and 99th are denoted by the subscripts 1, mean and 99 respectively. All statistics are calculated for the recent past (1981-2005) for the months of January-winter (DJF), April-spring (MAM), July-summer (JJA) and October-autumn (SON).

climate data (see Appendix A). Figure 2 shows the simulated ECDFs from seasonal bias of each of the 14 remaining GCM-RCMs for four months of the year (January, April, July and October). Overall, when compared to observations between 1981 and 2005. For temperature, the coldest days (T<sub>1</sub>) typically show a negative bias, particularly in winter, spring and autumn. Biases for T<sub>99</sub> are generally positive, but smaller in magnitude. The average absolute bias in mean seasonal temperature (T<sub>mean</sub>) is 1.4 °C, but the majority of GCM-RCMs show absolute biases <1.2 °C. Biases in seasonal incident solar radiation projections are almost exclusively positive with the largest biases associated with SW<sub>mean</sub> and SW<sub>99</sub>, particularly in spring and summer where they can exceed 100 W m<sup>-2</sup>. Total precipitation biases are typically largest in winter and autumn where proportionally, biases in P<sub>mean</sub> can exceed the magnitude of the observations (see SON for [EC-EARTH]-[HIRHAM5]). The largest biases however are seen in extremes (P<sub>99</sub>) which range from -86.9 to 77.5 mm d<sup>-1</sup>. While positive and negative precipitation biases are present throughout the ensemble, the sensitivity of precipitation simulations to the RCM is clear. For example, the GCM-RCM performance is good, broadly capturing the seasonal change in CCLM4-8-17 RCM has a systematic negative bias and the ECDFs, however some biases are notable. In particular, the highest precipitation events are generally overestimated and there is a clear positive bias in the incident solar radiation simulations HIRHAM5 RCM has a systematic positive bias. Conversely, temperature shows some negative bias, particularly for the coldest days in January and April.

### 2.2.3 Downscaling regional climate projections

The statistical delta-change downscaling approach was employed which has been widely applied in hydrological impact studies (Farinotti et al., 2012; Immerzeel et al., 2013; Kobierska et al., 2013; Huss et al., 2014; Lutz et al., 2016). While most studies have used monthly mean delta-change values to capture seasonal shifts in climate, several recent investigations have used advanced quantile-based approaches which account for changes in higher-order statistical properties of future climate by

<table>
<thead>
<tr>
<th>Observed</th>
<th>T&lt;sub&gt;1&lt;/sub&gt; (°C)</th>
<th>T&lt;sub&gt;mean&lt;/sub&gt; (°C)</th>
<th>T&lt;sub&gt;99&lt;/sub&gt; (°C)</th>
<th>SW&lt;sub&gt;1&lt;/sub&gt; (W m&lt;sup&gt;-2&lt;/sup&gt;)</th>
<th>SW&lt;sub&gt;mean&lt;/sub&gt; (W m&lt;sup&gt;-2&lt;/sup&gt;)</th>
<th>SW&lt;sub&gt;99&lt;/sub&gt; (W m&lt;sup&gt;-2&lt;/sup&gt;)</th>
<th>P&lt;sub&gt;1&lt;/sub&gt; (mm d&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>P&lt;sub&gt;mean&lt;/sub&gt; (mm d&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>P&lt;sub&gt;99&lt;/sub&gt; (mm d&lt;sup&gt;-1&lt;/sup&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CNR-MIM]C[CCLM4-8-17]</td>
<td>-12.1-10.5</td>
<td>-0.5</td>
<td>-4.6</td>
<td>4.7</td>
<td>0.3</td>
<td>3.1</td>
<td>8.6</td>
<td>5.9</td>
<td>12.2</td>
</tr>
<tr>
<td>[EC-EARTH]C[CCLM4-8-17]</td>
<td>-6.6</td>
<td>-0.5</td>
<td>-2.1</td>
<td>1.0</td>
<td>0.4</td>
<td>3.2</td>
<td>1.9</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>[HadGEM2-ES]C[CCLM4-8-17]</td>
<td>-6.6</td>
<td>-0.5</td>
<td>-2.1</td>
<td>1.0</td>
<td>0.4</td>
<td>3.2</td>
<td>1.9</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>[MPI-ESM-LR]C[CCLM4-8-17]</td>
<td>0.9</td>
<td>0.6</td>
<td>2.6</td>
<td>0.6</td>
<td>1.4</td>
<td>3.5</td>
<td>3.2</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>[CNR-MIM]C[RCAM4]</td>
<td>-9.5</td>
<td>-7.8</td>
<td>-4.1</td>
<td>-1.2</td>
<td>0.3</td>
<td>1.0</td>
<td>4.7</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>[EC-EARTH]C[RCAM4]</td>
<td>3.6</td>
<td>3.6</td>
<td>6.2</td>
<td>-0.6</td>
<td>1.1</td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>[MPI-ESM-LR]C[RCAM4]</td>
<td>0.2</td>
<td>0.2</td>
<td>-1.0</td>
<td>1.0</td>
<td>0.1</td>
<td>0.6</td>
<td>3.2</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>[CNR-MIM]C[RCAM4]</td>
<td>-9.5</td>
<td>-7.8</td>
<td>-4.1</td>
<td>-1.2</td>
<td>0.3</td>
<td>1.0</td>
<td>4.7</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>[EC-EARTH]C[RCAM4]</td>
<td>3.6</td>
<td>3.6</td>
<td>6.2</td>
<td>-0.6</td>
<td>1.1</td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>[MPI-ESM-LR]C[RCAM4]</td>
<td>0.2</td>
<td>0.2</td>
<td>-1.0</td>
<td>1.0</td>
<td>0.1</td>
<td>0.6</td>
<td>3.2</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>[CNR-MIM]C[RCAM4]</td>
<td>-9.5</td>
<td>-7.8</td>
<td>-4.1</td>
<td>-1.2</td>
<td>0.3</td>
<td>1.0</td>
<td>4.7</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>[EC-EARTH]C[RCAM4]</td>
<td>3.6</td>
<td>3.6</td>
<td>6.2</td>
<td>-0.6</td>
<td>1.1</td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>[MPI-ESM-LR]C[RCAM4]</td>
<td>0.2</td>
<td>0.2</td>
<td>-1.0</td>
<td>1.0</td>
<td>0.1</td>
<td>0.6</td>
<td>3.2</td>
<td>3.2</td>
<td>3.2</td>
</tr>
</tbody>
</table>

8
evaluating shifts in the ECDFs of climate variables. Including these higher-order changes has shown to be important for evaluating shifts in extreme high flows and sub-seasonal metrics of river flow projections (Jakob Themeßl et al., 2011; Immerzeel et al., 2013; Lutz et al., 2016). In addition, shifts in the day-to-day variability of temperature impact projections of glacier retreat as these variations control the periodic rising of temperature above the melting point (Beer et al., 2018). Accordingly, the advanced delta-change approach was adopted in this study. The approach is summarised in five steps which were applied to each combination of GCM-RCM, climate variable and ES separately:

1. The climate variable time-series was divided into four 25-year long periods including the recent past (1981 - 2005) and early (2006 - 2030), mid (2041-2065) and late (2076 - 2100) 21st century.

2. For each of the four periods, all daily data points were further divided into 12 sub-samples representing each month of the year. An ECDF was constructed for each month of each period.

3. For each month of each future period, ten deltas were calculated by taking the mean difference between the recent past and future ECDF for each 10% section (see grey bars in Fig. 3a for example).

4. Given the need for transient climate time-series to simulate glacier evolution over the 21st century, a daily delta time-series from 2006 to 2100 was constructed for each ECDF section of each month by linearly interpolating between the calculated deltas of each future period (e.g. as implemented by Farinotti et al., 2012), using the midpoints of the future periods as interpolation points (Figure 3b).

5. The hourly historic observation data for the recent past were randomly sampled (with replacement) on a year-by-year basis to generate an initial unperturbed future climate variable time-series (blue dash, Fig. 3c). The daily deltas were applied to this time-series for each month and ECDF section separately to generate a future perturbed climate time-series at an hourly resolution (orange dash Fig. 3c). It was noted upon visual inspection, that the inter-annual variability of the future climate time-series was very sensitive to the random sampling of the historic climate data. Accordingly, uncertainty associated with this aspect of the DS parameterisation was considered by using ten different random historic climate samples.

For temperature, catchment average daily deltas were applied evenly across the catchment and each daily period of the unperturbed time-series. Accordingly, diurnal temperature variability and lapse rates were assumed not to change in the future. For incident solar radiation and precipitation, proportional deltas were used to prevent negative values and preserve the sub-daily proportional distribution of these variables in space and time.

It was noted upon visual inspection, that the inter-annual variability of the future climate time-series was very sensitive to the random sampling of the historic climate data (step 5 above). Accordingly, uncertainty associated with this aspect of the DS parameterisation was considered by using ten different random historic climate samples. As such a total of 2 ES × 14 GCM-RCM × 10 DS parameterisations = 280 future climate time-series were generated for this study.
Figure 3. Example of advanced delta-change approach when applied to near surface air temperature data based on the projections from 1 of the 14 GCM-RCMs with the RCP8.5 ES projections using the CNRM-CM5 GCM and CCLM4-8-17 RCM. Deltas (grey bars) derived from ECDFs (black curves) for April in late 21st century (a); Daily delta time-series for each section of the April ECDFs (green line represents 40th - 50th percentile section) (b); Initial and perturbed future temperature time-series when deltas for all months and ECDF sections are applied (c).

2.3 Glacio-hydrological model

The distributed GHM code implemented by Mackay et al. (2018) was used in this study because it includes a dynamic, mass-conserving glacier evolution component and also allows the user to utilise different model structures of melt and runoff-routing processes. The GHM resolves glacio-hydrological processes over a regular 2D Cartesian grid of nodes (50 m cells) driven by hourly climate data including precipitation, temperature and incident solar radiation. Empirical ‘index-based’ equations simulate the melt of snow and ice. Snow redistribution by drift and avalanches is calculated using the curvature and slope of the surface (Huss et al., 2008) while the mass-conserving \( \Delta h \) parametrisation of glacier retreat (Huss et al., 2010) resolves changes in the glacier geometry. A soil infiltration and evapotranspiration model (Griffiths et al., 2006) (Griffiths et al., 2008) based on the well-established FAO56 soil moisture accounting procedure (Allen et al., 1998) solves the water balance for model nodes with no ice or snow coverage. This model has been applied extensively (Mackay et al., 2014, 2015; Jackson et al., 2016; Mansour et al., 2018) and has shown to compare favourably to physically-based models at the field scale where interception losses are small (Sorensen et al., 2014). Excess soil moisture, rainfall and melt are routed to the catchment outlet via a semi-distributed network.
of linear-reservoir cascades (Ponce, 1989) which represent the average water storage and release characteristics of the major hydrological pathways in the watershed (see Fig. 2 in Mackay et al., 2018).

2.3.1 Modification to Δh parameterisation of glacier retreat

Under periods of sustained positive mass balance, simulations from the Δh glacier evolution model may result in an unrealistic build up of ice at the glacier tongue without any simulated areal advance. Given the potential for periods of glacier advance under a changing climate, such behaviour is likely to result in significant projection biases. Recently, Seibert et al. (2018) presented an implementation of the original Δh parameterisation that provides more realistic simulations of glacier advance. They propose running the Δh parametrisation a priori outside of the GHM. A small negative mass balance is used to force the Δh model from an initial glacier profile (ideally its maximum observed extent) until the glacier has disappeared completely.

