Meteorol. Appl. 8, 127–135 (2001)

An intercomparison of results from three trajectory models

A Stohl[1], L Haimberger[2], M P Scheele[3] and H Wernli[4]
[1] *Lehrstuhl für Bioklimatologie und Immissionsforschung, Technische Universität München, Am Hochanger 13, D-85354 Freising-Weihenstephan, Germany*
[2] *Institut für Meteorologie und Geophysik, Universität Wien, Hohe Warte 38, A-1190 Wien, Austria*
[3] *Royal Netherlands Meteorological Institute, P.O. Box 201, 3730 AE De Bilt, The Netherlands*
[4] *Institute for Atmospheric Science, ETH Zürich, CH-8093 Zürich, Switzerland*

Three three-dimensional trajectory models (LAGRANTO, TRAJKS and FLEXTRA), all driven with analysis wind fields from the European Centre for Medium-Range Weather Forecasts, are intercompared. The comparison has three parts: first, a case study of strong ascent in a warm conveyor belt is performed; second, a large set of back trajectories from the tropopause region over Europe and the mid-latitude Atlantic Ocean is investigated; third, a set of low-level trajectories is compared. The intercomparison shows that all three models have been implemented correctly. The degree of model accordance depends on the interpolation methods used. Deviations between the results from a single model using different interpolation schemes are of the same magnitude as the deviations of different models. If all models use linear spatial interpolation, their respective trajectories closely agree with each other, with deviations of 2% or less for the average distance between the starting and the ending positions in the free atmosphere after 48 h. Close to the surface, where the differences in the model formulations are largest, average horizontal position deviations may be up to 10%. Compared with other sources of errors, such as inaccuracies in the wind fields or insufficient temporal and spatial resolution of the data set, these differences are much smaller. Non-linear spatial interpolation leads to stronger vertical motions than linear interpolation and, in the case study, enhanced the quality of the results.

1. Introduction

Trajectory models are popular tools for describing air mass motions and are applied in several fields of the atmospheric sciences, but their accuracy is limited. Errors in the trajectory calculation result from numerical truncation (e.g. Seibert, 1993), interpolation (Kuo *et al.*, 1985; Rolph & Draxler, 1990; Stohl *et al.*, 1995), treatment of the vertical velocity (for instance, use of isobaric or isentropic approximation) (Stohl & Seibert, 1998), errors in the underlying wind fields (Kahl *et al.*, 1989; Pickering *et al.*, 1994, 1996) and sometimes inaccurate specification of the starting positions and times and subsequent growth of error (Merrill *et al.*, 1985). For more references on error sources for trajectory calculations see Stohl (1998). Although isentropic trajectories are still widely used, with adequate wind field data being available, three-dimensional trajectories are clearly more accurate in the troposphere (Stohl & Seibert, 1998) and are studied in this paper.

Estimates of the errors through comparisons of calculated trajectories with the ‘ground truth’ are inaccurate because of the difficulty of defining or observing the ground truth. Accuracy studies have been carried out with balloon flights and meteorological (e.g. potential vorticity) and physical (e.g. perfluorocarbons, smoke) tracers, each with their own set of problems in defining the ground truth. Furthermore, the sensitivity of trajectory models to the input data and to its spatial and temporal resolution was tested. Errors found in a large number of such studies with different trajectory models and under a variety of meteorological conditions vary considerably. Errors of up to 100% have been reported (see Stohl, 1998, for references), but average errors of approximately 20% of the travel distance may be considered as more typical, at least for the northern hemisphere. The largest single source of errors in trajectories is certainly inaccuracies in the wind fields used for the trajectory calculation. For instance, Pickering *et al.* (1994, 1996) compared trajectories calculated from analyses of the European Centre for Medium-Range Weather Forecasts (ECMWF) and from analyses from the National Meteorological Center (NMC, now

NCEP). In data-void regions of the southern hemisphere, they found average deviations between the two sets of up to 60% of the distance travelled after 120 hours. Clearly such large errors dominate over errors caused by the trajectory models themselves. However, over areas with denser observations, the agreement between different analysis datasets is much better, and other error sources, resulting from the trajectory models themselves, become important.

