Response to Reviewer Comments

We would very much like to thank the editor and both referees for reading and commenting on this manuscript. We have provided a response to each of the referee comments below. The referee comments are in black and our responses are in blue. For each comment we have indicated how we have changed the manuscript to address the comments in the revised version.

Response to RC1 (Matthieu Lafaysse)

General comment

Pritchard et al presents a multiphysics spatialized snowpack modelling application in the challenging context of Himalaya with all the implied limitations in terms of data availability. This is a remarkable effort and this new application of FSM demonstrates the interest of multiphysics frameworks even in such complex contexts. To my mind, the level of importance of the different results presented in this paper is variable. For instance, the elevation-dependence of the snowpack model sensitivity (Section 4.2.4 and Figure 6) is a very innovative result compared to the existing literature and is of major importance. The analysis of the climate sensitivity of the different multiphysics options (Section 4.4 and Figure 9) is also a new and very promising method for snowpack model evaluation, and the different behaviours between options is especially interesting. Conversely some other sections present results which are either less robust (comparison of snowpack runoff with river discharge) either easily expected (best skill of a prognostic albedo to simulate albedo). The number of results presented in the paper is rather large and the paper may be more striking if more focused. Therefore, I definitely think that this paper deserves publication but I would suggest to remove some unnecessary results and a few other modifications as suggested below if it is possible (nothing being absolutely necessary).

We are very pleased that the reviewer found the manuscript to contain a number of interesting results and be deserving of publication subject to the adjustments suggested. After reflecting on the reviewer's points, we agree overall about the lesser interest of some of the prognostic vs diagnostic albedo evaluation and the difficulties in directly comparing modelled snowpack runoff with observed river flows in some parts of the results. As such, in the revised version we have removed some of the material relevant to these two points from the manuscript. Where the material might still provide a useful reference to the interested reader, we have put it into the Supplement. The details of these changes are given below in response to the "detailed comments", but the main changes may be summarised as follows:

- Moved the evaluation of albedo parameterisations using MODIS to the Supplement.
- Removed the final paragraphs in Section 4.1.1 and 4.1.2 comparing modelled snowpack runoff and observed river flows from the main text. We have retain the observed total (cumulative) runoff curve on Figure 2 as a reference for the reader on overall catchment response, but with an adjusted caption to highlight that it provides context only.
- Removed Section 4.3.1 quantifying simulated runoff differences relative to observations (and how they relate to SCA performance). Part of this material has been moved to the Supplement (along with some of the removed and adapted text from Section 4.1). However, the framing has been adjusted to consider what the differences between the modelled and observed series suggest about catchment functioning, in contrast to the emphasis on performance in the original manuscript.
- Adjusted Section 3.3 (introducing model evaluation data/methods) to explain better the more restricted use of observed river flow data in the revised manuscript.

Detailed comments

Page 1 Line 15 ‘atmospheric stability adjustment’ I know what the authors are talking about but I think this is unclear in an abstract for a standard reader who cannot be supposed to know in details how are computed the turbulent fluxes in such models.

The term ‘atmospheric stability adjustment’ has been removed from the revised abstract. In the relevant sentence it is now expressed as: “*Ensemble spread is dominated by interactions between parameterisations of albedo, snowpack hydrology, and atmospheric stability effects on turbulent heat fluxes.*” In general we have slightly adjusted the wording of results in the revised abstract to make it a little simpler and hopefully clearer for a range of readers.

Section 3.3 Snow models are known to exhibit a highly variable skill over time from one year to another. Therefore, it is critical to provide the evaluation period for the different evaluated variables to assess the robustness of the conclusions.

The evaluation periods have been added to this section. The relevant sentence for the remote sensing products now reads: “*The evaluation thus uses a number of MODIS remote sensing products (Collection 6) to assess the FSM ensemble over the full simulation period (October 2000 to September 2014).*” The sentence for observed river flows reads: “*The study also uses quality-controlled daily river flows over the period October 2000 to September 2010 recorded at the Doyian gauging station by WAPDA (Figure 1b).*”

Page 6 Line 25 If I understand well, the modelled SCA of a pixel can only be 0 or 1, is it correct?

Yes that is correct.

Page 6 Line 29 Can you provide the details of the normalization? Are there estimates of the magnitude of the remaining errors after normalization?

The ‘normalisation’ in the albedo comparison is done simply by subtracting the mean from the series of values. The modelled and (two) MODIS products considered show different mean albedos (as quantified in Section 4.2.1 in the original manuscript), so this approach permits a comparison of their time-variant behaviour after acknowledging their overall differences in the mean. On reflection we feel it would be clearer to describe this approach as the calculation of albedo anomalies, rather than as normalisation.

In response to the reviewer comments about the albedo evaluation being a less important feature of the results in this manuscript, the details of the comparison of modelled and MODIS albedo series have been moved to the Supplement (for details see response below). The method is now explained in its own section in the Supplement (Section S3) with the additional sentence: “*The modelled series are transformed to anomalies by subtracting the ensemble mean albedo (all members), while the two MODIS series are converted to anomalies by subtracting their respective means.*” The remaining differences in the series are summarised in the Supplement: “*In quantitative terms, the prognostic parameterisation outperforms the diagnostic option for the anomaly series, with an overall RMSD relative to the MOD10A1 product of 0.062 (prognostic) compared with 0.071 (diagnostic).*”

Page 8 Lines 6-14 I am not convinced that it really make sense to compare simulated snowpack runoff with observed river flows without any hydrological modelling to raise significant conclusions about the best-performing snow model configurations. The authors state that "such large differences in timing are unlikely to be accounted for by runoff routing or other hydrological processes at this time of year" but this is not demonstrated. Over mountainous basins of this surface in European Alps, it is common that evapotranspiration represents 30 to 60% of total discharge. It is also common that aquifers delay for 2 to 4 months a significant part of the runoff. How can we be sure that this is not the case in this study catchment? Several papers argue that these hydrological processes must not be ignored in snow-covered basins. Unfortunately, if some prevailing processes are missing to simulate the discharge, I am afraid that the sensitivity analysis should be limited to the 3 first paragraphs of

Section 4.1.1 because the conclusions of the last one may rely on unrealistic assumptions. Obviously, I understand that it would be nice to show that including all physical processes improve the resulting discharge but my feeling is that the evaluation variable is not direct enough for such a conclusion.

As outlined in response to the reviewer's general comments above, we agree that it may be better to remove from the manuscript some of the comparisons between modelled snowpack runoff and observed river flows. To this end we have removed the final paragraph in Section 4.1.1 as suggested by the reviewer. We have also removed the final paragraph in Section 4.1.2, which had a similar comparison.

The caption for Figure 2 has also been adjusted to highlight that the observed total runoff shown on the figure is only to provide context on the catchment behaviour. The caption now reads: "*Mean cumulative snowpack runoff for the high-flow season for each of the 32 ensemble members. In (a) each ensemble member is coloured according to the combination of albedo (AL) and liquid water (LW) parameterisations it uses. In (b) each ensemble member is coloured by its stability adjustment (ST) option. Observed total runoff (OBS, black dashed line) is shown for reference only (it is not directly comparable with snowpack runoff – see Section 3.3).*"

In addition, Section 3.3 on the data/methods for model evaluation has been adjusted to explain more explicitly the restricted ways in which observed river flows will be used in a revised manuscript. The paragraph now reads: "*The study also uses quality-controlled daily river flows over the period October 2000 to September 2010 recorded at the Doyian gauging station by WAPDA (Figure 1b). These data are used to provide some context on the volume, timing and variability of catchment runoff. As the adapted FSM model does not simulate full catchment hydrology (Section 3.1.2), the use of the observed data is restricted to two cases: (1) an indication of the timing of the rising limb of the annual hydrograph (and thus the timing of early snowpack runoff) for context; (2) an indication of the sensitivity of runoff to climate variability in the snow-dominated earliest part of the melt season (April), when flow pathways from the low elevation melting snowpack to the main channels are short, travel times are low and the influence of evapotranspiration is relatively small (Lundquist et al., 2005; Naden, 1992). The modelled quantity considered is termed snowpack runoff, which is defined as runoff from the base of the snowpack. It is thus different to surface melt, which may be subject to storage and refreezing processes before leaving the snowpack. Snowpack runoff is aggregated across all grid cells in the catchment (or across subsets of the cells for selected analyses) without any routing.*"

In addition, the related analyses quantifying the differences between simulated snowpack runoff and observed total runoff (and the temporal variability in these differences) in the original Section 4.3.1 have been removed from the main text (the original Figure 7 and related text). Part of this (relating to the original Figure 7a) has been moved into the Supplement (Section S4), with a reframing of the text to emphasise the possibility that the co-occurrence of lower modelled vs observed deviations in SCA (with MODIS) and snowpack runoff (with observed total runoff) tells us something about the nature of some of the hydrological processes not present in the adapted version of FSM used in this study. This contrasts with the original framing, which considered this co-occurrence more in terms of model performance. We think it is useful to have the adapted framing on this in the Supplement because of the questions it raises and possible hints it gives about how some of the hydrological processes in this type of catchment work. A new final sentence in Section 4.1.2 guides the reader to this material: "*Section S4 in the Supplement explores how differential SCA errors could relate to differences in the timing of early season simulated snowpack runoff and observed total runoff, as well as what this could imply about the hydrological behaviour of the catchment.*"

Page 8 Lines 21-22 Please add "(not shown)" as this statement is not supported by the presented results.

Added.

Page 9 Lines 1-2 "The slowest-responding model configurations (. . .) are too slow in their SCA decline from June onwards." Is this statement sensitive to the assumption of a binary modelled SCA? Did the authors try to use depletion curves which are largely used in the literature?

This statement should not be sensitive to the binary modelled SCA. Modelled pixels are classed as no-snow if their SWE is equal to zero. The MODIS SCA matches this definition by using a NDSI threshold of zero, which corresponds with no snow in a pixel on average (Salomonson and Appel, 2004). If sub-grid fractional snow cover in the model were taken into account, sub-pixel cover in MODIS would also need to be considered – leading to a shift of some degree (towards slightly earlier SCA decline) in both modelled and MODIS curves, not just the modelled curve. Physically, the slow SCA decline of the slowest-responding configurations is linked with the fact that they use the bulk Richardson stability adjustment. The results presented later in the manuscript (in Section 4.2.3) suggest that this adjustment likely suppresses sensible heat fluxes to the surface by too large a degree under stable conditions (see e.g. LST comparison in original Figure 5), which helps to explain why melting and SCA decline appear to be too slow for these configurations from June onwards. We have not used depletion curves, because the observed data needed to parameterise or verify curves for catchments in this region is very limited.

Section 4.2.1 Indeed the prognostic parameterization of albedo seems more consistent with observed time variations. However, is this result really surprising? I understand that the FSM framework does not consider the different parameterizations as equally probable but that they represent the variety of snow schemes used in climate models. I think that it could be expected to find that the diagnostic parameterization (based on surface temperature which is actually quite surprising) has a poorer skill than the prognostic one when considering only albedo evolution and I am not sure that we are really learning something here.

We agree that this is not really a surprising result. As such, we feel that Section 4.2.1 can be shortened by moving this comparison to the Supplement. Figure 4 and the second paragraph in Section 4.2.1 has been removed from the main text and added to the Supplement (new section) and is referred to very briefly at the end of the first paragraph in Section 4.2.1 with a new final sentence: "*Section S3 in the Supplement demonstrates that the prognostic parameterisation agrees better with the MODIS albedo products than the diagnostic option, as might be expected from previous studies (e.g. Essery et al., 2013).*"

Page 11 Lines 1-3 Indeed this is an interesting result. Actually, I am not so surprised. The classical formulation of turbulent fluxes is known to not be valid in stable conditions and especially in mountainous environments where turbulence is much more influenced by the local topography than by the snow roughness length. There is a variety of numerical artefacts to correct this behaviour in several models but even the utility of a stability correction in such conditions is subject to debate.

We agree with the reviewer about the ongoing challenges in defining turbulent flux formulations for stable atmospheric conditions, especially in complex terrain. We think it is useful to foreground this point in the main text, not least because it highlights the potential of remote sensing to identify such issues in data-sparse regions like the western Himalaya. The issue is highlighted in Section 5.1 (discussion).

Section 4.2.4 presents some of the most important findings of this paper. I am not sure that it is sufficiently highlighted.

We agree that this section should be better highlighted. To help do this we have moved it from being a sub-section of Section 4.2 into its own section (now 4.3). Details of what is shown on the relevant figure (now Figure 5) have now been added to aid reader interpretation (see response to RC2 specific comment on this section). These results are now also referred to in the abstract (new sentence "*The resulting ensemble structure is similar in different years, which leads to systematic divergence in* *ablation and mass balance at high elevations.*") and conclusion (new sentence "*These tendencies lead to notable differences in vertical patterns of snowpack ablation up to very high elevations, with substantial implications for understanding and modelling the evolution of the perennial cryosphere.*".

Page 12 Line 31 The introduction of the sensitivity to climate input in a section dedicated to the interannual variability of the skill of the different members is a bit confusing. It is hard to see why the authors choose to introduce the dependence to forcing uncertainty here while they have ignored it in all the previous sections. It is definitely useful to discuss the dependence and robustness of the results as a function of the forcing, but I would recommend to better isolate this discussion in a dedicated section instead of introducing it this way at the end of an already dense and complicated paragraph.

This section has been removed from the main text in response to the comments from RC1 and RC2 about the issues comparing modelled snowpack runoff and observed river flows. However, we agree that it would have been better not to introduce forcing uncertainty in this section. The forcing uncertainty is now treated solely in Section 5 (with reference to supporting material in the Supplement). The discussion in Section 5 (which included forcing uncertainty in the original manuscript) has also been split up into sub-sections to show more clearly where different issues and uncertainties are treated.

Section 4.4 Despite the limitations already mentioned between observed river discharges and simulated snowpack runoff, I really like this approach (in terms of methodology). I definitely agree with the implication discussed Page 15 lines 23-25. I would even say that analysing the correct climate sensitivity of the different members might be a more robust way of members selection than what have be done at the moment on direct evaluation variables. It might be a methodological recommandation for further multiphysics snow modelling as well as for the ongoing evaluations of the ESM-SnowMIP models, which can be mentioned in the discussion of this paper.

We are pleased that the reviewer supports this approach. As noted above, we have now explained in Section 3.3 that focusing just on the very early part of the melt season (April) helps to justify comparison with observed river flow anomalies (i.e. short flow paths and travel times from low elevation snowmelt to the river channels, lesser influence of evapotranspiration etc.). We have also added the suggested points on using the approach for member selection and recommending it as a method to the discussion in Section 5 with the revised sentences: "*As such, analysing the climate sensitivity of model configurations, based on their responses to historical climate variability, offers a complementary approach to traditional model evaluation methods, especially at scales where climate inputs are subject to large uncertainties. Such an approach could be useful in snow model inter-comparisons such as ESM-SnowMIP (Krinner et al., 2018), as well as for interpreting projections of snow dynamics and their wide-ranging implications in a warming world (e.g. Musselman et al., 2017; Palazzi et al., 2017; Pepin et al., 2015).*"

Response to RC2 (Anonymous Referee #2)

In this manuscript, Pritchard et al. present a new application of the Factorial Snow Model (FSM) as applied in a 2D configuration to investigate the snowmodel configuration on Himalayan snowpack simulation. First, I would like to say it was a delight to read a well written and generally clear manuscript. The figures are of a superb quality, and overall the manuscript is well done. The scientific content appears to be of generally high quality, and I believe it will be of interest to The Cryosphere readership. I would recommend this for publications, however there are a few points I have concerns with, detailed below.

We are very pleased that the reviewer found the manuscript to be of high quality and recommended for publication after addressing the few issues identified below.

My most pressing concern is that I am not convinced it is appropriate to compare the aggregate snowpack runoff with measured discharge when there is no hydrological routing in the model, nor any other hydrological processes, e.g., groundwater, soil storage, frozen soil infiltration, and ET. It is not clearly demonstrated that these processes can be ignored, especially for late seasons flows and over such a large area. I really like the story that the inclusion of improved process representation (e.g., inclusion of liquid water flows in snowpacks) improves the hydrology, however I believe this study is putting the cart before the horse, and that such claims are not supported. I would strongly suggest that these comparisons be removed.

This general point is similar to one of the main comments from RC1. Overall we agree that it would be better to remove most of the comparisons between simulated snowpack runoff and observed runoff. The details of these changes are given above in response to RC1's "detailed comments", but to reiterate the main changes may be summarised as follows:

- Removed the final paragraphs in Section 4.1.1 and 4.1.2 comparing modelled snowpack runoff and observed river flows from the main text. We have retained the observed total (cumulative) runoff curve on Figure 2 as a reference for the reader on overall catchment response, but with an adjusted caption to highlight that it provides context only.
- Removed Section 4.3.1 quantifying simulated runoff differences relative to observations (and how they relate to SCA performance). Part of this material has been moved to the Supplement (along with some of the removed and adapted text from Section 4.1). However, the framing has been adjusted to consider what the differences between the modelled and observed series suggest about catchment functioning, in contrast to the emphasis on performance in the original manuscript.
- Adjusted Section 3.3 (introducing model evaluation data/methods) to explain better the more restricted use of observed river flow data in a revised manuscript.

Second, I'm concerned that two, generally critical cold-regions processes, are ignored: blowing snow sublimation and horizontal mass redistribution. These combine to a) remove mass via sublimation – O(5-30%) of total precipitation depending on area; b) clear snow from steep slopes; c) cause avalanching that results in deep, persistent snowpacks at the base of the slopes. Personally, I've found that the Snowslide parameterization is insufficient without blowing snow processes to make a profound impact on snowcover heterogeneity, likely in line with the observations of the authors. I think this would be especially the case at the resolutions being used herein. I'm deeply sympathetic to the incredible challenge that operating a blowing snow and diagnostic avalanche model is over this type of basin, viz. uncertainty in wind and precipitation fields. I also realize the authors are also well above the snowdrift resolving scales of approx. 1 m to 150 m. However, I cannot help but worry that this confounds the SCA and albedo aggregation comparisons over the entire basin. Personally, I miss a more detailed treatment and explanation on these processes in the manuscript, and I believe it would be strengthened by highlighting these limitations to a greater degree.

We concur with the reviewer about the importance of avalanching and blowing snow processes, which we have now discussed more in the revised manuscript. It is certainly a major ongoing challenge to run distributed snow models over large areas at the appropriate resolutions mentioned by the reviewer. Deriving the required climate forcing fields at these resolutions is still problematic in data-sparse regions like the study area, where the best available climate datasets (like the HAR) are still using 10 km (or larger) grid spacing. Indeed, distributed snowpack simulations using energy balance models are still rare in the Himalayan region, such that the results of this study seem to be a necessary precursor to simulations involving additional snow redistribution (and sublimation) processes in follow-on work (when appropriate forcing datasets become available for the region, which is particularly limiting for the blowing snow component at present).

We have addressed this comment primarily by dedicating a section of expanded discussion to it within a restructured Section 5 (discussion section). We have taken the existing relevant part of Section 5 and explained the significance of these processes in more detail based on an expanded set of references. We have also discussed how not including both avalanching and blowing snow processes together is a limitation of this study, which focuses instead on the sensitivity of snow cover and runoff dynamics to snowpack process representations. Reviewing the literature certainly affirms the importance of these processes, but it remains difficult to infer the precise implications of omitting these processes just from the results of other studies (which are indeed non-existent in the Himalayan region). As such, we have tried to avoid speculating too much in the revised discussion. Given the emphasis of the study, its spatial scale and the combination of analyses in absolute/anomaly terms, we feel that the underlying conclusions are robust even if the precise (absolute) magnitude of sensitivities or performance errors were to be affected to a degree. However, better highlighting the omission of avalanching and blowing snow processes does allow the reader to make their own judgement on what the implications might be.

We have noted in the revised discussion that a key area for further work is the incorporation of these processes into distributed snow model simulations in the region. We suggest that this modelling advance is likely to require (currently unavailable) very high resolution dynamical downscaling to support model input, especially for the blowing snow modelling to be meaningful.

The relevant section of the discussion in Section 5 reads as follows:

> "*Two important influences on snow dynamics in high mountain catchments are avalanching and blowing snow processes. The latter includes snow redistribution by wind and associated sublimation during turbulent suspension. In conjunction with orographic precipitation and preferential deposition of snowfall, these processes have been shown to be important for local patterns of snow accumulation and subsequent ablation, especially in high elevation areas characterised by ridges, crests and steep slopes (e.g. Bernhardt and Schulz, 2010; Grünewald et al., 2010, 2014; MacDonald et al., 2010; Mott et al., 2010, 2014, 2018; Musselman et al., 2015; Strasser et al., 2008; Vionnet et al., 2017).*
>
> *However, evidence for the influence of these processes at larger scales is mixed. Some studies have suggested that accounting for them leads to improvements in catchment-scale model performance (e.g. Brauchli et al., 2017; Winstral et al., 2013). Yet, when considered together and when larger scales are examined, other results have indicated that these processes may have limited influence on the overall water balance and runoff dynamics (e.g. Bernhardt et al., 2012; Groot Zwaaftink et al., 2013; Vionnet et al., 2014; Warscher et al., 2013). The role of these processes thus appears to depend on the time and space scales analysed, as well as perhaps the states and fluxes under consideration.*
>
> *Initial testing showed the overall results of this study to be relatively insensitive to the SnowSlide avalanching parameterisation (Bernhardt and Schulz, 2010). Yet, it is likely that blowing snow processes need to be considered at the same time to truly capture the relevant interactions*

> *(Bernhardt et al., 2012). Therefore not including both avalanching and blowing snow processes together is a limitation of this study, which focuses instead on the sensitivity of snow cover and runoff dynamics to snowpack process representations. Similar to other mountain regions (e.g. Freudiger et al., 2017), the large mismatch in scale between available HAR climate forcing data and the requirements of blowing snow simulations makes it difficult to conduct such modelling at present, especially for catchments the size of the Astore. However, very high resolution dynamical downscaling represents a promising avenue to resolve this problem, at least on an event basis (e.g. Bonekamp et al., 2018; Vionnet et al., 2014). This will be an important area for further work.*"

I think the vertical analysis of sensitivity is a novel approach and deserves a stronger place in the paper. The sensitivity to the climate anomalies was also a nice contribution.

We agree that the section on vertical sensitivity analysis should be better highlighted. We have moved it from being a sub-section of Section 4.2 into its own section (now 4.3). Details of what is shown on the relevant figure (now Figure 5) have now been added to aid reader interpretation (see response to specific comment on this section). These results are now also referred to in the abstract (new sentence: "*The resulting ensemble structure is similar in different years, which leads to systematic divergence in ablation and mass balance at high elevations.*") and conclusion (new sentence "*These tendencies lead to notable differences in vertical patterns of snowpack ablation up to very high elevations, with substantial implications for understanding and modelling the evolution of the perennial cryosphere.*".

Specific comments:

P1. L.14-15 "These[. . .]adjustments. "This sentence is unclear

The sentence has been removed from the revised abstract. The point in the reworded abstract is now expressed as: "*Ensemble spread is dominated by interactions between parameterisations of albedo, snowpack hydrology, and atmospheric stability effects on turbulent heat fluxes.*" In general we have slightly adjusted the wording of results in the revised abstract to make it a little simpler and hopefully clearer for a range of readers.

P1. L.18 "anomaly space" is not clear and I would reword for the general reader

The term 'anomaly space' has been removed from the manuscript. The relevant sentence in the abstract now reads: "*While ensemble spread and errors may be notably lower in anomaly rather than absolute terms, FSM configurations show substantial differences in their absolute sensitivity to climate variation.*" As noted above, we have slightly adjusted the wording of results in the revised abstract to make it a little simpler and hopefully clearer for a range of readers.

P2. L.20 "This reflects". What, specially, does 'this' refer to? I have this point throughout where 'this' is used at the start of a sentence, but it is not entirely clear exactly what 'this' refers to.

The final sentences of the paragraph have been reworded for clarity to: "*Moreover, there has been little examination of how such approaches could support snow model inter-comparison for practical uses, such as water resources modelling and management, as previous inter-comparisons have focused primarily on site scales. More systematic studies are needed to understand differences in snow model behaviour at the larger scales relevant to water resources applications and regional climate modelling (Essery et al., 2009; Krinner et al., 2018).*"

P2 L. 21 "application-relevant" should be described – what application?

