# 1 General

The authors have implemented suggestions and thus improved their paper. However, not all recommendations have been sufficiently addressed, as detailed below. Specifically, certain problems related to meteorological reasoning, for example invoking turbulence as explanation for larger transport errors, have not been resolved.

Quotations from my orginal comments are set in *italics*, those from the authors' answers or revised manuscript in 'single quotes'.

# 2 Major remarks

1. clarified

2. *Another open question is whether RK4 with 60 s time step is a suitable reference method.* The authors accept this critism for the northern midlatitudes. However, they don't offer an improvement or a justification of keeping their reference. I would expect that the authors address this point either by repeating the calculations and evaluations with a shorter reference time step, or by adding an explanation and justification within their paper. Their remark on Hoffman et al. (2016) is not leading to an argument, it just repeats and confirms what I wrote.

3. clarified

4. OK.

5. *Finally, the results are certainly sensitive to the resolution advection module and ECMWF operational analysesof the wind field data. Results obtained for the specific case of 16 km / 3 h therefore cannot be generalised. Keeping in mind the conclusions of Stohl1995, Brioude2012, and Bowman2013, 3 h intervals for the wind fields are coarser than what would be desired at this horizontal resolution. As 1 h is provided by ECMWF, I am wondering why it was not used.*
   I don't feel satisfied by the answer given. It is not appropriate to call the 3 h 'officially approved resolution' (there is no such thing as official approval, all what is in MARS is usable). The question of how results would change with 1 h data is highly pertinent. Alternatively, one could test coarser temporal in combination with coarser spatial resolution.

6. New issue created by introducing the wording 'total error' in recognition of the fact that the transport errors obtained are not pure truncation errors: As shortening the time step eliminates only errors introduced by the truncation error, the 10-day error is not the total error. The total error would be larger than that as it has other contributions as well which are also amplified during the 10-day transport. The authors should make that clear and find another, more appropriate wording. See also major remark #1 of Reviewer #2.

# 3 Specific and Minor Remarks

1. Authors now propose the title 'Domain specific trajectory errors diagnosed with the MPTRAC advection module and ECMWF operational analyses'. First of all, one would have to hyphenate 'domain-specific'. Then, I think the word *domain* is not the best choice to express that evaluations were done separately for different regions (*domain* usually refers to a calculation domain and not to a climatological region.) Furthermore, MPTRAC and

ECMWF data are not on the same level, one is the model, the other model input. Better say something like 'Trajectory errors as a function of numerical scheme, time step, and region of the atmosphere diagnosed with the MPTRAC model'. Maybe you can add 'for ECMWF wind fields', but I think it is not so important to bring that into the title.

2. Page 1, line 1: *Abstract*. OK

3. *Page 1, line 4: kinematic equation of motion* (comes also in other places). *I don't feel comfortable with this wording. "Equations of motion" for me would refer to the Euler or Navier-Stokes equations. Why not call this the trajectoy equation?*
   Authors decline to change their wording without providing arguments. I will accept their wording if they provide a quotation from a well-established meteorological textbook which uses 'eqation of motion' for the kinematic trajectory equation.

4. OK

5. It is not true that no forecasts are produced from 06 and 18 UTC analysis. However, these are not long-term forecasts but just for providing background fields at the next major analysis step.
   I accept the argument that these details don't belong to the introduction. Therefore, I would suggest to provide all the information about resolution and kind of ECMWF input fields only in the Section 2, and to remove them completely from Section 1. Otherwise, one is wondering about partial information until one has reached the next section.

6. OK

7. Reformulation is ok. I don't have the impresson that Heng et al. (2016) provides a full model description.

8. OK

9. The explanation is ok but it should also go into the revised manuscript.

10. OK

11. *Page 6, line 4 ff.:* $\boldsymbol{k}_1 = \ldots$ *.It seems that you define certain velocities as* $\boldsymbol{k}$.
    I regret that the authors want to keep this notation. In any case, symbols have to be explicitly explained (in words) before or immediately after first usage, whether or not there is an equation which defines them. This is a standard in scientific publications.

