Review for

Regional and seasonal truncation errors of trajectory calculations using ECMWF high-resolution operational analyses and forecasts

by Rössler et al.

Synopsis:

In their study the authors look at the truncation errors of six explicit integration schemes of the Runge-Kutta family. The performance is studied based on real-case data from the operational ECMWF analysis and forecasts, whereby the sensitivity with respect to the sphere (troposphere, UTLS, stratosphere) is discussed. Further, the seasonal dependence of the errors is compared, and the computational efficiency is discussed. The paper is very well written, the argumentation very clear, and the number and quality of the figures support well the discussion. I think the paper fits well the interests of the GMD readership, and therefore I certainly can recommend its publication. Still, the authors might want to address the following concerns.

Concern:

1) The abstract (and manuscript) ends with a rather strong conclusion: "we recommend the 3rd-order Runge Kutta method with a time step of 170 s or the midpoint scheme with a time step of 100 s for efficient simulations of up to 10 days time based on ECMWF's high resolution meteorological data." This is, as the authors note, far below the time step that typically is applied in trajectory calculations based on ECMWF fields. I think the authors can clearly demonstrate that such a small timestep is indeed necessary to get a high degree of accuracy of a *single* trajectory -- where the convergence of the trajectories is assessed based on the AHTD and AVTD distance metric. However, I wonder whether we should trust any single trajectory anyway. Let me make my point more clear: Suppose that we have a calculated a *single* trajectory which reaches after 10 days a AHTD(single) of 100 km. Hence, the trajectory calculation is not perfect. But now also assume that we very slightly change the starting position of the trajectory and repeat the trajectory calculation. We can now compare the distance between the initial and shifted trajectory, and the resulting metric is AHTD(single-shifted) = 200 km. Of course, we could repeat this kind of experiment with several shifted starting positions. The point is that the AHTD(single) can now be seen in a better light, because it is smaller than the inherent spread AHTD(single-shift) due to a minor shift of the starting position. I would argue that the uncertainty of the single trajectory is negligible compared to the flow-inherent dispersion of the trajectories.

In short, I think that there is not too much meaning in considering single trajectories at all. We always have to look at an ensemble of trajectories started from nearby positions. The coherence of this trajectory ensemble then defines the time horizon until the trajectory is meaningful. Of course, there is also some subjectivity in this argument: The slight shift in starting positions has to be specified. Still, I think the authors should comment on this 'coherent trajectory bundle vs. single trajectory' concept.

2) In Figure 5 the winter 2015 stands out. The authors find a reasonable explanation for it: a sudden stratospheric warming and near splitting of the polar vortex. I think this explanation makes perfect sense, and actually points to a potentially interesting extension of the study. In fact, we can expect a varying degree of inter-annual variability not only in the stratosphere, but also in the troposphere and in the UT/LS. There are years with more or less cyclones passing along the storm tracks; there are years where the jet stream in the UT/LS meanders more than in other years (with a more zonal jet). This variability is reflected in climate indices (e.g., the NAO), but it could also be assessed by explicitly 'counting' the cyclones, anticyclones, or by considering a measure of jet zonality. In short, it would be rather interesting to see the trajectory accuracy in context of this inherent tropospheric, UT/LS, and stratospheric flow variability. I don't expect the authors to do that all in the current study! But, possibly they can think about it, and thus link their findings more to meteorology than 'abstract' statistical measures. If appropriate, I would appreciate if the authors comment on this perspective in their study.

Minor comments:

-P2,L25: "However, it needs to be stressed that appropriate ..." -> "However, the appropriate ..."

- P3,L1,3,4: Three sentences starting with 'We' - please rephrase!

- P4,L12: "Wind dynamics in the extratropical summer hemisphere are generally slow" -> Unclear what is meant by this statement? Do you want to say that winds in summer are slower? Or that they are not changing as much?

- P6,L27: "we calculated the horizontal distances as Cartesian distances of the air parcel positions projected to the Earth surface" -> It is not completely clear how the distance is calculated. What are Cartesian distances on a sphere?

- P8,L15: "it needs to be pointed out that this is undersampling" -> "this is still undersampling"

- Figure 1: In the upper-right panel the vertical wind velocity is shown. A rather large-scale wave pattern is discernible over northern Europe. I wonder whether this pattern is physical, or some kind of numerical artefact? The amplitude of the waves is rather small, as expected in the stratosphere.

- Figure 5,6: I wonder whether it would be better to reduce the number of panels, e.g., by only showing the results for the northern hemisphere? Of course, it would (for instance) also be interesting to compare the northern UT/LS with the southern UT/LS. But, at the moment the UT/LS is defined by means of fixed heights (8-16 km) and it is not clear whether the tropospheric fraction for the southern hemisphere is the same as for the northern hemisphere. If not, and this will certainly be the case to some degree, the two hemispheres are not really comparable.

- P11,L1-10: The values listed in the text are better presented as a table.

- P11,L29-30: "The tropospheric mid-latitudes were expected to cause the largest errors, because the most complex wind systems occur in this region due to a larger land surface ratio and more complex orography" -> What do you mean with 'complex wind systems'? What is the 'land surface ratio' - most likely you mean 'land-sea ratio'? Further, it is rather unspecific to attribute the flow variability to the orography and/or the land-to-sea fraction.

- P11,L33: Again, what is a 'complex wind pattern' and why is the turbulence higher in this region? I guess that the authors point to the higher jet variability, i.e., its north-south meandering structure. I would suggest to add some references to climatologies that quantify this variability.

- P12, L17: "As a rough indication for inter-annual variability" -> This is indeed a very rough measure for *inter-annual variability*! If I hear 'interannual variability', I would expect a study covering at least 10 years. Hence, I would simply say that the two years 2014 and 2015 are compared, and that they differ substantially -- indicating that the inherent flow properties have a considerable impact on the outcome.