Response to review

Snow model comparison to simulate snow depth evolution and sublimation at point scale in the semi-arid Andes of Chile

Annelies Voordendag, Marion Réveillet,
Shelley MacDonell, Stef Lhermitte

June 10, 2021

Dear editor and reviewers,

First, we would like to thank the editor and reviewers for their careful evaluation of our work and the valuable suggestions, comments and questions. We believe that the manuscript has substantially benefited from the editor's and reviewers' feedback. Below we adress our detailed responses to all the comments.

In this response-to-review document we try to clarify and address each of the suggestions, comments and questions made during the review. Therefore we have copied the comments in blue boxes and have addressed them one by one. In the response we use italic fonts to quote text from the revised manuscript. Additional to the revised manuscript, we have uploaded a supplementary version of the manuscript with highlighted track changes that indicate where the manuscript has changed (red=removed; blue=added).

The main changes in the manuscript include a new section presenting an idealised setup to acquire a better precipitation forcing, in response to a main concern of the reviewers. For that purpose, for both models snow depth or SWE have been assimilated to present idealized cases.

Second, the forcing uncertainty have been calculated with a bias, instead of random errors (Sect. 3.2.4), as initially presented. This choice has been made in agreement with a constructive comment made by a reviewer mentioning that the random errors can counterbalance each other. To avoid the underestimation of the forcing uncertainty, we chose to estimate forcing uncertainty according to Raleigh et al. (2015).

Finally, to better consider the snow roughness value uncertainty, we included a sensitivity to different values in the ensemble simulations in agreement with reviewer comments.

We also want to note that we have added Stef Lhermitte as corresponding author.

Yours sincerely, Annelies Voordendag & co-authors

Response to the Editor B. Noel

Dear Annelies Voordendag and co-authors,

Thank you for your submission to TCD. As you may know, papers accepted for TCD appear immediately on the web for comment and review. Before publication in TCD, all papers undergo a rapid access review undertaken by the editor with the aim of providing initial quality control. It is not a full review and the key concerns are fit to the journal remit, basic quality issues and sufficient significance, originality and/or novelty to warrant publication. As a result, even a manuscript ranked highly during access review can receive a low ranking during full peer review later. Evaluation criteria are found at: www.the-cryosphere.net. Grades are from 1 (excellent) to 4 (poor).

ORIGINALITY / NOVELTY (1-4): 2
The paper focuses on improving the representation of snow processes at a point location in the Chilean Andes in winter 2017. To that end, the authors estimate the sensitivity of two snow models, namely SNOWPACK and SnowModel, to various parameterizations (i.e., fresh snow density and albedo) and atmospheric forcing perturbations (notably measured precipitation). Models show that sublimation is the main driver of ablation

during the snow season (May-November), and that its relative contribution to total ablation (i.e., sublimation ratio) is highly sensitive to the selected albedo parameterization. Atmospheric forcing perturbations also strongly impact model results mainly through precipitation uncertainties. By conducting multiple sensitivity experiments and model evaluation against local in situ measurements, the authors provide valuable insights on model performance, and response to different parameterizations and forcing perturbations.

SCIENTIFIC QUALITY / RIGOR (1-4): 1
The authors conduct a large set of sensitivity experiments, and evaluate model results using three measured variables, i.e., snow depth, snow water equivalent and surface albedo. This evaluation work yields valuable insights on model performance, model differences (i.e., in terms of physical complexity), and response to different parameterizations and forcing perturbations. The authors refer to relevant literature and provide sufficient information on the two selected models, evaluation data sets, calibration methods and design of the sensitivity experiments.

SIGNIFICANCE / IMPACT (1-4): 2
In its current form, the manuscript does not clearly stress the motivations and objectives of the study, notably in the abstract and conclusions. For instance, conclusions remain unclear about e.g. which model parameterizations/forcing perturbations have the strongest impact on modelled ablation components, and thus snow depth evolution in the Andes.

PRESENTATION QUALITY (1-4): 1
The manuscript figures and tables well support the study, results and discussion. The paper could benefit from stylistic improvements and some clarifications.

In brief, this is an interesting and generally well-written modelling study. The editor invites the authors to better stress the motivations and objectives in the abstract and conclusions. This can be addressed in the first round of revisions. For now, the editor recommends publication in TCD.

Kind regards, Dr. Brice Noël

The authors thank the editor for his very positive feedbacks and encouraging remarks. As described in the introduction of this document, we made some changes to the study and as suggusted by the editor, we better stress the motivations and objectives in the abstract and conclusions.

Response to the anonymous reviewer #1

Main comments

R1-1: In this manuscript the authors present a sensitivity analysis of two commonly used snowmodels, SNOWPACK and Snowmodel, for a semi-arid Andes catchment. The authors aim to quantify the impact of various parameter and parameterization selections and impact of forcing uncertainty on a sublimation dominated catchment. Overall, this manuscript is clear, well written, and provides useful results. Research topics like this have a tendency to end up very location-centric and not widely applicable to the larger community. However, I do not believe that is the case here. There are sufficient linkages with existing work.

