**Fachhochschule Aachen**
Campus Jülich

Fachbereich: Medizintechnik und Technomathematik
Studiengang: Scientific Programming

FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

# Optimization and validation of atmospheric advective and diffusive transport simulations

**Bachelorarbeit**

*Autor:*
Thomas Franz-Peter Rößler

*Prüfer:*
Prof. Dr.-Ing. Andreas Terstegge
Dr. Lars Hoffmann

Jülich, 12. August 2015

**Eigenständigkeitserklärung**

Diese Arbeit ist von mir selbstständig angefertigt und verfasst. Es sind keine anderen als die angegebenen Quellen und Hilfsmittel benutzt worden.

| | |
|---|---|
| Ort, Datum | Thomas Rößler |

Diese Arbeit wurde betreut von:

1. Prüfer: Prof. Dr.-Ing. Andreas Terstegge
2. Prüfer: Dr. Lars Hoffmann

## Abstract

Lagrangian particle dispersion models are indispensable tools to study atmospheric transport processes based on the flow of individual air parcels. In operational uses cases they are used to simulate the spread of radionuclides or volcanic emissions in emergency situations. Different Lagrangian particle dispersion models have been developed for those studies, like FLEXPART or HYSPLIT. It is very important that the models are verified and their errors are estimated so that the results are reliable. In this thesis the advection and diffusion part of the new Lagrangian particle dispersion model Massive Parallel Trajectory Calculations (MPTRAC) was verified in idealized test cases and in real atmospheric conditions and optimized in terms of accuracy and performance.

Analytical test cases were applied to the model to validate the advection. The implementation and accuracy of integration schemes of different order, the error caused by the linear interpolation of the wind fields and the used coordinate system were discussed. All implementations lead to correct and very accurate results. A notable difference concerning the accuracy was only determined between the first order Euler method and other methods of higher order. Runge-Kutta methods of the order 2 to 4 and the Petterssen scheme led to very similar results, indicating that an order of larger than 2 can not increase the accuracy significantly. This is partly caused by the linear interpolation of the meteorological data. Even if tests have shown that the error caused by this linear interpolation is small, higher order methods may need a better interpolation to provide benefits.

Tests for real atmospheric conditions confirm these results, i.g., the midpoint method was found to be the most performant integration scheme. However, a small time step is required to yield small deviations from the reference solution. For the tests with real atmospheric conditions the atmosphere was separated into regions with similar conditions and the resulting transport deviations are analyzed in detail. Simulation errors in the stratosphere and in the tropics are significantly smaller than in the troposphere or at high latitudes.

The diffusion scheme of MPTRAC consists of three components, a random horizontal displacement, a random vertical displacement, and a mesoscale diffusion that depends on the variation of the wind field around an air parcel. All three components were tested individually without advection and produced results that correspond to analytical solutions. However, it is difficult to predict a solution under real atmospheric conditions, because the wind field has a significant influence. A sensitivity test was done, to get an overview of the impact of the diffusion components on the transport of air parcels. The mesoscale diffusion has the strongest influence, if default parameters are used, followed by the vertical diffusion. The horizontal diffusion has only a very small impact and does not seem to influence the transport simulations significantly.

In conclusion, the model MPTRAC can be used to simulate the advection and dispersion of air parcels correctly and efficiently. Simulations with altitudes below the free troposphere require relatively small time steps and the diffusion model is probably too simplistic, but the main task of the model are simulations in the free troposphere and stratosphere where the model shows good results.

# Contents

# Chapter 1

# Introduction

Lagrangian particle dispersion models are widely applied for atmospheric research and for operational use cases (Bowman et al., 2013). They simulate the flow of air parcels in the atmosphere and are, for instance, used to analyze atmospheric transport processes or the impact of events like forest fires (Forster et al., 2001) or volcanic eruptions (Maerker et al., 2008). They are also used in emergency simulations to simulate the spread of radionuclides or volcanic emissions (Heffter and Stunder, 1992) so that endangered people can be warned and flights can be redirected or canceled. These predictions have to be as accurate as possible to reduce economical loss (Mazzocchi et al., 2010).

The most important aspect in developing a Lagrangian particle dispersion model is its correctness. Models have to be verified so that their outcome can be trusted. This can be done by simulating the spread of aerosols or trace gases that were released in experiments or by natural events, for which observations are available. Another approach is to define test cases with analytical wind fields and compare simulations to solutions that were calculated with high accuracy (Walmsley and Mailhot, 1983) or analytically (Williamson and Rasch, 1989). These tests are used to check the implementation and estimate the models error in time. The error sources for different dispersion models and meteorological data products have been analyzed in many studies. They found that the average horizontal uncertainty in trajectory calculations is on the order of 100 km/day, but varies strongly under different conditions (Rolph and Draxler, 1990; Seibert, 1993). This uncertainty is caused by several error sources. One of the most important factors is the quality of the meteorological input data and its resolution. Especially the temporal resolution of the input files is an important error source, but in addition the resolution in space and time have to fit for good results (Stohl et al., 1995). The error caused by interpolation of the wind is smaller and can be reduced to every desired value by reducing the time step (Seibert, 1993).

The Simulation Laboratory "Climate Science" of the division "Computational Science" of the Jülich Supercomputing Centre has developed a new trajectory model called Massive Parallel Trajectory Calculations (MPTRAC) (Rößler, 2014; Hoffmann et al., 2015). The model was used, e.g., for reconstructing the initial emissions of trace gases, such as $SO_2$ and to simulate dispersion of volcanic emissions over the globe. In previous studies the model was validated by comparing satellite

observations of volcanic emissions to predictions. Since such a comparison is limited to individual events of volcanic eruptions and available observations, a more detailed analysis of the correctness of the model is done in the first part of this thesis. Therefore, test cases introduced by Williamson and Rasch (1989), which can be solved analytically, are applied to the model. Different numerical integration schemes are compared, to optimize the accuracy and performance of the calculation of the advection. Other error sources that can affect the accuracy, like interpolation errors or the choice of the coordinate system, are also analyzed. Furthermore, the accuracy of the model in different regions and seasons under real atmospheric condition is discussed, to obtain more general results. In the second part of this thesis the dispersion modules for the turbulent and mesoscale diffusion were analyzed and improved. Tests without advection were made to validate that the diffusion modules work correctly. Additionally a parameter study for the influence of different diffusion parameters was performed, to validate the current choice for cases with advection. In the end a performance and scaling analysis of the particle dispersion model as a whole is done. Performance improvements in the advection and diffusion module are also explained there.

# Chapter 2

# Methods and data

## 2.1 The Lagrangian particle dispersion model MPTRAC

### 2.1.1 Overview of the model

The Lagrangian particle dispersion model Massive Parallel Trajectory Calculations (MPTRAC) (Hoffmann et al., 2015) was developed to study atmospheric transport processes. There are other similar models available, e.g. FLEXPART (Stohl et al., 2005) or HYSPLIT (Draxler and Hess, 1998). The main reason why our new model was developed is that it is more suitable for large-scale simulations on supercomputers than existing modules. An own model can be customized easier to match the requirements of specific applications and studies.

A common usage of the model is to do ensemble simulations of volcanic eruptions with different initial conditions. This leads to sets of hundreds of simulations that all consider millions of parcels. Usual simulations are done with $10^6$ to $10^8$ air parcels and go on for many weeks. Therefore, parallel computing and a good performance is necessary. The model is kept as general as possible to allow simulations of different kinds. It can calculate forward and backward in time to predict the spread of emissions over the globe or to find out where the source of emissions is located. The simulated air parcels have optional properties such as a mass or other coefficients that are necessary to calculate sedimentation or decay. This allows us to model different kinds of parcels as required, e.g., gases like $SO_2$, radioactive dust or volcanic ash. Simulations can work with meteorological input data of different centers. In this study the ERA-Interim reanalysis data product provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) is used. More information about the meteorological data product is given in section 2.2. All physical processes are implemented in separate modules that consider advection, diffusion, decay, and sedimentation. They are described in the following section. This thesis concentrates on analyzing, testing and improving the advection and diffusion modules.

### 2.1.2 Physics

#### Advection

In meteorology “advection ”strictly refers to the horizontal transport of an air parcel, but in this case both horizontal and vertical movement of parcels with the wind is meant. MPTRAC does not calculate the wind itself, but uses external meteorological input data that contains all needed information about the atmosphere. In particular wind speed information in the zonal (west-east), meridional (north-south), and vertical direction are required. The meteorological data files are provided on a regular longitude-latitude grid in the horizontal and on fixed model layers in the vertical direction for discrete time steps. An interpolation in time and space is used to get the wind vector for the exact position of an air parcel at a model time step. A numerical integration scheme is applied that integrates the wind vector over a defined model time step and predicts the new position for the parcel. The time step in general controls performance and accuracy of the simulation, the smaller it is, the smaller are interpolation and integration errors, but the computational costs increase. It can be set for each simulation individually to allow an adjustment for the requested accuracy. Every time the advection is executed, the parcel's model time is increased by the time step. The advection scheme is called iteratively until the stop time is reached and the air parcels' flow has been fully computed. Meanwhile the parcel's position is written to disk periodically so that its trajectory is logged.

The trajectory model currently uses a linear interpolation in space and time and a second order integration scheme (midpoint method). Trajectory models in general use a low order interpolation method and a higher order integration scheme (Bowman et al., 2013), so the current compasses to other models. Five integration schemes of different orders are compared in the following and the integration error is analyzed, to find out if improvements of the current advection model can enhance the accuracy profitably.

#### Diffusion

Diffusion in general is the transport of mass from a region of high concentration to a region of low concentration. As a result, air parcels like dust, aerosols, or trace gases distribute uniformly in a medium with time. The diffusive mass transport of the parcels takes place without bulk motions like advection or convection, but is related to eddy mixing. Before the diffusion is described in detail an introduction to the structure of the atmosphere is given. The atmosphere is separated into multiple layers in height. The relevant layers for the model are described in the following from the bottom to the top.

The lowest layer is called atmospheric boundary layer (ABL) and extends from the Earth's surface to about 1.5 to 2 km in height. In this layer very strong fluctuations in velocity, moisture, temperature, and other physical quantities can be observed. Also the vertical mixing is very strong.

The ABL is the lowest part of the troposphere, which extends up to an average height of 9 km at the poles or 17 km at the equator. The part of the troposphere that is above the ABL is also called the free troposphere. The troposphere is the atmospheric layer where the greatest part of the weather takes place. Nearly all water vapor and clouds are located in this layer and convection occurs here. The temperature of the troposphere decreases with height. The decreasing rate is called environmental lapse rate (ELR) and is $10\,\mathrm{C}^\circ/\mathrm{km}$ in dry conditions and $6.5\,\mathrm{C}^\circ/\mathrm{km}$ in wet conditions, since the water vapor preserves the energy. When the so tropopause layer is reached, the temperature starts to grow with height.

The tropopause boundary layer marks the border to the stratosphere. Its height varies from place to place and from definition to definition. A more detailed description on how the tropopause is defined for the model is given in section 4.1. The stratosphere extends to a height of around 50 km. The stratosphere shows increasing temperature because it contains the ozone layer and absorbs ultraviolet radiation from the sun, which heats the layer up. The temperature profile of this layer suppresses vertical motion, turbulence, and mixing and creates very stable atmospheric conditions.

There are two basic approaches to describe diffusion. The first one is atomistic and statistical. It considers diffusion as a result of random walk of the individual parcels. Molecular diffusion is known as Brownian motion. The other approach is phenomenological and explains diffusion as direct movement from a region of high concentration to a region of low concentration. This behavior is defined by Fick's laws of diffusion. These laws imply that the diffusion flux is proportional to the negative gradient of concentrations, which means that it goes from regions of higher concentration to regions of lower concentration.

In both cases the diffusion takes place on the atomic or molecular level and, therefore, is only relevant for length scales between nanometers and millimeters. So the global transport of air parcels is not affected by this kind of diffusion directly. However, so called stirring processes cause the clouds of air parcels to form thin filaments or sheets. This deformation reduces the distance of the different air masses and the molecular diffusion is sped up. Nevertheless, another kind of diffusion, called turbulent diffusion, is more relevant on the global scale. Turbulent diffusion occurs when a system reaches critical conditions due to shear flow. It is caused by high velocities and steep concentration or density gradients. For example wind shears, a difference in relative speed between two adjacent air masses, can cause the so called clear-air turbulence that occurs near the tropopause and is hazardous to the comfort and even the safety of air travelers.

Most studies about dispersion modeling focus on the ABL, because this lowest layer of the atmosphere is where transport of emissions emitted by industry takes place (Liu et al., 1984). The border to the free troposphere limits mixing of those emissions in height and reflects them back to the ground. Therefore, if the impact of industry on the air quality is analyzed, the ABL is the most

important part of the atmosphere for dispersion. A common use case is that a constant emitting source is modeled and the expansion of the plume is simulated. A part of the theories, on which the simulations are based, work on a small scale and assume a constant wind in space and time. Other theories simulate turbulent diffusion based on an analogy to the molecular diffusion which depends on the concentration gradient. In general the models are semiempirical, because they were developed by fitting tentative theories to observations (Venkatram and Du, 2003). However, these theories can only be applied to the ABL that has very special conditions in comparison to the upper layers. Furthermore these theories usually describe the plume by a Gaussian distribution or concentration plumes and can not directly be applied to a Lagrangian dispersion model which simulates individual air parcels. Anyhow, the ABL plays a minor role for MPTRAC, because the air parcels that are simulated are usually above the ABL, where diffusion has to be modeled differently. So this complex part of the atmosphere is not modeled as detailed as in more specialized models.

Turbulence in the free troposphere and stratosphere has not been analyzed as detailed as the diffusion in the ABL. It is assumed that there is no three-dimensional background diffusion that causes a general spread of the air parcels in the fashion of a molecular diffusion (Haynes, 2003). Turbulence above the ABL is caused by small localized events like convection, shear flow or breaking of gravity waves. They form sporadically and last for short times. The used meteorological data does not include information about such events, they can only be estimated based on the wind field.

A common procedure to model the turbulent diffusion is to use two basic components. The first one is chaotic or random flow. When referring to "turbulent diffusion" this kind of diffusion is meant in the following. The air parcels are perturbed by assuming that they follow a Gaussian distribution. The exact implementation depends on the atmospheric layer of parcel. Diffusivity coefficients that define the standard deviation of the Gaussian distribution in horizontal and vertical direction are obtained from experiments. This diffusion component is explained in detail in section 4.1.

The second component is a time correlated motion that depends on the diffusivity of the wind field. This time correlated diffusion is called "mesoscale diffusion" and considers the variation of the wind field around the parcels position. The variation of the previous time step is also taken into account so that a time correlation is given. This diffusion affects the parcels in all atmospheric layers and considers horizontal and vertical movement. A more detailed description on how this part of the diffusion is modeled and implemented is given in section 4.1.

The diffusion model implemented in MPTRAC is based on the scheme of the Lagrangian particle dispersion model FLEXPART (Stohl et al., 2005).

#### Decay and Sedimentation

Two more physics modules are implemented in MPTRAC. The decay model can be used to simulate the exponential decay of mass of air parcels by giving them a half-life period and a mass. When

non-radioactive parcels are simulated this model can also be used to approximately simulate the chemical decomposition of trace gases. The sedimentation module considers the fall out of aerosol particles, e.g. omitted by volcanic eruptions. Details of these modules are ooutside the scope of this thesis and are not further discussed here.

### 2.1.3 Simulation evaluation and model output

The evaluation of simulations with MPTRAC can either be performed afterwards, by analyzing the output files or automatically at run time. The output can be written in three different styles for different use cases. The first one is a simple list of all parcels which match the format of the input files and can be used to create restart points. It can be written as a human readable ASCII- or binary netCDF-file. This is the standard format which shows the behavior of every single parcel and can be used to calculate transport deviations which is a good method to compare simulations. They are described in detail in section 2.3.

The second output scheme divides the Earth's atmosphere into a regular grid and sums up all parcels in the boxes. With the given box area and height, the column density of parcels or simulated quantities can be calculated. This output is comparable to an Eulerian models output. Eulerian models compute transport based on the transfer of concentrations of properties between fixed grid points that represent the atmosphere instead of individual air parcels that travel themselves. When a great number of parcels is simulated this output format is usually used, because it is faster and the density is more interesting than individual parcels.

The last output type filters the parcel and only writes them out if they are near a specific location. This scheme is used in combination with backward simulations to find out if a parcel was emitted by a specific source and when this happened. Many plots scripts are available that allow to illustrate the results.

The run time evaluation (Rößler, 2014) compares satellite observations of a simulated quantity to the simulated parcels by using the Critical Success Index (Schaefer, 1990). It returns a figure of merit that indicates how well the observations and predictions fit at a given time. Since this evaluation is not needed in this thesis it will not be described in detail. This run time evaluation can also be written as a gridded output file that shows which regions match and which differ most.

### 2.1.4 Hybrid parallelization

The Lagrangian particle dispersion model is completely written in ANSI C and has about 5000 lines of code. It is parallelized with Open Multi-Processing (OpenMP) and the Message Passing Interface (MPI). So the program supports multi core processors and can run on computers with multiple independent processors such as supercomputers. The OpenMP parallelization allows users to run relatively large simulations on usual desktop computers. Since air parcels do not interact with each other, no communication between different tasks is needed. So all air parcels are distributed equally

on all cores. The MPI parallelization is determined to execute ensemble simulations and works very similarly. Instead of air parcels, simulations are distributed on the available working units of a supercomputer. Detailed information about the performance and the OpenMP-speedup are given in section 4.5.

## 2.2 ERA-Interim Reanalysis

As mentioned earlier, the MPTRAC simulations are driven by external meteorological data. In this thesis the ERA-Interim data product (Dee et al., 2011), provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) is used. A meteorological reanalysis assimilates historical data from long time period with a single consistent scheme. The ERA-Interim reanalysis is computed with the ECMWF's Integrated Forecast System (IFS) in version Cy31r2 from 2006. The time span from 1979 to the present is covered by this data product. The resolution of the data is $1^\circ \times 1^\circ$ in the horizontal and 60 pressure levels from surface up to 0.1hPa (around 65 km) in the vertical. The temporal resolution is 6 h. It is the data product of choice for MPTRAC as a comparison of three reanalysis data products and one operational analysis has shown (Hoffmann et al., 2015; Rößler, 2014). In three simulations of the spread of volcanic eruptions the ERA-Interim dataset always lead to the best results.

## 2.3 Absolute and relative transport deviations

A good way to compare different transport simulations with the same initial set of trajectory seeds is to calculate transport deviations (Kuo et al., 1985; Stohl et al., 1995). They show how the trajectories of air parcels pairs diverge over time. One simulation usually is assumed to be an exact solution, so the deviations indicate how good another simulation is. Many simulations with different configurations can be compared to the solution that is assumed to be exact to find the best one. In this thesis transport deviations are used to compare the different integration schemes, find a good time step and analyze the impact of the interpolation of the wind field.

Transport deviations are calculated by averaging the distance of all parcel pairs at a specific time. The horizontal and vertical deviation is calculated separately because horizontal errors are usually many times higher. In contrast to the original definition of Kuo et al. (1985) or Stohl et al. (1995) the vertical deviations are calculated here in geometric altitude and not in pressure. The equations for the absolute horizontal and vertical transport deviations are

$$AHTD(t) = \frac{1}{N} \sum_{n=1}^{N} \sqrt{(X_n(t) - x_n(t))^2 + (Y_n(t) - y_n(t))^2}, \tag{2.1}$$

$$AVTD(t) = \frac{1}{N} \sum_{n=1}^{N} |Z_n(t) - z_n(t)|. \tag{2.2}$$

Here a parcels position at time t in Cartesian coordinates is given by $X_n(t)$, $Y_n(t)$ and $Z_n(t)$. This equation does not consider the Earth's curvature, so the calculated distance between two points is smaller than the real distance (arc length). As only relatively small deviations are evaluated here, the error can be neglected.