In this study, for the first time, we present an intercomparison of three frequently used trajectory model tools, which are all driven with wind fields from the same meteorological model. The aim is to quantify and explain the discrepancies caused by the different preparation of the input data and by the trajectory models themselves under conditions typical for the applications of the models. Another aim is to validate the three models.

2. Description of the models

The models used in this study are the TRAJKS model of the Royal Netherlands Meteorological Institute (Scheele *et al.*, 1996), the LAGRANTO model of the ETH Zurich (Wernli & Davies, 1997), and the FLEXTRA model of the University of Agricultural Sciences in Vienna and the Technical University of Munich (Stohl *et al.*, 1995). All three models have been mainly used with wind fields from the ECMWF for studying synoptic processes and air chemistry data, for planning of measurement campaigns and for operational purposes. LAGRANTO is currently used at six research institutions worldwide, TRAJKS at one and FLEXTRA at 17, and their output has been used in dozens of reviewed publications. The main features of the models are summarised in Table 1. Here, we only discuss the model properties relevant to this study; for details, the reader is referred to the above publications.

The comparisons shown below were performed for periods in 1997. At that time, the ECMWF model had a horizontal resolution of T213 (i.e. a maximum of 213 resolvable wavelengths around the globe) and 31 levels in the vertical. Analysed wind fields on a 1° latitude/longitude grid at six-hour intervals on the original 31 model levels were used to drive the trajectory models. We did not use a common dataset from ECMWF for this study, but each group had direct access to the ECMWF archives and used their own data retrieval procedures. Since all retrieval procedures are based on the ECMWF archival system, the retrieved data sets are – with one exception, see below – identical.

All three models use a latitude/longitude coordinate system in the horizontal, but employ different vertical coordinate systems. TRAJKS and LAGRANTO use pressure as vertical coordinates, whereas FLEXTRA uses the ECMWF hybrid η coordinates. But all three models take the ECMWF data on the original η model levels as input. The η coordinate system was described by Simmons & Burridge (1981) and Ritchie *et al.* (1995). η coordinates are defined such that the levels closest to the ground follow the topography, while the highest levels coincide with pressure surfaces. Intermediate levels have a gradual transition between these two extremes. With reasonably short time steps, FLEXTRA trajectories do not reach the ground, whereas LAGRANTO and TRAJKS trajectories may hit the surface because the pressure levels do not follow the topography.

Vertical velocity as stored in the ECMWF archives is given in pressure coordinates (i.e. in Pa s^{-1}), but is available also on the η model levels. LAGRANTO and TRAJKS can employ this vertical velocity directly. FLEXTRA needs the vertical wind in η coordinates, which is not available operationally. A conversion of the archived vertical velocities to η coordinates is inaccurate and produces unbalanced velocity fields. Therefore, FLEXTRA uses a pre-processor that determines the vertical velocity by vertically integrating the horizontal divergences (see Ritchie *et al.*, 1995), as available from ECMWF, and employing spherical harmonics data of the surface pressure for greater accuracy, which is important especially over steep orography.

Table 1. *Summary of the main features of the three models and their reference set-up.*

Model	TRAJKS	LAGRANTO	FLEXTRA
Horizontal coordinate system		Latitude/Longitude	
Vertical coordinate system	Pressure	Pressure	η
Spatial Resolution		1°/31 levels retrieved from T213L31 spectral data	
Time Resolution		6 h	
Integration method	Iterative, Petterssen (1940)	Petterssen (1940) with 3 iterations	Iterative, Petterssen (1940)
Time step	10 minutes	30 minutes	Flexible, < 1.2 h
Horizontal interpolation	Bilinear	Bilinear	Bicubic
Vertical interpolation	Linear with log pressure	Linear with pressure	Quadratic with η
Time interpolation	Quadratic	Linear	Linear
Vertical velocity	As retrieved from ECMWF archive	As retrieved from ECMWF archive	Determined from spherical harmonics data

All models have been found previously to yield only very small numerical truncation errors. This was confirmed with the TRAJKS model using time steps of 10 and 30 minutes, which yielded average discrepancies between the respective trajectories of less than 0.1% of the travel distance. Thus, discrepancies between the models must be due to differences in the interpolation procedures, in the vertical coordinate system (which in turn also results in different interpolation) and in the way the vertical velocity is determined.

