The paper is a solid and well-founded comparison between northern and southern hemispheric MLT observations on the one hand and the comparison of the model output of three whole atmosphere GMCs with the observations on the other hand. The structure of the paper is logical and clear, but resembles more a technical report than a scientific paper. The main presentation of the results is a mixture of contrasting observations from the two hemispheres for different locations and the same presentation for the model results (10 of 15 figures do this plus the five in the Appendix). As the technical methodology appears to be very sound, I only have two main remarks that might help modify the current manuscript and improve the presentation:

(a) I would suggest separating the physical comparison between the hemispherical observations from the comparison GCMs to the observations. This would allow the authors to formulate research questions that can be addressed and answered by the comparison. The comparison between the GCMs and the observations should constitute a second main part of the paper. Currently, it is a hodgepodge, hard to read and difficult to separate the individual results.

(b) The authors go to great lengths to create a homogeneous data set consisting of both observations and model outputs on comparable altitude-time grids. I wonder why the results are presented and discussed only qualitatively ("...agrees reasonably well ...", etc). Why don't the authors show differences of the climatological means MODEL vs. OBSERVATION? I admit that the authors use a lot of effort to turn the shocking disagreements into positive words (e.g. "... shows a better agreement with the radars for the hemispheric zonal summer wind reversal...") but for scientific usage a QUANTIFICATION of the differences would be really desirable!

Nevertheless, the study has its merits but really needs focus.