Reply to review comments

We thank the reviewers for the time and efforts spent on the manuscript. We considered all comments and hope that the revised draft properly addresses the remaining issues. Please find our point-by-point replies below (colored in blue).

Reviewer #1

1 General

This paper investigates the performance of a number of numerical schemes to integrate the trajectory equation. This is done using the LPDM MPTRAC. As the code is prepared for parallel computing, the performance is investigated also as a function of the number of threads (for one of the schemes only). For the tests, 10-day simulations were carried out using ECMWF data with 16 km grid spacing, and results are presented for different regions of the globe, layers of the atmosphere, and seasons. This study is a useful addition to the previous investigations, as it also tests higher-order methods rarely used in atmospheric transport modelling, and as parallel performance is included. I recommend publishing it after consideration of the following remarks. I think that the authors have some choices with respect to doing additional calculations and/or evaluations, and I hope that they would be able to consider my respective suggestions, as the value of this work could be significantly increased in that way.

2 Major remarks

1. In my opinion, there is one aspect in the setting of the numerical experiments which is not ideal. The regular LPDM code has been used, that is, including a stochastic wind component to represent turbulence. Existing similar studies have been carried out with simple trajectory models. It is not very clear what the consequence of adding stochastic wind components is for the deviations between the schemes tested. The authors propose turbulence as explanation for several of the observed variations in accuracy, but this remains hypothetic. I would strongly recommend to repeat at least a subset of the simulations with all kinds of stochastic influences (turbulence, mesoscale fluctuations, convection if it exists in the model) switched off, present and discuss these results as well.

The simulations are based on the advection module of MPTRAC solely, the modules for turbulence and mesoscale fluctuations were turned off. We added this information to the model description.

2. Another open question is whether RK4 with 60 s time step is a suitable reference method. If one extrapolates the RK3 or RK4 curves in Fig. 8 (bottom), one would arrive at an AHTD value of about 100 km at 60 s (probably against a hypothetical perfect simulation). The time step has to be reduced until a further reduction does not reduce

AHTD significantly in order to establish a reference simulation. (I see that Hoffmann et al. (2016) claim that convergence already was reached at 120 s, but this is in obvious contradiction with the results reported here.) This might change the apparent relative benefits of higher-order methods.

Fig. 8 (bottom) illustrates the convergence for the troposphere, where truncation errors are higher than elsewhere. In fact, the northern mid latitudes slow down the tropospheric convergence. We agree that the AHTD for this particular region suggests that a shorter time step of 30 s might be a better choice. However, the other regions and especially the combined set of all parcels show convergence already for a time step of 60 seconds. The convergence analysis of Hoffmann et al. (2016) is not applicable to this study for two reasons: First, the horizontal resolution was increased from 0.25° to 0.125° in this study, which reduces the convergence rate. Second, the simulations of volcanic emission dispersion by Hoffmann et al. (2016) covered only the UT/LS region, and the results cannot be generalized to the troposphere.

3. As the authors rightly point out, higher-order methods are unlikely to bring much gain if we use linear interpolation. This points to another option for a potentially optimal trajectory calculation, at least as a reference method: Linear interpolation should allow to solve the trajectory equation analytically within a grid cell and between two times of wind field availability (cf. Seibert, 1993). Admittedly, the need to bound each calculation step at grid-cell borders has a potential to make this method a bit cumbersome and computationally probably less efficient.

A reference simulation using linear interpolation would indeed be suited as reference. As our computations do not use such a method, this would also potentially allow for a more solid comparison. However, implementation is complicated due to the implied transformation of spherical and Cartesian coordinates and computation costs are relatively high, so we decided to keep the RK4 reference for the current study.

4. Another methodological issue is the questions on which transport times the final evaluation of schemes should be based. Even though not explicitly mentioned, Fig. 8 seems to be made with results after 10 days. I dont think this is the most appropriate choice. As discussed in Sect. 3.2, there is a strongly non-linear growth of the deviations with time. This growth has nothing to do with numerical errors, it is solely a function of atmospheric flow patterns (diffluent flows or bifurcations). Thus, a longer calculation mainly amplifies initial deviations which are due to the different truncation errors. The longer calculations only mean more calculational efforts, and the true truncation errors are obscured by the increasingly important atmospheric flow influences, probably exaggerating the difference between atmospheric regions or seasons (note also that for example polar-region trajectories mostly leave the polar domain within the 10 days). Please also look at results with much shorter transport times and consider replacing the 10-day results by them.

Our intention was to give estimates for the total uncertainties of trajectory calculations

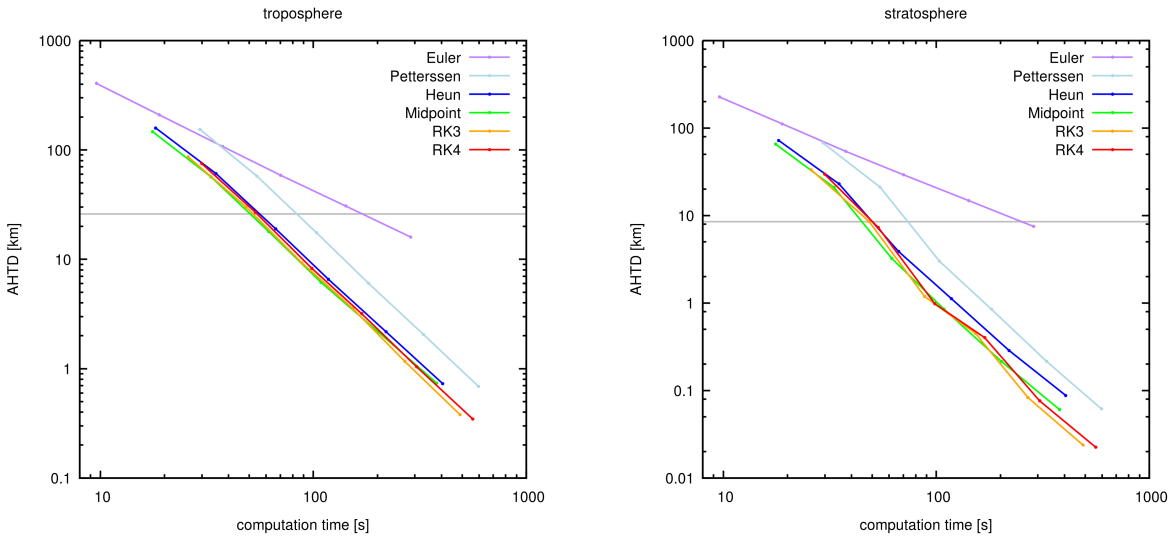

Figure 1: The figures show the efficiency of the used methods by relating the average error in specific altitudes after 24 h to the required computational time. The dots indicate the time steps from 3600 s on the left to 120 s on the right.

in different atmospheric conditions. We defined a limit for the spread of the parcels and wanted to find the cheapest method to adhere to the limit. However, to allow for distinction between initial truncation errors and those potentially perturbed by atmospheric influence, we added a Figure for the trade-off between computational accuracy and CPU-time after 1 day and discussed the results in Sect. 3.4 (see Figure 1 in this reply).

5. Finally, the results are certainly sensitive to the resolution of the wind field data. Results obtained for the specific case of 16 km / 3 h therefore cannot be generalized. Keeping in mind the conclusions of Stohl et al. (1995), Brioude et al. (2012), and Bowman et al. (2013), 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. This also diminishes the value of the results presented here, as most people would want to use the 1-h data if they go to the highest horizontal resolution. There would be a number of ways to produce more general results, such as trying out different resolutions or to parameterize the recommended time step by flow field properties such as (local) spatial and/or temporal derivatives at different orders.

Indeed, hourly operational forecast data can be downloaded from ECMWF since November 2011. However, the description of the operational products at `http://www.ecmwf.int/en/forecasts/datasets/set-i` implies 3-hourly forecast time steps for the first 144 hours of HRES operational forecasts. Similarly, the most recent (updated 2015) user guide to ECMWF forecast products available at `http://www.ecmwf.int/sites/default/files/User_Guide_V1.2_20151123.pdf` specifies the temporal retrieval of ECMWF forecasts as follows: 'All forecast parameters, both surface and upper air, based on 00 and 12 UTC

HRES and ENS, are available at 3-hourly intervals up to +144 hours and at 6-hourly intervals from +150 to +240 hours.‘ For the scope of this paper we decided to restrict ourselves to data with original and officially approved resolution only and therefore downloaded the operational forecasts with a forecast time step of three hours.

Specific and Minor Remarks

1. The title could be rephrased for example as ‘Truncation errors of trajectory calculations using ECMWF high-resolution data diagnosed with the MPTRAC Lagrangian particle dispersion model’

Following the suggestion we rephrased the title of the manuscript to: ‘Domain specific trajectory errors diagnosed with the MPTRAC advection module and ECMWF operational analyses’.

2. Page 1, line 1: Abstract. The abstract could be shortened by removing nonessential background and more concise wording.

The abstract has been shorted by removing some unnecessary or redundant background information.

3. Page 1, line 4: kinematic equation of motion (comes also in other places). I dont 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?

We think that the term ‘kinematic equation of motion’ is correctly used for Eq. (1). We do not intend to change the wording.

4. Page 2, line 6: Lagrangian particle dispersion models have proven. Under this chapeau, next to real LPDMs, LAGRANTO is listed which is a simple trajectory model and not an LPDM. I think it does no harm to enumerate it here, but not under a category that doesnt fit (and there is no reason to focus specifically on LPDMs here, as the truncation error problem occurs in the same way in trajectory models).

Our focus is on LPDMs, which is now also visible in the updated title of the study. Therefore we decided to skip the reference to LAGRANTO in the introduction but changed slightly in our conclusions: *Page 14, line 32: All integration methods discussed here are in principle suited and have been used for Lagrangian Particle dispersion and trajectory model simulations.*

5. Page 3, line 3: The T1279L137 ECMWF operational analysis data used here have 16 km effective horizontal resolution, about 180 - 750m vertical resolution at 2 - 32 km altitude, and are provided at 3 h synoptic time intervals. ‘Provided at 3 h ... ’ is not entirely correct - it is your choice. Analyses are available every 6 h and forecasts at steps of 1 h. It would be useful if you indicate what composite of AN and FC fields you were using here and not on the next page.

ECMWF analyses are produced every 6 hours, but forecasts are only calculated from

the analysis base times of 00 and 12 UTC. Thus we decided to use analyses at 00 and 12 UTC and the corresponding forecasts in between, as described on page 4, lines 5-6. To our opinion such detailed information does not belong to the introduction. Concerning the 1 h forecast steps, please refer to our reply to major remark #5.

6. Page 3, line 7: LPDM studies using this new data set. It is not clear what you mean by 'this new data set'. Obviously, ECMWF rules will not allow to make the ECMWF data set that you have used here available for general use.

We rephrased the sentence: *Page 3, lines 6-7: Using most recent meteorological data, the results will be of interest for many current and future LPDM studies using ECMWF operational data or data sets with comparable resolution.*

7. Page 3, line 26: meteorological wind fields. Just wind fields should good enough. If the model uses other fields as well (e.g, thermodynamic or surface fields), please explain in more detail. I am also wondering whether the model considers convection – it is invoked as a possible explanation later, but in Hoffmann et al. (2016) I did not find a reference to convection being a simulated process (if it isnt, it should also not be invoked). Maybe you want in general to provide a little bit more information about the model, especially considering that the only paper published so far is not open-access.

We rephrased this as suggested. Our model does not consider convection. In the text we refer to convection patterns visible in the meteorological input data. Note that more information on the MPTRAC model can also be found in Heng et al. (2016), which is referenced in our manuscript.

8. Page 3, line 31: While atmospheric reanalyses... typically have a horizontal resolution of ~ 100 km or less, the resolution of operational forecast products has been continuously improving during the last decades. Reanalysis products resolution has improved as well! And better write '$\approx$ 100 km' (\approx) or 'ca. 100 km' to not confuse with symbol for proportionality (symbols appear also on p. 7 and p. 10).

While it is true that also the resolution of global reanalyses has been improved over time, this has not been done as often as for the operational products. E.g., from ERA-15 (1996) to ERA-INTERIM (2011) the resolution of the ECMWF reanalyses has improved from 1.125° to 0.7° and from 31 to 60 vertical levels, while for the atmospheric operational analyses the resolution has improved from 0.56° to 0.14° and from 31 to 91 levels over the same time frame. Symbols for approximation have been changed throughout the text.

9. Page 4, line 4: For usage with MPTRAC, the wind fields have been interpolated horizontally to a longitude-latitude grid. Have they really been interpolated (from another, e.g. reduced Gaussian, grid), or were they just extracted at the given grid through MARS (by evaluation of the spectral data)?

Wind data on model levels have been directly extracted from MARS by indicating the desired horizontal resolution. The interpolation on pressure levels has been performed by using the model to pressure level interpolation operator ml2pl from the Climate Data

Operators (CDO).

10. Math vector notation: You are using upright bold letters for vectors. Standard notation would italic bold, accessible (with the amsmath package) for example through \boldsymbol{text}.

The notation has been changed accordingly.

11. Page 6, line 4 ff.: k1= ... It seems that you define certain velocities as k. It is very unusual to denote a velocity by k and not with a letter such as u or V, upper-or lowercase, and even more difficult as you dont give an explanation in words of these variables.

The vectors k_i are just auxiliary vectors at different nodes of the integration schemes, for which 'k' may be an acceptable choice of notation. The definitions of these vectors in Eqs. (7) to (15) make clear that wind vectors are meant. Calling the vectors u or v may cause confusion with the wind function that is already called $\boldsymbol{v}$. We kept this as is.

12. Page 8, line 5, 8: 5 latitude bands, 3 altitude layers. According to standard typesetting rules, numbers less or equal to twelve in running text should, in general, be written out (same for '2nd/3rd-order' elsewhere).

This has been fixed throughout the manuscript.

13. Page 11, line 30: land surface ratio. I guess that 'land-surface fraction' is meant.

This is correct. We changed the text accordingly.

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 distribution of continents and orography is relevant for the difference between the mid-latitudes of the two hemispheres, but not for differences between mid-latitudes and elsewhere - this latter effect is due to the structure of the global circulation which in the end is caused by the poleward increase of the Coriolis parameter, allowing for Rossby waves and baroclinic instability to occur there.

We would like to pick up the remarks of both reviewers to clarify our view on the errors occurring in the mid latitudes. In the original manuscript a hint to the meandering jet streams and the baroclinic structure of the atmosphere was missing, which is a important source of transport errors in our simulations for the mid latitudes. Text on page 11 has been changed as follows: *Page 11, lines 27-30: The troposphere has its largest errors at northern mid-latitudes with errors between 245 km and 470 km. Tropospheric mid-latitudes were expected to cause relatively large errors because of the nature of global circulation: Rossby waves and baroclinic instability occurring predominantly in this region come along with highly variable wind patterns. In addition, 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. The errors in the polar regions... Page 11, lines 32-34: The UT/LS region has its largest AHTDs in the northern mid-latitudes with 95 km to 177 km. These errors are caused by*

the north-south meandering of the jet (Woollings et al., 2014) and higher turbulence in the underlying region. The second largest...

15. Page 12, line 1: The south pole has the smallest errors. Probably you want to say that the smallest errors were found over Antarctica / the southern polar region.

We replaced 'South Pole' by 'Antarctica'.

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. Is turbulence due to convection resolved in MPTRAC? If not, it cant be invoked as an explanation here.

The term 'turbulence' was misleading in this context, we intended to refer to the grid-scale fluctuations that are given in the meteorological input data.

17. Page 12, line 9: During northern hemisphere wintertime land-sea temperature differences as well as the temperature gradient between the North pole and the equator are largest, which allows for more intense and complex dynamic patterns to occur than in summer. I would not refer to the meridional temperature gradient as the pole-equator temperature gradient – the pole is a single point and neither the pole nor the equator typically represent the locations of the extreme temperatures. Furthermore, the baroclinicity in mid-latitudes rather depends on the subtropical region temperatures than on equatorial ones.

On page 12, lines 8-10, we replaced 'North Pole' by 'Arctic' and 'Equator' by 'subtropical regions'.

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 dont need longer trajectories, but more frequent starts or more years. I any case, I dont 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.

We skipped the term 'seasonal truncation errors' from the title of the revised manuscript. Consequently, we changed the title of Sect. 3.3 to 'Regional and temporal truncation errors' in order to include both seasonal and intra-annual effects. The section on seasonal dependencies itself is already very short.

19. Page 12, line 27: Vertical transport deviations are about 800 - 1000 times smaller than the horizontal transport deviations. As the atmosphere in general is anisotropic (L $\approx$ 10,000 km, H $\approx$ 10 km), this is trivial and not worth mentioning.

We omitted this as suggested.

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?

See reply to major remark #1.

21. Page 13, line 15: As an example, Fig. 7 shows results of scaling tests using the midpoint scheme with a time step of 120 s for different numbers of particles and OpenMP threads. 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.

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.*

22. Page 13, line 18: the computing time is limited by an offset of ... s, which is due to the overhead of the OpenMP parallelization. Language-wise, I would prefer to speak about showing a plateau rather than 'being limited by an offset'. Do these times refer only to the time spent in the trajectory calculation, or to the model as a whole? In the latter case, there is not only overhead from parallelization but also from other parts of the model (the minor plateau even with a single a single thread seems to indicate some contribution.) One is also wondering here about your parallelization strategy – is there a barrier after each time step? Is that needed?

