Reply to review comments

Dear Dr. Pisso, dear Dr. Seibert,

we thank you for the time and efforts spent on the manuscript. We apologize that not all of the earlier comments have been addressed properly, but we hope that the revised draft solves the remaining issues. Please find our point-by-point replies below (colored in blue).

Reviewer #1

0.1 General

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

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

0.2 Major remarks

1. clarified

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

In order to better justify our choice, we used a time step of 30 s to recompute the reference trajectories for January 2015. The resulting absolute horizontal transport deviations after 10 days changed by 0-3.3 km in the stratosphere, by 0.6-8.5 km in the UT/LS region and by 4-14 km in the troposphere in this test case. These differences are much smaller than the AHTDs presented in Fig. 5 and we can therefore consider our results to be robust. The only exception would be the high and mid latitudes of the southern hemisphere of the stratosphere, but here the numerical errors are very small anyway (AHTD <6 km).

We think that a reference calculation with a 60 s time step is adequate for all presented cases. We changed the wording to make clear that by convergence, the convergence of the transport deviations relative to simulations with larger time steps is meant:

Sensitivity tests using variable time steps down to 30 s showed that the numerical solution from the RK4 method converges for time steps of 60 s or less, in the sense that transport deviations relative to simulations with a time step of 120 s do not change significantly.

The remark to Hoffmann et al. (2016b) was made to clarify that the two studies are not contradicting each other, because the specific region (UTLS) needs to be taken into account when comparing the results of the studies.

3. clarified

4. OK.

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

We agree that a study of truncation errors with a high temporal resolution, which fits better to the fine horizontal resolution, would be interesting. We admit that the 3 h time interval is a potential weakness of this study. Considering that substantially more ECMWF data would have to be downloaded, all simulations would have to be repeated and the complete paper would have to be rewritten, we think that this refinement is not feasible now. We discuss this potential weakness of the study in the concluding section.
A comparison of different resolutions is beyond the scope of this study. We already vary the region, integration method and time step, which are quite many variables for one study.

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

We agree that the term *total error* is misleading, it would include other error sources

that are not considered in this study. Therefore, we changed the wording to *global truncation error*, which implies that the numerical errors investigated here originate from the accumulation of local truncation errors that are introduced at each time step. This terminology is consistent with the literature of numerical mathematics.

0.3 Specific and Minor Remarks

1. Authors now propose the title 'Domain specific trajectory errors diagnosed with the MPTRAC advection module and ECMWF operational analyses'. First of all, one would have to hyphenate 'domain-specific'. Then, I think the word *domain* is not the best choice to express that evaluations were done separately for different regions (*domain* usually refers to a calculation domain and not to a climatological region.) Furthermore, MPTRAC and ECMWF data are not on the same level, one is the model, the other model input. Better say something like 'Trajectory errors as a function of numerical scheme, time step, and region of the atmosphere diagnosed with the MPTRAC model'. Maybe you can add 'for ECMWF wind fields', but I think it is not so important to bring that into the title.

We changed the term 'domain' to 'region' throughout the manuscript and removed it from the title. We also rephrased the title to clarify that the calculations are driven by external ECMWF operational data:

Trajectory errors diagnosed with the MPTRAC advection module driven by ECMWF operational analyses

2. Page 1, line 1: Abstract. OK

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

We traced back our usage of the phrase Equations of motion to the reference paper of Bowman et al. (2013). We also think this is a standard term in physics textbooks. We kept it as is but added the term *trajectory equation* as a synonym.

4. OK

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

one is wondering about partial information until one has reached the next section.

We removed information regarding the time resolution from the introduction to avoid confusion for the reader at this point. The horizontal resolution is still mentioned to show that relatively fine resolved meteorological data are used. The temporal resolution and information about the data processing are now given in Section 2:

For usage with MPTRAC, the wind fields were retrieved on model levels with a longitude-latitude grid with $0.125° \times 0.125°$ resolution and have been interpolated vertically to 114 pressure levels in the troposphere and stratosphere up to 5 hPa with help of the Climate Data Operators (CDO, 2015). 12-hourly analyses were combined with short-term forecasts in between to obtain data with a 3-hourly time step.

6. OK

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

This paper focuses on the advection module only, therefore we think no full model description is necessary. However, we added references to Heng et al. (2016) and Hoffmann et al. (2017) for readers that are interested in more information about the model.

8. OK

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

We added information about the data conversion to the manuscript, see answer to minor remark #5.

10. OK

11. *Page 6, line 4 ff.: k1 = ...It seems that you define certain velocities as k.* I regret that the authors want to keep this notation. In any case, symbols have to be explicitly explained (in words) before or immediately after first usage, whether or not there is an equation which defines them. This is a standard in scientific publications.

We added an explanatory phrase between Eq. (7) and (8).

12. OK

13. OK

14. *Page 11, line 30: The tropospheric mid-latitudes were expected to cause the largest errors, because the most complex wind systems occur in this region due to a larger land surface ratio and more complex orography.* The authors offer some improvement, but it is still insufficient.
(1) They should pay attention to their style. For example, wordings such as 'The troposphere has its largest error' are not correct (this example is not the only such mistake). The troposphere cannot have any error, only calculations can.

(1) We tried to improve the wording in the whole paragraph.

(2) We still find the phrase 'the evolution of northern mid-latitudes meteorological systems is more difficult to simulate than for the southern mid-latitudes due to the larger land-sea ratio and more complex orography of the northern hemisphere.' Apart from the question whether this is true or not, the difficulty to simulate the evolution of meteorological systems (in other words, the predictability) is not relevant for trajectory errors based on analyses.

(2) Our intention was to give reasons for larger trajectory errors in the northern mid-latitudes compared to the southern mid-latitudes. We changed our statement accordingly:

In addition, stronger fluctuations are expected in the northern mid-latitudes compared to the southern mid-latitudes due to the larger land-sea ratio and more complex orography of the northern hemisphere.

(3) 'These errors are caused by . . . and higher turbulence in the underlying region'. In their answer to minor remark 16 authors admit that turbulence is not relevant thus, why do they again come up with turbulence as an explanation of trajectory errors?

(3) We used the term turbulence for small-scale fluctuations of the wind field, which was wrong. We changed it throughout the manuscript.

15. OK

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

Our original explanations was misleading. We would like to explain why trajectory errors in the tropical stratosphere are larger than in other stratospheric regions and changed the paragraph to:

The errors of the simulations in the stratosphere are typically below 25 km, except for January 2015. Stratospheric trajectory errors in the tropics are larger than in the other regions, which is probably due to the close vicinity of the tropical tropopause, which reaches an average altitude of 16 km near the ITCZ.

17. OK

18. *Page 12, line 12: We need to stress that each simulation lasts only 10 days, which*

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

Our discussion of seasonal and year-to-year variability has already been shortened and we try to make clear that it is not meant to be statistically representative. However, Figure 5 of the manuscript clearly indicates that some seasonal effects are present and that year-to-year variability might be large. We do not want to average the few data points of each region, because then possible effects of seasons or year-to-year variability would be completely hidden. We can not claim these effects to be statistically significantly due to the limited sample size, but we think that they are relevant and want to underline that trajectory errors depend on many factors. As the average of the seasonal samples has been added to the figures following earlier review comments, we think that it does no harm to show all available information.
Each of the 5,000 simulations uses 500,000 trajectories which results in high computation costs. A study of seasonal and year-to-year variations would be much easier and cheaper if only one integration method and one time step would be used. If more representative results are necessary, a separate study could be conducted in the future.

19. OK

20. *Page 12, line 35: The median error gets somewhat larger in the troposphere, where particle paths are more likely being affected by atmospheric turbulence. Hoffmann et al. (2016a) says that MPTRAC uses the same diffusivity throughout troposphere and stratosphere. How is this compatible?*
The authors say that they answered this in their response to major comment 1. However, there they only explain that turbulence is not active in their simulations. But then, this argument simple collapses; however, the sentence in question has been left unchanged.

We were misleadingly referring to turbulence instead of small-scale fluctuations of the wind field. We changed the sentence to:

The median error is somewhat larger for simulations in the troposphere, where particle paths are more likely being affected by synoptic-scale fluctuations of the wind field.