At each step, the glacier mass and geometry are stored in the form of a lookup table. On running the GHM, the retreat/advance of the glacier is derived from the lookup table as a function of the simulated glacier mass. One important drawback of using this static lookup table is that it modifies the behaviour of the Δh formulation during periods of retreat. More specifically, this approach neglects the transient annual sequencing of glacier mass balance which influences simulated glacier geometry due to the non-linear structure of the Δh polynomial that defines the relationship between mass balance and glacier geometry.

Accordingly, a modified implementation of the Seibert et al. (2018) approach was used in this study which behaves like the original Δh formulation during periods of glacier retreat and allows for the simulation of glacier advance while accounting for mass balance sequencing effects on the model behaviour. For periods of negative glacier mass balance the original Δh formulation was used. The GHM was then modified so that for each simulation year, the simulated glacier mass and geometry were stored in memory. If a positive glacier mass balance (ΔM) was simulated, the GHM would log the current glacier mass (M_{current}) and then look for the most recent historical simulated glacier mass (M_{hist}) that exceeded M_{current} + ΔM. The Δh model was then run with a negative mass balance (ΔM^*) so that M_{hist} + ΔM^* = M_{current} + ΔM.

2.3.2 Melt and runoff-routing model structures

The selection of melt and runoff-routing model structures was based on the findings of Mackay et al. (2018). They applied nine different combinations of three melt model structures and three runoff-routing model structures of varied complexity in the GHM and evaluated their ability to capture a range of river discharge signatures derived from observation time-series at automatic stream gauge 1 (ASG1 in Fig. 1c) which has been operational since 2012. They also used observation data of ice melt and snow coverage to derive signatures that described different aspects of these data (see Appendix B). They showed that while introducing model complexity did improve simulations when evaluated against specific signatures, it did not necessarily result in better consistency across all signatures, emphasising model selection uncertainty. The most complex runoff-routing structure, however, was consistently the least efficient when compared to the two simpler alternatives, particularly in relation to capturing signatures representing high river flow events. As such, this model structure was discarded, and only the remaining six combinations of melt and runoff-routing models structures were used in this study. These included three melt model structures: i) the classic temperature index model (TIM1) where melt increases linearly with near surface air temperature above a critical
threshold (e.g. Braithwaite, 1995); ii) the enhanced temperature-index model (TIM$_2$) proposed by Hock (1999) which accounts for topographic effects on incident solar radiation including surface slope, aspect and shading from the surrounding landform; and iii) the enhanced temperature-index model (TIM$_3$) proposed by Pellicciotti et al. (2005) which accounts for topographic effects and also includes a dynamic snow-albedo parameterisation (Brock et al., 2000) which accounts for the drop in snow albedo as it ages. Each melt model structure was combined with the two runoff-routing structures: i) a single linear reservoir cascade (ROR$_1$) which routes runoff from all sources (ice melt, snowmelt, rainfall and excess soil water) simultaneously; and ii) two linear reservoir cascades in parallel (ROR$_2$) where the first represents the slow percolation of water through the snow and firn and the second represents faster flow of water through and over bare ice and overland. The simplest ROR$_1$ structure assumes all catchment water stores delay and diffuse downstream river response to runoff in the same way, effectively fixing the run-off routing behaviour of the catchment over time. The more complex ROR$_2$ structure accounts for temporal variations in the drainage efficiency of the catchment according to changes in snow and ice coverage.

2.4 Signatures of river flow regime

Table 1 lists the 25 signatures of river discharge used to evaluate future changes in river flow regime. The majority of signatures were selected from past studies (Yadav et al., 2007; Yilmaz et al., 2008; Shafii and Tolson, 2015; Schaeffli, 2016) and were chosen to reflect the types of changes that one might expect to see in snow and ice covered catchments. They also broadly follow those used in the model assessment study of Mackay et al. (2018). The signatures are grouped into seven different attributes and further categorised by the characteristic(s) of flow regime that they evaluate and their temporal scale. At the decadal timescale, two signatures were selected. These include the ‘peak water’, which defines the timing (by year) of maximum flow, as well as the inter-annual flow range which characterises long term flow variability. Changes in mean annual river flow were also evaluated, while mean monthly flows were used to evaluate changes to the seasonal timing and magnitude of river flow. The range in mean monthly flows was also chosen to evaluate intra-annual flow variability. In addition, eight signatures were selected which broadly describe the magnitude and variability of slow release low flows (99-95% exceedance flows), moderate flows (52-48% exceedance) and quick release high flows (5-1% exceedance). For these, the quantiles of the FDC were used to assess changes in the magnitude of these flow types. The standard deviation was also used to define flow variability of each flow type. Finally, the integral scale, which measures the lag time at which the autocorrelation function of the river flow time-series falls below $\frac{1}{e}$ was utilised as an indicator of the response time of the catchment to runoff events (flashiness).

2.5 GHM calibration

Given the focus on projecting changes in river discharge signatures, these were explicitly included in the GHM calibration procedure as this gives better signature simulations than using traditional global objective functions (Kiesel et al., 2017; Pool et al., 2017). Calibrating against river flow data alone can lead to unrealistic snow and glacier melt rates, inhibiting model consistency and increasing projection uncertainties (Konz and Seibert, 2010; Finger et al., 2011; Schaeffli and Huss, 2011; Hanzer et al., 2016). Accordingly, a novel signature-based calibration of the GHM was undertaken by evaluating the GHM against 20 of the river discharge signatures in Table 1 for which observation data exists calculated from hourly river discharge
measurements (Macdonald et al., 2016) at the automatic stream gauge (ASG1 in Fig. 1) in combination with 12 signatures of ice melt and snow coverage (Appendix B).

For each signature, model simulations were compared to observations using a continuous acceptability score that is analogous to those used in other signature-based hydrological studies (Coxon et al., 2014; Shafii and Tolson, 2015). This objective function explicitly accounts for uncertainty in the observation signatures, hereafter termed ‘limits of acceptability’ (LOA), so that decisions about model appropriateness can be made within the uncertainties of observation data. In this study the 95% confidence bounds were used to define the LOA for the river discharge signatures (Table 1) and the ice melt and snow coverage signatures (Table B1). Details of how these were derived can be found in the study of Mackay et al. (2018). The acceptability for signature \( j \) is defined as:

\[
s_j = \begin{cases} 
  0 & \text{low}_j \leq \text{sim}_j \leq \text{upp}_j \\
  \frac{\text{sim}_j - \text{low}_j}{\text{upp}_j - \text{obs}_j} & \text{low}_j > \text{upp}_j \\
  \frac{\text{upp}_j - \text{sim}_j}{\text{obs}_j - \text{low}_j} & \text{sim}_j < \text{low}_j 
\end{cases}
\]

(1)
where \( obs_j \) and \( sim_j \) are the observed and simulated values and \( upp_j \) and \( low_j \) are the upper and lower LOA. A score of zero indicates that the model captures the signature within the LOA. A non-zero score is given for any simulation that falls outside of the LOA with a sign that indicates the direction of bias and a magnitude that indicates the model’s performance relative to the LOA. A score of -3 would indicate that the model underestimates the signature by three times the observation uncertainty. This score therefore does not penalise a model if it falls within the observation uncertainty of a signature. It is also tolerant of projections that fall outside of the LOA where observation uncertainty is high; a desirable attribute given the range of signatures the GHM was evaluated against.

The aim of the calibration was to extract an ensemble of GHM compositions (TIM and ROR structure-parameter combinations) that were most acceptable across the river discharge signatures whilst broadly reproducing the snow coverage and ice melt signatures. This was achieved using a three-stage two-stage Monte-Carlo procedure which was devised so that the resultant GHM ensemble reflected the uncertainty in model selection given the known inconsistencies of the GHM across the signatures.

### 2.5.1 Stage 1: TIM calibration

The first stage aimed to extract the optimal TIM compositions (structure-parameter combinations) by calibrating them against the 12 snow coverage and ice melt signatures. Here, for each of the three TIM structures, 5000 TIM parameter sets were drawn from pre-defined uniform distributions (Table C1) using the quasi-random Sobol sampling strategy (Brately and Fox, 1988) to sample the parameter space as efficiently as possible. For each parameter set, the GHM was spun-up for three years from 1985 to 1988 with a static ice geometry fixed to a 1988 ice DEM (Magnússon et al., 2016). The GHM was then run from 1988 to the end of 2016 with a freely evolving glacier geometry.

Given the high degree of glaciation in the study catchment, and its recent rapid retreat, an initial emphasis of the calibration was put on the model’s ability to capture the long term glacier volume change signature. Accordingly, only those TIM compositions that captured this signature within the LOA were considered and the rest were discarded. These remaining compositions were then further refined by evaluating them against the remaining 11 snow and ice signatures. First, the TIM compositions were sorted by structure (TIM\(_1\), TIM\(_2\), TIM\(_3\)). Then, for a given TIM structure, the following steps were applied:

1. Find the TIM parameter set(s) that capture the signature within the LOA and discard the rest. If more than one parameter set captures the current signature, go to step 2. If none capture the current signature, discard none and go to step 2.

2. Of the remaining models, find that which best captures the remaining snow and ice signatures overall according to the weighted mean scores obtained in Eq. 1. The weights were applied to ensure that equal preference was given to ice melt and snow coverage signatures.

24 unique TIM compositions were obtained from this calibration stage made up of eight unique parameterisations of each of the three TIM structures. In some cases the same composition was selected more than once which was accounted for by weighting the simulations in the results presented throughout this study.
2.5.2 Stage 2: ROR calibration

The second calibration stage aimed to extract the optimal ROR compositions when used in combination with the 24 pre-selected TIM compositions by calibrating them against 20 of the river discharge signatures obtained from observations of river discharge for the years 2013 and 2014 (see signatures with calibration LOA in Table 1). Note, the inter-annual flow signatures and the mean December river flow signatures could not be calculated as there was insufficient observation data. Furthermore, the mean annual river flow and mean monthly flow range were not included as this information was already accounted for in the mean monthly flow signatures. Here, 5000 random ROR parameter sets were drawn for each ROR structure. Each was used in combination with the pre-selected TIM compositions in the GHM. Then, the two steps outlined in calibration stage 1 were applied using the 20 calibration river discharge signatures with two notable differences. Firstly, for each ROR structure and each river discharge signature, rather than selecting a unique ROR parameter set for each of the 24 TIM compositions, a single parameter set was selected based on its mean performance across the 24 TIM compositions. This was done to satisfy the ANOVA requirements so that the TIM and ROR composition uncertainty could be analysed separately. Furthermore, for step 2, the signatures were weighted so that each of the attributes in Table 1 were weighted equally. In total, 14 unique ROR compositions were selected made up of seven unique parameterisations of the ROR\textsubscript{1} and ROR\textsubscript{2} structures, giving a total of $24 \times 14 = 336$ unique GHM compositions.

2.5.3 Stage 3: Evaluation of calibrated GHM compositions

After calibration, the simulated river discharge time-series and signatures were evaluated against river discharge observations covering the years 2015 and 2016. Note, no data for mean January and February flows were available for these years. Figures 4a and b show the simulated ‘capture ratio’ (the ratio of the 336 GHM compositions that capture the observation data within their 95\% uncertainty bounds) time-series projected onto the mean observed river discharge for the years 2015 and 2016 respectively. Also shown is the ensemble mean simulated river discharge (black dash) which while not indicative of a single GHM simulation, does provide an indication of overall projection bias.