Furthermore, relative deviations can be calculated (RHTD & RVTD). They divide the absolute deviations by the horizontal or vertical length of the trajectory, respectively,

$$RHTD(t) = \frac{1}{N}\sum_{n=1}^{N}\frac{\sqrt{(X_n(t)-x_n(t))^2+(Y_n(t)-y_n(t))^2}}{\frac{1}{2}(L_n^a(t)+L_n^b(t))}, \tag{2.3}$$

$$L_n^a(t) = \sum_{n=1}^{N}\sqrt{(X_n(t)-X_n(t-1))^2+(Y_n(t)-Y_n(t-1))^2},$$

$$L_n^b(t) = \sum_{n=1}^{N}\sqrt{(x_n(t)-x_n(t-1))^2+(y_n(t)-y_n(t-1))^2},$$

$$RVTD(t) = \frac{1}{N}\sum_{n=1}^{N}\frac{|Z_n(t)-z_n(t)|}{H_n(t)}, \tag{2.4}$$

$$H_n(t) = \frac{1}{2}\sum_{n=1}^{N}|Z_n(t)-Z_n(t-1)|+|z_n(t)-z_n(t-1)|.$$

In the following, the mean transport deviations are not analyzed alone, because they are sensitive to outliers and sometimes falsified. In addition, the statistical distribution of the transport is often interesting. It shows how stable the model and an integration scheme is. Therefore, the median and certain quantiles ($p_{0.9}$, $p_{0.75}$, $p_{0.5}$, $p_{0.25}$ and $p_{0.1}$) are calculated. The minimum and maximum deviation is also tracked so that the complete error interval is known. This helps by detecting outliers and possible implementation errors.

It is necessary for a reasonable evaluation of transport deviations that the reference is proven to be an exact or good solution for the transport. A usual procedure is to use a simulation with a small time step as reference. Since the main purpose of this thesis is to verify the model, this reference will also be checked for correctness.

# Chapter 3

# Advection

## 3.1 Integration schemes

Trajectory models calculate the pathway of air parcels with time. In the Lagrangian approach, the flow of individual air parcels is calculated. Assuming that air parcels are infinitely small and only affected by wind, their pathway is defined by the kinematic equation

$$\dot{x} = v(x(t), t). \tag{3.1}$$

With a parcel's position $x(t)$ and time $t$, the velocity vector $\dot{x}$ of an air parcel is given by function $v(x, t)$, that returns the wind speed at a position and time. The exact solution for a parcel's trajectory can be written as

$$x(t_e) = x_0 + \int_{t_0}^{t_e} v(x(t), t)dt. \tag{3.2}$$

With start position $x_0$, start time $t_0$ and end time $t_e$. The wind field, returned by function $v$, is very complex and can not be integrated analytically. Based on the ordinary differential equation given by the kinematic equation of motion (3.1), the trajectory calculation can be treated as an Initial Value Problem (IVP). The initial condition is given by start times and locations of a set of air parcels and a wind field that changes with time. Many methods for numerical integration are known and can be used to solve this IVP. They basically solve Equation (3.2) by repeatedly integrating over short time steps. This is exemplary shown by the equation

$$\tilde{x}(t + \Delta t) = \tilde{x}(t) + \Delta t \cdot v(\tilde{x}(t), t). \tag{3.3}$$

Where $\tilde{x}$ is a approximation of $x$ and $\Delta t$ the model's time step over that is integrated. The derivative of $x$ is $v$ (and $\dot{x}$), which returns a vector that contains the wind speeds in all three directions at the parcels position at a specific time. These information are obtained from the wind field, which is a part of the initial condition. Since the wind vector is given by meteorological input files, the

approximation can easily be calculated. Better approximations can be achieved when higher order derivatives are used.

In this thesis the best method of five different integration schemes of different orders is established. The comparison will consider accuracy and performance. Calls to function $v$ are the most expensive part of the calculation, because a linear 4-D interpolation is done for every parcel. The number of calls to this function depends on the model time step and the number of function calls needed by the integration scheme. As a consequence, the choice of $\Delta t$ basically controls the performance. The smaller it is, the more time is needed for the simulation. A too large time step will lead to a bad approximation of the integration scheme and consequently the results are wrong. So a good tradeoff between performance and accuracy has to be found by choosing the right time step.

The error of an integration scheme can be estimated by the equation

$$|x(t_n) - x_n| \leq C \cdot h^p, \qquad h \to 0. \tag{3.4}$$

Where $h$ is the time step and $p$ the convergence order of the integration scheme. The function $x(t_n)$ is the exact solution and $x_n$ is the calculated approximation. The left side of Equation (3.4) is the global truncation error, that is the difference of the exact solution and the approximation at a time. The global truncation error is limited by a constant $C$ multiplied with the time step to the power of the convergence order. Global truncation error is caused by successive local truncation errors. A local truncation error is the difference of the exact solution and the approximation for a single iteration. It shows how strong the predicted locations of the current step differ, assuming that the initial condition was correct which usually is not the case. The global truncation error is greater than accumulated local truncation errors, because after the first time step the positions for all following calculations are already affected by errors. These errors in the position were caused by the local truncation error of the previous steps. Every step will have a more falsified initial condition, which will lead to an amplification of the error. The only chance to keep this error small is to minimize the local truncation error. Usually the global error is ignored and assumed to be unproblematic, if the local truncation error is small. Nevertheless the assumption can cause great problems in long simulations. Since both kinds of truncation errors can not be calculated, because they require the exact solution, another proceeding is not possible. The local truncation error, however, can easily be estimated by solving every time step with two different integration schemes. the error estimation is used by adaptive procedures to choose a good value for h in every time step. It does not take as much computation time as it seems on the first view, if some tricks are applied. Since MPTRAC always uses a constant time step, this kind of error estimation is not available and will not be discussed further. A third and last source of errors is the numerical error of the inaccurate floating point operations. This error is usually small enough to be ignored completely.

Thus, with relation to the most important local truncation error, a convergence order of 1 means

that the error is proportional to the time step size, while a second order integration scheme reduces the error by a factor of 4 when the time step is halved. From this it follows that a higher order scheme with a larger time step can be faster and more exact than a lower order scheme with a smaller time step. This also shows that, with enough computational effort, every desired accuracy can be reached, when the time step is reduced. In this thesis the explicit Euler method, the midpoint method, the iterative Petterssen scheme and Runge-Kutta methods of the order 3 and 4 are analyzed and compared. The best method shall be used in MPTRAC in the future.

### 3.1.1 Euler method

A first possibility to solve the given IVP is the explicit Euler method. It is the simplest method for numerical integration of ordinary differential equations. It is obtained from Equation 3.2 by a Taylor approximation with one element, which is equal to Equation 3.3. It is called a 'zero acceleration' solution. Mathematically the explicit Euler is written as

$$x_{n+1} = x_n + hv(t_n, x_n). \tag{3.5}$$

In the trajectory model the integration is done by adding the wind vector at the current position multiplied with the time step to the position. This procedure is easy to implement and leads to a fast calculation, but the accuracy is relatively low, because the method is of first order. A weak point of this method are functions with strong variations in space and time, like a wind field with a circular movement, because this method uses the conditions at the start point for the complete time step. Since the wind changes in time and space on the traveled path of a single time step, a very short time step has to be used to get reasonable results. Figure 3.1 shows an extreme example with an unrealistic long time step of 12 h.

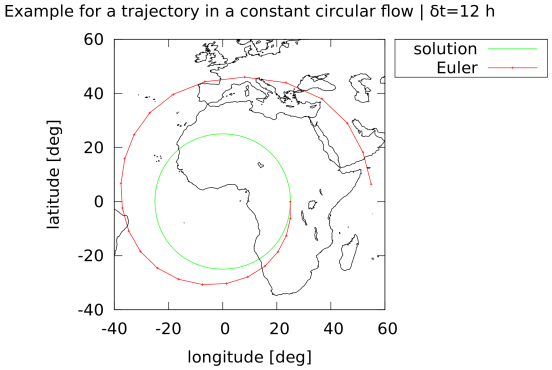


Figure 3.1: Example for a trajectory in a circular wind field calculated with the Euler method.

### 3.1.2 Midpoint method

MPTRAC currently uses the midpoint method as integration scheme. It is a second order method that uses a help point or 'mid-point' to obtain a better approximation of the wind vector. The recursive definition of the midpoint method is

$$x_{n+1} = x_n + hv(t_n + \frac{h}{2}, x_n + \frac{h}{2}v(t_n, x_n)). \tag{3.6}$$

The mid point is calculated with a Euler step using a half time step. In the trajectory model the wind vector of the mid point is then added to the initial location multiplied with the full time step. This method is still easy to implement and has good performance.

### 3.1.3 Petterssen scheme

The Petterssen scheme (Petterssen 1940) is a frequently used integration scheme in trajectory models (Bowman et al., 2013). In contrast to the other methods the Petterssen scheme works iteratively. The first iteration is similar to the Euler method, therefore the scheme is also called iterative Euler. The basic idea behind this iterative approach is that a wind field can vary strongly from place to place. While some air parcels are in a region with slow or constant winds, other parcels can be in complex wind fields. With this iterative scheme more iterations can be done, if necessary, for a good approximation of the wind. This increases the accuracy, but saves computing time. The recursive definition is

$$\begin{aligned} y_{n+1,0} &= y_n + dt \cdot v(t_n, y_n) \\ y_{n+1,i} &= y_n + \frac{dt}{2} \cdot (v(t_n, y_n) + v(t_{n+1}, y_{n,i-1})). \end{aligned} \tag{3.7}$$

At first a normal Euler step is done and a first prediction is calculated. The following $i$ iterations use the end point of the last prediction to calculate a better approximation. In detail, the mean wind of the wind vector at the start point $y_n$ at start time $t_n$ and the wind at the last predicted point $y_{n+1,i-1}$ at the time after the time step $t_{n+1}$, is used as advection vector. This is repeated until the wind vector converges. In MPTRAC a convergence is defined by a change in the wind vector of less than $10^{-5}$ m/s or after a maximum of 6 iterations. The order of this iterative method can not be specified as clear as for the other methods, but it is at least of the second order after two iteration.

### 3.1.4 Runge-Kutta methods

Runge-Kutta (RK) methods can theoretically be created for every desired order. In this thesis methods up to the 4th order are examined. Higher order methods are uncommon for trajectory models like MPTRAC, because at some point the input data's quality and other factors become more important. Two methods of the Runge-Kutta family were already introduced. The explicit

Euler presented in section 3.1.1 is a first order RK method. One possible second order method is the midpoint method shown in section 3.1.2. There are infinite possibilities to define RK methods. The methods used in this thesis are well known and used.

RK methods of higher orders will be defined in general form with the so called Butcher tableau, instead of an explicit definition. A Butcher tableau consists of a quadratic coefficient matrix $A$, called Runge-Kutta matrix, and two vectors $b$ and $c$. The size of the matrix $A$ defines how many stages the method has. The number of stages is equal to the number of function-calls, the higher it is, the higher is the possible order of the method. Its coefficients $a_{i,j}$ are used to calculate bases $k_i$ for the integration. They are weights that control the change in place for the function calls by combining other bases. In this thesis only explicit methods are used, therefore every element on the main diagonal and above is zero. This allows us to easily calculate every base, because they can only depend on previously calculated ones. Vector $b$ contains weights for calculated bases, its sum has to be one. The elements of vector $c$ are called nodes and are used to vary the time step for different function calls. So a Runge-Kutta method with s stages is defined by

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ a_{2,1} & 0 & 0 & 0 \\ \vdots & \ddots & 0 & 0 \\ a_{s,1} & \dots & a_{s,s-1} & 0 \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_s \end{pmatrix}, \quad c = \begin{pmatrix} c_1 \\ \vdots \\ c_s \end{pmatrix}$$

A Butcher tableau contains matrix $A$ and both vectors $b$ and $c$ and looks like

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array}. \tag{3.8}$$

The points are then calculated with

$$k_i = f(t + c_i \cdot h, y + h \sum_{j=1}^{i-1} a_{i,j} k_j), \qquad i = 1, ..., s \tag{3.9}$$

$$x_{n+1} = y + h \cdot \sum_{i=1}^{s} b_i \cdot k_i. \tag{3.10}$$

The values $k_i$ are results of function calls, the wind vector $v(x,t)$ in this case, $h$ is the model time step and $x_{n+1}$ is the new air parcel position that was predicted. All other coefficients are obtained from the Butcher tableau. How the matrix A and vectors b and c are determined in detail is not further discussed for the general case, but the Butcher tableau for the Euler and midpoint method are described as example.

The Butcher tableau for the first order RK, corresponding to the Euler method, is very simple. Since it has only one stage, matrix $A$ is zero. From this it follows that $b$ has only one element that has to be 1, since the sum of all elements has to be 1. The value in $c$ is 0 because the current time is used for the function call.

$$\begin{array}{c|c} 0 & \\ \hline & 1 \end{array} \tag{3.11}$$

For the second order RK or midpoint method the only non-zero element of $A$, $a_{1,1}$, is 0.5 because the midpoint is a half time step away from the origin. Vector $b$=(0, 1)$^T$ shows, that only the midpoint's wind vector is used to predict the new point. The value $c_0$ is zero, that means that the first function call for the value $k_0$ uses the start time. The value $c_1$ is 0.5, that increases the time by a half time step. This means that the time right between the start and end time is used at the midpoint for the function call. The Butcher tableau of the midpoint method is:

$$\begin{array}{c|cc} 0 & & \\ \frac{1}{2} & \frac{1}{2} & \\ \hline & 0 & 1 \end{array} \tag{3.12}$$

The definition of the used third order Runge-Kutta method is:

$$\begin{array}{c|ccc} 0 & & & \\ \frac{1}{2} & \frac{1}{2} & & \\ 1 & -1 & 2 & \\ \hline & \frac{1}{6} & \frac{4}{6} & \frac{1}{6} \end{array} . \tag{3.13}$$

The usual RK4 method, which is widely used, is defined by:

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array} . \tag{3.14}$$

This method is so popular because it is the highest order Runge-Kutta method for which the number of stages matches its order. A fifth order RK method requires at least six function calls. The accuracy and computation time ratio for the RK4 method is very good for that reason. The implementation of higher order RK methods in MPTRAC is also feasible, if the Butcher tableau is known.

## 3.2 Analytical test for the integration schemes

### 3.2.1 Definition of the analytical test case

The analytical tests for the model discussed here are most of all meant to check, if the implementation of the integration schemes is correct. There are different possibilities to define tests that can be used to validate the model. The tests introduced in this chapter where defined by Williamson and Rasch (1989). In general, they simulate a circular flow around the Earth oriented by different angles, without any vertical wind. The wind field of those test cases is defined by the following equation, which describes the zonal ($u$) and meridional ($v$) wind at a location:

$$u = u_0 \cdot (cos(\theta) \cdot cos(\alpha) + sin(\theta) \cdot cos(\lambda) \cdot sin(\alpha)) \tag{3.15}$$

$$v = u_0 \cdot sin(\lambda) \cdot sin(\alpha) \tag{3.16}$$

The location is given by a longitude and latitude, $\lambda$ and $\theta$, respectively. The maximum wind speed can be adjusted with the constant $u_0$. The parameter $\alpha$ defines the angle between the Earths rotation axis and the rotation axis of the wind field. An angle of $\alpha = 0°$ means that an air parcel's latitude coordinate will stay unchanged and all parcels travel parallel to the equator. This test case can also be called a rotation around the north and south pole. An angle of $\alpha = 90°$ leads to a rotation axis through the points (0° N, 0° W) and the matching point at the Earths other side (0° N, 180° W). This test case is more challenging because air parcels travel across the poles, where the singularity of the longitude-latitude grid at the poles can cause problems. Additional values for $\alpha$ are required to test the model sufficiently. The values 0, 0.05, $\frac{\pi}{2} - 0.05$ and $\frac{\pi}{2}$ were suggested by Williamson et al. (1992).

Since all wind fields are circular, every parcel should at some point in time return to its start location. The wind speed and simulation time were chosen so that the parcels arrive at their start locations at the end of the simulation. The simulation time is set to 12 days, that leads to a maximum wind speed $u_0$ of about 38.58 m/s by assuming that the Earth is a perfect sphere with a radius of 6367.421 km. The 12 days time period were suggested by Williamson et al. (1992) to create a wind field with realistic wind speeds.

The wind field defined by Equations 3.15 and 3.16 is constant in time and all advantages or disadvantages of the interpolation and integration methods, which depend on the temporal variation, are not taken into account. Since the temporal variation is a main reason why higher order methods are used, another wind field with a variation in time is proposed here. The wind speed is multiplied with a sinus-function depending on time. Due to the fact that the sinus function is non-linear, greater divergences of the integration schemes can be expected. After a period the parcel should also end up at its start position, without completing a complete circle. The time period for the sinus-function

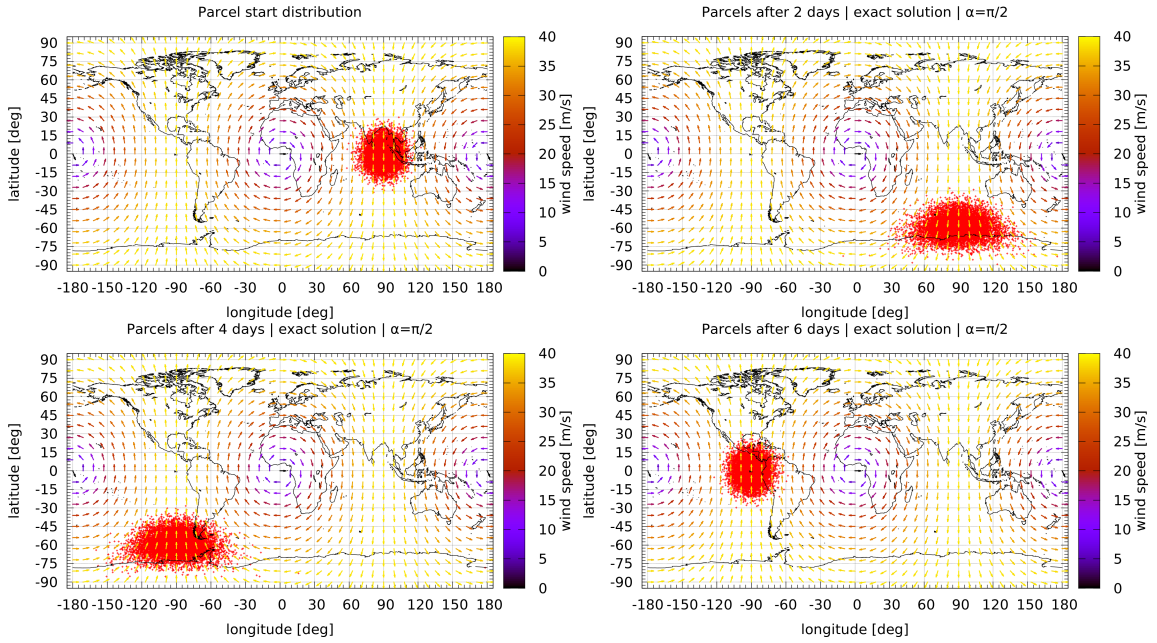


Figure 3.2: The wind field for $\alpha = \pi/2$ and the locations of all air parcels at the start and after 2, 4 and 6 days.

is set to 12 days. So the wind speed starts at zero and increases for the first 3 days, then decreases for the next 3 days until it is zero again. After that it is repeated with the opposite direction. The equation for the dynamic wind field is

$$u(t) = sin(\pi \cdot t/\delta t) \cdot u_0 \cdot (cos(\theta) \cdot cos(\alpha) + sin(\theta) \cdot cos(\lambda) \cdot sin(\alpha)) \tag{3.17}$$

$$v(t) = sin(\pi \cdot t/\delta t) \cdot u_0 \cdot sin(\lambda) \cdot sin(\alpha). \tag{3.18}$$

Where t is the simulation time and $\delta t$ the time period for a sine wave. The other variables have the same meaning as before.

As mentioned before, the initial value problem is defined by the wind field, a start time and start locations. All parcels start at the same time $t_0$, which can be chosen freely. The start locations are defined by a Gaussian distribution around the point (0° N, 90° E). The standard deviation for the radius is set to a third of the Earths radius, about 2122.47 km. These parameters are similar to those proposed by Williamson et al. (1992). A total number of 10000 parcels are distributed. Figure 3.2 shows the wind field and air parcel positions at time steps after 2, 4 and 6 days for $\alpha = \pi/2$.

### 3.2.2 Cosine singularity

At every time step the change in meridional and zonal direction is calculated with the integration schemes presented in section 3.1. Since the wind field is given in the unit m/s, the calculated change for the time step is given in meters. However the parcels positions are given by a longitude ($\lambda$), latitude ($\theta$) and an altitude coordinate, so a conversion from the Cartesian distance to an angular distance and vice versa is necessary. For this conversion the lengths of a circle of latitude and longitude are required. The length of a circle of longitude is always $l_\theta = 2\pi \cdot R_{Earth}$ and only depends on the Earth's radius $R_{Earth}$, which is set to a fixed value of 6367.421 km in MPTRAC. In contrast, the length of a circle of latitude also depends on its latitude: $l_\lambda(\theta) = 2\pi \cdot R_{Earth}/cos(\theta)$.