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

In our computing environment it is easier to conduct ensemble simulations (e.g. Heng et al. (2016)) if the code offers MPI parallelization.

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

We added the information that only the advection module is analyzed (as the other physics modules are not enabled). The advection module does not have a non-parallel part. The code structure shows that time measurement directly embraces the parallel section:

```
START_TIMER(timer_phys);
#pragma omp parallel for default(shared) private(dt,ip)
for (ip = 0; ip < atm->np; ip++)
 module_advection(ip, ...);
STOP_TIMER(timer_phys);
```

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

We added the explanation to the manuscript:

For smaller number of particles (10^4 or less) the speedup is limited by the overhead of the OpenMP parallelization and by load imbalances, which can also become significant for larger numbers of parcels if SMT is not enabled for all cores jointly.

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

As it is difficult to define a reasonable iteration cut-off, we executed new simulations with an implementation of the Petterssen method that uses exactly two iterations. The accuracy in most simulations is similar or slightly worse than the accuracy of the midpoint

method. In conclusion, none of our three implementations of the Petterssen scheme was more efficient than the midpoint method or the Runge-Kutta 3 method.

The accuracy of the Petterssen scheme with one iteration (Heun's method) is somewhat worse than the midpoint method. When two iterations of the Petterssen scheme are computed, the transport deviations are closer to those obtained with the midpoint method. The best efficiency, which we define as lowest computational costs when adhering to our error limit, is mostly obtained with the midpoint and RK3 methods. The additional iterations of the Petterssen scheme improve the accuracy, but they are too computational expensive for our model. In general, a well defined convergence limit for the number of iterations is needed for an efficient application.

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

We added a note on the influence of the hardware:

Note that the hardware, especially the memory cache, affects the six integration schemes differently. A single call to the wind interpolation function is up to 50% cheaper for a higher order method compared to Euler's method, because the cache is used more efficiently.

26. OK

27. OK

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

We corrected this and removed the reference to turbulence.

29. Summary and conclusions: Please spell out RK where it occurs for the first time in this section (some people may read only this section).

The first occurrence of Runge-Kutta has been spelled out in the conclusions.

'After 24 h the trajectory errors are quite similar in the troposphere and stratosphere' Figure 4 shows a difference by about a factor of 10 for AHTD, thus they are not 'quite similar'.

We agree that the difference is relevant and changed the sentence accordingly.

The trajectory errors after 24 h are shown, as they are expected to be less affected by individual meteorological conditions than the errors after 10 days. The errors of the higher order integration schemes with a time step of 120 s are in the order of 80-200 m in the troposphere. The errors of the simulations in the stratosphere are about ten times smaller and the discrepancy between the error in the troposphere and stratosphere becomes even larger to a factor of about 25 when the global truncation errors after ten days are analyzed.

When analyzing the errors after 24 h simulation time we noticed that there was a bug in the plot script, resulting in errors that are five times too large. Only the curves for the AHTDs after 24 h were affected in the bug. We corrected the figure and the paragraph about the suggested time step.

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 large time steps for the calculations. In fact, even the results obtained with the longest time step of 1 h adhere to the error limit for the higher order methods as shown in Fig. 8.

'We attribute this to larger small-scale variations caused by atmospheric turbulence.' Remove erroneous reference to turbulence.

The erroneous reference was corrected.

Statistics not being sufficiently robust: as said before, it would be desirable to increase the sample size.

More simulations could be executed in another study, after the decision for an integration method and time step was made, which was the aim of this study.

30. *Page 15, line 79: The study of Seibert (1993) . . . . To achieve truncation errors that are smaller than overall trajectory uncertainty, they found that the time step should fulfill the CFL criterion as a necessary condition for convergence. The recommendation there for a sufficiently small truncation error was 15% of the time step needed for convergence of the Petterssen scheme. If we assume that the reference accuracy has also improved in the meantime, an even smaller value would result. The CFL criterion is recommended to make sure that no small-scale features are skipped, not for convergence of the iterations in the Petterssen scheme.* The authors have amended their wording, but they have not changed the first sentence quoted above, which is not an accurate representation of Seibert (1993), as explained in my previous comment.

We apologize for the wrong representation of your conclusions. We decided to omit this paragraph as we are still unsure if we fully understand the statement. The paragraph was replaced by a more general discussion of the used time interval of 3 h as mentioned in the answer to major remark #5.

31. OK

32. OK

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

Unfortunately, the data are only available on model levels and not on standard pressure levels. We added a note to the caption to remind the reader that the scales vary.

34. OK

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

Annotations will be added if they are requested during the formatting phase.

36. *Figures 5 and 6: This figure should be simplified. You don't need to show the two years separately, and I think you also don't need to show seasons separately. Thus you could have just three subfigures (three levels) and the five regions inside of each one. Then use a log scale for the AHTD, and symbols instead of bars (which will bring out the median also more clearly).*
'We would like to show the simulation results separately because regional and temporal impacts on the error were a part of the motivation for this study.' This is not sufficient as a justification the question is whether there are enough interesting and relevant results to convey. As pointed out earlier, because of the small sample size (just two times one starting date per season) this is questionable. Most of the seasonal results are just representing general knowledge about seasonality of atmospheric variability. I don't see a good reason for presenting the two years separately due to sample size, this is neither a good representation of year-to-year variability nor a replacement for error bars. Also, the year-to-year variability is not of interest here. Make your samples as large as possible, and then present one value for each!

We tried to summarize our reasons to show all samples in reply to comment #18.

'We added a horizontal line for the average error to all subfigures (see Figures 4 and 5).' That is useful, but I would not call this 'average of the domain' but rather 'annual mean' or 'average over all seasonal samples'.

We followed the suggestion:

The errors obtained in the polar regions are second largest with an average over all seasonal samples of around 200 km and peak errors in polar summer of up to 380 km.

'We did not use a log scale, because it would hide the seasonal and regional differences.' It would not hide them, even though they would be less strongly visible. It is not a good

practice to present a set of figures where for the majority of them, more than 50% of the graph area is unused. Also, the stratospheric results are hard to see at all. If you don't want to use a log scale, at least you should optimise them and use different scales for each level.

We used identical scales to underline the differences between the regions, but we agree that readability is more important and optimized the scale for each subfigure. We think that the very small median values as depicted in Fig. 5 do not make it necessary to provide actual values to the reader. Therefore, we decided to keep the linear scale.

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

We increased the linewidth further which also reduced the problem with pale colors.

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

We included a paragraph about the most efficient configuration for different numbers of parcels:

The fastest simulations for a set of about 10^2 parcels are possible with 4 cores. 12 cores should be used when 10^3 parcels are simulated. For 10^4 parcels the simulations with 24 cores are fastest. For 10^5 or more parcels all 48 cores (which includes SMT) should be used.

39. OK

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.

Brioude, J., Angevine, W., McKeen, S., and Hsie, E.-Y.: Numerical uncertainty at mesoscale in a Lagrangian model in complex terrain, Geoscientific Model Development, 5, 1127–1136, 2012.

CDO: Climate Data Operators. Available at: http://www.mpimet.mpg.de/cdo, 2015.

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., Grimsdell, A. W., and Alexander, M. J.: Stratospheric gravity waves at southern hemisphere orographic hotspots: 2003 – 2014 AIRS/Aqua observations, Atmos. Chem. Phys. Discuss., 2016, doi: 10.5194/acp-2016-341, 2016a.

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, 2016b.

Hoffmann, L., Hertzog, A., Rößler, T., Stein, O., and Wu, X.: Intercomparison of meteorological analyses and trajectories in the Antarctic lower stratosphere with Concordiasi superpressure balloon observations, Atmospheric Chemistry and Physics, 17, 8045, 2017.

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.