The time measurements refer only to the part of the code spent in the advection module of MPTRAC. Due to the operator splitting approach used by our model, an OpenMP barrier occurs after the call of each operator (or 'module' of MPTRAC) and after each time step. Future work may focus on 'pipelining' of the operators, but this would require a major revision of the structure of our model. We will replace the word 'offset' by 'constant contribution that can be attributed to the OpenMP parallelization overhead'.

23. Page 13, line 23: It is also found that the code provides additional speedup if the simultaneous multithreading capabilities of the compute nodes are used, in particular for very large numbers of particles (on the order of 10^6 to 10^7). For smaller number of particles (10^4 or less) the speedup is limited due to the overhead of the OpenMP parallelization and by the limited work load of the problem itself. This is an interesting part of your results, but I dont agree completely with your description and interpretation. There is always a drop at first when the number of threads exceeds the number of 24 cores, which is quite typical (see also the indications given in your footnote source). The interesting feature is that for a large enough number of particles, it then rises again. Maybe your computing specialists have more detailed insights for this behaviour? Also, I was wondering why for the largest number of particles the first maximum is reached with 20 threads. Is this a

plotting error, or is this related to memory access? We should also note some irregular behaviour for moderate numbers of particles toward the maximum number of threads.

We consulted the IT experts at our center to get more information. According to their analysis, limited scalability (or 'drops' in speed-up) can be assigned to load imbalances. Our model implicitly uses a 'static' schedule for the OpenMP loop parallelization. For instance, for 10^6 particles on 28 threads there will be 4 cores that have to process two packages of 36 k particles using hyper threading (HT) while the other 20 cores only process one package without HT. This implies a significant load imbalance compared to a more balanced scaling using 24 threads, which corresponds to the number of physical cores. Nevertheless, speedup results at 48 threads compared to 24 threads show that running with HT is 45% more efficient than without.

24. Page 13, line 33: Among the 2nd-order methods the Petterssen scheme has the lowest computational efficiency, which is due to the fact that we tuned the convergence criteria for this method for accuracy rather than speed. So, it is not 'the Petterssen scheme' but your implementation of this scheme for which the statement holds! That is a bit of a pity, so we dont know how the Petterssen scheme would do with a more reasonable cut-off of the iterations. As this is quite a relevant issue, and some people might only look at the figure without reading the full text, I suggest to mention that also in the figure caption (or better do some more realistic tests for a revised version).

The Petterssen scheme with many iterations did not give significantly more accurate results than the second order methods, which include Heun's method, which is equal to the Petterssen scheme with one iteration. Therefore no further analysis of intermediate configurations was made. However, we share your concern and added the following note to the caption of Fig. 8: *Note that our implementation of the Petterssen scheme was optimized for numerical accuracy rather than speed.*

25. Page 14, line 1: The best efficiency, i. e., the best accuracy at the lowest computational costs, is mostly obtained with the midpoint and RK3 methods. This wording is not providing an operational definition of 'best efficiency', as best accuracy and lowest computations cost are mutually exclusive and you are not defining how exactly you want to measure the efficiency. A suitable measure would be the computation time to achieve a given AHTD. Do this for a value that is reasonable and then quantify the computation times, as just reading them out from a log-log diagram is not so easy (note also the unexplained minor tick intervals better use a full set of them). Thus, you may want to combine this paragraph with the following one. For the rating of Petterssen (vs. midpoint), see above. Another question which needs to be answered is with how many threads this result was obtained, and whether there is any difference between schemes with respect to speed-up.

The most efficient method was detected as suggested by the reviewer, and this has been made more clear in a revision of this paragraph. The computation used 48 cores and the methods profit differently from the parallelization. Figure 2 in this reply shows the relative

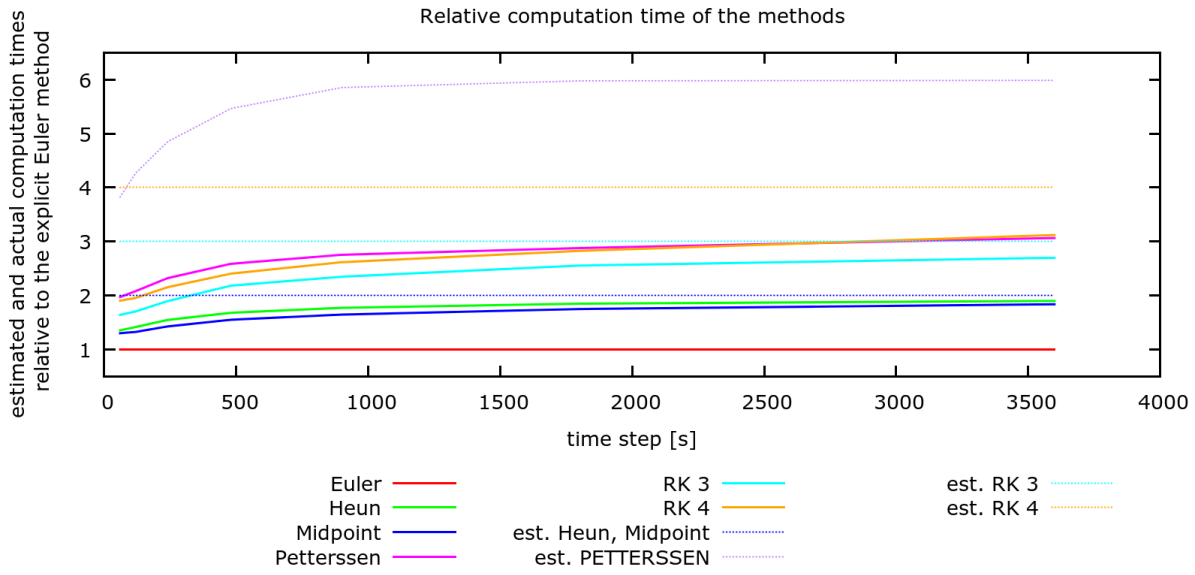

Figure 2: Relative computation time of the methods. The estimated times are based on the assumption that the time scales linearly with the number of calls to the wind interpolation function.

computation time of the methods in comparison to the Euler method. Theoretically there should be a linear dependency between computation time and the number of calls to the wind interpolation function (which is the most expensive part of the advection module). However, the higher order methods, which call to the wind interpolation more often, are faster than this estimate for the computational time. Note that the maximum number of iterations for the Petterssen scheme was six, which explains the plateau. We contribute the better speedup for more computationally expensive methods to cache usage, since the wind interpolation probably considers some grid points more than once, such that elements can be read from the cache instead from main memory. However, this makes the RK4 and Petterssen scheme even less attractive, since the computation time would be even larger without higher parallelization speedup.

26. Page 14, line 17: with an effective horizontal resolution of about 16 km. Mention also the 3 h here!

The information about the temporal resolution has been added accordingly.

27. Page 14, line 18: The truncation errors of the schemes were found to cluster into three groups that are related to the order of the method. Add ‘for a given time step’.

This has been added accordingly.

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!).

We omitted the wrong part of the explanation.

29. Page 14/15, line 336: [whole para]. I suggest to rephrase this paragraph in line with the remarks made above for Sect. 3.4, making sure it clearly conveys the relevant facts and definitions.

The paragraph has been rewritten taking into account your remarks #21 to #25.

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.

We adjusted the paragraph according to your comment and added the following: *Page 15, line 9: Their recommendation for a sufficiently small truncation error was 15% of the time step needed for convergence of the Petterssen scheme. Assuming that the reference accuracy has improved in the meantime, an even smaller value would result. The CFL criterion is used to make sure that no small-scale features are skipped.*

31. Page 15, line 19: However, the large variability of regional and seasonal truncation errors found here suggests that applications may benefit from more advanced numerical techniques. Adaptive quadrature could be an interesting topic for future research. Note that adaptive time steps have been recommended by Seibert (1993) and were used already in the 1980ies for atmospheric trajectories by Maryon and Heasman (1988) and Walmsley and Mailhot (1983).

We made a reference to the mentioned studies: *Page 15, lines 18-19: However, the large variability of regional and seasonal truncation errors found here suggests that applications may benefit from more advanced numerical techniques. Adaptive time stepping as recommended by Seibert (1993) was used already in the 1980s for atmospheric trajectories by Maryon and Heasman (1988) and Walmsley and Mailhot (1983). Such an adaptive quadrature could be taken up for future research.*

32. References: For Hoppe et al. (2014), quote the final paper and not the discussion version.

The reference has been updated in the final manuscript.

33. Figure 1: I would suggest to use the same scale for all pressure levels. I am wondering why odd pressure levels are used (32.6, 180, 488 hPa) instead of standard levels.

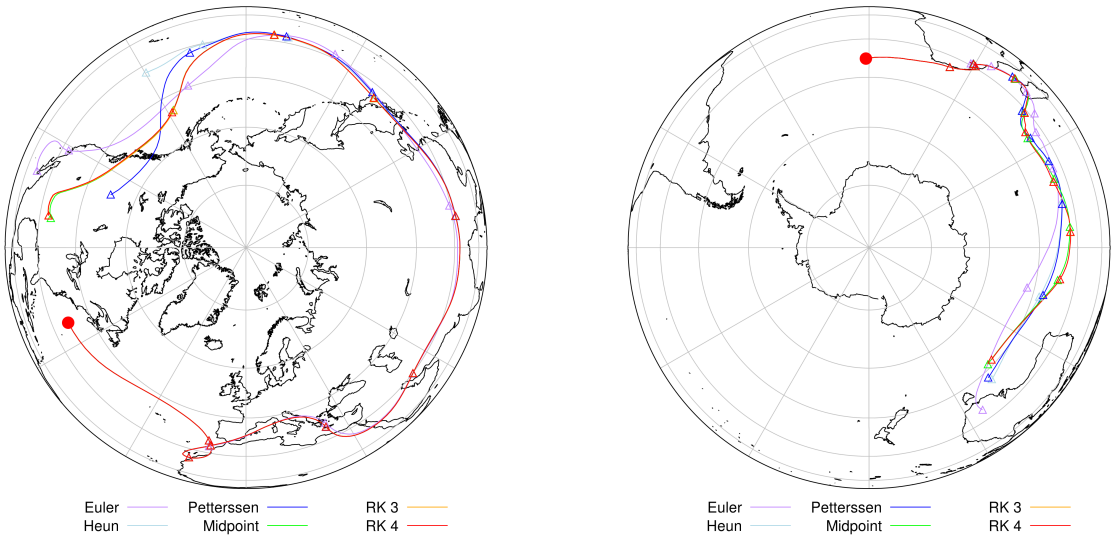

Figure 3: Examples of trajectory calculations using different numerical integration schemes. Circles mark the starting positions of the trajectories. Trajectories were launched at an altitude of 10.8 km (left) and 9.7 km (right). The starting time is 1 January 2014, 00:00 UTC in both examples. Triangles mark trajectory positions at 0 UTC on each day.

And I would suggest to reverse the colour coding for vertical velocity – meteorologists would find it more natural letting blue denote subsidence and red upward motion.

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. The chosen pressure levels are used in the model and correspond closely to the altitudes given in the figure caption. The colour coding for vertical velocity has been reversed following comments by reviewer #2.

34. Figure 2: I dont deem this figure necessary. If you want to keep it, use an appropriate viewing position in the projection for the Northern hemisphere, presently we are looking from a point located somewhere above the South pole, like peeking through the ground, not down from space! Also, use hollow symbols of different shapes so that we can easily recognize coincident positions as such.

We corrected the projection error and changed the symbols (see Figure 3).

35. Figures 3 ff.: It would help the reader if you annotate subfigures or at least columns of subfigures.

This will be done during copy-editing.

36. Figures 5 and 6: This figure should be simplified. You dont need to show the two years separately, and I think you also dont need to show seasons separately. Thus you could have just three subfigures (three levels) and the five regions inside of each one. Then

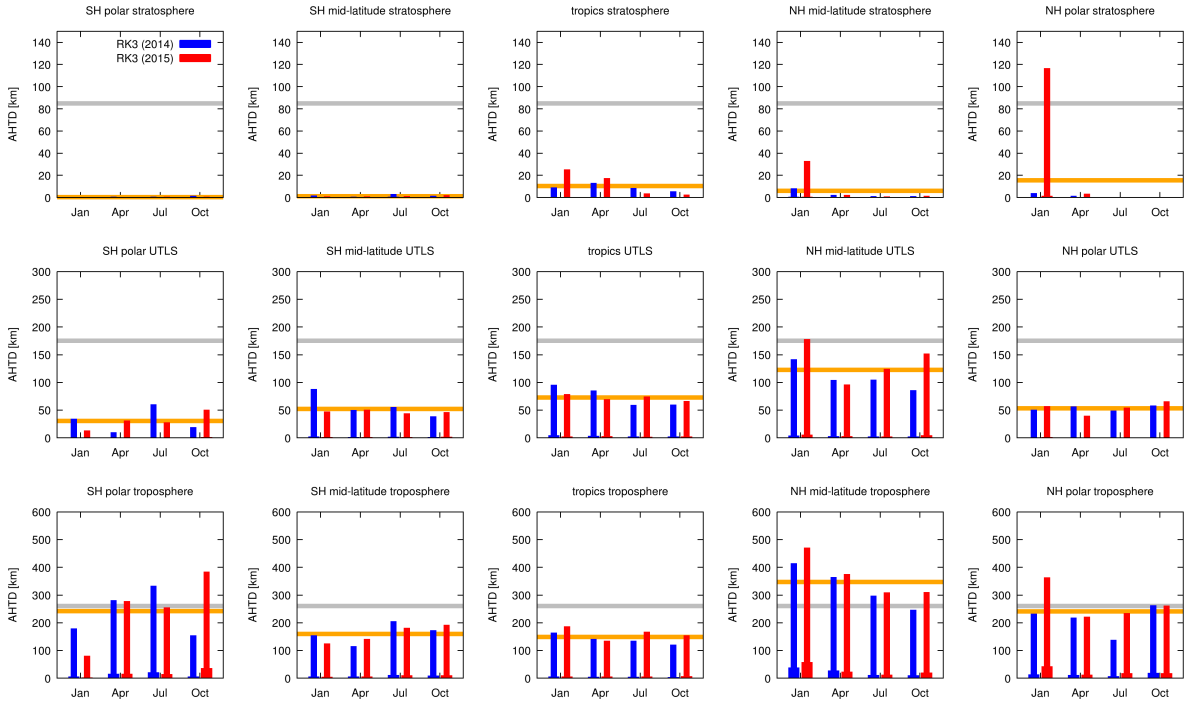

Figure 4: Average and median horizontal transport deviations after 10 days in different regions for the RK3 method. The orange horizontal line represents the average of the domain. The gray horizontal line indicates the error limit.

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. We added a horizontal line for the average error to all subfigures (see Figures 4 and 5). We did not use a log scale, because it would hide the seasonal and regional differences.

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.

38. Using an enlarged printout of the lower part of Fig. 7, I tried to figure out the number of cores which works fastest as a function of the number of particles. I arrived at something like this:

#particles #threads remark

<50 1

50 - 200 4

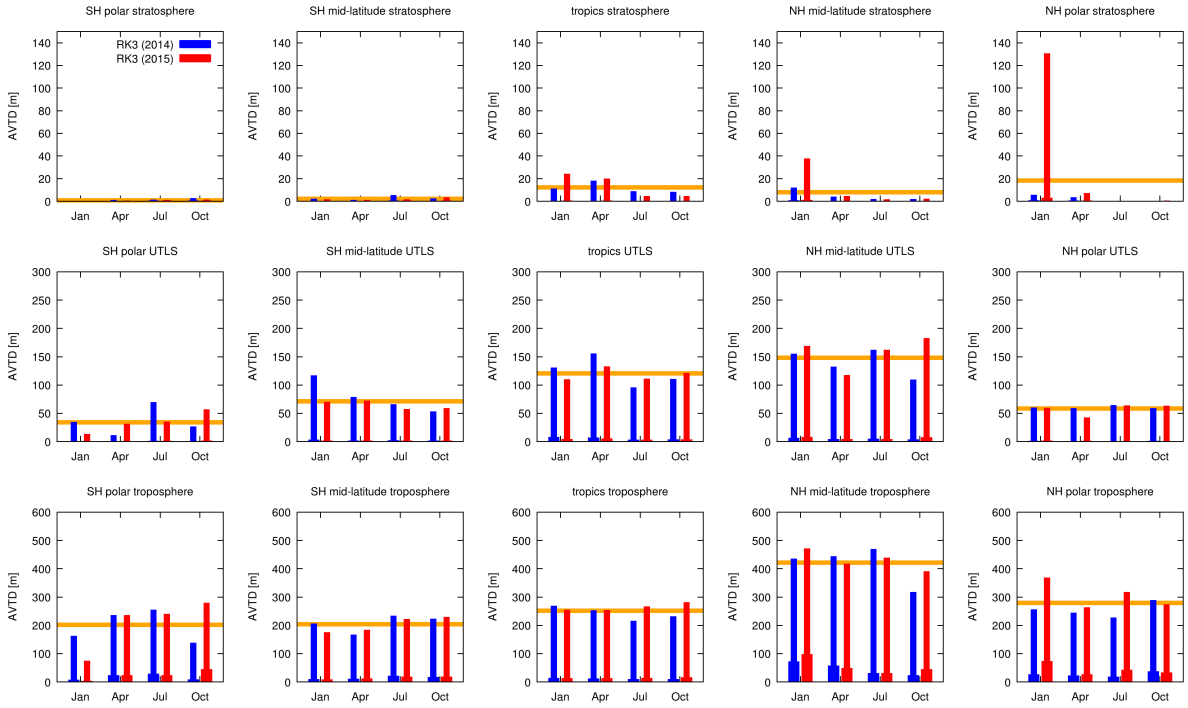

Figure 5: Average and median vertical transport deviations after 10 days in different regions for the RK3 method. The orange horizontal line represents the average of the domain.