The authors thanks the reviewer for its careful evaluation of our work and the valuable suggestions, comments and questions.

R1-2: I have two concerns: One, the description of precipitation measurements is unclear to me. Figure 1 suggests an unshielded Geonor gauge is used. However, the authors use an Alter-shielded correction factor. This should be clarified in the text. If an unshielded gauge was indeed used, then a) a different factor should be used and b) the uncertainty in the precipitation is massive and I am then not completely convinced. I also note that MacDonald (2007) is a grey-literature source (conference proceedings), and I am curious as to why the authors chose this correction versus some of the "more standard" WMO/Goodison corrections?

First, the description of precipitation measurements have been clarified. For instance, in the measurement description you can now read: *This gauge is an unshielted, unheated weighing bucket precipitation gauge filled with anti-freeze liquid and oil to prevent freezing and evaporation respectively.*
Second, in agreement with your comment and comments made by the other reviewers, we:

a) forced the models with snow depth or SWE to assimilate precipitation data that lead to more realistic results and we therefore add a specific and new section 'Idealised setup', to present these results.
b) we corrected measured precipitation with an other approach (i.e.Wolff et al. (2015)) which leads to an amount of precipitation closer to precipitation reconstructed from SWE.
To clarify the different precipitation used in this study, a figure presenting the diffent options has been added in the revised version of the supplementary material.

R1-3: Two, it seems to me the authors are using only the instrument measurement uncertainty. This should be noted in the text. However, I'm surprised the authors did not use the uncertainty ranges and distributions from Raleigh et al. (2015, Table 3) which include more 'real-world' uncertainty ranges. I believe the manuscript would benefit from using these distributions and ranges. I believe this would more cleanly link this work with existing studies and increase the contribution.

We agree with the reviewer that considering only the measurement uncertainty might underestimate the forcing uncertainty. We therefore apply a bias to the forcing data, according to the ranges described in Raleigh et al. (2015, Table 3) in the revised version of the manuscript. In addition, as the method chosen in the initial version (i.e. random errors) tended to counterbalance each other, we decided to only apply a bias in the revised manuscript. Furthermore, the TA and RH are assumed to already experience some random noise due to its interpolation.

Detailed notes

R1-4: L18 "complexities" should be elaborated on as to what the authors mean by this, as everyone has different definitions.
R1-5: L18 "physical approaches" do you mean "physically-based" or "physics-based" here? All approaches should be physical.

We clarified the text. You can now read:
Several models, with different complexities in the representation of different snow processes, from empirical to physically-based approaches, (..)

R1-6: L22 "These approaches, coupled with snow models" I don't understand what you mean by this.

We agree that this statement was unclear. We meant to say that our approach is the use of the energy balance equation. The text now reads:
The use of the energy balance equation, coupled with snow models, enables (..)

R1-7: L32 Should cite Essery (2015) as well

Done.
In addition to the development of new models, many studies have focused on model improvements offering different parameterizations in a single model (e.g. Douville et al., 1995; Dutra et al., 2010; Essery, 2015).

R1-8: L53 "this" refers to what?

"This" refers to the importance of snow as water resource. This in now clarified in the revised version:
Despite the importance of snow as water resource, quantifying (..)

R1-9: L54 I would also note dry air (important for sublimation)

We agree with your statement and add this information in the revised manuscript:
i) high sublimation rates related to high levels of incoming solar radiation, cold air temperatures, arid atmosphere, and high wind speeds

R1-10: L65 1000 seems arbitrary, perhaps note as much or describe why this number was chosen

This was indeed an arbitrary amount and a compromise between computation effort and reliable results. We have added the following:
We have chosen 1000 runs as a compromise between computational effort and a reliable confidence interval.

R1-11: L72 “to assess the sensitivity” of what?

We clarified this phrase:
To assess the sensitivity of the models to the representation of snow physics and meteorological forcing, we (..)

R1-12: L72 “a permanent station” clarify this is a meteorological station

We clarified this: *a permanent meteorological tower since 2009*

R1-13: L80 If this table can be included in the main text, I think you should do so

We agree with the reviewer and have added this table to the main text in the revised manuscript.

R1-14: L80 Describe shielding for Geonor here.

We extended the text with:
This gauge is an unshielted, unheated weighing bucket precipitation gauge filled with anti-freeze liquid and oil to prevent freezing and evaporation respectively.

R1-15: L96 “Physical equilibrium” what does this mean?

With physical equilibrium we mean the realistic climatology for the model (e.g. surface temperature) after the model is being started from other initial conditions, but in the revised manuscript this entire phrase will be removed.