3. Intercomparison method

The three models are compared with three test sets from the mid-latitudes which cover the typical applications of trajectory models. The sets are as follows.

(a) A case study.
(b) A large set of trajectories starting in the upper troposphere and lower stratosphere.
(c) A set of trajectories starting in the atmospheric boundary layer.

Both the second and third test set cover a variety of meteorological situations.

For the case study, a situation was selected that may be seen as a difficult one for trajectory models, namely strong ascent within a warm conveyor belt ahead of a cold front. Aircraft measurements over the Atlantic Ocean showed that boundary layer air was uplifted to at least 238 hPa at 0100 UTC on 28 May. The measurements and the synoptic situation during this event have been discussed extensively by Stohl & Trickl (1999). Since differences in model formulation and input data are largest for the vertical wind, relatively large discrepancies between the models may be expected for this case. Some 672 (only 670 being actually available for all models) 48-hour forward trajectories were started at 950 hPa and 850 hPa on a 1°×1° grid within the region 80° W to 60° W and 25° N to 40° N at 1800 UTC on 26 May 1997, covering the entrance region of the warm conveyor belt.

The second test covered the period from 13 August to 24 November 1997. During this period, measurements of ozone and nitrogen oxides were made aboard a commercial airliner (Jeker *et al.*, 2000) that cruised between North America and Europe. Some 18,800 120-hour backward trajectories were calculated that started along the aircraft's flight paths at two-minute intervals to support the interpretation of the measurement data. The aircraft flew mostly in the upper troposphere and lower stratosphere.

The third test set consisted of 125 48-hour forward trajectories started from a point close to the surface (950 hPa) at the position 4°E and 52°N every six hours during July 1997. This low starting position is likely to enhance the differences between the FLEXTRA model on the one hand and both LAGRANTO and TRAJKS on the other, because the differences between the η coordinate system and the pressure coordinate system are largest close to the ground.

In the following, we compare each model with each other, using frequency distributions of the following two measures, which have been used widely in the literature (e.g. Kuo *et al.*, 1985):

(a) Absolute horizontal transport deviation (AHTD): spherical distance (in kilometres) between the calculated ending points of two trajectories starting at the same position.
(b) Absolute vertical transport deviation (AVTD): pressure difference (in hPa) between the calculated ending points of two trajectories starting at the same position.

Since we found systematic differences in the intensities of the vertical motions, we also report the pressure ranges of the trajectories, defined as the difference between the maximum and the minimum pressure along a trajectory.

We also present the results for two additional FLEXTRA model runs. For the first, since the other models use linear interpolation in space, we repeated the FLEXTRA calculations (which are normally done with non-linear interpolation methods) using linear interpolation. A second run was carried out with wind field intervals of three hours, using three-hourly first guesses in addition to the analyses. Furthermore, for the first test set we also present the results of an additional TRAJKS calculation with linear time interpolation (normally quadratic time interpolation is used).

4. Results

4.1. Case study

We first discuss the results of the case study. To visualize the strong ascent within the warm conveyor belt, we applied the coherent ensemble of trajectories (CET) method of Wernli & Davies (1997). For this we plotted, for each model, those trajectories that ascended by more than 600 hPa (Figure 1). It is clear that all three trajectory models suggest the same path of the air, connecting the east coast of North America with Northern Europe, while ascending strongly. Even details of the transport seem to be predicted similarly by the three models. The largest differences are that the LAGRANTO CET has less members (42) than the TRAJKS CET (48), which in turn has fewer members than the FLEXTRA CET (64), suggesting that the FLEXTRA trajectories ascend more strongly than the others.