The final sentences of the paragraph have been reworded for clarity (see comment above, which refers to the same paragraph).

P2. L.23 “recent data” what data are these? Remote sensing? In situ?

The sentence has been reworded to clarify (removing reference to “recent data”): “*Although regional climate modelling and remote sensing offer increasing potential to support snow model inter-comparison in data-sparse regions, identifying appropriate model formulations remains challenging even in well-instrumented contexts.*”

P2. L.27 “there to be only groups[. . .]variable” Awkward, consider revising

The sentence has been reworded to: “*Similarly, recent inter-comparisons using more systematic ensemble frameworks have found that different model configurations tend to show consistently good, poor or variable performance, with no single best model identifiable (Essery et al., 2013; Lafaysse et al., 2017; Magnusson et al., 2015).*”

P2. L.28. “model complexity” Please define exactly what you mean by this

The sentence has been reworded to include a definition: “*Model complexity, in terms of the number of processes represented and the associated number of parameters, does not appear to be strongly (or necessarily positively) related to skill or transferability in space and time (see also Lute and Luce, 2017).*”

P3. L.6 “The aim [. . .]” awkward, revise.

This sentence has been reworded to: “*The aim of the study is to evaluate the sensitivity of simulated snow cover and runoff dynamics to different snowpack process representations, while also identifying possible causes and implications of model performance variation.*”

P3. L.8 “application-orientated”, same as above, what does this encapsulate?

The sentence has been reworded to remove “application-oriented”: “*This contributes to the need for more large (basin) scale model evaluations using unified frameworks (Clark et al., 2015; Essery et al., 2013), in a context where accurately simulating snow processes is essential for understanding cryospheric, hydrological and water resources trajectories in a changing climate.*”

P3. L.14 3500-4500 mASL missing. Throughout I would prefer units to be attached to the first number. As well, it is generally assumed m is ASL. Is ASL really needed here?

Units have been added after 3500-4500 but left as mASL for now.

P3. L.21-22 “this indicates [. . .] variability” is unclear to me. You’re saying that even at these cold, high elevations there is always sufficient energy to melt the entirety of the snowpacks?

We are saying here that there is sufficient energy to melt the winter snowpack over the vast majority of the basin, but not at the (much smaller areas of) very high elevations (hence the small glaciated area in the catchment). The references given (Archer and Fowler, 2004; Archer, 2003; Fowler and Archer, 2005) back this up by showing that winter precipitation observations can skilfully predict summer runoff totals, which strongly suggests that the majority of the runoff originates as winter snowfall. The annual hydrograph also tends to peak earlier in the snow-dominated Astore catchment than nearby glacier-dominated catchments like the Hunza, which supports this point.

We have reworded and expanded to clarify: “*Together with the fact that glacier cover is relatively limited, at around 6% according to the Randolph Glacier Inventory 5.0 (Arendt et al., 2015), the strong correlation between winter precipitation and summer river flows indicates that catchment runoff is primarily mass- rather than energy-limited (Archer, 2003; Fowler and Archer, 2005). However, energy constraints certainly affect intra-seasonal variability. The perennial snowpack that persists through the* *summer is confined to small areas of very high elevation, while the glaciated extent is sufficient only to provide a modest contribution to late-summer river flows (Forsythe et al., 2012).*"

P4. L.1 "This beings" what is this?

The sentence has been split and reworded to: "*The surface energy balance is solved first. Turbulent heat fluxes are estimated using the bulk aerodynamic approach.*"

P4. L.14 "physically realistic" define what you mean by this (similar to complexity)

The sentence has been reworded and expanded to define: "*For each process, the second option (1) may be considered generally more physically realistic (i.e. more in line with conceptual understanding of the physical processes governing snowpack evolution) than the first option (0). For example, in the case of snowpack hydraulic processes, it is more realistic to include liquid water retention, refreezing and drainage via a bucket model (option 1) than to permit instantaneous drainage of liquid water instead (option 0).*"

P4. L.27 Should add Marks, et al. 1999 for iSnobal

Reference added.

P5. L.19 A description on site locations would be beneficial. Are these valley sites? Cold air drainage susceptible?

A brief description of site locations has been added: "*Typical of data-sparse high mountain regions, the available stations are situated in valley locations. The HAR cold bias relative to these stations has been shown to be closely related to issues in snow cover representation in the WRF simulations underpinning the HAR (Pritchard et al., 2019), but some influence of local meteorological processes such as cold air drainage cannot be ruled out, at least at some sites.*"

It is of course very difficult to quantify the role of cold air drainage and other influences on the local temperature observations (and thus bias calculations) in this context. We have accounted for this as best we can by running the ensemble simulations with the two other forcing options presented in Section 3.3 (i.e. including one run without any bias correction).

P5. L.25 I found this section unclear with respect to, exactly, what Micromet algorithms were used. If I understood the text correctly you derived the lapse rates to use in the micromet algorithms? If this is correctly, just explicitly state this. I'm not familiar with the HAR dataset; does it provide prognostic variables at multiple pressure heights and these were what you used to derive the vertical gradients?

The main difference of our approach compared with Micromet is that the lapse rates were calculated at each time step, rather than having just one (climatological) lapse rate for each month. The vertical gradients from the HAR were calculated using surface variables (e.g. precipitation) or near-surface variables (e.g. 2 m air temperature) at all of the model grid points within the catchment and the corresponding grid cell elevations. So the gradients were calculated by regressing the simulated surface (or near-surface) values against the respective elevations of the model grid points. We have tried to refine the wording of this section to clarify as follows:

> "*For most of the climate variables, spatial disaggregation of the 10 km HAR fields to the 500 m FSM grid was conducted using methods similar to those in the MicroMet meteorological pre-processor of SnowModel (Liston and Elder, 2006b), as well as the approaches used by Duethmann et al. (2013). Specifically, for temperature, specific humidity, incoming longwave radiation, pressure and (log-transformed) precipitation, linear regression was used each time step to relate each variable to elevation, based on all HAR grid cells within the catchment (i.e. by regressing the simulated surface (or near-surface) values against the elevations of the*

corresponding model grid cells). If the gradient term in the regression was significant at the 95% confidence level, the values at each HAR cell (10 km grid) were interpolated to a reference level using the gradient. This spatial (horizontal) anomaly field was then interpolated to the high resolution FSM grid (500 m), and the elevation signal was subsequently reintroduced using the regression gradient. This approach thus differs from MicroMet primarily by using gradients that can vary at each time step, rather than applying a single climatological gradient for each calendar month. For time steps when the gradient term in the regression was not statistically significant, simple interpolation of the HAR field to the FSM grid was undertaken.

Due to the pronounced topography of the study area, clear-sky shortwave radiation at the surface was estimated for the 500 m resolution DEM using a vectorial algebra approach (Corripio, 2003). This approach accounts for the effects of slope, aspect, hill-shading and sky view factor. It has been successfully applied before in this region (e.g. Ragettli et al., 2013) and was additionally checked against station measurements. The calculated clear-sky shortwave radiation fields were adjusted to account for HAR-simulated cloud cover effects. The cloud cover effects were estimated by using spatially interpolated ratios of all-sky to clear-sky incoming shortwave radiation at the surface, with both quantities from the HAR. This approach maintains consistency between variables while capturing topographic influences, although direct/diffuse partitioning and cloud variability are simplified. In addition, MicroMet itself was used to downscale wind speed to the 500 m grid to take advantage of MicroMet's routines for modulating wind fields according to topographic slope and curvature."

P6. L.7 "To account for [. . .]" This is unclear to me what is model and what is observed. I would be more explicit here

The sentence has been split and reworded to: "*The calculated clear-sky shortwave radiation fields were adjusted to account for HAR-simulated cloud cover effects. The cloud cover effects were estimated by using spatially interpolated ratios of all-sky to clear-sky incoming shortwave radiation at the surface, with both quantities from the HAR.*"

P6. L.15 "other variables" Please state what these are here, or give some examples.

Variable names have been explicitly added to the sentence: "*The second strategy retains the same approaches for precipitation, incoming shortwave radiation and wind speed, but local observations from the Astore and other catchments in the north-west upper Indus basin (Figure 1b) are used to estimate fields for the other required variables (temperature, humidity and incoming longwave radiation – see Section S1 in Supplement).*"

P6. L.31 "This confirms [. . .]" what is this.

The sentence has been reworded to refer to the LST evaluation introduced in the preceding sentence: "*The evaluation shows that the product performs well compared with observed surface temperatures, with a relatively low mean bias of -0.55°C.*"

P6. L.31 "confirms" I would consider changing to "supports" or "agrees with" (comment applies throughout the text).

We have changed "confirms" to "shows" in the revised sentence: "*The evaluation shows that the product performs well compared with observed surface temperatures, with a relatively low mean bias of -0.55°C.*" Uses of "confirms" have been reduced throughout the manuscript.

P7. L.2 Define satisfactory quality

Now defined: "*Spatial aggregates for LST were only calculated when 90% of pixels had satisfactory quality retrievals, which are defined here as mandatory data quality flags of 00 (good).*"

P8. L.21 Blowing snow and avalanching, as per my comments earlier.

We have added some text here to reflect the importance of these processes, but longer discussion is saved until the discussion section (5) – see response to main comment above. The text added here reads: "*However, these high elevation areas are also particularly subject to the influences of blowing snow processes and avalanching, which typically alter high elevation snow cover patterns (see discussion in Section 5.2).*"

P8. L.25 "Again this [. . .]" what is this?

We have removed 'this' and reworded to: "*The similar responses of these two combinations of liquid water and stability adjustment parameterisations again reflect compensatory effects in the ensemble.*"

P8. L.25 I would lay out what compensatory effects are present here

A sentence has been added to make the compensatory effects explicit: "*Therefore, turning off the stability adjustment compensates for the tendency for slower SCA decay when using the liquid water parameterisation, whereas turning on the stability adjustment compensates for the tendency for faster SCA decline when assuming instantaneous drainage of liquid water from the snowpack.*"

P10. L.25, 26 16 and 26.8 need units

Units have been added for these numbers.

P10. L.30 What are mm/a ?

The units have been changed to mm/year.

P11. L.2 I perhaps missed it but ensure LST is defined

LST is defined earlier in the manuscript (Section 3.3).

P11. L.15+ This section is missing details described in the figure caption regarding opt0-opt1. The isotherm was not clear to me. I would ensure it is described in this text.

The details in the figure caption have been included in the main body of the text here with an additional short paragraph at the beginning of the section: "*Figure 5 examines how the tendencies identified in Sections 4.1-4.2 are manifest spatially, as well as how the influence of different processes depends on both space and time. Spatial (vertical) and temporal (monthly) differences in simulated snowpack runoff due to albedo, liquid water processes and stability option choices are shown. The differences are calculated as option 1 minus option 0 for each process, with the former being generally considered more realistic (Section 3.1.1). The lines on Figure 5 show mean differences, while ranges denote inter-annual variability. Monthly mean freezing isotherm elevations for daily minimum, mean and maximum temperatures are shown to help interpret the vertical patterns.*"

P12. L.4 RMDS, should this be RMSD? I would also ensure this is defined and the equation shown in the methodology

Yes this should have been RMSD, but the section has been removed from the main text (as one of the parts comparing simulated snowpack runoff with observed river flows – see response to general comments and RC1 comments). RMSD does not occur in the main text now (it will be defined in the Supplement though).

P12. L.10 Consider w/c for "confirms"

As above, this section has been removed and use of "confirms" has been checked/reduced throughout the manuscript.

P13. L.16 “Anomaly space” would benefit from a definition in the methodology

We have taken the phrase “anomaly space” out of the manuscript so it does not need to be defined in the methodology now. We have reworded the sentence originally containing “anomaly space” to: “*Despite these patterns of divergence and variation in absolute errors, Figure 6b indicates that, when transformed to anomaly series, the different FSM configurations are generally much more consistent, both with each other and with remote sensing.*”

P31. Figure 1 Is 0 m the correct lowest elevation for this site?

0 m is not the lowest elevation for the catchment, but it is the correct elevation for the lowest pixels and colour scale shown on Figure 1 (which covers the broader region, including neighbouring plains).

References

Marks, D., Domingo, J., Susong, D., Link, T., & Garen, D. (1999). A spatially distributed energy balance snowmelt model for application in mountain basins. Hydrological Processes, 13(12-13), 1935-1959.

Response Additional References (see manuscript for full references)

Archer, D.: Contrasting hydrological regimes in the upper Indus Basin, J. Hydrol., 274, 198–210, doi:10.1016/S0022-1694(02)00414-6, 2003.

Archer, D. R. and Fowler, H. J.: Spatial and temporal variations in precipitation in the Upper Indus Basin, global teleconnections and hydrological implications, Hydrol. Earth Syst. Sci., 8(1), 47–61, doi:10.5194/hess-8-47-2004, 2004.

Bernhardt, M. and Schulz, K.: SnowSlide: A simple routine for calculating gravitational snow transport, Geophys. Res. Lett., 37, L11502, doi:10.1029/2010GL043086, 2010.

Bernhardt, M., Schulz, K., Liston, G. E. and Zängl, G.: The influence of lateral snow redistribution processes on snow melt and sublimation in alpine regions, J. Hydrol., 424-425, 196–206, doi:10.1016/j.jhydrol.2012.01.001, 2012.

Bonekamp, P. N. J., Collier, E. and Immerzeel, W. W.: The Impact of Spatial Resolution, Land Use, and Spinup Time on Resolving Spatial Precipitation Patterns in the Himalayas, J. Hydrometeorol., 19, 1565–1581, doi:10.1175/JHM-D-17-0212.1, 2018.

Brauchli, T., Trujillo, E., Huwald, H. and Lehning, M.: Influence of Slope-Scale Snowmelt on Catchment Response Simulated With the Alpine3D Model, Water Resour. Res., 53, 723–739, doi:10.1002/2017WR021278, 2017.

Forsythe, N., Kilsby, C. G., Fowler, H. J. and Archer, D. R.: Assessment of Runoff Sensitivity in the Upper Indus Basin to Interannual Climate Variability and Potential Change Using MODIS Satellite Data Products, Mt. Res. Dev., 32(1), 16–29, doi:10.1659/MRD-JOURNAL-D-11-00027.1, 2012.

Fowler, H. J. and Archer, D. R.: Hydro-climatological variability in the Upper Indus Basin and implications for water resources, in Regional Hydrological Impacts of Climatic Change - Impact Assessment and Decision Making (Proceedings of symposium S6 held during the Seventh IAHS Scientific Assembly at Foz do Iguacu, Brazil, April 2005), IAHS Publ. 295, Foz do Iguacu, Brazil., 2005.

Freudiger, D., Kohn, I., Seibert, J., Stahl, K. and Weiler, M.: Snow redistribution for the hydrological modeling of alpine catchments, WIREs Water, e1232, 1–16, doi:10.1002/wat2.1232, 2017.

Groot Zwaaftink, C. D., Mott, R. and Lehning, M.: Seasonal simulation of drifting snow sublimation in Alpine terrain, Water Resour. Res., 49, 1581–1590, doi:10.1002/wrcr.20137, 2013.

Grünewald, T., Schirmer, M., Mott, R. and Lehning, M.: Spatial and temporal variability of snow depth and ablation rates in a small mountain catchment, Cryosph., 4, 215–225, doi:10.5194/tc-4-215-2010, 2010.

Grünewald, T., Bühler, Y. and Lehning, M.: Elevation dependency of mountain snow depth, Cryosph., 8, 2381–2394, doi:10.5194/tc-8-2381-2014, 2014.

MacDonald, M. K., Pomeroy, J. W. and Pietroniro, A.: On the importance of sublimation to an alpine snow mass balance in the Canadian Rocky Mountains, Hydrol. Earth Syst. Sci., 14, 1401–1415, doi:10.5194/hess-14-1401-2010, 2010.

Mott, R., Schirmer, M., Bavay, M., Grünewald, T. and Lehning, M.: Understanding snow-transport processes shaping the mountain snow-cver, Cryosph., 4, 545–559, doi:10.5194/tc-4-545-2010, 2010.

Mott, R., Scipión, D., Schneebeli, M., Dawes, N., Berne, A. and Lehning, M.: Orographic effects on snow deposition patterns, J. Geophys. Res. Atmos., 119, 1419–1439, doi:10.1002/2013JD019880, 2014.

Mott, R., Vionnet, V. and Grünewald, T.: The Seasonal Snow Cover Dynamics: Review on Wind-Driven Coupling Processes, Front. Earth Sci., 6(197), 1–25, doi:10.3389/feart.2018.00197, 2018.

Musselman, K. N., Pomeroy, J. W., Essery, R. L. H. and Leroux, N.: Impact of windflow calculations on simulations of alpine snow accumulation, redistribution and ablation, Hydrol. Process., 29, 3983–3999, doi:10.1002/hyp.10595, 2015.

Salomonson, V. V and Appel, I.: Estimating fractional snow cover from MODIS using the normalized difference snow index, Remote Sens. Environ., 89, 351–360, doi:10.1016/j.rse.2003.10.016, 2004.

Strasser, U., Bernhardt, M., Weber, M., Liston, G. E. and Mauser, W.: Is snow sublimation important in the alpine water balance?, Cryosph., 2, 53–66, doi:10.5194/tc-2-53-2008, 2008.

Vionnet, V., Martin, E., Masson, V., Guyomarc'h, G., Naaim-Bouvet, F., Prokop, A., Durand, Y. and Lac, C.: Simulation of wind-induced snow transport and sublimation in alpine terrain using a fully coupled snowpack/atmosphere model, Cryosph., 8, 395–415, doi:10.5194/tc-8-395-2014, 2014.

Vionnet, V., Martin, E., Masson, V., Lac, C., Bouvet, F. N. and Guyomarc'h, G.: High-Resolution Large Eddy Simulation of Snow Accumulation in Alpine Terrain, J. Geophys. Res. Atmos., 122, 11005–11021, doi:10.1002/2017JD026947, 2017.

Warscher, M., Strasser, U., Kraller, G., Marke, T., Franz, H. and Kunstmann, H.: Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps, Water Resour. Res., 49, 2619–2637, doi:10.1002/wrcr.20219, 2013.

Winstral, A., Marks, D. and Gurney, R.: Simulating wind-affected snow accumulations at catchment to basin scales, Adv. Water Resour., 55, 64–79, doi:10.1016/j.advwatres.2012.08.011, 2013.

Multi-physics ensemble snow modelling in the western Himalaya

David M. W. Pritchard[1], Nathan Forsythe[1], Greg O'Donnell[1], Hayley J. Fowler[1], Nick Rutter[2]

[1]School of Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom
[2]Department of Geography and Environmental Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom

Correspondence to: David M.W. Pritchard (david.pritchard@newcastle.ac.uk)

Abstract. Combining multiple data sources with multi-physics simulation frameworks offers new potential to extend snow model inter-comparison efforts to the Himalaya. As such, ~~T~~this study evaluates the sensitivity of simulated regional snow cover and runoff dynamics ~~importance and performance of~~to different snowpack process representations. ~~for simulating snow cover and runoff dynamics in the region.~~ The evaluation is based on ~~Focusing on the Astore catchment in the upper Indus basin, a~~a spatially distributed version of the Factorial Snowpack Model (FSM) set up for the Astore catchment in the upper Indus basin. The FSM multi-physics model was driven by climate fields from the High Asia Refined Analysis (HAR) dynamical downscaling product. Ensemble performance was evaluated primarily using ~~observed runoff and~~ MODIS remote sensing of snow-covered area, albedo and land surface temperature. In line with previous snow model inter-comparisons, ~~The results show that~~no single FSM configuration performs best in all of the years simulated. However, the results demonstrate that performance variation in this case is at least partly related to inaccuracies in the sequencing of inter-annual variation in HAR climate inputs, not just FSM model limitations. Ensemble spread is dominated by interactions between parameterisations of albedo, snowpack hydrology, and atmospheric stability effects on turbulent heat fluxes. The resulting ~~FSM~~ ensemble structure ~~spread~~is similar in different years, which leads ~~depends primarily on the interactions between parameterisations of albedo and snowpack hydrology when applied in the western Himalaya.~~to systematic divergence in ablation and mass balance at high elevations. ~~These interactions incur variation in the importance of other model choices, most notably the atmospheric stability adjustment. Although no single FSM configuration performs best in all years, applying the prognostic albedo parameterisation while accounting for liquid water retention, refreezing and drainage leads to the highest overall performance. Years when this is not the case tend to coincide with probable inaccuracies in HAR climate inputs.~~ While ~~the results indicate that~~ ensemble spread and errors ~~may be~~are notably lower ~~in anomaly space~~when viewed as anomalies, FSM configurations show ~~substantial~~ important differences in their absolute sensitivity to climate variation. ~~Therefore~~Comparison with observations suggests that~~,~~ a subset of the ensemble should be retained for climate change projections, namely those members including ~~both~~ prognostic albedo and ~~snowpack hydrology~~liquid water retention, refreezing and drainage processes.~~, while additional stability adjustment options should be tested.~~

Formatted: Not Highlight

1 Introduction

Snow plays a profound role in the climate system and supports water resources in many regions (Barnett et al., 2005; Hall and Qu, 2006). As such, several applications depend on process-based models that approximate snow physics while solving mass and energy balances. These include coupled land-atmosphere modelling, testing hypotheses about physical snow processes and catchment behaviour, and hydrological projections in non-stationary climates. Given the number of snow models in existence (e.g. Essery et al., 2013), understanding the relative skill of different models and their suitability for various uses is essential. This is reflected by the succession of snow model inter-comparison initiatives over recent decades (Essery et al., 2009; Etchevers et al., 2004; Krinner et al., 2018; Slater et al., 2001).

While snow models have been evaluated in various contexts, there has been little analysis of how different process-based models perform in the Himalayan region. Yet, vast populations downstream of these water towers are vulnerable to climate change impacts on snowmelt-derived river flows (Immerzeel et al., 2010; Viviroli et al., 2011). ~~Process-based s~~Snow model analysis and inter-comparison ~~in the region~~ would ultimately help to manage these risks, but model input and evaluation is severely impeded by data paucity. Indeed, deriving consistent, multivariate climate input fields in largely unobserved and highly variable mountain environments is a longstanding problem (Klemeš, 1990; Raleigh et al., 2015, 2016). In part this explains the proliferation of simple snow modelling in the region, namely through (sometimes enhanced) temperature index methods (e.g. Armstrong et al., 2019; Lutz et al., 2014; Ragettli et al., 2015; Stigter et al., 2017)~~(e.g. Lutz et al., 2014; Ragettli et al., 2015; Stigter et al., 2017)~~. There are~~is~~ a small but growing number of offline process-based, energy balance model applications (i.e. not coupled to an atmospheric model), but these have focused only on single model structures, without detailed snow process evaluation (e.g. Brown et al., 2014; Prasch et al., 2013; Shakoor and Ejaz, 2019; Shrestha et al., 2015).

In the face of these data challenges, developments in high resolution regional climate modelling and remote sensing increasingly offer partial solutions. Several studies have now successfully used dynamical downscaling to provide offline forcing for cryospheric and hydrological models in the Himalaya and other contexts (e.g. Biskop et al., 2016; Havens et al., 2019; Huintjes et al., 2015; Tarasova et al., 2016). Similarly, the potential for remote sensing to support model evaluation by providing some constraints on both snow cover dynamics and the surface energy balance has been demonstrated (Collier et al., 2013, 2015; Essery, 2013; Finger et al., 2011). Yet, studies combining these data sources for process-based snow model applications are still few in number (see e.g. Quéno et al., 2016; Revuelto et al., 2018), especially in the Himalaya. Moreover, there has been little examination of how such approaches could support snow model inter-comparison for practical uses, such as water resources modelling and management, as ~~. This reflects the emphasis of p~~previous inter-comparisons have focused primarily~~on~~ on site~~small~~ scales. ~~, such that m~~More systematic studies are needed to understand differences in snow model behaviour at the larger scales relevant to water resources applications and regional climate modelling ~~at larger scales with application-relevant datasets are needed~~ (Essery et al., 2009; Krinner et al., 2018).