56\% of the observation time-series were captured by at least half of the GHM compositions, while 41\% and 28\% of the observations were captured by at least 75\% and 90\% of the GHM compositions. 12\% of the observations were not captured by any of the GHM compositions. These included some of the low flows observed at the beginning of the year outside of the melt season, particularly in 2015, where the GHM showed consistent negative biases. Some rainfall induced summer peak flows were also not captured, particularly during the late summer months of August and September. Furthermore, the sustained summer melt runoff discharge in between rainfall induced peak flows tended to be overestimated (for example during July and August 2016). Even so, the flow duration curve (FDC) in Fig. 4c shows that almost the entire FDC was captured by all of the GHM simulations except for some of the lowest flows on record. Indeed, Fig. 4d reveals that GHMs were least efficient at capturing the low flow signatures, particularly the variability signature ($\sigma_{99-95}$), where simulations were positively biased by almost four times the observation uncertainty. For the remaining signatures though, the ensemble of GHMs were remarkably efficient, with the majority of simulations (and in most cases all of them) capturing these signatures within their
2.6 21st-century projections

For the 21st century runs, all 336 GHM compositions were run to the end of 2016 using the historic observed climate to capture the evolving ice geometry as accurately as possible. From 2017 to 2100, the 280 downscaled future climate time-series were used to drive the GHM compositions resulting in 94080 individual model chain runs. Prior to running the full ensemble, a preliminary subset of 2000 runs from the ensemble was run for the 21st century to ensure the model chain was behaving as expected. From this it was found that ~5% of the runs simulated sustained positive glacier mass balance over the future period. Given that the $\Delta h$ glacier evolution parametrisation is not designed to simulate glacier advance, these simulations resulted in an unrealistic build-up of ice at the glacier tongue without any simulated areal advance of the glacier. Recently, Seibert et al. (2018) presented an implementation of the original $\Delta h$ parameterisation that provides more realistic simulations of glacier evolution under periods of sustained positive mass balance. They propose running the $\Delta h$ parametrisation a priori outside of the GHM. A small negative mass balance is used to force the $\Delta h$ model from an initial glacier profile (ideally its maximum observed extent) until the glacier has disappeared completely. At each step, the glacier mass and geometry are stored in the form of a lookup table. On running the GHM, the retreat/advance of the glacier is derived from the lookup table as a function of the simulated glacier mass.

Seibert et al. (2018) note that this approach neglects any delays in the response of glacier advance to mass balance changes and is also limited to advancing the glacier as far as the maximum observed glacier extent. It should also be noted that the use of a ‘static’ lookup table modifies the behaviour of the $\Delta h$ formulation during periods of retreat. In the original formulation three consecutive years of $\pm 1$ m w.e. surface mass balance would not necessarily produce the same glacier geometry as a single year of $\pm 3$ m w.e. for example. This is because of the non-linear structure of the $\Delta h$ polynomial that defines the relationship between mass balance and glacier geometry. The simulated glacier geometry after a period of time is thus dependent on the cumulative mass balance and the annual sequencing of mass balance. The approach proposed by Seibert et al. (2018) only accounts for the former. Accordingly, a modified implementation of the Seibert et al. (2018) approach was used in this study. Here, the original $\Delta h$ formulation was used for periods of negative glacier mass balance. The GHM was then modified so that for each simulation year, the simulated glacier mass and geometry were stored in memory. If a positive glacier mass balance ($\Delta M$) was simulated, the GHM would log the current glacier mass ($M_{\text{current}}$) and then look for the most recent historical simulated glacier mass ($M_{\text{hist}}$) that exceeded $M_{\text{current}} + \Delta M$. The $\Delta h$ model was then run with a negative mass balance ($\Delta M^*$) so that $M_{\text{hist}} + \Delta M^* = M_{\text{current}} + \Delta M$. This approach therefore behaves like the original $\Delta h$ formulation during
periods of glacier retreat and allows for the simulation of glacier advance while accounting for mass balance sequencing effects on the model behaviour.

On running the full ensemble, it was found that <2% of the simulations showed periods of glacier advance of more than five consecutive years.

5 2.7 ANOVA uncertainty analysis

runs. For each model chain-run, projections of watershed snow and ice coverage and the 25 river discharge signatures were extracted for each decade from the 1990s (1991–2000) to the 2090s (2091–2100) six 21st century 25-year time-slices centred on the 2030s (2023–2047), 2040s (2033–2057), 2050s (2043–2067), 2060s (2053–2077), 2070s (2063–2087) and 2080s (2073–2097). Future changes in these were then calculated using the 1990s as the reference period and the eight decades influenced by future climate change (2020s to 2090s) as the change periods. The 1990s was chosen as the reference period given that the glacier extent was near its maximum and relatively stable over this period relative to a reference 25-year period (1991–2015). This reference period was chosen because ice-coverage data (used to initialise the GHM) were only available from 1988 and historic climate data were available up to the end of 2016. ANOVA was used to quantify the effect size of the five components of the model chain, hereafter termed factors, on each signature for each 21st century time-slice. Note, the peak water (PW) signature can only be calculated taking into account the full projection time-series and, as such, it was not possible to apply ANOVA to decadal metrics of change each time-slice for this signature. ANOVA was used to quantify the effect size of the five components of the model chain, hereafter termed factors, on each decadal projection of change. The five factors include the 2×future ES, 14×GCM–RCM combinations, 10×DS parameterisations, 24×TIM compositions and 14×ROR compositions. ANOVA offers an intuitive approach to estimate the effect size of each factor on each signature by partitioning the total sum of squares ($SS_{tot}$) in the response variable over all combinations of factor levels:

$$SS_{tot} = SS_a + SS_b + SS_c + SS_d + SS_e + SS_I + SS_e$$

where:

$$SS_{tot} = \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \sum_{k=1}^{n_c} \sum_{l=1}^{n_d} \sum_{m=1}^{n_e} (y_{i,j,k,l,m} - \bar{Y})^2$$

where $n_a$, $n_b$, $n_c$, $n_d$ and $n_e$ are the number of levels for each factor, $y$ is the response for a given treatment (i.e. combination of factor levels) and $\bar{Y}$ is the grand mean of the response variable over all treatments. $SS_a$, $SS_b$, $SS_c$, $SS_d$ and $SS_e$ in Eq. 2 are the sum of squares due to the main effects, i.e. the variability in the response variable due to varying a given factor on its own. For example:

$$SS_a = n_b n_c n_d n_e \sum_{i=1}^{n_a} (y_{i,0,0,0,0} - \bar{Y})^2$$

where 0 indicates averaging over an index. $SS_I$ includes all non-additive interaction terms where the combined effect of two or more factors is not the sum of their main effects. For a 5-factor ANOVA, one could include all unique $n$-tuple combinations of factors where $n = (2,3,4,5)$. Given the difficulty in interpreting these higher-order interactions, and computational
requirements, it was decided to investigate the nine first-order interactions only, so that:

$$SS_I = SS_{ab} + SS_{ac} + SS_{ad} + SS_{ae} + SS_{bc} + SS_{bd} + SS_{be} + SS_{cd} + SS_{ce} + SS_{de}$$  \hspace{1cm} (5)

The sum of squares for a first-order interaction are calculated as follows using factors $a$ and $b$ as an example:

$$SS_{ab} = n_c n_d n_e \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} (y_{i,j,o,o,o} - y_{i,o,o,o,o} - y_{o,j,o,o,o} + \bar{Y})^2$$  \hspace{1cm} (6)

Finally, the $SS_e$ term includes all unexplained variance i.e. error in the ANOVA model.

Having partitioned the sum of squares, the effect size, $\eta^2$ for any term in Eq. 3 can be taken as the proportion of the total sum of squares:

$$\eta^2 = SS_e / SS_{tot}$$  \hspace{1cm} (7)

where * can be any of the main effects, interactions or error term.

Bosshard et al. (2013) showed that because ANOVA is based on a biased variance estimator that underestimates the variance in small sample sizes, the calculated effect sizes are biased if a different number of levels are used for each factor. Given that the number of factor levels range from 2 to 24, a pure application of ANOVA using all possible treatments would lead to biased results. Bosshard et al. (2013) outlined a method to correct for this which involves sub-sampling the factor levels down to the smallest number levels across all factors. The procedure is repeated using every possible combination of factor levels with unbiased effect size taken as the mean across all sub-samples. However, given that there are $C(94080, 2) \approx 4 \times 10^5 \gg 10^8$ unique combinations of factor levels when sub-sampled down to two (and discarding factor level repetitions), it would have been infeasible to account for every possible combination. Instead, it was decided to calculate the effect sizes in this manner using five different sub-sample sizes ($10^1, 10^2 \ldots 10^5$). The results were then analysed to see if the effect sizes converged. It was found that $10^3$ sub-samples were sufficient to converge the effect sizes for all river discharge signatures and projections of snow and ice coverage. Accordingly, this sub-sampling strategy was adopted in this study.

3 Results

3.1 Evaluation of calibrated GHM compositions

The simulated river discharge time-series and signatures using the calibrated GHM compositions were evaluated against river discharge observations covering the years 2015 and 2016. Note, no data for mean January and February flows were available for these years. Figures 4a and b show the simulated ‘capture ratio’ (the ratio of the 336 GHM compositions that capture the observation data within their 95% uncertainty bounds) time-series projected onto the mean observed river discharge for the years 2015 and 2016 respectively. Also shown is the ensemble mean simulated river discharge (black dash) which while not indicative of a single GHM simulation, does provide an indication of overall projection bias.

56% of the observation time-series were captured by at least half of the GHM compositions, while 41% and 28% of the observations were captured by at least 75% and 90% of the GHM compositions. 12% of the observations were not captured
by any of the GHM compositions. These included some of the low flows observed at the beginning of the year outside of the melt season, particularly in 2015, where the GHM showed consistent negative biases. Some rainfall-induced summer peak flows were also not captured, particularly during the late summer months of August and September. Furthermore, the sustained summer melt runoff discharge in between rainfall-induced peak flows tended to be overestimated (for example during July and August 2016). Even so, the flow duration curve (FDC) in Fig. 4c shows that almost the entire FDC was captured by all of the GHM simulations except for some of the lowest flows on record. Indeed, Fig. 4d reveals that GHMs were least efficient at capturing the low flow signatures, particularly the variability signature ($\sigma_{95-95}$), where simulations were positively biased by almost four times the observation uncertainty. For the remaining signatures though, the ensemble of GHMs were remarkably efficient, with the majority of simulations (and in most cases all of them) capturing these signatures within their 95% observation uncertainty bounds.

### 3.2 Future climate projections

Projections of temperature for the late 21st century (2076-2100) consistently show an increase relative to the recent past (1981-2005). The largest increases are predicted projected for the coldest days of the year during the winter (Figure 5a), spring (Figure 5d) and autumn (Figure 5j) months as shown by the positive skew in the lower sections of the ECDFs. However, these changes are also typically associated with the greatest projection uncertainty. The future climate ensemble projects RCP4.5 projects annual mean near surface air temperature to rise by between 0.8 and 1.1 and 3.6°C by the late 21st century relative to the recent past with an ensemble mean projection of +2.7°C. RCP8.5 projects an equivalent rise of between 2.3 and 4.9°C with an ensemble mean projection of +3.3°C.

Projected changes in incident solar radiation span positive and negative values, but the median projections are consistently negative indicating reductions in incident solar radiation are most likely. Uncertainty in the magnitude of change is highest during the spring and summer months (Figures 5e and h) when incident solar radiation peaks. By late 21st century, Under RCP4.5 annual mean incident solar radiation is projected to change by between -18.5 and 10.7 to +13.0% -0.8% by the late 21st century with an ensemble mean projection of -6.3% - 4.4%. Under RCP8.5 changes of between -15.3 to 0.4% are projected with an ensemble mean projection of -7.7%.