The following equation converts the distance to travel meridional $\Delta y$ given in meters into a change of degrees for the latitude:

$$\Delta\theta = \frac{360 \cdot \Delta y}{2\pi \cdot R_{Earth}} \tag{3.19}$$

The conversion for the zonal movement into a change of the longitude is done with:

$$\Delta\lambda = \frac{360 \cdot \Delta x}{2\pi \cdot R_{Earth} \cdot cos(\theta)} \tag{3.20}$$

A problem that comes with the second conversion is that the cosine function is part of the denominator and becomes zero at the poles, causing a singularity. To ensure that no division by zero occurs, a 4th order Taylor series was used in the original version of MPTRAC instead of the cosine function. The Taylor series used as approximation is

$$cos_{Taylor4}(x) = 1 - \frac{x^2}{2} + \frac{x^4}{24}. \tag{3.21}$$

The approximation's roots are at $\pm 1.5925$, which would be angles of about $\pm 91.2°$. Since the latitude is always kept between -90° and 90°, a division by zero can not occur. Nevertheless the approximation was assumed to be sufficiently exact on average. Figure 3.3 shows the inverse of the cosine function and the approximation. The absolute error of the approximation is around 0.02 for angles of $\pm 90°$. However the relative error increases rapidly for latitudes above 80°. This large error does not count too much, though, because an error of one degree on a point with a high latitude causes a smaller error in distance than near the equator, because the circles of latitude are shorter. Especially at the pole itself, the longitude does not matter at all, because it refers to the same point.

To check if the error caused by this Taylor series is negligible, an additional test case was defined. Air parcels start at every latitude with longitude 0 and on every longitude with latitude 0 in distances of 1 degree. The green cross shown by Figure 3.4 at day 12 is equal to the start distribution. This

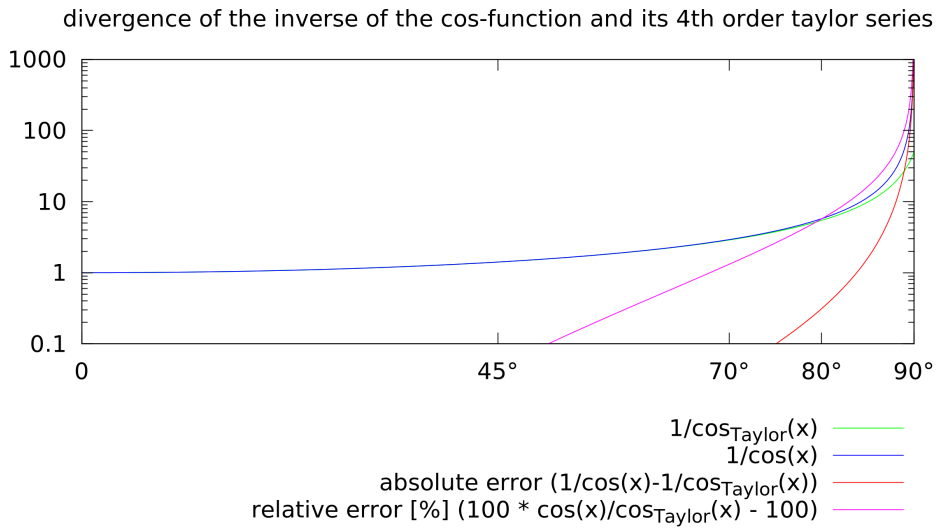


Figure 3.3: The inverse of the cosine function and its 4th order Taylor series approximation of the cosine by its 4th order Taylor series.

initial condition was used instead of the previous introduced ones, because deviations are easier to notice with a small number and uniformly distributed parcels. In a great point-cloud, the behavior of a single parcel can not be seen and the problem with the cosine singularity could not be illustrated as well as with this test, since this problem only affects air parcels which pass by close to the pole. In the simulation the air parcels rotate around the axis between the points (0° N, 0° W) and (0° N, 180° E) which leads to a movement across the poles. So the wind field is equal to the case with $\alpha = \pi/2$ mentioned in section 3.2.1. After 12 days all parcels should have fulfilled the rotation and return to their start locations. Two simulations with the RK4 integration scheme and a time step of 30 s were done. Both simulations use the direct equations to calculate the wind instead of the interpolation with a gridded input file. This reduces integration and interpolation errors to a minimum. One simulation uses the Taylor series and the other the cosine function with a latitude check to increase the accuracy without causing a devision by zero. This improved version will use the cosine function if the latitude of the point is in the interval $[-89.999 : 89.999]$, otherwise 0 is returned for the change of the longitude. Figure 3.4 shows the positions of all air parcels after 7 and 12 days. The red points are parcels of the simulation with the Taylor approximation and the green dots belong to an improved version.

After six days all parcels have fulfilled over a half of the rotation, so all parcels have been near a pole once. Most parcels positions of both simulations match, but four regions with strong deviation are noticeable. The ends of the line with the start longitude of zero diverge strongly. Those parcels that started with a longitude of around ±90° also show great deviations.

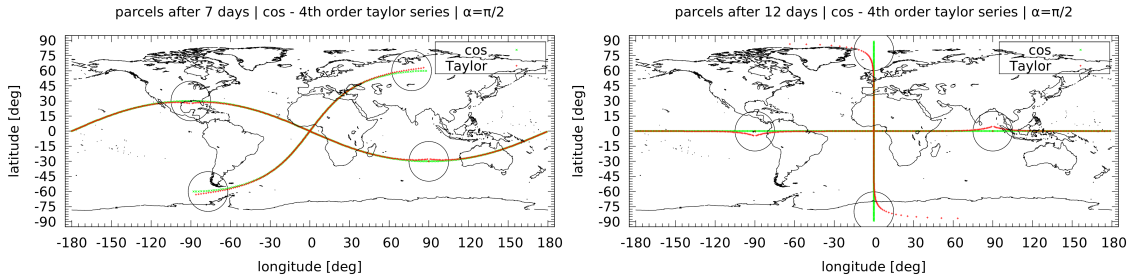


Figure 3.4: Comparison of a simulation with the cosine function and its Taylor approximation. The circles mark regions with great differences.

The other plot shows the end points of both simulations, which should match the start points. This criteria is only satisfied by the simulation with the cosine function. The mean distance of the endpoints of both simulations is 52.87 km. This mean value is strongly influenced by the outliers that are marked in the figure. The median distance is just 2.93 km and indicate that most parcels fit together. The time series has shown that all parcels which diverge strongly start with a longitude around $\pm 90°$ or with latitude near the north or south pole. The highest diversion is 488.07 km. Those parcels' trajectories pass by very close to the poles, where the cosine and the Taylor series diverge most.

This test demonstrated that the Taylor series is not accurate enough and has to be replaced. A possible solution could be to use a higher order approximation. Since every second term is 0, a 6th or 8th order approximation is conceivable. The 6th order Taylor series can not be used because its roots are in the interval $[-\frac{\pi}{2} : \frac{\pi}{2}]$ and would not solve the division by zero problem. The computation of the 8th order approximation already takes longer than the cosine function itself, so in the end the solution with the real cosine function and the latitude check was used. The mean distance between the start and end locations for this improved version is only 0.44 km, the median is 0.22 km and the maximal deviation is 1.76 km. So this implementation will be used for MPTRAC hereafter.

### 3.2.3 Exact solution

An exact analytical solution for the trajectories of air parcels is possible but relatively complicated to calculate for every $\alpha$. Therefore a simulation with a 4th order Runge-Kutta integration scheme and a short time step of 30 seconds is used as reference for the transport deviations in the following. The wind data for the reference simulation does not come from an input file, as in usual simulations with meteorological data, but is computed exactly by solving the introduced equations for a parcels location like in the previous test.

Since the reference has to be a correct solution, the reference is validated by two ways. The first validation was done with an analytical solution. The test cases $\alpha = 0$ and $\alpha = \frac{\pi}{2}$ are special

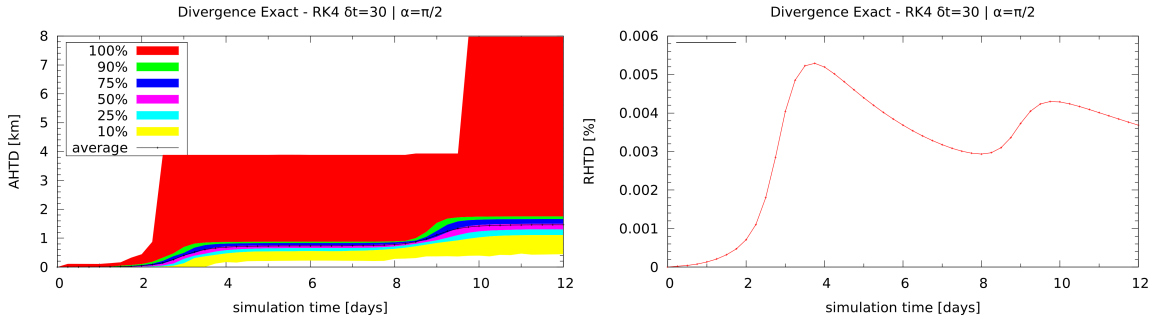


Figure 3.5: Absolute and relative transport deviations of the exact solution and a RK 4 simulation for the test case $\alpha = \pi/2$ .

cases that can relatively easy be solved analytically. For the test case with $\alpha = 0$ all parcels travel along a circle of latitude, which means that the latitude coordinate stays constant. As the time for a complete circle and its length is known, the position can be calculated for every time. Since this test case is not really challenging for the model, no comparison between the exact solution and the reference calculated with the model is done. The test case with $\alpha = \pi/2$ is more pretentious but can also be solved. In this test case all parcel rotate around the axis between the points (0° N, 0° W) and (0° N, 180° W). A rotation matrix can be used to rotate the parcels around it. To do so the longitude and latitude are converted into a Cartesian coordinates. After that the location vector is multiplied with the rotation matrix and converted back. The exact solutions for all parcels in time steps of 6 hours were calculated and compared to the reference simulation by calculation transport deviations. The deviation is so small that the simulation can be used as reference without doubts. The maximum absolute error is around 8 km at the end and the relative error stays below 0.006 %. The temporal evolution of the error is shown in Figure 3.5. The other test cases are not analyzed as detailed as this most challenging test, the knowledge that the most complicated test works correctly and the following second validation are sufficient to validate all references.

The second validation is done for every test case. As all tests simulate a circular movement, the start and end locations should match. Tables 3.1 and 3.2 show the calculated distances of start and end locations of the reference simulation for all four test cases with the constant and dynamic wind field. The mean distance, the quartiles and the minimum and maximum error are listed. The largest distance is 9.2 km, which leads to a relative error of 0.025 % with the average travel distance of about 40,000 km. Together with the first validation the correctness of the reference is proven.

| $\alpha$ | **mean** | max | $\mathbf{p}_{0.75}$ | $\mathbf{p}_{0.5}$ | $\mathbf{p}_{0.25}$ | min |
|---|---|---|---|---|---|---|
| 0.0 | 5.40341e-09 | 5.46033e-09 | 5.45173e-09 | 5.43305e-09 | 5.38642e-09 | 4.64343e-09 |
| 0.05 | 0.000516505 | 0.0027768 | 0.000741253 | 0.000429134 | 0.000204522 | 1.5021e-08 |
| $\frac{\pi}{2} - 0.05$ | 1.00575 | 9.19443 | 1.47876 | 1.27437 | 0.15577 | 0.00135969 |
| $\frac{\pi}{2}$ | 1.46186 | 7.9789 | 1.66144 | 1.49747 | 1.30009 | 0.466808 |

Table 3.1: Distances of start and end locations for the constant wind field.

| $\alpha$ | **mean** | max | $\mathbf{p}_{0.75}$ | $\mathbf{p}_{0.5}$ | $\mathbf{p}_{0.25}$ | min |
|---|---|---|---|---|---|---|
| 0.0 | 3.22893e-12 | 1.42472e-11 | 4.30361e-12 | 2.87607e-12 | 1.36424e-12 | 0 |
| 0.05 | 0.0048611 | 0.0284295 | 0.00694483 | 0.00394026 | 0.00182477 | 1.76539e-06 |
| $\frac{\pi}{2} - 0.05$ | 1.05973 | 4.54776 | 1.27968 | 1.05566 | 0.836196 | 0.017448 |
| $\frac{\pi}{2}$ | 1.06902 | 3.49856 | 1.28574 | 1.06747 | 0.852579 | 0.166446 |

Table 3.2: Distances of start and end locations for the dynamic wind field.

### 3.2.4 Integration scheme comparison

#### Time step choice

As mentioned before, the time step is a main factor that controls the trajectory model's performance. Therefore the greatest time step is searched that still allows a sufficient accuracy. A reasonable limit for the time step can be obtained from the Courant-Friedrichs-Lewy (CFL) condition. This condition ensures that no grid box of the meteorological data is skipped by an air parcel within one time step. So all available wind information that affect an air parcel are really taken into account. The mathematical definition of the CFL condition is: $\Delta t \leq \Delta x / v_{max}$, with a time step $\Delta t$, an input data grid box size $\Delta x$ and a maximal wind speed of $v_{max}$. The maximum wind speed for this test case is about 38 m/s, but for real wind fields greater speeds are possible, therefore the maximum speed is assumed to be 100 m/s. The meteorological input data box sizes at the equator are around 277, 111, and 55 km for the typical resolutions of 2.5°, 1°, and 0.5°. This leads to maximum time steps of around 9, 18, and 46 minutes. These limits actually only apply to the explicit Euler, because the other methods use help points that probably are in a box that would be skipped by the Euler method. So the tests were run with time steps of 2, 4, 8, 15, 30, and 60 minutes.

An additional set of simulations was run with time steps that are intended to lead to a balanced computation time of all integration schemes. This was done to compare the possible accuracy with a fixed computation time to find the most efficient integration method. The Euler time step was set to 120 seconds. The RK 2 method requires two calls to the interpolation function, but needs less than twice as much time, a time step of 216 seconds was used. For the RK 3 method a 300 second time step and for the RK 4 method a time step of 400 seconds were chosen. The Petterssen scheme

on average needs 2.3 to 3.7 calls to the interpolation function for the different time steps. A time step of 360 seconds costs around as much computation time as the other integration schemes.

**Comparison of all schemes**

In this section simulations of all integration schemes with an identical time step are compared. This shows basic differences in the accuracy of the different schemes. Time steps of 2 min to 1 hour were used. A time step of 30 minutes is used for the comparison in the following. Figure 3.6 shows the absolute horizontal transport deviations between the deciles $p_{0.1}$ and $p_{0.9}$. The test cases $\alpha = 0.05$ and $\alpha = \pi/2$ are shown exemplarily, the test cases with $\alpha = 0$ and $\alpha = \pi/2 - 0.05$ are very similar and are left out. Results using other time steps show the same curve shape and only differ in the magnitude. The Runge-Kutta methods show similar results, so that only the RK4 method is shown in the figure.

The test cases that do not pass the poles ($\alpha = 0$ and $\alpha = 0.05$) show much smaller deviations. For the constant wind field there is practically no error, with relative errors less than 0.01 %. The dynamical wind leads to greater deviations of up to 45 km for the Euler or 20 km for the other methods. For the higher order methods this error grows faster when the wind speeds are higher. The Euler method shows no clear correlation to the sine-shaped wind progression, it is more likely that it depends on the derivation of the wind progression. The maximum slope is at the beginning, where the cosine function also has its maximum. This is actually the weak spot of the Euler, changes in the wind field over a time step are not considered. In both cases the errors that were made in the first 6 days are balanced out over the following 6 days with the opposite wind direction.

The test cases that go across the poles show greater deviations. This time the constant and dynamic wind fields do not lead to such different curve shapes, because another error source has a greater impact this time. There are two regions with a strong increase of the deviations after 3 and 9 days or 4 and 8 days respectively. All integration schemes show this feature distinctly. A smaller time step weakens the effect, but even a time step of 30 seconds can not solve the problem fully (cf. Figure 3.12). In consequence, the integration schemes seem not to be the source of this error. The problem will be discussed later in section 3.2.4. The spread of the air parcels is higher with the dynamical wind field than with the constant one. Nevertheless the maximal absolute deviation is about 50 km (15 %) smaller. The smaller absolute deviations can be explained by the lower wind speeds. On average the wind speeds are about 1.6 times slower for the dynamic sine-shaped wind. The greater spread is caused by the higher variation of the wind field. The average and mean deviations of the individual simulations do not differ much, showing that large outliers are unlikely. Both wind fields lead to similar relative deviations. All RK methods and the Petterssen scheme have relative deviations of less than 0.4 %. The Petterssen scheme shows minimal greater deviations than the RK methods of order 2 to 4, but that's probably due to the fact that the reference was calculated with a RK method. The Euler method shows greater relative deviations for the dynamic wind field at

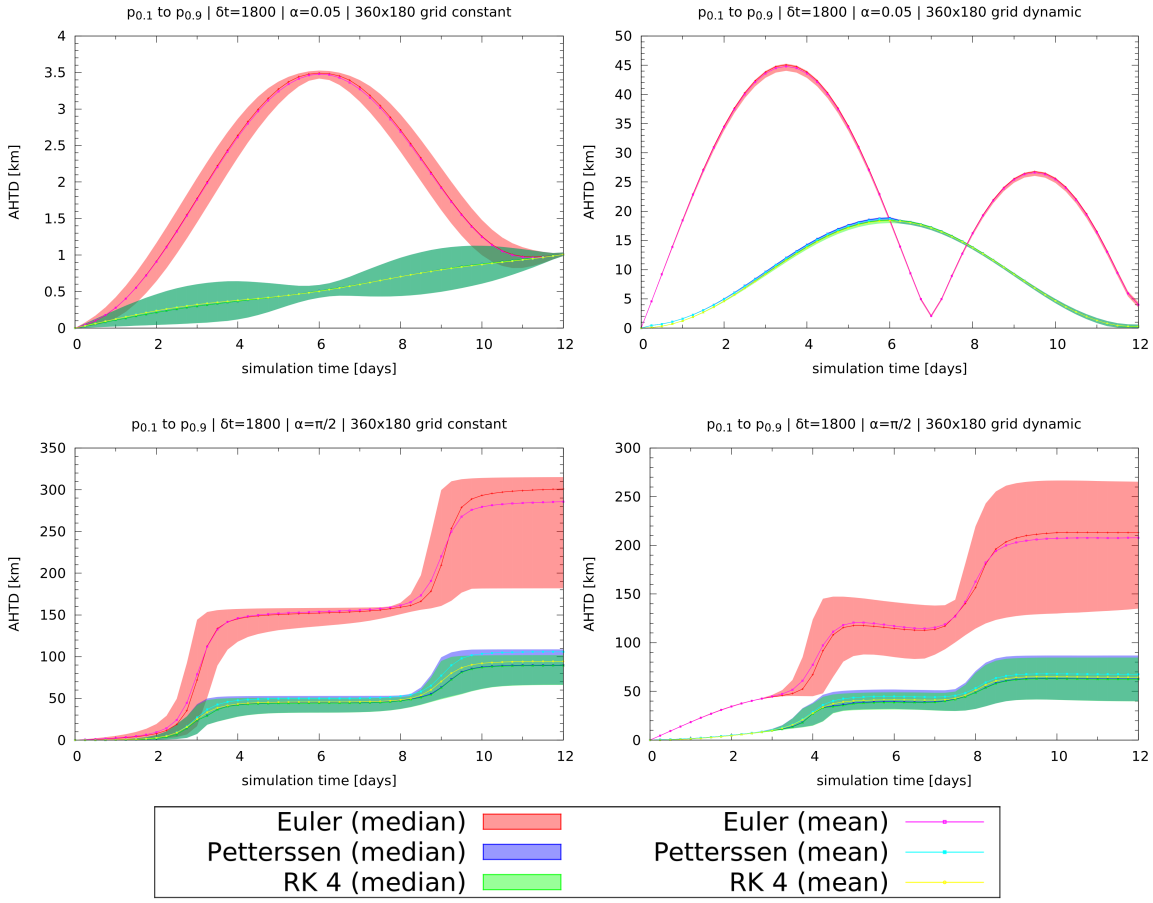


Figure 3.6: Transport deviations for $\alpha = 0.05$ and $\alpha = \pi/2$ with a time step of 30 min. The lines display the average and mean deviation. Runge-Kutta methods of order two and three show no visible difference to the forth order and are left out. The test case with $\alpha = 0$ or $\alpha = \pi/2 - 0.05$ are very similar.