200 - 300 8 very small interval!

300 - 1000 16

1,000 - 50,000 24 = #cores!

> 50,000 48 = max. #threads

I think that this evaluation would be useful for users. What is really striking is the fact that only (integer) powers of two show up as recommendable number of threads until 16. Then we can add 8 to arrive at the maximum number of cores (the question is open whether on a 32-core machine, 24 would show up or not), and then we can double once with hyper-threading. This is really a lesson for users, and if you have IT colleagues who are able to relate this behaviour to the hardware layout of your nodes, it would be even more useful.

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. Unfortunately, our IT experts were not able to provide a simple explanation of how the number of threads is linked to the hardware layout. The findings may depend specifically on the computing architecture and should not be generalized too much.

39. Page 15, line 20: Code and data availability. - ECMWF data (of the kind used here) are not simply 'distributed' by the centre. In general they would be available only for member-state NMS (or institutions authorized by them) and special-project holders. I suggest that the limited availability of these data is indicated. (I also thought that data provision could be mentioned in the acknowledgements.) - It would be useful to indicate the availability of the preprocessor which transforms ECMWF data to MPTRAC input data. - Does the version of the MPTRAC code available on github include the variety of integration schemes used here? If not, please make a statement about their availability. - It would be useful to provide the starting points of the trajectories as supplementary material so that the calculations become more reproducible.

There are several options to obtain ECMWF operational data, all of them are described in http://www.ecmwf.int/en/forecasts/accessing-forecasts. We crated a separate repository containing the MPTRAC code for the various integration schemes as well as the starting points of the trajectories. Section 5 has been changed as follows: *Page 15, lines 18-19: Operational analyses and forecasts can be obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF), see http://www.ecmwf.int/en/forecasts (last access: 3 May 2017) for further details on data availability and restrictions. ECMWF data have been processed for usage with MPTRAC by means of the Climate Data Operators (CDO, https://code.zmaw.de/projects/cdo, last access: 3 May 2017). The version of the MPTRAC model that was used for this study along with the model initializations is available under the terms and conditions of the GNU General Public License, Version 3 from the repository at https://github.com/slcs-jsc/mptrac-advect (last access: 3 May 2017).*

Reviewer #2

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 ECMWFs 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.

Usual simulations with MPTRAC follow your approach and many parcels are randomly distributed around a starting point. Alternatively, in this study many different starting points are used, such that the impact of the average atmospheric conditions of the domains

on the error can be estimated. The analysis of the error would become very costly, if groups of parcels were created for each starting point. We tried to show that deviations of individual trajectories are not very meaningful by additionally computing the median deviation of the parcels. However, our impression is that the AHTD/AVTD metric is the common method for trajectory evaluation and we wanted to make the results comparable to existing studies. Also, with this setup , we wanted to reach convergence for the trajectories to compare the errors of different methods.

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.

We would like to thank the reviewer for bringing up this interesting starting point for further research on trajectory accuracy in context of inter-annual variability of tropospheric and stratospheric atmospheric flow. Indeed, we did not expect such large variations between two NH winters in the first place and it would be intriguing to extend such a study to multiple years once input data at sufficient and constant resolution will be available.

Minor comments:

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

Text has been changed accordingly.

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

Text has been rephrased.

- 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?

We tried to make this point more clear by saying 'Stratospheric wind speeds in the extratropical summer hemisphere are generally slow compared to the winter hemisphere.'

- P6,L27: "we calculated the horizontal distances as Cartesian distances of the air parcel

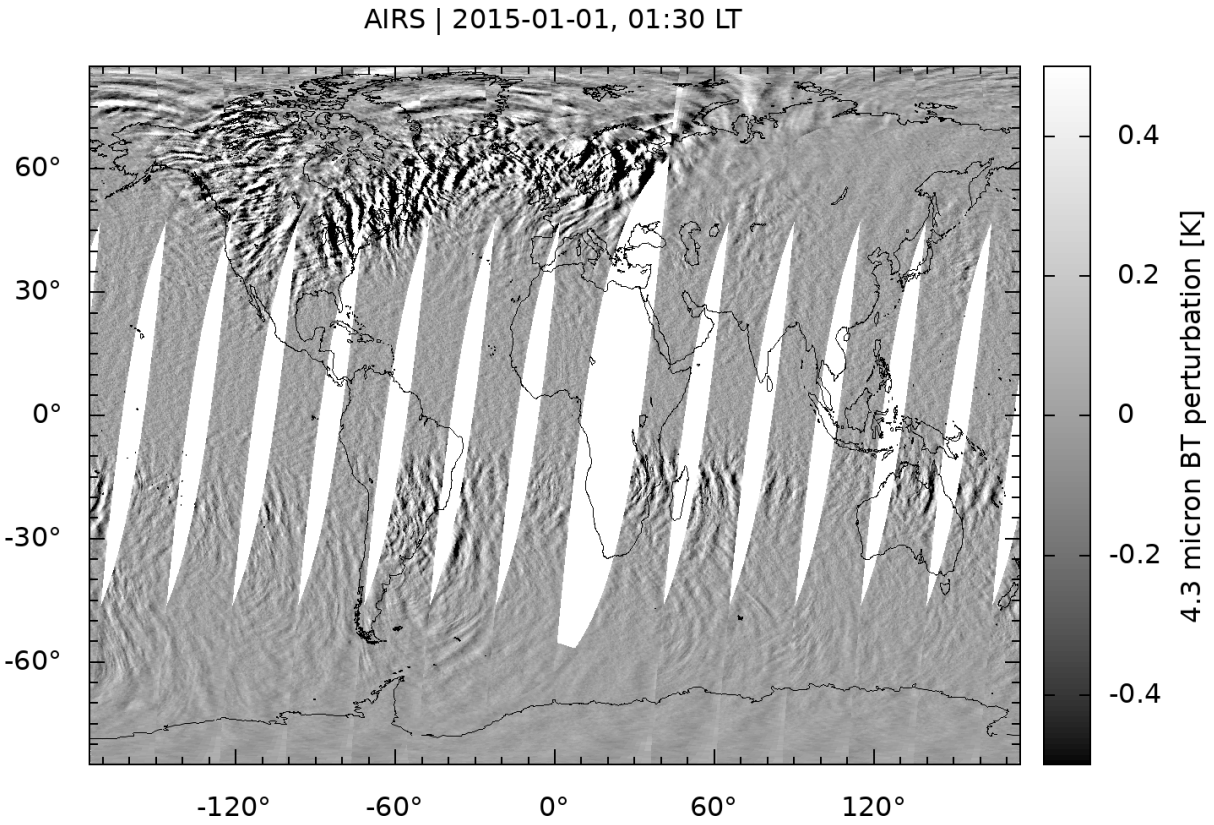

Figure 6: Atmospheric InfraRed Sounder (AIRS/Aqua) satellite observations of stratospheric gravity waves (following Hoffmann et al., 2013).

positions projected to the Earth surface" → It is not completely clear how the distance is calculated. What are Cartesian distances on a sphere?

To clarify we rephrased this: 'To calculate the horizontal distances we converted the spherical coordinates of the air parcels to Cartesian coordinates and calculated the Euclidean distance of the Cartesian coordinates.' Note that this approach approximates spherical distances quite well, as long as those distance are smaller than about 3000 km.

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

We rephrased accordingly.

- 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 artifact? The amplitude of the waves is rather small, as expected in the stratosphere.

Although we can not exclude that some numerical artifacts of the ECMWF IFS model are present in the vertical velocity map, there is evidence that the wave structures are

physical, because they occur in the same places where an infrared nadir sounder observed stratospheric gravity waves. See Figure 6 in this reply.

- 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.

The intention of our Figures 5 and 6 is to give a comprehensive impression of spatial and temporal variability of transport deviations on the global scale. Although the altitude classification namely in the UT/LS region does not exactly reflect the real tropospheric and stratospheric fractions and their hemispheric variations, it still shows substantial differences between the hemispheres.

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

We added a new table (Table 2) which comprises the values originally given on page 11, lines 1-8.

- 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.

We decided to rephrase the whole paragraph and would like to refer to our reply to minor remark #14 of reviewer #1. 'Land surface ratio' has been changed to 'land-sea ratio'.

- 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.

Please see our reply to minor remark #14 of reviewer #1.

- 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.

We share your concern. Although our study used two years to give an initial indication of inter-annual variability, we would rather speak of differences between the two test years

instead of an inter-annual variability. We changed the text accordingly.

Executive editor comment

Dear authors,

in my role as Executive editor of GMD, I would like to bring to your attention our Editorial version 1.1:

http://www.geosci-model-dev.net/8/3487/2015/gmd-8-3487-2015.html

This highlights some requirements of papers published in GMD, which is also available on the GMD website in the 'Manuscript Types' section:

http://www.geoscientific-model-development.net/submission/manuscript_types.html

In particular, please note that for your paper, the following requirements have not been met in the Discussions paper:

- The main paper must give the model name and version number (or other unique identifier) in the title.
- If the model development relates to a single model then the model name and the version number must be included in the title of the paper. If the main intention of an article is to make a general (i.e. model independent) statement about the usefulness of a new development, but the usefulness is shown with the help of one specific model, the model name and version number must be stated in the title. The title could have a form such as, "Title outlining amazing generic advance: a case study with Model XXX (version Y)".

So please add the model name and/or its acronym (MPTRAC) and its respective version number in the title of your article in your revised submission to GMD.

Yours, Astrid Kerkweg

We rephrased the title of the manuscript according to suggestions made by reviewer #1. The name of the model, MPTRAC, was included. Unfortunately, a specific version number was not assigned for the code used here. However, to allow others to reproduce our results, we made the code available in a separate repository, as described in the the revised section on 'code and data availability' in our manuscript.

References

Heng, Y., Hoffmann, L., Griessbach, S., Rößler, T., and Stein, O.: Inverse transport modeling of volcanic sulfur dioxide emissions using large-scale simulations, Geosci. Model Dev., 9, 1627–1645, 2016.

Hoffmann, L., Xue, X., and Alexander, M. J.: A global view of stratospheric gravity wave hotspots located with Atmospheric Infrared Sounder observations, J. Geophys. Res., 118, 416–434, 2013.

Hoffmann, L., Rößler, T., Griessbach, S., Heng, Y., and Stein, O.: Lagrangian transport simulations of volcanic sulfur dioxide emissions: impact of meteorological data products, J. Geophys. Res., doi: 10.1002/2015JD023749, 2016.

Maryon, R. and Heasman, C.: The accuracy of plume trajectories forecast using the UK Meteorological Office operational forecasting models and their sensitivity to calculation schemes, Atmos. Environment, 22, 259–272, 1988.

Seibert, P.: Convergence and accuracy of numerical methods for trajectory calculations, J. Appl. Met., 32, 558–566, 1993.

Walmsley, J. L. and Mailhot, J.: On the numerical accuracy of trajectory models for long-range transport of atmospheric pollutants, Atmosphere-Ocean, 21, 14–39, 1983.

Woollings, T., Czuchnicki, C., and Franzke, C.: Twentieth century North Atlantic jet variability., Quart. J. Roy. Meteorol. Soc., 140, 783791, doi: 10.1002/qj.2197, 2014.

~~Regional and seasonal truncation errors of~~ Domain specific trajectory ~~calculations using~~ errors diagnosed with the MPTRAC advection module and ECMWF ~~high-resolution~~ operational analyses~~and forecasts~~

Thomas Rößler[1,2], Olaf Stein[1], Yi Heng[3], Paul Baumeister[1], and Lars Hoffmann[1]

[1]Forschungszentrum Jülich, Jülich Supercomputing Centre, Jülich, Germany
[2]now at: Department of Mathematics, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin, United States
[3]School of Chemical Engineering and Technology, Sun Yat-sen University, Guangzhou, China

Correspondence to: Thomas Rößler (t.roessler@fz-juelich.de)

Abstract. The accuracy of trajectory calculations performed by Lagrangian particle dispersion models (LPDMs) ~~are indispensable tools to study atmospheric transport processes. The accuracy of trajectory calculations, which form an essential part of LPDM simulations,~~ depends on various factors. ~~Here we focus on truncation errors that originate from the use~~ The optimization of numerical integration schemes used to solve the ~~kinematic equation of motion. The optimization of numerical integration schemes to minimize truncation errors and to maximize computational speed is of great interest regarding~~ trajectory equation helps to maximize the computational efficiency of large-scale LPDM simulations. ~~In this study we~~ We analyzed truncation errors and total errors of six explicit integration schemes of the Runge Kutta family, which we implemented in the Massive-Parallel Trajectory Calculations (MPTRAC) ~~model~~advection module. The simulations were driven by wind fields ~~of the latest~~ from operational analysis and forecasts of the European Centre for Medium-range Weather Forecasts (ECMWF) at T1279L137 spatial resolution and 3 h temporal sampling. We defined separate test cases for 15 distinct domains of the atmosphere, covering the polar regions, the mid-latitudes, and the tropics in the free troposphere, in the upper troposphere and lower stratosphere (UT/LS) region, and ~~in the lower and~~ the mid stratosphere. ~~For each domain we performed simulations for~~ In total more than 5000 different transport simulations were performed, covering the months of January, April, July, and October for the years of 2014 and 2015. ~~In total more than 5000 different transport simulations were performed.~~ We quantified the accuracy of the trajectories by calculating transport deviations with respect to reference simulations using a ~~4th-order~~ fourth-order Runge-Kutta integration scheme with a sufficiently fine time step. We assessed the transport deviations with respect to error limits based on turbulent diffusion. Independent of the numerical scheme, the ~~truncation~~ total errors vary significantly between the different domains~~and seasons. Especially the differences in altitude stand out~~. Horizontal transport deviations in the stratosphere are typically an order of magnitude smaller compared with the free troposphere. We found that the truncation errors of the six numerical schemes fall into three distinct groups, which mostly depend on the numerical order of the scheme. Schemes of the same order differ little in accuracy, but some methods need less computational time, which gives them an advantage in efficiency. The selection of the integration scheme and the appropriate time step should possibly take into account the typical altitude ranges as well as the total length of the simulations to achieve the most efficient simulations. However, trying to

~~generalize~~summarize, we recommend the ~~3rd-order Runge Kutta~~ third-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~~ ten days simulation time for the specific ECMWF high-resolution ~~meteorological data~~ data set considered in this study.

1 Introduction

Lagrangian particle dispersion models (LPDMs) have proven to be useful for understanding the properties of atmospheric flows, particularly for problems related to transport, dispersion, and mixing of tracers and other atmospheric properties (e. g. Lin et al., 2012; Bowman et al., 2013). Commonly used LPDMs include the Flexible Particle (FLEXPART) model (Stohl et al., 2005), the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Draxler and Hess, 1998), the ~~Lagrangian Analysis Tool (LAGRANTO) (Wernli and Davies, 1997; Sprenger and Wernli, 2015), the~~ Numerical Atmospheric-dispersion Modelling Environment (NAME) (Jones et al., 2007), and the Stochastic Time-Inverted Lagrangian Transport (STILT) model (Lin et al., 2003). While all these models are applied to solve similar tasks, they differ in specific choices such as the numerical methods or vertical coordinates that are used. In this study we apply the rather new model Massive-Parallel Trajectory Calculations (MPTRAC) (Hoffmann et al., 2016), which was recently developed at the Jülich Supercomputing Centre, Germany. MPTRAC was primarily designed to conduct trajectory calculations for large-scale simulations on massive-parallel computing architectures. Computational efficiency is an important aspect during the development of such a model.

LPDMs simulate transport and diffusion of atmospheric tracers based on trajectory calculations for many air parcels that move with the fluid flow in the atmosphere. The accuracy of these calculations has been the subject of numerous studies (e. g., Kuo et al., 1985; Rolph and Draxler, 1990; Seibert, 1993; Stohl et al., 1995; Stohl and Seibert, 1998; Stohl et al., 2001; Davis and Dacre, 2009). According to reviews of Stohl (1998) and Bowman et al. (2013), trajectory calculations have errors that arise from three sources: (i) errors in the gridded winds themselves, which could result from measurement error that enter the analyzed fields through the data assimilation process or from Eulerian model approximations, such as subgrid-scale parameterizations; (ii) sampling errors that follow from the fact that velocity fields are available only at finite spatial and temporal resolution and must be interpolated to particle locations; and (iii) truncation errors that originate from the use of an approximate numerical scheme to integrate the kinematic equation of motion in time. Bowman et al. (2013) point out that (i) and (ii) are usually the limiting factors for the accuracy of trajectory calculations, whereas high numerical accuracy and significant reduction of truncation errors can be achieved by reducing the size of the time step of the numerical integration scheme. The size of the time step is usually the most important factor that controls the trade-off between numerical accuracy and computation time. However, ~~it needs to be stressed that~~ the appropriate selection of the numerical scheme and optimization of the size of the time step is still mandatory to maximize computational efficiency. This is particular important for large-scale simulations, like Lagrangian transport simulations aiming at emission estimation by means of inverse modeling (e. g. Stohl et al., 2011; Heng et al., 2016) or long-term simulations coupled to chemistry climate models ~~(e. g. ?)~~(e. g. Hoppe et al., 2014).