Projected changes in total precipitation are negligible for the four lowest 10% sections of the precipitation ECDFs, but significant for the two highest sections. In the winter (Figure 5c) and autumn (Figure 5l) months, absolute changes exceed 40 mm d⁻¹. The sign direction of change is uncertain apart from autumn where median projections are consistently positive for the upper sections of the ECDF. The magnitude of change is also uncertain and by the late 21st century annual mean precipitation will change by anywhere between -42.8 between -13.5 to +45.5 21.6% relative to the recent past with an ensemble by the late 21st century with an ensemble mean projection of +1.2% - 1.7%. Under RCP8.5 changes of between -25.7 to 25.1% are projected with an ensemble mean projection of +1.4%.

Figure 6 shows the correlation matrix calculated between seasonal average climate variables for late 21st century. For all climate variables, between-season changes (scores within green borders in Fig. 6) are positively correlated indicating that an increase in summer temperature typically corresponds with an increase in winter temperature for example. Temperature
Figure 4. Capture ratio projected onto observed river discharge data during evaluation period for 2015 (a); 2016 (b); and over the FDC (c). The weighted ensemble mean simulation is shown as a black dash. Also shown are the range of acceptability scores for each of the available river discharge signatures over the evaluation period (d). Acceptable simulations in (d) are those contained within the black dash lines.
Figure 5. Seasonal average projected changes in ECDFs for near surface air temperature (a,d,g,j), incident solar radiation (b,e,h,k) and total precipitation (c,f,i,l) for the late 21st century (2076-2100) relative to the recent past (1981-2005). Changes are plotted for each 10% section of the ECDFs. For each section, blue and yellow dots represent each of the 140 downscaled future climate time-series for the RCP4.5 and RCP8.5 ES respectively (280 in total). Winter = Dec, Jan, Feb; Spring = Mar, Apr, May; Summer = Jun, Jul, Aug; Autumn = Sep, Oct, Nov.
Figure 6. Correlation matrix between seasonal average climate variables calculated for late 21st Century (2076-2100) using the 280 downscaled future climate time-series. Within-variable, between-season correlation scores are contained within the green borders and within-season, between-variable correlation scores are contained within the purple borders. Those regions of the correlation matrix that do not cover these two groups are shaded in black.

has the greatest between-season correlation while precipitation is the least well correlated. Within-season, between-variable correlation scores (within purple border in Fig. 6) show that precipitation and incident solar radiation are negatively correlated and that the correlation between precipitation and temperature depends on the time of year. For the cooler winter, spring and autumn months, temperature and precipitation are positively correlated, but there is a weak negative correlation for the summer months. Temperature and incident solar radiation are negatively correlated, most strongly for the cooler winter, spring and autumn months.

3.3 Future evolution of snow and ice coverage

The ensemble mean projection of annual mean watershed snow coverage (yellow line in Fig. 7a) indicates that it will decrease from 12.2 km² in 2016 to 7.6 km² in 2100 (38% reduction) under RCP4.5 and 6.0 km² (51% reduction) under RCP8.5. The 95% projection confidence intervals indicate that by 2100 the watershed could
Figure 7. Projected annual mean watershed snow coverage (a) and ice coverage (b) including the projection confidence intervals (blue bands) and ensemble mean projections (yellow thick solid lines) for the RCP4.5 (blue) and RCP8.5 (yellow) projections. Also shown are projection confidence levels for a reduction in coverage relative to 2000 and 2016 (red thin solid lines, right hand axis).

be almost entirely free of snow (2.9-2.5 km² remaining) under RCP8.5 or could have a coverage equivalent to exceeding present levels (12.5-13.3 km²) under RCP4.5.

Beyond 2050, there is high confidence (≥ 95%) that snow coverage will reduce relative to 2000 levels throughout the 21st century (solid red line - 2016 levels under RCP8.5 (thin yellow line in Fig. 7a). Equally and equally high levels of projection confidence apply to projected reductions in snow coverage beyond 2050 relative to 2016 (dashed red line - 2066 under the cooler RCP4.5 (thin blue line in Fig. 7a).

The ensemble mean projection of ice coverage (Figure 7b) predicts a 31% reduction relative to 2016 by 2100 under RCP4.5 and a more severe 63% reduction under RCP8.5. There is high confidence (≥95%) that ice coverage will be less than 2000 and 2016 levels through the remainder of the 21st century, from 2037 onwards under RCP4.5 and from 2030 onwards under RCP8.5 but the magnitude of change is uncertain under both emission scenarios. By 2100, the 95% confidence interval spans 2.3 km² to 11.3 km² coverage (12-82% reduction relative to intervals for both RCP4.5 and RCP8.5 are 6.5 km² wide (more than half the 2016 watershed ice coverage).

The simulations that projected the mean and minimum ice coverage by 2100 both project a under the RCP8.5 emission scenario shows sustained retreat of the glacier between 2000 and 2100 (Figure 8). The minimum ice coverage simulation projects that the watershed will be resulting in a watershed that is almost entirely ice free by the end of the century (Figure 8). In contrast, the maximum ice coverage simulation under the RCP4.5 emission scenario projects two periods of
Figure 8. Simulated ice thickness between 2000 and 2100 based on simulations that projected the maximum (RCP4.5) and minimum (RCP8.5) ice coverage by 2100. Watershed outline shown in magenta.

Glacier advance between 2010 and 2030 and between 2060 and 2100. By the end of the century, this simulation projects ice coverage will be similar to that in 2000.

Figures 9a, b and c show the climate projection time-series that produced the minimum (dotted lines) and maximum (dashed lines) snow (blue lines) and ice (red lines) coverage by 2100. The minimum coverage simulations were forced with some of the highest temperature time-series while the maximum coverage simulations were forced with some of the lowest. The maximum coverage simulations show higher than average incident solar radiation inputs (Figure 9b) and lower precipitation volumes than the minimum coverage simulations. Indeed, correlation scores calculated between seasonal average climate variables and the simulated snow and ice coverage by 2100 (Figure 9d) show that there is a strong negative correlation between mean temperature and projected snow and ice coverage and a weaker positive correlation between snow and ice coverage and incident solar radiation. An even weaker negative correlation exists between autumn and winter precipitation and snow and ice coverage.

3.4 Sources of uncertainty in snow and ice coverage projections

The effect size of the main, interaction and error terms calculated using ANOVA for projected changes in decadal snow and ice coverage are shown in Fig. 10. Note, ROR effects are not included here as this model chain component has no influence on cryospheric processes in the GHM. The effect size of each ANOVA term changes through the decades and also varies between snow and ice coverage. Throughout the 21st century, TIM uncertainty contributes <3% to the total projection uncertainty of snow coverage. For projections of ice coverage, $\eta_{TIM}^2 > 0.12$ for the first half of the 21st century, 0.11 up to and including the 2060s, but then gradually falls to 0.06 by the 2090s. $\eta_{DS}^2$ and $\eta_I^2$ never exceed 0.2–0.1 for snow and ice coverage and as with $\eta_{TIM}^2$, they gradually reduce through the latter half of the 21st century. GCM-RCM uncertainty is the
Figure 9. Relationship between driving climate data and projected snow and ice coverage including annual mean downscaled climate time-series of temperature (a), incident solar radiation (b) and total precipitation (c) with time-series that produced the minimum (dotted lines) and maximum (dashed lines) snow and ice coverage by the end of 2100. Also included are correlation scores calculated between seasonal average climate variables over the entire future period (2017-2100) and simulated snow and ice coverage by the end of 2100 (d).
Figure 10. Effect size ($\eta^2$) of main effects (ES, GCM-RCM, DS and TIM), interactions (I) and remaining error calculated using ANOVA ($\epsilon$) on projected changes in snow and ice coverage calculated using ANOVA for the six 21st century time-slices. Note, the ROR main effect is not included here as it does not influence cryospheric processes in the GHM.

The largest contributor to ice coverage projection uncertainty in the 2020s with effect 2030s with an effect size of 0.47. For snow coverage, ES and GCM-RCM have similar effect sizes of 0.45 and 0.51 for snow and ice coverage 0.4 respectively. However, for the mid and latter half of the 21st century ES uncertainty dominates, contributing 71.73% and 72% of the total projection uncertainty of snow and ice coverage by the 2090s.

5 3.5 Future evolution of primary runoff components

As an initial indication of the potential for downstream river flow regime change, Fig. 11 shows the decadal-21st century evolution of changes in the four primary runoff components relative to the 1990s including the ensemble mean projections of change (yellow dots) and their respective projection confidence intervals (blue bands) reference period. The ensemble mean indicates means (solid coloured lines) indicate that by the end of the century rainfall will increase relative to the 1990s for all months except August, but particularly under both emission scenarios except for August where RCP8.5 shows a small decrease in rainfall on average (Figure 11a). The largest increases are shown during the autumn (SON) and winter (DJF) months (Figure 11a) under RCP8.5. The confidence in the sign of change for these months is direction of change by the end of the century is $\geq$ 85% by the 2090s 90% for six months under RCP8.5 (as indicated by the coloured bands), but only for two months (March and April) under RCP4.5. However, $\geq$ 75% of the projections from both RCPs project an increase in rainfall between October and April (as indicated by the markers in Figure 11a). A comparison of the 1990s and 2090s reference and 2080s monthly ensemble means (black lines inset in Figure 11a) indicates that rainfall will more than double for majority of the autumn and winter months the peak rainfall month will shift from September to October.

For snowmelt, the greatest changes are projected to occur in the summer months (JJA) of July and August under RCP8.5 where there is $\geq$ 85% confidence that melt will reduce relative to the 1990s throughout the 21st century reference period from the 2040s onwards (Figure 11b). The ensemble mean projects that summer RCP4.5 also projects decreases in summer
melt, but the magnitude of change is smaller. In the winter months, both RCPs project a small increase in melt on average by the end of the century. The ensemble means project that total summer (JJA) melt will reduce by 40–50% by the 2090s and that the annual 19% under RCP4.5 and 37% under RCP8.5 by the 2080s (inset in Figure 11b). Annual melt will decrease by 28% under RCP4.5 and 26% under RCP8.5. A similar pattern of change is projected for ice melt (Figure 11c) with the largest proportional reduction in melt projected for June (61%) by the 2090s. A where total summer (JJA) melt will reduce by 33% under RCP4.5 and 58% under RCP8.5 by the 2080s. There is high confidence (≥90%) that mean monthly ice melt will reduce for all months except December under RCP8.5. Under RCP4.5 a small increase in winter ice melt is projected for the early and mid 21st century relative to the 1990s, but by the 2090s–2080s, winter melt is projected to reduce to near 1990s–near to reference levels on average with an overall decrease in annual melt of 45%. Under RCP8.5, winter ice melt is projected to reduce relative to reference levels for the latter half of the 21st century.