Although ~~recent data advances are significant~~regional climate modelling and remote sensing offer increasing potential to support snow model inter-comparison in data-sparse regions, identifying appropriate model formulations remains challenging even in well-instrumented contexts. In the SnowMIP and SnowMIP2 inter-comparison studies, no single model emerged as optimal, with performance varying between locations and years (Essery et al., 2009; Etchevers et al., 2004; Rutter et al., 2009). Similarly, recent inter-comparisons using more systematic ensemble frameworks have found ~~there to be only groups of~~that different model configurations tend to show~~ that tend to perform~~ consistently ~~well~~good, poor~~ly~~ or variabl~~y~~e performance, with no single best model identifiable (Essery et al., 2013; Lafaysse et al., 2017; Magnusson et al., 2015). Model complexity, in terms of the number of processes represented and the associated number of parameters, does not appear to be strongly (or necessarily positively) related to skill or transferability in space and time (see also Lute and Luce, 2017). Nevertheless, the systematic frameworks underpinning recent inter-comparisons are an important advance over earlier studies. By synthesising the array of models in existence and eliminating implementation differences, these frameworks permit better quantification of the behaviour of different parameterisations and identification of where improvements may be possible (Clark et al., 2015; Krinner et al., 2018). The full potential of this approach has not yet been realised, especially in data-sparse mountain regions.

Therefore, this study uses recent progress in high resolution regional climate modelling, remote sensing and ensemble-based snow modelling to assess ~~the performance and behaviour of~~ snowpack model formulations in a western Himalayan catchment. ~~To do this, t~~The study proceeds by using the High Asia Refined Analysis (HAR, Maussion et al., 2014) dynamical downscaling product ~~is used~~ to drive the Factorial Snowpack Model (FSM) multi-physics ensemble (Essery, 2015). The spatially distributed, high resolution FSM simulations ~~are~~ are then evaluated using a combination of ~~local observations and~~ multiple MODIS remote sensing products and local observations. The aim of the study is to ~~characterise~~ evaluate the sensitivity of simulated snow cover and runoff dynamics to different snowpack process representations, ~~how model performance varies between formulations and in time,~~ while also identifying possible causes and implications of model performance variation~~ in simulating snow cover dynamics and snowpack runoff~~. This contributes to the need for more ~~application-oriented~~large (basin) scale model evaluations using unified frameworks (Clark et al., 2015; Essery et al., 2013), in a context where accurately simulating snow processes is essential for understanding cryospheric, hydrological and water resources trajectories in a changing climate.

2 Study Area

The study focuses on the steep, mountainous Astore catchment, a 3988 km^2 gauged sub-basin of the upper Indus basin (Figure 1). The mean elevation of the catchment is around 4000 mASL, with 57% and 87% of the area lying in the elevation ranges 3500–4500 mASL and 3000–5000 mASL, respectively. River flows are typically very low in winter, before rising in spring and peaking in summer (June/July). Spring/summer runoff is dominated by snowmelt, which is derived primarily from orographically enhanced snowfall in the preceding winter and early spring (Archer, 2003). Showing notable spatial correlation at the seasonal scale, ~~M~~much of this snowfall originates from synoptic scale low pressure systems known as westerly disturbances (Archer and Fowler, 2004). ~~; m~~Monsoon influences are small compared with the central Himalaya. ~~Showing notable spatial correlation at the seasonal scale, winter precipitation observations can predict summer runoff totals reasonably for forecasting purposes (Archer and Fowler, 2004; Archer, 2003; Fowler and Archer, 2005).~~ Together with the fact that glacier cover is relatively limited, at around 6% according to the Randolph Glacier Inventory ~~(RGI)~~ 5.0 (Arendt et al., 2015), the strong correlation between winter precipitation and summer river flows~~is~~ indicates that catchment runoff is primarily mass- rather than energy-limited (Archer, 2003; Fowler and Archer, 2005). However, energy constraints certainly affect intra-seasonal variability. The perennial snowpack that persists through the summer is confined to small areas of very high elevation, while the glaciated extent is sufficient only to provide a modest contribution to late-summer river flows (Forsythe et al., 2012). ~~However, energy constraints certainly affect intra-seasonal variability.~~ The ESA GlobCover 2009 product (Arino et al., 2012) ~~confirms~~ indicates that vegetation cover is relatively sparse, with the catchment dominated by a mixture of bare ground, herbaceous plants, and perennial snow and ice.

3 Data and Methods

3.1 Model

3.1.1 Factorial Snowpack Model (FSM) Overview

FSM is an intermediate complexity, systematic framework for evaluating alternative representations of snowpack processes and their interactions (Essery, 2015). Constructed around a coupled mass and energy balance scheme, FSM was originally formulated as a one-dimensional column model with up to three layers, depending on the total snow depth. The model is based around the sequential solution of a set of linearised equations at each time step. The ~~is begins with the~~ surface energy balance is solved ~~,~~first. T~~in which t~~urbulent heat fluxes are ~~are~~ estimated using the bulk aerodynamic approach. Heat conduction between layers and to the substrate is calculated using an implicit scheme, before layer ice and water masses are updated to account for simulated rainfall, melt, sublimation, refreezing, drainage to successive layers, and runoff from the base of the snowpack. FSM then updates layer densities and thicknesses, accounting for snowfall and conservation of total ice and liquid water masses, as well as internal energy.

Within this framework, FSM offers alternative parameterisations of different snowpack processes. The five processes are: (1) albedo evolution, (2) thermal conductivity variation, (3) snow density change by compaction, (4) adjustment of turbulent heat fluxes for atmospheric stability, and (5) liquid water retention, refreezing and drainage. With two parameterisation options (0/1) for these five processes, the FSM ensemble includes 32 possible model configurations. Summarised in Table 1, the ~~parameterisation~~ options synthesise ~~a number of~~ approaches found in a range of widely applied models. These include CLASS (Verseghy, 1991), CLM (Oleson et al., 2013), Crocus (Vionnet et al., 2012), HTESSEL (Dutra et al., 2010), ISBA (Douville et al., 1995), JULES (Best et al., 2011), MOSES (Cox et al., 1999) and Noah-MP (Niu et al., 2011). For each process, the second option (1) may be considered generally more physically realistic (i.e. more ~~physically realistic~~in line with conceptual understanding of the physical processes governing snowpack evolution) than the first option (0). For example, in the case of snowpack hydraulic processes, it is more realistic to include liquid water retention, refreezing and drainage via a bucket model (option 1) than to permit instantaneous drainage of liquid water instead (option 0).

Analyses of FSM to date have shown that it gives ensemble performance and spread comparable to larger multi-model ensembles (Essery, 2015). As such, it has been used to support study design and inter-comparison in the ESM-SnowMIP initiative (Krinner et al., 2018). Its value for testing new process representations in forest environments has also been demonstrated (Moeser et al., 2016), while Günther et al. (2019) used FSM to delineate the influences of input data errors, model structure and parameter values on simulation performance at an Alpine site.

3.1.2 Adaptations and Implementation

While the core FSM subroutines used in this study remain as described in Essery (2015), the model was adapted to perform spatially distributed simulations on a regular grid, using climate inputs that vary in space and time. In this adaptation, each grid cell is treated as independent of all other cells. Inter-cell mass and energy transfers are not considered, but the default parameterisation of sub-grid variability of snow cover as a function of snow depth is retained. In effect, each grid cell is considered as a site for which the original FSM formulation is run. The adapted version of the model is thus similar in principle to other widely applied, distributed snow models when used without their snow transport options (Lehning et al., 2006; Liston and Elder, 2006a; Marks et al., 1999)~~(e.g. Lehning et al., 2006; Liston and Elder, 2006a)~~. In line with the focus on snow cover dynamics and snowpack processes, the model does not simulate other aspects of catchment hydrology, such as evapotranspiration from snow-free cells, the soil water balance, or hydrological routing (subsurface, overland and channel flows).

The simulations reported here use a 500 m horizontal resolution grid and an hourly time step. Topography is derived from the SRTM 90 m DEM v4.1 (Jarvis et al., 2008). The 500 m spatial resolution is representative of hydrological and cryospheric modelling applications in the large basins of the Himalaya (e.g. Lutz et al., 2016), as well as extremely high resolution ~~regional~~ climate modelling (e.g. Bonekamp et al., 2018; Collier and Immerzeel, 2015). It is also consistent with several of the MODIS products used for evaluation (Section 3.3). Spatial variation in land surface properties is ignored on the basis that glacier and vegetation (including forest) cover is low (Section 2) and information on substrate properties is unavailable. The simulations use the default FSM model parameters, which have been shown to reproduce much of the spread in previous model inter-comparisons (Essery, 2015).

3.2 Climate Inputs

3.2.1 High Asia Refined Analysis (HAR)

Spatiotemporally varying input fields of rainfall, snowfall, air temperature, relative humidity, wind speed, surface air pressure, and incoming shortwave and longwave radiation are based on the HAR (Maussion et al., 2014). The HAR is a 14-year dynamical downscaling of coarser global analysis to 10 km over the Himalaya and Tibetan Plateau using the Weather Research and Forecasting model (WRF, Skamarock et al., 2008). Although a seasonally varying cold bias is present in the upper Indus basin, the HAR shows substantial performance in capturing many spatial and temporal patterns in the near-surface climatology (Maussion et al., 2014; Pritchard et al., 2019). The HAR has also exhibited a good representation of climate in several hydrological and glaciological modelling studies in neighbouring regions (Biskop et al., 2016; Huintjes et al., 2015; Tarasova et al., 2016).

3.2.2 Bias Correction and Downscaling

Near-surface air temperature fields were bias-corrected to reduce the HAR's cold bias in the study area. The mean bias was estimated using quality-controlled local observations from stations in the Astore catchment (Figure 1b), which are maintained by the Pakistan Meteorological Department (PMD) and the Water and Power Development Authority (WAPDA). Typical of data-sparse high mountain regions, the available stations are situated in valley locations. The HAR cold bias relative to these stations has been shown to be closely related to issues in snow cover representation in the WRF simulations underpinning the HAR (Pritchard et al., 2019), but some influence of local meteorological processes such as cold air drainage cannot be ruled out, at least at some sites. More advanced bias correction approaches were tested, including quantile-based methods, but these did not lead to improvements either in cross-validation at station locations or model performance at the catchment scale. This is likely due to the difficulty of fully characterising spatial and temporal variation in biases based on limited observations, especially for less commonly measured variables. The minimal approach taken thus represents a realistic application of the data. It also has the advantage of retaining much of the physical consistency of the HAR fields in terms of both inter-variable and spatiotemporal dependence structures.

For most of the climate variables, ~~S~~spatial disaggregation of the 10 km HAR fields to the 500 m FSM grid was conducted using methods ~~adapted from~~similar to those in the MicroMet meteorological pre-processor of SnowModel (Liston and Elder, 2006b) and ~~, as well as~~ those~~e approaches~~ used by Duethmann et al. (2013). Specifically, ~~F~~for temperature, specific humidity, incoming longwave radiation, pressure and (log-transformed) precipitation, linear regression was used each time step to relate each variable to elevation, based on all HAR grid cells within the catchment (i.e. by regressing the simulated surface (or near-surface) values against the elevations of the corresponding model grid cells). If the gradient term in the regression was significant at the 95% confidence level, the values at each HAR cell (10 km grid) were interpolated to a reference level using the gradient. This spatial (horizontal) anomaly field was then interpolated to the high resolution FSM grid (500 m), and the elevation signal was subsequently reintroduced using the regression gradient. This approach thus differs from MicroMet primarily by using gradients that can vary at each time step, rather than applying a single climatological gradient for each calendar month. For time steps w~~ Where~~hen the gradient term in the regression was not statistically significant, simple interpolation of the HAR field to the FSM grid was undertaken.

~~Wind speed was interpolated and modulated by topographic slope and curvature using MicroMet.~~

Due to the pronounced topography of the study area, clear-sky shortwave radiation at the surface was estimated for the 500 m resolution DEM using a vectorial algebra approach (Corripio, 2003). This approach accounts for the effects of slope, aspect, hill-shading and sky view factor. It has been successfully applied before in this region (e.g. Ragettli et al., 2013) and was additionally checked against station measurements.~~ To account for cloud cover, t~~The calculated clear-sky shortwave radiation fields ~~-~~were adjusted to account for HAR-simulated cloud cover effects. The cloud cover effects were estimated by using spatially interpolated ratios of all-sky to clear-sky ~~to received~~ incoming shortwave radiation at the surface, with both quantities from~~ according to~~ the HAR. This approach maintains consistency between variables while capturing topographic influences, although direct/diffuse partitioning and cloud variability are simplified. In addition, MicroMet itself was used to downscale wind speed to the 500 m grid to take advantage of MicroMet's routines for modulating wind fields according to topographic slope and curvature.

3.2.3 Uncertainty

Given the low density of climate observations, biases and other errors undoubtedly remain in the climate input fields. As such, two alternative input strategies were tested. The first strategy uses the same approach outlined in Section 3.2.2 but simply forgoes bias correction of temperature. The second strategy retains the same approaches for precipitation, incoming shortwave radiation and wind speed, but local observations from the Astore and other catchments in the north-west upper Indus basin (Figure 1b) are used ~~as the basis for estimation of~~to estimate fields for the other required variables (temperature, humidity and incoming longwave radiation – see ~~-~~Section S1 in Supplement). The purpose of these two alternative strategies is to indicate whether the main conclusions reached on snowpack process representations are unduly affected by the approaches described in Section 3.2.2.

3.3 Model Evaluation

The model evaluation focuses primarily on snow cover ~~and snowpack runoff~~dynamics, with other variables considered to evaluate underlying processes. Typical of the remote Himalaya, no local snow measurements for site-scale evaluation were available. The evaluation thus uses a number of MODIS remote sensing products (Collection 6) to assess the FSM ensemble over the full simulation period (October 2000 to September 2014). These include snow-covered area (SCA) derived from the MOD10A1 daily snow cover product (Hall and Riggs, 2016). Data with quality flags of 0 (best), 1 (good) and 2 (OK) were retained and~~,~~ ~~following application of~~ the widely used cloud infilling method developed by Gafurov and Bárdossy (2009) was then applied. The analysis focuses primarily on SCA corresponding with a Normalised Difference

Snow Index (NDSI) threshold of zero. This indicates very limited or no snow cover in a pixel (Salomonson and Appel, 2004), which is consistent with the no-snow threshold used to ~~estimate~~ calculate modelled SCA.

~~The evaluation also draws on the MCD43A3 and MOD10A1 surface albedo products. These datasets have been validated in different settings (Gascoin et al., 2017; Liu et al., 2009; Wang et al., 2012), but additional challenges are posed by complex terrain (Wen et al., 2018). The evaluation thus considers the model results and datasets in both absolute and normalised terms.~~ In addition, comparisons with MOD11A1 land surface temperature (LST) are undertaken. To extend previous validations (e.g. Wan, 2014; Wan et al., 2004), an evaluation of MOD11A1 at the Concordia site (Figure 1b) in the neighbouring Karakoram range is given in the Supplement (Section S2). The~~is~~ evaluation ~~confirms~~ shows that the product performs well compared with observed surface temperatures, with a relatively low mean bias of -0.55°C. ~~For both albedo and LST, s~~Spatial aggregates for LST were only calculated when 90% of pixels had satisfactory quality retrievals, which are defined here as mandatory data quality flags of 00 (good). Additional evaluation of the FSM albedo parameterisations presented in the Supplement also draws on the MCD43A3 and MOD10A1 surface albedo products, as detailed in Section S3.~~.~~

The study also uses quality-controlled daily river flows over the period October 2000 to September 2010 recorded at the Doyian gauging station by WAPDA (Figure 1b). These data are used to provide some context on the volume, timing and variability of catchment runoff. As the adapted FSM model does not simulate full catchment hydrology (Section 3.1.2), the use of the observed data is restricted to two cases: (1) an indication of the timing of the rising limb of the annual hydrograph (and thus the timing of early snowpack runoff) for context; (2) an indication of the sensitivity of runoff to climate variability in the snow-dominated earliest part of the melt season (April), when flow pathways from the low elevation melting snowpack to the main channels are short, travel times are low and the influence of evapotranspiration is relatively small (Lundquist et al., 2005; Naden, 1992). The modelled quantity considered is termed snowpack runoff, which is defined as runoff from the base of the snowpack. It is thus different to surface melt, which may be subject to storage and refreezing processes before leaving the snowpack. Snowpack runoff is aggregated across all grid cells in the catchment (or across subsets of the cells for selected analyses) without any routing.~~The study also uses quality-controlled daily river flows recorded at the Doyian gauging station by WAPDA (Figure 1b). These data allow for some inferences about the volume and timing of modelled snowpack runoff, which is defined here as runoff from the base of the snowpack. Snowpack runoff is thus different to surface melt, which may be subject to storage and refreezing processes before leaving the snowpack. Snowpack runoff is aggregated across all grid cells in the catchment for comparison with the observed data.~~

4 Results

4.1 Mean Ensemble Structure

4.1.1 Snowpack Runoff

The evaluation begins by considering how the FSM ensemble is structured on average at the catchment scale. For snowpack runoff (as defined in Section 3.3), the ensemble shows notable spread, which takes the form of groupings of ensemble members (Figure 2). Three groups of cumulative snowpack runoff curves are distinguishable early in the melt season (April and early-May) in Figure 2, but the groups split and their spread increases to varying degrees thereafter, as melt rates accelerate. Differences in snowpack runoff timing between groups are substantial, with variation of around one month in the date at which the 25^{th}, 50^{th} and 75^{th} percentiles of total seasonal snowpack runoff are exceeded.

Figure 2a indicates that the development of groups in the ensemble is primarily controlled by interactions between parameterisations of albedo and liquid water processes within the snowpack. The earliest, most rapid snowpack runoff occurs for members combining diagnostic albedo with instantaneous liquid water drainage (grey in Figure 2a). In contrast, the slowest-responding members (purple in Figure 2a) use prognostic albedo and apply the parameterisation of liquid water retention, refreezing and drainage processes (hereafter referred to as the liquid water parameterisation). The remaining two combinations of albedo and liquid water representations result in similar cumulative runoff curves, especially early in the season (orange and green in Figure 2a). This indicates that a propensity for earlier, more rapid runoff when applying diagnostic albedo is offset by a delaying effect of the liquid water parameterisation. Conversely, a tendency ~~delaying~~ to delay runoff in the prognostic albedo parameterisation is counteracted by faster runoff due to instantaneous drainage. Interactions between the albedo and liquid water parameterisations thus appear to govern whether a given option's tendency to accelerate or slow snowpack runoff is compensated for or exacerbated.

The next most important determinant of ensemble structure for snowpack runoff is the stability adjustment option, whose significance increases later in the melt season, especially in July. As noted above, Figure 2a indicates that the spread in the ensemble groups increases with time. Cross-referencing this with Figure 2b ~~confirms~~ illustrates that the stability adjustment is the main driver of the divergence. The separation is particularly pronounced for the slowest-responding ensemble members (purple in Figure 2a). In this case, not applying a stability adjustment leads to more rapid snowpack runoff from mid-June and earlier convergence with the other ensemble members, as evident from comparing the lowermost orange and blue curves in Figure 2b. In contrast, the adjustment effect is much less pronounced for the faster-responding groups of ensemble members in Figure 2a (grey curves). Therefore, not only does ~~the significance of~~sensitivity to the stability adjustment vary notably through the melt season, it is also a function of the choices of albedo and liquid water parameterisations.

~~Figure 2a also provides a preliminary indication that the timing of snowpack runoff in the slowest-responding ensemble members, which combine prognostic albedo with the liquid water parameterisation, is most similar to the timing of observed catchment runoff on average. This can be inferred particularly from the snowmelt-dominated early part of the season (April,~~

~~May and early June), when the other groups of ensemble members tend to lead the observed total runoff curve by between two and four weeks. While the full hydrology of the Astore catchment is not modelled (Section 3.1.2), such large differences in timing are unlikely to be accounted for by runoff routing or other hydrological processes at this time of year. Based on findings from snow-dominated mountainous catchments elsewhere and other settings, runoff travel times from the base of the snowpack to the catchment outlet in spring and early summer are expected to be small relative to the ensemble spread in Figure 2a, even in a catchment of this size (Lundquist et al., 2005; Naden, 1992).~~

4.1.2 Snow-Covered Area (SCA)

Figure 3 indicates that the albedo, liquid water and stability adjustment parameterisations are also the main influences on mean ensemble spread and structure in SCA. However, the dominance of albedo and liquid water processes is lesser compared with snowpack runoff, especially later in the melt season. By the annual SCA minimum in August, the stability adjustment comes to control ensemble structure (Figure 3b). Model configurations applying the adjustment exhibit a markedly slower decline in SCA over the melt season, which leads to an annual SCA minimum approximately 5% higher. Spatially, this difference is largely found at ~~relatively~~ high elevations (not shown), which could have substantial implications for modelling the evolution of ice mass in the catchment over longer periods. However, these high elevation areas are also particularly subject to the influences of blowing snow processes and avalanching, which typically alter high elevation snow cover patterns (see discussion in Section 5.2).

For much of the melt season, the model configurations most closely matching MODIS SCA (using an NDSI threshold of zero – see Section 3.3) apply prognostic albedo, along with either no liquid water representation but a stability adjustment, or the liquid water parameterisation but no stability adjustment. ~~Again this~~The similar responses of these two combinations of liquid water and stability adjustment parameterisations again reflect~~s~~ compensatory effects in the ensemble. Switching on the liquid water representation slows the rate of SCA decline, while turning off the stability adjustment speeds it up, and vice versa. Therefore, turning off the stability adjustment compensates for the tendency for slower SCA decay when using the liquid water parameterisation, whereas turning on the stability adjustment compensates for the tendency for faster SCA decline when assuming instantaneous drainage of liquid water from the snowpack. For most of the other ensemble members, the SCA curves exhibit a relatively rapid snow cover decline during the melt season compared with MODIS (NDSI threshold of zero). Section S4 in the Supplement explores how differential SCA errors could relate to differences in the timing of early season simulated snowpack runoff and observed total runoff, as well as what this could imply about the hydrological behaviour of the catchment.

~~Comparing Figure 2 and Figure 3 suggests that model configurations where snowpack runoff is closer to observed total runoff also agree better with MODIS SCA, although this is complicated by the larger spread in ensemble groups for SCA. The clearest case is for ensemble members that both apply diagnostic albedo and omit a liquid water representation. These configurations show snowpack runoff that appears to be too early and rapid (Figure 2a), as well as too fast a decline in SCA (Figure 3a). The other combinations of albedo and liquid water parameterisations perform better for both variables overall,~~

~~but this does depend substantially on the stability adjustment for SCA. Notably, the slowest-responding model configurations (using prognostic albedo, the liquid water parameterisation and the stability adjustment) are too slow in their SCA decline from June onwards (Figure 3). This may well highlight too slow a response in snowpack runoff later in the melt season that is ambiguous from Figure 2 alone, which highlights the value of MODIS SCA for reinforcing and extending inferences on model configuration performance (Finger et al., 2011).~~

4.2 Process Evaluation

The analysis now explores the processes behind the structure of ensemble spread identified in Section 4.1, as well as how far they can be assessed with independent data. This assessment is based in part on Table 2, which summarises the mean influences of albedo, liquid water and stability adjustment parameterisation choices on key catchment-scale states and fluxes in selected months. The influences are delineated as the monthly mean differences between those ensemble members applying one option for a given process (e.g. prognostic albedo) and those members applying the other option (e.g. diagnostic albedo). Density and thermal conductivity parameterisations are not considered, as it can be inferred from Section 4.1 that their effects are comparatively minor for the foci of this study, namely SCA and snowpack runoff ~~and SCA~~.

4.2.1 Albedo

Table 2 shows that prognostic albedo in FSM tends to be higher than diagnostic albedo in the first part of the melt season. The mean difference between the two albedo options ranges from 0.12 to 0.15 between April and June at solar noon. The resulting difference in net shortwave radiation helps to explain why prognostic albedo initially delays and slows melt, snowpack runoff and SCA decline relative to diagnostic albedo, which ultimately permits more snowpack runoff later in the season (see Section 4.1 and Table 2). One factor in the faster melt in spring and early summer using diagnostic albedo is its pronounced diurnal range. This is linked to the diurnal surface temperature cycle, partly through a positive feedback. The mean diurnal range in albedo for daylight hours rises from 0.18 to 0.27 between April and June when using the diagnostic parameterisation, whereas the equivalent range for the prognostic parameterisation stays much lower, at around 0.02. While albedo does vary diurnally with solar zenith angle, it does not necessarily follow that the diagnostic parameterisation captures the magnitude of variation appropriately. Section S3 in the Supplement demonstrates that the prognostic parameterisation agrees better with the MODIS albedo products than the diagnostic option, as might be expected from previous studies (e.g. Essery et al., 2013).