R1-16: “June 11:00 and 31 Oct” replace ‘and’ with ‘to’

Done.
(23 June 11:00 to 31 October 10:00 due to sensor failure)

R1-17: L110 describe shielding

We solved this in comment R1-14.

R1-18: L118 Blowing snow sublimation will also lower on ground swe and should be noted as this further adds uncertainty to the reconstruction.

We agree with the reviewer that sublimation influences the amount of SWE and we therefore reconstructed the precipitation from SWE at moments that also precipitation was registered at the precipitation gauge to ensure that the change in SWE is caused by precipitation.

R1-19: L165 add “snow” roughness

Done.
Atmospheric stability and snow roughness length (z_0)

R1-20: L165 I’m surprised the roughness length is so small. This is well on the lower end of what is reported in the literature. It seems to me to be calibrating sensitivity to the turbulent fluxes, suggesting that they are being overestimated with more ‘reasonable’ z_0 values. Could this be related to the stability parameterization being insufficient in this area?

We agree with your comment. However as no measurements are available to better calibrated this value we chose to consider different values in the evaluation of the model sensitivity to the parameterization choice (i.e. $z_0 = 1\,\mathrm{mm}$ as in the intial version and $z_0 = 1\,\mathrm{cm}$). We also better discuss this point in the discussion section of the revised manuscript:
The turbulent fluxes parameterization is sensitive to the z_0 value and observations, such as Eddy Covariance measurements, are essential to accurately parameterize the turbulent fluxes (e.g. Conway and Cullen, 2013; Litt et al., 2017; Réveillet et al., 2018). Due to the absence of such measurements, the variability of this value over time (e.g. MacDonell et al., 2013b; Pellicciotti et al., 2005; Nicholson et al., 2016) and due to other values in literature at other locations showing a wide variety (two orders of magnitude) of the snow roughness length (Gromke et al., 2011;

Poggi, 1977; Bintanja and Broeke, 1995; Andreas et al., 2005), it was decided to use two different values for z_0 (1 mm and 1 cm). Similar sensitity ranges for SNOWPACK and SnowModel were found (e.g. Fig. 4), along with similar sublimation rates, but this directly depended on the value for z_0. For both models, a z_0 of 1 cm led to better simulations (Fig. 3,4), but as applied more often, this can also be seen as optimizing parameter (e.g. Stigter et al., 2018). In future work, the z_0 can be verified with eddy covariance measurements.

> R1-21: L230 all RMSE values need a unit, even if it's (-) for albedo.

Done and also solved this in the rest of the document.

> R1-22: L230 "compared to the albedo" what is compared?

We calculated the *RMSE* and R^2 between the modelled and the observed albedo.
(i.e. RMSE of 0.09 (-) and R^2 of 0.86 calculated with the observed and simulated albedo)

> R1-23: L241 rmse units

Done.

> R1-24: Figure 3: Legend should be added. Albedo yaxis needs more ticks, at least in the 0.5-1 range.

We added extra ticks to the y-axis and the legend has been added.

> R1-25: L321 remove "total"
> R1-26: L321 add "in other alpine"

Done.
This conclusion, found here in an arid environment, is in agreement with the studies performed in other alpine areas.

> R1-27:L351 fix reference ( )

Done.

Response to the reviewer #2: Michael Lehning

Main comments

> R2-1: The paper investigates two snow models with respect to their sensitivities on input and parameterizations in simulating snow in a dry high-elevation environment. This is in principle an useful exercise given the importance of snow and snow melt as a water resource in these ecosystems. I share the motivation that snow model evaluation is needed for the particularly dry environment in the Andes. The paper is well-written and the presentation of the material is clear.

We are very thankful for the constructive and positive comments of reviewer #2.

> R2-2: However, I have major concerns about the execution of the study. The main problem is that the main features of the mass balance are not reproduced by neither model despite the "calibration" attempts. Since there is a strong influence of total mass (SWE) and depth of the snow on processes such as sublimation and melt, which are in the focus of the study, the point of departure is insufficient. From the SWE and snow depth curves presented in the paper, I would hypothesize that you have significant snow accumulation and occasional erosion from snow transport at your site during the main winter. Since snow transport very heavily influences snow sublimation and total mass influences melt, the results obtained without reproducing at least approximately the local mass balance appear not trustworthy. A good mass balance could in principle be simulated with SNOWPACK by using the transport module. As a minimum, I would request that SNOWPACK is used to first generate a best estimate mass input (by using the snow depth forcing feature) and then start the sensitivity

analysis.

We totally agree with the underestimation of the precipitation input. In agreement with this statement and also comments made by the other reviewers, we added a new section *Idealised setup*, where we use both SNOWPACK and SnowModel to assimilate a more 'realistic' precipitation set. In addition, it raised the awareness that we needed another correction for the precipitation. As also mentioned in response to reviewer #1, we therefor chose an other method (based on the study by Wolff et al. (2015)) to correct the measured precipitation which better corresponds to the total SWE. In the revised version the simulations with parameterization ensembles and forcing ensembles are therefore performed with a more appropriate precipitation amount.