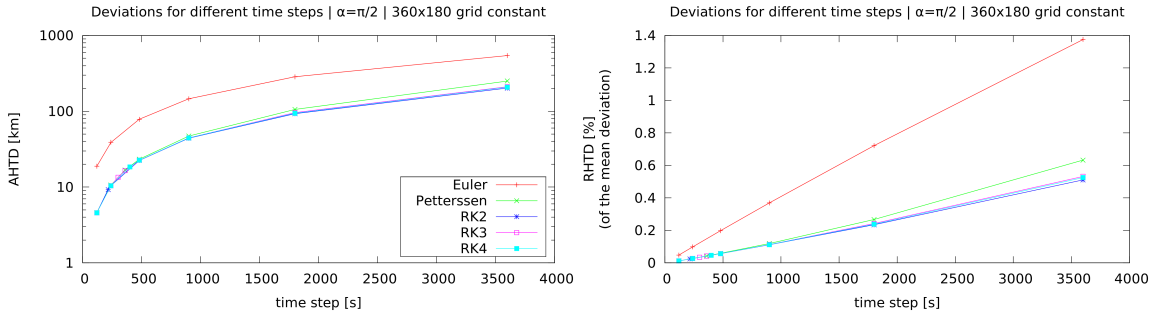


Figure 3.7: Deviations of simulations with different time steps for the most challenging test case with $\alpha = \pi/2$.

the beginning and at the already mentioned peaks, but the mean relative deviation is still around 1 %.

All in all, this comparison has not shown large differences between the integration schemes. Only the Euler method differs noticeably in terms of the average and median horizontal transport deviations. This could be caused by the relatively simple wind field that does not challenge the integration schemes. Figure 3.7 shows the maximal transport deviations and the end of the simulation with the constant wind field and different time steps. The deviations show that even a large time step of 60 minutes gives results with an acceptable relative deviation. The time step can not be enlarged much further without affecting the CFL condition. Moreover, wind fields with higher variations and wind speeds will require a shorter time step as the test with a real wind field will show. The wind data was obtained from data files with the usual grid resolutions of 2.5°, 1° and 0.5°. The three resolutions do not show relevant differences, so the error caused by the resolution and the linear interpolation is negligibly small compared to other error sources. The comparison has also shown that all Runge-Kutta methods of order 2 and above are very similar, no real deviations between the curves are visible. An order of above two does not result into further improvements. Considering the increasing computational effort, a RK method of a low order seems to be the method of choice.

### Maximal errors

Previously, only mean errors and the behavior of the bulk of parcels were analyzed. For a validation of the correctness maximal deviations are also interesting, because they can reveal outliers or weak spots of the model. In simulations with real wind fields, maximum errors usually belong to parcels that run away so far from the correct trajectory that completely different winds affect them. Small errors in the trajectories add up fast and soon a pair of trajectories has nothing in common. So a comparison of trajectories only makes sense if their distance is not too big.

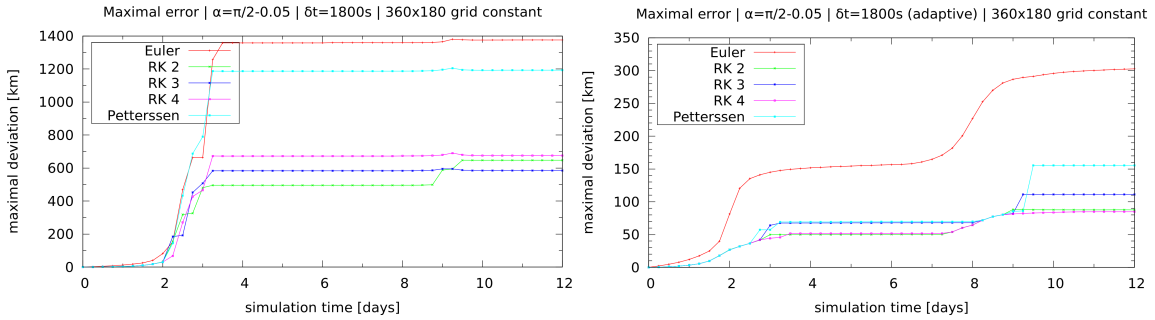


Figure 3.8: Maximum transport deviations of a simulation with a constant time step of 1800 s in comparison to a simulation with an adaptive time step near the poles in a constant wind field.

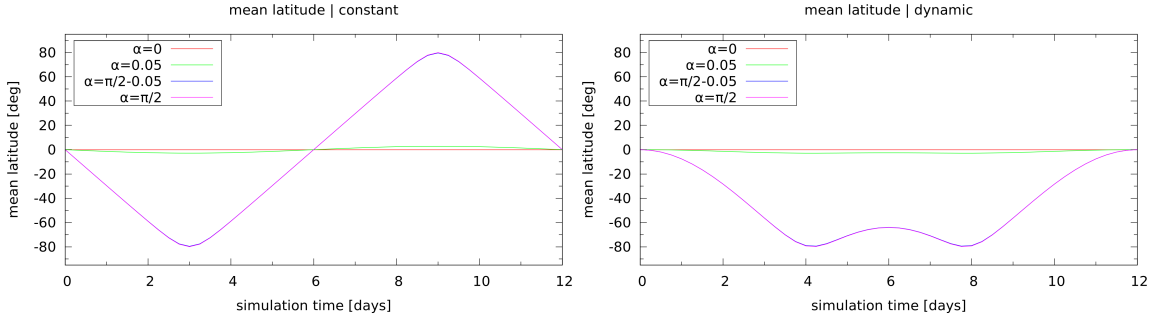


Figure 3.9: Mean latitude of the air parcels over time for a constant and dynamic wind field.

In this case, the wind field stabilizes the trajectories, because it has a similar behavior at every point. All trajectories are circles around a center, so a small error does not lead to an amplification of the error, but can only change the radius of the trajectory. From this it follows that maximal errors are relevant for the test cases and should be acceptably small.

For the test cases with $\alpha = 0$ and $\alpha = 0.05$ the maximal deviations are only slightly greater than the mean deviation and no outliers occur at all. The maximal deviations from the exact solution for the test case with $\alpha = \pi/2$ with a time step of 1800 s and a constant wind field is shown in Figure 3.8. The right plot belongs to a simulation of an improved MPTRAC version, that will be described later. The relative transport deviations of the maximum error is still small with 2 to 5%, but the steep ascent of the curves at some points suggest that there is a problem in the model. The maximal errors and the previously analyzed transport deviations show a strong increase for day 3 and 9 for the constant wind field and day 4 and 8 for the dynamic wind field.

Figure 3.9 shows the mean latitude of all air parcels over time with a constant or dynamic wind field. The time when most parcels are near a pole matches the peak of the gradient of the deviation. So it is very likely that the reason for this strong deviation at the poles has to do with the coordinate

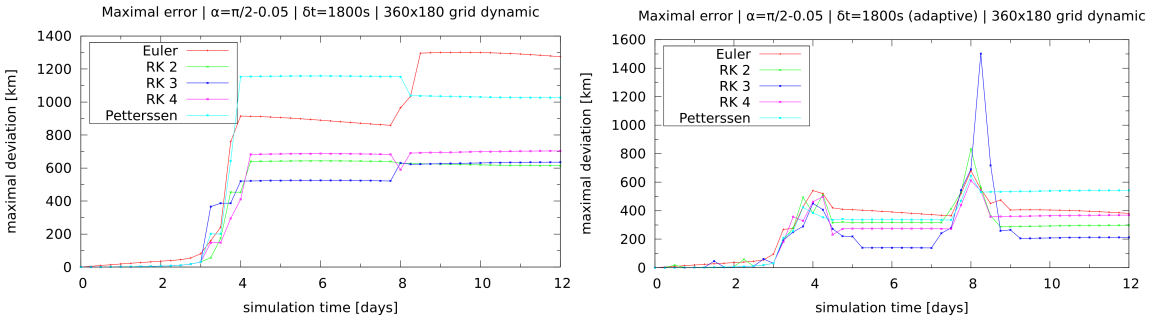


Figure 3.10: Maximum transport deviations for a dynamic wind field with a time step of 1800 s.

system. The problem is associated with the singularity of the cosine. The higher the latitude, the smaller is the value of the cosine function and the higher is the change in longitude for a constant Cartesian distance. An eastward distance of 100 meters leads to a change of around 2 degrees for the longitude at a latitude of 89.9°. At a latitude of 89.99° this distance already leads to a change of over 20 degrees. This rapid increase of the change in longitude can cause problems concerning the floating point representation. Moreover, integration errors are worsened, when the longitude is changed so much at one step and the CFL condition is violated.

Since there is no easy solution to fix this problem with the used coordinate system, an attempt was made to at least minimize the integration errors at the poles to get better results. Therefore an adaptive advection procedure was developed that can use every presented integration scheme. It reduces the time step by a factor of 2 when the absolute latitude coordinate is greater than 86°. If the absolute latitude coordinate is greater than 88° the time step is divided by 4. Finally the time step is divided by 8, if the absolute latitude coordinate is greater than 89°. When a time step was reduced, multiple iterations are done at one model step to keep the time for all air parcels consistent. This workaround shows good results for the constant wind field, without increasing the computation time noticeably. Figure 3.8 contrasts the adaptive and usual model. The maximum errors are about 4 times lower than before. For the dynamic wind field also an improvement is visible, but the difference is much smaller (Figure 3.10). The problem with this workaround is that it is possible that results are worse than before, because more model steps are done near the pole, which can lead to greater numerical errors because of the inaccurate floating point numbers. The test case with $\alpha = \pi/2 - 0.05$ shows a peak in deviation for the Petterssen scheme that is higher than all deviations of the original method. In this case the deviation decreases again, but it can not be assumed that this is a usual behavior. In the end the deviation are also smaller than for the original method.

The greatest errors of the analytical tests are caused by the coordinate system. The interpolation and integration are unproblematic compared to the numerical and accuracy problems at the poles. An adaptive advection module can at least reduce the accuracy problem and helps to satisfy the

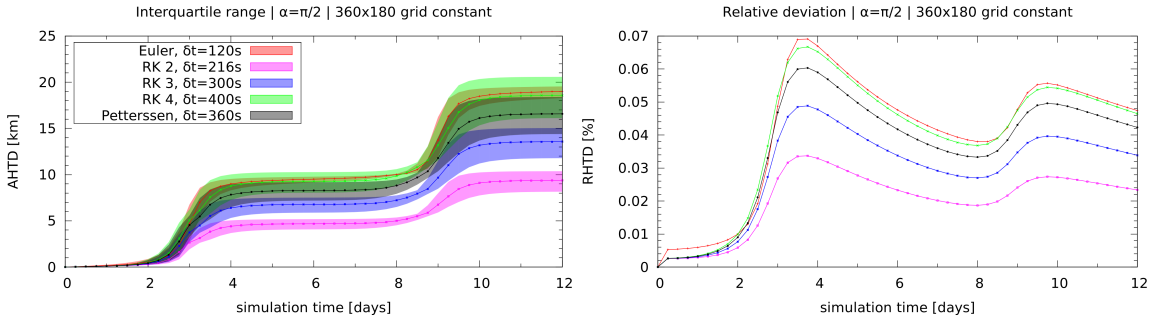


Figure 3.11: Interquartile range and median of the transport deviations with a constant wind field (left). Relative transport deviations with a constant wind field (right).

CFL condition. Nevertheless the problem can not be prevented completely. A different coordinate system could be used near the poles to remove this error source. The following simulations do not use the adaptive advection to keep the number of integration iterations identical.

**Runtime-dependent comparison**

Since most methods give similar results, the computation time can be used as evaluation criterion. To do so, simulations with individual time steps for the different integration schemes were done that lead to comparable run times. So the reached accuracy with a fixed computational time is compared. Figure 3.11 shows the transport deviations of the exact solutions and the other simulations. The midpoint method (RK2) is more exact than every other integration method if the computation time is equal. Second most efficient is the RK3 method. The highest order RK4 scheme is only a bit better than the first order Euler method, regarding the performance. The iterative Petterssen scheme is third best in this comparison. Since the convergence criteria and the maximum number of iterations can be varied, it is possible that a better configuration can be found. However, it makes no sense to optimize the configuration for this test, because real wind fields show completely different conditions and a new setup has to be found.

Another interesting point is that the Euler method has the shortest interquartile range. The distance of the $p_{0.9}$ and $p_{0.1}$ deciles show the same results. Therefore the spread of the parcels is smaller even if the deviation is bigger. The higher the order of an integration scheme, the greater is the spread. This could be caused by a faster increasing global truncation error. Small deviations in the parcels position have a greater influence if the local integration of the wind vector is more exact. Nevertheless all methods give very good results with their time steps, like the relative errors of less than 0.1 % shown in Figure 3.11 prove.

All test suggest that every integration scheme can be used and works well with a fitting time step. The small differences shown above are too small for a qualitative evaluation. However this

test did not consider vertical wind and variations. The differences between the methods could also be increased by greater time steps. Simulations on different computers have shown that the selected time steps to equalize the computation time are not valid in general. As the accuracy does not differ much at all, the midpoint method should remain the most efficient method on every computer.

### 3.2.5 Test for the linear interpolation and grid resolution

As mentioned before, the trajectory model interpolates the gridded input wind field linear in time and space. Trajectory models usually use a low order interpolation for the wind field and a higher order implementation for the integration (Bowman et al., 2013). The error caused by this interpolation is unknown and is hard to estimate in normal simulations because the wind is given on a grid and not exactly known. Since it could be a great error source, that could be weakened by implementing a higher order interpolation procedure, the interpolation errors are analyzed in this section. The wind of all analytical test cases can be exactly calculated for every location and time. Therefore a simulation that does not use gridded input files but the exact wind field can be run and is thus not affected by interpolation errors.

To verify that a linear interpolation is sufficient, the deviations of this simulation and other simulations, that use gridded wind data of different resolutions, were calculated. The gridded wind data has a horizontal resolution of 2.5°, 1° and 0.5°, which are typical resolutions of global meteorological data products. A temporal resolution of 6 h and 3 h was used for the time depended dynamical wind. Since the test only considers horizontal movement, it gives no information about possible vertical interpolation errors. To reduce the numerical integration error to a minimum, the Runge-Kutta scheme of order 4 was used as integration scheme with a time step of 30 s. As a consequence all remaining deviations are caused by the interpolation method and not by the integration scheme.

Figure 3.12 shows the transport deviations of the simulation with the exact wind information and the simulation with the gridded data for the constant wind field in the upper plot. The shown divergence has to be caused by the linear interpolation of the wind data, because all other factors are identical for both simulations. All four test cases show equivalent behavior. Notable differences in the plots are the stronger increments around day 3 and 9 of the test cases with $\alpha = \pi/2 - 0.05$ and $\alpha = \pi/2$. The reason for this behavior was already discussed in section 3.2.4. It is caused by the coordinate system and not by the interpolation. Apart from that all curves show a linear growth. The maximum error with a grid resolution of 144×72 points is about 6.2 km, a grid resolution of 360×180 points leads to a maximum deviation of about 1.1 km and the finest resolution of 720×360 points has a maximum deviation of about 0.3 km. The 360×180-grid contains 6.25 times more data than the coarsest grid. This is very close to the factor of the error. The factor between the data amount of the 720×360-grid and the middle resolution is 4, which again is the factor between the errors. Consequently the error caused by the linear interpolation and the amount of data seems to be linked by a linear dependence. A mean divergence of 6 km after 12 days is very good, considering a

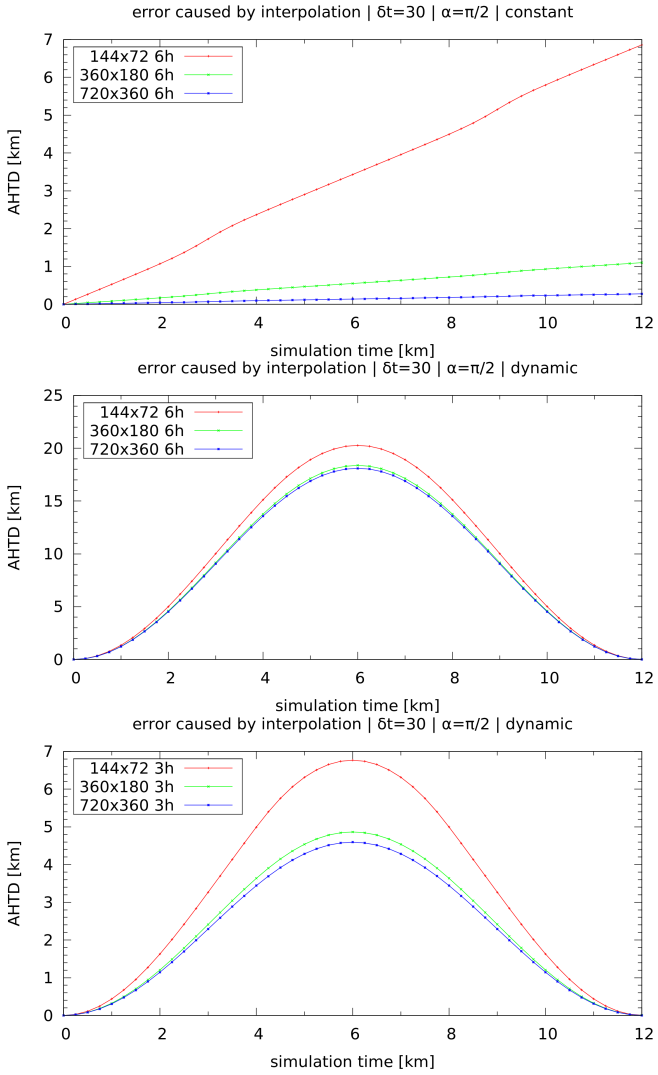


Figure 3.12: Transport deviations of the interpolated and exact wind with a constant or dynamic wind field and different grid resolutions. Only deviations for the case $\alpha = \pi/2$ are shown, all other test cases with other values for $\alpha$ are very similar.

mean travel distance of about 40,000 km for the constant wind and about 25,000 km for the dynamic wind.

The deviations for a dynamic wind field are shown in the middle plot of Figure 3.12. The curves indicate that the wind speed and direction is an important factor for the interpolation error. The gradient of the error curves behaves like a sinus curve. It increases until the third day. After that is goes back to zero at day 6. The curve reached the greatest negative increase on day 9 and goes back to zero till day 12. The error itself reaches its maximum at day 6. After that the error decreases again. Finally no significant deviation is left after 12 days. The reason for the recession of the error is that the errors of the first half travel and those of the second half cancel each other out. Since a perfect periodic wind field is unrealistic, only the first 6 days are taken into account for evaluating the interpolation error. The maximum error is about three times higher with a dynamic wind field. This greater error can be caused by the temporal interpolation. Even a fine grid of $720 \times 360$ points does not produce advantages if the temporal resolution is large. Therefore a second set of simulations with a 3 h temporal resolution was made for the dynamic wind field. Its deviations are shown in the lower plot of Figure 3.12. The finer resolution doubles the amount of data, but the error shrinks by a factor of about 3. This indicates that the temporal resolution is the greatest error source.

All in all the interpolation error is very small and should not affect the analytical test cases in this chapter. However the resolution still plays a major role for simulations with real wind fields, because the variations are greater. Nevertheless it seems that a grid resolution of $360 \times 180$ and 6 h is a good tradeoff between the amount of data and the interpolation error. For finer resolutions a reduction of the temporal resolution to 3 h should be considered first, as the tests for the dynamical wind have shown. The accuracy can only be slightly improved with a finer resolution in space. Reducing the temporal resolution brings about better improvement with the same amount of data.

This agrees to the results of my seminar thesis (Rößler, 2014) and the resulting paper (Hoffmann et al., 2015). The best simulation results were achieved with a horizontal resolution of $360 \times 180$. The second best data product had a resolution of only $288 \times 144$, but a temporal resolution of 3 h. Both data products lead to significant better results than the data product with a very high resolution of $1440 \times 720$. These differences are of cause not only due to the grid resolution, but it is a contributing factor.

## 3.3 Test with real atmospheric conditions

### 3.3.1 Definition of the test case

Real wind fields are much more complex than those used for the analytical test cases. This causes completely different conditions for the integration schemes. To find out if the conclusions of the analytical test are valid more generally a test case with real wind field was defined.

