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 would change the manuscript in a 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 a revised version we will remove 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 will put it into the Supplement. The details of these changes are given below in response to the "detailed comments", but the main proposed changes may be summarised as follows:

- Move the evaluation of albedo parameterisations using MODIS to the Supplement.
- Remove 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 propose to 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.
- Remove Section 4.3.1 quantifying simulated runoff differences relative to observations (and how they relate to SCA performance). Part of this material will be moved to the Supplement (along with some of the removed and adapted text from Section 4.1). However, the framing will be adjusted to emphasise what the differences between the modelled and observed series suggest about catchment functioning, in contrast to the emphasis on performance in the original manuscript.
- Adjust Section 3.3 (introducing model evaluation data/methods) to explain better the more restricted use of observed river flow data in a 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' will be removed from the revised abstract (i.e. the second clause of the sentence will be deleted).

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 will be added to this section. The relevant sentence for the remote sensing products will read: "*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 will read: "*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 will be moved to the Supplement (for details see response below). The method will now be explained in its own section in the Supplement 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 will also be summarised in the Supplement: "*In quantitative terms, the prognostic parameterisation outperforms the diagnostic option for the anomaly series, with an overall root-mean-square deviation (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 will remove the final paragraph in Section 4.1.1 as suggested by the reviewer. We will also remove the relevant parts of the final paragraph in Section 4.1.2, which has a similar comparison.

The caption for Figure 2 will also be adjusted to highlight that the observed total runoff shown on the figure is only to provide context on the catchment behaviour. The caption will now read: "*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 will be adjusted to explain more explicitly the restricted ways in which observed river flows will be used in a revised manuscript. The paragraph will now read: "*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), in general it is not possible to compare a modelled quantity with the observed river flows directly. Therefore 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). In these cases the modelled quantity considered is runoff from the base of the snowpack. This quantity is termed snowpack runoff, which is different to surface melt, as the latter 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 will be removed from the main text (the original Figure 7 and related text). Part of this (relating to original Figure 7a) will be moved into the Supplement, with a reframing of the text to emphasise the possibility that the co-occurrence of better correspondence in SCA (with MODIS) and snowpack runoff (with observed total runoff) tells us something about the nature/significance 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 about how some of the hydrological processes in this type of catchment work, rather than remove this analysis from the manuscript entirely.

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

This will be 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 curves for these catchments 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 will be removed from the main text and added to the Supplement (new section) and referred to very briefly at the end of the first paragraph in Section 4.2.1 with a new final sentence: "*Section SX 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 (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 will move it from being a sub-section of Section 4.2 into its own section (now 4.3). We will also expand the text slightly to add some more detail. These results will now also be referred to in the abstract and conclusion.

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 will be 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 will now be 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) will also be split up into sub-sections to show more clearly where different issues and uncertainties are treated (e.g. forcing uncertainty will be moved into its own sub-section – Section 5.2.1).

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 will explain now 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 will also add 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 different model configurations based on responses to historical climate variability offers a complementary approach to traditional model evaluation methods. Such an approach could be used in the ongoing evaluation of snow models in the ESM-SnowMIP inter-comparison (Krinner et al., 2018) and for understanding the plausibility of 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).*"

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 proposed changes may be summarised as follows:

- Remove 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 propose to 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.
- Remove Section 4.3.1 quantifying simulated runoff differences relative to observations (and how they relate to SCA performance). Part of this material will be moved to the Supplement (along with some of the removed and adapted text from Section 4.1). However, the framing will be adjusted to emphasise what the differences between the modelled and observed series suggest about catchment functioning, in contrast to the emphasis on performance in the original manuscript.
- Adjust 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 agree could be discussed in more detail in the 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 also 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 propose to address this comment primarily by dedicating a section of expanded discussion to it within a restructured Section 5 (discussion section). We will take the existing relevant part of Section 5 and explain the significance of these processes in more detail based on an expanded set of references. We will also discuss the main results of the manuscript in terms of how they could be affected by the inclusion of avalanching and blowing snow processes. These processes will certainly be important in some areas (especially at high elevations) and the effects of their omission present in some simulated quantities at some times of year, and the discussion will highlight how/where we can confidently derive results and where the uncertainty is larger. For example, the SCA annual cycle comparison may be particularly robust in the early part of the melt season, when SCA decline is dominated by low and middle elevation melt. This segment of the annual SCA cycle is still enough to support the results on the behaviour of different FSM configurations etc.

We will highlight in the discussion that a key area for further work is the incorporation of avalanching and blowing snow processes into distributed snow model simulations in the region. We will 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

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 will move it from being a sub-section of Section 4.2 into its own section (now 4.3). We will also expand the text slightly to add some more detail. These results will now also be referred to in the abstract and conclusion.

Specific comments:

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

We will reword this sentence and remove the second clause.

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

The term 'anomaly space' will be removed from the manuscript.

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 will be 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 emphasised primarily small 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 will be 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 emphasised primarily small 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.23 “recent data” what data are these? Remote sensing? In situ?

Sentence will be 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

Sentence will be 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

Sentence will be reworded to include 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 will be reworded.

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

Sentence will be 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 will be 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 will reword and expand to clarify: "*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 indicates that catchment runoff is primarily mass- rather than energy-limited. 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?

Sentence will be 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)

Sentence will be reworded and expanded to define: "*For each process, the second option (1) may be considered generally more in line with conceptual understanding of the 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 assume instantaneous drainage of liquid water instead (option 0).*"

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

Reference will be added.

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

Brief description of site locations will be added: "*The available stations are situated in valley locations, as is typical in data-sparse high mountain regions. 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. We will revisit the wording in this section to try to clarify.

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. So the gradients were calculated by regressing the simulated surface (or near-surface) values against the respective elevations of the model grid points.

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

Sentence will be 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 taken from the HAR.*"

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

Variable names will be 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 as the basis for estimation of 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.

Sentence will be 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 will change "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.*"

P7. L.2 Define satisfactory quality

Now will be defined in additional sentence: "*Satisfactory quality is defined here for MOD11A1 LST as a mandatory data quality flag of 00 (good).*"

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

We will add some text here to reflect the importance of these processes, but longer discussion will be saved until the discussion section (5). Text added here will be: "*However, these high elevation areas are also particularly subject to the influences of blowing snow processes and avalanching. These processes could substantially alter high elevation snow cover patterns (see discussion in Section 5.X).*"

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

We will remove 'this' and reword 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

Sentence will be 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 will be added for these numbers.

P10. L.30 What are mm/a ?

The units will be 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 will be included in the main body of the text here.

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 will be 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 will 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 will be removed.

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

We will take the phrase "anomaly space" out of the manuscript so it does not need to be defined in the methodology now. We will reword the sentence originally containing "anomaly space" to: "*Despite these patterns of divergence and variation in absolute errors, Figure 8b 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 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:https://doi.org/10.5194/hess-8-47-2004, 2004.

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.

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.