~~Domain specific trajectory~~ Trajectory errors diagnosed with the MPTRAC advection module ~~and~~ driven by ECMWF operational analyses

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) depends on various factors. The optimization of numerical integration schemes used to solve the trajectory equation helps to maximize the computational efficiency of large-scale LPDM simulations. We analyzed ~~truncation errors and total errors~~ global truncation errors of six explicit integration schemes of the ~~Runge-Kutta~~ Runge-Kutta family, which we implemented in the Massive-Parallel Trajectory Calculations (MPTRAC) advection module. The simulations were driven by wind fields 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~~ regions 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 mid stratosphere. 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. We quantified the accuracy of the trajectories by calculating transport deviations with respect to reference simulations using a fourth-order Runge-Kutta integration scheme with a sufficiently fine time step. ~~We assessed the transport deviations~~ Transport deviations were assessed with respect to error limits based on turbulent diffusion. Independent of the numerical scheme, the ~~total~~ global truncation errors vary significantly between the different ~~domains~~regions. 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 summarize, we recommend the 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 ten days simulation time for the specific ECMWF high-resolution data set considered in this study. Purely stratospheric simulations can use significantly larger time steps of 1100 s and 800 s for the midpoint scheme and the third-order Runge-Kutta method, respectively.

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 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) local and global truncation errors that originate from the use of an approximate numerical scheme to integrate the kinematic equation of motion in time ~~.~~ at a single time step or integrated over multiple time steps, respectively.

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 local truncation errors can be achieved by reducing the size of the time step of the numerical integration scheme, which will also reduce the global truncation errors. The size of the time step is usually the most important factor that controls the trade-off between numerical accuracy and computation time. ~~However, the~~ An 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. 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~~ and 180 – 750 m vertical resolution at 2 – 32 km altitude~~, and are nominally provided at 3h synoptic time intervals~~. We estimated the ~~total simulation errors and the~~ global truncation errors of the numerical methods for five latitude bands and three altitude ranges of the atmosphere, covering the free troposphere, the upper troposphere and lower stratosphere (UT/LS) region, and the mid stratosphere. 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. Using most recent meteorological data, the results will be of interest for many current and future LPDM studies using ECMWF operational data or other meteorological data sets with comparable resolution.

In Sect. 2 we present the advection module of 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 error growth rates and ~~domain specific~~ region-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; Wu et al., 2017). An intercomparison of meteorological analyses and MPTRAC trajectory calculations with superpressure balloon observations in the Antarctic lower stratosphere was presented by Hoffmann et al. (2017) . The primary task of MPTRAC is to solve the kinematic equation of motion for atmospheric air parcels. ~~The 'advection module '~~ Its advection module calculates air parcel trajectories based on given wind fields. 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 modules can simulate the sedimentation of air parcels and the decay of particle mass. 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 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 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 up to 0.01 hPa. For usage with MPTRAC, the wind fields ~~have been interpolated horizontally to~~ were retrieved on model levels with a longitude-latitude grid with $0.125° \times 0.125°$ resolution and have been interpolated vertically to 114 pressure levels in the troposphere and stratosphere up to 5 hPa with help of the Climate Data Operators (CDO, 2015). 12-hourly analyses ~~are~~ were combined with short-term forecasts in between to obtain data with a ~~3-hour~~ 3-hourly time step. Hoffmann et al. (2016) showed that ~~this data set outperforms existing~~ ECMWF operational analyses and forecasts outperform current reanalysis data products in terms of transport deviations for simulations of volcanic sulfur dioxide emissions in the upper troposphere and stratosphere. Hoffmann et al. (2017) showed the same for the trajectory calculations for superpressure observations in the Antarctic lower 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. 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 $100\,\mathrm{m\,s^{-1}}$ 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 or trajectory equation,

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

Here $\boldsymbol{x} = (x, y, z)$ denotes the position and $\boldsymbol{v} = (u, v, \omega)$ 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 $\boldsymbol{v}$ at any position $\boldsymbol{x}$ 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

$$\boldsymbol{x}(t_1) = \boldsymbol{x}_0 + \int_{t_0}^{t_1} \boldsymbol{v}\left(\boldsymbol{x}(t), t\right) dt, \tag{2}$$

with initial position $\boldsymbol{x}_0$ 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 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

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \Delta t\, \boldsymbol{v}\left(\boldsymbol{x}_n, t_n\right), \tag{3}$$

where $\Delta t = t_{n+1} - t_n$ refers to the time step. The Euler method is a 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,

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \Delta t\, \boldsymbol{v}\left(\boldsymbol{x}_n + \frac{\Delta t}{2} \boldsymbol{v}\left(t_n, \boldsymbol{x}_n, t_n\right), t_n + \frac{\Delta t}{2}\right). \tag{4}$$

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 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

$$\boldsymbol{x}_{n+1,0} = \boldsymbol{x}_n + \Delta t\, \boldsymbol{v}\left(\boldsymbol{x}_n, t_n\right), \tag{5}$$

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

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 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 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. An additional test was conducted with a fixed number of two iterations to assess possible improvements of the Petterssen scheme with respect to Heun's method.

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

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \Delta t\left(\frac{1}{6}\boldsymbol{k}_1 + \frac{4}{6}\boldsymbol{k}_2 + \frac{1}{6}\boldsymbol{k}_3\right), \tag{7}$$

with wind vectors k_i at the specific quadrature nodes,

$$\boldsymbol{k}_1 = \boldsymbol{v}\left(\boldsymbol{x}_n, t_n\right), \tag{8}$$

$$\boldsymbol{k}_2 = \boldsymbol{v}\left(\boldsymbol{x}_n + \frac{\Delta t}{2}\boldsymbol{k}_1, t_n + \frac{\Delta t}{2}\right), \tag{9}$$

$$\boldsymbol{k}_3 = \boldsymbol{v}\left(\boldsymbol{x}_n - \Delta t\,\boldsymbol{k}_1 + 2\,\Delta t\,\boldsymbol{k}_2, t_n + \Delta t\right). \tag{10}$$

The classical fourth-order Runge-Kutta method is defined by

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \Delta t\left(\frac{1}{6}\boldsymbol{k}_1 + \frac{2}{6}\boldsymbol{k}_2 + \frac{2}{6}\boldsymbol{k}_3 + \frac{1}{6}\boldsymbol{k}_4\right), \tag{11}$$

with wind vectors

$$\boldsymbol{k}_1 = \boldsymbol{v}\left(\boldsymbol{x}_n, t_n\right), \tag{12}$$

$$\boldsymbol{k}_2 = \boldsymbol{v}\left(\boldsymbol{x}_n + \frac{\Delta t}{2}\boldsymbol{k}_1, t_n + \frac{\Delta t}{2}\right), \tag{13}$$

$$\boldsymbol{k}_3 = \boldsymbol{v}\left(\boldsymbol{x}_n + \frac{\Delta t}{2}\boldsymbol{k}_2, t_n + \frac{\Delta t}{2}\right), \tag{14}$$

$$\boldsymbol{k}_4 = \boldsymbol{v}\left(\boldsymbol{x}_n + \Delta t\,\boldsymbol{k}_3, t_n + \Delta t\right). \tag{15}$$

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 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 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. To calculate the horizontal distances we first converted the spherical coordinates of the air 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~~ Sensitivity tests using variable time steps down to 30 s showed that the numerical ~~solutions~~ solution from the RK4 method ~~converge~~ converges for time steps of 60 s or less~~.~~, in the sense that transport deviations relative to simulations with a time step of 120 s do not change

significantly. 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 for up to 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 \approx$16 km and a maximum horizontal wind speed of $u_{max} \approx$120 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 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 $D_z \approx 1$ m^2 s^{-1} in the free troposphere (Pisso et al., 2009) and $D_z \approx 0.1$ 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 (Pisso et al., 2009), corresponding horizontal diffusivity coefficients are $D_x \approx 40\,000$ m^2 s^{-1} in the troposphere and $D_x \approx 4000$ m^2 s^{-1} in the stratosphere. The corresponding horizontal spread after ten days is $\sigma_x \approx 260$ km in the troposphere and $\sigma_x \approx 85$ km in the stratosphere. The vertical spread is $\sigma_z \approx 1300$ m in the troposphere and $\sigma_z \approx 415$ 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 errors of trajectory calculations in 15 ~~domains~~ regions of the atmosphere, covering rather distinct conditions in terms of pressure, temperature, and winds. The globe was divided into 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 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~~ regions are of major interest regarding various applications of transport simulations using MPTRAC and other LPDMs. The planetary boundary layer was not considered here, because MPTRAC lacks a 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 ten days. In each ~~domain~~ region 500,000 trajectory seeds were uniformly distributed. Although this is already a large number of trajectory seeds, this is still undersampling ~~as~~ of the effective resolution of the ECMWF data by as much as a factor of 4.5 in the polar troposphere and 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~~ regions 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, consisting of more than 2.5×10^9 individual trajectories.

