**General**

Numerical Weather Prediction (NWP) models, atmospheric reanalyses and Earth System Models (ESMs) are being used increasingly to study the weather and climate of the polar regions. However, the land surface and surface exchange schemes used in such models have generally been optimised for performance in mid-latitudes and, while they may include a representation of snowpack processes, the snow models used may not correctly represent processes at work in the cold, dry and persistent snowpacks of the polar regions. This interesting and well-written paper reports an intercomparison of stand-alone snowpack models using forcing and validation data from Dome C, Antarctica, a location which is representative of the high Antarctic plateau. The intercomparison, carried out as part of the GEWEX GABLS project, included snow models that are components of a number of leading ESMs and NWP models and the results will be of value to users of those models, particularly those using these models in the polar regions. The paper is (mostly) written very clearly and presents useful results. I recommend that it should be published in The Cryosphere after attention has been given to the points listed below.

**Major points**

- Section 2.2: I found this section quite difficult to follow. In particular, I found the different ways in which the various runs are referred to - “reference simulation”, “XP0”, “XP1”, “_new” – rather confusing. Please try to adopt a clear and consistent terminology. Are the parameters shown in table 3 those that were used in the
“reference simulation” (XP0)? I was also quite confused about how XP1 differed from XP0 – are the changes only applicable to single-layer models? As this section is describes a central part of the methodology of the study it needs to be rewritten to improve its clarity. It might help to add additional tables to highlight the differences between the XP0 and XP1 runs. You don’t say anything in this section about how the models were initialised (presumably using observed snow temperatures?) or whether any spin-up was undertaken to avoid initial transients.

- Section 3.4: The measured values of $Q_h$ and $Q_{le}$ on figures 9 & 10 indicate a Bowen ratio, $Q_h/Q_{le}$ of O(1), which seems remarkably low at such low temperatures (see discussion in King et al, 2006, doi:10.1029/2005JD006130). Measuring $Q_{le}$ at such low humidities is very challenging and I suspect that there are very large uncertainties in the measurements.

- Section 3.4, fig. 11: Looking at this figure, it is apparent that most of the models specify $C_h$ as a universal function of $R_i$, but GDPS4 and CLM4 seem not to do so. What other factors are used in the calculation of $C_h$ in these models? Would the curves for the different models collapse onto a universal curve if, instead of plotting $C_h$, you plotted $C_h/C_h(R_i=0)$ – i.e. the ratio of $C_h$ to its value under neutral stratification (which should only depend on the roughness lengths)? It would also be interesting to plot $C_h$ calculated from the Dome C observations on this figure. From my own experience, I would expect to see a lot of scatter if you plotted points for each 30 minute observation, but if you averaged these together in bins of $R_i$ you might get a useful set of points for comparison with the models.

- Section 4, concluding remarks. This study has focussed largely on snow models that are used within global models and have not been specifically optimised for polar conditions. It might be worth mentioning here work that has been done on developing polar-optimised snow/firn models for use within regional climate/NWP models, such as Polar WRF (Hines and Bromwich, 2008, 10.1175/2007MWR2112.1), MAR (e.g. Agosta et al, 2019, 10.5194/tc-13-281-2019) and RACMO2 (e.g. van Wessem et al, 2018, 10.5194/tc-12-1479-2018).

**Minor points and typographical corrections**

Line 40: “firn”, not “firns”

Line 146: “UK Met Office”

Table 1: Maybe add a column indicating which NWP/ESM/reanalysis models use the snow model that is being tested. I assume that “LMDZ” refers to the snow submodel used within the LMDZ global atmosphere model – doesn’t the submodel have its own name?

Lines 171-174: Maybe include a table that gives the snow layer depths, densities, etc.?

Line 211: “...enabled us...” instead of “...allowed...”? 
Figure 1: Pressure seems to have been recorded with only 1 hPa resolution? “Direct solar radiation...” in the caption should (I think) be “downward solar radiation”. "Direct" would usually mean direct solar beam only, i.e. not including the diffuse component.

Table 3: Caption could be made a bit more informative. [...] indicates the range of a parameter that is calculated within the model. Presumably single values are where a fixed value is specified? Explain why several values depart from the control run values listed in section 2.2.

Line 286: Change “Fig. 4 shows the evolution of the broadband albedos that vary over time. Indeed, the four models presented consider the variation of the albedo...” to “Fig. 4 shows the modeled broadband albedos in the four models that model the albedo...”

Line 321: Change “radiation” to “energy”.

Lines 389-390: “...that varies from simple to double” Please clarify

Lines 390-392: “It should be noted that the single layer models (D95, CHTESSEL, EBA) have sometimes better results than the more sophisticated models which have to represent more physical processes, such as the evolution of albedo with time, the increase of snow density by compaction, among others.” Maybe you could make this clearer on figure 8 by using a different shape of symbol for the multi-layer models?

Line 442: “…gradient Richardson number...”. Model parametrisations are usually based on a bulk Richardson number, calculated from the temperature difference between the lowest model level and the surface and the wind speed at the lowest model level. Include an equation that defines how the Richardson number is calculated.