Review of “Spatiotemporal Distribution of Tropopause Folds over the Arctic during Spring Season: A 24-Year Analysis (2000-2023)” by Chen et al.

In this study, Chen et al. investigated the spatiotemporal distribution of tropopause folds over the Arctic in spring over 24 years (2000-2023). Based on ERA5 reanalysis data, they identified tropopause folds following Škerlak et al. (2015). Then, they analyzed the 24-year mean spatial distribution and temporal variations in different time scales of the tropopause over the Arctic. They found poleward decreases in both fold frequency and the upper percentiles of fold depth. They also found positive trends in total fold frequency at four stations. This study is well motivated, and some outputs are scientifically meaningful. However, the paper can be improved, as suggested below.

This paper includes a few figures, such as Figure 5b, that directly show a link between fold frequency and upper-tropospheric dynamics. However, most of the discussion is merely descriptive of the figures, and most figures show only spatial distributions of different variables, which makes the paper more like a report than a scientific paper. The authors are encouraged to expand the discussion in depth, compare their study with previous work (especially Škerlak et al. 2015), and highlight the new and important discoveries. Most importantly, add more analysis similar to Figure 5b to connect variations in tropopause folds to their underlying drivers.

Line 100-104: In this study, “four high-latitude monitoring sites” are selected. Indeed, the data from these sites can be used to validate ERA5 wind and other meteorological variables. However, since stratospheric folding is identified using gridded ERA5 data, it would be better to select four or more regions for analysis rather than these four stations. These four stations represent only part of the Arctic, not all.

Line 110-112: Since the authors focus on tropopause folding, validating EAR-5 with only surface data would not be sufficient. The authors could validate tropopause-folding events or temperature profiles using IGRA2 data. Tropopause folding from IGRA2 data can be determined following the approach of Xian and Homeyer (2019) (https://doi.org/10.5194/acp-19-5661-2019).

Line 128-130: ERA5 reanalysis data on model levels have higher vertical resolution and would therefore be more accurate in determining tropopause folding events. However, the authors used ERA5 data on pressure levels instead. While using the model levels would significantly increase workloads, we suggest selecting a few sites or grid points to compare the differences between the model-level and pressure-level results

Line 156-157: Please provide explanations with references for why these variables were

selected to characterize the upper-tropospheric dynamical background associated with tropopause folding.

Line 186: Section 3.1 presents the spatial distribution of tropopause folds in multiple figures but lacks comparisons with previous studies. What are the new discoveries in this study? One of them may be the poleward decrease in fold frequency. The authors related this poleward decrease to the eddy growth rate. However, the explanation of the eddy growth rate and its physics is missing.

Line 234-235: Are there any previous studies or materials in the Supplement that support the statement?

Line 246-263: This summary would benefit from some references supporting these dynamical interpretations, which would make the physical explanations more convincing.

Line 306-320: This paragraph does not clearly explain a quantitative relationship between changes in fold frequency and changes in the upper-tropospheric background. The interpretation relies mainly on visual comparisons among several spatial trend distributions. The authors should provide a clearer illustration, similar to Figure 5b. These illustrations could include scatter plots, correlation/regression analysis, or spatial pattern correlations between fold-frequency variability and relevant dynamical variables.

Figure 6 and Figure 7: The figures' titles are incorrectly labeled as Figure 4 and Figure 5.

Figure 8: Only positive trends in Villum site are statistically significant. The wording regarding increasing fold frequency should clearly distinguish between statistically significant and nonsignificant trends.

Figure 12 and Figure 13: Please provide the mean, standard deviation, or other relevant statistics in the top-right corner of each panel. This would be more helpful.