Reply to Referee #2, Richard L.H. Essery

We thank Prof. Richard L.H. Essery for his thoughtful review and comments.
The reviewer's comments are reported below in bold font while our replies are in regular text:

Although limited to a single site and short on mechanistic explanations, the evaluations of several models in simulating snow mass, depth and density with several forcing datasets in this paper are of value (another with similar aims that should be cited is https://journals.ametsoc.org/doi/full/10.1175/JHM-D-15-0013.1).
We thank the reviewer for this suggestion, the citation has been included in the revised paper.

Measurements of outgoing shortwave and longwave radiation are mentioned but not used in the model evaluations; these might provide more insight.

We agree that the evaluation of modelled outgoing shortwave and longwave radiation could provide interesting additional insights, as suggested by the reviewer. These variables are not provided by all the models considered in this study: they are missing completely in the simplest model considered, S3M, while the SMASH model provides the outgoing longwave radiation only. Therefore, we evaluated the simulated outgoing shortwave and longwave radiation at the surface for all remaining models (SNOWPACK, GEOTOP, HTESSEL and UTOPIA).

Figure 7 shows the difference (left) and the scatterplot (right) of the simulated and observed daily-averaged outgoing shortwave radiation for the SNOWPACK, GEOTOP, HTESSEL and UTOPIA models in the CTL run.

All the models tend to moderately underestimate the outgoing shortwave radiation, with SNOWPACK and HTESSEL showing the best agreement with observations both in terms of bias and of coefficient of determination R^2, compared to GEOTOP and UTOPIA. Differences between the simulated and observed outgoing shortwave radiation are mainly dependent on the representation of the albedo. These results suggest to check, in the UTOPIA model, the albedo, which is function of surface temperature and snow age as well as in the HTESSEL model, which nevertheless provides better outgoing shortwave radiation estimates.

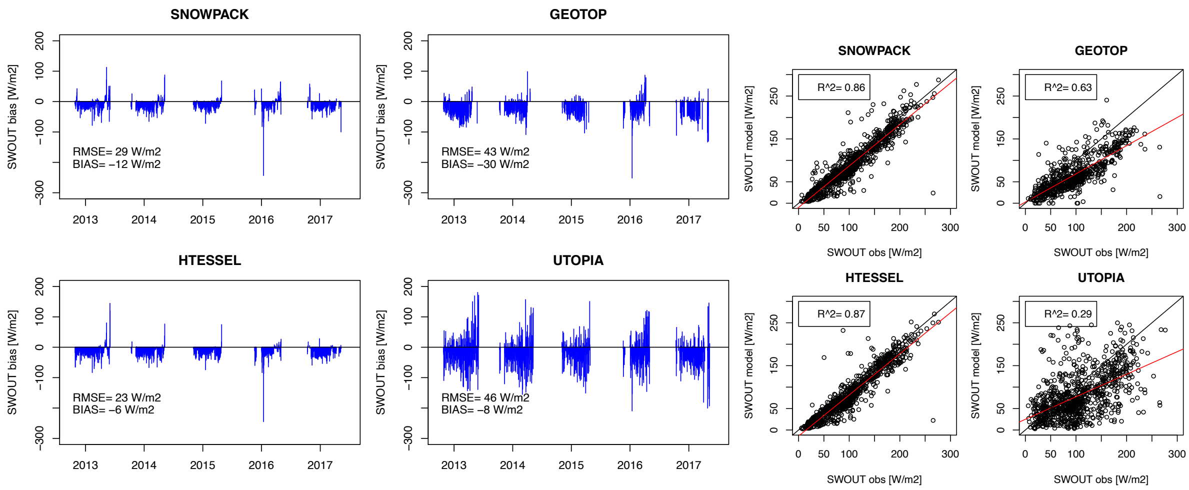

Figure 1 Difference (left figure) and scatterplot (right figure) of the simulated and observed outgoing shortwave radiation (CTL experiment) at the Torgnon site. Only four out of the six models considered in the paper provide the outgoing shortwave radiation among the output variables.

Similarly to Figure 7, Figure 8 shows the difference (left) and the scatterplot (right) of the simulated and observed daily-averaged outgoing longwave radiation for the SNOWPACK, GEOTOP, HTESSEL, UTOPIA and SMASH models in the CTL run. The simulation of the net longwave radiation mainly affects the representation of the snow-melt dynamics. Both SNOWPACK and SMASH underestimate the outgoing longwave radiation, causing an excess in the snowpack energy available for the melting. However, none of the two models remarkably underestimates the snow depth, so other mechanisms might compensate for this behavior. GEOTOP, HTESSEL and UTOPIA outgoing longwave radiation do not show systematic biases.