Finally, in the revised version, the snow transport in SNOWPACK was activated. However, as simulations are performed at point scale, the impact on snow transport is poor, likely related to the complexity of snow transport (erosion/depostion) at point scale. Further studies at larger scale would be very interesting, especially in such windy areas, but this is out of the scope of the present work. This is however discussed in the discussion of the revised manuscript.

R2-3: Another major point is that I cannot see in how your analysis supports your conclusion that parameterization would be more important than model choice or structure. The two striking differences between the two models are 1) the strong settling/melt of SnowModel already during the main winter and 2) the rapid melt-out at the end of the season. These two characteristics are qualitatively not changed by any of the parameterization changes. I would in fact assume that they have to do with model physics (settling law / water transport / refreezing) and model structure (number of layers).

According to your comment we decided to compare and discuss independently the model parameterization uncertainties and the model structure/physics differences. For that purpose, first the model parameterization choice is quantified for both models (Fig. 4). Then in the discussion we show why the model choice is of last importance, as the sensitivity to parameterizations and to forcing is of similar magnitude.

R2-4: Let me further suggest that the choice of model variants (such as picking the sub-model for albedo) is not called calibration. A calibration is a procedure, by which you determine the value of a free parameter. What you do is not the same as calibration.

Thank you for this clarification. For the specific case of the albedo, it is now mentioned that different sub-models for the albedo are tested. In addition, parameterization is generally preferred throughout the manuscript instead of 'calibration'.

Specific Comments

R2-5: l. 11: "varies EVERY eight days" is not clear

This section had been removed from the abstract, as the scope of the study slightly changed.

R2-6: l. 48: not sure I agree with "ESPECIALLY in regions where sublimation ...." You should give a justification here.

By 'especially' we meant that not a lot of snow models have been applied to semi-arid regions. We modified the sentence in agreement with your statement:
... in particular in regions where sublimation is the main ablation process, due to the lack of snow modelling studies in semi-arid regions (Gascoin et al., 2013; Réveillet et al., 2020; MacDonell et al., 2013a; Mengual Henríquez, 2017).

R2-7: l. 77: In the picture, I can see some lichen vegetation

We adjusted the text:
At this elevation, vegetation is extremely sparse. .

R2-8: l. 85 ff: A good recent paper discussing errors in this type of SWE measurement is Gugerli et al. (2019).

Gugerli et al. (2019) used a cosmic ray sensor mounted on the glacier surface. While this paper provides valuable information about the uncertainty of this sensor in a glaciarized area, we are not sure about the relevance in comparing this uncertainty to our study, as in our case a different sensor was used, which was mounted above the snow surface.

R2-9: l. 95/96: This only makes sense if you had included soil layers below the snow.

The text now reads:
The period between 5 May and 30 Nov 2017 has been covered to model the snow evolution in the austral winter.

R2-10: l. 101: Why using a moist adiabatic lapse rate is such a dry environment? Give a justification!

We agree that using a moist adiabatic lapse rate migh not be appropriate in dry environment. However, in our study the lapse rase wasn't a moist adiabatic one. It has been calculated with the available data. This point point have been clarified in the revised manuscript as follows:
For TA, a daily temperature lapse rate (Blandford et al., 2008) was calculated using TA measured at La Laguna and Paso Agua Negra AWSs (1565 m elevation difference) between 2014 and 2017.

R2-11: l. 114: Very small roughness length!

In agreement with your comment a roughness length of 1 cm has been chosen. We added this to the Sect. 4.1:
Eq. (12) from Wolff et al. (2015) with WS corrected to gauge height using a logarithmic wind profile (e.g. Lehning et al., 2002) and a z_0 of 0.01 m is used as precipitation data in the further study, (..)
However, we made a sensitivity test on this value and does not lead to significant difference in the precipitation correction.

R2-12: l. 122: Such spatial variability has been investigated by [Grunewald and Lehning, 2015]

We moved this phrase to the section *Idealised setup* and added the reference.
Preliminary results showed simulated SWE and SD to be more than two times lower than the observed SWE. This is caused by an underestimation of the precipitation measurements, as the AWS is placed in a concave area that collects more snow than the Geonor precipitation gauge. This is in correspondence with research by Grünewald and Lehning (2014) on the spatial variability of SD measurements.

R2-13: l. 194 ff: If I understand the text and Eq. (2) correctly, this must produce a value at every time step including negative values. Can you please clarify? From your figure S2.1 this appears not to be the case but then the presentation may be wrong. Please check / clarify.