To create a start condition that is as fair and generic as possible, parcels are spread over the entire globe. Therefore, different regions divided by latitude and altitude were defined. The latitude ranges are the north and south pole [-90°:-65°] and [65°:90°], mid-latitudes [-65°:-20°] and [20°:65°], and the equator [-20°:20°]. The altitude was divided into the ABL [0:2 km], free troposphere [2:8 km], the border between troposphere and stratosphere [8:16 km], and the stratosphere [16:32 km]. This leads to 20 different regions that have different conditions in terms of temperature, pressure, and wind speeds. In every region 100,000 parcels were distributed randomly, so that they are weighted equally in the mean transport deviations. Since the atmospheric conditions also depend on the season of the year, the first of January and the first of July 2011 were used as starting times. The duration of the simulation was 21 days. The ERA-Interim data product used as meteorological input data was already described in section 2.2.

Simulations with all five integration schemes and all time steps mentioned in section 3.2.4 (2, 4, 8, 15 and 30 minutes) were done. Transport deviations were used to evaluate the simulations again. As reference, a simulation with a time step of 30 seconds and the RK 4 integration scheme was made. The small time step and high order scheme were assumed to lead to a reasonably good reference solution. The mean deviations and quantiles in combination with the run time were used to find a stable and fast configuration.

### 3.3.2 Analysis of the accuracy

At first the horizontal transport deviations are compared. Figure 3.13 shows the absolute and relative transport deviations for a time step of 30 minutes. Again all higher order methods diverge similarly strongly form the reference solutions. The curves for the RK4 and RK3 method and the curves of the RK2 and Petterssen scheme are on top of each other. Both curve pairs are side by side, their absolute deviation at the end is only 100 km or 0.5 % for the relative error. It is clear that the Euler method leads to much bigger deviations with such a large time step, but other time steps show the same manner. The relative transport deviations are always twice as big as the other methods' deviations.

All curves have a small gradient at the beginning, which grows polynomially until day 12 or 13. This is caused by the local truncation errors that sum up. The small deviations at the beginning still allow a good prediction of the new position, but when the errors become too big, the parcel is affected by complete different winds and drifts off very fast.

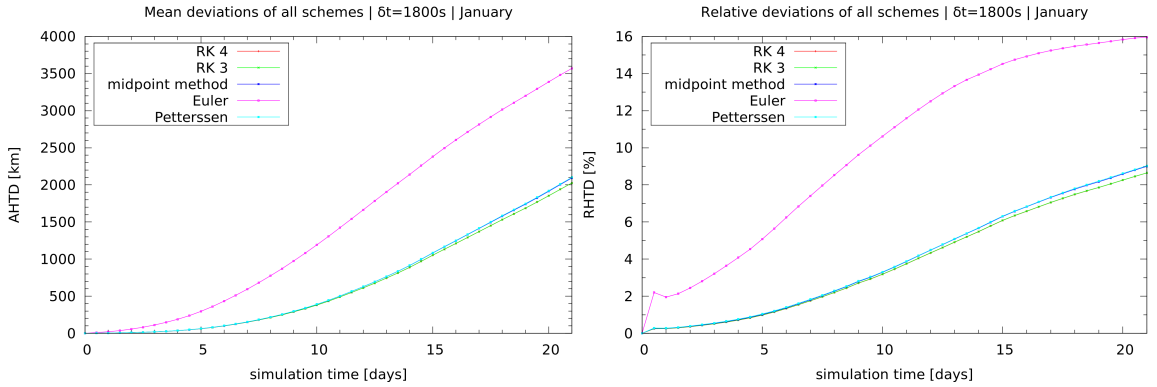


Figure 3.13: Absolute and relative horizontal transport deviations for all integration schemes with a real wind field and a time step of 30 minutes.

After day 16 the gradients of the curves slow down. This is in particular shown by the curve of the relative error of the Euler method, which seems to reach a plateau at around 16 %. The reason for this behavior of the slope are the outliers, that once diverge too strongly and fast drift off very far. The distance of such parcels to the reference then is rather random. The deviation of those outliers can not grow forever, since it is limited by the half circumference of the earth. Since the Cartesian distance is calculated in this case the maximum distance is two times the Earth's radius. When the parcels reach their maximal distances, the increase of the absolute and relative deviations drop again. The relative error will reach a maximum, because the absolute deviation is limited but the traveled distance continues to grow.

The previous relative errors show that a time step of 30 minutes is too long for a 21 day simulation. For a comparison of the spread of the parcels a more realistic time step of 8 minutes is used. It leads to a relative deviation of 1 % after 7 days, 3 % after 14 days, and 6 % after 21 days for the higher order methods. The Euler's deviations for the same times are 3.5, 9 and 12 %. Figure 3.14 shows the median of the transport deviations of all methods and the distribution of the transport deviations for the midpoint scheme. The curves of the median of the transport deviations show a constant polynomial growth. This robust measure is not falsified by the outliers like the mean value. The figure shows curves for time steps of 8 an 30 minutes. Only the Euler and RK 4 method are shown, because no difference between the higher order schemes is visible. In comparison to the mean deviations of Figure 3.13, the median is notably lower than the average (roughly 500 km or 25%).

The right plot shows the distribution of the deviation for the midpoint method and a time step of 8 minutes. All other integration schemes of a higher order show only marginal differences to this plot and are left out. The median is the border between the purple and blue region. Its value at

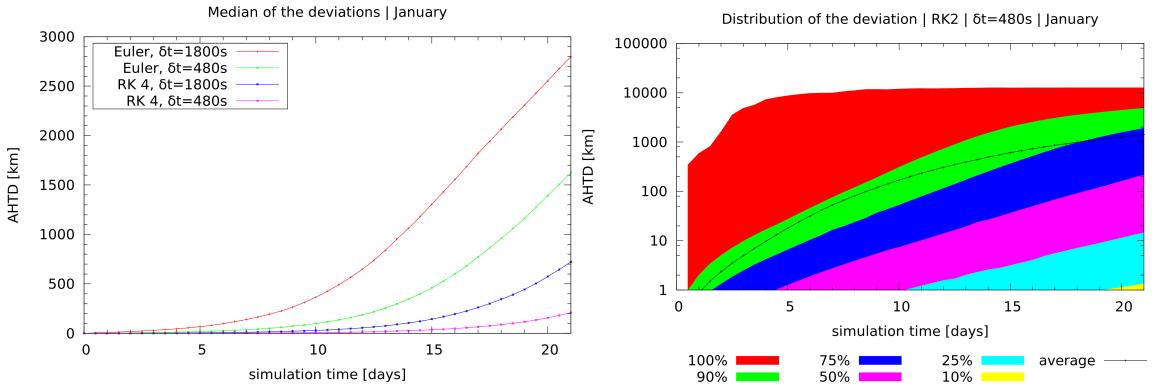


Figure 3.14: Median of the horizontal transport deviations for the Euler and RK 4 method and time steps of 8 and 30 minutes (left). The median deviation of the RK 2, RK 3 and Petterssen scheme are nearly equal to the RK 4 method. Distribution of the absolute transport deviations for the midpoint method and a time step of 8 minutes (right).

the end is around 200 km. The interquartile range is about 20 to 1,100 km. The average traveled distance is around 15,700 km. By assuming that all parcels have traveled comparably far, 75% of all parcels have a relative error of below 7% after 21 days. The black line shows the average deviation and is again notably higher than the median. The red area shows that the maximal error increases very fast and settles down at around 13,000 km.

The horizontal deviation can again not really be used to determine which higher order scheme is the best concerning the accuracy. Long simulations with great time steps are a bit better with the RK 4 and RK 3 methods, but the relative error caused by great time steps is too big. The time step had to be reduced anyhow to get reasonable results and in that case other methods are comparably good. Thus higher order schemes do not show relevant benefits, even with large time steps. The median and distribution of the deviation have shown, that a relatively great part of the parcels diverge strongly. To keep the number of outliers small, the time step should be chosen smaller than the mean deviations suggest. Moreover, very long simulations are difficult and require even smaller time steps than the minimal time step of 2 minutes.

Now vertical transport deviations are finally taken into account. The vertical deviations are at least as important as the horizontal transport deviations, because the wind field varies strongly in both directions and a change in altitude can cause great changes in the horizontal wind. Figure 3.15 shows the vertical absolute and relative transport deviations for the same simulations with a time step of 30 minutes. Apart from the outliers in the beginning of the relative errors, the plots show the same behavior than the horizontal wind. This shows that all methods are as good in the vertical as in the horizontal.

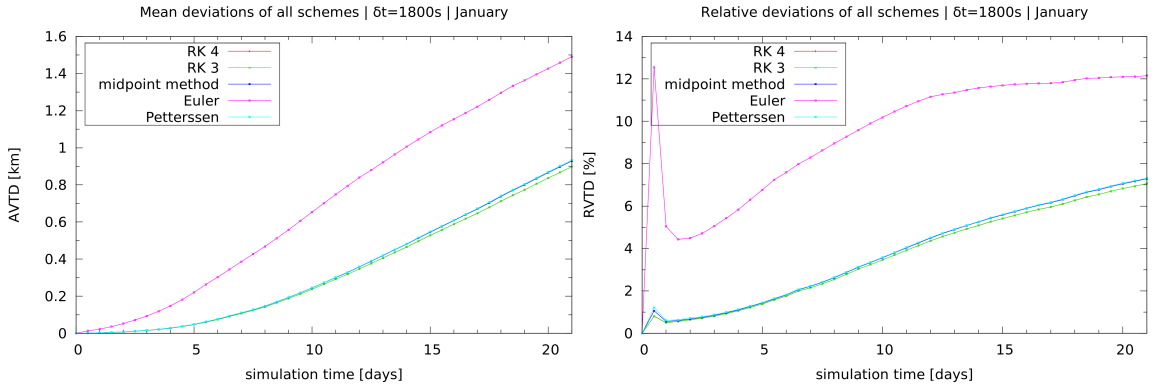


Figure 3.15: Absolute and relative vertical transport deviations for all integration schemes with a real wind field and a time step of 30 minutes.

### Time step choice

When a time step is chosen, the duration of the simulation and the used integration method have to be considered. Previous transport deviations have shown that long simulations can not use long time steps because the error sums up too fast. Now an overview on fitting time steps for simulations with different durations is given. Figure 3.16 shows the absolute and relative horizontal and vertical transport deviations for the Euler and RK2 method after 7, 14, and 21 days for different time steps. The RK3 method, RK4 method, and the Petterssen scheme are very similar to the RK2 method and are left out. The relative horizontal and vertical deviation are similar in general. For a one week simulation the horizontal deviations are a little higher and lower for 2 and 3 week long simulations. The relative deviations for the RK2 method stays below 2.5 % with time steps up to 30 minutes for short simulations of 1 week. After 2 weeks the relative errors are about 6 %, so the time step should be reduced to 480 s to get relative errors of around 3.5 %. If the simulation is 3 weeks long, the time step should be reduced to at least 240 s, to keep the errors below 5 %. Simulations that last longer than one week should not be made with the Euler method, because even a time step of 120 s leads to relative deviations of 5 %.

In conclusion, test with analytical and real atmospheric conditions have only shown accuracy differences between the first order Euler method and all other methods of orders from 2 to 4. This is likely caused by the linear interpolation of the wind field. Higher order methods can not profit from more function calls, because the wind field shows only minor changes in time and space. A higher order interpolation method could be implemented to test, if greater differences between the integration schemes can be found. Since the temporal resolution of ERA-Interim is only 6 h it is questionable, if higher order interpolation methods are noticeably better. So all in all the RK2 or

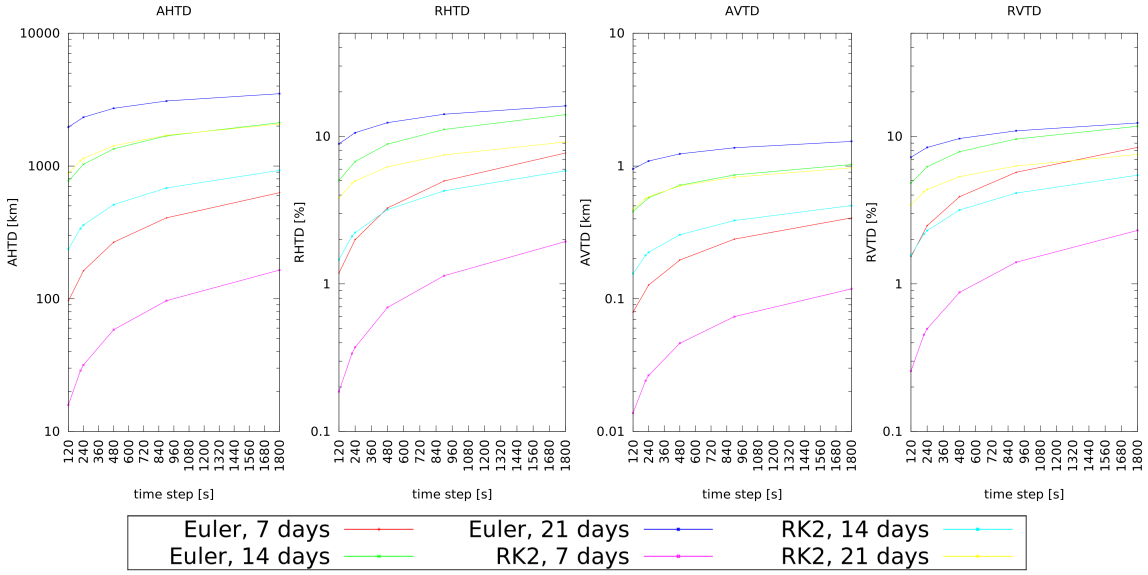


Figure 3.16: Relative horizontal and vertical deviations of simulations with the Euler or RK 2 method for different time steps after 7, 14, and 21 days. The other higher order schemes behave like the RK 2 method.

midpoint method shows the best performance in analytical and realistic wind fields and will be used for further simulations. The suggested time steps for a 1, 2, or 3 week long simulations are 480, 120 and roughly about 30 s respectively, to achieve a relative error of around 1 %.

| | | altitude | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | method and | [0:2] | | [2:8] | | [8:16] | | [16:32] | |
| lat. | time | RHTD | RVTD | RHTD | RVTD | RHTD | RVTD | RHTD | RVTD |
| [65:90] | Euler, Jan. | 12.9 | 11.2 | 13.1 | 9.4 | 4.8 | 4.9 | 1.0 | 1.8 |
| | Euler, July | 8.5 | 8.5 | 12.2 | 9.6 | 5.8 | 6.2 | 0.0 | 0.1 |
| | RK2, Jan. | 5.3 | 5.3 | 6.0 | 5.0 | 2.2 | 2.4 | 0.3 | 0.6 |
| | RK2, July | 2.4 | 3.1 | 5.0 | 4.7 | 2.7 | 3.1 | 0.0 | 0.0 |
| [20:65] | Euler, Jan. | 9.9 | 8.3 | 9.3 | 7.9 | 3.7 | 4.0 | 0.7 | 1.2 |
| | Euler, July | 8.1 | 7.6 | 8.7 | 7.9 | 5.9 | 6.0 | 0.1 | 0.2 |
| | RK2, Jan. | 2.1 | 2.1 | 2.4 | 2.5 | 1.0 | 1.2 | 0.1 | 0.2 |
| | RK2, July | 1.3 | 1.4 | 1.9 | 2.1 | 1.4 | 1.7 | 0.0 | 0.0 |
| [20:-20] | Euler, Jan. | 3.0 | 3.3 | 2.6 | 2.7 | 3.1 | 3.5 | 0.2 | 0.4 |
| | Euler, July | 3.5 | 3.9 | 2.5 | 2.6 | 2.6 | 2.9 | 0.2 | 0.5 |
| | RK2, Jan. | 0.1 | 0.2 | 0.1 | 0.1 | 0.1 | 0.2 | 0.0 | 0.0 |
| | RK2, July | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 | 0.2 | 0.0 | 0.0 |
| [-65:-20] | Euler, Jan. | 6.8 | 7.9 | 6.7 | 6.8 | 4.5 | 5.0 | 0.2 | 0.4 |
| | Euler, July | 8.6 | 8.6 | 8.1 | 7.7 | 3.5 | 3.5 | 0.9 | 1.5 |
| | RK2, Jan. | 0.9 | 1.3 | 1.5 | 1.9 | 1.1 | 1.5 | 0.0 | 0.0 |
| | RK2, July | 2.1 | 2.3 | 2.4 | 2.6 | 1.1 | 1.2 | 0.1 | 0.2 |
| [-90:-65] | Euler, Jan. | 4.5 | 3.4 | 7.2 | 6.4 | 2.3 | 3.5 | 0.1 | 0.3 |
| | Euler, July | 6.7 | 7.1 | 11.6 | 9.1 | 6.7 | 6.0 | 0.2 | 0.5 |
| | RK2, Jan. | 0.9 | 0.8 | 2.3 | 2.3 | 0.8 | 1.4 | 0.0 | 0.1 |
| | RK2, July | 2.4 | 2.9 | 4.8 | 4.3 | 3.0 | 3.0 | 0.1 | 0.3 |

Table 3.3: Relative horizontal and vertical transport deviations for simulations itemized for different altitude and latitude ranges.

### 3.3.3 Impact of different atmospheric conditions

Table 3.3 shows the relative horizontal and vertical transport deviations for simulations after 2 weeks with a time step of 120 s, like suggested before, in January and July separated by the previously defined regions in altitude and latitude. Before general statements about the model's accuracy are made, the integration methods are compared. The table shows deviations for the Euler method and RK2 method, that only differs faintly from the other methods. The differences vary strongly with the different regions. The RK2 method shows at least 2 times smaller deviations, but is up to 30 times better around the equator. Roughly the Euler method is worst in the ABL and lower troposphere and improves with height. The RK2 method shows the same behavior. This is caused by the decreasing variability of the wind and increasing distance of the vertical layers of the meteorological data.

In the following the RK2 method is analyzed. The differences of the horizontal and vertical regions are significant. The model's accuracy highly depends on the conditions given by the wind

field. At first the vertical layers are compared. In the stratosphere the smallest deviations occur, the maximal deviation is 0.6 %. The standard deviations presented in Figure 3.17 show that the wind is less variable at higher altitude. The transition region of the troposphere and lower stratosphere shows the second smallest deviations, the maximal error is 3 % here. The regions above 8 km show greater vertical than horizontal deviations, while the lower layers have a higher horizontal deviation. The ABL comes next in this simulations. Its maximal deviations are 5 %, while the lower troposphere shows deviation of up to 6 %. The ABL was expected to cause the largest errors due to its variability, a reason why the errors are a little smaller than in the troposphere could be, that the Earth's geographies affects the wind field used in MPTRAC. MPTRAC currently does not consider the Earth's geography and assumes that the surface is always at 0 km altitude. When mountains are present, no wind field is given in the input files and the wind from above is extrapolated down to the surface. This reduces the variability of the wind in the ABL and could reduce the average deviations. So it can be assumed, that the higher variability, due to the resolution and physical phenomena, is the main factor for the deviations.

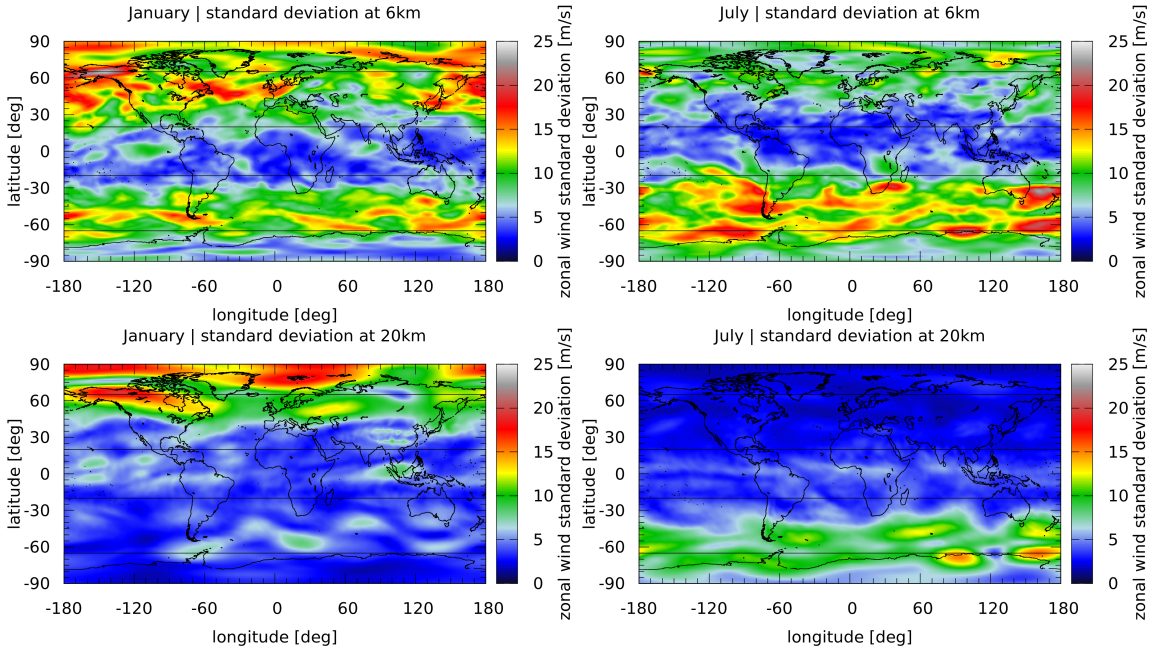


Figure 3.17: Standard deviations of the zonal wind of the first 3 weeks of January and July 2011 at 6 and 20 km altitude.