In the following we present an assessment of six numerical integration schemes, all belonging to the class of explicit Runge-Kutta methods (e. g., Press et al., 2002; Butcher, 2008), for atmospheric trajectory calculations. Seibert (1993) studied the

truncation errors of some of these schemes based on analytic flow types such as purely rotational flow, purely deformational flow, wave flow, and accelerated deformational flow. Here we decided to focus on tests with realistic wind fields obtained from high-resolution operational analyses and forecasts provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). The T1279L137 ECMWF operational analysis data used here have 16 km effective horizontal resolution, about 180 – 750 m vertical resolution at 2 – 32 km altitude, and are nominally provided at 3 h synoptic time intervals. We estimated the total simulation errors and the truncation errors of the numerical methods for ~~5~~ five latitude bands and ~~3~~ three altitude ranges of the atmosphere, covering the free troposphere, the upper troposphere and lower stratosphere (UT/LS) region, and the mid stratosphere. ~~We studied the seasonal and inter-annual variability of the truncation errors~~ The simulations were used to study the seasonal error variability for the years 2014 and 2015. We systematically assessed trade-offs between accuracy and computation time to infer the computational efficiency of the integration methods. ~~The results of our study will be transferable to most other currently used LPDMs.~~ Using most recent meteorological data, the results will be of interest for many current and future LPDM studies using ~~this new data set~~ECMWF operational data or other meteorological data sets with comparable resolution.

In Sect. 2 we present the Lagrangian particle dispersion model MPTRAC together with an overview on the meteorological data. The selected numerical integration schemes and the diagnostic variables are introduced and the experimental set-up is described. Section 3 shows transport deviations from case studies followed by a general analysis of the error behavior in terms of ~~seasonal and regional~~ error growth rates and domain specific characteristics. Scalability and performance on a high-performance computing system are discussed. In Sect. 4 we conclude with suggestions for best-suited integration schemes and optimal time step choice in order to achieve most effective simulations of large-scale problems on current high-performance computing systems.

2 Methods and Data

2.1 Lagrangian particle dispersion model

In this study we apply the Lagrangian particle dispersion model MPTRAC (Hoffmann et al., 2016) to conduct trajectory calculations. MPTRAC has been developed to support the analysis of atmospheric transport processes in the free troposphere and stratosphere. In recent studies it has been used to perform transport simulations for volcanic eruptions and to reconstruct time- and height-resolved emission rates for these events ~~(Heng et al., 2016; Hoffmann et al., 2016)~~(Heng et al., 2016; Hoffmann et al., 2016; Wu
The primary task of MPTRAC is to solve the kinematic equation of motion for atmospheric air parcels. ~~It~~ The 'advection module' calculates air parcel trajectories based on given ~~meteorological~~ wind fields. ~~Turbulent~~ In another module turbulent diffusion and subgrid-scale wind fluctuations are simulated by adding stochastic perturbations to the trajectories, following the approach of the FLEXPART model (Stohl et al., 2005). Additional ~~but unused~~ modules can simulate the sedimentation of air parcels and the decay of particle mass. ~~The model~~ For this study only the advection module was activated. MPTRAC is particularly suited for ensemble simulations on supercomputers due to its efficient Message Passing Interface (MPI) / Open Multi-Processing (OpenMP) hybrid parallelization.

2.2 ECMWF operational analysis

Air parcel transport in MPTRAC is driven by ~~meteorological~~ given wind fields. In principle any gridded data produced by general circulation models, atmospheric reanalyses, or operational analyses and forecasts can be used for this purpose. Reanalyses and forecasts benefit from well-established meteorological data assimilation methods (Rabier et al., 2000; Buizza et al., 2005) which help to better constrain the modelled circulation fields to reality. While atmospheric reanalyses (e. g., Kalnay et al., 1996; Dee et al., 2011; Rienecker et al., 2011) typically have a horizontal resolution of ~~~ 100~~≈ 100 km or less, the resolution of operational forecast products has been continuously improving during the last decades. In this study we use horizontal and vertical winds from European Centre for Medium-range Weather Forecasts (ECMWF) operational analyses and forecasts[1] for the years 2014 and 2015 produced in spectral truncation T1279, which corresponds to a horizontal resolution of about 16 km. Vertically, the data consists of 137 levels reaching from the surface to 0.01 hPa. For usage with MPTRAC, the wind fields have been interpolated horizontally to a longitude-latitude grid with $0.125° \times 0.125°$ resolution and vertically to 114 pressure levels in the troposphere and stratosphere up to 5 hPa. 12-hourly analyses are combined with short-term forecasts in between to obtain data with a 3-hour time step. Hoffmann et al. (2016) showed that this data set outperforms existing reanalysis data products in terms of transport deviations for simulations of volcanic sulfur dioxide emissions in the upper troposphere and stratosphere.

Example wind fields from the operational data are presented in Figure 1. Horizontal and vertical wind velocities from the ECMWF operational analysis for 1 January 2015, 00:00 UTC are shown for three pressure levels in the stratosphere, in the UT/LS region, and in the free troposphere. At about 24 km altitude the global wind fields are dominated by a meandering band of high horizontal wind speed at high northern latitudes indicating the wintertime polar vortex, together with weaker tropical easterlies. ~~Wind dynamics~~ Stratospheric wind speeds in the extratropical summer hemisphere are generally slow compared to the winter hemisphere. Enhanced horizontal wind speeds at about 12 km altitude are connected with UT/LS jet streams over both hemispheres and are highest for the subtropical jet stream situated at around 30°N with maxima over the western Pacific reaching more than 100 m s^{-1} locally. In the free troposphere typical weather patterns from the moving high and low pressure systems over the mid latitudes exhibit the highest horizontal wind speeds, but with stronger spatial variability than in the stratosphere. The vertical wind velocities mostly vary on short spatial scales of several 100 km or less, often associated with atmospheric gravity waves (e. g. Preusse et al., 2009; Hoffmann et al., 2013). In the troposphere, also contiguous areas of high vertical velocities with extension of 1000 km or more occur close to strong pressure systems. Other high vertical wind speeds are connected with the polar vortex and the jet streams. Strong vertical winds are also observed at the Inter-Tropical Convergence Zone (ITCZ) which is located around 10°N-20°S for January. Note that many of the small-scale features identified here cannot be found in lower resolution data sets such as global meteorological reanalyses.

[1] See http://www.ecmwf.int/en/forecasts/datasets (last access: 8 December 2016).

2.3 Numerical methods for trajectory calculations

Lagrangian particle dispersion models calculate the trajectories of individual particles or infinitesimally small air parcels over time. The trajectory of each air parcel is defined by the kinematic equation of motion,

$$\frac{d\mathbf{x}}{dt}\ \frac{d\boldsymbol{x}}{dt} = v\left(x(t), t\right). \tag{1}$$

Here ~~$\mathbf{x} = (x, y, z)$~~ $\boldsymbol{x} = (x, y, z)$ denotes the position and ~~$\mathbf{v} = (u, v, \omega)$~~ $\boldsymbol{v} = (u, v, \omega)$ the velocity of the air parcel at time t. In MPTRAC the horizontal position (x, y) of the air parcel is defined by longitude and latitude, which requires spherical coordinate transformations to relate it to the horizontal wind (u, v). The vertical coordinate z is related to pressure p by the hydrostatic equation, and the vertical velocity is given by $\omega = dp/dt$. The wind vector ~~$\mathbf{v}$~~ $\boldsymbol{v}$ at any position ~~$\mathbf{x}$~~ $\boldsymbol{x}$ is obtained by means of a 4-D linear interpolation of the meteorological data, which is a common approach in many LPDMs (Bowman et al., 2013). The analytic solution of the kinematic equation of motion is given by

$$x(t_1) = x_0 + \int_{t_0}^{t_1} v\left(x(t), t\right) dt, \tag{2}$$

with initial position ~~$\mathbf{x}_0$~~ $\boldsymbol{x}_0$ at start time t_0 and end time t_1. In this study the performance of six numerical schemes to solve the kinematic equation of motion is analyzed. All schemes belong to the class of explicit Runge-Kutta methods, for an overview of these methods see, e. g., Butcher (2008).

The explicit Euler method likely poses the most simple way to solve the kinematic equation of motion. The numerical solution is obtained from Equation (2) by means of a ~~1st-order~~ first-order Taylor series approximation. Hence, it is also referred to as 'zero acceleration' scheme. The iteration scheme of the explicit Euler method (referred to as the Euler method below) is given by

$$x_{n+1} = x_n + \Delta t\, v\left(x_n, t_n\right), \tag{3}$$

where $\Delta t = t_{n+1} - t_n$ refers to the time step. The Euler method is a ~~1st-order~~ first-order Runge-Kutta method, i. e., the local truncation error for each time step is on the order of $\mathcal{O}(\Delta t^2)$, whereas the total accumulated error at any given time is on the order of $\mathcal{O}(\Delta t)$.

MPTRAC currently uses the explicit midpoint method as its default numerical integration scheme,

$$x_{n+1} = x_n + \Delta t\, v\left(x_n + \frac{\Delta t}{2} v\left(t_n, x_n\right), t_n + \frac{\Delta t}{2}\right). \tag{4}$$

First the 'mid point' is calculated using an Euler step with half the time step, $\Delta t/2$. The final step is calculated using the wind vector at the mid point of the Euler step. The midpoint method is a ~~2nd-order~~ second-order Runge-Kutta method. The local truncation error is on the order of $\mathcal{O}(\Delta t^3)$, giving a total accumulated or global error on the order of $\mathcal{O}(\Delta t^2)$. The method is computationally more expensive than the Euler method, but errors generally decrease faster in the limit $\Delta t \to 0$.

The scheme of Petterssen (1940) is popular in many LPDMs (e. g. Stohl, 1998; Bowman et al., 2013). It is defined by

$$x_{n+1,0} = x_n + \Delta t\, v\left(x_n, t_n\right), \tag{5}$$

$$x_{n+1,l} = x_n + \frac{\Delta t}{2}\left(v\left(x_n, t_n\right) + v\left(x_{n+1,l-1}, t_{n+1}\right)\right), \tag{6}$$

with l being an index counting the number of inner iterations carried out as part of each time step. If no inner iterations are performed, the scheme is equivalent to the Euler method. If one inner iteration is carried out, the method is also known as Heun's method, another type of a ~~2nd-order~~ second-order explicit Runge-Kutta method. An increasing number of inner iterations can help to improve the accuracy of the solution in situations with rather complex wind fields. If the local wind field is smooth, it results in fewer iterations and less computing time. We applied the Petterssen scheme with up to ~~7~~ seven inner iterations and did not tune the convergence limit for the inner iterations for efficiency, as we were mostly interested in good accuracy of the solutions.

In this study we also evaluated ~~3rd- and 4th-order~~ specific third- and fourth-order explicit Runge-Kutta methods (RK3 and RK4). The ~~3rd-order~~ third-order method used here is defined by

$$x_{n+1} = x_n + \Delta t\left(\frac{1}{6}k_1 + \frac{4}{6}k_2 + \frac{1}{6}k_3\right), \tag{7}$$

$$k_1 = v\left(x_n, t_n\right), \tag{8}$$

$$k_2 = v\left(x_n + \frac{\Delta t}{2}k_1, t_n + \frac{\Delta t}{2}\right), \tag{9}$$

$$k_3 = v\left(x_n - \Delta t\, k_1 + 2\,\Delta t\, k_2, t_n + \Delta t\right). \tag{10}$$

The classical ~~4th-order~~ fourth-order Runge-Kutta method is defined by

$$x_{n+1} = x_n + \Delta t\left(\frac{1}{6}k_1 + \frac{2}{6}k_2 + \frac{2}{6}k_3 + \frac{1}{6}k_4\right), \tag{11}$$

$$k_1 = v\left(x_n, t_n\right), \tag{12}$$

$$k_2 = v\left(x_n + \frac{\Delta t}{2}k_1, t_n + \frac{\Delta t}{2}\right), \tag{13}$$

$$k_3 = v\left(x_n + \frac{\Delta t}{2}k_2, t_n + \frac{\Delta t}{2}\right), \tag{14}$$

$$k_4 = v\left(x_n + \Delta t\, k_3, t_n + \Delta t\right). \tag{15}$$

For these methods the local truncation error is on the order of $\mathcal{O}(\Delta t^{p+1})$, while the total accumulated error is on the order of $\mathcal{O}(\Delta t^{p})$, with p referring to the order of the method. The classical ~~4th-order~~ fourth-order Runge-Kutta method is the highest order Runge-Kutta method for which the number of function calls matches its order. It typically provides a good ratio of accuracy and computation time. Any ~~5th-order~~ fifth-order method requires at least six function calls, which causes more overhead.

2.4 Evaluation of trajectory calculations

A common way to compare sets of test and reference trajectories is to calculate transport deviations (Kuo et al., 1985; Stohl et al., 1995; Stohl, 1998). Transport deviations are calculated by averaging the individual distances of corresponding air parcels from the test and reference data sets at a given time. The reference data set could be the known analytical solution for an idealized test case, it could be based on observations like balloon trajectories, or it could be obtained by using a numerical integration method known to be highly accurate for real wind data. Absolute horizontal and vertical transport deviations at time t are calculated according to

$$\mathrm{AHTD}(t) = \frac{1}{N}\sum_{i=1}^{N}\sqrt{[X_i(t)-x_i(t)]^2+[Y_i(t)-y_i(t)]^2}, \tag{16}$$

$$\mathrm{AVTD}(t) = \frac{1}{N}\sum_{i=1}^{N}|Z_i(t)-z_i(t)|. \tag{17}$$

with X_i(t), Y_i(t), and Z_i(t) as well as x_i(t), y_i(t), and z_i(t) referring to the air parcel coordinates of the test and reference data set, respectively. Each data set contains N air parcels. ~~Here we calculated~~ To calculate the horizontal distances ~~as Cartesian distances~~ we first converted the spherical coordinates of the air ~~parcel positions projected to the Earth surface~~parcels to Cartesian coordinates and then calculated the Euclidean distance of the Cartesian coordinates. This approach approximates spherical distances with $\geq 99\%$ accuracy for distances up to 3000 km. Vertical distances are calculated based on pressure and the hydrostatic equation. Relative horizontal transport deviations (RHTD) and relative vertical transport deviations (RVTD) are calculated by dividing the absolute transport deviations by the horizontal or vertical path lengths of the trajectories, respectively.

According to the definition, the transport deviations are calculated as mean absolute deviations of the air parcel distances. Although the mean absolute deviation is a rather intuitive approach to measure statistical dispersion, we note that it is not necessarily the most robust measure, as it can be influenced significantly by outliers. Such outliers of rather large individual transport deviations exist in some of our simulations. Strong error growth of individual trajectories can occur once the test and reference trajectories are significantly separated from each other, meaning that the air parcels are located in completely different wind regimes. To mitigate this issue we decided to report also the median of the absolute and relative transport deviations of the individual air parcels as an additional statistical measure. The median absolute deviation is a much more robust statistical measure. In all cases considered here we found that the median absolute deviation is smaller than the mean absolute deviation. This indicates that the distributions of transport deviations are skewed towards larger outliers. Note that skewed distributions of transport deviations have also been reported in other LPDM intercomparison and validation studies (e. g., Stohl et al., 2001).

2.5 Considerations on time steps and error limits

Since our test cases are based on real meteorological data, we obtained the reference data to calculate the transport deviations using the most accurate integration method available to us with a sufficiently short time step. Tests showed that the numerical solutions from the RK4 method converge for time steps of 60 s or less. In particular, comparing simulations with time steps of 120 s and 60 s, the median horizontal deviation is less than 7 km and the median vertical deviation is less than 10 m up to ~~10~~

ten days of simulation time. Alternatively, following Seibert (1993), we may also evaluate the Courant-Friedrichs-Lewy (CFL) criterion, $\Delta t \leq \Delta x / u_{max}$, to establish a time step estimate for the reference simulations. Based on an effective horizontal resolution of ~~$\Delta x \sim$~~ $\Delta x \approx 16$ km and a maximum horizontal wind speed of ~~$u_{max} \sim$~~ $u_{max} \approx 120\,\mathrm{m\,s^{-1}}$, we find that $\Delta t \leq 130$ s is needed to ensure sufficiently fine sampling of the ECMWF data. Therefore, we selected a time step of 60 s to calculate the reference trajectories.