According to comments made by reviewer #1, the method to calculate the precipitation uncertainty has been modified. We now use a uniformly distributed bias for the precipitation ranging between positive values. In case of the other forcing variables, a normally distibuted bias has been applied (and reported in Table 2) and if negative values (e.g. for RH, WS, $S_{\downarrow}$) occurred here, the value was set to 0 (or 0.1 m s^{-1} for WS). This is now clarified in the revised version:
To assess the model sensitivity to meteorological measurement uncertainties, a bias has been applied to the meteorological forcing presented in Sect. 2.2 to generate an ensemble of 1000 forcing files. Raleigh et al. (2015) have shown that the model outputs are more sensitive to forcing biases than random errors. Therefore, all input variables except P were modified by adding hourly biases with a normal distribution $N(\mu = 0, \sigma^2)$ with σ the uncertainty range taken from Raleigh et al. (2015) and reported in Table 2. The biases has been kept in the range (Table 2) by assuming that the 99.7% of the bias, thus 3σ, is within this range. This positive component of the range is divided by three and multiplied with a normal distributed random number and added to the observed forcing. We have chosen 1000 runs as a compromise between computational effort and a reliable confidence interval.

R2-14: l. 223 ff: Note that there is a strong cross-sensitivity with settling. Also all of these results will look differently when snow transport is properly taken into account. For the diverse parameterizations, I would emphasize that they are listed and named in S4.

We agree with your comment. However, the aim of the study is to compare the sensitivity of both models to diverse parameterizations. As snow transport cannot be activated in SnowModel at point scale, we decided to make the comparison with SNOWPACK without the snow transport. We aware that this is a limitation, and this is therefore discussed in the revised version.

Otherwise, in agreement with your main comment, the snow transport has been activated for the idealized case. As mention above, as simulations are performed at point scale, the impact on snow transport is poor, likely related to the complexity of snow transport (erosion/depostion) at point scale. Note also that in the revised version, we referred again to the equations and references of the parameterizations that can be found in the supplementary material.

R2-15: l. 235: See, this is where the mass balance is important: if you had a correct mass balance then you could see which runs do reproduce the melt-out date, which is an important quantity to model.

We agree with your statement. Therefore in agreement with your main comments and comments made by the other reviewers, we run an 'Idealised setup', and better corrected the precipitation (please refer to answer to comment R1-2 for more details), to obtain a more realistic mass balance. The figure has been modified accordingly in the manuscript. In addition, the comparison between the observed and simulated SD and SWE has been adjusted and better described in the revised version. This allows a better model comparison.

R2-16: l. 247: Justify the statement!

We clarified this:
Where SnowModel relies on two albedo models based on TA and albedo ranges, SNOWPACK relies on empirical relations calibrated with measurements in Switserland and not adapted to the arid Tapado climate.

R2-17: l. 303: latent heat is also a turbulent flux in the ABL!

We totaly agree with your comment. However according to the comments made by the 3 reviewers, we decided to restructure the manuscript and the energy balance section is no longer in the manuscript.

R2-18: l. 331: Really Richardson number is not used much any more. Almost everybody uses MO similarity with corresponding stability corrections.

We agree with your statement. However, as the aim of the study was to compare the two models and as SnowModel only offers the correction based on the Richardson number, this correction has been prefered. This is clarified in the revised version:
The representation of turbulent fluxes in snow models is commonly based on the bulk method, and the atmosphere stability corrections are made based on different possible appraoches such as the the Monin–Obukhov similarity theory or the Richardson number (e.g. Liston and Hall, 1995; Vionnet et al., 2012). In this study, the method based on the Richardson number, was preferred, as both models offered this option, while it might be not the most common one used.

R2-19: l. 373: Vögeli et al. (2016) have demonstrated how one can get to good spatial mass input.

Thanks for mentioning this study added. It is a very good example of a correction of spatial mass input. Nevertheless, it would not have been possible in our study area as they used Airborne Digital Sensors (ADS) to scale precipitation input data. We now also use the approach described in the comment to R1-2.

R2-20: l. 350 ff: A very complete and systematic study on input uncertainties (but using a distributed snow model) is Schlögl et al. (2016).

We added this reference to the text:
Likewise, results presented here show that the main sensitivity remains in the forcing uncertainty, in agreement with previous studies (Magnusson et al., 2015; Günther et al., 2019; Raleigh et al., 2015; Schlögl et al., 2016).

Response to the anonymous reviewer #3

Main comments

R3-1: The manuscript "Snow model comparison to simulate snow depth evolution and sublimation at point scale in the semi-arid Andes of Chile" by Annelies Voordendag et al. discusses the application of two snow models at a site with an automatic weather station in a dry mountainous area in the Chilean Andes. The snow models provide mass balance components, which is of importance for investigating water resources in such areas. The model simulations suggest a very strong sublimation flux to the atmosphere, depleting significant amount of snow mass on the ground. By perturbing model settings and forcing data, uncertainties are quantified. Generally, the paper is well written and the topic is highly relevant, and the approach by including two snow models and doing the sensitivity study by perturbing model settings and forcing data is very solid. This makes for excellent ingredients for the manuscript.