~~In absolute terms, FSM mean snow albedo from March to August exceeds MODIS by 0.07 and 0.14 for MOD10A1 and MCD43A3, respectively. This may be in part due to challenges in fully characterising albedo in complex terrain with remote sensing (Wen et al., 2018). However, after normalising to remove differences in the mean, Figure 4 demonstrates that the prognostic albedo parameterisation is in reasonable agreement with MODIS. Acknowledging some timing offsets and points of divergence between the MODIS products, prognostic albedo more skilfully captures the sharp albedo increases following snowfall in the melt season. Field studies have shown these events to be an important for regional melt rate variability,~~

~~especially early in the season, in accordance with the latitude (~35°N) and continentality of the area (Hewitt, 2014). The prognostic parameterisation also generally reproduces the rate of albedo decay during melting periods in Figure 4, whereas the diagnostic parameterisation induces frequent, sharp and pronounced albedo fluctuations. These fluctuations give rise to a comparatively low albedo in the early melt season. In quantitative terms, the prognostic parameterisation outperforms the diagnostic option for the normalised series, with an overall root-mean-square deviation (RMSD) relative to the MOD10A1 product of 0.062 compared with 0.071. Process-level evaluation with MODIS thus corroborates the better performance of prognostic albedo for simulating snowpack runoff and SCA (Section 4.1).~~

4.2.2 Liquid Water Retention, Refreezing and Drainage

Table 2 shows that the net effect of switching on the liquid water parameterisation is to delay snowpack runoff, even though it accelerates surface melt rates. For example, in April (May), snowpack runoff is on average 1.9 (1.6) mm/d lower when using the liquid water parameterisation, even though surface melt is 1.0 (1.9) mm/d higher. With the option on, liquid water from melting is allowed to refreeze, leading to latent heat release, which maintains a higher snowpack temperature (for example by 3.5°C in April). Retention and delayed release of liquid water in storage are part of the reason why these higher temperatures lead to higher melt rates but not higher snowpack runoff rates initially. However, multiple diurnal cycles of melting and refreezing may also be required before a given unit of snow is entirely converted to snowpack runoff. At any rate, the delaying effect of switching on the liquid water option outweighs its tendency to increase melt rates in this setting. By allowing snow to persist for longer, this enhanced storage ultimately leads to higher melt and runoff rates later in the season (by July), as later-lying snow becomes subject to increasing energy inputs (e.g. Musselman et al., 2017).

4.2.3 Stability Adjustment

Table 2 also demonstrates that switching on the stability adjustment leads to lower melt rates later in the season, primarily due to a smaller sensible heat flux towards the snow surface in stable atmospheric conditions. During the early part of the season (e.g. April), the differences in net turbulent fluxes arising from the stability adjustment choice (16 W/m^2) are largely offset by differences in net radiation (12.3 W/m^2). The larger sensible heat flux to the surface with no stability adjustment leads to a higher surface temperature (by 2.9°C in April) and thus higher outgoing longwave radiation, ~~as well as~~which incurs lower net longwave radiation. However, as the gradients between the snow surface (limited to 0°C) and near-surface air temperatures increase in spring and summer, the differences in net turbulent fluxes ~~difference~~ ultimately ~~becomes a key~~ drive~~r of~~ differences in the surface energy balance and melt rates. By June and July, the differences in net radiation (16 W/m^2 and 20 W/m^2) no longer offset the differences in net turbulent fluxes (36.8 W/m^2 and 52.7 W/m^2). Yet, Table 2 also indicates that the resulting differences in melt rates do not necessarily alter snowpack runoff on average. This reinforces the point that modelled snowpack runoff sensitivity to the stability adjustment is contingent on the representations of other processes, namely albedo and liquid water retention, refreezing and drainage (Section 4.1.1). It is also noteworthy that switching on the stability adjustment approximately halves average sublimation from 80 to 45 mm/year~~a~~, which correspond with around 8% and 4% of total catchment snow ablation, respectively.

While ~~no~~ data to evaluate turbulent fluxes directly are not available, Figure 4~~Figure 5~~ shows that switching off the stability adjustment leads to higher LST, which is in fact in closer agreement with MODIS LST. This is somewhat counter to initial expectations, as applying a stability adjustment would typically be considered more physically realistic. The largest differences in vertical LST profiles occur at night and increase with elevation, for the clear-sky conditions when MODIS retrievals are available. These differences may suggest too strong a suppression of turbulent fluxes under stable conditions using the bulk Richardson number correction. Such suppression may well contribute to the slow ~~runoff rise and~~ SCA decline when combining the stability adjustment with the otherwise realistic configuration of prognostic albedo and the parameterisation of liquid water processes (Section 4.1). Ensemble spread in day-time LST is smaller and generally in good agreement with MODIS, although the extent and influence of sub-pixel snow cover variation on MODIS LST likely increases during melting periods, giving it some positive bias in summer (Section S2 in Supplement).

~~4.2.4~~4.3 ~~Vertical Profiles~~Spatial Variation in Process Sensitivity

Figure 5~~Figure 6~~ examines how the tendencies identified ~~above~~ in Sections 4.1-4.2 are manifest spatially, as well as how the influence of different processes depends on both space and time. ~~Focusing on the vertical dimension, Figure 6 shows that~~ Spatial (vertical) and temporal (monthly) differences in simulated snowpack runoff due to albedo, liquid water processes and stability option choices are shown. The differences are calculated as option 1 minus option 0 for each process, with the former being generally considered more realistic (Section 3.1.1). The lines on Figure 5 show mean differences, while ranges denote inter-annual variability. Monthly mean freezing isotherm elevations for daily minimum, mean and maximum temperatures are shown to help interpret the vertical patterns.

~~Focusing on the vertical dimension,~~ Figure 5~~Figure 6~~ shows that

S-shaped vertical profiles of snowpack runoff differences develop and migrate upwards as the melt season progresses. The profiles take this form because the 0 options in FSM for albedo, liquid water and stability adjustment ~~processes~~ parameterisations all lead to earlier and ~~faster~~ larger snowpack runoff and snow water equivalent (SWE) depletion relative to the 1 options (see Table 1 and Sections 4.1-4.2). This gives negative differences earlier in the season ~~-~~at higher elevations when energy availability exerts more control over melt rates. However, it also means that the 1 options (i.e. prognostic albedo, liquid water processes represented and stability adjustment applied) are associated with larger SWE later in the season, allowing more ~~snowpack~~ runoff at lower elevations (positive differences in Figure 5~~Figure 6~~ increasing through the season). This is consistent with the catchment responses described above, although the inter-annual variation in the magnitude of differences is notable.

The S-shaped profiles for the different processes migrate upwards through the melt season in sequence. ~~, with t~~The profile for (negative) differences due to liquid water processes peak~~sing~~ at the highest elevations, followed by the albedo and then the stability adjustment profiles. The choice of liquid water option ~~choice~~ is particularly critical around the freezing

Formatted: Heading 2

Field Code Changed

isotherm for daily maximum temperatures, determining whether early melt is released from the snowpack or subject to storage through refreezing/melting cycles (Section 4.2.2). In comparison, the lower elevation of peak negative differences caused by the albedo parameterisation ~~is likely~~ reflects the sensitivity of the diagnostic option to~~consistent with~~ ~~a~~the higher snow surface temperature found under ~~slightly~~ higher daily mean air temperatures. The peak negative differences due to the stability adjustment option are found at lower elevations again, reflecting their dependence on the development of large near-surface temperature gradients (Section 4.2.3). Notably, for both albedo and particularly liquid water processes, differences in snowpack runoff are present up to the highest elevations. Therefore, how these processes are represented is critical for simulating the fate of high elevation perennial snow and ice.

~~4.3~~4.4 ~~Performance~~ Temporal Variation in Ensemble Structure and Performance

~~**4.3.1 Absolute Performance and Trade-Offs**~~

~~**The analysis now considers how model performance varies in time with inter-annual climate variations. Figure 7a shows the relationship between RMDS for cumulative snowpack runoff and SCA, with the ensemble aggregated by albedo, liquid water and stability adjustment options. For snowpack runoff, RMSD was calculated for each year based on the cumulative runoff curves for the period between April and June. For each year, the curves were first normalised by (dividing by) their respective total runoff volumes between April and September, in order to focus on differences in timing rather than total volumes. For SCA, RMSD was calculated for the period between April and September using an NDSI threshold of zero for MODIS (see Section 3.3).**~~

~~**Figure 7a confirms that, for individual years as well as on average, ensemble groups exhibiting closer correspondence between snowpack runoff and observed total runoff also tend to show more consistency with MODIS SCA. This provides additional support for the suggestion that snowpack runoff dominates river flows in spring and early summer, with routing effects and other influences being relatively small. Figure 7a also confirms that using both prognostic albedo and the liquid water parameterisation generally leads to the best performance (Section 4.1). However, the group omitting liquid water processes but applying the stability correction also shows low mean RMSD overall, especially for SCA. As Section 4.2.3 strongly suggests the stability adjustment to be too strong in damping turbulent fluxes in stable conditions, it is possible to identify these compensatory effects as unphysical.**~~

~~**Inter-annual variability in RMSD for all groups is high, as reflected by the wide and overlapping error bars in Figure 7a. Although substantial asymmetries and trade-offs between runoff and SCA RMSDs are present, the range of RMSD tends to be smaller for groups performing better on average. However, in some years, configurations tending to perform worse on average may outperform more realistic configurations. To examine this, the relationships between performances of different groups are shown in Figure 7b. Focusing on runoff RMSD for the three best-performing groups overall, Figure 7b shows that trade-offs resembling a Pareto front develop in some cases. This means that, for a number of the years simulated, performance in one group cannot increase without a corresponding reduction in performance of another group. Specifically, there are a number of years in which good performance with the liquid water option off (and the stability adjustment applied) is associated with performance reductions in more physically realistic configurations where the liquid water option is switched on (blue and orange points on Figure 7b).**~~

~~**In other cases, such as when prognostic albedo and the liquid water parameterisation are both applied but the stability adjustment is varied (grey points on Figure 7b), a Pareto-like front is absent and a more linear relationship between runoff RMSDs occurs. This relationship is approximately centred on the 1:1 line in Figure 7b, albeit with notable scatter, most of which is associated with the simulations using observation-based climate inputs,**~~

Formatted: Heading 3

~~**which may be the least accurate input strategy (Section 3.2.3 and Section S1 in the Supplement). This suggests that, overall, model performance limitations are not due to varying the stability adjustment option in this case. Rather, any structural constraint on performance is common to both configurations, namely through the dominance of the albedo and snowpack hydrology parameterisations and their response to the climate conditions of a particular year. More such inter-group comparisons are provided in the Supplement (Section S3). This shows that a range of linear relationships and Pareto-like fronts – as well as combinations of the two – are present across the full ensemble.**~~

~~**4.3.2 Time Series**~~

~~To explore further performance variation in time,~~ Figure 6~~Figure 8~~a shows time series of absolute catchment SCA errors (calculated as FSM minus MODIS). This series demonstrates that the best-performing group of configurations ~~-~~varies both between and within years, but that the structure of the ensemble remains fundamentally similar between years. Specifically, while the groups of FSM configurations all converge in winter, their divergence in spring and summer follows a similar pattern each year. The group using prognostic albedo and the liquid water parameterisation is consistently the uppermost series in Figure 6~~Figure 8~~a, while the group using diagnostic albedo and no liquid water parameterisation is consistently the lowermost. The rank order of the groups remains the same through time, although the magnitude of divergence varies notably between years. In some years, fast-responding combinations (Section 4.1) exhibit the lowest overall errors in at least part of the melt season (for example 2001, 2002, 2005 and 2007). Autumn convergence of groups is often associated with larger SCA errors, due to ~~difficulties~~ limitations of the climate inputs in capturing specific snowfall and melt events.

Despite these patterns of divergence and variation in absolute errors, Figure 6~~Figure 8~~b indicates that, when transformed to anomaly series, ~~, in anomaly space,~~ the different FSM configurations are generally much more consistent, both with each other and with remote sensing. Anomalies were calculated by subtracting the mean SCA for each month from the absolute SCA, separately for MODIS and the ensemble groups (Section 3.3). The sign and magnitude of SCA anomalies are generally well simulated, and the range in anomalies shown in Figure 6~~Figure 8~~b is clearly much smaller than the range of absolute errors during spring and summer in different configurations (Figure 6~~Figure 8~~a). While the focus here is on SCA, as the comparison with MODIS is fairly direct, ~~the consistency of different configurations in anomaly space is~~similar findings are also evident for snowpack runoff and LST, as shown in the Supplement (Section S~~4~~5). With potentially important implications for seasonal forecasting, this ~~reflects the~~reflects the critical role ~~dominance~~ of climatic drivers in shaping the response of diverse snow models.

However, Figure 6~~Figure 8~~b also shows that there are some instances where errors are still large in anomaly space. To explore whether these errors could be due to climate inputs, rather than model response, Figure 6~~Figure 8~~c provides time series of seasonal precipitation and temperature anomalies from the HAR input data product and local observations. As noted in Section 2, local observations show strong spatial correlation at the seasonal scale (Archer and Fowler, 2004; Archer, 2004). While the HAR provides reasonable climatological performance in many respects (Section 3.2.1), Figure 6~~Figure 8~~c suggests that its agreement with observed seasonal climate anomalies is quite variable. Taking the underestimation of the negative SCA anomaly in spring/summer 2001 in Figure 6~~Figure 8~~b as an example (labelled as (i) on Figure 6~~Figure 8~~),

Figure 6~~Figure 8~~c indicates that HAR precipitation anomalies in the (preceding) winter and (contemporaneous) spring/summer are reasonable, but that the positive observed temperature anomalies are substantially underestimated by the HAR. The erroneously cool conditions would be conducive to relatively slow simulated melt rates. This helps to explain the anomaly error~~ in anomaly space~~, as well as the large absolute SCA error in the slower-responding, more physically realistic configurations in Figure 6~~Figure 8~~a (i.e. prognostic albedo with representation of liquid water processes).

Table 3 details more such examples. Together, these cases strongly suggest that the larger discrepancies between modelled and remotely sensed SCA ~~in anomaly space~~anomalies may be related to issues with the sequences of climate ~~(~~input~~)~~ anomalies. Importantly, the examples also generally imply that nudging the climate anomalies towards observations would lead to reduction of absolute errors for the more physically realistic configurations, similar to example (i) discussed above. Thus, there is evidence that errors in climate input anomalies are a substantial factor in performance variation for FSM configurations in the Astore catchment, which partly explains poor performance of more realistic models in some years.

~~4.4~~4.5 Climate Sensitivity

While simulating specific sequences of climate anomalies and snowpack response is critical for applications such as forecasting, correctly representing the overall sensitivity of snow processes to climate variations and perturbations is crucial for offline and online climate change projections. Given the limitations in the climate input anomaly sequencing identified in Section~~ 4.3.2~~ 4.4, this section makes some inferences about the climate sensitivity of FSM configurations on the basis of inter-annual variability. The focus here is on snowpack runoff in the early part of the melt season (April), when snow is typically abundant, such that runoff is primarily constrained by energy rather than mass availability (Section 2) and dominated by snowmelt (see also Sections 3.3, 4.1 and 4.3~~ 4.3.1~~).

Figure 7~~Figure 9~~ shows the relationship between simulated 10-day air temperature and snowpack runoff anomalies for the four combinations of albedo and liquid water parameterisations. The equivalent relationship using observed temperatures and total runoff is also shown. Figure 7~~Figure 9~~ indicates that the sensitivity of snowpack runoff to climate (temperature) anomalies varies significantly for different model configurations. While scatter in the relationships are notable, likely reflecting the significant influence of other climate variables on the surface energy balance, the relationships are reasonably well approximated by linear regression (all with positive and statistically significant gradients at the 95% confidence interval). Notably, the shallowest gradient, which is associated with configurations using prognostic albedo and the liquid water parameterisation, agrees best with observations. The consistency of the ranking of the four groups (and observations) can be confirmed with bootstrapping, which shows that 77% of realisations have the same order as in Figure 7~~Figure 9~~ (89-98% of realisations if looking at consecutive groups in terms of pair-wise rankings). While the multivariate relationships between snow model response and other climate variables could be explored, the example in Figure 7~~Figure 9~~ ~~already~~ demonstrates how fundamental ~~and important~~ differences in absolute climate sensitivity can be inferred and assessed from (observable) variability.

5 Discussion

5.1 Comparison with Previous Snow Model Evaluations

~~5~~

~~The results in Section 4 confirm that several findings from previous snow model inter-comparisons hold in the western Himalaya. Prognostic albedo and inclusion of snowpack hydrology tend to be associated with the best overall model performance in the Astore catchment, which is similar to findings from the European Alps (Essery et al., 2013; Magnusson et al., 2015). In addition, the results reinforce the importance of interactions between process parameterisations, which leads to contingency of sensitivity estimates for a given process on other aspects of model configuration (Günther et al., 2019; Lafaysse et al., 2017). There is also further evidence here that turbulent heat fluxes can be overly suppressed under stable atmospheric conditions when using a stability adjustment based on the bulk Richardson number (e.g. Andreas, 2002; Collier et al., 2015; Slater et al., 2001), such that testing additional approaches would be valuable (Andreadis et al., 2009; Lapo et al., 2019). T~~Using multiple remote sensing products to augment sparse local observations, ~~he results reflect t~~the results in this study support the consensus from previous site-based inter-comparisons that no single snow model configuration performs best in all conditions, but that subsets of typically better-performing models are identifiable (Essery et al., 2013; Lafaysse et al., 2017; Magnusson et al., 2015). Yet, given both the structural similarity in the FSM ensemble between years and the close ensemble agreement on simulated anomalies, there is a strong indication that errors in the sequencing of climate input anomalies are part of the reason for year-to-year variability in catchment-scale performance in this study. As such, analysing the climate sensitivity of model configurations, based on their responses to historical climate variability, offers a complementary approach to traditional model evaluation methods, especially at scales where climate inputs are subject to large uncertainties. Such an approach could be useful in snow model inter-comparisons such as ESM-SnowMIP (Krinner et al., 2018), as well as for interpreting projections of snow dynamics and their wide-ranging implications in a warming world (e.g. Musselman et al., 2017; Palazzi et al., 2017; Pepin et al., 2015).

In agreement with previous studies, ~~Further work could explore whether this is the case elsewhere in the Himalaya and neighbouring ranges, for example in heavily monsoon-influenced areas.~~ ~~T~~the cold bias in simulated night-time LST provides~~re is also~~ ~~further~~ evidence ~~here~~ that turbulent heat fluxes can be overly suppressed under stable atmospheric conditions when using a stability adjustment based on the bulk Richardson number (e.g. Andreadis et al., 2009; Andreas, 2002; Slater et al., 2001)~~(e.g. Andreas, 2002; Collier et al., 2015; Slater et al., 2001)~~. Longstanding and f~~, such that testing additional approaches would be valuable~~undamental questions thus remain over the validity of current theories of turbulent exchange under stable conditions, especially in complex terrain ~~(Andreadis et al., 2009; Lapo et al., 2019)~~. Therefore in future work it would be useful to test alternative or amended stability adjustment options, such as the introduction of a minimum conductance (or maximum flux dampening) parameter (e.g. Collier et al., 2015; Lapo et al., 2019). These tests would ideally be done~~,~~ after adding into FSM other processes omitted in this study that could influence the surface energy balance. These include ~~In addition to alternative stability adjustments, further model development could consider~~ terrain

Formatted: Heading 2

enhancement of longwave radiation (Sicart et al., 2006), refined treatment of sub-grid variability (Clark et al., 2011), and sensible and latent heat advection (Harder et al., 2017)~~. Incorporation of these processes could potentially lead to some model performance gains and further insights into additional factors affecting the interplay of radiative fluxes, turbulent fluxes and surface temperatures. However, they would also increase complexity, dimensionality and the likelihood of error compensation.~~ Although additional processes increase the potential for error compensation, the results in this study reinforce the importance of understanding interactions between snow process parameterisations (e.g. Günther et al., 2019; Lafaysse et al., 2017) and how they vary in time and space.

~~By extending previous site-based inter-comparisons to a western Himalayan catchment, the study demonstrates the potential for combining dynamical downscaling products, remote sensing and local observations to drive, evaluate and constrain process-based ensemble snow models at large, application-relevant scales in data-sparse mountain regions. In doing so, the results suggest that process-based models could ultimately be deployed in more situations in the Himalaya. For example, it would now seem possible to assess whether estimates of the contributions of snowmelt to river flows using temperature index models (e.g. Armstrong et al., 2019) hold when more processes are incorporated. The results also reveal the potential for structural similarity in a systematic snow model ensemble in different years, as well as the hierarchy of process interactions leading to this consistent behaviour. Local climate observations can help to identify when poor performance in more physically realistic model configurations at catchment scales is likely driven by input errors, rather than model response errors. Moreover, using variability as a guide to the climate sensitivity of different model configurations can provide additional confidence when selecting ensemble members for applications like climate change projections. This may have substantial implications for projections of snow dynamics in a warming world, which are associated with important feedbacks at a range of scales (e.g. Musselman et al., 2017; Palazzi et al., 2017; Pepin et al., 2015).~~

~~Uncertainty in these results stems largely from input and evaluation data limitations, applying representative parameter values used in previous studies rather than attempting local calibration, and omitting some snow processes. In terms of input data, biases and other errors inevitably limit model performance in such a data-sparse context. However, Section S5 in the Supplement shows that the structure of the ensemble and overall performance variation remain similar when applying two alternative input strategies (Section 3.2.3 and Section S1). Although the distinction of groupings in the ensemble does reduce when using the more observation-based strategy (Figure S5 in the Supplement), this may well be the least accurate approach, due to the small number of stations available for extrapolation. Therefore, the results show some robustness to alternative, commonly applied methods for deriving climate input fields in mountain regions for practical applications. Further work could attempt a more quantitative uncertainty analysis, but defining meaningful bias ranges and error distributions to test is challenging given currently available local observations.~~

5.2 ~~For parameter values, the FSM defaults appear reasonable for the Astore catchment, based on the overall performance of more physically plausible configurations, as well as process-level evaluation where possible, such as for albedo (Section 4.2.1). Indeed, the results show there to be notable agreement in multiple simulated and observed variables in both absolute and normalised/anomaly terms, which is strengthened by physical consistency between different observed and remotely sensed variables. Further work could undertake calibration and sensitivity analyses, but this would need to guard against overfitting, error compensation and potentially unphysical behaviour given local data limitations, especially for less realistic parameterisations. In any case, recent work suggests that parameter choice may be of lower importance than both input errors and process parameterisations (Günther et al., 2019).~~Blowing Snow Processes and Avalanching

Two important influences on snow dynamics in high mountain catchments are avalanching and blowing snow processes. The latter includes snow redistribution by wind and associated sublimation during turbulent suspension. In conjunction with orographic precipitation and preferential deposition of snowfall, these processes have been shown to be important for local patterns of snow accumulation and subsequent ablation, especially in high elevation areas characterised by ridges, crests and steep slopes (e.g. Bernhardt and Schulz, 2010; Grünewald et al., 2010, 2014; MacDonald et al., 2010; Mott et al., 2010, 2014, 2018; Musselman et al., 2015; Strasser et al., 2008; Vionnet et al., 2017).

However, evidence for the influence of these processes at larger scales is mixed. Some studies have suggested that accounting for them leads to improvements in catchment-scale model performance (e.g. Brauchli et al., 2017; Winstral et al., 2013). Yet, when considered together and when larger scales are examined, other results have indicated that these processes

Formatted: Heading 2

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

may have limited influence on the overall water balance and runoff dynamics (e.g. Bernhardt et al., 2012; Groot Zwaaftink et al., 2013; Vionnet et al., 2014; Warscher et al., 2013). The role of these processes thus appears to depend on the time and space scales analysed, as well as perhaps the states and fluxes under consideration.