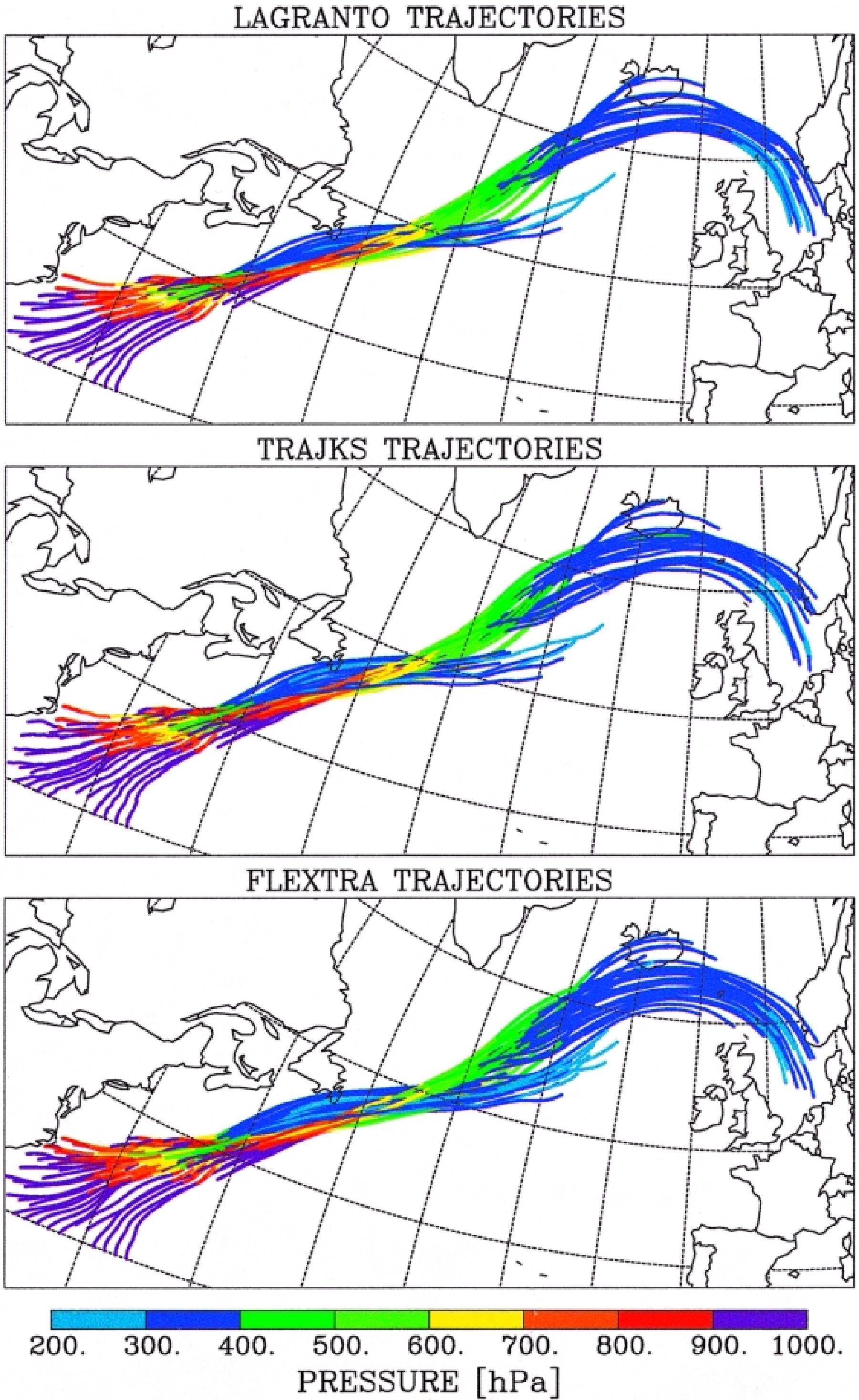

Figure 1. *Trajectories that ascended by more than 600 hPa within the warm conveyor belt using LAGRANTO (top), TRAJKS (middle) and FLEXTRA (bottom).*