We would like to thank the reviewer for his/her careful evaluation of our work and the valuable feedback and the very positive comments.

R3-2: However, it is surprising to see how poorly the models are able to reproduce the snow cover at the site. This deserves some more attention, because in my opinion, the agreement is so poor, that the trust I have in the simulated mass balance components is also severely limited. Intuitively when looking at the results and the discussion by the authors, a problem could be that the place is so wind affected, that the models underestimate density. That could explain the much stronger settling in the simulations than observed. As the authors discuss themselves, snow density is an important factor for sublimation, since the snow surface temperature depends on it. So I wonder if it is maybe a better approach to run the models with a fixed fresh snow density of let's say 350 kg m^{-3}, to see if the agreement improves? Or also to analyze the combined SWE increase and snow depth increase to get an estimate of the fresh snow density? This value could be compared to the parameterizations from SNOWPACK and SnowModel (add observed density from SWE/SD to Fig. S6.1 for example). It is definitely an aspect that is not sufficiently discussed in the current manuscript where the discrepancies between models and observations come from. It's an almost more interesting aspect of the study that the models are apparently very poorly able to capture the processes at this site.

We agree that the models very poorly simulated the SD and SWE in the initial version. This was principally related to the strong underestimation of the mass as both simulated SWE and SD where understimated in the initial version. Therefore, in agreement with your comments (R3-2 and R3-3), but also with the comments made by the two other reviewers, we adjusted the mass with a better precipitation correction. Despite this adjustment, the SD is still underestimated at the end of the season likely related to a too high snow density as this underestimation is not observed for the SWE (see Fig. 4 in response to R2-2). As no density measurements were available, the fresh snow density parameterization was simulated using the 5 different parameterizations available in SNOWPACK and 3 options for SnowModel. We have added the density calculated with SWE and SD to Fig. 4 and this shows two distinct regimes in the snow density calculation. From May to mid-June, high compaction rates were found, whereas the compaction afterwards was more moderate. SNOWPACK is able to model both moderate and high compaction, depending on the parameterization chosen, but the mode of compaction remains the same over the season. SnowModel simulates high compaction rates for all parameterizations, which is correct for the start of the season but an overestimation after mid-June. However we fully agree that this is a limitation, especially because of the impact on the snow surface temperature and therefore to the turbulent fluxes and sublimation rates. This point is discussed in the discussion of the revised manuscript.

R3-3: I'm also somewhat confused that the total SWE in the models is so much underestimated. The models roughly produce 175 mm w.e. in sublimation/evaporation and 100 mm w.e. runoff. So the total mass input to the models (around 275 mm w.e.) is less than the maximum SWE observed at the site (300-350 mm w.e.). The maximum SWE is about half in the simulations than what is observed. In L129-123, authors discuss that a precipitation reconstruction based on the SWE time series overpredicts total mass, but I don't find any compelling reason presented why an underprediction from the precipitation gauge is preferred over an overestimation from the SWE reconstruction. Furthermore, I think generally undercatch corrections vary so much, and are often found to be site-specific and setup-specific, that I think authors should take some freedom to improve the undercatch correction for this specific setup.

We agree that this underestimation was a limitation in the first version of the manuscript. In agreement with your comments and the comments made by the other reviewers we first add a specific section of an idealised case. For that purpose the simulation used assimilated SD or SWE to reconstruct the precipitation. However, as such measurements are not always avalable and as by using it as input, it cannot be used as validation, we run other simulations based on corrected precipitation in the further study.

Following your comment, we tried different options to correct precitation and also made a comparison between simulations if a precipitation set reconstruted from SWE (PSWE) was used as input. The precipitation correction chosen is a total of 423 mm w.e. compared to 477 mm w.e. at the end of season for the PSWE, which is feasible, given that PSWE might be overestimated due to some snow drift at particular dates.

Specific comments

R3-4: L255-258: this is not a correct description of how the default version of SNOWPACK works. Unless the authors modified the source code specifically regarding this, SNOWPACK will only use air temperature to distinguish between rainfall and snowfall when driven by a precipitation time series. This wouldn't alter the SWE response, unless runoff occurs. The other criteria are only used when SNOWPACK is driven by a snow height time series.

Thank you for your constructive comment. In response to comments made by the other reviewers the method to calculate the forcing uncertainty has changed and this senstence is no longer in the revised manuscript.

R3-5: The introduction may need citation of the recent SnowMIP study, with some recent publications: Krinner et al. (2018); Menard et al. (2021)

We added these interesting studies in the revised version of the manuscript:
Furthermore, the Earth System Model - Snow Model Intercomparison Project (ESM-SnowMIP) compared several snow models to improve the models in the context of local- and global scale modelling (Krinner et al., 2018) and indicated scientific and human errors in snow model intercomparisons (Menard et al., 2021), but the study sites did not include semi-arid regions.