Now the different latitudes and seasons are compared. The horizontal and vertical deviations at the equator are with around 0.2 % notably lower than in the other regions, because the wind is

less turbulent in that region, as Figure 3.17 shows. The wind at the equator is not affected by the seasons and shows the same deviations for January and July.

The other regions show greater deviations of up to 2.5 % at mid-latitudes and 6 % near the poles. They also show a dependency on the season. In January the northern hemisphere shows greater deviations and in July and the southern hemisphere shows greater deviations in July. The comparison of the zonal standard deviations (Figure 3.17) show that the wind changes with the seasons and causes the different behavior of the deviations. In general the northern hemisphere tends to have a slightly greater deviations.

The high variation of the relative errors indicate, that the accuracy is too small in some regions. The iterative Petterssen scheme should theoretically show advantages under variable conditions, because a minimal accuracy of the wind vector is granted. But the method does not differ from the RK2 method. This shows again, that reducing the time step is the only way to increase the accuracy, when a second order integration scheme is used.

### 3.3.4 Performance of the integration schemes on JURECA

The simulations were carried out on the new supercomputer JURECA of the Forschungszentrum Jülich. It has 1872 compute nodes with two 12 core Intel Xeon E5-2680 v3 Haswell CPUs with a clock speed of 2.5 GHz. Those CPUs support simultaneous multithreading with the Intel Hyperthreading Technology, so each node has 24 physical and 48 logical threads. Most nodes have 128 GiB of DDR4 memory with a clock speed of 2133 MHz. Some nodes are equipped with 256 or even 512 GiB memory. All simulations were run on a single node, so only the OpenMP parallelization was used. Since the trajectory model has to execute the same operations for a large number of parcels, that do not interact with each other, a good speedup can be expected.

The previous comparisons have shown that the midpoint-method is the integration method of choice, so the general Runge-Kutta implementation was replaced by an optimized version for the midpoint method. Instead of loops, direct function calls were made. The use of local variables was also reduced to a minimum. With these changes the midpoint function itself needs only 40 % of the computational time that was needed before. Since the most expensive operations are done in sub functions, where a linear 4-D interpolation is done, the overall computation time of the advection was only improved marginally by 1-3 %. A comparison of all integrations schemes in a short test simulation with a constant time step of 480 s for all integration schemes has shown, that the number of function calls does not slow the program down as badly as expected on the supercomputer JURECA. Due to optimizations of the compiler and caching, the time needed per function calls shrinks with the number of necessary function calls. Table 3.4 shows the individual commutation time and time needed for one function call. The slowdown for a second function call is only 30 % instead of 100 %. If the Petterssen, RK3, or RK4 method had shown notable advantages, these methods could be implemented without slowing MPTRAC down badly. However, the midpoint-method has proven

to be as exact as the other methods, if the time step is adjusted to bring reasonable results, so its optimized implementation will be used in the future. A performance and scaling analysis of MPTRAC as a whole is done in section 4.5, after the changes in the diffusion module are introduced.

| **method** | **time [s]** | **slowdown [%]** | **function calls** | **time per call [s]** |
|---:|---:|---:|---:|---:|
| Euler | 28.1 | 0 | 1 | 28.1 |
| opti. midpoint method | 36.7 | 30 | 2 | 18.4 |
| Petterssen scheme | 53.1 | 89 | (avg.) 2.9 | 17.7 |
| RK1 | 29.6 | 5 | 1 | 29.6 |
| RK2 | 41.1 | 36 | 2 | 20.6 |
| RK3 | 49.4 | 76 | 3 | 16.4 |
| RK4 | 56.7 | 102 | 4 | 14.2 |

Table 3.4: Comparison of the computation time of different integration methods and called subfunctions. The slowdown is given in comparison to the Euler method.

# Chapter 4

# Diffusion

## 4.1 Dispersion modeling

As mentioned in the general description of the diffusion in section 2.1.2, MPTRAC has two different diffusion components. The first component simulates small-scale turbulent motions by adding a random stochastic perturbation to the position. The second component occurs on a larger scale. Mesoscale turbulence simulates turbulent motions that are caused by eddies of the wind field but not covered in meteorological data files due to the lack of resolution. Both diffusion components are discussed in the following in more detail.

**Turbulent diffusion**

The modeled random perturbation is independent from the wind field, but considers the structure of the atmosphere. The different conditions in the atmosphere's layers are described by different models for the diffusion in the troposphere and stratosphere. The complex diffusion of the atmospheric boundary layer, that was described before, is not modeled in particular and is treated like every other point in the troposphere. MPTRAC focuses on the free troposphere and the stratosphere, where transport of volcano emissions takes place, so this proceeding is unproblematic.

The tropopause between both layers, however, is considered and modeled as a 1 km high transition layer, where a combination of the troposphere and stratosphere diffusion is applied. The height of the tropopause has to be calculated for each position to find out if a parcel is in the troposphere, stratosphere, or in the transition layer. MPTRAC follows the definition used by the World Meteorological Organization (WMO) to determinate the height of the tropopause: "The [tropopause is the] boundary between the troposphere and the stratosphere, where an abrupt change in lapse rate usually occurs. It is defined as the lowest level at which the lapse rate decreases to 2 °C/km or less, provided that the average lapse rate between this level and all higher levels within 2 km does not exceed 2 °C/km.". The altitude dependent diffusion was implemented in this thesis, previously both diffusion components were always applied. The reason why the diffusion is split up in the FLEXPART model that was used as reference is not discussed in detail in their documentation. Most likely it has to do with the vertical stratification of air masses that suppresses vertical mixing below

the tropopause. Turbulence is most likely caused by wind shear between those stratified layers and causes horizontal diffusion. In the stratosphere these motions are less common and effects that cause vertical motion like gravity waves or high convection events are more important.

The magnitude of the diffusion is controlled by two parameters. The first is $D_h$, the horizontal diffusivity and the second parameter is $D_v$, which is the vertical diffusivity. Horizontal and vertical turbulent diffusion are applied by adding a random displacement to the position of each air parcel:

$$\Delta x = \sqrt{\omega_{trop} \cdot 2D_h/\Delta t} \cdot \zeta_x \cdot \Delta t \tag{4.1}$$

$$\Delta y = \sqrt{\omega_{trop} \cdot 2D_h/\Delta t} \cdot \zeta_y \cdot \Delta t \tag{4.2}$$

$$\Delta z = \sqrt{\omega_{strat} \cdot 2D_v/\Delta t} \cdot \zeta_z \cdot \Delta t \tag{4.3}$$

With horizontal and vertical diffusivity parameters $D_h$ and $D_v$ and Gaussian distributed random numbers $\zeta$ with zero mean and standard deviation one. It is assumed that the diffusion of the three wind components are independent from each other. Existing cross corelations are neglected, because they have little impact on large scale simulations (Stohl, 1998). The weight factors $\omega_{trop}$ and $\omega_{strat}$ steer the weight of the horizontal and vertical diffusion component. The parcel's altitude is compared to the height of the tropopause at is longitude and latitude. If it is below the tropopause, we set $\omega_{trop} = 1$ and $\omega_{strat} = 0$. So in the troposphere only horizontal diffusion takes place and only Equations (4.1) and (4.2) have an impact on the diffusion. If the parcel is at least 1 km above the troposphere, the weights are set to $\omega_{trop} = 0$ and $\omega_{strat} = 1$. So in the stratosphere only vertical diffusion is applied. If the parcel is in the transition layer, the altitude ($\Delta z = z - z_{trop}$) over the tropopause is used as $\omega_{strat} = \Delta z/1\,\text{km}$ and the weighting factor for the troposphere is $\omega_{trop} = 1 - \omega_{strat}$.

The default value for the parameter $D_h$ is $50\,\text{m}^2/\text{s}$ and the default value for $D_v$ is $0.1\,\text{m}^2/\text{s}$. Both default values are suggested by Stohl et al. (2005) and used in the dispersion model FLEXPART. Observations of radon in the atmosphere have resulted in average vertical eddy coefficients ($D_v$) in the range of 0.1-0.5 $\text{m}^2/\text{s}$ (Liu et al., 1984). Since the constant $D_v$ is applied independent of the presence of eddies, a low approximation for the diffusivity is probably good. Another study that used backwards trajectories to reconstruct observed diffusion in the lower stratosphere, suggests smaller values for $D_v$ in the range of 0.001-0.1 $\text{m}^2/\text{s}$ (Legras et al., 2005). The stratification of the atmosphere limits the vertical motion. Especially the diffusion inside the polar vortex is notably lower than outside. Pisso et al. (2009) also analyzed the dispersion of CO in the free troposphere and found out that values in the ranges of 0.3-1 $\text{m}^2/\text{s}$ and 1-1000 $\text{m}^2/\text{s}$ for $D_v$ and $D_h$ respectively lead to the best simulation results.

A different approach to define the diffusion coefficients is the Smagorinsky diffusion, as used

in the COSMO model (Baldauf and Zängl, 2012). The Smagorinsky diffusion is a pure horizontal diffusion that calculates the diffusion coefficient in dependence on the zonal and meridional wind and the meteorological input data resolution. In this thesis this kind of diffusion or variable diffusion coefficients in general are not further discussed.

The great range for the parameters suggested by the different studies and the different model approaches show that the impact of the diffusion varies under different conditions and is not easy to model. In the following a sensitivity test for the diffusivity coefficients is presented. A comparison of the mesoscale and turbulent diffusion is also included in the sensitivity test.

**Mesoscale diffusion**

The term "mesoscale "refers to the size of typical weather phenomena and is divided into three subclasses, meso-gamma (2-20 km), meso-beta (20-200 km) and meso-alpha (200-2000 km). Wind motions caused by events on the meso-gamma or meso-beta scale are not resolved in the meteorological data files due to the limited resolution. The turbulent diffusion discussed earlier also does not reproduce these effects correctly. Since mesoscale motions can have a significant impact on the dispersion, it should be taken into account at least in an approximate way. Following Stohl et al. (2005), the mesoscale dispersion can be described by the Langevin equation, e.g.

$$\Delta\psi_{n+1} = r \cdot \Delta\psi_n + \sigma_\psi \cdot \sqrt{1 - r^2} \cdot \zeta \tag{4.4}$$

$$r = 1 - 2 \cdot \Delta t / \tau_L. \tag{4.5}$$

This equation is solved for all three wind components, so $\Delta\psi_{n+1}$ is the movement caused by the mesoscale dispersion in the direction of a chosen wind component. The first part of the term considers the fluctuation of the previous time step, the factor r controls the temporal correlation. It depends on the model time step $\Delta t$ and the Lagrangian time scale $\tau_L$, that is given by the temporal resolution of the input data. The value $\zeta$ is a Gaussian distributed random number with mean 0 and standard deviation 1. The only unknown variable is $\sigma_\psi$, which is the subscale standard deviation of the wind component. To estimate this standard deviation, the empirical variance of the surrounding grid points is calculated. The corresponding standard deviation is multiplied with a parameter $D_{MESO}$ and then is used to solve the mesoscale Langevin equation. So finally no real mesoscale phenomena are described, but simulations have shown that the accuracy can be improved by this empirical approximation (Stohl, 1998).

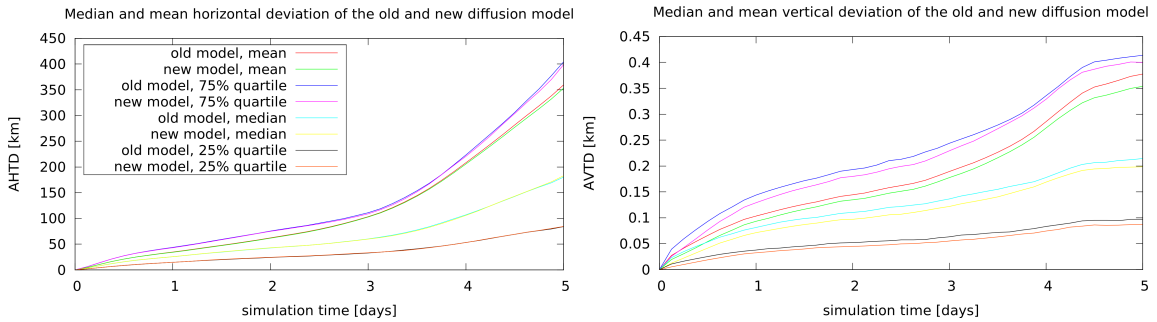


Figure 4.1: Mean and quartile values of the horizontal and vertical transport deviations.

## 4.2 Test for the transition layer above the tropopause

This test case was made to analyze the diffusion near the tropopause, the border between the troposphere and stratosphere. The old implementations of uniform diffusion did not separate between troposphere and stratosphere. To find out how strong the new implementation differs from the old one, a test case that concentrates on the tropopause was defined. Parcels are distributed regularly in a rectangle westwards of Europe given by the longitudes [-15:-5], latitudes [35:45] and heights [8:12]. The tropopause in that region is located in 9 to 11 km altitude at the start time 4th of January 2011. The simulation only takes 5 days, so that the deviation due to advection is neglectable. A simulation without diffusion, a simulation with the mesoscale diffusion and the new separated diffusion model and a simulation with the mesoscale diffusion and the old uniform diffusion was made.

The transport deviations of the simulations to a reference without diffusion were calculated. They are presented in Figure 4.1. It shows that the new model is a little less diffusive than the old one. The mean horizontal and vertical transport deviations are always higher for the old model. Also all quartiles show higher deviations for the old model. This result was expected, since the new model implements the same diffusion terms, but uses only one diffusion term or a combination of both, depending on the parcels altitude. In contrast the old model always applies horizontal and vertical diffusion and thus was assumed to cause a greater diffusion. The differences are so small, because a factor of 2 was added to the diffusivity constant that was missing in the diffusion calculation of the old implementation. So the weaker diffusion caused by the separation was compensated by a $\sqrt{2}$ times higher standard deviation. Test cases that have an imbalanced number of air parcels below and above the tropopause may show greater differences from the old model in the horizontal and vertical direction.

## 4.3 Consistency tests without advection

### 4.3.1 Turbulent diffusion

An easy consistency test is to turn off the advection and validate that the parcels distribute as expected. This is done for the turbulent diffusion in this section. The mesoscale diffusion will be analyzed separately, since it behaves differently. In this test a point source for 10,000 air parcels is defined and the mean distance of the air parcels to the origin and the standard deviations in all directions is calculated. The expected behavior is, that the standard deviations of the zonal, meridional and vertical direction increase proportionally to the square root of time. The exact standard deviation is $\sqrt{2 \cdot D \cdot t}$, where $D$ is a diffusivity parameter and $t$ the elapsed time.

Three initial conditions are used that differ in the release height for the air parcels. In the first case all air parcels are released in the free troposphere, so that only horizontal diffusion takes place. The second one only simulates parcels in the stratosphere and only vertical diffusion occurs. The last test checks if the tropopause is implemented correctly. All parcels start 0.5 km above the tropopause, which is constant at 12 km altitude for this test case, so that turbulent diffusion in both directions takes place equally strong. The simulations last 10 days and are repeated for three origin locations (0° N, 0° W), (45° N, 45° E) and (90° N, 0° W), which should not show any differences.

Three aspects can be validated with this test. Firstly, the correct separation of the turbulent diffusion models of the different layers can be reviewed. The second aspect is that the diffusion coefficients $D_h$ and $D_v$ steer the dispersion correctly. The mean deviation has to be linear depended form the square root of $D_h$ and $D_v$. The third aspect is that the dispersion has to be independent from the model time step, so that different time steps lead to equal results.

At first the differences caused by the origin's positions are discussed. For the vertical diffusion in the stratosphere and tropopause all three positions show the same results. For the horizontal diffusion the locations (0° N, 0° W) and (45° N, 45° E) show identical results for all test cases. They only differ minimally due to different random numbers that were generated. A difference is that the start point (90° N, 0° W) causes a little smaller horizontal standard deviations. The reason for this is the singularity of the poles. The diffusion is applied to the longitude and latitude coordinate separately and since the circle of latitude has a length of zero at the poles, the diffusion actually only perturbs in the latitude direction. After some time steps this behavior is weakened because parcels have delocalized from the pole and horizontal diffusion can occur. As it is very unlikely that parcels are exactly at a pole, this small inconsistency is acceptable. A solution for this problem would be to select a random angle between 0° and 360° and move the parcel by a random distance in that direction. To illustrate that the change is not necessary, a fourth start location (89° N, 0° W) was tested. It shows that a distance of 1° to the poles is already enough to bring the same results like the correct working test cases.

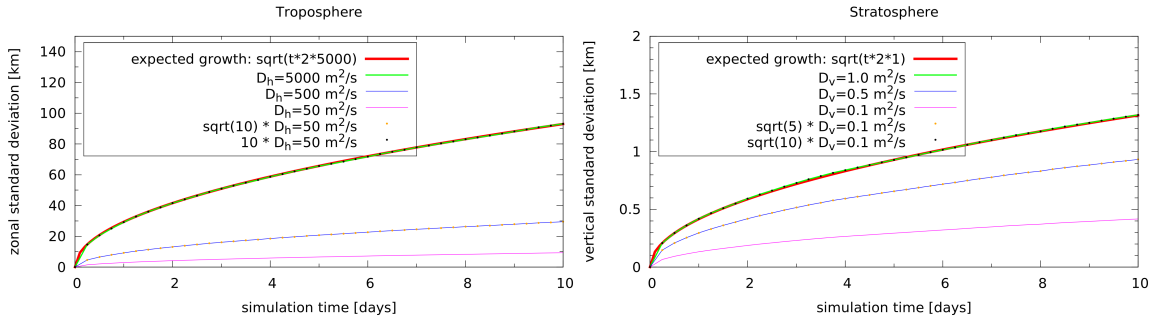


Figure 4.2: Zonal and vertical standard deviation of the distance to the origin. The meridional standard deviation is equal to the zonal standard deviation.

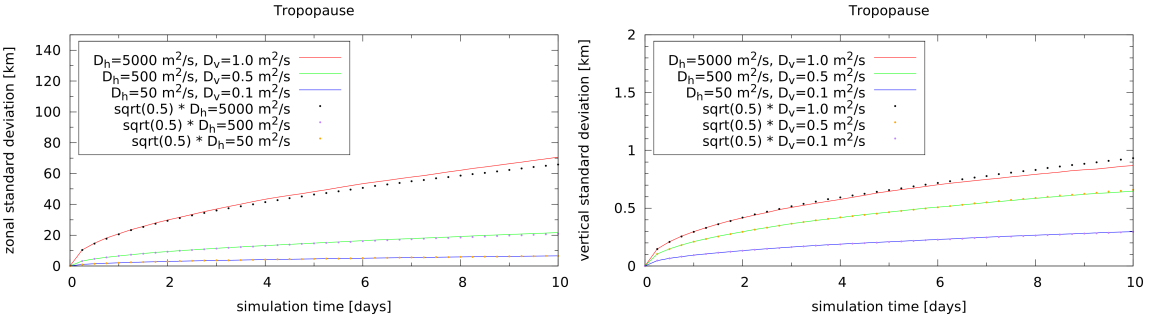


Figure 4.3: Zonal and vertical standard deviation in the tropopause compared to the troposphere or stratosphere respectively.

Now the effect of the time step is analyzed. This test was done separately for the horizontal and vertical component of the diffusion. Simulations with time steps of 2, 8 and 60 minutes were made and compared with each other. They show all the same standard deviations with minimal differences due to the random numbers.