The maximum tolerable error limits for trajectory calculations depend on the individual application of course. However, as a guideline, we here provide physically motivated error limits that are of particular interest regarding LPDM simulations. LPDMs consider both, advection and diffusion, to ~~calculate~~ simulate dispersion. Clearly, the numerical errors of the trajectory calculations, representing the advective part, should be smaller than the particle spread caused by diffusion. Considering a simple model of Gaussian diffusion, the standard deviations of the horizontal and vertical particle distributions are given by $\sigma_x = \sqrt{2 D_x t}$ and $\sigma_z = \sqrt{2 D_z t}$, respectively. Typical vertical diffusivity coefficients are ~~$D_z \sim 1$~~ $D_z \approx 1\,\mathrm{m^2\,s^{-1}}$ in the free troposphere (Pisso et al., 2009) and ~~$D_z \sim 0.1$~~ $D_z \approx 0.1\,\mathrm{m^2\,s^{-1}}$ in the lower stratosphere (Legras et al., 2003). Assuming a typical scale ratio of horizontal to vertical wind velocity of ~~~ 200~~ ≈ 200 (Pisso et al., 2009), corresponding horizontal diffusivity coefficients are ~~$D_x \sim 40\,000$~~ $D_x \approx 40\,000\,\mathrm{m^2\,s^{-1}}$ in the troposphere and ~~$D_x \sim 4000$~~ $D_x \approx 4000\,\mathrm{m^2\,s^{-1}}$ in the stratosphere. The corresponding horizontal spread after ~~10 days is $\sigma_x \sim 260$~~ ten days is $\sigma_x \approx 260$ km in the troposphere and ~~$\sigma_x \sim 85$~~ $\sigma_x \approx 85$ km in the stratosphere. The vertical spread is ~~$\sigma_z \sim 1300$~~ $\sigma_z \approx 1300$ m in the troposphere and ~~$\sigma_z \sim 415$~~ $\sigma_z \approx 415$ m in the stratosphere. However, note that these values should only be considered as a guideline. Local diffusivities may be an order of magnitude smaller or larger than these values, depending on the individual atmospheric conditions.

2.6 Experiment configuration

In this study we analyzed the ~~truncation~~ errors of trajectory calculations in 15 domains of the atmosphere, covering rather distinct conditions in terms of pressure, temperature, and winds. The globe was divided into ~~5~~ five latitude bands: polar latitudes (90°S to 65°S and 65°N to 90°N; 23.9×10^6 km^2 surface area in each hemisphere), mid-latitudes (65°S to 20°S and 20°N to 65°N; 143.9×10^6 km^2 surface area in each hemisphere), and tropical latitudes (20°S to 20°N; 174.2×10^6 km^2 total surface area). The selected ~~3~~ three altitude layers cover the free troposphere (2 to 8 km; 24 ECMWF model levels), the UT/LS region (8 to 16 km; 24 levels), and the lower and mid stratosphere (16 to 32 km; 31 levels). These domains are of major interest regarding various ~~kinds of transport simulation applications~~ applications of transport simulations using MPTRAC and other LPDMs. The planetary boundary layer was not considered here, because MPTRAC lacks more sophisticated parametrization schemes for diffusion needed for simulations in this layer. As the atmospheric conditions depend on the season and vary from year to year, we selected 1 January, 1 April, 1 July, and 1 October of the years 2014 and 2015 as start times for the simulations. All simulations cover a time period of ~~10~~ ten days. In each domain 500,000 trajectory seeds were uniformly distributed. Although this is already a large number of trajectory seeds, ~~it needs to be pointed out that this is undersampling~~ this is still undersampling as the effective resolution of the ECMWF data by as much as a factor of 4.5 in the polar troposphere up to a factor of 42 in the tropical stratosphere. Nevertheless, initial tests with different numbers of trajectory seeds showed that our

results are statistically significant. In all domains we tested time steps of 60, 120, 240, 480, 900, 1800, and 3600 s for each of the six integration schemes. In total more than 5000 individual transport simulations were performed.

~~In this study we defined~~ Here the atmospheric domains have been defined by means of fixed latitude and altitude boundaries. This is arguably a rather simple approach compared to physically motivated separation criteria based on equivalent latitudes or the dynamical tropopause. However, the simple approach may still reflect how the model is initialized and used in different applications in practice. An important consequence of ~~the simple~~ our approach is that part of the air parcels ~~left~~ leave their initial domain during the course of simulation. Table 1 provides the fraction of air parcels that remain in their initial domain after ~~5 and 10~~ five and ten days simulation time. In the stratosphere we found fractions of 48 – 88% after ~~5~~ five days and 36 – 78% after ~~10~~ ten days in the different latitude bands. In the UT/LS region the fractions are lower, i. e., 32 – 55% after ~~5~~ five days and 14 – 40% after ~~10~~ ten days. In the troposphere the ~~fraction is~~ fractions are even lower, i. e., 32 – 48% after ~~5~~ five days and 10 – 24% after ~~10~~ ten days. The lowest fractions are found for the polar latitudes for all altitude layers, being the smallest regions in terms of surface area. The horizontal wind maps shown in Fig. 1 suggest that planetary wave activity and meandering of the westerly jets between mid and high latitudes are responsible for the low fractions at polar latitudes. We also found that the fractions decrease from the stratosphere to the troposphere. This may be attributed to stronger turbulent transport associated with deep convection and eddy diffusivity in the troposphere. Although a substantial fraction of air parcels may leave their initial region during the simulations, we decided to not filter and exclude those trajectories in our analyses. The trajectories that leave the domains are more likely related to ~~larger winds and vertical~~ higher wind and velocities. Excluding those trajectories would cause a strong bias towards short trajectories, representing only the lower ~~winds and~~ wind velocities in the statistical analysis.

3 Results

3.1 Case studies of trajectory calculations

~~In this section~~ First we present two case studies that illustrate some of the common features related to trajectory calculations using different numerical integration schemes. Figure 2 shows maps of trajectories that were calculated using the six numerical schemes introduced in Sect. 2.3 with a time step of 120 s. Figure 3 provides the corresponding absolute transport deviations with respect to the reference calculations (RK4 method with 60 s time step). Both case studies show trajectories that were launched on 1 January 2014 at about 10 km altitude. In the example for the northern hemisphere the trajectories calculated using the different schemes agree well (AHTD $\leq$ 200 km and AVTD $\leq$ 600 m) for the first six days. After this point the Euler solution shows rapidly growing errors, with an AHTD up to 3900 km and an AVTD up to 4800 m after ~~8~~ eight days. The Petterssen scheme and Heun's method yield AHTDs $\leq$ 200 km and AVTDs $\leq$ 800 m for about ~~8~~ eight days, before they diverge from the reference calculation. The midpoint and RK3 method provide AHTDs $\leq$ 200 km and AVTDs $\leq$ 800 m until the end of the simulation (after ~~10~~ ten days). The example for the southern hemisphere illustrates that the onset of rapid error growth may occur much earlier in time. Here an AHTD of 200 km and an AVTD of 800 m is already exceeded after ~~3~~ three days by the Euler solution and after ~~4 to 6~~ four to six days by the solutions from Heun's method and ~~Petterssen's~~ the

Petterssen scheme. However, although error growth starts earlier, in the southern hemisphere example the maximum AHTD remains below 2200 km and the AVTD below 2200 m, which is ~~factor of 2~~ by a factor of two lower compared with the northern hemisphere example. Relative transport deviations between the examples are more similar, as the horizontal trajectory length is about 36,400 km in the northern hemisphere, but only 14,400 km in the southern hemisphere.

A common feature of the trajectory calculations we found in the case studies and also in many other situations is that the numerical integration schemes yield solutions that typically agree well up to a specific point in time before rapid error growth begins. Errors grow slowly in the beginning, but at some point, e. g., if there is strong wind shear locally, the trajectories may begin to diverge significantly. Shorter time steps or high-order integration schemes are needed to properly cope with such situations. The case studies also show that transport deviations do not necessarily grow monotonically over time. Trajectories may first diverge from and then reapproach the reference data. Individual local wind fields can bring trajectories back together by chance. The case studies also seem to suggest that vertical errors start to grow earlier than horizontal errors. Furthermore, we note that the Petterssen scheme mostly provides smaller errors than Heun's method. This was expected, because the Petterssen scheme provides iterative refinements compared with Heun's method. In both case studies the midpoint method performs better than the other ~~2nd-order~~ second-order methods. However, this is not valid in general, we also found counter-examples with the midpoint method performing worse than the other ~~2nd-order~~ second-order methods. Both examples generally exhibit large variability of the errors. This indicates that transport deviations need to be calculated for large numbers of air parcels to obtain statistically meaningful results.

3.2 Growth rates of ~~truncation~~ trajectory errors

In this section we discuss the temporal growth rates of the ~~truncation errors of the trajectory calculations~~ trajectory calculation errors from a more general point of view. Although the magnitude of the truncation errors varies largely between the schemes and with the time step used for numerical integration, we found that the transport deviations typically grow rather monotonically over time, if large numbers of particles are considered. Hence, we decided to present here ~~the truncation~~ errors using a fixed time step of 120 s for the numerical integration as a representative example. As the magnitude of the ~~truncation~~ calculation errors varies largely between the troposphere and stratosphere, we present the analysis for both regions separately. The results for the UT/LS region are not shown, as they just fall in between. We calculated combined transport deviations considering all the seasons and all the latitude bands in the given altitude range. A more detailed analysis of the ~~truncation~~ total errors in individual latitude bands and for different seasons will follow in Sect. 3.3. The influence of the choice of the time step on the accuracy and performance of the trajectory calculations will be discussed in Sect. 3.4.

Figure 4 shows the AHTDs and AVTDs of the trajectory calculations for the troposphere and stratosphere as obtained ~~by~~ with the different numerical schemes. A common feature is the clustering of the results into three groups, which we attribute to the numerical order of the integration schemes. The largest truncation errors are produced by the Euler method, which is a ~~1st-order~~ first-order scheme. After ~~10~~ ten days simulation time we found absolute (relative) horizontal transport deviations of 1450 km (14.6%) in the troposphere and 170 km (1.4%) in the stratosphere as well as vertical transport deviations of 1150 m (13.3%) in the troposphere and 194 m (3.5%) in the stratosphere. The ~~truncation errors of the 2nd-order~~ errors derived with the

second-order methods (midpoint, Heun, and Petterssen scheme) are typically 1 – 2 orders of magnitude smaller compared to the Euler method. For the midpoint method we found horizontal transport deviations of up to 320 km (3.4%) in the troposphere and 11 km (0.086%) in the stratosphere as well as vertical transport deviations of up to 361 m (3.9%) in the troposphere and 14 m (0.18%) in the stratosphere. The RK3 and RK4 methods cluster in the third group, with truncation errors being another factor 2 – 4 lower than for the ~~2nd-order~~ second-order schemes. For the RK3 method we found horizontal transport deviations of up to 228 km (2.5%) in the troposphere and 6.7 km (0.048%) in the stratosphere as well as vertical transport deviations of up to 272 m (2.9%) in the troposphere and 8 m (0.099%) in the stratosphere. We attribute the fact that there are nearly no differences between the RK3 and RK4 method to the use of the 4-D linear interpolation scheme for the meteorological data. ~~A~~ Any high-order numerical integration schemes is not expected to provide ~~any~~ large benefits in combination with a low-order interpolation scheme.

From the data presented in Fig. 4 we can also estimate the temporal growth rates of the ~~truncation~~ calculation errors as well as the leading polynomial order of the error growth. We found that error growth typically starts off linear, i. e., with a polynomial order close to one, but gets non-linear already after 1 – 2 days, with the polynomial order getting significantly larger than ~~1.~~ one. For the troposphere we found a maximum polynomial order of ~~$\sim$ 3 after 5~~ $\approx$ 3 after five days for the AHTDs and of an order ~~$\sim$ 2 after 4~~ $\approx$ 2 after four days for the AVTDs for the Euler method. The higher order methods show non-linearity at even higher levels, with a maximum polynomial order of ~~$\sim$ 5 after 8~~ $\approx$ 5 after eight days for the AHTDs and of ~~$\sim$ 4 after 6~~ $\approx$ 4 after six days for the AVTDs for the RK3 and RK4 method. The ~~2nd-order~~ second-order methods are in between. Due to ~~the~~ this non-linear error growth, the ~~growth rates of the truncation errors~~ error growth rates also increase rapidly over time until they reach their maxima after ~~10 days. For the Euler method we found horizontal growth rates of up to 334 km day^{-1} (2.1 percentage points per day) in the troposphere and 43 km day^{-1} (0.26 pp day^{-1}) in the stratosphere as well as vertical growth rates of up to 181 m day^{-1} (1.0 pp day^{-1}) in the troposphere and 35 m day^{-1} (0.41 pp day^{-1}) in the stratosphere. For the midpoint method (representing the 2nd-order methods) we found horizontal growth rates of up to 115 km day^{-1} (0.93 pp day^{-1}) in the troposphere and 3.2 km day^{-1} (0.024 pp day^{-1}) in the stratosphere as well as vertical growth rates of 105 m day^{-1} (0.89 pp day^{-1}) in the troposphere and 3.3 m day^{-1} (0.042 pp day^{-1}) in the stratosphere. For the RK3 method we found horizontal growth rates of up to 87 km day^{-1} (0.73 pp day^{-1}) in the troposphere and 1.9 km day^{-1} (0.013 pp day^{-1}) in the stratosphere as well as vertical growth rates of 86 m day^{-1} (0.74 pp day^{-1}) in the troposphere and 1.9 m day^{-1} (0.024 pp day^{-1})in the stratosphere~~ ten days, maximal error growth rates for three selected methods representing first to third order Runge-Kutta schemes are given in Table 2. During the course of the simulations, the observed error growth is largely dependent on atmospheric flow patterns. To allow for distinction between initial truncation errors and those potentially affected by atmospheric influence, we will in the following discriminate between truncation errors (analyzed after one day) and total calculation errors (analyzed after ten days).

3.3 ~~Regional~~ Spatial and ~~seasonal truncation~~ temporal variations of trajectory errors

For a more detailed analysis of the regional and seasonal ~~truncation~~ variations of the total trajectory errors we focus on the errors after ~~10~~ ten days simulation time for simulations using the ~~3rd-order Runge Kutta~~ third-order Runge-Kutta method with

a single time step of 120 s. This is considered to be a representative example as other schemes and time steps show similar ~~regional and seasonal~~ variations. We calculated individual transport deviations for all 15 altitude-latitude domains and for simulations starting at the beginning of January, April, July, and October 2014 and 2015, respectively. The results are shown in Figs. 5 and 6.

Our simulations show that horizontal errors increase from typically 20 km in the stratosphere to 100 km in the UT/LS region and about 200 km in the troposphere. The corresponding maximum AHTDs are 116 km, 177 km, and 470 km, respectively. The corresponding relative errors increase from 0.0 to 0.4 % in the stratosphere, to around 0.1 to 1.0 % in the UT/LS region, and 1.0 to 4.0 % in the troposphere. As shown in Fig. 4, the ~~truncation~~ calculation errors in the stratosphere comply on average with the error limit defined in Sect. 2.5, while the error limit in the troposphere is reached after about ~~10~~ ten days. However, as can be seen from Figs. 5 and 6, the ~~truncation~~ total errors can vary considerably seasonally and ~~inter-annually~~ from year to year as well as between the domains, causing maximum errors of specific domains to exceed the defined limit. In the stratosphere, generally the horizontal transport deviations are smaller than 40 km, far below the error limit. An exception is found for the NH polar stratosphere in January 2015 with an AHTD of 116 km. Errors are growing from the stratosphere towards the troposphere. While stratospheric wind fields are rather uniform, the turbulent wind fraction becomes stronger and more frequent at lower altitudes. Wind speeds and wind directions can vary strongly in turbulent regions. So, even if travel distances in the troposphere may be relatively short, transport deviations typically increase with decreasing altitude.

~~Truncation~~ Calculation errors at all altitude layers vary with latitude. We focus on the horizontal errors in this case, but vertical errors show similar results. The troposphere has its largest errors at northern mid-latitudes with errors between 245 km and 470 km. ~~The tropospheric~~ Tropospheric mid-latitudes were expected to cause ~~the largest errors , because the most complex wind systems occur~~ relatively large errors because of the nature of global circulation: Rossby waves and baroclinic instability occurring predominantly in this region ~~due to a larger land surface~~ come along with highly variable wind patterns. In addition, 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. The errors in the polar regions are second largest with average errors of around 200 km and peak errors in polar summer of up to 380 km. The tropics and southern mid-latitudes have small errors of less than 200 km and adhere to the error limit in all test cases. The UT/LS region has its largest AHTDs in the northern mid-latitudes with 95 ~~-~~km to 177 ~~-~~km. These errors are caused by ~~complex wind patterns~~ the north-south meandering of the jet (Woollings et al., 2014) and higher turbulence in the underlying region. The second largest errors occur in the tropics with about 75 km on average followed by the ~~north pole~~ Arctic and southern mid-latitudes with about 50 km on average. ~~The south pole has~~ Antarctica exhibits the smallest errors in this altitude layer with about 30 km on average. Errors in the stratosphere are generally small, with some larger average errors in the tropics. All test cases show errors between 2 km and 25 km, which ~~is~~ are mostly much larger than in the other stratospheric regions. 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 variation of the horizontal ~~integration~~ errors also exhibits some seasonal dependencies. This is most prominent for the northern mid latitudes, where maximum errors in all cases occur in January. During northern hemisphere wintertime land-sea temperature differences as well as the temperature gradient between the ~~North pole and the equator~~ Arctic and the subtropical regions are largest, which allows for more intense and complex dynamic patterns to occur than in summer. Our test cases for the southern hemisphere and for the arctic region do not show a seasonal behavior as clearly as one could expect. We need to stress that each simulation lasts only ~~10~~ ten 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. The small error differences between polar summer and winter additionally can be attributed to the small fraction of parcels that stay in that region. Only 13% of the parcels that are represented by the statistic remained in the polar regions after ~~10~~ ten days of simulation, which weakens our statistics.