R3-6: Section 2.2: please specify if the rain gauge was heated or not.

The rain gauge was not heated. This as been clarified in the revised manuscript:
This gauge is an unshielted, unheated weighing bucket precipitation gauge filled with anti-freeze liquid and oil to prevent freezing and evaporation respectively.

R3-7: L97: I assume that the full period between 23 June 11:00 and 31 October 10:00 is missing for TA and RH, and not only those two specific times.

Indeed, the entire period was missing. This has been clarified as follows: *(23 June 11:00 to 31 October 10:00 due to sensor failure)*

R3-8: Why was only one year studied while the measurement site has operated for a much longer period (installed in 2009 apparently)? Please also discuss to what extend this year 2017 is representative for the climate at the site (i.e., compare with the full data set in terms of temperature, precipitation and wind speed).

2017 was chosen as a SWE sensor was available for this year. This is now mentioned in the revised manuscript:
The period between 5 May and 30 November 2017 has been covered to model the snow evolution in the austral winter, as this was a season where SWE data was available to validate the models.
In addition, 2017 wasn't affected by an ENSO event, and can therefore be considered as neutral (despite quite high temperatures).

References

Andreas, E. L., Jordan, R. E., and Makshtas, A. P.: Parameterizing turbulent exchange over sea ice: the ice station weddell results, Boundary-Layer Meteorology, 114, 439–460, doi: 10.1007/s10546-004-1414-7, 2005.

Bintanja, R. and Broeke, M. R. V. D.: Momentum and scalar transfer coefficients over aerodynamically smooth antarctic surfaces, Boundary-Layer Meteorology, 74, 89–111, doi: 10.1007/bf00715712, 1995.

Blandford, T. R., Humes, K. S., Harshburger, B. J., Moore, B. C., Walden, V. P., and Ye, H.: Seasonal and

Synoptic Variations in Near-Surface Air Temperature Lapse Rates in a Mountainous Basin, Journal of Applied Meteorology and Climatology, 47, 249–261, doi: 10.1175/2007jamc1565.1, 2008.

Conway, J. P. and Cullen, N. J.: Constraining turbulent heat flux parameterization over a temperate maritime glacier in New-Zealand, Annals of Glaciology, 54, 41–51, doi: 10.3189/2013aog63a604, 2013.

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, Climate Dynamics, 12, 21–35, doi: 10.1007/s003820050092, 1995.

Dutra, E., Balsamo, G., Viterbo, P., Miranda, P. M. A., Beljaars, A., Schär, C., and Elder, K.: An Improved Snow Scheme for the ECMWF Land Surface Model: Description and Offline Validation, Journal of Hydrometeorology, 11, 899–916, doi: 10.1175/2010jhm1249.1, 2010.

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

Gascoin, S., Lhermitte, S., Kinnard, C., Bortels, K., and Liston, G. E.: Wind effects on snow cover in Pascua-Lama, Dry Andes of Chile, Advances in Water Resources, 55, 25–39, doi: 10.1016/j.advwatres.2012.11.013, 2013.

Gromke, C., Manes, C., Walter, B., Lehning, M., and Guala, M.: Aerodynamic Roughness Length of Fresh Snow, Boundary-Layer Meteorology, 141, 21–34, doi: 10.1007/s10546-011-9623-3, 2011.

Grünewald, T. and Lehning, M.: Are flat-field snow depth measurements representative? A comparison of selected index sites with areal snow depth measurements at the small catchment scale, Hydrological Processes, 29, 1717–1728, doi: 10.1002/hyp.10295, 2014.

Gugerli, R., Salzmann, N., Huss, M., and Desilets, D.: Continuous and autonomous snow water equivalent measurements by a cosmic ray sensor on an alpine glacier, The Cryosphere, 13, 3413–3434, doi: 10.5194/tc-13-3413-2019, 2019.

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 Resources Research, 55, 2779–2800, doi: 10.1029/2018wr023403, 2019.

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, Geoscientific Model Development, 11, 5027–5049, doi: 10.5194/gmd-11-5027-2018, 2018.

Lehning, M., Bartelt, P., Brown, B., and Fierz, C.: A physical SNOWPACK model for the Swiss avalanche warning: Part III: Meteorological forcing, thin layer formation and evaluation, Cold Regions Science and Technology, 35, 169–184, doi: 10.1016/s0165-232x(02)00072-1, 2002.

Liston, G. E. and Hall, D. K.: An energy-balance model of lake-ice evolution, Journal of Glaciology, 41, 373–382, doi: 10.3189/s0022143000016245, 1995.

Litt, M., Sicart, J.-E., Six, D., Wagnon, P., and Helgason, W. D.: Surface-layer turbulence, energy balance and links to atmospheric circulations over a mountain glacier in the French Alps, The Cryosphere, 11, 971–987, doi: 10.5194/tc-11-971-2017, 2017.