Here the atmospheric ~~domains~~ regions 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 our approach is that part of the air parcels leave their initial ~~domain~~ region during the course of simulation. Table 1 provides the fraction of air parcels that remain in their initial ~~domain~~ region after five and ten days simulation time. In the stratosphere we found fractions of 48 – 88% after five days and 36 – 78% after ten days in the different latitude bands. In the UT/LS region the fractions are lower, i. e., 32 – 55% after five days and 14 – 40% after ten days. In the troposphere the fractions are even lower, i. e., 32 – 48% after five days and 10 – 24% after 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~~ fluctuations in the wind field 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~~ regions are more likely related to higher wind ~~and~~ velocities. Excluding those trajectories would cause a strong bias towards short trajectories, representing only the lower wind velocities in the statistical analysis.

3 Results

3.1 Case studies of trajectory calculations

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 eight days. The Petterssen scheme and Heun's method yield AHTDs $\leq$ 200 km and AVTDs $\leq$ 800 m for about 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 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 three days by the Euler solution and after four to six days by the solutions from Heun's method and 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 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 second-order methods. However, this is not valid in general, we also found counter-examples with the midpoint method performing worse than the other 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 trajectory errors

In this section we discuss the temporal growth rates of the 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 errors using a fixed time step of 120 s for the numerical integration as a representative example. As the magnitude of the 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~~ seasons and latitude bands in the given altitude range. A more detailed analysis of the 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 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 first-order scheme. After 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 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 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. Any high-order numerical integration schemes is not expected to provide 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 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 one. For the troposphere we found a maximum polynomial order of ≈ 3 after five days for the AHTDs and of an order ≈ 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 ≈ 5 after eight days for the AHTDs and of ≈ 4 after six days for the AVTDs for the RK3 and RK4 method. The second-order methods are in between. Due to this non-linear error growth, the error growth rates also increase rapidly over time until they reach their maxima after 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~~ better assess the influence of individual atmospheric conditions on the trajectory errors, we will in the following discriminate between ~~truncation errors (analyzed after one day) and total calculation errors (analyzed after ten days)~~errors after one and ten days,respectively.

3.3 Spatial and temporal variations of trajectory errors

For a more detailed analysis of the regional and seasonal variations of the total trajectory errors we focus on the errors after ten days simulation time for simulations using the 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 variations. We calculated individual transport deviations for all 15 altitude-latitude ~~domains~~ regions 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 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 ten days. However, as can be seen from Figs. 5 and 6, the total errors can vary considerably seasonally and from year to year as well as between the ~~domains~~regions, causing maximum errors of specific ~~domains~~ regions 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~~ fluctuations of the wind field become stronger and more frequent at lower altitudes ~~. Wind~~ causing wind speeds and wind directions ~~can vary strongly in turbulent regions~~to vary more strongly. So, even if travel distances in the troposphere may be relatively short, transport deviations typically increase with decreasing altitude.

~~Calculation~~ The trajectory 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~~ largest trajectory errors in the troposphere are found at northern mid-latitudes with errors between 245 ~~km~~ and 470 km. ~~Tropospheric~~ The meteorological conditions in 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~~ stronger fluctuations are expected in the northern mid-latitudes ~~meteorological systems is more difficult to simulate than for~~ compared to the southern mid-latitudes due to the larger land-sea ratio and more complex orography of the northern hemisphere. The errors obtained in the polar regions are second largest with ~~average errors~~ an average over all seasonal samples of around 200 km and peak errors in polar summer of up to 380 km. The simulations for the tropics and southern mid-latitudes ~~have small~~ show smaller errors of less than 200 km and adhere to the error limit in all test cases. The simulations for the UT/LS region ~~has its~~ have their 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~~ small-scale variability of the wind field in this region. The second largest errors of the simulations in the UT/LS region occur in the tropics with about 75 km on average followed by the Arctic and southern mid-latitudes with about 50 km on average. ~~Antarctica exhibits~~ Simulations that cover the Antarctica show the smallest errors in ~~this altitude layer~~ the UT/LS region with about 30 km on average. ~~Errors~~ The errors of the simulations in the stratosphere are ~~generally small, with some larger average~~ typically below 25 km, except for January 2015. Stratospheric trajectory errors in the tropics ~~. All test cases show errors between 2km and 25km, which are mostly much~~ are 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 16km, since the tropopause~~ regions, which is probably due to the close vicinity of the tropical tropopause, which 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 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 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 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 ten days of simulation, which weakens our statistics.

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~~regions. 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~~ is somewhat larger for simulations in the troposphere, where particle paths are more likely being affected by ~~atmospheric turbulence~~synoptic-scale fluctuations of the wind field.

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 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~~ computation 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)~~ (Jülich Supercomputing Centre, 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 multi-

threading (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 of the advection module using the midpoint scheme with a time step of 120 s for different numbers of particles and OpenMP threads. 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 (less or equal to 10^4) the minimum computing time is limited by a constant offset of 6.3×10^{-5} s to 4.3×10^{-3} s (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~~ advection module of MPTRAC 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~~ SMT 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~~ by the overhead of the OpenMP parallelization and ~~load imbalances~~by load imbalances, which can also become significant for larger numbers of parcels if SMT is not enabled for all cores jointly. The fastest simulations for a set of about 10^2 parcels is possible with 4 cores. 12 cores should be used when 10^3 parcels are simulated. For 10^4 parcels the simulations with 24 cores are fastest. For 10^5 or more parcels all 48 cores (which includes SMT) should be used. LPDM simulations will typically use large numbers of particles (more than 10^5) to obtain statistically significant results. MPTRAC ~~would~~ will not be affected by significant 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. Note that the hardware, especially the memory cache, affects the six integration schemes differently. A single call to the wind interpolation function is up to 50% cheaper for a higher order method compared to Euler's method, because the cache is used more efficiently. Results are shown separately for the troposphere and stratosphere after 24 h and ten days. ~~Tropospheric and stratospheric errors are quite similar~~ The trajectory errors after 24 h ~~, indicating that these errors are most representative for the truncation~~ are shown, as they are expected to be less affected by individual meteorological conditions than the errors after 10 days. The errors of the ~~integration methods. In contrast, analyzing the total~~ higher order integration schemes with a time step of 120 s are in the order of 80-200 m in the troposphere. The errors of the simulations in the stratosphere are about ten times smaller and the discrepancy between the error in the troposphere and stratosphere becomes even larger to a factor of about 25 when the global truncation errors after ten days ~~, the~~ are analyzed. The errors in the troposphere using a time step of 120 s are in the order of 200-350 km after 10 days, which shows the nonlinear error