12. OK

13. OK

14. *Page 11, line 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.* The authors offer some improvement, but it is still insufficient.
    (1) They should pay attention to their style. For example, wordings such as 'The troposphere has its largest error' are not correct (this example is not the only such mistake). The troposphere cannot have any error, only calculations can.
    (2) We still find the phrase 'the evolution of northern mid-latitudes meteorological systems is more difficult to simulate than for the southern mid-latitudes due to the larger land-sea ratio and more complex orography of the northern hemisphere.' Apart from the

question whether this is true or not, the difficulty to simulate the evolution of meteorological systems (in other words, the predictability) is not relevant for trajectory errors based on analyses.
(3) 'These errors are caused by ... and higher turbulence in the underlying region'. In their answer to minor remark 16 authors admit that turbulence is not relevant – thus, why do they again come up with turbulence as an explanation of trajectory errors?

15. OK

16. *Page 12, line 5: The relative high errors in the tropics are probably caused by a stronger turbulence in that region. The lower bound of the stratospheric region of our test cases is 16 km, since the tropopause reaches an average altitude of 16 km near the ITCZ, turbulent movements due to deep convection can occur more frequently in the lower stratosphere above the tropics.*
The authors admit that the term 'turbulence' is misleading here, but they have not changed their wording. The explanation by turbulence is wrong and has to be removed. If they want to refer to 'fluctuations in the meteorological input data', I think they have to be more specific what is different in the tropics compared to mid-latitudes. Mid-latitude wind fields also show fluctuations.

17. OK

18. *Page 12, line 12: We need to stress that each simulation lasts only 10 days, which is a relatively short time interval to analyze seasonal effects. Fast temporal variations and changes in medium-range weather patterns can blur out the impact of seasons that is observed here. To better resolve the seasons you don't need longer trajectories, but more frequent starts or more years. I any case, I don't think that the seasonal effects are so interesting, you could discuss this just briefly. It is obvious that stronger variations in the wind fields will lead to larger truncation errors, and the dependence of the variability of wind fields on the seasons is well known.*
Most of this comment is not at all addressed in the reply by the authors. It is really questionable whether seasons are represented in a statistically adequate way with a single day on which trajectories were started. Could you not just add some more? They did 5000 (10-day) trajectory simulations, with a parallelised model on a supercomputer. Doing 20000 or 50000 instead would not be a serious burden in terms of computing work. It is true that the section on seasonal results is not very long, but if we add also the figures, it is also not so short. Generally speaking, I consider the merits of this paper lying in the realm of numerical methods, showing a systematic comparison of a number of numerical integration schemes. The layer of meteorological interpretation according to season, region etc. which has been put around that has much less scientific substance, and in its present formulation even is partly mistaken (see the turbulence issue). It would be better to de-emphasise this part.

19. OK

20. *Page 12, line 35: The median error gets somewhat larger in the troposphere, where particle paths are more likely being affected by atmospheric turbulence. Hoffmann et al 2016 says that MPTRAC uses the same diffusivity throughout troposphere and stratosphere. How is this compatible?*
The authors say that they answered this in their response to major comment 1. However,

there they only explain that turbulence is not active in their simulations. But then, this argument simple collapses; however, the sentence in question has been left unchanged.

21. *It would be useful to explain why you are only testing OpenMP and a single node if MPTRAC is capable to work on distributed-memory systems as well.* Answer: 'We added the following sentence in Sect. 3.4: The MPI parallelization is only used for ensemble simulations, which are conducted independently on multiple nodes. Therefore, the scalability of the MPI parallelization is mostly limited by I/O issues, which are out of scope of this study.'
   I don't understand the argument. If the strategy for distributed-memory machines is trivial parallelisation by multiple runs started concurrently, why does the code offer MPI-based parallelisation?

22. The explanation 'The time measurements refer only to the part of the code spent in the advection module of MPTRAC.' should be included also in the manuscript text. Furthermore, if this is the case, the contribution independent of the number of threads also stems from non-parallel parts of the code, not only from the OpenMP overheads.

23. The explanation given should be included in the manuscript text.

24. As the Pettersson method is widely used, it would really be desirable that the authors test its efficiency with a reasonable iteration cut-off compared for example to the midpoint method.

25. It would be good to put the explanations, e.g. about the role of cache for different numerical schemes, into the manuscript text.