MacDonell, S., Kinnard, C., Mölg, T., Nicholson, L., and Abermann, J.: Meteorological drivers of ablation processes on a cold glacier in the semi-arid Andes of Chile, The Cryosphere, 7, 1513–1526, doi: 10.5194/tc-7-1513-2013, 2013a.

MacDonell, S., Nicholson, L., and Kinnard, C.: Parameterisation of incoming longwave radiation over glacier surfaces in the semi-arid Andes of Chile, Theoretical and Applied Climatology, 111, 513–528, doi: 10.1007/s00704-012-0675-1, 2013b.

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A., and Jonas, T.: Evaluating snow models with varying process representations for hydrological applications, Water Resources Research, 51, 2707–2723, doi: 10.1002/2014wr016498, 2015.

Menard, C. B., Essery, R., Krinner, G., Arduini, G., Bartlett, P., Boone, A., Brutel-Vuilmet, C., Burke, E., Cuntz, M., Dai, Y., Decharme, B., Dutra, E., Fang, X., Fierz, C., Gusev, Y., Hagemann, S., Haverd, V., Kim, H., Lafaysse, M., Marke, T., Nasonova, O., Nitta, T., Niwano, M., Pomeroy, J., Schädler, G., Semenov, V. A., Smirnova, T., Strasser, U., Swenson, S., Turkov, D., Wever, N., and Yuan, H.: Scientific and Human Errors in a Snow Model Intercomparison, Bulletin of the American Meteorological Society, 102, E61–E79, doi: 10.1175/bams-d-19-0329.1, 2021.

Mengual Henríquez, S. A.: Caracterización de la nieve de distintas localidades de Chile mediante el uso del modelo SNOWPACK, Master's thesis, Universidad de Chile, 2017.

Nicholson, L. I., Pętlicki, M., Partan, B., and MacDonell, S.: 3-D surface properties of glacier penitentes over an ablation season, measured using a Microsoft Xbox Kinect, The Cryosphere, 10, 1897–1913, doi: 10.5194/tc-10-1897-2016, 2016.

Pellicciotti, F., Brock, B., Strasser, U., Burlando, P., Funk, M., and Corripio, J.: An enhanced temperature-index glacier melt model including the shortwave radiation balance: development and testing for Haut Glacier d'Arolla, Switzerland, Journal of Glaciology, 51, 573–587, doi: 10.3189/172756505781829124, 2005.

Poggi, A.: Heat Balance in the Ablation Area of the Ampere Glacier (Kerguelen Islands), Journal of Applied Meteorology (1962-1982), 16, 48–55, doi: 10.1175/1520-0450(1977)016<0048:hbitaa>2.0.co;2, URL http://www.jstor.org/stable/26177592, 1977.

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, Hydrology and Earth System Sciences, 19, 3153–3179, doi: 10.5194/hess-19-3153-2015, 2015.

Réveillet, M., Six, D., Vincent, C., Rabatel, A., Dumont, M., Lafaysse, M., Morin, S., Vionnet, V., and Litt, M.: Relative performance of empirical and physical models in assessing the seasonal and annual glacier surface mass balance of Saint-Sorlin Glacier (French Alps), The Cryosphere, 12, 1367–1386, doi: 10.5194/tc-12-1367-2018, 2018.

Réveillet, M., MacDonell, S., Gascoin, S., Kinnard, C., Lhermitte, S., and Schaffer, N.: Impact of forcing on sublimation simulations for a high mountain catchment in the semiarid Andes, The Cryosphere, 14, 147–163, doi: 10.5194/tc-14-147-2020, 2020.

Schlögl, S., Marty, C., Bavay, M., and Lehning, M.: Sensitivity of Alpine3D modeled snow cover to modifications in DEM resolution, station coverage and meteorological input quantities, Environmental Modelling & Software, 83, 387–396, doi: 10.1016/j.envsoft.2016.02.017, 2016.

Stigter, E. E., Litt, M., Steiner, J. F., Bonekamp, P. N. J., Shea, J. M., Bierkens, M. F. P., and Immerzeel, W. W.: The Importance of Snow Sublimation on a Himalayan Glacier, Frontiers in Earth Science, 6, doi: 10.3389/feart.2018.00108, 2018.

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

Vögeli, C., Lehning, M., Wever, N., and Bavay, M.: Scaling Precipitation Input to Spatially Distributed Hydrological Models by Measured Snow Distribution, Frontiers in Earth Science, 4, doi: 10.3389/feart.2016.00108, 2016.

Wolff, M. A., Isaksen, K., Petersen-Øverleir, A., Ødemark, K., Reitan, T., and Brækkan, R.: Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study, Hydrology and Earth System Sciences, 19, 951–967, doi: 10.5194/hess-19-951-2015, 2015.