Initial testing showed the overall results of this study to be relatively insensitive to the SnowSlide avalanching parameterisation (Bernhardt and Schulz, 2010). Yet, it is likely that blowing snow processes need to be considered at the same time to truly capture the relevant interactions (Bernhardt et al., 2012). Therefore not including both avalanching and blowing snow processes together is a limitation of this study, which focuses instead on the sensitivity of snow cover and runoff dynamics to snowpack process representations. Similar to other mountain regions (e.g. Freudiger et al., 2017), the large mismatch in scale between available HAR climate forcing data and the requirements of blowing snow simulations makes it difficult to conduct such modelling at present, especially for catchments the size of the Astore. However, very high resolution dynamical downscaling represents a promising avenue to resolve this problem, at least on an event basis (e.g. Bonekamp et al., 2018; Vionnet et al., 2014). This will be an important area for further work.

~~Some omitted snow processes that may be of less importance for studies such as this one, which examines a relatively short period primarily at the catchment scale, could become important in other applications in the Himalaya. These processes could include avalanching, as well as wind redistribution of snow and associated sublimation (e.g. Litt et al., 2019; Stigter et al., 2018). Although initial testing showed the overall results of this study to be relatively insensitive to the SnowSlide avalanching parameterisation (Bernhardt and Schulz, 2010), these processes have been demonstrated to be important at some scales in Alpine contexts (Bernhardt et al., 2012; Musselman et al., 2015; Strasser et al., 2008). Coupling FSM with existing snow redistribution models would be possible (e.g. Lehning et al., 2006; Liston and Elder, 2006a), but ideally such simulations would be driven by higher resolution dynamical downscaling products than the HAR (e.g. Vionnet et al., 2014). For some applications it could also useful to run FSM at higher spatial resolutions (e.g. Baba et al., 2019; Sohrabi et al., 2019; Winstral et al., 2014), although testing showed similar results at both 250 and 1000 m resolutions.~~

5.3 Uncertainty in Data and Parameters

Formatted: Heading 2

Input data biases and other errors inevitably limit model performance to some extent in such a data-sparse context (Raleigh et al., 2016). However, Section S6 in the Supplement indicates that the structure of the ensemble and overall performance variation remain similar if two alternative input strategies are applied (see also Section 3.2.3 and Section S1). Although the distinction of groupings in the ensemble does reduce when using the more observation-based input strategy (Figure S7 in the Supplement), this may well be the least accurate approach, due to the small number of stations available. The results therefore show some robustness to alternative, commonly applied methods for deriving climate input fields in mountain regions for practical applications. Section 4.4 also shows how the larger of the possible input anomaly errors may be identified, which facilitates a better understanding of the occasional performance drops for more physically realistic, typically well-performing model configurations (i.e. identification of potential input rather than structural errors, such as underestimation of the warm temperature anomaly in the 2001 ablation season – Section 4.4). Further work could attempt a

Formatted: Not Highlight

Formatted: Not Highlight

more quantitative uncertainty analysis, but defining meaningful bias ranges and error distributions to test is challenging given currently available local observations.

The FSM parameter defaults appear reasonable for the Astore catchment, based on the performance of more physically plausible model configurations across multiple variables in both absolute and normalised/anomaly terms. Further work could undertake calibration and sensitivity analyses, but this would need to guard against overfitting, error compensation and potentially unphysical behaviour given local data limitations, especially for less realistic parameterisations. In any case, recent work suggests that parameter choice in FSM may be of lower importance than both input errors and process parameterisations (Günther et al., 2019).~~In addition to alternative stability adjustments, further model development could consider terrain enhancement of longwave radiation (Sicart et al., 2006), refined treatment of sub-grid variability (Clark et al., 2011), and sensible and latent heat advection (Harder et al., 2017). Incorporation of these processes could potentially lead to some model performance gains and further insights into additional factors affecting the interplay of radiative fluxes, turbulent fluxes and surface temperatures. However, they would also increase complexity, dimensionality and the likelihood of error compensation.~~

6 Conclusion

This study demonstrates that combining local observations, dynamical downscaling and remote sensing products with multi-physics ensemble frameworks facilitates the identification of skilful process-based snow model configurations in the western Himalaya. Although the importance of different snowpack processes varies with time, space and other model options, the results show that the structure of the FSM ensemble is fundamentally similar between years. Different process parameterisations consistently act to accelerate or delay snowpack runoff and SCA decay. These tendencies lead to notable differences in vertical patterns of snowpack ablation up to very high elevations, with substantial implications for understanding and modelling the evolution of the perennial cryosphere.~~.~~ In agreement with other studies, the results indicate that the prognostic albedo parameterisation should be preferred (Essery et al., 2013; Magnusson et al., 2015). Representation of liquid water retention, refreezing and drainage in the snowpack is also generally required, unless compensatory effects are introduced by other aspects of model configuration, especially the atmospheric stability adjustment option. However, there is evidence that turbulent fluxes are overly suppressed in some conditions by applying the stability adjustment based on the bulk Richardson number implemented in FSM. Correctly simulating these fluxes is a major ongoing challenge in land surface modelling (Lapo et al., 2019), especially in complex terrain.

While no model configuration performs best in all years, the results~~re is evidence to~~ suggest that errors in climate input anomalies play a key role in this, not just model structural limitations. Model snow cover dynamics and runoff responses to climate variations are more consistent ~~in~~ after being transformed to anomalies~~anomaly space~~, which may be useful in forecasting applications, but there is substantial spread within the ensemble in terms of absolute climate sensitivity. This variation in climate sensitivity is critical for model selection in both offline and online simulations supporting climate change impact projections. Together, these points suggest that an ensemble modelling approach should be used in applications where possible. However, a subset of the full FSM ensemble could be taken forward, namely those members which use prognostic albedo and account for snowpack hydrology. Further work could examine input uncertainties in more detail, as well as alternative process parameterisations (especially for the stability adjustment), parameter value sensitivity, and additional unrepresented processes (such as ~~snow redistribution~~avalanching and blowing snow processes). In the complex terrain of the western Himalaya, these tasks would all benefit from higher resolution dynamical downscaling products to help drive snow models. Advances in these areas ~~This~~ could ultimately lead to improved modelling tools to support water resources management in the Himalaya and other mountain regions in a changing climate.

7 Code Availability

The open source FSM model code is available here: https://github.com/RichardEssery/FSM. An extended version with additional physics and options for multi-site simulations is found at: https://github.com/RichardEssery/FSM2. The scripts used to prepare climate inputs and conduct spatially distributed simulations can be requested from the corresponding author.

8 Data Availability

The HAR data product is publicly available at this site: http://www.klima-ds.tu-berlin.de/har/. MODIS data are publicly available; instructions for downloading the products used in this study can be obtained from: https://lpdaac.usgs.gov/products/mcd43a3v006/, https://lpdaac.usgs.gov/products/mod11a1v006/, and https://nsidc.org/data/mod10a1. Most of the observed climate and river flow data are not publicly available, but they can be requested from PMD and WAPDA. Tables summarising station metadata and climatology are also available in various publications (e.g. Archer and Fowler, 2004; Archer, 2003; Mukhopadhyay and Khan, 2016; Sharif et al., 2013; Waqas and Athar, 2018). The climate data from the EvK2CNR Concordia site can be downloaded from: http://share.evk2cnr.org/.

9 Author Contributions

With input from NF, GOD, HF and NR, DP designed the study, conducted the modelling and led the analysis. All authors contributed to the interpretation of results. DP drafted the manuscript, with NF, GOD, HF and NR all providing ideas and alterations.

10 Competing Interests

The authors declare that they have no conflict of interest.

11 Acknowledgements

The authors are grateful to PMD, WAPDA and EvK2CNR for the in-situ observations used in this study. We would also sincerely like to thank Fabien Maussion and colleagues for making the HAR dataset publicly available, as well as Richard Essery for making the FSM program open source. During this study Hayley Fowler was funded by the Wolfson Foundation and the Royal Society as a Royal Society Wolfson Research Merit Award (WM140025), as well as the European Research Council Grant, INTENSE (ERC-2013-CoG-617329). Nathan Forsythe was supported by the Royal Society (CH160148 and CHG\R1\170057) and David Pritchard by an EPSRC doctoral training award (EP/M506382/1).

References

Andreadis, K. M., Storck, P. and Lettenmaier, D. P.: Modeling snow accumulation and ablation processes in forested environments, Water Resour. Res., 45, W05429, doi:10.1029/2008WR007042, 2009.

Andreas, E. L.: Parameterizing Scalar Transfer over Snow and Ice: A Review, J. Hydrometeorol., 3, 417–432, doi:10.1175/1525-7541(2002)003<0417:PSTOSA>2.0.CO;2, 2002.

Archer, D.: Contrasting hydrological regimes in the upper Indus Basin, J. Hydrol., 274, 198–210, doi:10.1016/S0022-1694(02)00414-6, 2003.

Archer, D. R.: Hydrological implications of spatial and altitudinal variation in temperature in the Upper Indus Basin, Nord. Hydrol., 35(3), 209–222, doi:10.2166/nh.2004.0015, 2004.

Archer, D. R. and Fowler, H. J.: Spatial and temporal variations in precipitation in the Upper Indus Basin, global teleconnections and hydrological implications, Hydrol. Earth Syst. Sci., 8(1), 47–61, doi:10.5194/hess-8-47-2004, 2004.

Arendt, A., Bliss, A., Bolch, T., Cogley, J. G., Gardner, A. S., Hagen, J.-O., Hock, R., Huss, M., Kaser, G., Kienholz, C., Pfeffer, W. T., Moholdt, G., Paul, F., Radić, V., Andreassen, L., Bajracharya, S., Barrand, N. E., Berthier, E., Bhambri, R., Brown, I., Burgess, E., Burgess, D., Cawkwell, F., Chinn, T., Copland, L., Davies, B., De Angelis, H., Dolgova, E., Earl, L., Filbert, K., Forester, R., Fountain, A. G., Frey, H., Giffen, B., Guo, W. Q., Gurney, S., Hagg, W., Hall, D., Haritashya, U. K., Hartmann, G., Helm, C., Herreid, S., Kapustin, G., Khromova, T., König, M., Kohler, J., Kriegel, D., Kutuzov, S., Lavrentiev, I., Lebris, R., Nosenko, G., Negrete, A., Nuimura, T., Nuth, C., Pettersson, R., Racoviteanu, A., Ranzi, R., Rastner, P., Rau, F., Raup, B., Rich, J., Rott, H., Sakai, A., Schneider, C., Seliverstov, Y., Sharp, M., Sigurðsson, O., Stokes, C., Way, R. G., Wheate, R., Winsvold, S., Wolken, G., Wyatt, F. and Zheltyhina, N.: Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 5.0: Technical Report, Colorado, USA., 2015.

Arino, O., Ramos Perez, J. J., Kalogirou, V., Bontemps, S., Defourny, P. and Van Bogaert, E.: Global Land Cover Map for 2009 (GlobCover 2009), , doi:https://doi.org/10.1594/PANGAEA.787668, 2012.

Armstrong, R. L., Rittger, K., Brodzik, M. J., Racoviteanu, A., Barrett, A. P., Khalsa, S.-J. S., Raup, B., Hill, A. F., Khan, A. L., Wilson, A. M., Kayastha, R. B., Fetterer, F. and Armstrong, B.: Runoff from glacier ice and seasonal snow in High Asia: separating melt water sources in river flow, Reg. Environ. Chang., 19, 1249–1261, doi:10.1007/s10113-018-1429-0, 2019.

Barnett, T. P., Adam, J. C. and Lettenmaier, D. P.: Potential impacts of a warming climate on water availability in snow-dominated regions, Nature, 438, 303–309, doi:10.1038/nature04141, 2005.

Bernhardt, M. and Schulz, K.: SnowSlide: A simple routine for calculating gravitational snow transport, Geophys. Res. Lett., 37, L11502, doi:10.1029/2010GL043086, 2010.

Bernhardt, M., Schulz, K., Liston, G. E. and Zängl, G.: The influence of lateral snow redistribution processes on snow melt and sublimation in alpine regions, J. Hydrol., 424-425, 196–206, doi:10.1016/j.jhydrol.2012.01.001, 2012.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Menard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B. and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1 : Energy and water fluxes, Geosci. Model Dev., 4, 677–699, doi:10.5194/gmd-4-677-2011, 2011.

Biskop, S., Maussion, F., Krause, P. and Fink, M.: Differences in the water-balance components of four lakes in the southern-central Tibetan Plateau, Hydrol. Earth Syst. Sci., 20, 209–225, doi:10.5194/hess-20-209-2016, 2016.

Bonekamp, P. N. J., Collier, E. and Immerzeel, W. W.: The Impact of Spatial Resolution, Land Use, and Spinup Time on Resolving Spatial Precipitation Patterns in the Himalayas, J. Hydrometeorol., 19, 1565–1581, doi:10.1175/JHM-D-17-0212.1, 2018.

Brauchli, T., Trujillo, E., Huwald, H. and Lehning, M.: Influence of Slope-Scale Snowmelt on Catchment Response Simulated With the Alpine3D Model, Water Resour. Res., 53, 723–739, doi:10.1002/2017WR021278, 2017.

Brown, M. E., Racoviteanu, A. E., Tarboton, D. G., Sen Gupta, A., Nigro, J., Policelli, F., Habib, S., Tokay, M., Shrestha, M. S., Bajracharya, S., Hummel, P., Gray, M., Duda, P., Zaitchik, B., Mahat, V., Artan, G. and Tokar, S.: An integrated modeling system for estimating glacier and snow melt driven streamflow from remote sensing and earth system data products in the Himalayas, J. Hydrol., 519, 1859–1869, doi:10.1016/j.jhydrol.2014.09.050, 2014.

Clark, M. P., Hendrikx, J., Slater, A. G., Kavetski, D., Anderson, B., Cullen, N. J., Kerr, T., Hreinsson, E. Ö. and Woods, R. A.: Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review, Water Resour. Res., 47, W07539, doi:10.1029/2011WR010745, 2011.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J. and Rasmussen, R. M.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resour. Res., 51(4), 2498–2514, doi:10.1002/2015WR017200.A, 2015.

Collier, E. and Immerzeel, W. W.: High-resolution modeling of atmospheric dynamics in the Nepalese Himalaya, J. Geophys. Res. Atmos., 120, 9882–9896, doi:10.1002/2015JD023266, 2015.

Collier, E., Mölg, T., Maussion, F., Scherer, D., Mayer, C. and Bush, A. B. G.: High-resolution interactive modelling of the mountain glacier-atmosphere interface: an application over the Karakoram, Cryosph., 7, 779–795, doi:10.5194/tc-7-779-2013, 2013.

Collier, E., Maussion, F., Nicholson, L. I., Mölg, T., Immerzeel, W. W. and Bush, A. B. G.: Impact of debris cover on glacier ablation and atmosphere-glacier feedbacks in the Karakoram, Cryosph., 9, 1617–1632, doi:10.5194/tc-9-1617-2015, 2015.

Corripio, J. G.: Vectorial algebra algorithms for calculating terrain parameters from DEMs and solar radiation modelling in mountainous terrain, Int. J. Geogr. Inf. Sci., 17(1), 1–23, doi:10.1080/713811744, 2003.

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R. and Smith, J.: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, Clim. Dyn., 15(3), 183–203, doi:10.1007/s003820050276, 1999.

Douville, H., Royer, J.-F. and Mahfouf, J.-F.: A new snow parameterization for the Météo-France climate model. Part

I: validation in stand-alone experiments, Clim. Dyn., 12(1), 21–35, doi:10.1007/BF00208760, 1995.

Duethmann, D., Zimmer, J., Gafurov, A., Güntner, A., Kriegel, D., Merz, B. and Vorogushyn, S.: Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modelling, Hydrol. Earth Syst. Sci., 17, 2415–2434, doi:10.5194/hess-17-2415-2013, 2013.

Dutra, E., Balsamo, G., Viterbo, P., Miranda, P. M., Beljaars, A., Schär, C. and Elder, K.: An improved snow scheme for the ECMWF land surface model: Description and offline validation, J. Hydrometeorol., 11(4), 899–916, doi:10.1175/2010JHM1249.1, 2010.

Essery, R.: Large-scale simulations of snow albedo masking by forests, Geophys. Res. Lett., 40, 5521–5525, doi:10.1002/grl.51008, 2013.

Essery, R.: A factorial snowpack model (FSM 1.0), Geosci. Model Dev., 8, 3867–3876, doi:10.5194/gmd-8-3867-2015, 2015.

Essery, R., Rutter, N., Pomeroy, J., Baxter, R., Stähli, M., Gustafsson, D., Barr, A., Bartlett, P. and Elder, K.: SNOWMIP2: An Evaluation of Forest Snow Process Simulations, Bull. Am. Meteorol. Soc., 90(8), 1120–1135, doi:https://doi.org/10.1175/2009BAMS2629.1, 2009.

Essery, R., Morin, S., Lejeune, Y. and Ménard, C. B.: A comparison of 1701 snow models using observations from an alpine site, Adv. Water Resour., 55, 131–148, doi:10.1016/j.advwatres.2012.07.013, 2013.

Etchevers, P., Martin, E., Brown, R., Fierz, C., Lejeune, Y., Bazile, E., Boone, A., Dai, Y., Essery, R., Fernandez, A., Gusev, Y., Jordan, R., Koren, V., Kowalczyk, E., Nasonova, N. O., Pyles, R. D., Schlosser, A., Shmakin, A. B., Smirnova, T. G., Strasser, U., Verseghy, D., Yamazaki, T. and Yang, Z.-L.: Validation of the energy budget of an alpine snowpack simulated by several snow models (SnowMIP project), Ann. Glaciol., 38, 150–158, doi:https://doi.org/10.3189/172756404781814825, 2004.

Finger, D., Pellicciotti, F., Konz, M., Rimkus, S. and Burlando, P.: The value of glacier mass balance, satellite snow cover images, and hourly discharge for improving the performance of a physically based distributed hydrological model, Water Resour. Res., 47, W07519, doi:10.1029/2010WR009824, 2011.

Forsythe, N., Kilsby, C. G., Fowler, H. J. and Archer, D. R.: Assessment of Runoff Sensitivity in the Upper Indus Basin to Interannual Climate Variability and Potential Change Using MODIS Satellite Data Products, Mt. Res. Dev., 32(1), 16–29, doi:10.1659/MRD-JOURNAL-D-11-00027.1, 2012.

Fowler, H. J. and Archer, D. R.: Hydro-climatological variability in the Upper Indus Basin and implications for water resources, in Regional Hydrological Impacts of Climatic Change - Impact Assessment and Decision Making (Proceedings of symposium S6 held during the Seventh IAHS Scientific Assembly at Foz do Iguacu, Brazil, April 2005), IAHS Publ. 295, Foz do Iguacu, Brazil., 2005.

Freudiger, D., Kohn, I., Seibert, J., Stahl, K. and Weiler, M.: Snow redistribution for the hydrological modeling of alpine catchments, WIREs Water, e1232, 1–16, doi:10.1002/wat2.1232, 2017.

Gafurov, A. and Bárdossy, A.: Cloud removal methodology from MODIS snow cover product, Hydrol. Earth Syst. Sci., 13, 1361–1373, doi:10.5194/hess-13-1361-2009, 2009.

Groot Zwaaftink, C. D., Mott, R. and Lehning, M.: Seasonal simulation of drifting snow sublimation in Alpine terrain, Water Resour. Res., 49, 1581–1590, doi:10.1002/wrcr.20137, 2013.

Grünewald, T., Schirmer, M., Mott, R. and Lehning, M.: Spatial and temporal variability of snow depth and ablation rates in a small mountain catchment, Cryosph., 4, 215–225, doi:10.5194/tc-4-215-2010, 2010.

Grünewald, T., Bühler, Y. and Lehning, M.: Elevation dependency of mountain snow depth, Cryosph., 8, 2381–2394, doi:10.5194/tc-8-2381-2014, 2014.

Günther, D., Marke, T., Essery, R. and Strasser, U.: Uncertainties in Snowpack Simulations—Assessing the Impact of Model Structure, Parameter Choice, and Forcing Data Error on Point-Scale Energy Balance Snow Model Performance, Water Resour. Res., 55, 2779–2800, doi:10.1029/2018WR023403, 2019.

Hall, A. and Qu, X.: Using the current seasonal cycle to constrain snow albedo feedback in future climate change, Geophys. Res. Lett., 33, L03502, doi:10.1029/2005GL025127, 2006.

Hall, D. K. and Riggs, G. A.: MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 6. [2000-2015], , doi:10.5067/MODIS/MOD10A1.006, 2016.

Harder, P., Pomeroy, J. W. and Helgason, W.: Local-Scale Advection of Sensible and Latent Heat During Snowmelt, Geophys. Res. Lett., 44, 9769–9777, doi:10.1002/2017GL074394, 2017.

Havens, S., Marks, D., FitzGerald, K., Masarik, M., Flores, A. N., Kormos, P. and Hedrick, A.: Approximating Input Data to a Snowmelt Model Using Weather Research and Forecasting Model Outputs in Lieu of Meteorological Measurements, J. Hydrometeorol., 20(5), 847–862, doi:10.1175/JHM-D-18-0146.1, 2019.

Huintjes, E., Sauter, T., Schröter, B., Maussion, F., Yang, W., Kropáček, J., Buchroithner, M., Scherer, D., Kang, S. and Schneider, C.: Evaluation of a coupled snow and energy balance model for Zhadang glacier, Tibetan Plateau, using glaciological measurements and time-lapse photography, Arctic, Antarct. Alp. Res., 47(3), 573–590, doi:10.1657/AAAR0014-073, 2015.

Immerzeel, W. W., van Beek, L. P. H. and Bierkens, M. F. P.: Climate Change Will Affect the Asian Water Towers, Science (80-.)., 382, 1382–1385, doi:10.1126/science.1183188, 2010.

Jarvis, A., Reuter, H. I., Nelson, A. and Guevara, E.: Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org), 2008.

Klemeš, V.: The modelling of mountain hydrology: the ultimate challenge, Hydrol. Mt. Areas (Proceedings Štrbské Pleso Work. Czechoslov. June 1989), IAHS 190, 29–44, 1990.

Krinner, G., Derksen, C., Essery, R., Flanner, M., Hagemann, S., Clark, M., Hall, A., Rott, H., Brutel-Vuilmet, C., Kim, H., Ménard, C. B., Mudryk, L., Thackeray, C., Wang, L., Arduini, G., Balsamo, G., Bartlett, P., Boike, J., Boone, A., Chéruy, F., Colin, J., Cuntz, M., Dai, Y., Decharme, B., Derry, J., Ducharne, A., Dutra, E., Fang, X., Fierz, C., Ghattas, J., Gusev, Y., Haverd, V., Kontu, A., Lafaysse, M., Law, R., Lawrence, D., Li, W., Marke, T., Marks, D., Ménégoz, M., Nasonova, O., Nitta, T., Niwano, M., Pomeroy, J., Raleigh, M. S., Schaedler, G., Semenov, V., Smirnova, T. G., Stacke, T., Strasser, U., Svenson, S., Turkov, D., Wang, T., Wever, N., Yuan, H., Zhou, W. and Zhu, D.: ESM-SnowMIP: assessing snow models and quantifying snow-related climate feedbacks, Geosci. Model Dev., 11, 5027–5049, doi:10.5194/gmd-11-5027-2018, 2018.

Lafaysse, M., Cluzet, B., Dumont, M., Lejeune, Y., Vionnet, V. and Morin, S.: A multiphysical ensemble system of numerical snow modelling, Cryosph., 11, 1173–1198, doi:10.5194/tc-11-1173-2017, 2017.

Lapo, K., Nijssen, B. and Lundquist, J. D.: Evaluation of Turbulence Stability Schemes of Land Models for Stable Conditions, J. Geophys. Res. Atmos., 124, 3072–3089, doi:10.1029/2018JD028970, 2019.

Lehning, M., Ingo, V., Gustafsson, D., Nguyen, T. A., Stähli, M. and Zappa, M.: ALPINE3D: a detailed model of mountain surface processes and its application to snow hydrology, Hydrol. Process., 20, 2111–2128, doi:10.1002/hyp.6204, 2006.

Liston, G. E. and Elder, K.: A Distributed Snow-Evolution Modeling System (SnowModel), J. Hydrometeorol., 7, 1259–1276, doi:10.1175/JHM548.1, 2006a.

Liston, G. E. and Elder, K.: A Meteorological Distribution System for High-Resolution Terrestrial Modeling (MicroMet), J. Hydrometeorol., 7, 217–234, doi:10.1175/JHM486.1, 2006b.