3.6 Future evolution of river flow regime

Figure 12 shows the projected deadal changes in river discharge signatures relative to the 1990s reference period (except peak water for which the raw projections are shown). The 95% projection bounds of the peak water (PW) signature indicate that it may have occurred in 2002 or could occur as late as 2089. Under RCP4.5, the ensemble mean projection of peak water is 2045, while under RCP8.5 peak water is projected to occur 17 years earlier in 2028. Indeed, the sign of change of RCP8.5 projections of the mean annual flow signature (Q) relative to the 1990s is also uncertain as shown by the confidence intervals which are approximately evenly distributed between positive and negative changes. By the show a consistent decline through the 21st century with ≥90% confidence that flows will reduce by the end of the 21st century though; annual flow is projected to remain close to 1990s on average century by 19% on average. In contrast, under RCP4.5 the magnitude of the decline is smaller (ensemble mean projects 87% decrease for 2090s). In contrast, there is much higher confidence (≥75%) that the direction of change is more uncertain. Both RCPs project an increase in inter-annual flow range (RANN) will be higher throughout the 21st century relative to the 1990s. The (≥75% under RCP8.5). Under RCP4.5 the ensemble mean projects a 0.6 m s⁻¹ (60%) 47% increase in RANN by the 2090s–2080s while RCP8.5 shows a 71% increase.

Seasonally, monthly winter (DJF) flows are projected to increase with high confidence (≥85%) while ≥90% of the ensemble project a decrease in summer (JJA) flows by the 2090s under both RCPs. The absolute change in mean monthly flows is larger for summer flows on average, but proportionally, the winter flows are projected to change most, particularly in February where the ensemble mean projects an increase of 212% a 60% and 67% increase under RCP4.5 and RCP8.5 respectively by the end of the century. The combined effect of increased winter flows and decreased summer flows results in decreased intra-annual flow variability. More than 95% Under both RCPs, more than 90% of the ensemble project a decrease in Rmnth relative to the 1990s and the ensemble mean projects reference period from the 2050s onwards. The ensemble mean projections under RCP8.5 show a decade-on-decade reduction in Rmnth with time and a 41% reduction by the end of the century.
Figure 11. Projections of monthly mean runoff components including rainfall (a), snowmelt (b), ice melt (c) and evapotranspiration (d) for RCP4.5 (blue) and RCP8.5 (yellow). Each plot includes projected decadal changes for each month, the trajectory of the ensemble mean change over the 21st century time-slices (2020s-2030s to 2090s) relative to the 1990s with projection confidence intervals. Reference period (blue bands 1991-2015) is shown by the solid coloured lines. These lines are marked for each time-slice where there is >75% confidence in the direction of change. They are bounded by the 10th and ensemble mean 90th percentiles of the projections (yellow dotted lines bands). Inset in each plot are ensemble mean monthly runoff volumes averaged over the 1990s reference period (black solid line) and 2090s-2080s (black dash lines dashed lines). Note, the monthly runoff volumes are plotted on the same scale for all runoff components to allow for cross-comparison.
Of those signatures with units m³ s⁻¹, the high flow $Q_{01}$ signature shows the largest ensemble mean change of just under an increase of 2.8 m³ s⁻¹ by and 2.5 m³ s⁻¹ for RCP4.5 and RCP8.5 respectively by the end of the 2090s, an increase of 28% relative to the 1990s century. There is high confidence ($≥ 8575\%$) that $Q_{01}$ will increase relative to the 4900s reference period under RCP8.5 but the magnitude of change is uncertain ($-2.5$ to $15.7$ m³ s⁻¹). The sign of change for $Q_{01}$ is less certain and by the 2090s the 95% confidence bounds indicate it could increase by 4.4 m³ s⁻¹ or decrease by 3.2 m³ s⁻¹. Projections under both RCPs. For $Q_{05}$, the ensemble means from both RCPs both show a reduction throughout the 21st century, however the 10th and 90th percentile span positive and negative values of change for all decades. The ensemble mean projections of changes to high flow variability ($σ_{05-01}$) are consistently positive with an average increase of 1.4 m³ s⁻¹ (91%) by the 2090s positive throughout the 21st century under both RCPs. In the latter half of the century, $≥ 75\%$ of the projections under RCP4.5 show an increase in $σ_{05-01}$ while $≥ 90\%$ of the projections under RCP8.5 show an increase.

For moderate flows, the ensemble mean projects a small increase of the RCP4.5 projections show a small reduction in $Q_{50}$ of 0.6 approximately 0.15 m³ s⁻¹ for the 2020s which then gradually trends towards a small, negative change by the 2090s. The sign of change is uncertain, particularly towards end of century. However, moderate through the 21st century while the RCP8.5 ensemble mean projects a decade-on-decade reduction in $Q_{50}$ and by the end of the century there is high confidence ($≥ 90\%$) that moderate flows will reduce under this emission scenario. Moderate flow variability ($σ_{52-48}$) is consistently projected to reduce with high confidence under both RCPs, albeit by only 0.05-0.03 m³ s⁻¹ (31%), by the 2090s and 0.06 m³ s⁻¹ by the 2080s under RCP4.5 and RCP8.5 respectively.

For the slow-release low flow signatures, projections are consistently $≥ 90\%$ of the projections are positive throughout the 21st century under both RCPs indicating an increase in the magnitude of low flow events (or equivalently a reduction in the frequency of these flow events) and variability of low flows. The absolute change in the ensemble mean never exceeds 0.2 means never exceed 0.1 m³ s⁻¹ for these signatures, although proportionally, they show the largest degree of change, particularly for $Q_{90}$ where the proportional change exceeds 6000%-2000% under RCP4.5.

Finally, the response time to runoff ($τ$) is projected to decrease by the 2090s throughout the 21st century under both RCPs ($≥ 7590\%$ confidence) indicating the catchment will likely become more flashy. The magnitude of change is small where the ensemble mean projects a small reduction in $τ$ of 2.6 hours-3.9 hours under RCP4.5 and a slightly greater reduction of 4.7 hours under RCP8.5.

### 3.7 Sources of uncertainty in river flow regime projections

Figure 13 shows the ANOVA effect sizes averaged across the future decades (2020s to 2090s) calculated for the 2030s and 2080s for each river discharge signature. The error term ($η²_{ε}$) never exceeds 0.09 and for 21 of the 25 signatures is $< 0.03$ indicating that the main effects and first order interaction terms explain the majority of projection uncertainty. Of these For the 2030s, ES uncertainty contributes $7-294-27\%$ of the total projection uncertainty across the signatures. The By the 2080s, ES contributes up to 65% of total projection uncertainty. In fact, for all but four signatures, ES contributes proportionally more to total projection uncertainty in the 2080s than the 2030s. By the 2080s the five signatures with the highest $η²_{ε}$ are include the mean monthly flows between May and August, and October (from May to August and the mean monthly flow range (R$_{mnth}$)
Figure 12. Projected decadal changes in river discharge signatures for 2020s. For each signature, the trajectory of the ensemble mean change over the 21st century time-slices (2030s to 2090s) relative to 1990s including projection confidence intervals. The reference period (blue bands 1991-2015) is shown by the solid coloured lines. These lines are marked for each time-slice where there is ≥75% confidence in the direction of change. They are bounded by the 10th and 90th percentiles of the projections (yellow dotted line bands). Also shown are 2000s-2080s ensemble mean change expressed as a percentage of simulated signatures for 1990s the reference period (text). Note, the peak water (PW) signature is not expressed as decadal change, but as the overall raw projections.

signature (Table 2) for which the effect sizes are at least 0.250.47. GCM-RCM uncertainty is the largest contributor to total projection uncertainty for 49.21 of the 25 river discharge signatures $\eta^2_{GCM-RCM}$ exceeds 0.4 for the for the 2030s and it still remains a significant contributor to projection uncertainty by the 2080s with a mean effect size across the signatures of 0.3. Four of the five most sensitive signatures to GCM-RCM uncertainty for the 2030s remain in this top five for the 2080s (Table 2) which include mean annual flow ($Q$), mean monthly flows for November and January and these include the mean monthly winter flows in January and February and two of the quick-release high flow signatures ($Q_{01}$ and $Q_{05}$). Uncertainty stemming from

On average, the DS parameterisation contributes 42-41% of the total projection uncertainty across the signatures $\eta^2_{DS}$ is also the largest contributor to projection uncertainty for the for the 2030s. In fact, $\eta^2_{DS}$ is relatively consistent across the signatures, ranging from 0.1-0.2 for 18 of the 25 signatures. For the 2080s, $\eta^2_{DS}$ reduces for all signatures except mean November and December flows and the inter-annual flow range ($R_{ANN}$) projections. Indeed, it exceeds the combined effect
Table 2. Top five river discharge signatures ranked according to the average effect size for each of the main effects, interactions and remaining error on projected decadal changes for the 2030s and 2080s. Effect sizes are included in brackets.

<table>
<thead>
<tr>
<th>Decade</th>
<th>Rank</th>
<th>ES ($\eta^2_{ES}$)</th>
<th>GCM-RCM ($\eta^2_{GCM-RCM}$)</th>
<th>DS ($\eta^2_{DS}$)</th>
<th>TIM ($\eta^2_{TIM}$)</th>
<th>ROR ($\eta^2_{ROR}$)</th>
<th>I ($\eta^2_{I}$)</th>
<th>$\tau$ ($\eta^2_{\tau}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2030s</td>
<td>1</td>
<td>$Q_{\text{DEC}}$ (0.27)</td>
<td>$Q_{\text{AN}}$ (0.59)</td>
<td>$Q_{\text{UN}}$ (0.39)</td>
<td>$R_{\text{ANN}}$ (0.35)</td>
<td>$Q_{99}$ (0.43)</td>
<td>PW (0.27)</td>
<td>PW (0.09)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$R_{\text{ANN}}$ (0.23)</td>
<td>$Q_0$ (0.56)</td>
<td>$Q_{\text{NOV}}$ (0.35)</td>
<td>$Q_{\text{MAY}}$ (0.23)</td>
<td>$Q_{95}$ (0.22)</td>
<td>$\tau$ (0.23)</td>
<td>$\tau$ (0.06)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$Q_{\text{OCT}}$ (0.21)</td>
<td>$Q_{90}$ (0.53)</td>
<td>$Q_{\text{MAR}}$ (0.26)</td>
<td>$\tau$ (0.20)</td>
<td>$Q_{99-95}$ (0.20)</td>
<td>$R_{\text{ANN}}$ (0.05)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>$Q_{\text{MAR}}$ (0.20)</td>
<td>$Q_{01}$ (0.52)</td>
<td>$\sigma_{52.48}$ (0.21)</td>
<td>$\sigma_{52.48}$ (0.18)</td>
<td>$\sigma_{99-95}$ (0.13)</td>
<td>$R_{\text{ANN}}$ (0.05)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>$\eta_{05-01}$ (0.20)</td>
<td>$Q_{\text{FEB}}$ (0.52)</td>
<td>$Q$ (0.20)</td>
<td>$Q_{\text{APR}}$ (0.17)</td>
<td>$\sigma_{05-01}$ (0.06)</td>
<td>$R_{\text{ANN}}$ (0.17)</td>
<td>$Q_{99}$ (0.02)</td>
</tr>
</tbody>
</table>

of all the remaining main effects. For $R_{\text{ANN}}$, DS has the largest effect size, contributing 43% of the total projection uncertainty. Autumn and winter monthly mean flows for September, November, December and February make up the remainder of the top five signatures most effected by DS uncertainty for the 2080s. On average TIM uncertainty contributes ≥20% of the total projection uncertainty across the different signatures for the 2030s. For this period it is the largest contributor to $R_{\text{ANN}}$ uncertainty ($\eta^2_{\text{TIM}} = 0.35$) and it also shows significant contributions to mean monthly flow projection uncertainty between April and June for April ($\eta^2_{\text{TIM}} = 0.17$) and May ($\eta^2_{\text{TIM}} = 0.23$) at the beginning of the melt season. It is also the largest contributor to uncertainty of projections of response time to runoff ($\tau$) where $\eta^2_{\text{TIM}} = 0.22$, $\eta^2_{\text{TIM}} = 0.20$. For the 2080s the average $\eta^2_{\text{TIM}}$ falls slightly to 7%, but TIM uncertainty remains an important contributor to total projection uncertainty for $\tau$, April and May flows and two of the low flow signatures ($Q_{95}$ and $\sigma_{99-95}$) where $\eta^2_{TIM} > 0.12$. Uncertainty stemming from the ROR structure-parameterisation has a negligible influence on the decadal signatures (PW and $R_{\text{ANN}}$) or those signature characterising annual and monthly mean flows. It does, however, show to be for the 2030s and 2080s. For the 2030s it is important for projections of low flow magnitude ($Q_{99}$ and $Q_{95}$, $\eta^2_{\text{ROR}} = 0.36$, $\eta^2_{\text{ROR}} = 0.43$ and 0.20 respectively) and variability ($\sigma_{99-95}$, $\eta^2_{\text{ROR}} = 0.11$, $\eta^2_{\text{ROR}} = 0.13$). In fact, for $Q_{99}$, ROR is the single largest contributor to total projection uncertainty. 46% of the For the 2080s, its influence on low flow quantiles remains significant and it is the single largest contributor to both the $Q_{99}$ and $\tau$ projection uncertainty also stems from the ROR structure-parameterisation. It also remains a significant contributor to the high flow variability signature, $\eta_{05-01}$ where $\eta^2_{ROR} = 0.12$.