~~As a rough indication for inter-annual variability, we simulated the same domains and periods for 2014 and 2015.~~ Most of our simulations for the corresponding months in 2014 and 2015 differ by less than 20 %, only deviations of a few individual months differ more strongly but in a similar range than the seasonal variations. Most striking differences occur in January in the stratosphere of the northern polar region. The simulation of 2014 shows small errors of 4 km, while the simulation of 2015 reaches an error of up to 116 km and exceeds the stratospheric error limit. This particular behavior (which is also present in Fig. 6 with an AVTD of 132 m) may be related to a specific meteorological situation during the winter 2014/2015, where a sudden stratospheric warming event occurred during the first days of January 2015 and temporarily caused nearly a split of the arctic vortex in the lower stratosphere (Manney et al., 2015). Significant disturbances of the wind field during this event may be a reason why trajectory calculations exhibit larger errors.

Vertical and horizontal errors behave very similar, extrema are found in the same domains. ~~Vertical transport deviations are about 800 – 1000 times smaller than the horizontal transport deviations.~~ The errors in the stratosphere are usually very small and below 10 m. Typical errors in the UT/LS region and in the troposphere are about 100 m and 250 m, respectively. Corresponding maximum errors are 130 m in the stratosphere, 168 m in the UT/LS region, and 470 m in the troposphere. The vertical error limits of 415 m in the stratosphere and 1300 m in the troposphere are easily adhered to. Relative vertical errors range 0.0 – 0.9 % in the stratosphere, 0.2 – 1.6 % in the UT/LS region, and 1.2 – 4.4 % in the troposphere.

We also calculated the horizontal and vertical median errors for the regions. In general, horizontal and vertical median errors are much smaller than the mean errors. Small median deviations shows that most trajectories follow closely to the reference. Those parcels that part from the reference usually diverge strongly, which leads to a high average deviation. The median error gets somewhat larger in the troposphere, where particle paths are more likely being affected by atmospheric turbulence.

To summarize, the relative errors of 2 – 4 % in the troposphere show that this layer is more difficult to solve and that relatively large uncertainties remain even if the absolute error limit is adhered to. The stratospheric relative errors of of about 1 % are less critical for the integration method. The large difference of the ~~truncation errors of the~~ trajectory errors between altitude regions suggests that lower order integration schemes or larger time steps could be used in the stratosphere to save computation time without causing significant errors. Tropospheric northern mid-latitudes are most challenging areas for numerical integration.

3.4 Computational efficiency

In this section we focus on the computational efficiency of the numerical integration schemes, which is assessed in terms of the trade-off between computational accuracy of and the computational time required for the trajectory calculations. As the computational efficiency depends, to some extent, on the problem size and the computer architecture that is applied, we will discuss the scalability of the application first. Our scalability tests were performed on the Jülich Research on Exascale Cluster Architectures (JURECA) supercomputer (Krause and Thörnig, 2016). JURECA is equipped with two Intel Xeon E5-2680 v3 Haswell central processing units (CPUs) per compute node. Each node is equipped with $2 \times 12 = 24$ physical compute cores, operating at 2.5 GHz clock-speed. The CPUs support 2-way simultaneous multithreading (SMT), i. e., each node provides up to 48 logical cores. A runtime improvement of up to 50% can be expected due to the SMT feature.[2]

As an example, Fig. 7 shows results of scaling tests using the midpoint scheme with a time step of 120 s for different numbers of particles and OpenMP threads. ~~For large numbers of particles (on the order of 10^4 to 10^7)~~ Note that the MPI parallelization of MPTRAC is only used for ensemble simulations, which are conducted independently on multiple nodes. The scalability of the MPI parallelization is mostly limited by I/O issues, which is beyond the scope of this study. For the OpenMP parallelization we found that the CPU time scales linearly with the number of particles for large numbers of particles (on the order of 10^4 to 10^7). The computation per time step and particle requires between 0.31×10^{-6} and 9.0×10^{-6} s computing time, depending on the number of the OpenMP threads. For small numbers of particles (~~on the order of 1~~ less or equal to 10^4) the minimum computing time is limited by ~~an~~ a constant offset of 6.3×10^{-5} s to 4.3×10^{-3} s ~~, which is due to the overhead of~~ (depending on the the number of threads) that can be attributed to the OpenMP parallelization overhead and load imbalances. Figure 7 also shows the speedup of the OpenMP parallelization for growing numbers of threads. We found that the trajectory code provides good to excellent parallel efficiency for large numbers of particles. The computational efficiency is about 83% for up to 24 physical threads and for 10^5 to 10^6 particles. It is also found that the code provides additional speedup if the simultaneous multithreading capabilities of the compute nodes are used, in particular for very large numbers of particles (on the order of 10^6 to 10^7). For smaller number of particles (10^4 or less) the speedup is limited due to the overhead of the OpenMP parallelization and ~~by the limited work load of the problem itself~~load imbalances. LPDM simulations will typically use large numbers of particles (more than 10^5) to obtain statistically significant results. MPTRAC would not be affected by scaling issues on the JURECA supercomputer in this regime.

As a measure of computational efficiency, Fig. 8 illustrates the trade-off between computational accuracy, in terms of the AHTD, and computational time. In particular, Fig. 8 illustrates how this trade-off depends on the selection of the time step for the different integration schemes. Results are shown separately for the troposphere and stratosphere ~~, as we already discussed in Sects. 3.2 and 3.3~~ after 24 h and ten days. Tropospheric and stratospheric errors are quite similar after 24 h, indicating that these errors are most representative for the truncation errors of the integration methods. In contrast, analyzing the total errors after ten days, the troposphere is much more challenging for the integration methods than the stratosphere, as already discussed in Sects. 3.2 and 3.3. From this analysis we find that despite being the fastest method, the Euler method usually has the lowest

[2]See http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JURECA/UserInfo/SMT.html (last access: 12 December 2016).

computational efficiency because of its low accuracy. The ~~2nd-order~~ second-order methods as well as the RK3 and RK4 methods yield much smaller truncation errors, in particular for short time steps. Among the ~~2nd-order methods~~ second-order methods our implementation of the Petterssen scheme has the lowest computational efficiency, which is due to the fact that we tuned the convergence criteria for this method for accuracy rather than speed. The best efficiency, ~~i. e., the best accuracy at the~~ which we define as lowest computational costs when adhering to our error limit, is mostly obtained with the midpoint and RK3 methods. The RK4 method does not provide any benefits in combination with the low-order 4-D linear interpolation scheme for the meteorological data. In fact, the RK4 method is slightly less efficient than the RK3 method due to the higher numerical costs.

Figure 8 also allows us to more accurately establish the individual optimal time steps of the integration methods with respect to the error limits defined in Sect. 2.5. This approach is similar to the well-known discrepancy principle (Engl et al., 1996), where the time step is considered as a tuning factor so that the truncation errors of the methods match an a priori known error bound. To provide estimates for all methods, we use linear extra- and interpolation to determine the largest time step that ~~just fulfills the error limit~~still adheres to the error limit. After 24 h, when trajectory errors are mostly influenced by truncation errors, the diffusivity-based error limit is not particularly strict, which allows us to use rather large time steps for the calculations. We estimated time steps of about 2100 s for the Petterssen scheme, Heun's scheme, and the midpoint method, and about 3500 s for the RK3 and RK4 methods for both the troposphere and stratosphere from Fig. 8. After ten days the diffusivity-based error limit is a lot more difficult to achieve. In addition to the truncation errors, the trajectory errors are also significantly affected by the atmospheric flow patterns (e. g., difluent flows or bifurcations). For the troposphere we derived time steps of about 100 s for the Petterssen scheme, Heun's scheme, and the midpoint method, and about 170 s for the RK3 and RK4 methods. For the stratosphere we found time steps of about 800 s for the Petterssen scheme, Heun's scheme, and the midpoint method, and time steps of about 1100 s for the RK3 and RK4 methods.

4 Summary and conclusions

In this study we characterized the ~~regional and seasonal truncation errors~~ truncation errors and total errors of trajectory calculations in the free troposphere, in the UT/LS region, and in the stratosphere. Transport simulations were conducted with the LPDM MPTRAC, driven by wind fields from T1279L137 ECMWF operational analyses and forecasts in 2014 and 2015, with an effective horizontal resolution of about 16 km and 3 h time intervals. We analyzed the computational performance of the simulations in terms of accuracy and CPU-time costs of six explicit integration schemes that belong to the ~~Runge-Kutta~~ Runge-Kutta family. The truncation errors of the schemes for a given time step were found to cluster into three groups that are related to the order of the method: (i) the ~~1st-order~~ first-order Euler method, (ii) the ~~2nd-order~~ second-order methods (midpoint method, Heun's method, and Petterssen's scheme), (iii) the higher order methods, which are the common RK3 and RK4 methods. Different methods within each group provide similar accuracy in terms of error growth rates and transport deviations.

Based on more than 5000 individual transport simulations, we further analyzed horizontal and vertical transport deviations in relation to altitude, latitude, as well as seasonal and year-to-year variability. After 24 h the trajectory errors are quite similar in

the troposphere and stratosphere and mostly affected by truncation errors. After ten days the trajectory errors vary substantially in different domains because of the influence of atmospheric flow patterns. We found that tropospheric simulations require more accurate integration methods or significantly shorter time steps to keep errors within physically motivated error limits than simulations for the stratosphere. We attribute this to larger small-scale variations caused by atmospheric turbulence ~~and mixing in the troposphere. Truncation~~ . Calculation errors also depend on the latitude band, with the northern mid-latitudes having the largest errors in each altitude layer. Seasonal ~~and inter-annual error variations~~ error variations and differences from year to year are clearly visible from our simulations, but in some cases the number of samples still seems to be too small to deduce robust statistics. One example are large errors that are associated with a sudden stratospheric warming in the northern stratosphere in January 2015, which suggests that part of the ~~truncation errors~~ total error is due to situation-dependent factors. However, a robust feature seams to be a northern mid-latitude winter maximum in the troposphere and stratosphere, existent in both years, 2014 and 2015.

All integration methods discussed here are in principle suited and have already been used for ~~LPDM~~ Lagrangian Particle dispersion and trajectory model simulations. To decide which method is most efficient on state-of-the-art high performance computing systems, we analyzed the trade-off between computational accuracy and computational time. This trade-off is largely controlled by the time step used for numerical integration. The Euler method requires very short time steps to achieve reasonably accurate results and is ~~therefore~~ generally not considered to be an efficient method. Heun's method and the iterative Petterssen scheme are more accurate at the same computational costs. The midpoint method and the RK3 method usually ~~lead to~~ provided the most efficient simulations with MPTRAC, i. e., these methods provide the most accurate results at the lowest computational costs. Note that the RK4 method is slightly more expensive than the RK3 method if it is applied together with a low-order linear interpolation scheme for the meteorological data.

The study of Seibert (1993) addressed the choice of the numerical integration scheme and choice of the time step based on idealized test cases and for realistic wind fields. 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. ~~This new study~~ To achieve sufficiently small truncation errors they recommended to use 15% of the time step needed for convergence of the Petterssen scheme. Here we used meteorological data with a ~~very fine~~ much finer spatial resolution as ~~used in~~ provided by current global weather forecast models, which requires adjustment of the time step. Time steps of ~~10 minutes to 1~~ ten minutes to one hour as used in former trajectory studies (e. g. Seibert, 1993; Stohl et al., 1998, 2001; Harris et al., 2005) are far beyond yielding convergence with high-resolution meteorological data. Given ~~a~~ an effective horizontal resolution of 16 km and applying the CFL criterion, the time step needs to be shorter than about 130 s to achieve convergence. From our simulations we found that time ~~step~~ steps of 100 s ~~(midpoint method)~~ for the midpoint method and 170 s ~~(~~for the RK3 method ~~)~~ provide accurate results ~~for the troposphere~~ in the troposphere for up to ten days. Purely stratospheric applications can be solved with time steps of 800 s (midpoint method) and 1100 s (RK3 method) because of lower ~~truncation~~ total errors in this altitude layer.

In this study we considered a range of popular and well-established integration schemes for trajectory calculations in LPDMs. However, the large variability of regional and seasonal ~~truncation~~ errors found here suggests that applications may benefit from more advanced numerical techniques. Adaptive quadrature ~~could be an interesting topic~~ by means of variable time stepping as

recommended by earlier studies (Walmsley and Mailhot, 1983; Maryon and Heasman, 1988; Seibert, 1993) could be taken up for future research.

5 Code and data availability

Operational analyses and forecasts ~~are distributed by~~ can be obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF), see http://www.ecmwf.int/en/forecasts (last access: ~~21 December 2016~~3 May 2017) for further details ~~. The code of the Massive-Parallel Trajectory Calculations (MPTRAC)model~~ on data availability and restrictions. ECMWF data have been processed for usage with MPTRAC by means of the Climate Data Operators (CDO, https://code.zmaw.de/projects/cdo, last access: 3 May 2017). The version of the MPTRAC model that was used for this study along with the model initializations is available under the terms and conditions of the GNU General Public License, Version 3 from the repository at https://github.com/slcs-jsc/mptrac-advect (last access: ~~21 December 2016~~3 May 2017).

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors acknowledge the Jülich Supercomputing Centre (JSC) for providing computing time on the supercomputer JURECA. YH ~~also~~ acknowledges support from the "100 Talents Program" of Sun Yat-sen University, Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase), and the National Supercomputer Center in Guangzhou. We thank Xue Wu for helpful comments on an earlier draft of this manuscript.

References

Bowman, K. P., Lin, J. C., Stohl, A., Draxler, R., Konopka, P., Andrews, A., and Brunner, D.: Input Data Requirements for Lagrangian Trajectory Models, B. Am. Meteorol. Soc., 94, 1051–1058, 2013.

Buizza, R., Houtekamer, P. L., Toth, Z., Pellerin, G., Wei, M., and Zhu, Y.: A comparison of the ECMWF, MSC, and NCEP Global ensemble prediction systems, Mon. Wea., 133, 1076–1097, 2005.

Butcher, J. C.: Numerical methods for ordinary differential equations, John Wiley & Sons, 2008.

Davis, L. S. and Dacre, H. F.: Can dispersion model predictions be improved by increasing the temporal and spatial resolution of the meteorological input data?, Weather, 64, 232–237, 2009.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kãllberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quart. J. Roy. Meteorol. Soc., 137, 553–597, 2011.

Draxler, R. R. and Hess, G. D.: An overview of the HYSPLIT_4 modeling system of trajectories, dispersion, and deposition, Aust. Meteor. Mag., 47, 295–308, 1998.

Engl, H. W., Hanke, M., and Neubauer, A.: Regularization of Inverse Problems, Kluwer Academic Publishers, Dordrecht, 1996.

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

Heng, Y., Hoffmann, L., Griessbach, S., Rößler, T., and Stein, O.: Inverse transport modeling of volcanic sulfur dioxide emissions using large-scale simulations, Geosci. Model Dev., 9, 1627–1645, 2016.

Hoffmann, L., Xue, X., and Alexander, M. J.: A global view of stratospheric gravity wave hotspots located with Atmospheric Infrared Sounder observations, J. Geophys. Res., 118, 416–434, 2013.

Hoffmann, L., Rößler, T., Griessbach, S., Heng, Y., and Stein, O.: Lagrangian transport simulations of volcanic sulfur dioxide emissions: impact of meteorological data products, J. Geophys. Res., doi:10.1002/2015JD023749, 2016.

Hoppe, C. M., Hoffmann, L., Konopka, P., Grooß, J.-U., Ploeger, F., Günther, G., Jöckel, P., and Müller, R.: The implementation of the CLaMS Lagrangian transport core into the chemistry climate model EMAC 2.40.1: application on age of air and transport of long-lived trace species, Geosci. Model Dev., 7, 2639–2651, 2014.

Jones, A., Thomson, D., Hort, M., and Devenish, B.: The UK Met Office's next-generation atmospheric dispersion model, NAME III, in: Air Pollution Modeling and its Application XVII, pp. 580–589, Springer, 2007.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–471, 1996.

Krause, D. and Thörnig, P.: JURECA: General-purpose supercomputer at Jülich Supercomputing Centre, J. Large-Scale Res. Facilities, 2, 62, 2016.

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. Wea. Rev., 113, 1852–1867, 1985.

Legras, B., Joseph, B., and Lefèvre, F.: Vertical diffusivity in the lower stratosphere from Lagrangian back-trajectory reconstructions of ozone profiles, J. Geophys. Res., 108, 4562, doi:10.1029/2002JD003045, 2003.

Lin, J., Gerbig, C., Wofsy, S., Andrews, A., Daube, B., Davis, K., and Grainger, C.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, J. Geophys. Res., 108, 4493, doi:10.1029/2002JD003161, 2003.

Lin, J., Brunner, D., Gerbig, C., Stohl, A., Luhar, A., and Webley, P., eds.: Lagrangian modeling of the atmosphere, vol. 200 of *Geophysical Monograph Series*, American Geophysical Union, Washington DC, 2012.