Lundquist, J. D., Dettinger, M. D. and Cayan, D. R.: Snow-fed streamflow timing at different basin scales: Case study of the Tuolumne River above Hetch Hetchy, Yosemite, California, Water Resour. Res., 41, W07005, doi:10.1029/2004WR003933, 2005.

Lute, A. C. and Luce, C. H.: Are Model Transferability and Complexity Antithetical? Insights from Validation of a Variable-Complexity Empirical Snow Model in Space and Time, Water Resour. Res., 53(11), 8825–8850, doi:10.1002/2017WR020752, 2017.

Lutz, A. F., Immerzeel, W. W., Shrestha, A. B. and Bierkens, M. F. P.: Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation, Nat. Clim. Chang., 4, 587–592, doi:10.1038/NCLIMATE2237, 2014.

Lutz, A. F., Immerzeel, W. W., Kraaijenbrink, P. D. A., Shrestha, A. B. and Bierkens, M. F. P.: Climate Change Impacts on the Upper Indus Hydrology: Sources, Shifts and Extremes, PLoS One, 11(11), 1–33, doi:10.1371/journal.pone.0165630, 2016.

MacDonald, M. K., Pomeroy, J. W. and Pietroniro, A.: On the importance of sublimation to an alpine snow mass balance in the Canadian Rocky Mountains, Hydrol. Earth Syst. Sci., 14, 1401–1415, doi:10.5194/hess-14-1401-2010, 2010.

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A. and Jonas, T.: Evaluating snow models with varying process representations for hydrological applications, Water Resour. Res., 51, 2707–2723, doi:10.1002/2014WR016498, 2015.

Marks, D., Domingo, J., Susong, D., Link, T. and Garen, D.: A spatially distributed energy balance snowmelt model for application in mountain basins, Hydrol. Process., 13, 1935–1959, doi:10.1002/(SICI)1099-1085(199909)13:12/13<1935::AID-HYP868>3.0.CO;2-C, 1999.

Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J. and Finkelnburg, R.: Precipitation Seasonality and Variability over the Tibetan Plateau as Resolved by the High Asia Reanalysis, J. Clim., 27, 1910–1927, doi:10.1175/JCLI-D-13-00282.1, 2014.

Moeser, D., Mazzotti, G., Helbig, N. and Jonas, T.: Representing spatial variability of forest snow: Implementation of a new interception model, Water Resour. Res., 52, 1208–1226, doi:10.1002/2015WR017961, 2016.

Mott, R., Schirmer, M., Bavay, M., Grünewald, T. and Lehning, M.: Understanding snow-transport processes shaping the mountain snow-cver, Cryosph., 4, 545–559, doi:10.5194/tc-4-545-2010, 2010.

Mott, R., Scipión, D., Schneebeli, M., Dawes, N., Berne, A. and Lehning, M.: Orographic effects on snow deposition patterns, J. Geophys. Res. Atmos., 119, 1419–1439, doi:10.1002/2013JD019880, 2014.

Mott, R., Vionnet, V. and Grünewald, T.: The Seasonal Snow Cover Dynamics: Review on Wind-Driven Coupling Processes, Front. Earth Sci., 6(197), 1–25, doi:10.3389/feart.2018.00197, 2018.

Mukhopadhyay, B. and Khan, A.: Altitudinal variations of temperature, equilibrium line altitude, and accumulation-area ratio in Upper Indus Basin, Hydrol. Res., 48(1), 214–230, doi:10.2166/nh.2016.144, 2016.

Musselman, K. N., Pomeroy, J. W., Essery, R. L. H. and Leroux, N.: Impact of windflow calculations on simulations of alpine snow accumulation, redistribution and ablation, Hydrol. Process., 29, 3983–3999, doi:10.1002/hyp.10595, 2015.

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K. and Rasmussen, R.: Slower snowmelt in a warmer world, Nat. Clim. Chang., 7, 214–220, doi:10.1038/NCLIMATE3225, 2017.

Naden, P. S.: Spatial variability in flood estimation for large catchments: the exploitation of channel network structure, Hydrol. Sci. J., 37(1), 53–72, doi:10.1080/02626669209492561, 1992.

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M. and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1 . Model description and evaluation with local-scale measurements, J. Geophys. Res., 116, D12109, doi:10.1029/2010JD015139, 2011.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, W. J., Subin, Z. M., Swenson, S. C., Thornton, P. E., Bozbiyik, A., Fisher, R., Kluzek, E., Lamarque, J.-F., Lawrence, P. J., Leung, L. R., Lipscomb, W., Muszala, S., Ricciuto, D. M., Sacks, W., Sun, Y., Tang, J. and Yang, Z.-L.: Technical Description of version 4.5 of the Community Land Model (CLM). NCAR Technical Note NCAR/TN-503+STR, Boulder, Colorado, USA., 2013.

Palazzi, E., Filippi, L. and Hardenberg, J. Von: Insights into elevation-dependent warming in the Tibetan Plateau-Himalayas from CMIP5 model simulations, Clim. Dyn., 48(11), 3991–4008, doi:10.1007/s00382-016-3316-z, 2017.

Pepin, N., Bradley, R., Diaz, H. F., Baraer, M., Caceres, E. B., Forsythe, N., Fowler, H. J., Greenwood, G., Hashmi, M. Z., Liu, X. D., Miller, J. R., Ning, L., Ohmura, A., Palazzi, E., Rangwala, I., Schoener, W., Severskiy, I., Shahgedanova, M., Wang, M. B., Williamson, S. N. and Yang, D. Q.: Elevation-dependent warming in mountain regions of the world, Nat. Clim. Chang., 5, 424–430, doi:10.1038/nclimate2563, 2015.

Prasch, M., Mauser, W. and Weber, M.: Quantifying present and future glacier melt-water contribution to runoff in a central Himalayan river basin, Cryosph., 7, 889–904, doi:10.5194/tc-7-889-2013, 2013.

Pritchard, D. M. W., Forsythe, N., Fowler, H. J., O'Donnell, G. M. and Li, X.-F.: Evaluation of Upper Indus Near-Surface Climate Representation by WRF in the High Asia Refined Analysis, J. Hydrometeorol., 20(3), 467–487, doi:10.1175/JHM-D-18-0030.1, 2019.

Quéno, L., Vionnet, V., Dombrowski-Etchevers, I., Lafaysse, M., Dumont, M. and Karbou, F.: Snowpack modelling in the Pyrenees driven by kilometric- resolution meteorological forecasts, Cryosph., 10, 1571–1589, doi:10.5194/tc-10-1571-2016, 2016.

Ragettli, S., Pellicciotti, F., Bordoy, R. and Immerzeel, W. W.: Sources of uncertainty in modeling the glaciohydrological response of a Karakoram watershed to climate change, Water Resour. Res., 49, 6048–6066, doi:10.1002/wrcr.20450, 2013.

Ragettli, S., Pellicciotti, F., Immerzeel, W. W., Miles, E. S., Petersen, L., Heynen, M., Shea, J. M., Stumm, D., Joshi, S. and Shrestha, A.: Unraveling the hydrology of a Himalayan catchment through integration of high resolution in situ data and remote sensing with an advanced simulation model, Adv. Water Resour., 78, 94–111, doi:10.1016/j.advwatres.2015.01.013, 2015.

Raleigh, M. S., Lundquist, J. D. and Clark, M. P.: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework, Hydrol. Earth Syst. Sci., 19, 3153–3179, doi:10.5194/hess-19-3153-2015, 2015.

Raleigh, M. S., Livneh, B., Lapo, K. and Lundquist, J. D.: How Does Availability of Meteorological Forcing Data Impact Physically Based Snowpack Simulations?, J. Hydrometeorol., 17, 99–120, doi:10.1175/JHM-D-14-0235.1, 2016.

Revuelto, J., Lafaysse, M., Zin, I., Charrois, L., Vionnet, V., Dumont, M., Rabatel, A., Six, D., Condom, T., Morin, S., Viani, A. and Sirguey, P.: Multi-Criteria Evaluation of Snowpack Simulations in Complex Alpine Terrain Using Satellite and In Situ Observations, Remote Sens., 10(1171), 1–32, doi:10.3390/rs10081171, 2018.

Rutter, N., Essery, R., Pomeroy, J., Altimir, N., Andreadis, K., Baker, I., Barr, A., Bartlett, P., Boone, A., Deng, H., Douville, H., Dutra, E., Elder, K., Ellis, C., Feng, X., Gelfan, A., Goodbody, A., Gusev, Y., Gustafsson, D., Hellström, R., Hirabayashi, Y., Hirota, T., Jonas, T., Koren, V., Kuragina, A., Lettenmaier, D., Li, W., Luce, C., Martin, E., Nasonova, O., Pumpanen, J., Pyles, R. D., Samuelsson, P., Sandells, M., Schädler, G., Shmakin, A., Smirnova, T. G., Stähli, M., Stöckli, R., Strasser, U., Su, H., Suzuki, K., Takata, K., Tanaka, K., Thompson, E., Vesala, T., Viterbo, P., Wiltshire, A., Xia, K., Xue, Y. and Yamazaki, T.: Evaluation of forest snow processes models (SnowMIP2), J. Geophys. Res., 114, D06111, doi:10.1029/2008JD011063, 2009.

Salomonson, V. V and Appel, I.: Estimating fractional snow cover from MODIS using the normalized difference snow index, Remote Sens. Environ., 89, 351–360, doi:10.1016/j.rse.2003.10.016, 2004.

Shakoor, A. and Ejaz, N.: Flow Analysis at the Snow Covered High Altitude Catchment via Distributed Energy Balance Modeling, Sci. Rep., 9(4783), 1–14, doi:10.1038/s41598-019-39446-1, 2019.

Sharif, M., Archer, D. R., Fowler, H. J. and Forsythe, N.: Trends in timing and magnitude of flow in the Upper Indus

Basin, Hydrol. Earth Syst. Sci., 17, 1503–1516, doi:10.5194/hess-17-1503-2013, 2013.

Shrestha, M., Koike, T., Hirabayashi, Y., Xue, Y., Wang, L., Rasul, G. and Ahmad, B.: Integrated simulation of snow and glacier melt in water and energy balance-based, distributed hydrological modeling framework at Hunza River Basin of Pakistan Karakoram region, J. Geophys. Res. Atmos., 120(10), 4889–4919, doi:10.1002/2014JD022666, 2015.

Sicart, J. E., Pomeroy, J. W., Essery, R. L. H. and Bewley, D.: Incoming longwave radiation to melting snow: observations, sensitivity and estimation in northern environments, Hydrol. Process., 20, 3697–3708, doi:10.1002/hyp.6383, 2006.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W. and Power, J. G.: A Description of the Advanced Research WRF Version 3, NCAR Technical Note NCAR/TN–475+STR., 2008.

Slater, A. G., Schlosser, C. A., Desborough, C. E., Pitman, A. J., Henderson-Sellers, A., Robock, A., Vinnikov, K. Y., Mitchell, K., Boone, A., Braden, H., Chen, F., Cox, P. M., de Rosnay, P., Dickinson, R. E., Dai, Y.-J., Duan, Q., Entin, J., Etchevers, P., Gedney, N., Gusev, Y. M., Kim, J., Koren, V., Kowalczyk, E. A., Nasonova, O. N., Noilhan, J., Schaake, S., Shmakin, A. B., Smirnova, T. G., Verseghy, D., Wetzel, P., Xue, Y., Yang, Z.-L. and Zeng, Q.: The Representation of Snow in Land Surface Schemes: Results from PILPS 2 (d), J. Hydrometeorol., 2, 7–25, doi:10.1175/1525-7541(2001)002<0007:TROSIL>2.0.CO;2, 2001.

Stigter, E. E., Wanders, N., Saloranta, T. M., Shea, J. M., Bierkens, M. F. P. and Immerzeel, W. W.: Assimilation of snow cover and snow depth into a snow model to estimate snow water equivalent and snowmelt runoff in a Himalayan catchment, Cryosph., 11, 1647–1664, doi:10.5194/tc-11-1647-2017, 2017.

Strasser, U., Bernhardt, M., Weber, M., Liston, G. E. and Mauser, W.: Is snow sublimation important in the alpine water balance?, Cryosph., 2, 53–66, doi:10.5194/tc-2-53-2008, 2008.

Tarasova, L., Knoche, M., Dietrich, J. and Merz, R.: Effects of input discretization, model complexity, and calibration strategy on model performance in a data-scarce glacierized catchment in Central Asia, Water Resour. Res., 52, 4674–4699, doi:10.1002/2015WR018551, 2016.

Verseghy, D. L.: CLASS—A Canadian land surface scheme for GCMs. I. Soil model, Int. J. Climatol., 11(2), 111–133, doi:10.1002/joc.3370110202, 1991.

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E. and Willemet, J.-M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, Geosci. Model Dev., 5, 773–791, doi:10.5194/gmd-5-773-2012, 2012.

Vionnet, V., Martin, E., Masson, V., Guyomarc'h, G., Naaim-Bouvet, F., Prokop, A., Durand, Y. and Lac, C.: Simulation of wind-induced snow transport and sublimation in alpine terrain using a fully coupled snowpack/atmosphere model, Cryosph., 8, 395–415, doi:10.5194/tc-8-395-2014, 2014.

Vionnet, V., Martin, E., Masson, V., Lac, C., Bouvet, F. N. and Guyomarc'h, G.: High-Resolution Large Eddy Simulation of Snow Accumulation in Alpine Terrain, J. Geophys. Res. Atmos., 122, 11005–11021, doi:10.1002/2017JD026947, 2017.

Viviroli, D., Archer, D. R., Buytaert, W., Fowler, H. J., Greenwood, G. B., Hamlet, A. F., Huang, Y., Koboltschnig, G., Litaor, M. I., López-Moreno, J. I., Lorentz, S., Schädler, B., Schreier, H., Schwaiger, K., Vuille, M. and Woods, R.: Climate change and mountain water resources: overview and recommendations for research, management and policy, Hydrol. Earth Syst. Sci., 15, 471–504, doi:10.5194/hess-15-471-2011, 2011.

Wan, Z.: New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product, Remote Sens. Environ., 140, 36–45, doi:10.1016/j.rse.2013.08.027, 2014.

Wan, Z., Zhang, Y., Zhang, Q. and Li, Z.-L.: Quality assessment and validation of the MODIS global land surface temperature, Int. J. Remote Sens., 25(1), 261–274, doi:10.1080/0143116031000116417, 2004.

Waqas, A. and Athar, H.: Observed diurnal temperature range variations and its association with observed cloud cover in northern Pakistan, Int. J. Climatol., 38(8), 3323–3336, doi:10.1002/joc.5503, 2018.

Warscher, M., Strasser, U., Kraller, G., Marke, T., Franz, H. and Kunstmann, H.: Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps, Water Resour. Res., 49, 2619–2637, doi:10.1002/wrcr.20219, 2013.

Winstral, A., Marks, D. and Gurney, R.: Simulating wind-affected snow accumulations at catchment to basin scales, Adv. Water Resour., 55, 64–79, doi:10.1016/j.advwatres.2012.08.011, 2013.

Process Description	Short Name	Parameterisation 0	Parameterisation 1
Snow albedo variation	Albedo (AL)	Diagnostic – function of surface temperature	Prognostic – decays with time and increases with snowfall
Density of fresh snow and snowpack density evolution	Density (DE)	Constant	Specified fresh snow density and compaction increases with time
Liquid water storage, drainage and refreezing	Liquid Water (LW)	Instant drainage, no refreezing	Bucket model (drainage to layer below if liquid holding capacity exceeded), with refreezing (and latent heat release) accounted for
Atmospheric stability adjustment for turbulent heat fluxes	Stability (ST)	No adjustment for atmospheric stability	Stability factor is a function of the bulk Richardson number, which quantifies the extent to which buoyancy suppresses shear production of turbulent fluxes
Thermal conductivity for heat conduction	Thermal Conductivity (TC)	Constant	Function of density

Table 1. Summary of the process parameterisation options available in FSM. Full details are provided in Essery (2015). The short names and abbreviations by which the processes are referred to in the text and figures are given.

Variable	Month / Process											
	April			May			June			July		
	AL	LW	ST	AL	LW	ST	AL	LW	ST	AL	LW	ST
Albedo (-)	0.12	0.00	0.00	0.15	0.00	0.00	0.12	0.00	0.00	0.07	0.00	0.00
Melt (mm/d)	-3.5	1.0	0.2	-5.2	1.9	-0.3	-4.3	2.7	-4.6	-0.1	3.5	-7.7
Snowpack Runoff (mm/d)	-2.2	-1.9	0.0	-2.7	-1.6	-0.2	0.0	0.6	-1.0	2.8	1.4	0.1
SCA (-)	0.04	0.02	0.02	0.08	0.04	0.04	0.12	0.06	0.07	0.10	0.05	0.09
Snowpack Temperature (°C)	-0.6	3.5	-1.9	-0.4	6.3	-1.6	0.1	7.3	-1.5	0.3	8.1	-1.1
Surface Temperature (°C)	0.1	0.9	-2.9	0.3	1.4	-3.2	0.4	1.5	-3.9	0.5	1.3	-3.6
Sensible Heat Flux (W m^{-2})	0.6	-1.2	-22.0	0.8	-1.6	-25.9	1.5	-0.7	-46.6	3.1	0.4	-49.4
Net Turbulent Fluxes (W m^{-2})	1.2	-2.4	-16.0	1.3	-3.0	-18.9	2.0	-1.8	-36.8	4.4	-0.3	-52.7
Net Radiation (W m-2)	-13.0	-1.8	12.3	-20.1	-2.9	13.8	-18.1	-2.0	16.2	-3.7	0.3	20.0

Table 2. Catchment-scale mean differences (option 1 – option 0) for key states and fluxes in selected months. Differences are calculated separately for the albedo (AL), liquid water (LW) and stability adjustment (ST) parameterisations. Albedo differences are at solar noon and average snowpack temperatures are weighted by snow depth.

Error ID	SCA Errors	Explanation
i	Underestimation of the negative spring/summer SCA anomaly in 2001.	The contemporaneous and preceding (negative) precipitation anomalies are reasonable, but the HAR does not capture the strongly positive temperature anomalies. Under these erroneously cool simulated conditions, the faster-responding configurations result in lower errors than the more physically realistic, slower-responding configurations (i.e. prognostic albedo and a representation of liquid water processes).
ii	Positive simulated SCA anomaly in summer 2002 exceeds the neutral anomaly suggested by MODIS.	The HAR inputs provide a positive spring precipitation anomaly that far exceeds the observed anomaly, while again offering too negative a temperature anomaly. These conditions are conducive to slow melt of the excessive spring snowfall, which helps to explain the poorer performance of the slower-responding configurations.
iii	Positive simulated SCA anomaly in late-summer/autumn 2005 contrasts with a negative anomaly from MODIS.	This is likely due at least partly to HAR overestimation of precipitation in the preceding winter and spring, as well as potentially in the concurrent summer. The HAR may also underestimate the temperature anomaly in the crucial summer months.
iv	Positive late-summer/autumn SCA anomaly in 2008 in MODIS is not reproduced by the model.	The magnitude of the positive precipitation anomaly at this time seems to be underestimated by the HAR, which also provides too positive a temperature anomaly. This would be conducive to melt of early snowfall and underestimation of the positive SCA anomaly. Unlike the summer examples, the resulting absolute errors are similar for all configurations, which reflects persistent challenges in simulating the timing of early autumn/winter snowfall.
vi	The magnitude of the strong positive summer SCA anomaly in 2010 is underestimated.	This coincides with the largest spread in simulated anomalies in the series. Precipitation anomalies in the preceding winter and spring are underestimated, along with the large summer anomaly, which coincided with devastating floods in Pakistan). Summer temperature anomalies are notably overestimated. All configurations underestimate summer SCA by varying degrees.

Table 3. Assessment of SCA time series errors in Figure 6~~Figure 8~~ and their relationships with climate anomalies. Error IDs correspond with Figure 6~~Figure 8~~.

Formatted: Font: Not Bold

Formatted: Font: Not Bold

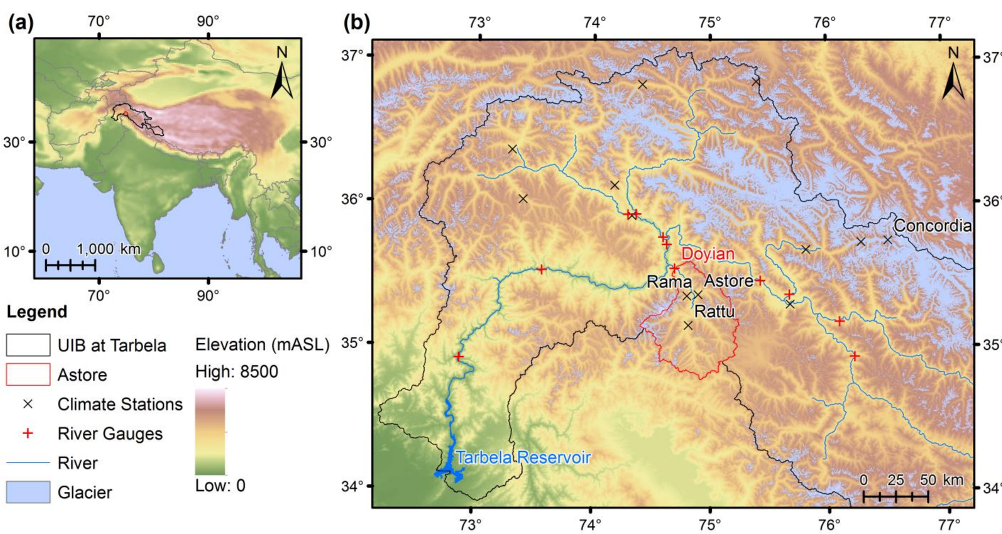


Figure 1. Location of study area and local measurement points. The regional context is indicated in (a). The Astore catchment and observation locations (with labels for the most important sites in this study) are shown with topography and glacier extent in (b). The SRTM 90 m DEM v4.1© (Jarvis et al., 2008) and Randolph Glacier Inventory 5.0© (Arendt et al., 2015) datasets are both licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Formatted: Font: Not Bold

Formatted: Font: Not Bold

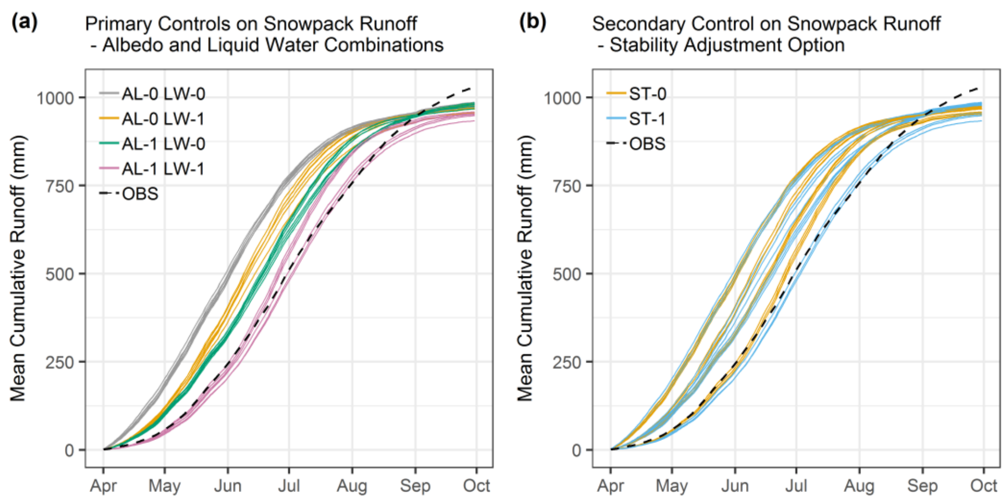


Figure 2. ~~Comparison of m~~Mean cumulative snowpack runoff for the high-flow season for each of the 32 ensemble members~~ with observed total runoff (OBS, black dashed line)~~. In (a) each ensemble member is coloured according to the combination of albedo (AL) and liquid water (LW) parameterisations it uses. In (b) each ensemble member is coloured by its stability adjustment (ST) option. Observed total runoff (OBS, black dashed line) is shown for reference only (it is not directly comparable with snowpack runoff – see Section 3.3).