Unlike ice and snow coverage, interactions between model components significantly contribute the total projection uncertainty across the signatures where $\eta^2$ ranges between 0.1 and 0.26. There is no clear pattern in the relative effect size of the interaction term across the signatures, but the five signatures with the largest $\eta^2$ include PW, $\tilde{Q}_{\text{OCT}}$, $R_{\text{ANN}}$, $\tilde{Q}_{\text{DEC}}$ and =0.07 and 0.27 for the 2030s and between 0.07 and 0.32 for the 2080s. Figure 14 shows the decomposition of the five interaction terms with the largest effect sizes on average. Generally, it is the for the 2030s (a) and 2080s (b). The interaction between the ES and GCM-RCM model chain components that dominate the contribution to projection uncertainty. However, interactions between the climate model chain components and the GHM (e.g. DS-TIM) may also contribute to the projection uncertainty. For $R_{\text{ANN}}$, DS-TIM interaction contributes ~57% of total projection uncertainty for the 2030s and 2080s. Furthermore interactions
Figure 13. Effect size ($\eta^2$) of all main effects (ES, GCM-RCM, DS, TIM and ROR), interactions (I) and remaining error ($\epsilon$) on projected decadal changes in the 25 river discharge signatures averaged across all decades affected by future climate change at the start (2020s to 2030s, a) and end (2080s, b) of the 21st century.
between the TIM and ROR in the GHM contribute some (albeit small) amounts to the total projection uncertainty. For 16 of the signatures, the contribution from interactions between model chain components increases from the 2030s to the 2080s. These include all of the signatures that characterise, high, moderate and low flow magnitude and variability, but the largest increases are shown for March and October mean monthly flows.

4 Discussion

4.1 Future evolution of river flow regime

There is high confidence that near-surface air temperature will increase by the late 21st century (2076-2100) relative to conditions in the recent past (1981-2005). Precipitation and incident solar radiation were projected to slightly increase and decrease respectively on average: a finding that is consistent with other analyses of the EURO-CORDEX projections for northern Europe (Bartók et al., 2017). The primary driver of changes in snow and ice is near-surface air temperature, while precipitation and incident solar radiation are of secondary importance. Because of this, there is high confidence that glacier ice and snow will continue to retreat as near-surface air temperature rises throughout the 21st century which would leave the river basin almost free of snow and ice by 2100 under the warmest climate projections.

The propagation of change from climate and cryosphere to the hydrosphere is clear for the Virkisá river basin and the signature-based analysis revealed that these changes will ultimately undertaken in this study has revealed how climate change will impact the magnitude, timing and variability of downstream river flows over a range of timescales. Seasonally, in the Virkisá river basin, Projected changes in flow seasonality broadly follow those shown for other mid-latitude alpine river basins where the loss of snow and ice will reduce meltwater inputs in the summer months (-37% by 2090s) where flows are currently highest on average. Additionally, and a phase shift of precipitation from snowfall to rainfall combined with enhanced melt during the colder months will increase winter runoff (113% by the 2090s) where monthly flows are currently lowest on average (Addor et al., 2014; Huss et al., 2014; Mandal and Simonovic, 2017; Jobst et al., 2018). Summer runoff is projected to decrease by 24% under RCP4.5 and 40% under RCP8.5 by the 2080s while winter runoff is projected to increase by 59% under RCP4.5 and 57% under RCP8.5 by the 2080s. The consequence of these seasonal shifts in runoff is that intra-annual (monthly) flow variability will reduce (-36% by 2090s) 25% (RCP4.5) and 41% (RCP8.5) by the 2080s. Furthermore, the magnitude of very low flow events ($Q_{90}$), which typically occur during the winter months, are likely to increase. Equivalently, the frequency of very low flow events, as defined by the 1990s reference $Q_{90}$ flow, will decrease.

On average, the projections indicated that the seasonal redistribution of runoff will have little influence on mean annual flows (-8% by 2090s under RCP4.5 -7% by the 2080s) as changes in summer and winter flows approximately compensate one-another. However, under RCP8.5, however, the more pronounced reduction in summer melt inputs results in a 19% reduction by the 2080s. The loss of a consistent melt input to the river basin and its evolution to a hydrological regime governed by rainfall-runoff processes means inter-annual flow variability (R$_{ANN}$) will increase (+60% by 2090s) 47% (RCP4.5) and 71% (RCP8.5) by the 2080s. The increase in rainfall inputs, particularly during the storm-prone autumn and winter months, likely explains the projected increased magnitude of very high flow events ($Q_{01}$) and, a finding that is in agreement with other
Figure 14. Effect size ($\eta^2$) of the five most significant interactions on projected decadal changes in the 25 river discharge signatures averaged across all decades affected by future climate change at the start (2020s to 2090s) and end (2080s) of the 21st century.
studies that have investigated changes in high flow magnitudes in glaciated river basins (Lutz et al., 2016). It is likely that the intensification of peak flow magnitudes will be further exacerbated by the projected decrease in river flow response time to runoff ($\tau$): an artefact of losing the runoff-regulating ice and snow water stores. Accordingly, the river basin will become more flashy and flood-prone in the future.

Increased flood frequency has major implications to local infrastructure in the vicinity of the Virkisá river basin. In particular, the southern section of the Route 1 highway which passes over the Skeiðarársandur floodplain navigates a large number of glacier-fed rivers including the Virkisá. Due to the unconsolidated nature of the floodplain lithology, the morphology of these rivers can change rapidly, particularly during high flow events (Marren, 2005) and often at considerable expense to the road authority (Björnsson and Pálsson, 2008). Accordingly, the projected increase in frequency and severity of high flow events will likely incur further expenses to maintain this transport link in the future.

Beyond local implications, one should be cautious in extrapolating the findings from this study to other glaciated catchments in Iceland or beyond as the sign and magnitude of river flow regime response is likely to be dependent on catchment characteristics such as climate, glacier hypsometry and geology. Even so, it can be concluded from this study, but it is clear that the timing, magnitude and variability of glacier-fed river flows over a range of timescales are sensitive to climate change. For Iceland, these changes could impact glacier-fed hydroelectric dams: a primary source of electricity for the country. Increased frequency and magnitude of high flow events could render current dams unsafe if their designed flood capacity can no longer meet regulation requirements (Thorsteinsson and Björnsson, 2012). The redistribution and ‘levelling out’ of seasonal flows, however, could actually have a beneficial effects on the running costs and capacity to produce electricity from such projects (Jóhannesson et al., 2007).

Outside of Iceland, changes to low-flow magnitudes are undoubtedly important for regions that rely on river water for drinking and irrigation. In this study a small absolute increase in low flow magnitude was projected indicating climate change and deglaciation could help to limit periods of water scarcity. However, in more arid regions, where rainfall cannot compensate reductions in melt, one might expect to see the opposite effect (Ragettli et al., 2016). One might also expect to see much greater changes in the river flow response time to runoff as snow and ice retreat in other river basins. For the Virkisá river basin, a relatively small reduction in response time ($\tau$) was projected on average by the end of the 21st century. This, perhaps, should not be surprising given the small size of the river basin and the fact that previous investigations have shown that Virkisjökull has a well-developed conduit drainage system that routes runoff efficiently year-round (Phillips et al., 2014; Flett et al., 2017). For larger river basins with more expansive cryospheric water stores, changes in the response time to runoff could be much greater, substantially increasing the risk of pluvial flooding. It is therefore vital that signature-based evaluations like the one undertaken in this study are applied to other glaciated river basins in the future so that regional variations in river flow regime change can be evaluated.

4.2 Uncertainties in projections of river flow regime

Projections of the sign direction of change relative to the 1990s-reference period were well constrained for the majority of river discharge signatures, particularly towards the end of the 21st century and for the warmer RCP8.5 emission scenario.
Even so, there was considerable spread in the projected magnitude of these changes due to uncertainties in the driving climate data (ES, GCM-RCM, DS) and representation of glacio-hydrological processes (TIM, ROR) in the model chain. Generally, sources of climate uncertainty had a larger influence on the spread of projections than those stemming from the GHM, but the relative importance of the five different model chain components was dependent on which signature of river discharge was being evaluated. Uncertainty in future snow and ice coverage primarily stemmed from the ES due to its control on future near-surface air temperature. In fact, the proportional contribution of the ES to projection uncertainties increased throughout the 21st century and consequently, and consequently the ES was also found to be the dominant source of uncertainty in projections of mean monthly flows during the melt season. In contrast, the spread of projected mean monthly winter flows stemmed mainly from GCM-RCM uncertainty, presumably by the 2080s. The growing influence of the ES over time was also shown by Addor et al. (2014) for six alpine catchments in Switzerland and by Duethmann et al. (2016) for two mountain river basins in the Tian Shan. Interestingly though, these studies along with the recent study of Jobst et al. (2018) found that climate model uncertainty was still the dominant source for projections of monthly river flows. Jobst et al. (2018) postulated that this was likely because of the considerable spread in projections of the highest precipitation magnitudes between the different GCM-RCM. High uncertainty in future precipitation across the climate models. Indeed, this would also explain why uncertainties in the projections of signatures representing quick-release high flows were most influenced by the choice of others have also attributed future runoff uncertainty in glaciated river basins to variability in precipitation projections (Lutz et al., 2016), a finding which is compounded by an increasingly warm and thus rainfall-dominated precipitation input. In this study, however, the GCM-RCM model chain component only dominated river flow projection uncertainty during the winter months while summer flow uncertainty was dominated by the ES. There are two key reasons that could explain this. Firstly, precipitation uncertainty across the GCM-RCMs showed to be especially high during winter (Fig. 5) which coupled with the fact that rainfall is the primary source of runoff during winter, likley explains the dominant role GCM-RCM plays in projection uncertainty during the winter months. Furthermore, it should be noted that the Virkisá river basin has a much higher proportional glacier coverage (60%) compared to the aforementioned studies (1.8%-22.3%). Therefore, it is postulated that the influence of the ES in the summer is related to the relatively high proportion of melt runoff that the Virkisá river receives during these months and the fact that the ES showed to be the dominant contributor to future ice coverage uncertainty. Importantly, this finding also serves to highlight the need to represent atmosphere-cryosphere-hydrosphere feedbacks adequately in future studies, particularly where glacier coverage is high, through the inclusion of a dynamic glacier evolution model in the model chain like that implemented in this study.