The considerations about the evaluation of the simulated outgoing shortwave and longwave radiation have been reported in the main text of the manuscript. We thank the reviewer for the suggestion.

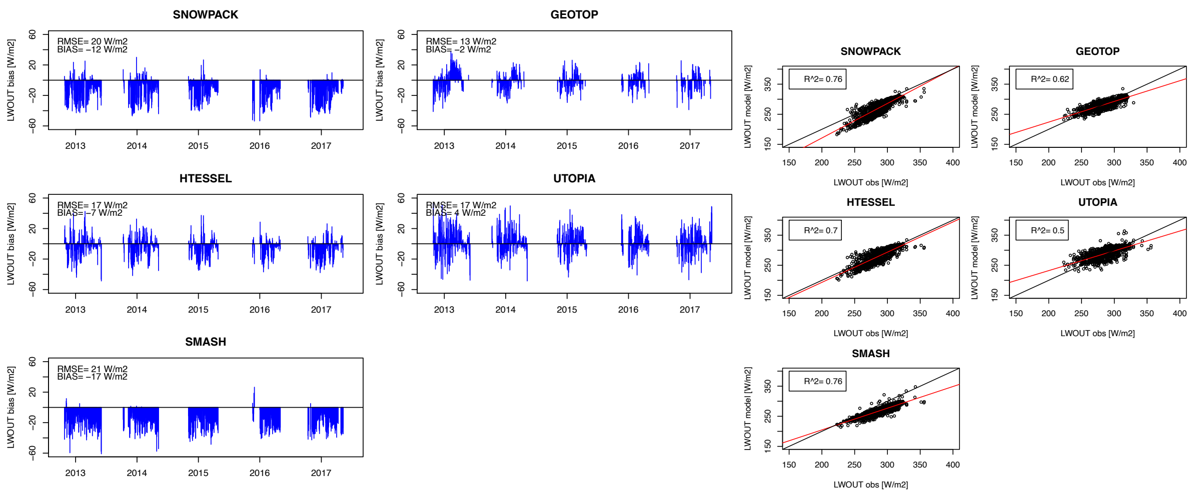

Figure 2 Difference (left figure) and scatterplot (right figure) of the modelled and observed outgoing longwave radiation (CTL experiment) at the Torgnon site. Only five out of the six models considered in the paper provide the outgoing longwave radiation among the output variables.

- **p2, line 28 There is no need for feedbacks for early differences in snow simulations to persist throughout the winter if the imposed conditions remain too cold for melt.**

We thank the reviewer for the suggestion; the text has been modified accordingly.

- **p3, line 22 delete "air" in "open air sites"**

Done, thank you.

- **p3, line 25 Specifically, Rutter et al. (2009) found that benefits from calibration at forest sites did not transfer to nearby non-forested sites. Direct calibration at the non-forested sites would almost certainly have improved simulations.**

We thank the reviewer for the suggestion; the text has been modified accordingly.

- **p6, line 18 I do not think that wind direction is needed to force the snow models, and please clarify whether any of them use surface temperature.**

Indeed none of the models employs the wind direction, thanks for the correction. With "surface temperature" we actually meant "ground temperature at 2 cm depth", and this variable is needed by the SNOWPACK model only. We have now clarified both points in the text.

- **p7, line 1 "both liquid and solid fractions" means that total precipitation is measured, not separate snowfall and rainfall.**

Yes, exactly. At the Torgnon station the total precipitation amount is measured. We modified the text to better clarify it, thank you.

- **p7, line 23 After reading this, I expected Appendix A to give details of the vertical gradients used for temperature and precipitation interpolation.**

Yes, we added this information in the text, thanks for the suggestion.

- **p8, Table 1 Information on the elevation of the reanalysis grid points would be interesting here or in the text. Also, how much gap-filling was required in the station data?**

Information on the elevation of the reanalyses at the Torgnon gridpoint, as well as the % of missing value for each input variable provided by the Torgnon station have been added in Table 2.

- **p9 I am confused by SWIN-CLS. If R is measured radiation and SWIN is modelled clearsky radiation, I don't see where the MSG cloud masks are being used. If R is incident solar radiation in cloudy conditions, isn't equation 2 the wrong way round?**

Yes, the equation was wrong and it has been corrected, many thanks for pointing it out. MSG cloud mask is used to identify the radiometers under clouds and compute an average attenuation factor. We have better explained this in the text.