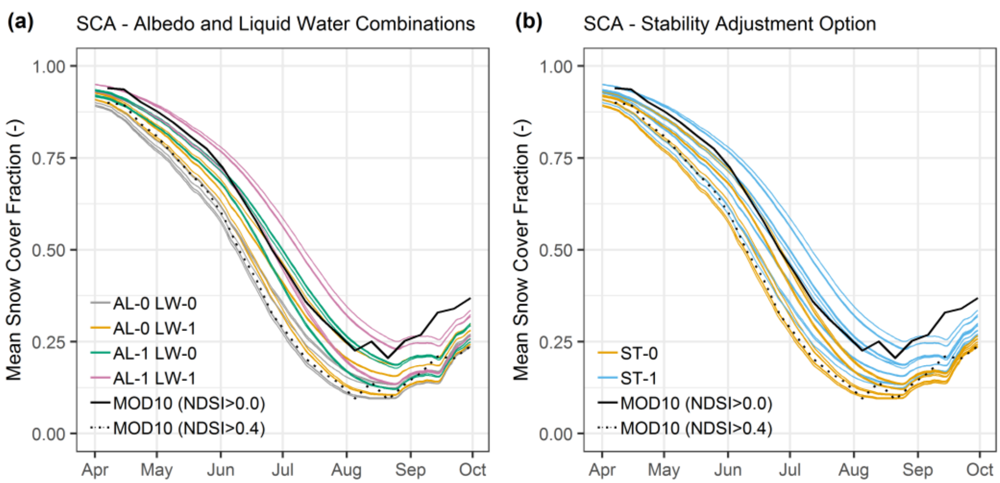
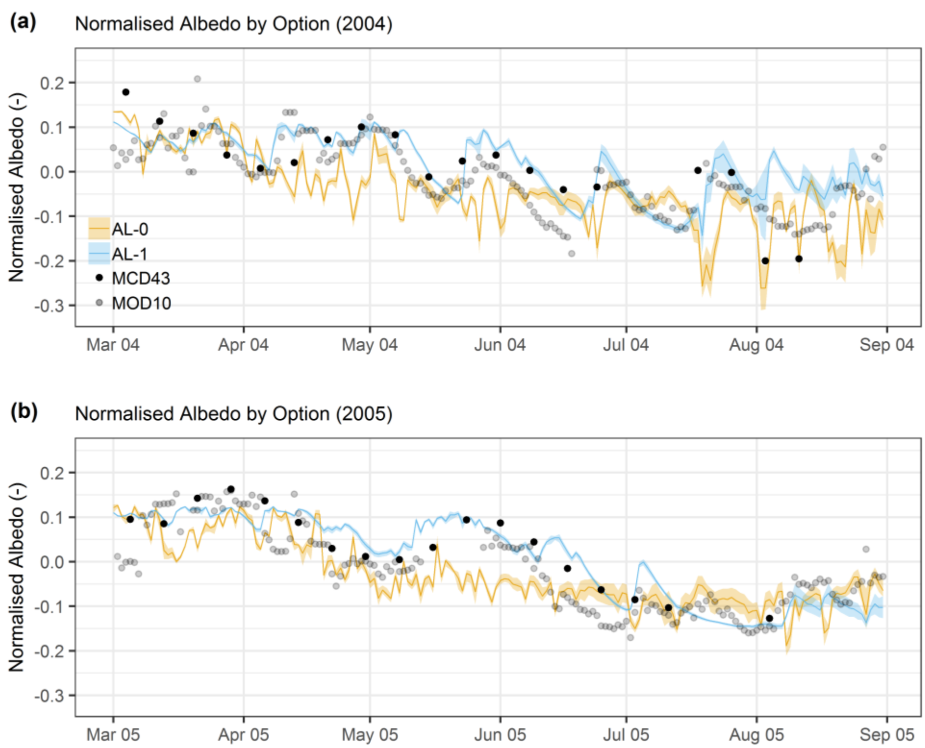


Figure 3. Similar to Figure 2~~Figure 2~~ but for catchment snow-covered area (SCA). The two MODIS MOD10A1 series shown are based on Normalised Difference Snow Index (NDSI) thresholds of 0.0 (solid line) and 0.4 (dotted line).

Formatted: Font: Not Bold



~~Figure 4. Comparison of modelled catchment average snow albedo with MODIS remote sensing for two example melt seasons. Modelled albedo is grouped by the diagnostic (0) and prognostic (1) options in orange and blue, respectively. The mean (line) and range (shading) for the two groups are shown. MCD43A3 and MOD10A1 (8-day moving average) estimates are denoted with black and grey dots, respectively. The modelled series are normalised by subtracting the ensemble mean albedo (all members), while the MODIS series are normalised by subtracting their respective means.~~

Formatted: Normal, Line spacing: single

Formatted: Justified, Indent: First line: 0.75 cm, Line spacing: 1.5 lines

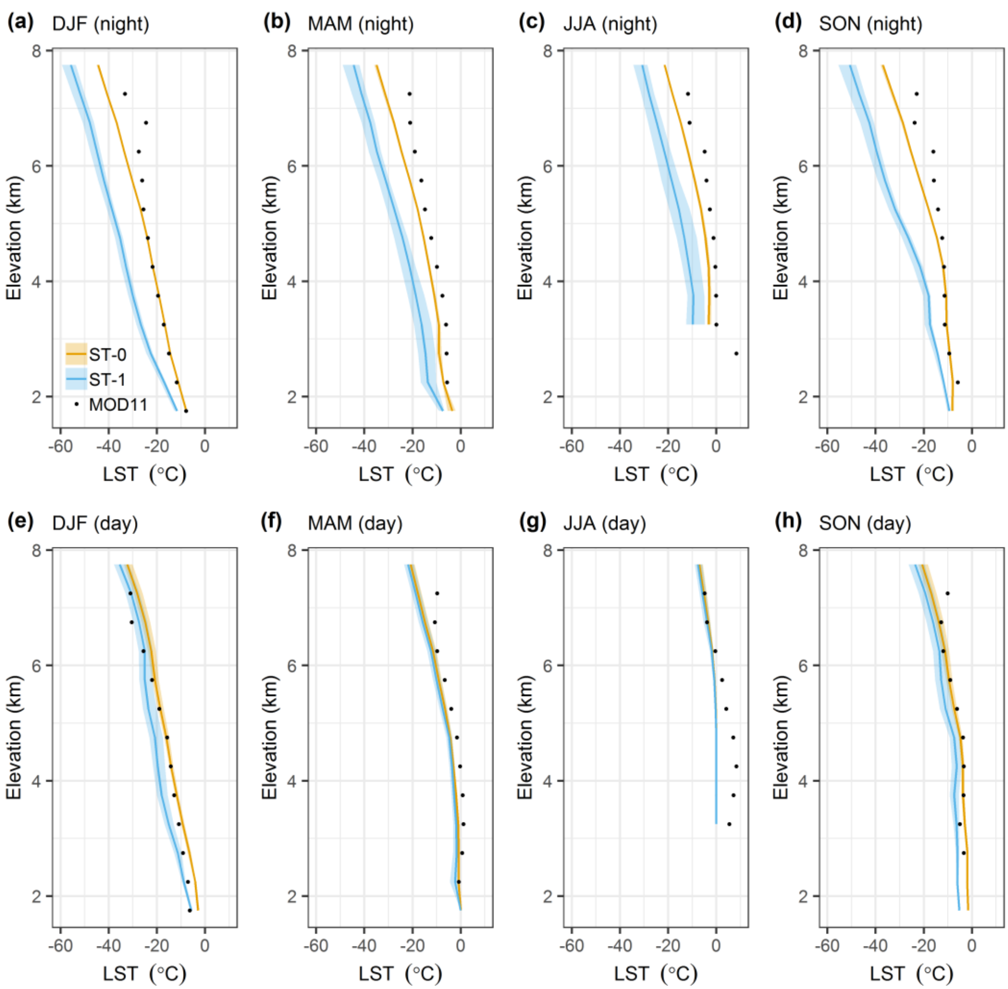


Figure 4~~5~~. Comparison of modelled seasonal mean elevation profiles of LST with MODIS MOD11A1 remote sensing. The ensemble members are grouped by stability adjustment option (mean and range of groups shown). The top and bottom rows show night-time and day-time temperatures, respectively. Model results correspond with the closest model time step to the Terra platform overpass times, as well as only days for which MODIS retrievals are available (i.e. clear-sky conditions).

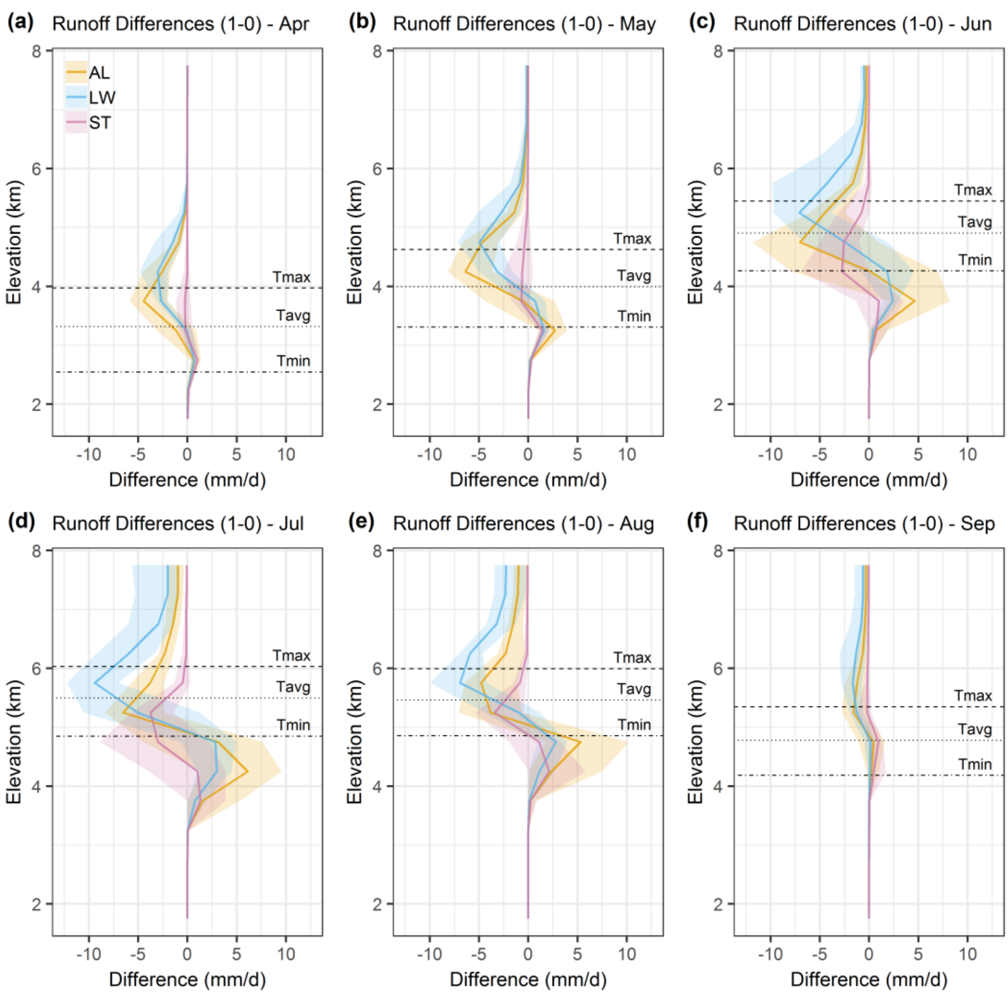
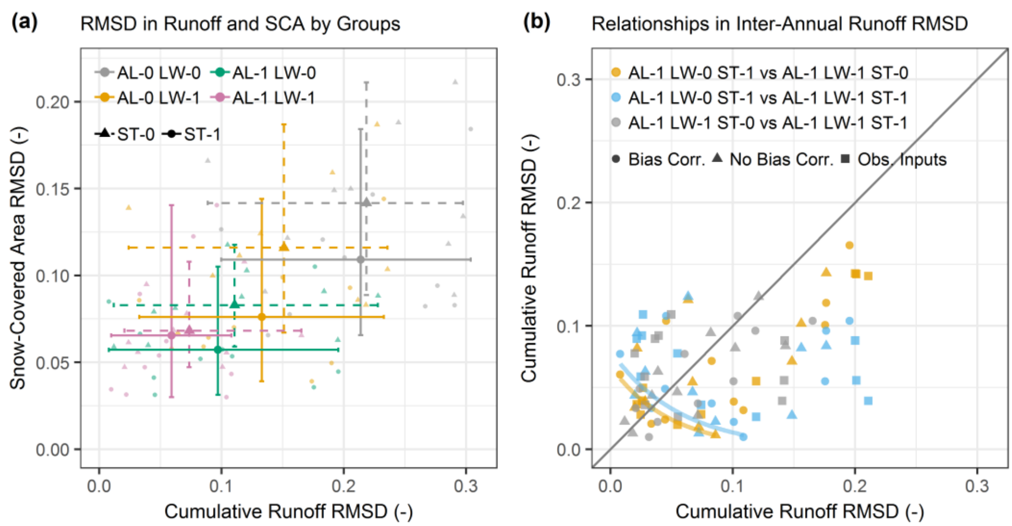


Figure 5~~6~~. Spatial (vertical) and temporal (monthly) differences in simulated snowpack runoff as a result of albedo (AL), liquid water (LW) and stability option (ST) choices. The differences are calculated ~~by~~ as option 1 – option 0 for each process. Lines show mean differences, while ranges denote inter-annual variability. Monthly mean freezing isotherm elevations for daily minimum, mean and maximum temperatures are also shown.



~~Figure 7. **Inter-annual variation in model performance. Cumulative runoff RMSD is plotted against SCA RMSD in (a) for each of the albedo (AL) and liquid water (LW) option combinations (i.e. averaging respective ensemble members), with differentiation by stability (ST) option (shape and line type) also shown. Cumulative snowpack runoff RMSD is for the April to June period and normalised, whereas SCA RMSD is for April to September and based on an NDSI threshold of 0. In (b), the relationships between cumulative runoff RMSD for the three best-performing combinations are shown for each of the climate input strategies (see Section 3.2.3).**~~

Formatted: Normal, Line spacing: single

Formatted: Justified, Indent: First line: 0.75 cm, Line spacing: 1.5 lines

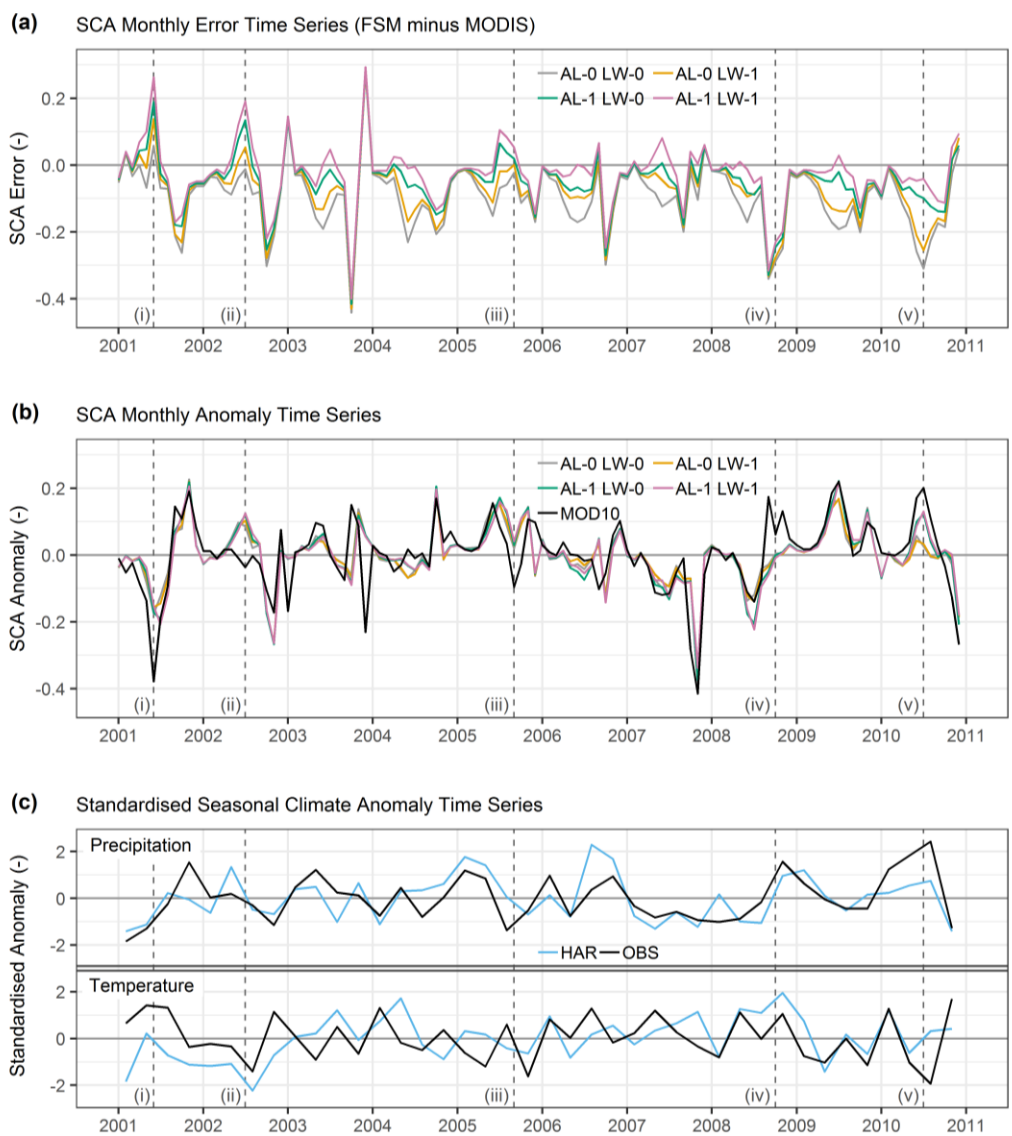


Figure 6~~8~~. Comparison of model SCA performance in absolute and anomaly space with climate inputs. Monthly time series show (a) SCA errors relative to MOD10A1 (FSM minus MODIS) and (b) anomalies (subtracting monthly means from absolute SCA) after aggregating the ensemble by the combinations of albedo and liquid water parameterisations. Standardised seasonal precipitation and temperature anomalies based on observations aggregated from the climate stations in Figure 1~~Figure 1~~b and the HAR are given in (c).

Formatted: Font: Not Bold

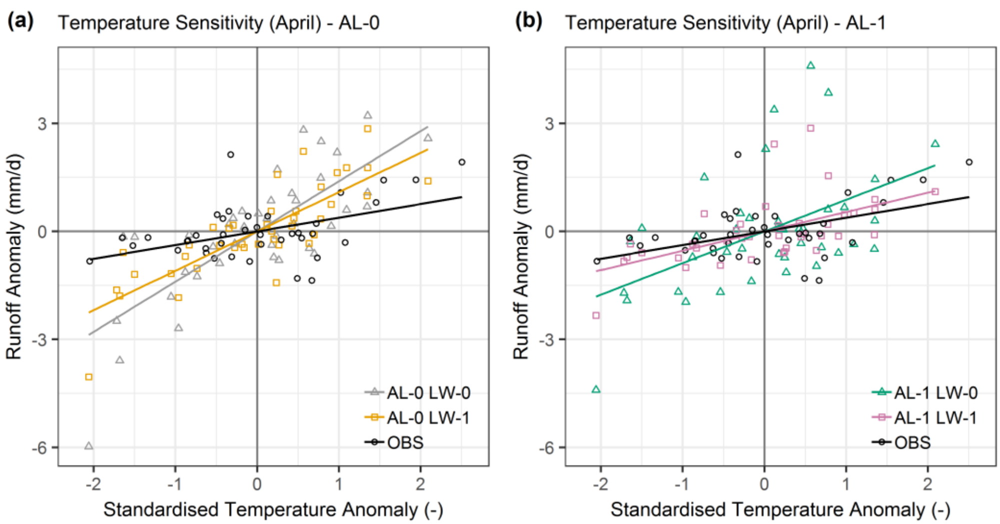


Figure 7~~9~~. Sensitivity of simulated snowpack runoff (and observed total runoff) in April to temperature anomalies for combinations of albedo and liquid water process options (split across two panels for clarity). Points represent 10-day anomalies and lines are from linear regression.

Supplement to: Multi-physics ensemble snow modelling in the western Himalaya

S1 Climate Input Strategies

The baseline climate inputs for this study are based primarily on downscaling and bias correction (for temperature) of the High Asia Refined Analysis (HAR, Maussion et al., 2014) dynamical downscaling product, as described in Section 3.2. Given the uncertainties in climate input fields in this data-sparse context, simulations were also performed using two alternative input derivation strategies (Section 3.2.3). These strategies are summarised in Table S1 below. The strategies are not independent, as their main purpose is to indicate whether the conclusions reached on snowpack process representations, the focus of this study, are unduly affected by the downscaling and bias correction approaches described in Section 3.2.2. Precipitation is kept consistent between strategies, as the HAR represents by far the best available source of distributed precipitation fields (Pritchard et al., 2019). The focus is thus on climate variables used in surface energy balance calculations. The implications of using these alternative input strategies are discussed in Section 5 and Section S5.

Input Strategy	Description
1	Downscaled HAR using approach described in Section 3.2.2, with simple bias correction of temperature fields. This is considered to be the best approach with available data and thus the baseline for the study.
2	Downscaled HAR using approach described in Section 3.2.2, but without any bias correction of temperature fields. The primary purpose of this strategy (2) is to check whether the temperature bias correction applied in (1) alters inter-variable relationships in a way which affects ensemble structure.
3	Downscaled HAR precipitation as per Section 3.2.2, but with other climate fields estimated primarily from observations. Specifically, temperature is lapsed based on observations (separately for daily minima and maxima, using monthly lapse rates). Daily temperatures are disaggregated to an hourly interval based on normalised climatological hourly diurnal cycle from EvK2CNR stations for each month. Relative humidity is estimated from daily minimum and maximum observations, and disaggregated to an hourly time step using a similar approach. Incoming shortwave radiation is calculated as per strategies (1) and (2). However, rather than estimating cloud transmissivity from the HAR (Section 3.2.2), a parameterisation of cloud transmissivity based on diurnal temperature range is used following calibration with local data (Pellicciotti et al., 2011). Incoming longwave radiation is estimated using the formulation from MicroMet (Liston and Elder, 2006). Wind speed is ~~taken from~~based on the HAR in the absence of observations, as per strategies (1) and (2).

Table S1. Summary of baseline and alternative climate input data sources and strategies.

S2 MODIS Land Surface Temperature (LST) Validation

~~Figure S2~~Figure S1 compares MODIS MOD11A1 Collection 6 land surface temperature (LST) with observations from the EvK2CNR Concordia site (Figure 1b), in order to provide further validation of the remote sensing product in this region (Section 3.3). Observed LST was estimated from measured outgoing longwave radiation at Concordia for the hours closest to the MODIS overpass times. The corresponding MODIS LST values were based on the average of the 9 pixels surrounding a station location and were corrected for elevation differences using local MODIS LST lapse rates (estimated from linear regression). Figure S1~~Figure S2~~ shows that the MODIS LST shows good correspondence with observations overall, as reflected by the values lying generally close to the 1:1 line. The summary statistics in Table S2 ~~Table S1~~ confirm that the MODIS bias is low at the annual scale (-1.6°C for night-time and 0.5°C for daytime), although it may be slightly larger for individual seasons. Nevertheless, Table S2~~Table S1~~ indicates that MOD11A1 is likely accurate enough to estimate climatological LST to within 2-3°C, depending on season.