Manney, G. L., Lawrence, Z. D., Santee, M. L., Read, W. G., Livesey, N. J., Lambert, A., Froidevaux, L., Pumphrey, H. C., and Schwartz, M. J.: A minor sudden stratospheric warming with a major impact: Transport and polar processing in the 2014/2015 Arctic winter, Geophys. Res. Lett., 42, 7808–7816, doi:10.1002/2015GL065864, 2015.

Maryon, R. and Heasman, C.: The accuracy of plume trajectories forecast using the UK Meteorological Office operational forecasting models and their sensitivity to calculation schemes, Atmospheric Environment (1967), 22, 259–272, 1988.

Petterssen, S.: Weather analysis and forecasting, McGraw-Hill, New York, 1940.

Pisso, I., Real, E., Law, K. S., Legras, B., Bousserez, N., Attié, J. L., and Schlager, H.: Estimation of mixing in the troposphere from Lagrangian trace gas reconstructions during long-range pollution plume transport, J. Geophys. Res., 114, D19301, doi:10.1029/2008JD011289, 2009.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical Recipes in C, The Art of Scientific Computing, vol. 1, Cambridge University Press, 2. edn., 2002.

Preusse, P., Eckermann, S. D., Ern, M., Oberheide, J., Picard, R. H., Roble, R. G., Riese, M., Russell III, J. M., and Mlynczak, M. G.: Global ray tracing simulations of the SABER gravity wave climatology, J. Geophys. Res., 114, D08126, doi:10.1029/2008JD011214, 2009.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A.: The ECMWF operational implementation of four-dimensional variational assimilation, Q.J.R. Meteorol. Soc., 126, 1143–1170, 2000.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Clim., 24, 3624–3648, 2011.

Rolph, G. D. and Draxler, R. R.: Sensitivity of three-dimensional trajectories to the spatial and temporal densities of the wind field, J. Appl. Met., 29, 1043–1054, 1990.

Seibert, P.: Convergence and accuracy of numerical methods for trajectory calculations, J. Appl. Met., 32, 558–566, 1993.

Sprenger, M. and Wernli, H.: The LAGRANTO Lagrangian analysis tool–version 2.0, Geoscientific Model Development, 8, 2569–2586, 2015.

Stohl, A.: Computation, accuracy and applications of trajectories – a review and bibliography, Atmos. Environment, 32, 947–966, 1998.

Stohl, A. and Seibert, P.: Accuracy of trajectories as determined from the conservation of meteorological tracers, Quart. J. Roy. Meteorol. Soc., 124, 1465–1484, 1998.

Stohl, A., Wotawa, G., Seibert, P., and Kromp-Kolb, H.: Interpolation errors in wind fields as a function of spatial and temporal resolution and their impact on different types of kinematic trajectories, J. Appl. Met., 34, 2149–2165, 1995.

Stohl, A., Hittenberger, M., and Wotawa, G.: Validation of the Lagrangian particle dispersion model FLEXPART against large-scale tracer experiment data, Atmospheric Environment, 32, 4245–4264, 1998.

Stohl, A., Haimberger, L., Scheele, M., and Wernli, H.: An intercomparison of results from three trajectory models, Meteorol. Applic., 8, 127–135, 2001.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmos. Chem. Phys., 5, 2461–2474, 2005.

Stohl, A., Prata, A. J., Eckhardt, S., Clarisse, L., Durant, A., Henne, S., Kristiansen, N. I., Minikin, A., Schumann, U., Seibert, P., Stebel, K., Thomas, H. E., Thorsteinsson, T., Tørseth, K., and Weinzierl, B.: Determination of time- and height-resolved volcanic ash emissions and their use for quantitative ash dispersion modeling: the 2010 Eyjafjallajökull eruption, Atmos. Chem. Phys., 11, 4333–4351, 2011.

Walmsley, J. L. and Mailhot, J.: On the numerical accuracy of trajectory models for long-range transport of atmospheric pollutants, Atmosphere-Ocean, 21, 14–39, 1983.

Wernli, H. and Davies, H. C.: A Lagrangian-based analysis of extratropical cyclones. I: The method and some applications, Quart. J. Roy. Meteorol. Soc., 123, 467–489, 1997.

Woollings, T., Czuchnicki, C., and Franzke, C.: Twentieth century North Atlantic jet variability., Quart. J. Roy. Meteorol. Soc., 140, 783–791, doi:10.1002/qj.2197, 2014.

Wu, X., Griessbach, S., and Hoffmann, L.: Equatorward dispersion of high-latitude volcanic plume and its relation to the Asian summer monsoon: a case study of the Sarychev eruption in 2009, Atmos. Chem. Phys. Discuss., 2017, doi:10.5194/acp-2017-425, 2017.

Table 1. Fractions of ~~Air Parcels Remaining~~ air parcels remaining in ~~Initial Regions~~ initial regions during ~~Course~~ the course of ~~Simulations~~the simulations.

	SH Polar Lat. (90°S – 65°S)	SH Mid Lat. (65°S – 20°S)	Tropical Lat. (20°S – 20°N)	NH Mid. Lat. (20°N – 65°N)	NH Polar Lat. (65°N – 90°N)
After 5 Days Simulation Time:					
Stratosphere (16 – 32 km)	66%	86%	88%	78%	48%
UT/LS Region (8 – 16 km)	42%	55%	44%	53%	32%
Troposphere (2 – 8 km)	32%	46%	48%	44%	32%
After 10 Days Simulation Time:					
Stratosphere (16 – 32 km)	54%	77%	78%	67%	36%
UT/LS Region (8 – 16 km)	25%	40%	19%	36%	14%
Troposphere (2 – 8 km)	13%	21%	24%	20%	10%

Table 2. Maximal error growth rates of trajectories. Relative growth rates in pp day^{-1} are given in parenthesis.

	Troposphere		Stratosphere	
	horizontal [km d^{-1}]	vertical [m d^{-1}]	horizontal [km d^{-1}]	vertical [m d^{-1}]
Euler method	334 (2.1)	181 (1.0)	43 (0.26)	35 (0.41)
Midpoint method	115 (0.93)	105 (0.89)	3.2 (0.024)	3.3 (0.042)
RK3 method	87 (0.73)	86 (0.74)	1.9 (0.013)	1.9 (0.024)

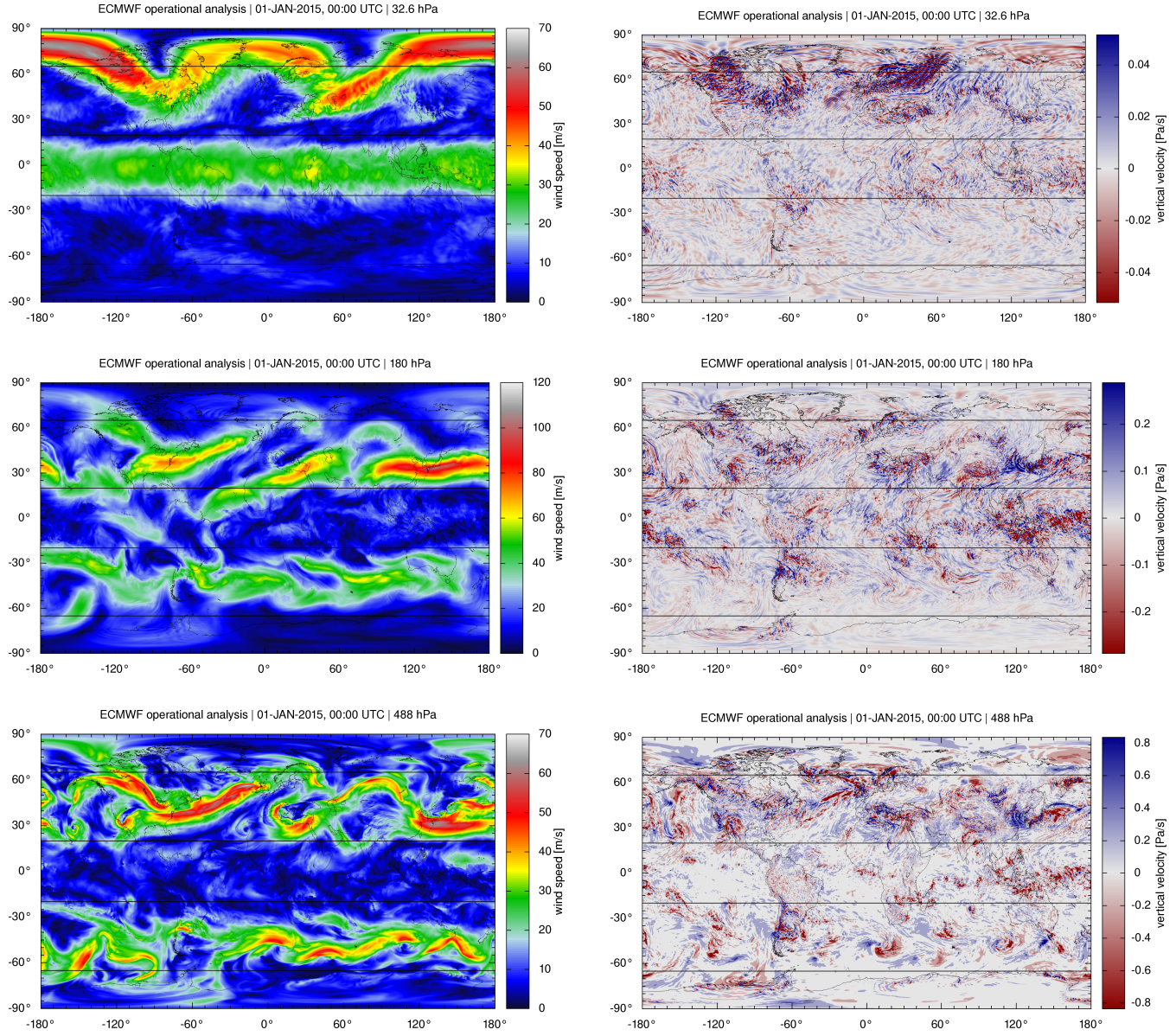

Figure 1. ECMWF operational analysis horizontal wind speed (left) and vertical velocity (right) at about 24 km (top), 12 km (middle), and 5 km (bottom) altitude on 1 January 2015, 00:00 UTC. Black lines indicate the latitude bands considered in our analysis.

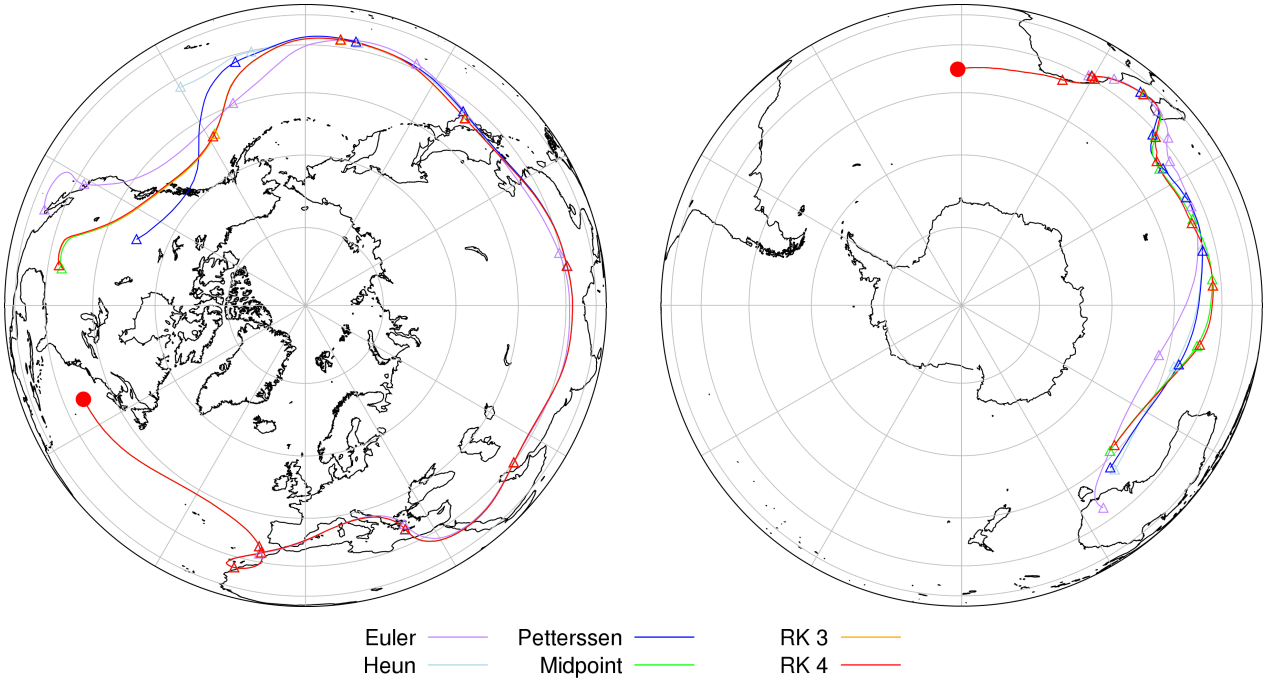

Figure 2. Examples of trajectory calculations using different numerical integration schemes. Circles mark the start positions of the trajectories. Trajectories were launched at an altitude of 10.8 km (left) and 9.7 km (right). The start time is 1 January 2014, 00:00 UTC for both. Triangles mark trajectory positions at 00:00 UTC on each day.

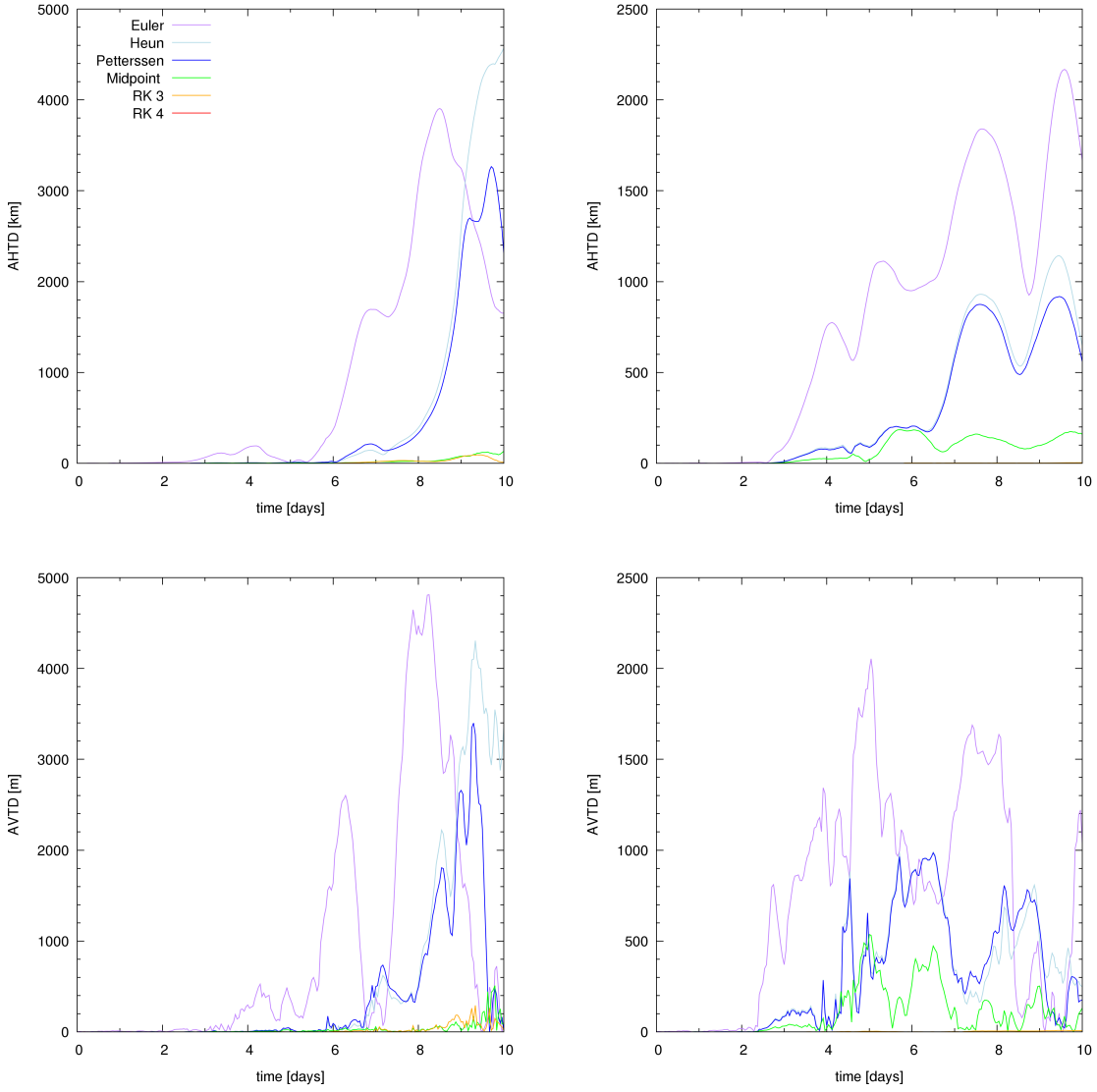

Figure 3. Absolute horizontal (top) and vertical (bottom) transport deviations for the case studies for the northern hemisphere (left) and southern hemisphere (right) presented in Fig. 2. Please note different ranges of y-axes. Results of the RK3 and RK4 method are close to zero in most cases.

