Reply to Referee 2

We thank referee 2 for the constructive, helpful criticism and the suggestion for revision. We have thoroughly revised the manuscript based on the comments given by the referees. A detailed point-by-point response to the comments by referee 2 is given below.

This study evaluated various compression models at different compression rates applied to ERA5 reanalysis data. The compressed ERA5 output was then used to drive MPTRAC for trajectory calculations. The paper investigates how both the choice of compression scheme and compression rate affect the accuracy of trajectories computed by MPTRAC.

The manuscript is well-structured overall, but the writing could benefit from some revision for clarity and flow, as it's currently a bit difficult to read. Below are my comments:

Major Comments

Fig 5-Fig6

- Do you have an explanation for why the deviation is larger near the surface.

  The stratosphere is more stable which suppresses vertical mixing. Therefore, in the stratosphere, air parcels follow smoother, more predictable paths. In contrast, the troposphere is more unstable and thus affected by turbulence and convection which makes trajectories harder to predict over time. Thus, it is not astonishing that in the lowest considered altitude range (2-8 km) the largest deviations are found.

  We have added the following sentence in Sect. 3.2.1 to explain this: "This is because the stratosphere is stable stratified and the wind fields are much smoother, allowing for more accurate trajectory calculations. In contrast, the troposphere is much more dynamically variable with turbulence, convection and vertical mixing, making accurate trajectory calculation more challenging."

- I am curious to know why the deviation is negligible in the days < 4 regardless of the compression method and compression level? Why does the deviation grow in time? The similar think is also observed in Fig 7 as well.

  Trajectory calculations suffer from several error sources and the errors accumulate and amplify over time (Kuo, 1985; Stohl, 1998; Harris, 2005; Engström, 2009). The longer the simulation or the more steps are taken, the more these errors can cause deviations from the true trajectory. This is why the deviations get larger as longer the trajectories are calculated.

  We have added the following sentence to Sect. 3.2.1: "This development in time is typical for transport deviations, since trajectory calculations suffer from several errors that accumulate over time (e.g. Kuo, 1985; Stohl, 1998; Harris, 2005; Engström, 2009)"

Minor Comments

Line 149-151: Is there a reason for choosing the tolerance option for geopotential height and Temperature while choosing the precision mode for all other variables?

The set-up for ZFP that we use was derived in previous studies made by one of our former group members and was found to work reasonable well with this data set.

- Line 224-226: It was mentioned that the correlation coefficient reached 0.99999 when the precision mode was used. What's the reason for choosing the tolerance option for T?

Although we could achieve a higher correlation coefficient in this test case, we kept our original set-up. This is because (1) we tested this only for ZFP16 and not for ZFP12 and ZFP8 and (2) it did not change the results concerning the trajectory calculations.

Line 233-235: According to Table 1, file size after the compression by PCK is larger than the file size after the compression by ZFP. Why do you think reading input file takes the shortest time for PCK and not ZFP compressed files?

This is because of the way how the data is compressed and then again decompressed. PCK was specifically designed for netCDF files and stores the data as arrays of scale and offset values. ZFP on the other hand comes from the image processing and devides the data into fixed size blocks of dimensions $4 \times 4 \times 4$ that are each stored using the same, userspecified number of bits. Although the compression ratios are higher for ZFP and the file sizes are smaller, the decompression is faster with PCK.

Fig4: Lines are labeled as physics, input, output, and total. But the caption or the main text is missing the explanations for those labels.

We have added the following sentence to the figure caption and to the main text to explain the used labels.

"Timers were used for reading the input data (Input), performing the trajectory calculation (Physics), writing the trajectory data to an ascii or netcdf file (Output) and the total time spent for the trajectory calculation (Total)."

I would suggest labelling each panel in each figure.

Thanks for the suggestion. We have labelled all figures. Additionally we have, due to a comment by the Editorial Support Team that came up during file validation, added to the linestyle markers, so that the figures are readable for readers suffering from all kinds of coulour blindness.

Is Fig5-6 are ensemble mean? Or from a single trajectory?

This is a mean over all globally distributed 10^6 trajectories that we have calculated. To make this more clear to the reader we have rewritten the sentence in Sect. 3.2.1 and refer also to Sect. 2.5 where the equation and some further explanations are given.

"Figures 5 and 6 and Fig. B1 and B2 show the absolute and relative horizontal and vertical transport deviations (AHTD, RHTD, AVTD and RHTD, respectively) for the 10^6 10-day trajectories (mean over all trajectories, see Sect. 2.5) that have been calculated with MPTRAC using the compressed reanalysis data files (PCK, ZFP16, ZFP12 and ZFP8)"

Line 285-287: Do you mean that the maximum deviation frequencies are similar between the

trajectories started at tropospheric and stratospheric altitudes?

Yes, for ZFP8 the mximum deviation frequencies for trajectories started in the stratosphere are almost as high as the ones started at tropospheric altitudes. We changed the sentence as follows to be more clear:

"For the trajectories calculated based on the ZFP8 compressed data the maximum deviations for the trajectories started at stratospheric altitudes occur almost as frequent as for the ones started at tropospheric altitudes."

References

Engström, A. and Magnusson, L.: Estimating trajectory uncertainties due to flow dependent errors in the atmospheric analysis, Atmos. Chem. Phys., 9, 8857–8867, 10.5194/acp-9-8857-2009, 2009.

Harris, J. M., Drexler, R. R., and Oltmans, S. J.: Trajectory model sensitivity to differences in input data and vertical transport method, J. Geophys. Res., 110, D14109, 10.1029/2004JD005750, 2005.

Kuo, Y.-H., Skumanich, M., Haagenson, P. L., and Chang, J. S.: The accuracy of trajectory models as revealed by the observing system simulation experiments, Mon. Weather Rev., 113, 1852–1867, 10.1175/1520-0493(1985)113¡1852:TAOTMA¿2.0.CO;2, 1985.

Stohl, A.: Computation, accuracy and applications of trajectories—a review and bibliography. Atmospheric Environment, 32, 947–966, 10.1016/S1352-2310(97)00457-3, 1998.