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

growth and the large impact of the atmospheric conditions on trajectory errors. 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 computational efficiency because of its low accuracy. The second-order methods as well as the RK3 and RK4 methods yield much smaller truncation errors, in particular for short time steps. Among the 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 accuracy of the Petterssen scheme with one iteration (Heun's method) is somewhat worse than the midpoint method. When two iterations of the Petterssen scheme are computed, the transport deviations are closer to those obtained with the midpoint method. The best efficiency, which we define as lowest computational costs when adhering to our error limit, is mostly obtained with the midpoint and RK3 methods. The additional iterations of the Petterssen scheme improve the accuracy, but they are too computational expensive for our model. In general, a well defined convergence limit for the number of iterations is needed for an efficient application. 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 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 2100s for the Petterssen scheme, Heun's scheme, and the midpoint method, and about 3500~~In fact, even the results obtained with the longest time step of 1 ~~s for the RK3 and RK4 methods for both the troposphere and stratosphere from~~ h adhere to the error limit for the higher order methods as shown in 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 truncation errors and total errors of trajectory calculations~~ global truncation errors after one and ten days 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 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 first-order Euler method, (ii) the 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, each consisting of 500,000 trajectories, we further analyzed horizontal and vertical transport deviations in relation to altitude, latitude, as well as seasonal and year-to-year variability. ~~After~~ The trajectory errors after 24 h ~~the trajectory errors are quite similar~~ were analyzed as they are expected to be less affected by individual flow patterns. The errors of the simulations in the troposphere ~~and stratosphere and mostly affected by truncation errors~~ have ten times larger errors compared to the simulations in the stratosphere. After ten days the trajectory errors vary ~~substantially in different domains~~ more substantially inside the climatological regions because of the ~~influence of~~ stronger influence of individual 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~~in the high-resolution meteorological input data. Calculation errors also depend on the latitude band, with the northern mid-latitudes having the largest errors in each altitude layer. Seasonal 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 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 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 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 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. To achieve sufficiently small truncation errors they recommended to use 15of the time step needed for convergence of the Petterssen scheme. Here we used meteorological data with a much finer spatial resolution as provided by current global weather forecast models, which~~ This study uses up to date meteorological data as provided by current global weather forecast models, with a spatial resolution that

is much finer than in former trajectory studies (Seibert, 1993; Stohl and Seibert, 1998; Stohl et al., 2001; Harris et al., 2005) . Previous studies suggest that a time interval of 3 h for the meteorological data may be too large for the fine spatial resolution (Stohl et al., 1995; Brioude et al., 2012; Bowman et al., 2013) . Pisso et al. (2010) found that using hourly wind fields in their ensemble LPDM reconstructions did not add significant information on the large scale atmospheric dynamics and suggested to use three-hourly wind fields for ECMWF operational data in T511 spectral resolution (corresponding to a horizontal resolution of about 40 km). Given the higher spatial resolution of meteorological data and the short integration time steps used here, it would be interesting to study the effect of shorter time intervals for the driving data on the trajectory errors, however for the time being we decided to restrict ourselves to the temporal resolution as specified in the user guide to ECMWF forecast products even if finer data is available.

The high resolution requires adjustment of the time step~~. Time~~ , as the commonly used time steps of 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 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 steps of 100 s for the midpoint method and 170 s for the RK3 method provide accurate results 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 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 errors found here suggests that applications may benefit from more advanced numerical techniques. Adaptive quadrature 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 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).

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 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.

Brioude, J., Angevine, W., McKeen, S., and Hsie, E.-Y.: Numerical uncertainty at mesoscale in a Lagrangian model in complex terrain, Geoscientific Model Development, 5, 1127–1136, 2012.

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.

CDO: Climate Data Operators. Available at: http://www.mpimet.mpg.de/cdo, 2015.

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.

Hoffmann, L., Hertzog, A., Rößler, T., Stein, O., and Wu, X.: Intercomparison of meteorological analyses and trajectories in the Antarctic lower stratosphere with Concordiasi superpressure balloon observations, Atmospheric Chemistry and Physics, 17, 8045, 2017.

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.

Jülich Supercomputing Centre: JURECA: General-purpose supercomputer at Jülich Supercomputing Centre, Journal of large-scale research facilities, 2, doi:10.17815/jlsrf-2-121, http://dx.doi.org/10.17815/jlsrf-2-121, 2016.

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.

Pisso, I., Marécal, V., Legras, B., and Berthet, G.: Sensitivity of ensemble Lagrangian reconstructions to assimilated wind time step resolution, Atmospheric Chemistry and Physics, 10, 3155–3162, doi:10.5194/acp-10-3155-2010, 2010.

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.

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.

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 in initial regions during the course of 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)

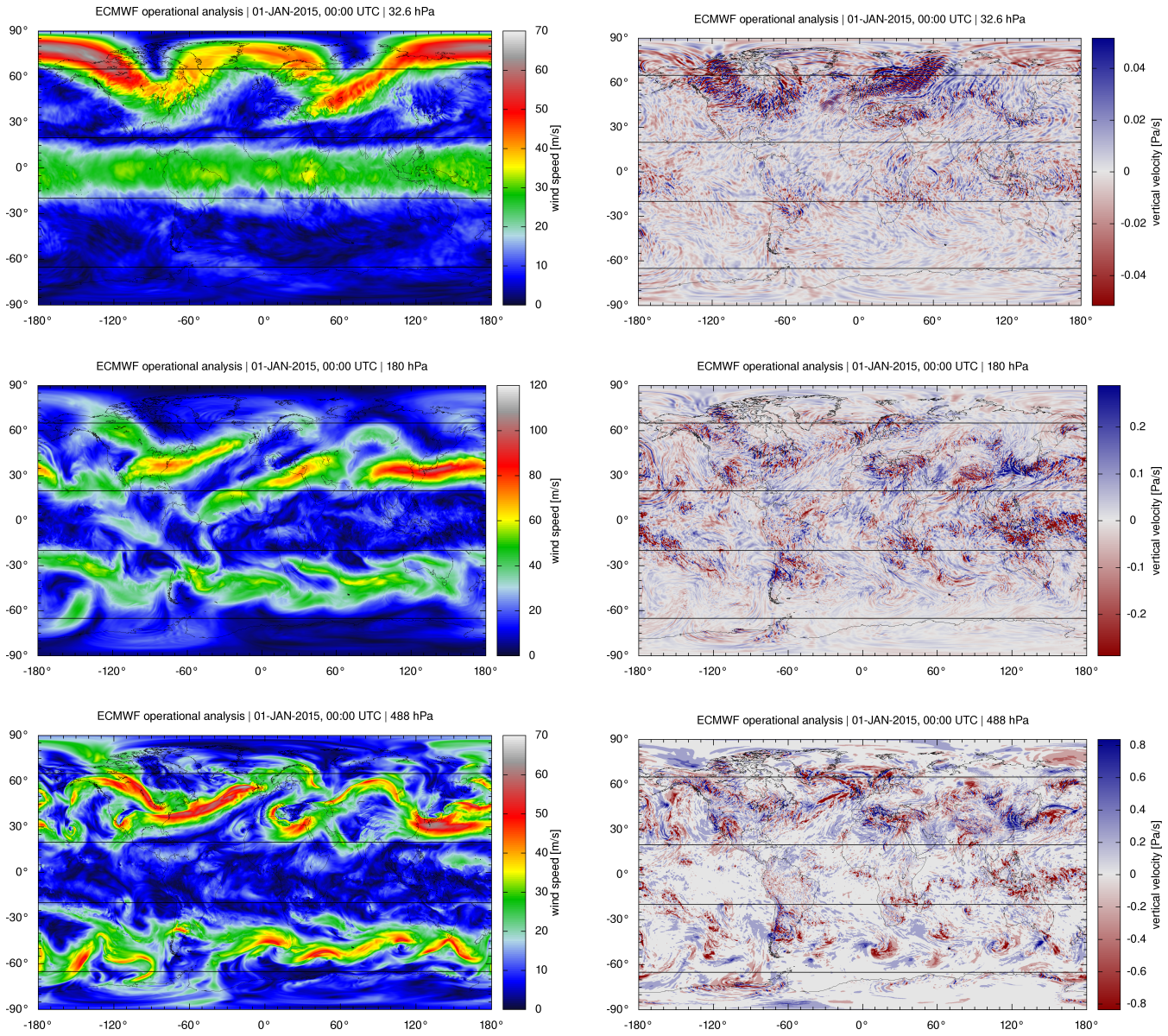

Figure 1. ECMWF operational analysis horizontal wind speed (left) and vertical velocity (right) at about 24 km (top), 12 km (middle), and 5 km (bottom) altitude on 1 January 2015, 00:00 UTC. Black lines indicate the latitude bands considered in our analysis. Note that color bar ranges have been adjusted for each height level, to make local structures visible.