Since the time step and location have no influence, the quantitative analysis of the diffusion is done with the start location (0° N, 0° E) and a time step of 8 minutes with various diffusivity constants. The horizontal diffusivities $D_h$ were increased from 50 to 500 and $5000\,m^2/s$. This should increase the standard deviation by a factor of $\sqrt{10}$ each. Vertical diffusivities $D_v$ of 0.1, 0.5, and $1\,m^2/s$ were used, so the standard deviations should increase by factors of $\sqrt{5}$ and $\sqrt{2}$. Figure 4.2 shows the zonal horizontal and vertical standard deviations for the test cases in the troposphere and stratosphere. The dots in the figure show that the requested factors concerning the diffusivity parameters really present are between the curves. The standard deviations grow linearly with the square root of time and match the analytical solution.

The standard deviations for the test in the tropopause are expected to be approximately smaller

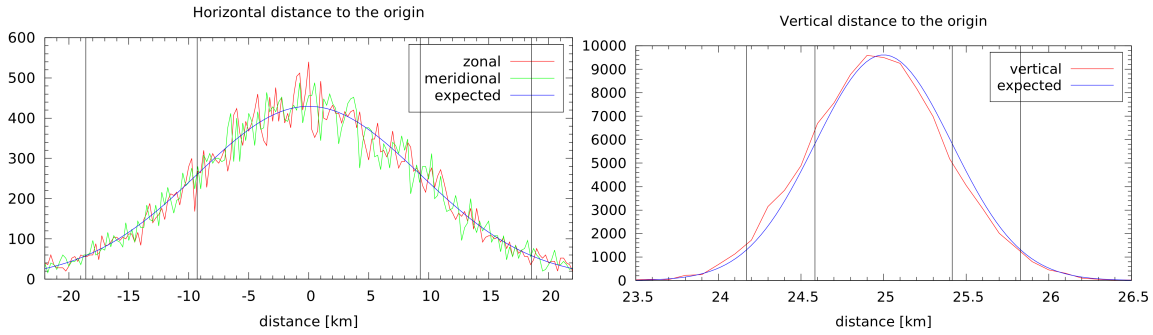


Figure 4.4: Histogram for the distribution of the parcels in zonal, meridional, and vertical direction in comparison to the expected Gaussian distribution. The one and two sigma range is is shown by vertical lines.

by a factor of $\sqrt{0.5}$, because the parcels start in the center of the tropopause where both diffusion components are applied with a factor of $\sqrt{0.5}$ for the magnitude. However, the parcels move in regions with stronger horizontal or vertical diffusion and the final standard deviation may vary a little. Figure 4.3 shows the zonal and vertical standard deviations for the tropopause in comparison to the troposphere or stratosphere respectively. The dots in the lower plots show that the standard deviations in the tropopause are smaller by around $\sqrt{0.5}$ in both directions and the combination of the components work correctly.

To check if the parcels are Gaussian distributed, the distances to the center in zonal, meridional, and vertical direction after 10 days were analyzed separately. Figure 4.4 shows a histogram for the zonal, meridional, and vertical distribution and the expected distribution. The distribution of the simulation fits very well to the exact solution.

In conclusion, all test cases have shown that the turbulent diffusion works correctly in the troposphere, stratosphere, and the transition layer at the tropopause. The default parameters of $D_h = 50\,\mathrm{m}^2/\mathrm{s}$ and $D_v = 0.1\,\mathrm{m}^2/\mathrm{s}$ lead to a mean horizontal standard deviation of 9.3 km and vertical standard deviation of 420 m. How great the influence of this diffusion is on a trajectory has to be tested in simulations with advection.

### 4.3.2 Mesoscale diffusion

At first the same test like for the turbulent diffusion was executed with only mesoscale diffusion activated. The consistency check for the time step and influence of the mesoscale parameter $D_{MESO}$ has shown that, the time step has again no influence on the diffusion. The value of $D_{MESO}$ has approximately a linear influence on the mean distances of the parcels to the origin. The distances are compared instead of standard deviations in this case, because the distribution of the air parcels

caused by mesoscale diffusion is not Gaussian and the standard deviation does not describe the spread correctly. Since the wind field differs in space and time different parameters lead to different conditions and a strict linear behavior can not be expected. In contrast to the turbulent diffusion, the different location and height lead to different results. For the start locations with latitudes of 0° and 45°, the horizontal distance is strong for altitudes of 12.5 and 25 km, while the distances at 4 km altitude are only half as big. For the start location at the north pole this is the other way around. The highest horizontal deviations take place at an altitude of 4 km, while the other two start locations show only half as high distances. The maximal horizontal distances match at the different locations. Their vertical distances show even greater variations.

The great differences show, that the mesoscale diffusion is affected by the wind field, that varies at different locations and heights. To find out how strong the influence of this kind of diffusion varies under different condition, a new test case was defined. Air parcels start on different circles of latitude with a fixed altitude, so that only the longitude coordinate changes. The start longitude goes from -180° to 179.9° in steps of 0.05°. Heights between 4 and 25 km with steps of 1 km and latitudes between -85° and 85° with steps of 5° were used. This leads to 770 test cases with around 7200 parcels each. The mean horizontal and vertical distance to the origin is calculated after 10 days. To get an impression if the results are valid in general, two starting times, the first of January and the first of July, are used.

Figure 4.5 shows that in some regions the distances are significantly higher. The distances are in the range of 40 to 180 km in the horizontal and 100 to 400 m in the vertical direction. Regions near the tropopause at altitudes of 4 to 20 km show increased distances. The noticeable region between latitude 30° N and 60° N or 30° S and 70° S is caused by the subtropical and polar jet streams. Usually jet streams are only around 100 km wide, since the distances are averaged over parcels at longitudes between -180° and 180° and the jet streams vary in latitude, the jet streams appears much bigger than they really are. The standard deviations presented in the previous section are an upper bound for the mean distances of the parcels. This shows that the default value $D_{MESO} = 0.16$ leads to a at least 4 times stronger diffusion considering the mean distance.

Since diffusion is partly caused by wind shear and velocity gradients and differs in magnitude in dependence of wind speeds and variability, the results of the mesoscale diffusion are realistic.

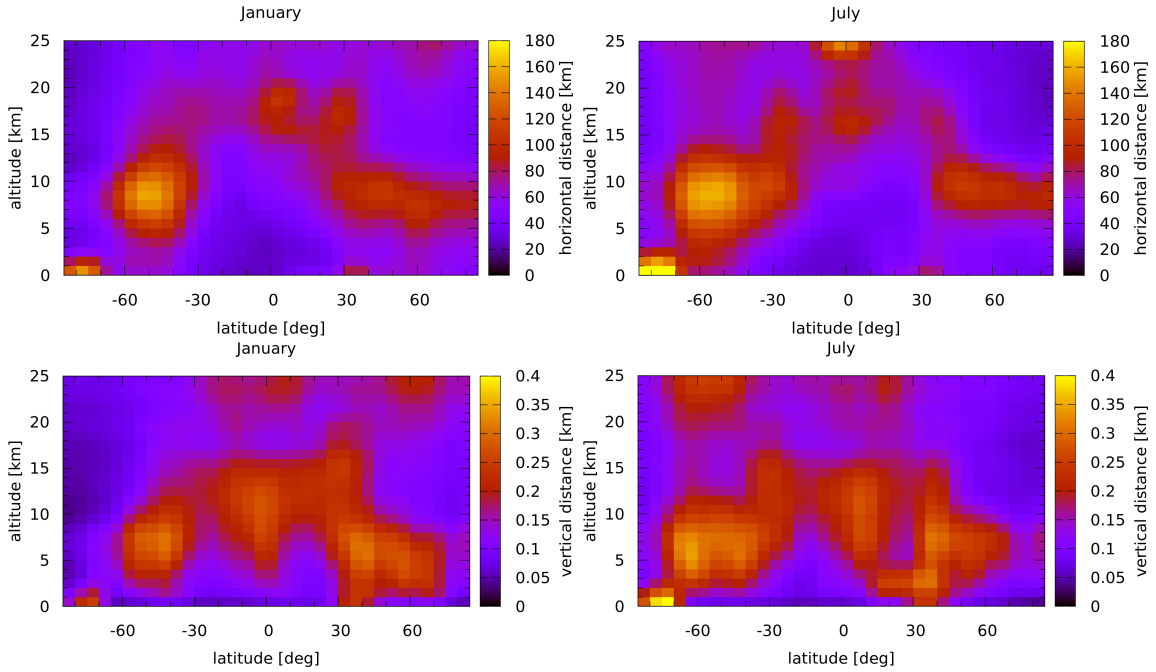


Figure 4.5: Mean horizontal and vertical distance to the origin in dependence of the start latitude and altitude.

## 4.4 Sensitivity test under real atmospheric conditions

It is expected that the diffusion shows different behavior in real meteorological wind fields. Therefore three start points at places with different meteorological conditions were defined. They are (65° S, 77.5° W) near the south pole, (0° N, 70° W) in South America and (42.5° S, 14° W) in Europe. The simulations start on the first of July with 44,000 air parcels with altitudes between 4 and 25 km. Groups of 2,000 parcels were released every 1 km in height. This was done because the wind varies strongly in height, but only the spread of the parcels due to the diffusion is meant to be checked. When parcels were released evenly distributed in height, they would be scattered strongly by the advection because of the variation of the wind field in height. If a single height level was used the different diffusion components could not be tested at once. Furthermore, the test was intended to have realistic and general conditions.

A simulation with the default diffusion parameters ($D_h = 50m^2/s, D_v = 0.1m^2/s, D_{MESO} = 0.16$) and one without advection were used as references for transport deviations. But transport deviations are falsified by the error caused by advection. Therefore the mean coordinate for every group of 2,000 parcels with the same start location was calculated. The vertical and horizontal distances and standard deviations to the center points are individually calculated for all 22 parcel-groups. After that all 22 mean distances and standard deviations are averaged. The mean distances to the center points and the standard deviations of this distance are a good indicator for the extent of a point source.

The mesoscale and turbulent diffusion were analyzed separately by turning modules on and off, respectively. This was done to find out how the individual diffusion parameters steer the diffusion in real case scenarios. Another aim of this test is to find out which part of the diffusion has the greatest impact in real case scenarios. Therefore different simulations were executed that had one diffusion parameter increased. The simulation with the greatest spread shows which parameter has the strongest influence.

Table 4.1 shows the horizontal and vertical distances and standard deviations after 5 and 10 days for the start location (42.5° S, 14° W). The three locations show a large variability in the distances and standard deviations, but special characteristics of individual regions are hard to detect. No clear statement about which kind of diffusion is most relevant in which region can be made. It is also hard to tell which region is most diffusive. The horizontal distances and standard deviation over Europe are relatively small, but on the other hand the vertical deviations are the greatest in most cases. The distances of the start location and the south pole (65° S, 77.5° W) are for most cases about twice as large as the start location over south America (0° N, 70° W) after 5 days, but after 10 days the start location in south America has as large horizontal distances as the other test cases and only 20% smaller vertical distances. This shows that the impact of diffusion in simulations with advection highly depends on the wind field.

| parameter | | | horizontal | | | | vertical | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $D_h$ | $D_v$ | $D_{MESO}$ | $\bar{h}_5$ | $\sigma_{h5}$ | $\bar{h}_{10}$ | $\sigma_{h10}$ | $\bar{v}_5$ | $\sigma_{v5}$ | $\bar{v}_{10}$ | $\sigma_{v10}$ |
| 50 | 0.1 | 0.16 | 396.8 | 266.8 | 1886.0 | 836.3 | 0.427 | 0.544 | 0.819 | 1.024 |
| 50 | 0.1 | 0 | 154.2 | 117.8 | 1199.3 | 598.9 | 0.259 | 0.332 | 0.648 | 0.814 |
| 200 | 0.4 | 0 | 329.5 | 253.5 | 1846.4 | 918.6 | 0.476 | 0.611 | 0.962 | 1.192 |
| 5000 | 10 | 0 | 1239.2 | 708.9 | 3320.3 | 1639.9 | 1.842 | 2.339 | 2.804 | 3.545 |
| 50 | 0 | 0 | 29.4 | 21.5 | 484.2 | 338.4 | 0.048 | 0.062 | 0.290 | 0.379 |
| 0 | 0.1 | 0 | 129.7 | 97.7 | 1066.0 | 485.5 | 0.218 | 0.279 | 0.428 | 0.544 |
| 0 | 0 | 0.16 | 343.4 | 227.2 | 1727.3 | 713.6 | 0.317 | 0.407 | 0.667 | 0.825 |
| 0 | 0 | 0.32 | 527.0 | 331.1 | 2359.4 | 927.9 | 0.507 | 0.648 | 0.871 | 1.083 |
| 0 | 0 | 1.6 | 1399.4 | 829.1 | 3101.1 | 1609.0 | 1.451 | 1.840 | 1.918 | 2.396 |
| 500 | 0.1 | 0.16 | 409.6 | 277.5 | 1920.7 | 846.1 | 0.430 | 0.550 | 0.822 | 1.029 |
| 50 | 0.4 | 0.16 | 509.9 | 351.6 | 2340.0 | 1001.3 | 0.582 | 0.735 | 1.055 | 1.304 |
| 50 | 0.1 | 0.32 | 564.0 | 355.5 | 2458.3 | 982.5 | 0.577 | 0.734 | 0.969 | 1.206 |

Table 4.1: Mean horizontal and vertical distance and standard deviation to the mean point of the parcel-groups for the start location (42.5° S, 14° W) after 5 and 10 days.

In general the start locations show the same behavior when parameters are modified, so relative statements about differences of the modules are possible with this analysis. At first the relative change of the turbulent diffusion after 5 days is discussed (row 2 to 4). The horizontal and vertical deviations behave similar. When the parameters are multiplied by 4 distances and standard deviations increase by a factor of around 2. When the parameters were increased by a factor of 100, the deviations increased by a factor of 8 in average, that is less than expected.

The distances after 10 days don't increase as strong, when the parameters are changed. However, the parcels are dispersed much faster in the second half of the simulation, which is most likely caused by the wind field that differs for the already distributed parcels. The smaller the distances after 5 days, the greater is the gap to the distances after 10 days. The factors for this gap are between 2.2 and 10 for the horizontal distance and between the 1.5 and 3 for the vertical distance.

Now the horizontal and vertical diffusion components are compared (row 5 and 6). Even if the $D_h$ is 500 times larger than $D_v$, the vertical diffusion has a four times greater influence on the horizontal and vertical dispersion. This shows that the variance of the wind is much higher in the vertical than the horizontal direction. The standard deviations of the vertical distance also demonstrate that vertical wind is more variable than the horizontal wind. They are relatively high compared to the horizontal standard deviations and indicate a high variation. This could be caused by the finer vertical resolution of the meteorological input data.

The mesoscale diffusion has a stronger influence than both turbulent diffusion components (row

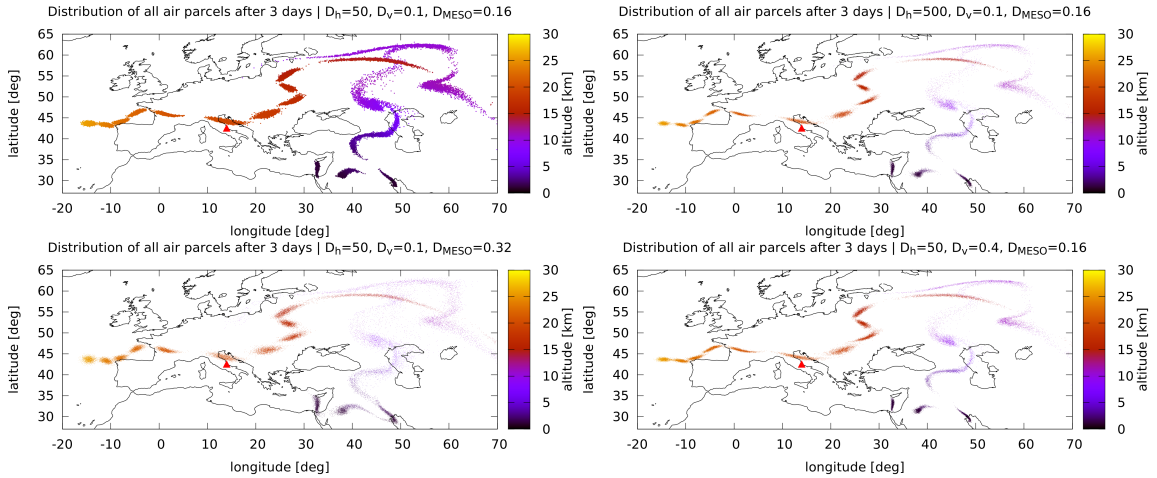


Figure 4.6: Distribution of the air parcels over Europe with different diffusion parameters after 3 days.

7). The simulation which uses all three components and default parameters shows not much greater distances than the simulation with only mesosale diffusion (row 1 and 7). This shows that the dispersion is dominated by the mesoscale diffusion in this case. The test case in South America shows only small deviations between the influence of the vertical and mesoscale diffusion (see Table 4.2). However, the mesoscale diffusion is still the most diffusive component.

When the parameter $D_{MESO}$ is changed, it does not really show a linear influence on the distances and standard deviations. This was already mentioned before in the test case without advection, in real scenarios with a variable wind field, different parameters lead to different conditions at different times and so the distance is changed completely. The three start locations show this as well. E.g., the simulations in this section lead to horizontal distances of 230.3, 343.4, and 575.4 km after 5 days.

Now the simulations with one increased parameter are analyzed (row 10 to 12) and compared to the default simulation. Figure 4.6 shows the spread of the air parcels for these simulations after three days, to describe the differences of the components more descriptively. The horizontal dispersion can be increased to 500 without any great changes in the diffusivity, in comparison to the simulation with default parameters as the two upper figures show. When the horizontal dispersion is increased to 0.4, the horizontal and, more clearly, the vertical errors increase. The lower plot on the right shows that the individual clouds are elongated. Doubling the $D_{MESO}$ parameter has the greatest effect, like the wide clouds of the lower plot on the left show. The influence of the other diffusion components with default parameters is so small that it makes no great difference if they are activated or not (row 9 and 12).

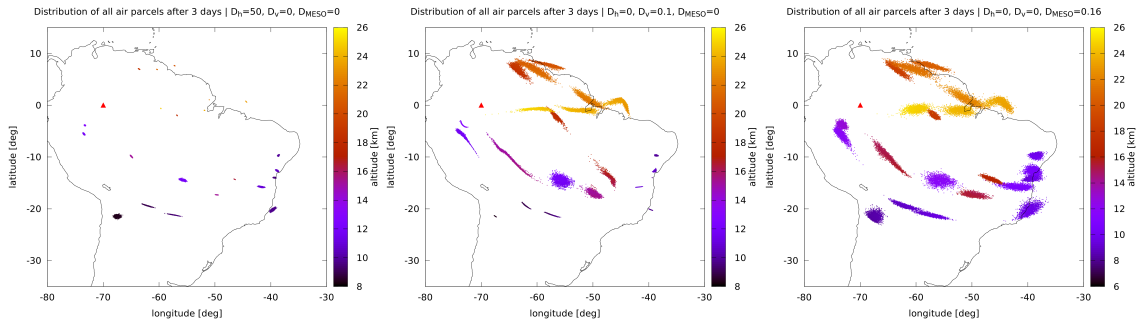


Figure 4.7: Distribution of the air parcels over South America with only one diffusion component activated after 3 days.

| **parameter** | | | **horizontal** | | | | **vertical** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $D_h$ | $D_v$ | $D_{MESO}$ | $\bar{h}_5$ | $\sigma_{h5}$ | $\bar{h}_{10}$ | $\sigma_{h10}$ | $\bar{v}_5$ | $\sigma_{v5}$ | $\bar{v}_{10}$ | $\sigma_{v10}$ |
| 50 | 0 | 0 | 21.3 | 15.2 | 177.0 | 112.5 | 0.024 | 0.030 | 0.134 | 0.172 |
| 0 | 0.1 | 0 | 190.3 | 138.0 | 1208.4 | 678.4 | 0.182 | 0.227 | 0.447 | 0.561 |
| 0 | 0 | 0.16 | 230.3 | 159.0 | 1458.9 | 771.2 | 0.201 | 0.254 | 0.556 | 0.698 |

Table 4.2: Mean horizontal and vertical distance and standard deviation to the mean point of the parcel-groups for the start location (0° N, 70° W) after 5 and 10 days for the parameters used in Figure 4.7.

Figure 4.7 shows the distribution of the air parcels of simulations that had only one diffusion component activated with the default value. The simulations with only one component show more distinctly that the horizontal diffusion has nearly no influence on the simulations. In this case the vertical diffusion is nearly as powerful as the mesoscale diffusion. Table 4.2 shows the mean distances and standard deviations after 5 and 10 days for the parameters shown in Figure 4.7 for a better comparison with the previously presented simulation. The horizontal and mesoscale diffusion leads to smaller distances than before, but the vertical diffusion has a significantly larger impact, even if the vertical distance is comparable to the other simulation.