26. OK

27. OK

28. *Page 14, line 25: We attribute this to larger small-scale variations caused by atmospheric turbulence and mixing in the troposphere. The first part of the explanation is correct, but the second part not. These variations are not caused by turbulence (16 km is not turbulence scale !!) and certainly not by mixing (this would reduce and not amplify variability!).*
   The authors have removed the reference to mixing, but they keep the reference to turbulence. As I have tried to explain in various parts of the paper where turbulence is invoked, this is not appropriate. Atmospheric turbulence does not create variability at the high-frequency end of the resolved motion scales. It will rather tend to undo existing gradients.

29. *Summary and conclusions*: Please spell out RK where it occurs for the first time in this section (some people may read only this section).
   'After 24 h the trajectory errors are quite similar in the troposphere and stratosphere' – Figure 4 shows a difference by about a factor of 10 for AHTD, thus they are not 'quite similar'.
   'We attribute this to larger small-scale variations caused by atmospheric turbulence.' – Remove erroneous reference to turbulence.
   Statistics not being sufficiently robust: as said before, it would be desirable to increase the sample size.

30. *Page 15, line 7–9: The study of Seibert (1993) .... To achieve truncation errors that are smaller than overall trajectory uncertainty, they found that the time step should fulfill the CFL criterion as a necessary condition for convergence. The recommendation there*

*for a sufficiently small truncation error was 15% of the time step needed for convergence of the Petterssen scheme. If we assume that the reference accuracy has also improved in the meantime, an even smaller value would result. The CFL criterion is recommended to make sure that no small-scale features are skipped, not for convergence of the iterations in the Petterssen scheme.*
The authors have amended their wording, but they have not changed the first sentence quoted above, which is not an accurate representation of Seibert (1993), as explained in my previous comment.

31. OK

32. OK

33. 'For better visibility of the circulation patterns we decided to use different scales for the three pressure levels. In case of vertical velocities the maximum values differ by more than a magnitude between the levels.'
I think a part of the motivation for showing sample vertical motion patterns at three different levels is exactly this fact that vertical motions are much smaller in the stratosphere, and this is hidden by changing the scale between the figures.
'The chosen pressure levels are used in the model and correspond closely to the altitudes given in the figure caption.'
If standard pressure levels also are available, it would be preferable to shown them.
'The colour coding for vertical velocity has been reversed following comments by reviewer #2.'
Thank you.

34. OK

35. I hope this annotation will really happen – why not doing it now?

36. *Figures 5 and 6: This figure should be simplified. You don't need to show the two years separately, and I think you also don't need to show seasons separately. Thus you could have just three subfigures (three levels) and the five regions inside of each one. Then use a log scale for the AHTD, and symbols instead of bars (which will bring out the median also more clearly).*
'We would like to show the simulation results separately because regional and temporal impacts on the error were a part of the motivation for this study.'
This is not sufficient as a justification – the question is whether there are enough interesting and relevant results to convey. As pointed out earlier, because of the small sample size (just two times one starting date per season) this is questionable. Most of the seasonal results are just representing general knowledge about seasonality of atmospheric variability. I don't see a good reason for presenting the two years separately – due to sample size, this is neither a good representation of year-to-year variability nor a replacement for error bars. Also, the year-to-year variabilty is not of interest here. Make your samples as large as possible, and then present one value for each!
'We added a horizontal line for the average error to all subfigures (see Figures 4 and 5).'
That is useful, but I would not call this 'average of the domain' but rather 'annual mean' or 'average over all seasonal samples'.
'We did not use a log scale, because it would hide the seasonal and regional differences.'
It would not hide them, even though they would be less strongly visible. It is not a good practice to present a set of figures where for the majority of them, more than 50% of the graph area is unused. Also, the stratospheric results are hard to see at all. If you don't

want to use a log scale, at least you should optimise them and use different scales for each level.

37. *Figures 2, 3, 4, 7, 8: Please make sure that line width, colour intensity and marker size are sufficient to read all the content easily.*
    'We tried to improve the figures accordingly'
    I am sorry, but I fail to see real improvement. Lines are still tiny and for some, very pale colours are used. For example, the Heun line is hard to recognise in the plots.

38. 'This is a helpful evaluation and we added a statement in the paper summarizing the findings regarding the number of threads providing the minimum computation time with respect to the number of particles'.
    Can you please point out where this statement is found? If you also find this useful, please make sure that it is properly presented.

39. OK