- **p9, 13 Linear interpolation of sampled radiation fluxes rather than solar elevation-based interpolation of accumulated fluxes will be biased. How do average fluxes compare? (briefly mentioned in 5.3 and turns out to be a source of error)**

We agree that linear interpolation of sampled shortwave radiation fluxes introduces errors in the forcing data, which can be large when the sampling time step is 12h (TIME-12h experiment). For this case, we tested a different method to estimate 30 minutes shortwave incoming radiation (SWIN) from 12h samplings, based on the rescaling of the potential radiation with respect to the measurement at 12:00.

The details of this exercise, as well as the comparison between the linearly interpolated SWIN and the modified SWIN forcing, are provided in the reply to Referee#1, at point 4. In summary, while the linearly interpolated SWIN forcing shows an average bias of +97 W/m^2 compared to observations in the period of investigation, the modified SWIN has significantly reduced the bias to a value close to zero (-0.87 W/m^2), so the average flux is conserved. We ran an additional experiment (TIME-12h-SWIN-POT) using the modified SWIN forcing (see details in the reply to Referee#1, point 4).

A detailed discussion of this point and the results of the new experiment have been included in the manuscript in Sections 4 and 5.3

- **p10 Is the partitioning of total precipitation into snowfall and rainfall only applied to station measurements (in which case I expected to read about it in section 3) or also to the reanalyses, even though they provide separate snowfall and rainfall?**

We applied the same method to separate rainfall and snowfall for all the forcing datasets, including reanalyses. We better clarified it in Section 3 and 4.

- **p12, line 31 Even a model that could account for impurities would not do so in this case because dust deposition was not provided as an input.**

Yes, thank you for this comment. We modified the text as suggested

- **p18, line 33 MeteoIO errors are relatively small for temperature and snowfall, but errors in other forcing variables are not shown. Correct spelling of "systematically" throughout**

The MeteoIO forcing biases with respect to the Torgnon measurements are relatively small on average not only for temperature and snowfall but for all variables (Figure 9). In order to explain the discrepancies between the simulated SWE and snow depth in the MeteoIO experiment and observations (see i.e. Table 4 in the manuscript), we investigated the temporal variability of the air temperature bias (Figure 10) and we related it to the simulated SWE and snow depth in the MeteoIO experiment (Figure 11). The temperature bias is about -1°C on average over the considered time period, however in winter the cold bias is generally stronger and it can reach values exceeding -4°C (Figure 10). Concerning MeteoIO-driven snow model simulations, the main issue is the overestimation of snow depth in winter (in selected snow seasons) and in spring (in all seasons). A plausible explanation for these errors is that colder-than-observed winter temperatures might favor the development of a cold snowpack which melts too slowly. Consequently, the models tend to overestimate the snow at surface and to predict a delayed ablation date.
We added these comments in Section 5.4 of the manuscript.
The typo "systematically" has been corrected, thank you.

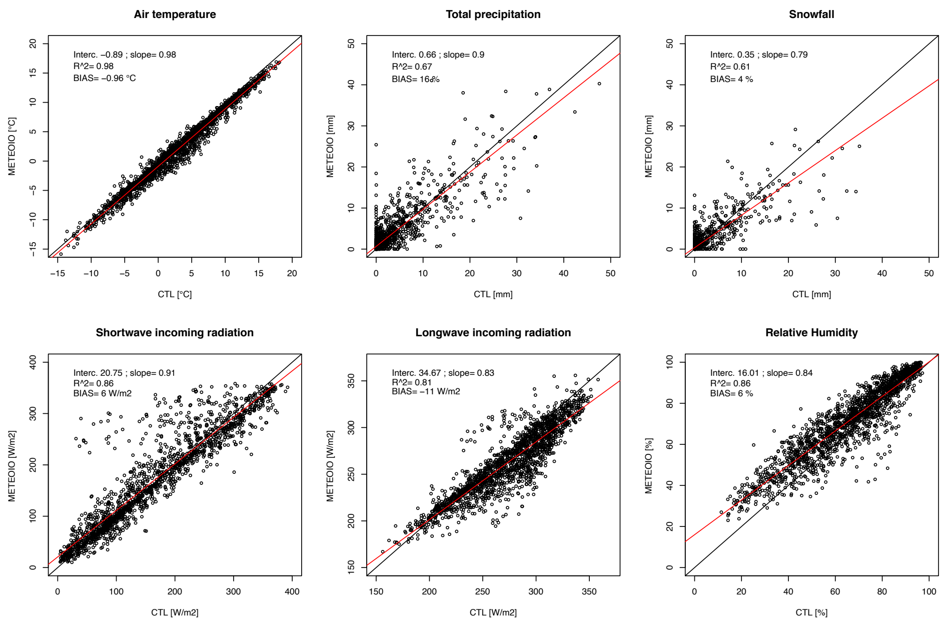

Figure 3. Scatterplot of the meteorological forcing of the MeteoIO experiment with respect to the CTL run.