This test has shown, that diffusion in combination with advection behaves differently from the theoretical spread. The different wind fields that affect the dispersed parcels, lead to an unpredictable dispersion. Instead of a Gaussian distribution around a point, long drawn filaments are formed. These filaments are mostly caused by the vertical diffusion component. A pure horizontal diffusion leads to results that are much more similar to the simulations without advection. The emissions released by a point source are very dispersive, so that it can be assumed that a spread of a point source leads to clouds similar in size than those shown in Figure 4.6. However, a validation of the

realistic behavior of the diffusion module of MPTRAC is not made in this thesis.

## 4.5 Performance of the diffusion model and MPTRAC

### 4.5.1 Mesoscale diffusion

The original MPTRAC version calculates the subscale standard deviation for every parcel individually for every time step. Since it is calculated for three wind components that have four dimensions (zonal, meridional, vertical and temporal), the computation is relatively cost intensive.

A second implementation was made to improve the performance for a large number of air parcels and short time steps. When the meteorological data is read, the subscale standard deviation is computed for every grid point. To make sure that a meteorological input data with a high resolution does not worsen the performance, this is only done if the number of parcels multiplied with the number of model steps needed for the time difference of the meteorological data files is greater than the number of grid points.

A test case with ERA-Interim data and around 710,000 air parcels has shown that the new implementation is over four times faster than the old implementation, this does not consider the longer time spend in the input function. This part is slowed down by 80%, but the computation is so much faster, that the slowdown is compensated. The more parcels are simulated, the higher is the profit. In this case the mesoscale diffusion was speed up by a factor of 3, considering the longer time spent in the input function.

### 4.5.2 Turbulent diffusion

In the old model, no separation between the troposphere, tropopause or startosphere was done. In the consequence it is hard to keep the old performance. Since this diffusion has only a relatively small influence, it should not take disproportional computation time. The problem is that the height of the tropopause is needed for the air parcels' positions. The heights are calculated for every point of the longitude-latitude grid of the meteorological data file when it is read. After that an exponential smoothing is done. The smoothing is done by substituting every data point by a weighted mean of all initial tropopause heights. The weight factor decreases exponentially with distance, so that only neighboring data points have a great impact. When the original tropopause heights contain data gaps, they can be removed by this smoothing. Since all data points are considered for the smoothing of every single point, this algorithm has a complexity class of $O(n^2)$. With the high number of data points, this causes the input function to take around 37 times as long as before. As the weight factor decreases exponentially, far off points have a vanishing small impact. To increase the performance, points that would have a weight factor of below $10^{-6}$ are left out completely. With this improvement the slowdown is neglectable.

The calculation of the turbulent diffusion is 85% slower than before and needs nearly as much time as the mesoscale diffusion. Considering that a 3-D interpolation is done for the tropopause height at the parcels position and that a combination of the horizontal and vertical diffusion components is applied in the tropopause, this slowdown is acceptable. In addition the mesoscale and turbulent diffusions still need 50% less time than the old implementation, considering the test case with 710,000 parcels.

### 4.5.3 gprof analysis

The performance analysis tool gprof was used to analyze the overall performance of the model and find functions that could be improved. As predicted, the interpolation of the wind is the most expensive part. The interpolation needs 55.8% of the full computation time and is divided on the three functions intpol_met_time (1.4%), intpol_met_space (6.75%) and intpol_met_help (47.65%). The second most expensive is the locate function that needs 19.65% of the time. After that comes the mesoscale diffusion with 8% and the Turbulent diffusion with 5.5%. Only 2% of the computational time is spend in the midpoint method. The method dx2deg, that converts a zonal Cartesian distance into a change for the longitude, needs 1.5% of the time. Thus changes that were made is this function to prevent inaccuracy and the division by zero problem, did not lead to performance loss. All mentioned functions are already implemented efficiently and are executed very fast, but they are called highly frequently. The only function that may allow some improvements is the locate function, that returns the grid indexes of the nearest data point of the meteorological grid for an air parcel's position. Since the function has not directly to do with the advection or diffusion, it was not tried to optimize it in this thesis.

### 4.5.4 Open-MP scaling

The optimized version with the RK2 method and the new diffusion scheme was used for a scaling analysis on JURECA. Simulations with 2 and 8 million air parcels were done with 1 to 48 cores. A good speedup is expected, because the air parcels' trajectories can be calculated independently. Half of the threads are not running on real physical cores, but hyperthreading can approximately increase the performance by 50%. So a speedup of 36 would be ideal. Figure 4.8 shows that the speedup is not as good as expected. When 8 cores are used the speedup is at around 7. When more cores are used the efficiency decreases further and a final speedup of around 20 is reached. Technically hyperthreading should not be used with less that 24 cores, so the overhead that comes with the parallelization seem to cause the slow increase of the speedup. Different schedule methods were tried out, dynamic scheduling is slower, while static (shown in the figure) and guided scheduling lead to comparable results. The absolute runtime reaches form 1.2 to 26 s and 4.5 to 92 s for the simulations with 2 or 8 million parcels respectively, so maybe even larger simulations can lead to

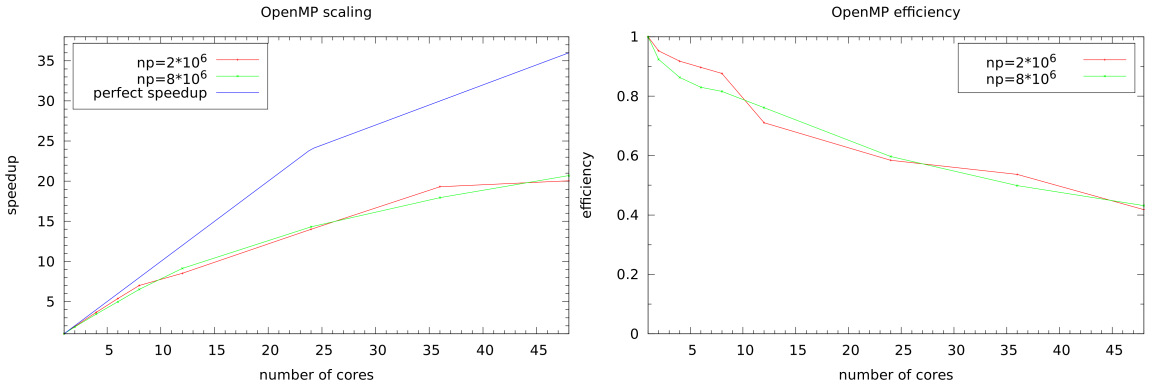


Figure 4.8: Speedup and efficiency of the OpenMP parallelization. The optimal speedup assumes that hyperthreading leads to a 50% performance increase.

a better speedup. The time measurement was done around the physics loop and is therefore not affected by delay due to input and output.

Previous scaling tests on JUROPA have shown a nearly perfect speedup of the physics calculations. The old version of MPTRAC was compared with the new one on the new supercomputer JURECA. The old model was significantly slower than the new model. Two cores take as much time as a single core with the new version. The speedup for the old version does also show a limited speedup. So the bad scaling is due to the new environment or the too small size of the problem and not due to changes in the code.

# Chapter 5

# Conclusion and Discussion

## 5.1 Summary

The aim of this thesis was to validate and improve the advection and diffusion modules of the trajectory dispersion model MPTRAC. The accuracy of the advection scheme and possible error sources were analyzed in analytical tests and for realistic atmospheric conditions. A validation of the correctness of the diffusion scheme was done for simulations without advection. Additionally,a parameter study for diffusivity coefficients was done in combination with advection. Finally, the performance of newly implemented functions was optimized and the performance of the MPTRAC model as a whole and the Open-MP scaling was analyzed.

The analytical tests have shown that the interpolation and integration methods are implemented correctly and that they provide a very good accuracy in the analytical wind field. The linear interpolation of the wind field works best with the midpoint method (a second order Runge-Kutta method). The model time step is primarily responsible for the accuracy this implies that the order of the integration scheme plays a minor role. However, a notable accuracy difference between the first order Euler method and the other methods was shown. Considering the integration schemes' performance, a second order method is the best choice for analytical test cases.

The analytical tests also revealed some error sources which worsened the accuracy of the model notably. Firstly, the used coordinate system causes problems at high latitudes. The circles of latitude which decrease in length with increasing latitude lead to numerical problems and the singularity at the pole itself has to be intercepted. These problems were solved as good as possible, by using the cosine function instead of a Taylor series and by assuming that very close to the pole no movement in zonal direction is necessary. Furthermore an adaptive time step choice near the pole was implemented to reduce the numerical error and help to satisfy the CFL condition.

A comparison of the interpolation error caused by using gridded wind informations on three typical resolutions has shown that matching resolutions in space and time is more important than the effective number of data points. It is better to use a horizontal resolution of 1° with a temporal

resolution of 3 h instead of a 0.25° horizontal and 6 h temporal resolution. However, the gridded analytical wind fields lead to only marginal deviations from the exact calculated wind and results for realistic wind fields can differ.

The tests for real atmospheric conditions have shown that integration errors depend on the conditions given by the wind field. A comparison of the integration schemes has not pointed to a sharper distinction of their accuracy, but the time step under real atmospheric condition has to be much shorter. Integration errors near the equator are very small at every altitude with relative errors of around 0.2 %. Possible reasons for small deviations at the equator are the lowest variation of the wind, the lowest wind speeds and the roughest effective resolution of the meteorological data in that region. The stratosphere in general shows good results with only marginally larger deviations at other latitudes. In the free troposphere and region around the tropopause an increase of the error with a decreasing height can be observed, while the errors of the ABL lay between them. All altitudes below the stratosphere show a dependence of the season. The deviations are greater on the northern hemisphere in January and greater on the southern hemisphere in July. The errors increase notably when the latitude comes closer to the poles. This is caused by the general increased perturbation of the wind field in the winter hemisphere and the finer effective resolution of the meteorological data files.

The midpoint method has shown the best performance in analytical and real atmospheric conditions for the integration of the wind. Methods of higher order do not lead to better accuracy, because the linear interpolation of the wind gives not enough information. In consequence, the time step is the main factor that steers the accuracy and not the order of the integration scheme. Since the global truncation error starts to increase rapidly after some time, it is very important that the time step is short enough to keep the local truncation error small in long simulations. A duration of 2 weeks should use a time step of 2 minutes or less, depending on the region the simulation is focused on. Simulations at the stratosphere or near the equator would allow a greater time step, but the other regions require a higher accuracy to be stable.

The common proceeding to use a linear interpolation of the wind field and a higher order integration method only makes sense, if a method of order two is used. The iterative Petterssen scheme that is widely used should not do more than two iterations, because attempts to increase the order above 2 would not increase the accuracy. Suggested time steps of up to an hour are clearly too big for simulations with durations of weeks. This comparison of the accuracy did not consider the errors that are contained in the meteorological wind field. It is possible that relative errors in the order of 10 % are acceptable, because the error caused by the wind field is even bigger. Validation for test cases with real wind fields are not considered in this thesis. Balloon trajectories could be used as validation, but they are problematic because they are also affected by other influences than the wind. Observations of emissions of volcanic eruptions lead to better reference trajectories, but the initial condition is not exactly known and the influence of diffusion is hard to determine. Therefore,

a real prove of correctness is difficult to achieve in real case scenarios.

Modeling of diffusion in Lagrangian dispersion models has turned out to be more complex and empirical than expected. Many theories rely on large simplifications and many uncertainties are left, e.g. the choice of subscale standard deviation of the wind or the diffusivity coefficients for turbulent diffusion. The partition of the horizontal and vertical turbulent diffusion has made the model a little less diffusive, but since a factor of 2 was missing in the old implementation, the difference is minor. The tests for the diffusion without advection have shown that the diffusion is implemented correctly and behaves like the theories predicts. The turbulent diffusion causes a Gaussian perturbation and the mesoscale diffusion takes care of missing wind perturbations that cause turbulence that is not covered by the meteorological data. In general, the mesoscale diffusion has a stronger impact on the spread of the air parcels. The vertical turbulent diffusion has the second strongest impact. Nevertheless the default parameters that were chosen according to FLEXPART still lead to a relatively small spread of the air parcels. Especially the horizontal diffusion parameter $D_h = 50m^2/s$ has shown little impact at all. It could be increased to $500\,\mathrm{m}^2/\mathrm{s}$ to have an impact of approximately the same order as the other components. Since the values suggested by other studies vary strongly, a general statement about which values are good is difficult. How the diffusivity parameters should be chosen most likely varies from case to case. Moreover, different diffusion models are used in different dispersion models that, however, should lead to comparable results. A validation of a correct diffusion model is difficult. The spread of emission can be compared to observations, but it is questionable, if the real fluctuation events that occur in the atmosphere are described by the model. In general, all diffusion models that obey the "well-mixing condition "(Thomson, 1987) are assumed to be physically correct. Like the figures for the diffusion with and without advection have shown, MPTRAC seems to satisfy this condition.

In conclusion, the main target to validate the correctness of the MPTRAC model as a whole and the advection and diffusions modules in detail was achieved. The advection module works correctly. Its accuracy can be controlled by the time step. The diffusion module causes a spread that can be adjusted with different parameters. The accuracy near the poles was improved and the performance of the advection and diffusion scheme was optimized. A general error estimation was done to give guidelines for the time step choice for simulations in different regions with different durations. So MPTRAC can be used for dispersion simulations above the ABL with small and large time scales and provides reliable results.

## 5.2 Outlook

For a better understanding of the diffusivity parameters in real scenarios, tests that can be compared to observations could be done. Volcanic eruptions can be considered as excellent test cases. The source and release time of the emissions are relatively well known and satellite measurements can detect $SO_2$ and volcanic ash. Multiple simulations with different parameters could be made and the shape of the emission clouds could be compared. If the modeled cloud is too small, this could be improved with a stronger mesoscale or horizontal diffusion, if the cloud is too short, it could be elongated with a higher vertical diffusion. Since the wind field has a strong influence on the effect of the diffusion, general results for the parameters require multiple test cases with different conditions. In addition, this test would act as a validation of the model in terms of realistic behavior.

The optimization of MPTRAC could also be continued. Since the advection and diffusion modules are independent from each other, they could be executed with different time steps. The advection requires a short time step for accuracy, but the diffusion is nearly independent of the time step. Since the mesoscale diffusion depends on the wind field and the parcels position, the CFL condition should still be accomplished. This also applies to the turbulent diffusion, because it requires the tropopause height at the parcel's position. The model time step is in general small enough so that a two or four times greater time step could be used for the diffusion without violation of the CFL condition.

Further improvements of the physic part in MPTRAC should include a modeling of the Earth's surface. Simulations in the ABL are problematic because parcels can reach positions were no real wind information is given. The extrapolated wind from above that is used in that regions prevents the model from crashing, but does probably not lead to correct results. Moreover, sedimentation of aerosols could be modeled if trajectories are stopped when they reach the ground. Some models also include geographic information in the calculation of diffusion. For instance, mountains can cause gravity waves which lead to a higher variation and mixing.

The integration methods seem to be limited by the linear interpolation in space and time. Especially the temporal component causes errors. A higher order interpolation could increase the accuracy and could give the possibility to use higher order integration schemes profitably. Significant changes in MPTRAC would be necessary for a higher order interpolation. The whole interpolation part had to be rewritten and the input function had to be changed. It could also lead to memory problems when meteorological data with a high resolution is used and more data has to be in the memory at once. Another possibility to increase the accuracy of the interpolation is to interpolate the meteorological data beforehand with external programs to create intermediate files.

# Bibliography


Baldauf, M. and Zängl, G.: Horizontal nonlinear Smagorinsky diffusion, COSMO newsletter, 12, 3–7, 2012.

Bowman, K. P., Lin, J. C., Stohl, A., Draxler, R., Konopka, P., Andrews, A., and Brunner, D.: Input data requirements for Lagrangian Trajectory Models, American Meteorological Society, pp. 1051–1058, 2013.

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., Kållberg, 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.

Forster, C., Wandinger, U., Wotawa, G., James, P., Mattis, I., Althausen, D., Simmonds, P., O'Doherty, S., Jennings, S. G., Kleefeld, C., Schneider, J., Trickl, T., Kreipl, S., Jäger, H., and Stohl, A.: Transport of boreal forest fire emissions from Canada to Europe, Journal of Geophysical Research: Atmospheres, 106, 22 887–22 906, doi:10.1029/2001JD900115, URL `http://dx.doi.org/10.1029/2001JD900115`, 2001.

Haynes, P.: {TURBULENCE} {AND} {MIXING}, in: Encyclopedia of Atmospheric Sciences, edited by Holton, J. R., pp. 2446 – 2450, Academic Press, Oxford, doi:http://dx.doi.org/10.1016/B0-12-227090-8/00439-5, URL `http://www.sciencedirect.com/science/article/pii/B0122270908004395`, 2003.

Heffter, J. L. and Stunder, B. J. B.: Volcanic Ash Forecast Transport And Dispersion (VAFTAD) Model, Weather and Forecasting, 8, 533–541, 1992.

Hoffmann, L., Rößler, T. F. P., Griessbach, S., Heng, Y., and Stein, O.: Lagrangian transport simulations of volcanic sulfur dioxide emissions: impact of meteorological data products, Journal of Geophysical Research, submitted, 2015.

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, Monthly Weather Review, 113, 1852–1867, 1985.

Legras, B., Pisso, I., Berthet, G., and Lefèvre, F.: Variability of the Lagrangian turbulent diffusion in the lower stratosphere, Atmospheric Chemistry and Physics, 5, 1605–1622, 2005.

Liu, S., McAfee, J., and Cicerone, R.: Radon 222 and tropospheric vertical transport, Journal of Geophysical Research: Atmospheres (1984–2012), 89, 7291–7297, 1984.

Maerker, C., Seidenberger, K., Erbertseder, T., Rix, M., Valks, P., and van Geffen, J.: Trajectory matching and dispersion modeling of volcanic plumes utilising space-based observations, in: Use of Remote Sensing Techniques for Monitoring Volcanoes and Seismogenic Areas, 2008. USEReST 2008., pp. 1–5, 2008.

Mazzocchi, M., Hansstein, F., Ragona, M., et al.: The 2010 volcanic ash cloud and its financial impact on the European airline industry, in: CESifo Forum, vol. 11, pp. 92–100, Ifo Institute for Economic Research at the University of Munich, 2010.

Pisso, I., Real, E., Law, K. S., Legras, B., Bousserez, N., Attié, J., and Schlager, H.: Estimation of mixing in the troposphere from lagrangian trace gas reconstructions during long-range pollution plume transport, Journal of Geophysical Research: Atmospheres (1984–2012), 114, 2009.

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

Rößler, T. F. P.: Vergleich meteorologischer Re-Analysen im Hinblick auf die Simulation der Ausbreitung von Vulkanemissionen, Seminararbeit, 2014.

Schaefer, J. T.: The Critical Success Index as an Indicator of Warning Skill, Weather and Forecasting, 5, 570–575, 1990.

Seibert, P.: Convergence and Accuracy of Numerical Methods for Trajectory Calculations, Journal of Applied Meteorology, 32, 558–566, 1993.

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

Stohl, A., Wotawa, G., Seibert, P., and Kromp-Kolb, H.: Interpolation Errors in Wind Field as a Function of Spatial and Temporal Resolution and Their Impact on Different Types of kinematic Trajectories, American Meteorological Society, 34, 2149–2165, 1995.

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.

Thomson, D.: Criteria for the selection of stochastic models of particle trajectories in turbulent flows, Journal of Fluid Mechanics, 180, 529–556, 1987.

Venkatram, A. and Du, S.: {TURBULENT} {DIFFUSION}, in: Encyclopedia of Atmospheric Sciences, edited by Holton, J. R., pp. 2455 – 2466, Academic Press, Oxford, doi:http://dx.doi.org/10.1016/B0-12-227090-8/00441-3, URL `http://www.sciencedirect.com/science/article/pii/B0122270908004413`, 2003.

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.

Williamson, D. L. and Rasch, P. J.: Two-Dimensional Semi-Lagrangian Transport with Shape-Preserving Interpolation, Monthly Weather Review, 117, 102–129, 1989.

Williamson, D. L., Drake, J. B., Hack, J. J., Jakob, R., and Swarztrauber, P. N.: A standard test set for numerical approximations to the shallow water equations in spherical geometry, Journal of Computational Physics, 102, 211 – 224, 1992.