Comment on acp-2022-150
Anonymous Referee #2

Referee comment on "Interannual variability of winds in the Antarctic mesosphere and lower thermosphere over Rothera (67°S, 68°W) in radar observations and WACCM-X" by Phoebe E. Noble et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2022-150-RC2, 2022

Interannual variability of winds in the Antarctic mesosphere and lower thermosphere over Rothera (67°S, 68°W) in radar observations and WACCM-X

by Noble et al.

This paper presents results of long-term meteor wind radar observations at Rothera (67°S, 68°W), Antarctic, and also WACCM-X modelling results for corresponding area and period, focusing on the interannual variability at MLT region. Monthly median zonal and meridional winds are displayed as time-height cross-section between 2005 and 2020 and difference between observed and modelled wind is discussed. The authors also applied linear regression analysis for the period between 2005-2015 on five indices of Solar cycle, ENSO, (two heights of ) QBO, and SAM. The analysis is extended to gravity wave tendency (zonal drag).

The data set and analysis of long-term observation over the Antarctic station and corresponding WACCM-X data are rare and valuable in understanding MLT dynamics and interannual variability. Results shown in '4. Results: The winds in radar observations and WACCM-X' is overall fine. However, the linear regression analysis results shown in '5. Results: Linear regression analysis’ need more careful consideration and evaluation.
My main concern is the number of samples in linear regression analysis. The sample for linear regression analysis is three-month window times 11 years, i.e. 33 samples (data points) for each time and height bin. The estimated coefficient ($\beta$) is as many as six, as shown in Equation (1). I suppose this number of coefficient is too many for 11 years of data. The variance in these 33 samples is not only due to the interannual (or year-to-year) variability, but also intra-seasonal (or month-to-month) variability, because variation in three-month window is also included. Statistical significance of each regression coefficient must be overestimated because the interannual variability is not included in three sample in the same year window (of three month). Data number of 33 are not number of independent samples in the sense of interannual variation, and therefore freedom of 27 is overestimated. Multicollinearity (3.1.1) check by VIF should be carefully evaluated because under the presence of both intra-seasonal and interannual variability VIF should be underestimated. Also, if the autocorrelation in 3.1.2 is calculated for all 11-year time-series, this will smooth out different auto-correlation function at different season (or three-month window) so the value should be underestimated. Effect of such seasonal smoothing (or averaging) should affect DW test, and this should be checked. Therefore, my recommendation is that authors clarify the above questions before publishing the results of Chapter 5, 6 and 7.2.

Specific comments:

L 121-3

‘These zonal gravity-wave tendencies in WACCM-X were found to be noisy when examined over the meteor collecting region. We therefore calculated tendencies as zonal-means in a band of 300 km latitudinal width, centred over the latitude of Rothera.’

This assumes that the gravity wave drag is uniform at all longitude if monthly averaged. Is this an appropriate assumption? It is known that Andes and Antarctic Peninsula is the region of strong gravity wave generation. I do not think the zonally uniform assumption is correct.
'In (a), it can be seen that the interdecile range of the zonal wind maximises over the summer, when the zonal wind reversal occurs, due to the considerable variability in the strength and timing of this reversal.'

Please check whether this difference of the interdecile range could be due to the difference of zonal wind magnitude between summer and winter. If the fluctuation is a certain % of the amplitude, this also cause the variation of interdecile range measured by m/s.

'Hatched regions show where the relationship is statistically significant at the 90% level, using the Student’s t-test.'

I am suspicious about this, considering the point described as ‘main concern’ above.