For projections of the inter-annual flow range, the DS procedure was the largest contributor to projection uncertainty by the end of the 21st century, which should be expected given that the perturbation of this procedure accounted for uncertainty in the random year-by-year sampling of the historic climate data. Uncertainty in the TIM structure-parameterisation was the dominant contributor to the spread in projections of moderate monthly flows during the transition from the cold to melt season which corroborates with the model comparison study of Mackay et al. (2018) who found that the structural representation of melt was important for controlling the initiation of the melt season due to the contrasting sensitivity of the models to temperature and incident solar radiation. Mackay et al. (2018) also concluded that signatures derived from the flow duration curve as well
as those representing flashiness were most sensitive to the configuration of the ROR component of the GHM. Indeed, here it was found that uncertainty in the ROR structure-parameterisation significantly contributed to the total projection uncertainty of slow-release low flow signatures as well as the response time (flashiness) of the catchment to runoff. Similar sensitivities in low flow metrics to the choice of hydrological model have been shown for non-glaciated river basins Yuan et al. (2017) and they postulated that these might stem from differences in water storage-release processes in the models. However, a key drawback of this study and other studies that have investigated the role of hydrological model uncertainty in glaciated river basins (e.g. Giuntoli et al., 2015; Vetter et al., 2017) is that they have implemented multiple model codes and therefore cannot make any definite conclusion about the source of the projection uncertainties. For example, Addor et al. (2014), concludes that the sensitivity to the choice of hydrological model could stem from any number of differences between model codes including the structure, climate interpolation method and calibration strategy. In this study, it has been demonstrated that by using a single but flexible model code, it is possible to separate out the sources of projection uncertainties down to the process level. Such insights can be used to help prioritise those aspects of the GHM that require: i) additional refinement (e.g. through model development); and ii) adequate representation of their uncertainty to improve projection robustness.

These findings have two key implications for the design of model experiments that seek to project changes in river flow regime in glaciated river basins. Firstly, studies should seek to avoid generating different sources of uncertainty from the GHM or the climate projections is dependent on which signature of river discharge is being evaluated. It is clear, therefore, that signature-based analyses could be used to help prioritise uncertainty sources based on the characteristic of flow one is interested in. For example, the results from this study indicate that for evaluating changes in monthly melt season runoff only, it may be beneficial to ignore ROR uncertainty and focus time and computational resources on quantifying uncertainties stemming from the remaining model components. In this respect, the time frame of the projections should also be considered, given the apparent change in effect sizes with time demonstrated for projections of snow and ice coverage and river flow signatures (see Appendix D).

More broadly, the results from this study emphasise the need for impact studies to represent uncertainties stemming from model chain components that control future climate and glacio-hydrological behaviour, the second of which has been widely neglected. The need for this is compounded by the fact that interactions between model chain components exceeded individual main effects for some river discharge signatures. Accordingly, an ensemble that includes perturbations of multiple components of the model chain simultaneously will provide the most rigorous quantification of projection uncertainty.

Secondly, if one is interested in projecting specific characteristics of

4.3 Limitations

While some characteristics of projected river flow regime change are broadly in agreement with other studies in similar mid-latitude alpine settings (e.g. changes in flow seasonality and projected increase in high flow magnitude), it is important to emphasise that the projected river flow regime shifts should not be generalised across glaciated river basins. Indeed, recent regional (Ragettli et al., 2016) and global (Huss and Hock, 2018) studies have shown that local catchment characteristics such as climate and glacier hypsometry largely influence seasonal river flow response to 21st century climate change. In this study
a small absolute increase in low flow magnitude was projected indicating climate change and deglaciation could help to limit periods of water scarcity. However, in more arid regions, where rainfall cannot compensate reductions in melt, the opposite effect has been shown (Stewart et al., 2015). One might also expect to see much greater changes in the river flow response time to runoff as snow and ice retreat in other river basins. For the Virkisá river basin, a relatively small reduction in response time ($\tau$) was projected on average by the end of the 21st century. This, perhaps, should not be surprising given the small size of the river basin and the fact that previous investigations have shown that Virkisjökull has a well developed conduit drainage system that routes runoff efficiently year-round (Phillips et al., 2014; Flett et al., 2017). For larger river basins with more expansive cryospheric water stores, changes in the response time to runoff could be much greater, substantially increasing the risk of pluvial flooding.

Similar inter-catchment variability should also be expected with regards to the experiment may be designed in such a way as to prioritise quantification of the dominant sources of projection uncertainty. For example, Indeed, as already noted in this discussion, some of the results from this study indicate that for evaluating changes in monthly melt season runoff only, it may be beneficial to ignore ROR uncertainty and focus time and computational resources on quantifying uncertainties stemming from the remaining model components. In this respect, the time frame of the projections should also be considered, given the apparent change in effect sizes with time demonstrated for projections of snow and ice coverage. Similar time variance in effect sizes were also found for the river flow signatures (see Appendix D). contrast the limited number of studies that have investigated uncertainty sources in other glaciated basins. Addor et al. (2014) suggests that catchment elevation influences the importance of the ES on projection uncertainty whereby runoff from higher elevation catchments with more snow and ice are more sensitive to the ES. It is therefore vital that signature-based evaluations like the one undertaken in this study are applied to other glaciated river basins in the future so that regional variations in river flow regime change and uncertainty sources can be evaluated.

It is also important to consider potential deficiencies in the calibrated GHMs. In fact, the model evaluation demonstrated that they were able to capture the majority of the observed river discharge signatures within their observation uncertainty bounds. Even so, it should be noted that there are several limitations in the calibration approach that could have hindered the efficiency of the calibrated models. Firstly, given the distributed structure of the GHM and the fact that it runs on an hourly time-step, running the GHM over multiple years required considerable computation time which limited the number of runs that could be undertaken in the Monte-Carlo calibration procedure. 5000 runs was adopted as an appropriate compromise, balancing the density of parameter sampling with available computational resources. Even so, it is recognised that particularly for the more complex model structures which employ more calibration parameters, a denser parameter sampling could help to find more efficient model parameterisations. It should also be noted that the models were calibrated and evaluated on four years of river flow data only. This detail is particularly important given the conceptual nature of the GHM and thus the potential for the calibration parameters to become less applicable when applied to periods outside of the calibration data. Additionally, it is important to highlight possible model deficiencies brought about by the two-step GHM calibration procedure employed in which the TIM and ROR model chain components were calibrated independently. This was necessary so that the main effects (Eq. 4) and interaction terms (Eq. 6) for both components could be calculated separately (thus achieving the second
aim of the study). However, the drawback of implementing this step-wise calibration procedure over one that calibrates both model components simultaneously is that it neglects any interactions between the TIM and ROR models. Of course, it should be noted that the ANOVA results showed that TIM and ROR interactions are negligible except for two of the 25 signatures evaluated.

In the previous model evaluation study undertaken by Mackay et al. (2018), they highlighted the historic observed precipitation data as source of model deficiencies. They noted the lack of available precipitation data at higher elevations, making the gridded dataset employed in this study less reliable near the basin summit. They also analysed the effectiveness of the bias-correction procedure applied to the precipitation dataset and showed that it resulted in time-series that were well correlated to the AWS data over a 3-day time step, but that this correlation degraded at shorter daily and hourly time steps which could have contributed to the model’s inability to capture snow coverage observations higher up in the catchment and river discharge signatures relating to the timing of flows.

Indeed, uncertainties in the historic precipitation data were not included as part of this study, partly because there was almost no information that would have allowed one to quantify these uncertainties (e.g. rain gauge errors), particularly higher up in the catchment where measurements are least reliable. Additionally, though, it would have meant further increasing the size of the model chain ensemble which was already at the very limit of what was computationally feasible. This, however, raises an important broader limitation of the study in that the total projection uncertainties reported are not indicative of the ‘true’ uncertainty. Further insights could undoubtedly be gained by perturbing other model chain components including the historic climate time-series and components related to key glacio-hydrological processes such as the snow redistribution routine and glacier evolution model. Certainly, Jobst et al. (2018) calculated that the bias-correction of precipitation contributed up to 22% of seasonal streamflow projection uncertainty.

Furthermore, the representation of uncertainty in the five components evaluated in this study are themselves not exhaustive. It is well established that uncertainties in climate model ensembles are under-represented (Daron and Stainforth, 2013) and steps were taken in this study to limit the total ensemble size so that the experiments were computationally feasible. For example, only 10 random DS sequences were generated, and indeed other aspects of the downscaling procedure could have also been modified (e.g. replacing the linear interpolation of change factors with a moving average model). Additionally, the melt and runoff-routing model structures implemented represent a sub-set of a much larger population of available models. For example, we adopted simplified energy balance models and the concept of linear reservoirs to route runoff. However, other model structures that employ more complex physically-based energy balance approaches and hydraulic models that simulate discrete flow pathways through the glacier (e.g. Arnold et al. 1998) could also be implemented to provide a more accurate representation of the ‘true’ projection uncertainty.

5 Conclusions

21st century climate change is projected to alter the magnitude, timing and variability of river flows over decadal to sub-daily timescales in the Virkisá river basin. Relative to the 1990s reference period, there was high confidence in the sign of direction
of change for the majority of the 25 river discharge signatures over the 21st century. The magnitude of change, however, was more uncertain. The application of ANOVA demonstrated that the climate model chain components (ES,GCM-RCM,DS) were the main sources of this uncertainty. However, uncertainty relating to glacio-hydrological process representation in the model chain (TIM,ROR) were the dominant source of projection uncertainty for some river discharge signatures. Furthermore, interactions between model chain components can exceed individual main effects. Based on these findings, we make several recommendations for future studies that aim to assess climate change impact on glacier-fed river flows:

1. Studies should seek to evaluate multiple characteristics of river flow regime change (magnitude, timing and variability) over different timescales where possible so that a more thorough understanding of potential environmental and socio-economic impacts can be deduced from projections. Signatures of river discharge provide the ideal tool to evaluate these changes quantitatively. Changes in the magnitude of river flows over decadal to seasonal timescales are already known to be highly site-specific and therefore we should expect that other signatures of regime change will also show considerable inter-catchment variation.

2. Studies should account for uncertainties stemming from both the climate projections and glacio-hydrological process representations so that more robust projections of river flow regime change are produced. The latter has largely been neglected in studies to date.

3. Careful consideration of which model chain components are the dominant sources of projection uncertainty (through the use of methods such as ANOVA) would help to prioritise resources (e.g. computational) to further enhance projection robustness. The results from this study indicate that such decisions will depend on the signatures of river flow regime change that one is interested in projecting.