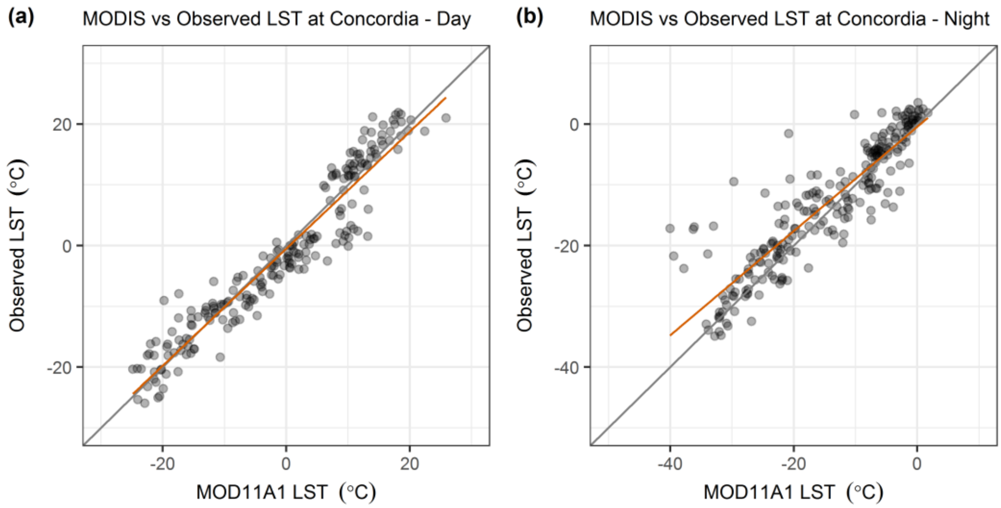


Figure S1. Comparison of (a) daytime and (b) night-time LST from MODIS remote sensing (MOD11A1) with observations at the Concordia site. Observed LST was derived from measured outgoing longwave radiation for the hour closest to the MODIS overpass time. The orange lines represent best fits from linear regression.

Season	**Bias (°C)**		**RMSE (°C)**	
	Night	Day	Night	Day
Annual	-1.6	0.5	5.0	3.7
DJF	-3.5	-1.4	3.4	7.2
MAM	-0.2	2.1	3.6	4.7
JJA	-0.7	1.4	5.0	3.7
SON	-2.8	-1.4	2.2	4.3

Table S2. Summary statistics for MOD11A1 performance at the EvK2CNR Concordia site.

S3 Simulated Albedo Evaluation

Formatted: Heading 1, Line spacing: single

This section provides an evaluation of albedo simulated by the diagnostic and prognostic albedo parameterisations using the MODIS MCD43A3 and MOD10A1 surface albedo products. These datasets have been validated in different settings (Gascoin et al., 2017; Liu et al., 2009; Wang et al., 2012), but additional challenges are posed by complex terrain (Wen et al., 2018). The evaluation thus considers the model results and datasets in both absolute and anomaly terms. The modelled series are transformed to anomalies by subtracting the ensemble mean albedo (all members), while the two MODIS series are converted to anomalies by subtracting their respective means. The evaluation focuses on catchment-scale albedo. Spatial aggregates were only calculated when 90% of pixels had satisfactory quality data, which are defined here as any processed inversion for MCD43A3 albedo.

In absolute terms, FSM mean snow albedo from March to August exceeds MODIS by 0.07 and 0.14 for MOD10A1 and MCD43A3, respectively. This may be in part due to challenges in fully characterising albedo in complex terrain with remote sensing (Wen et al., 2018). However, in anomaly terms, Figure S2 demonstrates that the prognostic albedo parameterisation is in reasonable agreement with MODIS. Acknowledging some timing offsets and points of divergence between the MODIS products, prognostic albedo more skilfully captures the sharp albedo increases following snowfall in the melt season. Field studies have shown these events to be an important for regional melt rate variability, especially early in the season, in accordance with the latitude (~35°N) and continentality of the area (Hewitt, 2014). The prognostic parameterisation also generally reproduces the rate of albedo decay during melting periods in Figure S2, whereas the diagnostic parameterisation induces frequent, sharp and pronounced albedo fluctuations. These fluctuations give rise to a comparatively low albedo in the early melt season. In quantitative terms, the prognostic parameterisation outperforms the diagnostic option for the anomaly series, with an overall RMSD relative to the MOD10A1 product of 0.062 (prognostic) compared with 0.071 (diagnostic). Process-level evaluation with MODIS thus corroborates the better performance of prognostic albedo for simulating catchment SCA (Section 4.1).

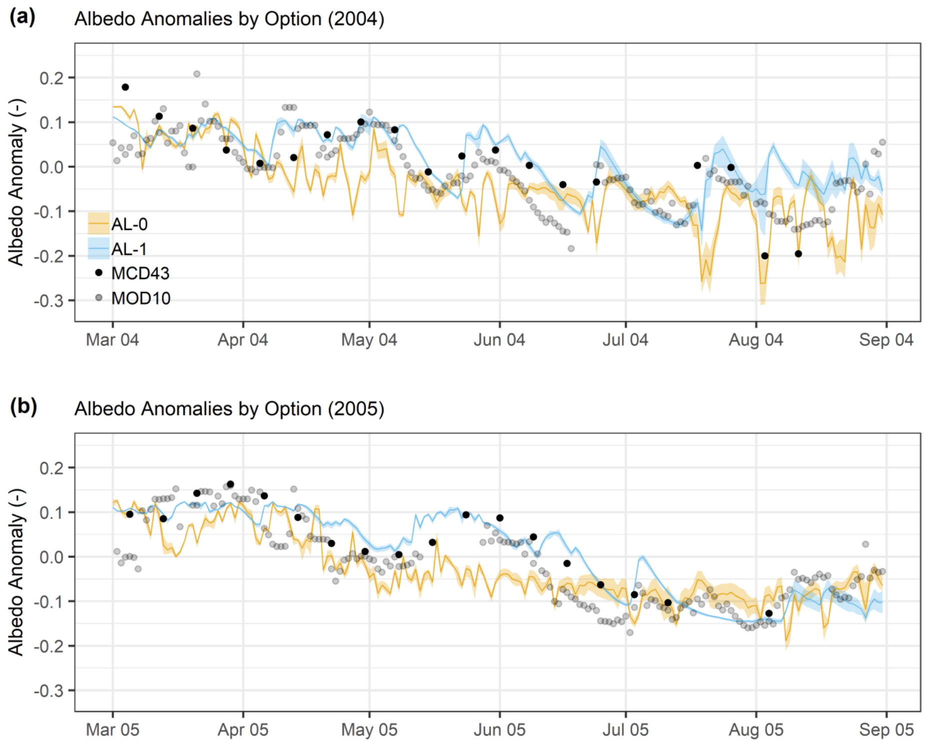


Figure S2. Comparison of modelled catchment-average snow albedo anomalies with MODIS remote sensing for two example melt seasons. Modelled albedo is grouped by the diagnostic (0) and prognostic (1) options in orange and blue, respectively. The mean (line) and range (shading) for the two groups are shown. MCD43A3 and MOD10A1 (8-day moving average) estimates are denoted with black and grey dots, respectively. The modelled series are converted to anomalies by subtracting the ensemble mean albedo (all members), while the MODIS series are converted by subtracting their respective means.

S4 ~~Model Performance Trade-Offs~~Model Deviations from Observations

Figure S3 shows the relationship between the root-mean-square deviation (RMSD) for cumulative snowpack runoff (simulated relative to observed total runoff) and SCA (simulated relative to MODIS), with the ensemble aggregated by albedo, liquid water and stability adjustment options. For snowpack runoff, RMSD was calculated for each year based on the cumulative runoff curves for the period between April and June. For each year, the curves were first normalised by (dividing by) their respective total runoff volumes between April and September, in order to focus on differences in timing rather than total volumes. For SCA, RMSD was calculated for the period between April and September using an NDSI threshold of zero for MODIS (see Section 3.3).

Figure S3 confirms that, for individual years as well as on average, ensemble groups exhibiting closer correspondence between snowpack runoff and observed total runoff also tend to show more consistency with MODIS SCA. This provides support for the suggestion that snowpack runoff dominates river flows in spring and early summer, with routing effects and other influences being relatively small. Figure S3 also confirms that using both prognostic albedo and the liquid water parameterisation generally leads to the best performance (Section 4.1). However, the group omitting liquid water processes but applying the stability correction also shows low mean RMSD overall, especially for SCA. As Section 4.2.3 strongly suggests the stability adjustment to be too strong in damping turbulent fluxes in stable conditions, it is possible to identify these compensatory effects as unphysical. Inter-annual variability in RMSD for all groups is high, as reflected by the wide and overlapping error bars in Figure S3. Although substantial asymmetries and trade-offs between runoff and SCA RMSDs are present, the range of RMSD tends to be smaller for groups performing better on average. However, in some years, configurations tending to perform worse on average may outperform more realistic configurations, as explored in Sections 4.4 and S5.

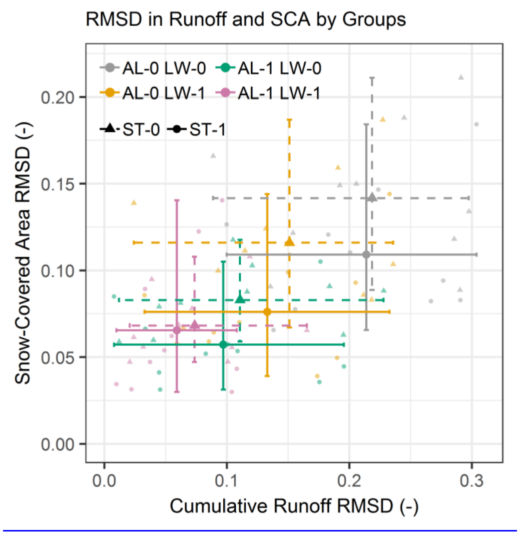


Figure S3. Cumulative runoff RMSD is plotted against SCA RMSD in for each of the albedo (AL) and liquid water (LW) option combinations (i.e. averaging respective ensemble members), with differentiation by stability (ST) option (shape and line type) also shown. Cumulative snowpack runoff RMSD is for the April to June period and normalised, whereas SCA RMSD is for April to September and based on an NDSI threshold of 0.

The relationships between deviations from observations for different groups of model configurations are shown in Figure S4. These relationships are portrayed as a scatterplot matrix comparing RMSD for ensemble groups using different

Formatted: Normal

configurations of albedo (A), liquid water / drainage (D) and stability adjustment (S) options. The upper right of the scatterplot matrix shows runoff RMSD (noting the difference between simulated snowpack runoff and observed total runoff explained in Section 3.3), whereas the lower left shows SCA RMSD. Each point represents the RMSD for a single year, and results from all three climate input strategies are plotted (Section 3.2.3 and Section S1).

~~**S3**~~

~~Figure S2 shows a scatterplot matrix comparing root mean square deviation (RMSD) for ensemble groups using different configurations of albedo (A), liquid water / drainage (D) and stability adjustment (S) options. This expands the number of groups shown in Section 4.3.1. The upper right of the scatterplot matrix shows runoff RMSD, whereas the lower left shows SCA RMSD. Each point represents the RMSD for a single year, and results from all three climate input strategies are plotted (Section 3.2.3 and Section S1). Figure S2 highlights the variation of performance relationships between ensemble groups. For both runoff and SCA, the most linear relationships tend to be associated with configurations using diagnostic albedo (A0). This suggests that this choice dominates performance when selected. In some other cases, possible Pareto-like fronts are accompanied by almost linear relationships away from the front (examples for runoff include A1D1S0 vs A1D0S0 and A1D0S1). This indicates that the relative importance of differences between ensemble groups may vary between years in some cases.~~

Figure S4~~2~~. Scatterplot matrix comparing root-mean-square deviation (RMSD) for ensemble groups using different configurations of albedo (A), drainage (D) and stability adjustment (S) options. Scatterplots for runoff RMSD are shown in the upper right of the matrix and scatterplots for SCA RMSD are shown in the lower left. RMSD is calculated following the description in Section 4.3.1. Each point represents the RMSD for a single year, and results from all three climate input strategies are plotted (Section 3.2.3 and Section S1).

Figure S4 highlights the variation of performance relationships between ensemble groups. For both runoff and SCA, the most linear relationships tend to be associated with configurations using diagnostic albedo (A0). This suggests that this choice dominates performance when selected. A relatively linear relationship (albeit with notable scatter) also occurs in the generally well-performing case where prognostic albedo and the liquid water parameterisation are both applied but the stability adjustment is varied (grey points on Figure S4). This suggests that, overall, model performance limitations are not due to varying the stability adjustment option in this case. Rather, any structural constraint on performance is common to both configurations, namely through the dominance of the albedo and snowpack hydrology parameterisations and their response to the climate conditions of a particular year.

In other cases, trade-offs resembling a Pareto front develop. Thus, for a number of the years simulated, performance in one group cannot increase without a corresponding reduction in performance of another group. For example, there are a number of years in which good performance with the liquid water option off (and the stability adjustment applied) is associated with performance reductions in more physically realistic configurations where the liquid water option is switched on. Sometimes possible Pareto-like fronts are accompanied by almost linear relationships away from the front (examples for runoff include A1D1S0 vs A1D0S0 and A1D0S1). This indicates that the relative importance of differences between ensemble groups may vary between years in some cases.

Formatted: Normal, Line spacing: single

~~S4~~S5 Runoff and LST Anomaly Time Series

Figure S5~~Figure S3~~ compares simulated catchment-scale snowpack runoff and LST monthly anomalies with observed runoff and MODIS remote sensing, respectively. This demonstrates that the simulated anomalies are reasonably consistent with the reference datasets in anomaly space. The spread amongst the four major ensemble groups considered in Section 4.~~3.2~~4 is also fairly small relative to the amplitude of inter-annual variability. Some discrepancies are of course evident, which may partly reflect the limitations of the HAR climate product in capturing the sequencing of inter-annual climate anomalies (Section 4.~~3.2~~4). Indeed, some of the discrepancies likely fit with the SCA errors discussed in Table 3. Overall then, Figure S5~~Figure S3~~ supports the finding in Section 4.~~3.2~~4 that the FSM ensemble may be somewhat reliable in anomaly space when climate input anomalies are sufficiently represented.

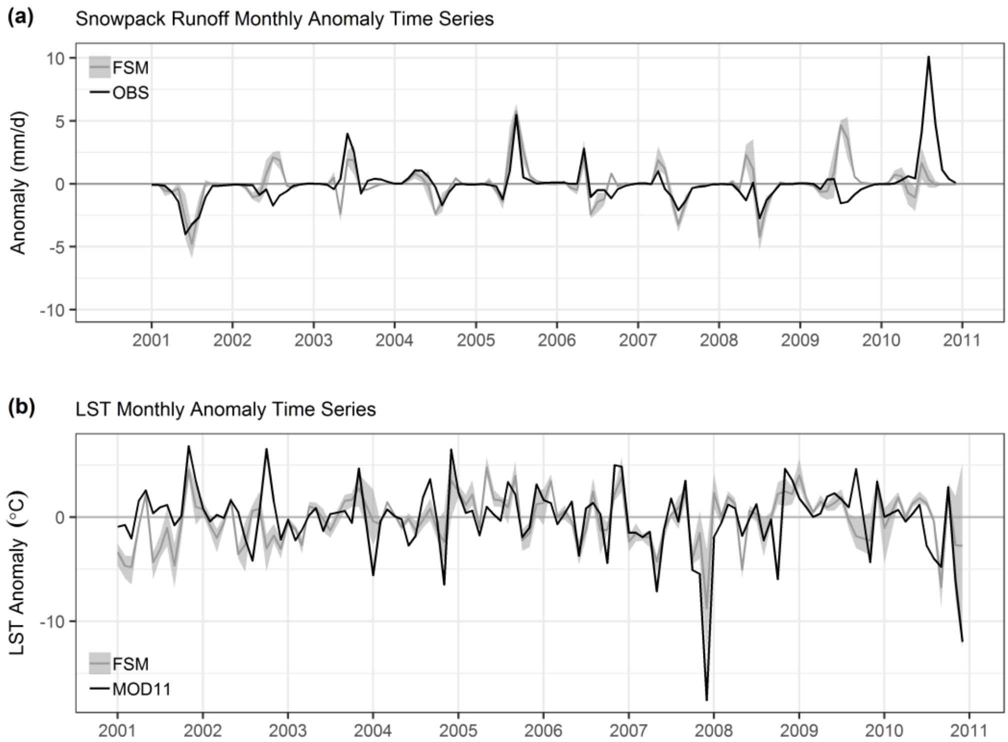


Figure S5~~3~~. Comparison of simulated (a) runoff and (b) LST catchment-scale monthly anomalies with observations and remote sensing, respectively. The mean and range of the four primary FSM ensemble groups considered in Section 4.~~3.2~~4 are shown by the grey lines and shading. The observations in (a) are total catchment runoff anomalies, while the simulated values are for snowpack runoff (Section 3.3).

~~S5~~S6 Climate Input Sensitivity

Figure S6~~Figures S4~~ and Figure S7~~S5~~ show the implications of using the two alternative climate input strategies described in Section 3.2.3 and Section S1. These strategies are essentially HAR-based inputs with no bias correction (Figure S6~~Figure S4~~), and using local observations to derive climate fields as far as possible (Figure S7~~Figure S5~~). Figure S6~~Figure S4~~ indicates that omitting bias correction of HAR temperatures leads to a very similar ensemble structure to that presented for the baseline climate inputs in Section 4.1. There is a small shift of the cumulative snowpack runoff curves to the right in Figure S6~~Figure S4~~ compared with Figure 2, which reflects the effect of the cold temperature bias in delaying runoff. However, the structure of ensemble groups matches closely in the two figures. Therefore temperature bias correction does not fundamentally alter FSM response in this case.

Figure S7~~Figure S5~~ also shows notable similarity with Figure 2 in Section 4.1 in terms of overall ensemble structure. However, there is more spread within the principal groups within the ensemble in Figure S7~~Figure S5~~, especially for slow-responding combinations (prognostic albedo and a representation of liquid water refreezing, retention and drainage). This leads to a wider overall ensemble spread when applying primarily observation-based inputs. Nevertheless, the rank order of the primary and secondary groups within the ensemble is the same as for the baseline climate inputs. This strengthens the notion that the findings in Section 4 are likely to be similar when using various commonly applied climate input strategies.

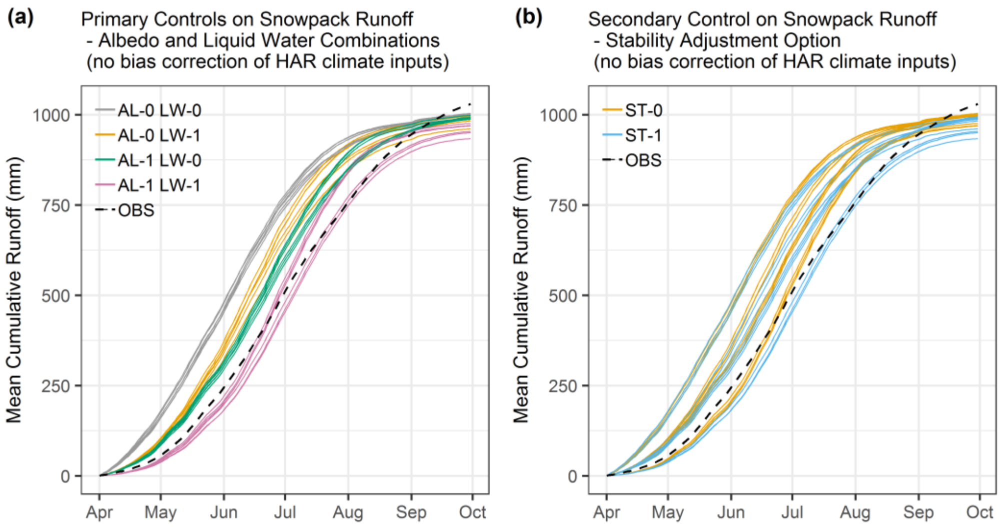


Figure S6~~4~~. Comparison of mean cumulative snowpack runoff for the high-flow season for each of the 32 ensemble members with observed total runoff (OBS, black dashed line). In (a) each ensemble member is coloured according to the combination of albedo (AL) and liquid water (LW) parameterisations it uses. In (b) each ensemble member is coloured by its stability adjustment (ST) option. The results are based on input strategy (2) described in Section S1 (i.e. HAR-based but without temperature bias correction).

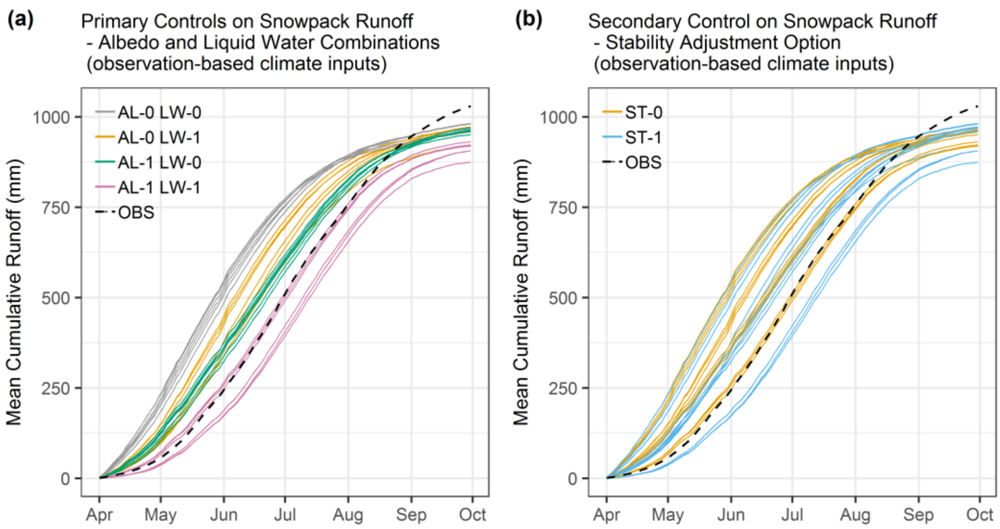


Figure S7~~5~~. As Figure S6 ~~Figure S4~~ but for input strategy (3) described in Section S1 (i.e. observation-based as far as possible).

Field Code Changed

Formatted: Font: Not Bold

Formatted: Normal, Line spacing: single

References

Gascoin, S., Guðmundsson, S., Aðalgeirsdóttir, G., Pálsson, F., Schmidt, L., Berthier, E. and Björnsson, H.: Evaluation of MODIS Albedo Product over Ice Caps in Iceland and Impact of Volcanic Eruptions on Their Albedo, Remote Sens., 9(399), 1–18, doi:10.3390/rs9050399, 2017.

Hewitt, K.: Glaciers of the Karakoram Himalaya, Springer Netherlands., 2014.

Liston, G. E. and Elder, K.: A Meteorological Distribution System for High-Resolution Terrestrial Modeling (MicroMet), J. Hydrometeorol., 7, 217–234, doi:10.1175/JHM486.1, 2006.

Liu, J., Schaaf, C., Strahler, A., Jiao, Z., Shuai, Y., Zhang, Q., Roman, M., Augustine, J. A. and Dutton, E. G.: Validation of Moderate Resolution Imaging Spectroradiometer (MODIS) albedo retrieval algorithm: Dependence of albedo on solar zenith angle, J. Geophys. Res., 114, D01106, doi:10.1029/2008JD009969, 2009.

Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J. and Finkelnburg, R.: Precipitation Seasonality and Variability over the Tibetan Plateau as Resolved by the High Asia Reanalysis, J. Clim., 27, 1910–1927, doi:10.1175/JCLI-D-13-00282.1, 2014.

Pellicciotti, F., Raschle, T., Huerlimann, T., Carenzo, M. and Burlando, P.: Transmission of solar radiation through clouds on melting glaciers: a comparison of parameterizations and their impact on melt modelling, J. Glaciol., 57(202), 367–381, doi:https://doi.org/10.3189/002214311796406013, 2011.

Pritchard, D. M. W., Forsythe, N., Fowler, H. J., O'Donnell, G. M. and Li, X.-F.: Evaluation of Upper Indus Near-Surface Climate Representation by WRF in the High Asia Refined Analysis, J. Hydrometeorol., 20(3), 467–487, doi:10.1175/JHM-D-18-0030.1, 2019.

Wang, Z., Schaaf, C. B., Chopping, M. J., Strahler, A. H., Wang, J., Román, M. O., Rocha, A. V, Woodcock, C. E. and Shuai, Y.: Evaluation of Moderate-resolution Imaging Spectroradiometer (MODIS) snow albedo product (MCD43A) over tundra, Remote Sens. Environ., 117, 264–280, doi:10.1016/j.rse.2011.10.002, 2012.

Wen, J., Liu, Q., Xiao, Q., Liu, Q., You, D., Hao, D., Wu, S. and Lin, X.: Characterizing Land Surface Anisotropic Reflectance over Rugged Terrain: A Review of Concepts and Recent Developments, Remote Sens., 10(370), 1–30, doi:10.3390/rs10030370, 2018.