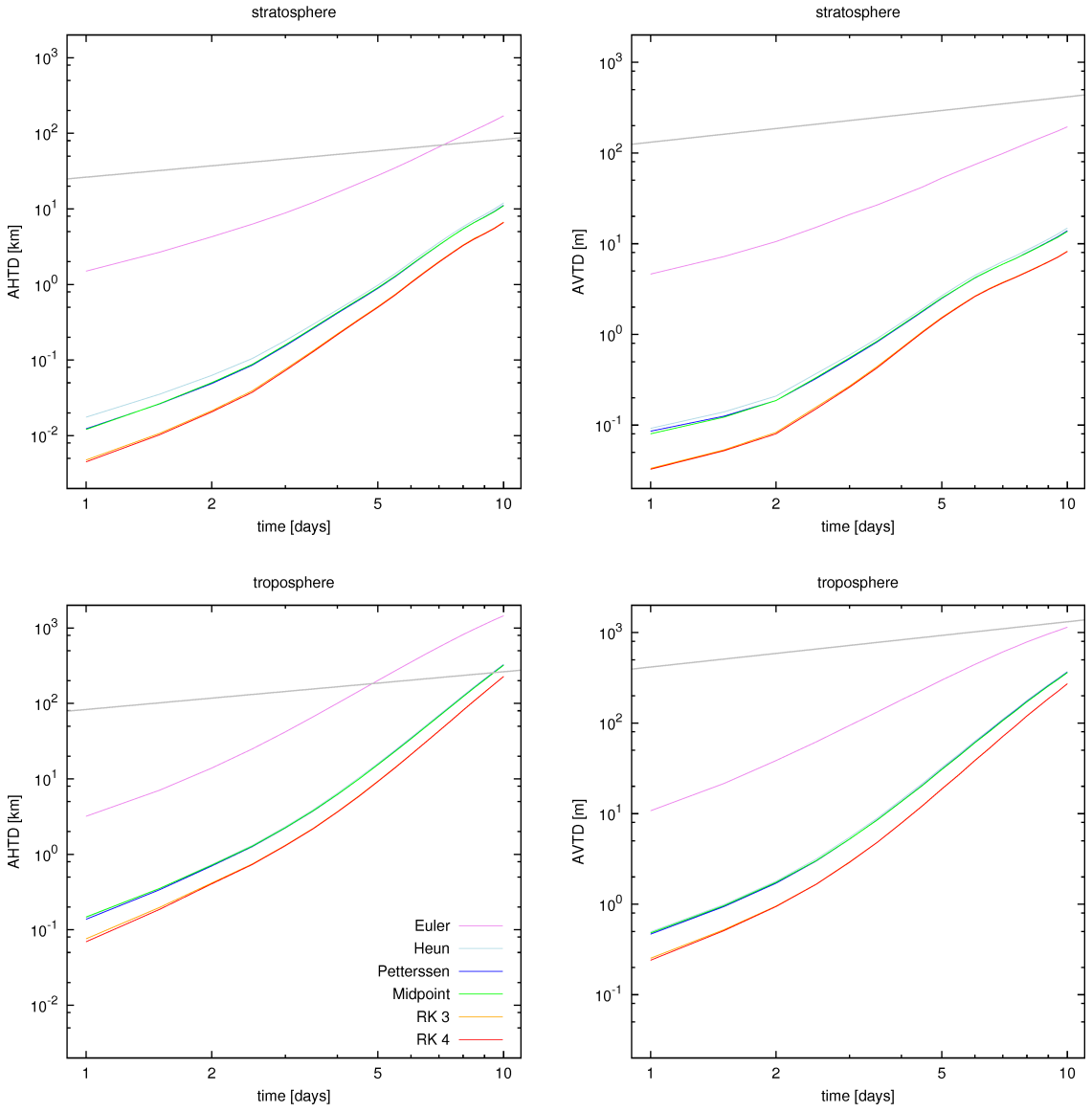

Figure 4. Absolute horizontal (left) and vertical (right) transport deviations of trajectory calculations for the stratosphere (top) and troposphere (bottom). The trajectory calculations are based on different numerical schemes, but use the same time step ($\Delta t = 120\,\mathrm{s}$). Grey lines show error limits based on particle diffusion.

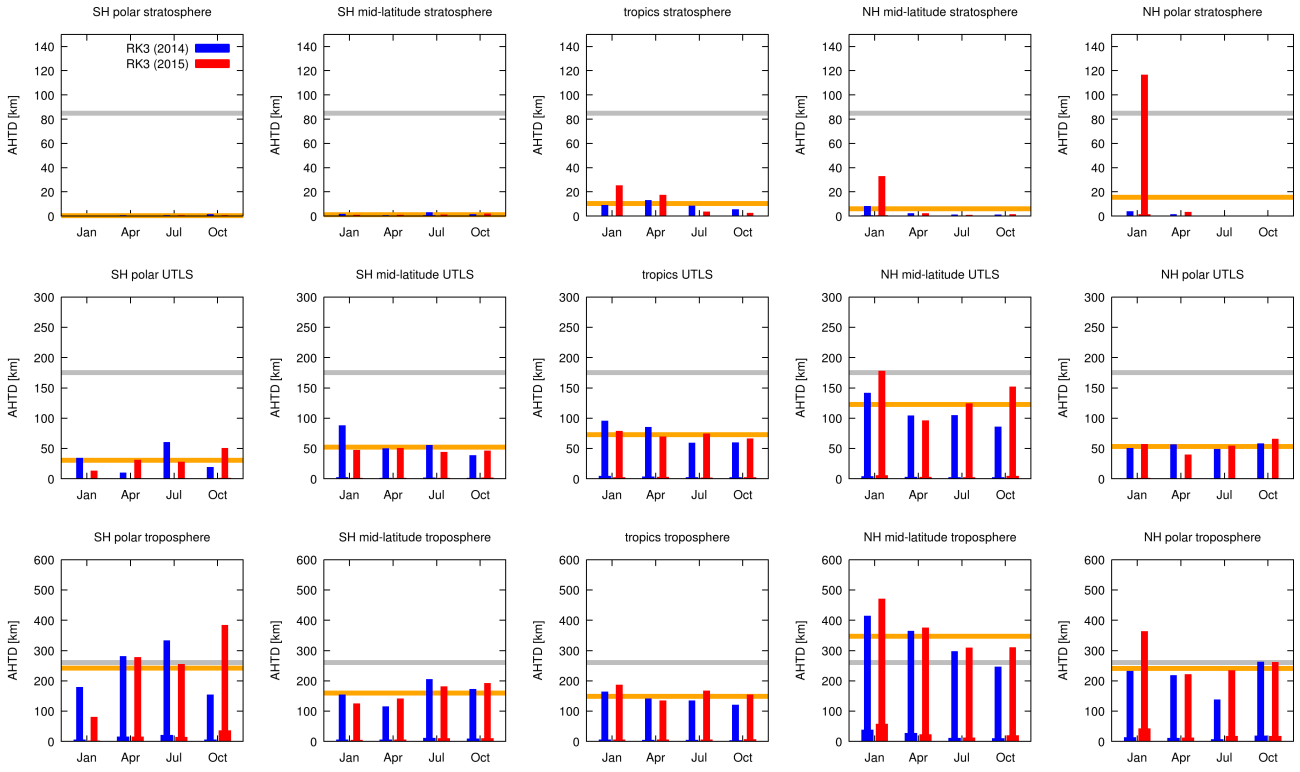

Figure 5. Mean (thin bars) and median (thick bars) horizontal transport deviations after ~~10~~ ten days simulation time in different domains for the RK3 method and 120 s time step. Orange lines show the averages of the four months (January, April, July, and October) and both years (2014 and 2015). Gray lines show error limits based on diffusion.

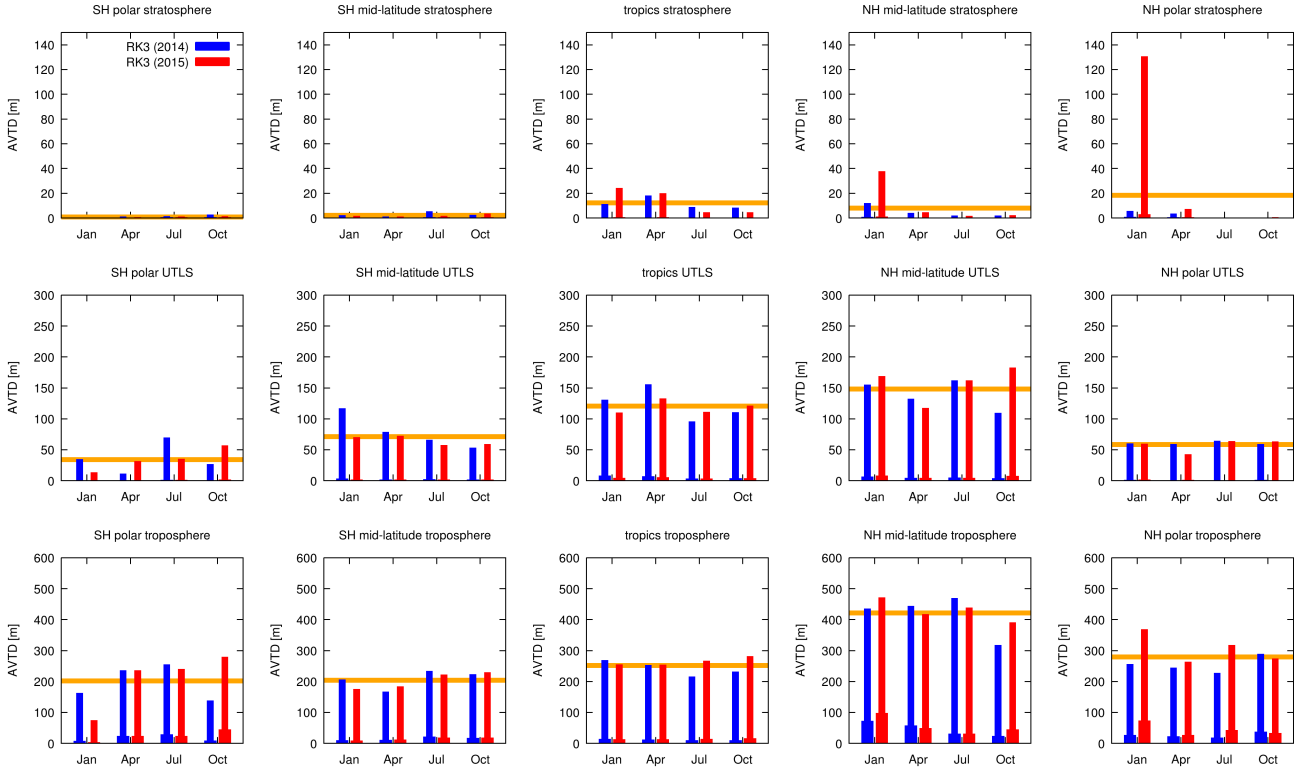

Figure 6. Same as Fig. 5, but for vertical transport deviations. Error limits based on diffusion are about 1300 m for the troposphere and 415 m for the stratosphere, which is beyond the AVTD ranges shown here.

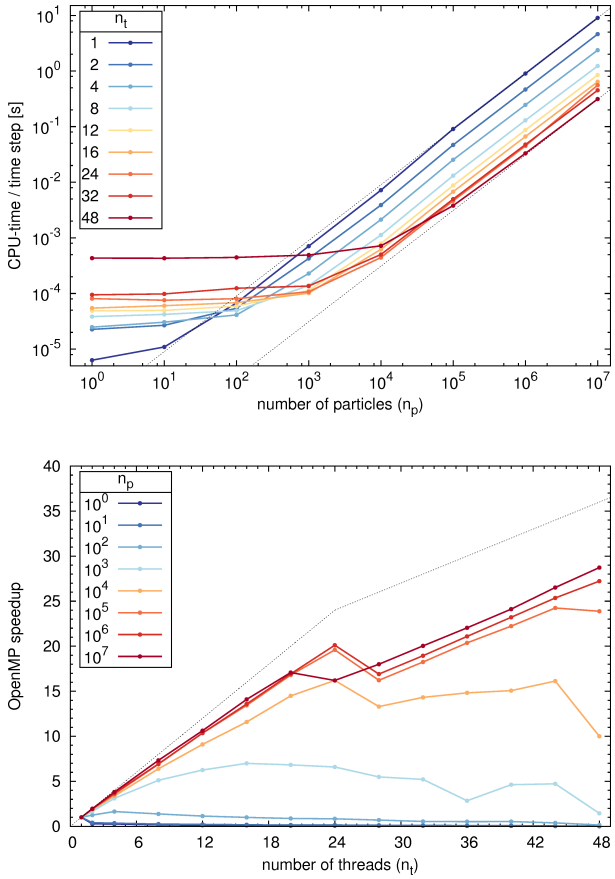

Figure 7. Scaling behaviour in terms of CPU-time (top) and speedup of the code (bottom) used to calculate trajectories with the ~~mid-point~~ midpoint method and a time step of 120 s for different numbers of particles (n_p) and OpenMP threads (n_t). Colored curves refer to different numbers of OpenMP threads (top) or different total numbers of particles (bottom). Dotted lines show ideal scaling behavior.

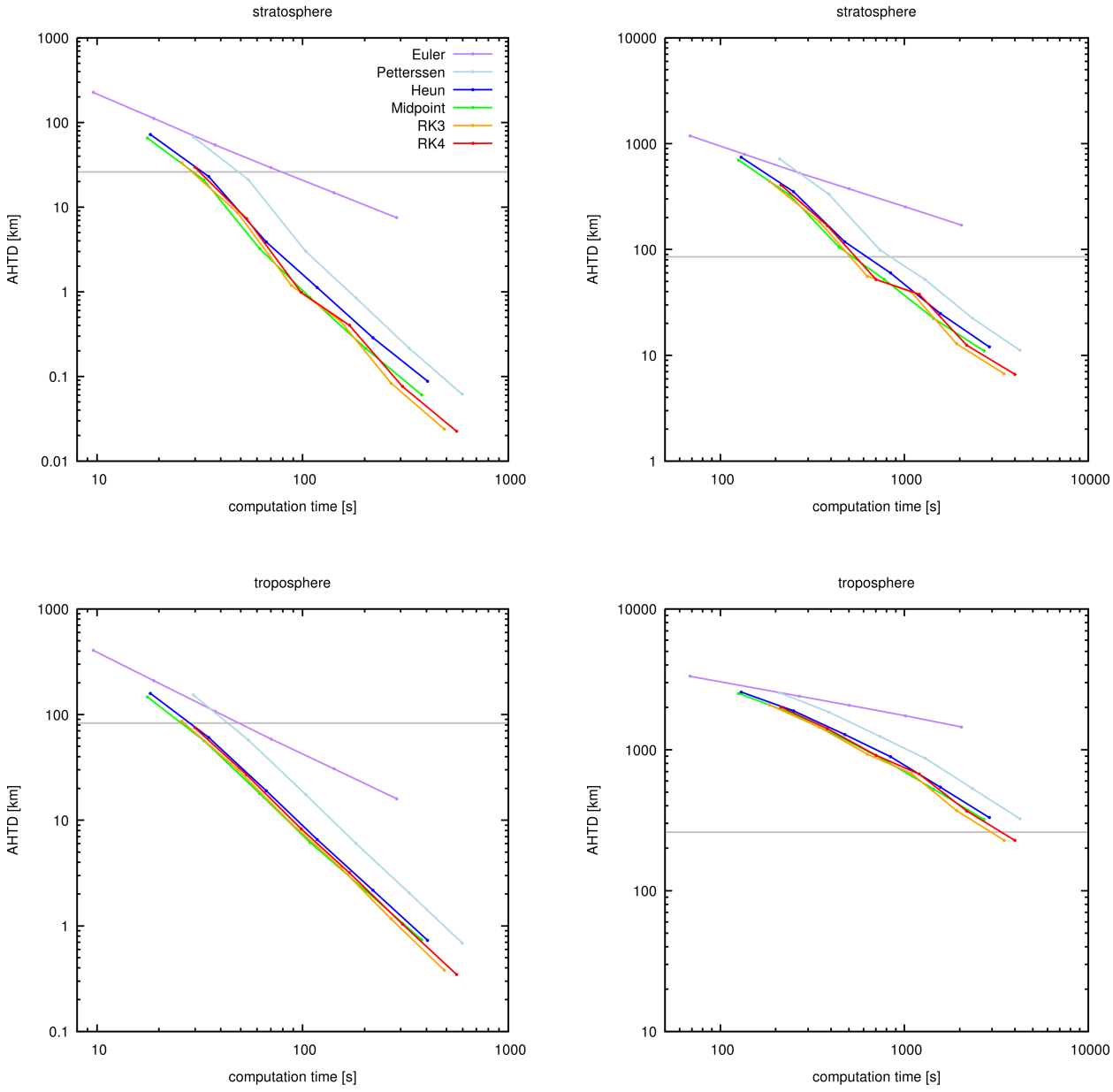

Figure 8. Trade-off between computational accuracy and total CPU-time requirements of the trajectory calculations after 24 h (left) and 10 days (right). Colored curves refer to different integration schemes. Dots along the curves indicate time steps of 3600, 1800, 900, 480, 240, and 120 s (from left to right). Horizontal lines refer to the maximum tolerable error limits as defined in Sect. 2.5. Note that our implementation of the Petterssen scheme was optimized for numerical accuracy rather than speed.