Review “Snow model comparison to simulate snow depth evolution and sublimation at point scale in the semi-arid Andes of Chile” by Annelies Voordendaag et al.

By Michael Lehning

General:
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 a 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. 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.

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).

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.

Specific Comments:
l. 11: “varies EVERY eight days” is not clear
l. 48: not sure I agree with “ESPECIALLY in regions where sublimation ….” You should give a justification here
l. 77: In the picture, I can see some lichen vegetation
l. 85 ff: A good recent paper discussing errors in this type of SWE measurement is [*Gugerli et al.*, 2019].
l. 95/96: This only makes sense if you had included soil layers below the snow.
l. 101: Why using a moist adiabatic lapse rate is such a dry environment? Give a justification!

l. 114: Very small roughness length!
l. 122: Such spatial variability has been investigated by [*Grunewald and Lehning*, 2015]
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.
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.
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.
l. 247: Justify the statement!
l. 303: latent heat is also a turbulent flux in the ABL!
l. 331: Really Richardson number is not used much any more. Almost everybody uses MO similarity with corresponding stability corrections
l. 373: [*Vogeli et al.*, 2016] have demonstrated how one can get to good spatial mass input
l. 350 ff: A very complete and systematic study on input uncertainties (but using a distributed snow model) is [*Schlogl et al.*, 2016]

References:

Grunewald, T., and M. Lehning (2015), 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*(7), 1717-1728, doi:10.1002/hyp.10295.
Gugerli, R., N. Salzmann, M. Huss, and D. Desilets (2019), Continuous and autonomous snow water equivalent measurements by a cosmic ray sensor on an alpine glacier, *Cryosphere*, *13*(12), 3413-3434, doi:10.5194/tc-13-3413-2019.
Schlogl, S., C. Marty, M. Bavay, and M. Lehning (2016), Sensitivity of Alpine3D modeled snow cover to modifications in DEM resolution, station coverage and meteorological input quantities, *Environ Modell Softw*, *83*, 387-396, doi:10.1016/j.envsoft.2016.02.017.
Vogeli, C., M. Lehning, N. Wever, and M. Bavay (2016), Scaling Precipitation Input to Spatially Distributed Hydrological Models by Measured Snow Distribution, *Frontiers in Earth Science*, *4*, doi:UNSP 108 10.3389/feart.2016.00108.