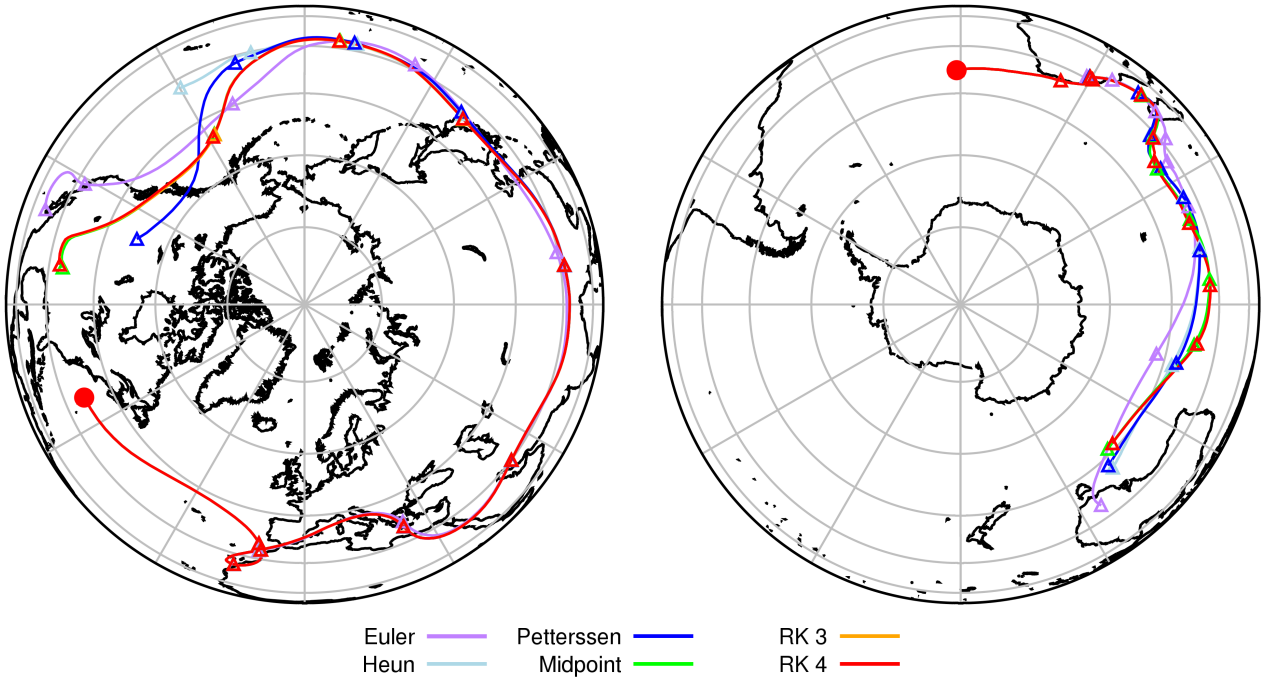

Figure 2. Examples of trajectory calculations using different numerical integration schemes. Circles mark the start positions of the trajectories. Trajectories were launched at an altitude of 10.8 km (left) and 9.7 km (right). The start time is 1 January 2014, 00:00 UTC for both. Triangles mark trajectory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

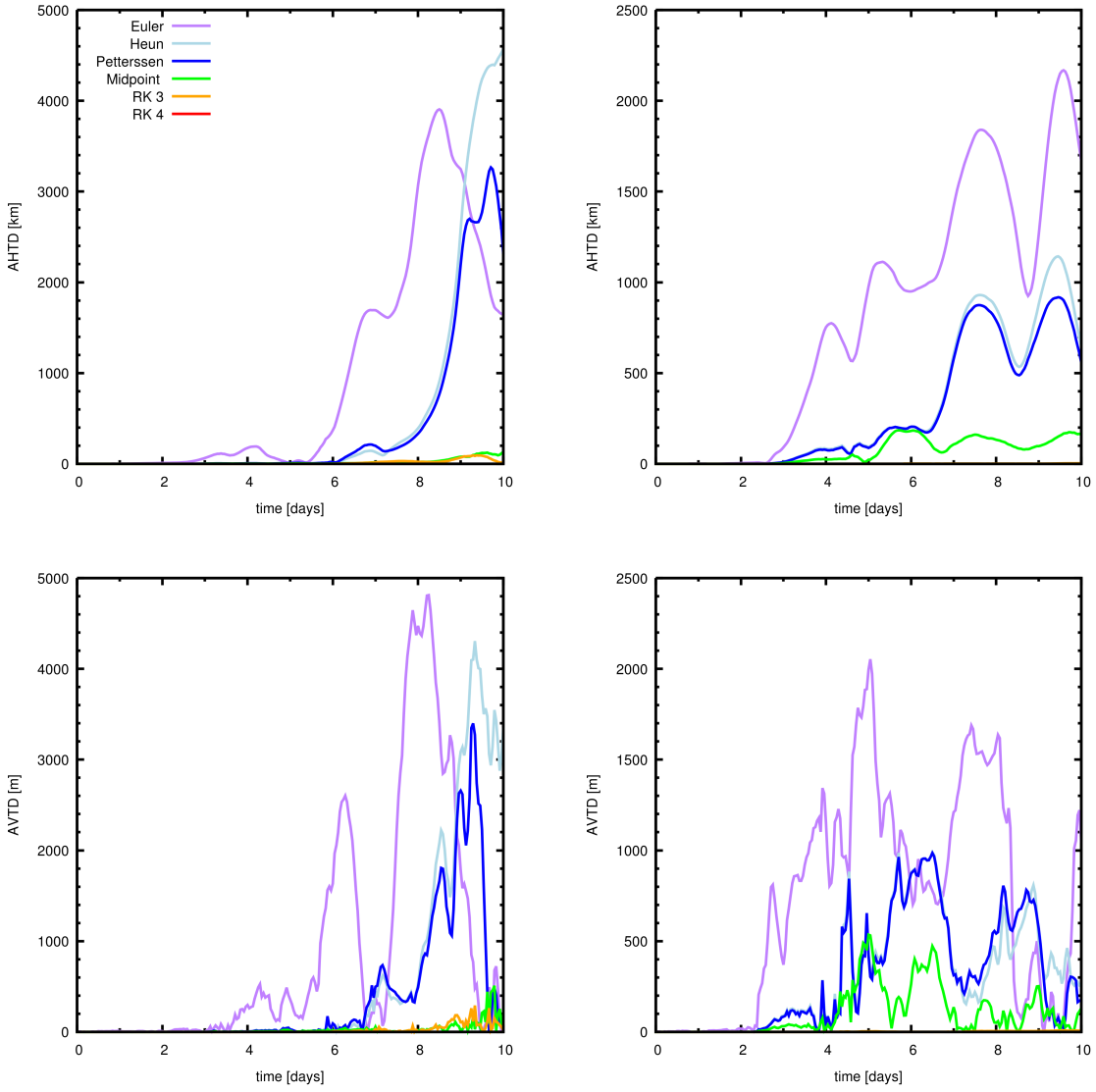

Figure 3. Absolute horizontal (top) and vertical (bottom) transport deviations for the case studies for the northern hemisphere (left) and southern hemisphere (right) presented in Fig. 2. Please note different ranges of y-axes. Results of the RK3 and RK4 method are close to zero in most cases.