Appendix A: EURO-CORDEX models

A total 15 unique GCM-RCMs using six GCMs and seven RCMs were available to use in this study (Table A1). Figure A1 shows the EURO-CORDEX 0.11° RCM grids. After comparing monthly average simulations from each GCM-RCM over the recent past (1981-2005) against the observed climate data, it was found that the [CNRM-CM5]-[ALADIN53] GCM-RCM has anomalously large negative temperature biases, particularly during the winter months of the year (see red line in Fig. A2d). In addition to this, a root mean squared error (RMSE) score was calculated for each climate variable by comparing monthly observed and simulated empirical distribution functions constructed from catchment-average daily climate data (Fig. A2a-c). When ranked according to their RMSE scores, the [CNRM-CM5]-[ALADIN53] GCM-RCM ranked 14, 13 and 15 out of 15. Given the anomalously high biases in temperature and the importance of temperature for driving hydrological change in the catchment (both in terms of melt rate and the proportion of precipitation falling as rainfall), coupled with the fact that the model was relatively poor across all three climate variables, it was deemed appropriate to remove this model from the ensemble.
Table A1. List of GCMs and RCMs used in this study.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Institution</th>
<th>Type</th>
<th>Driving GCMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNRM-CM5</td>
<td>National Centre for Meteorological Research</td>
<td>GCM</td>
<td>-</td>
</tr>
<tr>
<td>EC-EARTH</td>
<td>Europe-wide consortium</td>
<td>GCM</td>
<td>-</td>
</tr>
<tr>
<td>IPSL-CM5A-MR</td>
<td>Institut Pierre-Simon Laplace</td>
<td>GCM</td>
<td>-</td>
</tr>
<tr>
<td>HadGEM2-ES</td>
<td>Met Office Hadley Centre</td>
<td>GCM</td>
<td>-</td>
</tr>
<tr>
<td>MPI-ESM-LR</td>
<td>Max Planck Institute for Meteorology</td>
<td>GCM</td>
<td>-</td>
</tr>
<tr>
<td>NorESM1-M</td>
<td>Norwegian Climate Center</td>
<td>GCM</td>
<td>-</td>
</tr>
<tr>
<td>CCLM4-8-17</td>
<td>Climate Limited-area Modelling Community</td>
<td>RCM</td>
<td>CNRM-CM5, EC-EARTH, HadGEM2-ES, MPI-ESM-LR</td>
</tr>
<tr>
<td>ALADIN53</td>
<td>National Centre for Meteorological Research</td>
<td>RCM</td>
<td>CNRM-CM5</td>
</tr>
<tr>
<td>RCA4</td>
<td>Swedish Meteorological and Hydrological Institute, Rossby Centre</td>
<td>RCM</td>
<td>CNRM-CM5, EC-EARTH, HadGEM2-ES, MPI-ESM-LR</td>
</tr>
<tr>
<td>HIRHAM5</td>
<td>Danish Meteorological Institute</td>
<td>RCM</td>
<td>EC-EARTH, NorESM1-M</td>
</tr>
<tr>
<td>RACMO22E</td>
<td>Royal Netherlands Meteorological Institute</td>
<td>RCM</td>
<td>EC-EARTH, HadGEM2-ES</td>
</tr>
<tr>
<td>WRF331F</td>
<td>Institut Pierre Simon Laplace and Institut National de l Environnement industriel et des RISques</td>
<td>RCM</td>
<td>IPSL-CM5A-MR</td>
</tr>
<tr>
<td>REMO2009</td>
<td>Helmholtz-Zentrum Geestacht, Climate Service Center, Max Planck Institute for Meteorology</td>
<td>RCM</td>
<td>MPI-ESM-LR</td>
</tr>
</tbody>
</table>

Figure A1. EURO-CORDEX 0.11° RCM grid lines. RCM nodes are situated at grid line intersects. All RCMs utilise the green grid except for REMO2009 which uses the blue grid.
Appendix B: Ice melt and snow coverage signatures used for model calibration

12 signatures of ice melt and snow coverage which were previously derived by Mackay et al. (2018) were used for model calibration and are shown in Table B1. These signatures include: i) measurements of winter and summer ice melt in the main ablation zone between 2012 and 2014 which were derived from ablation stake data; ii) an estimate of long-term glacier volume change calculated using two DEMs of the ice for 1988 and 2011; and iii) estimates of the average seasonal snow coverage (spring, early summer, late summer) for spring (March and April) early summer (May and June) and late summer (July and August) in the lower, mid and upper (77 - 587 m asl), middle (587 - 776 m asl) and upper (776 - 1123 m asl) sections of the study basin glacier-free basin area. These were calculated from the remotely-sensed MOD10A1 MODIS product for the years 2001 to 2015 inclusive (Riggs and Hall, 2015).

Appendix C: GHM calibration parameters

Table C1 lists all of the calibration parameters and their pre-defined calibration ranges for the melt and runoff-routing model structures used during the GHM calibration procedure. The three melt model structures include the classic temperature index model (TIM₁), the enhanced temperature-index model proposed by Hock (1999) (TIM₂) and the enhanced temperature-index model proposed by Pellicciotti et al. (2005) (TIM₃). The two runoff-routing model structures include the single linear reservoir cascade (ROR₁) and two linear reservoir cascades in parallel (ROR₂).
Table B1. Summary of 12 ice melt and snow coverage signatures used to calibrate the GHM with their limits of acceptability. Note, snow coverage is expressed as a proportion of the glacier-free basin area.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Signature</th>
<th>Limits of acceptability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seasonal ice melt on tongue</td>
<td>2013 Summer ice melt</td>
<td>5.22 – 6.44 m we</td>
</tr>
<tr>
<td></td>
<td>2012-2013 Winter ice melt</td>
<td>0.64 – 0.78 m we</td>
</tr>
<tr>
<td>Long term glacier volume change</td>
<td>Change in ice volume (1988-2011)</td>
<td>-0.36 – -0.28 km³</td>
</tr>
<tr>
<td>Snow coverage in lower catchment</td>
<td>Mean snow coverage in spring</td>
<td>0.32 – 0.45</td>
</tr>
<tr>
<td></td>
<td>Mean snow coverage in early summer</td>
<td>0.02 – 0.08</td>
</tr>
<tr>
<td></td>
<td>Mean snow coverage in late summer</td>
<td>0.00 – 0.03</td>
</tr>
<tr>
<td>Snow coverage in mid catchment</td>
<td>Mean snow coverage in spring</td>
<td>0.70 – 0.80</td>
</tr>
<tr>
<td></td>
<td>Mean snow coverage in early summer</td>
<td>0.17 – 0.27</td>
</tr>
<tr>
<td></td>
<td>Mean snow coverage in late summer</td>
<td>0.00 – 0.04</td>
</tr>
<tr>
<td>Snow coverage in upper catchment</td>
<td>Mean snow coverage in spring</td>
<td>0.81 – 0.90</td>
</tr>
<tr>
<td></td>
<td>Mean snow coverage in early summer</td>
<td>0.51 – 0.64</td>
</tr>
<tr>
<td></td>
<td>Mean snow coverage in late summer</td>
<td>0.02 – 0.09</td>
</tr>
</tbody>
</table>

Table C1. Calibration parameters for the melt and runoff-routing model structures.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Parameter</th>
<th>Description</th>
<th>Calibration range</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIM₁</td>
<td>a&lt;sub&gt;ice&lt;/sub&gt;</td>
<td>Temperature factor for bare ice</td>
<td>2.0e-4 - 7.0e-4</td>
<td>m we °C⁻¹ hr⁻¹</td>
</tr>
<tr>
<td></td>
<td>a&lt;sub&gt;snow/firn&lt;/sub&gt;</td>
<td>Temperature factor for snow/firn</td>
<td>4.0e-7 - 2.0e-4</td>
<td>m we °C⁻¹ hr⁻¹</td>
</tr>
<tr>
<td>TIM₂</td>
<td>a&lt;sub&gt;ice&lt;/sub&gt;</td>
<td>Temperature factor for bare ice</td>
<td>2.0e-4 - 7.0e-4</td>
<td>m we °C⁻¹ hr⁻¹</td>
</tr>
<tr>
<td></td>
<td>a&lt;sub&gt;snow/firn&lt;/sub&gt;</td>
<td>Temperature factor for snow/firn</td>
<td>4.0e-7 - 2.0e-4</td>
<td>m we °C⁻¹ hr⁻¹</td>
</tr>
<tr>
<td></td>
<td>b&lt;sub&gt;ice&lt;/sub&gt;</td>
<td>Radiation factor for bare ice</td>
<td>4.0e-7 - 2.0e-6</td>
<td>m³ we W⁻¹ °C⁻¹ hr⁻¹</td>
</tr>
<tr>
<td></td>
<td>b&lt;sub&gt;snow/firn&lt;/sub&gt;</td>
<td>Radiation factor for snow/firn</td>
<td>4.0e-8 - 4.0e-7</td>
<td>m³ we W⁻¹ °C⁻¹ hr⁻¹</td>
</tr>
<tr>
<td>TIM₃</td>
<td>a&lt;sub&gt;ice&lt;/sub&gt;</td>
<td>Temperature factor for bare ice</td>
<td>1.5e-4 - 3.0e-4</td>
<td>m we °C⁻¹ hr⁻¹</td>
</tr>
<tr>
<td></td>
<td>a&lt;sub&gt;snow/firn&lt;/sub&gt;</td>
<td>Temperature factor for snow/firn</td>
<td>6.0e-5 - 2.0e-4</td>
<td>m we °C⁻¹ hr⁻¹</td>
</tr>
<tr>
<td></td>
<td>b&lt;sub&gt;ice&lt;/sub&gt;</td>
<td>Radiation factor for bare ice</td>
<td>1.0e-5 - 8.0e-5</td>
<td>m³ we W⁻¹ hr⁻¹</td>
</tr>
<tr>
<td></td>
<td>b&lt;sub&gt;snow/firn&lt;/sub&gt;</td>
<td>Radiation factor for snow/firn</td>
<td>2.0e-7 - 4.0e-6</td>
<td>m³ we W⁻¹ hr⁻¹</td>
</tr>
<tr>
<td></td>
<td>p₂</td>
<td>Dynamic snow albedo parameter for Brock et al. (2000) model</td>
<td>0.01 - 0.4</td>
<td></td>
</tr>
<tr>
<td>ROR₁</td>
<td>k</td>
<td>Mean residence time of reservoir</td>
<td>1 - 30</td>
<td>hr</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>Number of reservoirs</td>
<td>1 - 5</td>
<td></td>
</tr>
<tr>
<td>ROR₂</td>
<td>k&lt;sub&gt;ice/soil&lt;/sub&gt;</td>
<td>Mean residence time of runoff from ice and soil</td>
<td>0.1 - 5</td>
<td>hr</td>
</tr>
<tr>
<td></td>
<td>k&lt;sub&gt;snow/firn&lt;/sub&gt;</td>
<td>Mean residence time of runoff from snow and firn</td>
<td>20 - 100</td>
<td>hr</td>
</tr>
<tr>
<td></td>
<td>n&lt;sub&gt;ice/soil&lt;/sub&gt;</td>
<td>Number of reservoirs in ice/soil cascade</td>
<td>1 - 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n&lt;sub&gt;snow/firn&lt;/sub&gt;</td>
<td>Number of reservoirs in snow/firn cascade</td>
<td>1 - 5</td>
<td></td>
</tr>
</tbody>
</table>
Appendix D: Decadal changes in effect size for river discharge signatures

Author contributions. JDM undertook all practical elements of this study including regional climate projection downscaling, GHM calibration, 21st century projections, ANOVA, analysis of results and production of figures. He also led the writing of this manuscript. JE and ARB managed the design, commissioning and operation of the hydro-meteorological monitoring used in this study. All co-authors contributed to formulation and discussion of methods used as well as writing of the manuscript.

Competing interests. The authors declare they have no competing interests.

Acknowledgements. This work was supported by a NERC studentship awarded to JDM via the Central England NERC Training Alliance (CENTA).
Figure D1. Effect size of all main effects, interactions and remaining error on projected decadal changes in the 25 river discharge signatures for all decades affected by future climate change (2020s time-slices centred on the 2030s to 2050s) the 2080s.
References