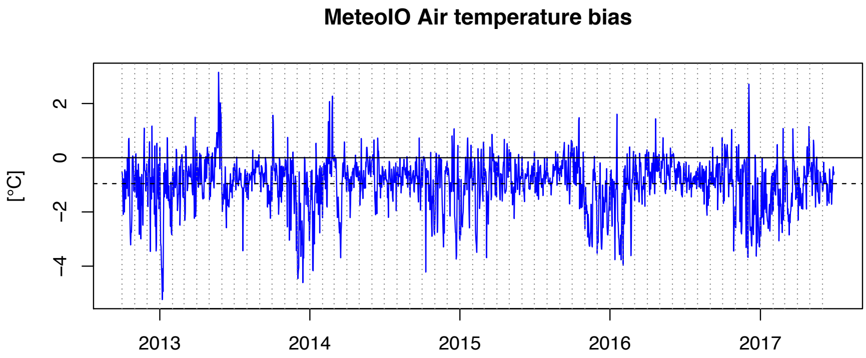

Figure 4 MeteoIO air temperature bias with respect to Torgnon station measurements.

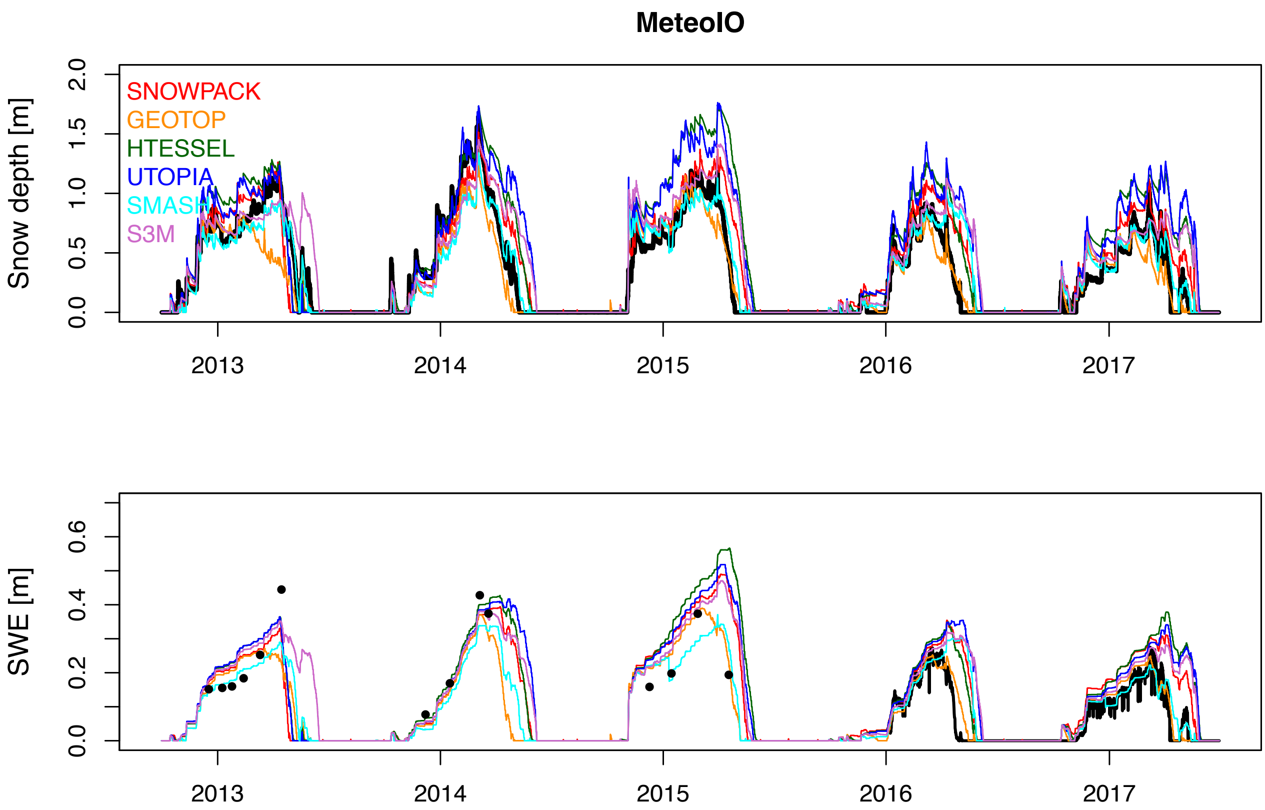

Figure 5 Simulated snow depth (top) and snow water equivalent for the MeteoIO experiment compared to observations.