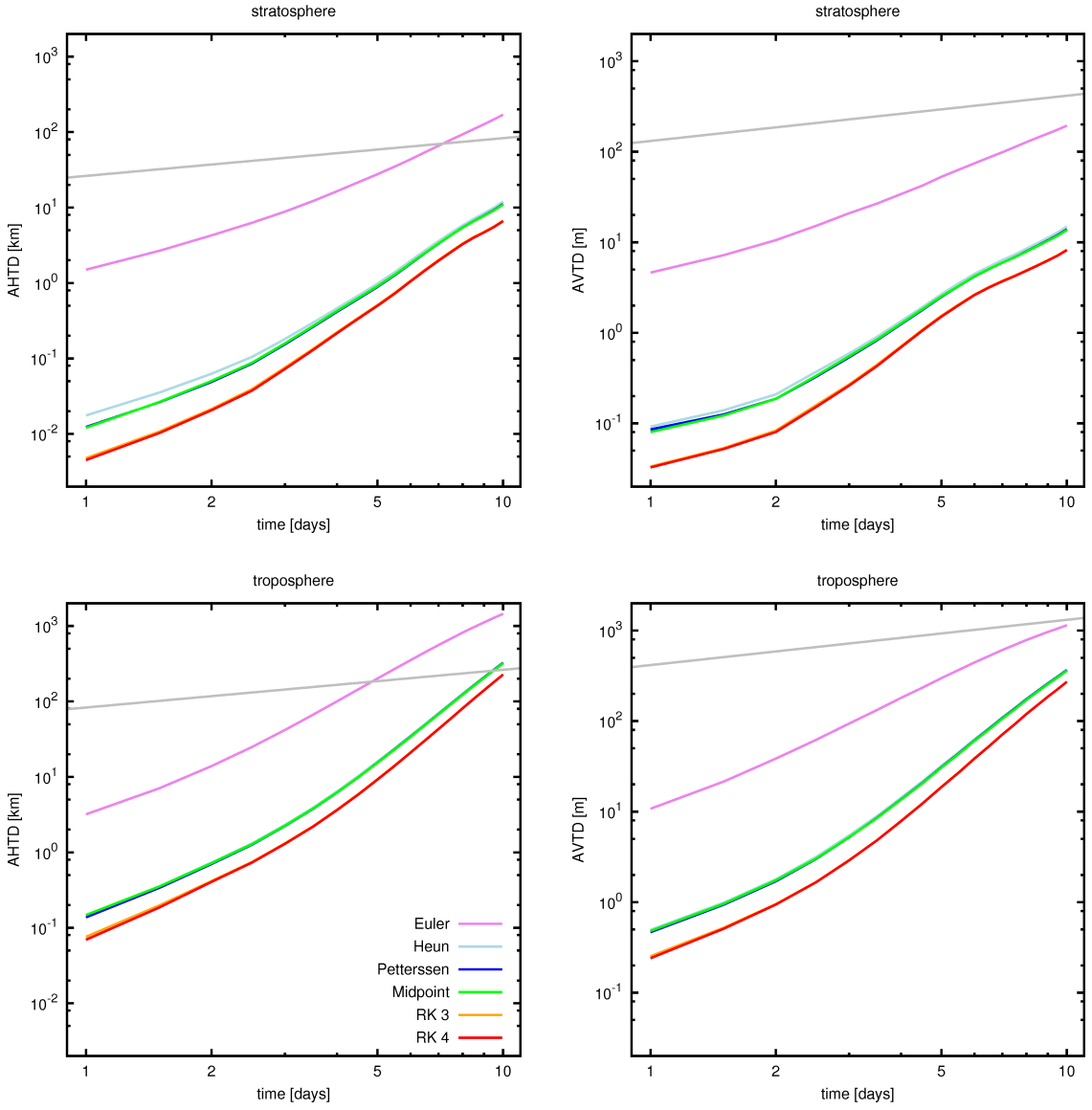

Figure 4. Absolute horizontal (left) and vertical (right) transport deviations of trajectory calculations for the stratosphere (top) and troposphere (bottom). The trajectory calculations are based on different numerical schemes, but use the same time step ($\Delta t = 120\,\mathrm{s}$). Grey lines show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

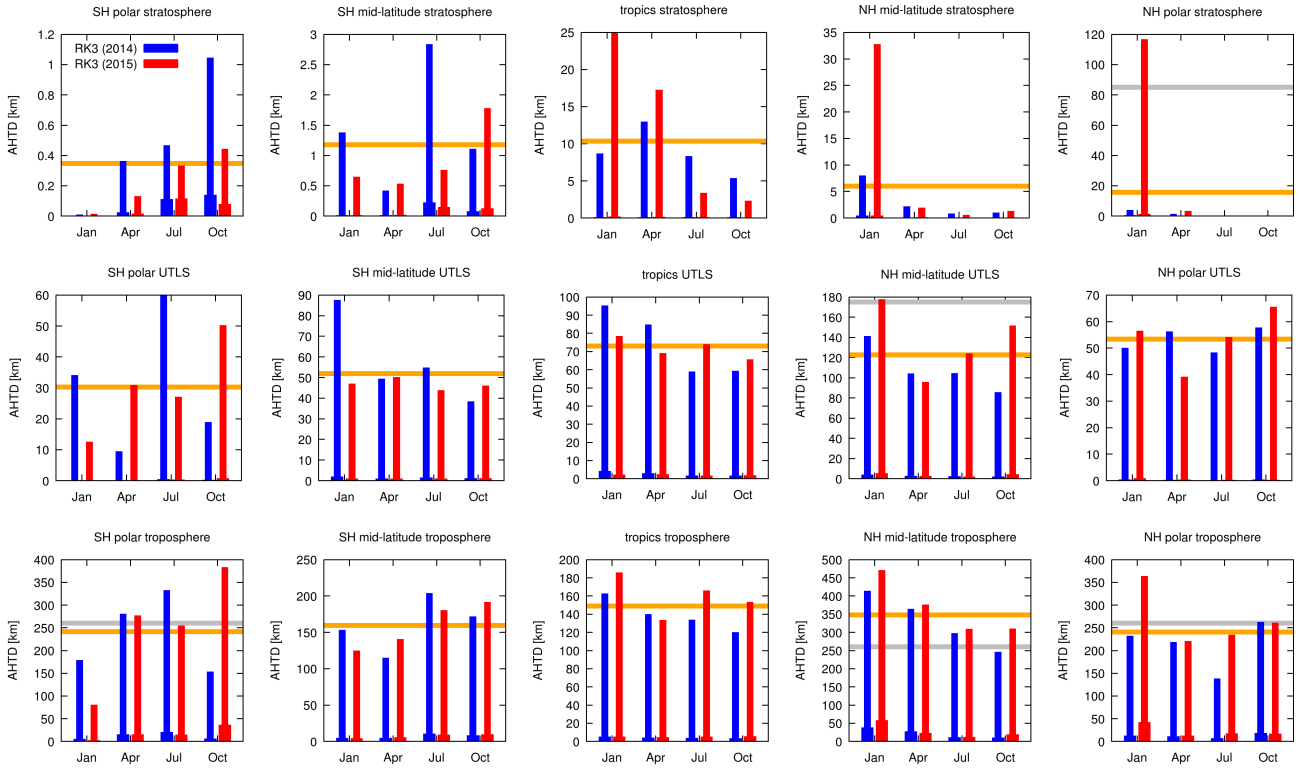

Figure 5. Mean (thin bars) and median (thick bars) horizontal transport deviations after ten days simulation time in different ~~domains~~ regions for the RK3 method and 120 s time step. Orange lines show the averages of the four months (January, April, July, and October) and both years (2014 and 2015).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

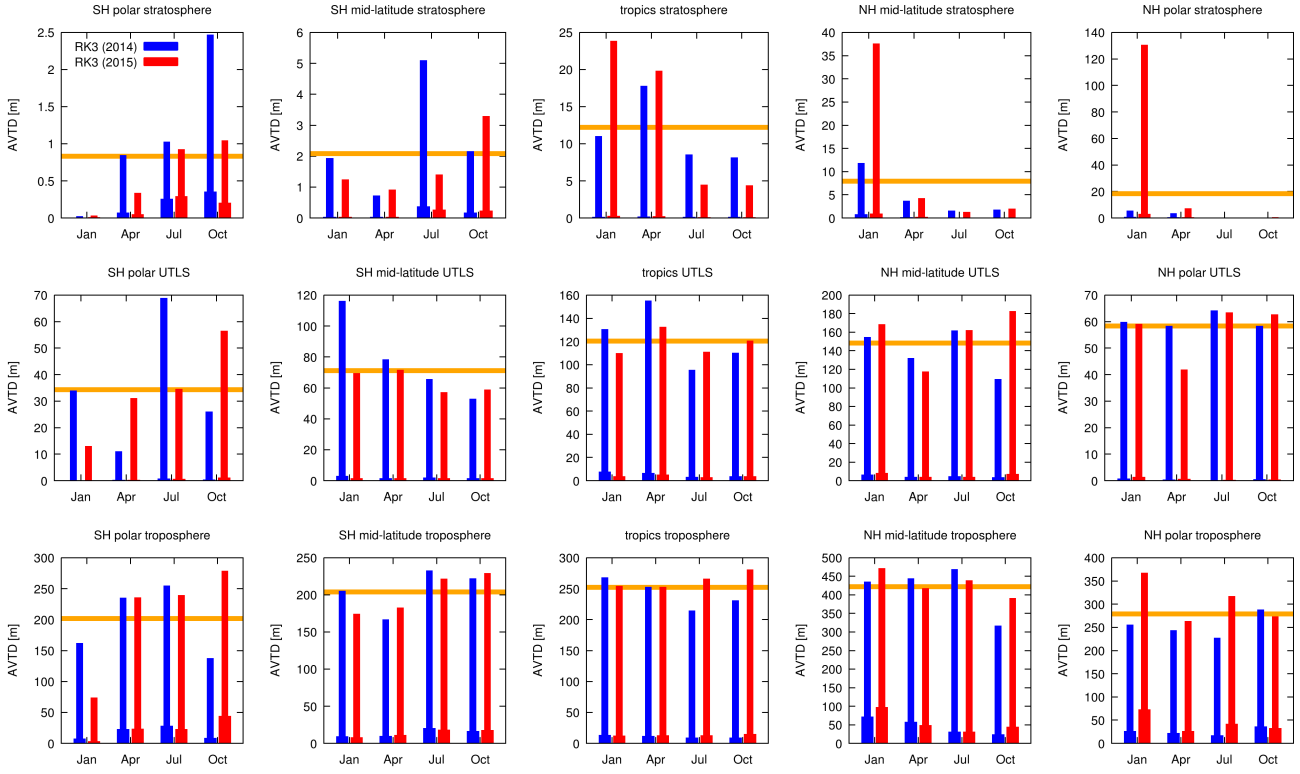

Figure 6. Same as Fig. 5, but for vertical transport deviations. Error limits based on diffusion are about 1300 m for the troposphere and 415 m for the stratosphere, which is beyond the AVTD ranges shown here.

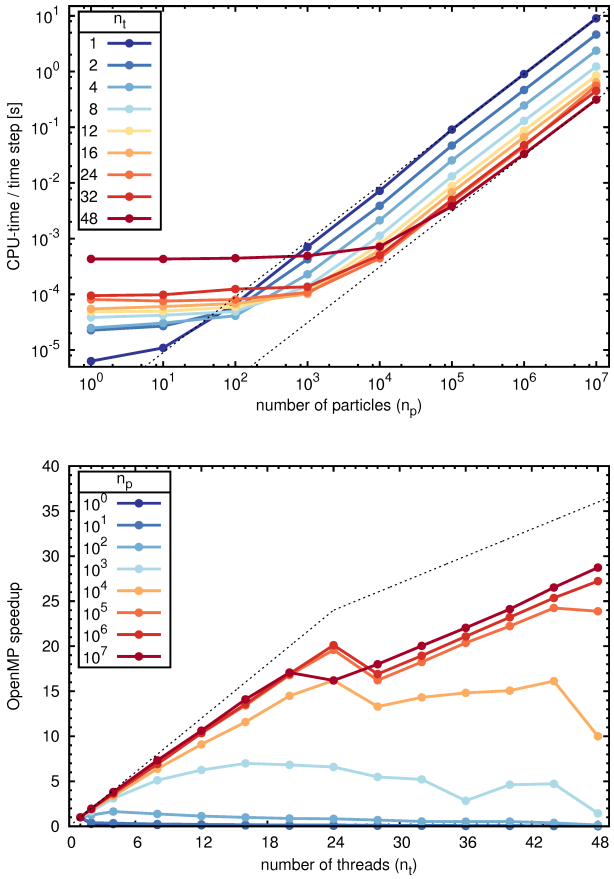

Figure 7. Scaling ~~behaviour~~ behavior in terms of CPU-time (top) and speedup of the code (bottom) used to calculate trajectories with the midpoint method and a time step of 120 s for different numbers of particles (n_p) and OpenMP threads (n_t). Colored curves refer to different numbers of OpenMP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

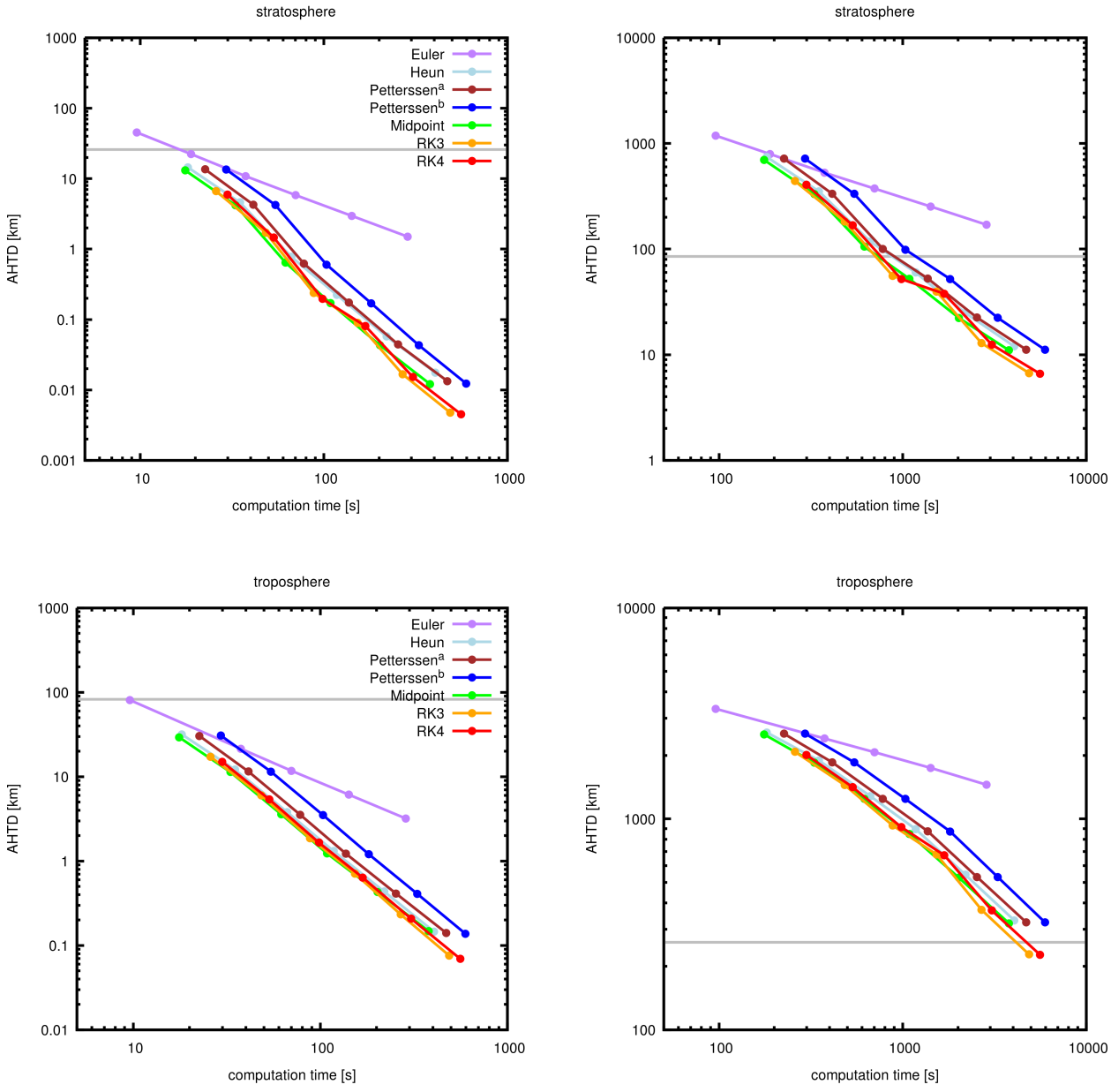

Figure 8. Trade-off between computational accuracy and total CPU-time requirements of the trajectory calculations after 24 h (left) and 10 days (right). Colored curves refer to different integration schemes. Dots along the curves indicate time steps of 3600, 1800, 900, 480, 240, and 120 s (from left to right). Horizontal lines refer to the maximum tolerable error limits as defined in Sect. 2.5. ~~Note that our implementation of the~~ The Petterssen[a] scheme ~~was optimized for numerical accuracy rather~~ always uses two inner iterations, which is one more than ~~speed~~for Heun's method. The Petterssen[b] scheme uses up to seven iterations with a strict convergence criteria.