Table 2 shows the frequency distributions of the AHTD as obtained from the intercomparisons for all the trajectories (i.e. not only the CET trajectories shown in Figure 1). The overall average distance on the sphere between the starting and the ending points is 2282 km. Compared to this, the average AHTD are much smaller: 61, 135 and 162 km (i.e. 3%, 6% and 7% of the overall travel distance, respectively) for the intercomparison of the three reference models. These values are clearly smaller than errors found typically in tracer studies (≈20%). Given that we have selected a 'difficult' case, this result is encouraging. Table 2 also shows, however, that the AHTD frequency distributions are strongly skewed: most of the trajectories agree reasonably well, while a few outliers show very large AHTD. The maximum AHTD are almost twice the average travel distance! The average vertical distance travelled is 272 hPa (Table 3), the average AVTD between the models are 9.0, 20.8 and 24.6 hPa (i.e. 3%, 8% and 9% of the overall vertical travel distance). The maxi-

Table 2. *Average, minimum and maximum values and percentiles (P10, P25, P50, P75, P90) of the AHTDs (in km) for all model comparisons of the three experiments. The average horizontal distances travelled by the trajectories are indicated for every experiment.*

Comparison	Average	Minimum	P10	P25	P50	P75	P90	Maximum
Experiment 1, 48 hours			Average distance travelled: 2282 km					
LAGRANTO vs. FLEXTRA	162	0	21	40	85	179	361	4175
TRAJKS vs. FLEXTRA	135	0	17	34	80	147	264	4180
LAGRANTO vs. TRAJKS	61	0	5	10	25	59	126	4045
LAGRANTO vs. FLEXTRA (linear)	47	0	7	15	27	50	87	4040
TRAJKS vs. FLEXTRA (linear)	68	0	9	19	39	71	131	4044
FLEXTRA vs. FLEXTRA (linear)	137	0	15	30	66	147	278	4175
TRAJKS (linear) vs. FLEXTRA (linear)	39	0	7	13	25	45	80	1100
TRAJKS (linear) vs. LAGRANTO	15	0	2	3	5	9	18	4040
FLEXTRA (6 h) vs. FLEXTRA (3 h)	291	8	26	54	124	289	661	4671
Experiment 2, 48 hours			Average distance travelled: 3467 km					
LAGRANTO vs. FLEXTRA	138	0	18	34	71	143	276	9992
TRAJKS vs. FLEXTRA	115	0	14	27	55	118	244	4320
LAGRANTO vs. TRAJKS	129	0	16	32	65	131	258	9969
LAGRANTO vs. FLEXTRA (linear)	83	0	8	18	40	79	154	9977
TRAJKS vs. FLEXTRA (linear)	101	0	13	24	49	104	212	4871
FLEXTRA vs. FLEXTRA (linear)	84	0	11	21	42	87	166	4852
FLEXTRA (6 h) vs. FLEXTRA (3 h)	346	0	46	90	185	395	819	5716
Experiment 2, 120 hours			Average distance travelled: 6730 km					
LAGRANTO vs. FLEXTRA	685	0	48	108	274	766	1860	9937
TRAJKS vs. FLEXTRA	673	0	44	99	260	713	1842	9446
LAGRANTO vs. TRAJKS	674	0	42	95	253	736	1856	9912
LAGRANTO vs. FLEXTRA (linear)	447	0	26	58	146	424	1189	9332
TRAJKS vs. FLEXTRA (linear)	617	0	38	85	223	648	1671	9949
FLEXTRA vs. FLEXTRA (linear)	503	0	32	72	181	514	1309	9992
FLEXTRA (6 h) vs. FLEXTRA (3 h)	1447	0	144	341	815	1966	3689	9977
Experiment 3, 48 hours			Average distance travelled: 908 km					
LAGRANTO vs. FLEXTRA	45	0	10	18	33	59	89	244
TRAJKS vs. FLEXTRA	46	3	10	20	34	57	100	275
LAGRANTO vs. TRAJKS	20	0	3	5	11	23	48	143
LAGRANTO vs. FLEXTRA (linear)	95	4	18	32	63	122	194	697
TRAJKS vs. FLEXTRA (linear)	89	2	17	33	59	123	183	604
FLEXTRA vs. FLEXTRA (linear)	86	3	15	31	60	105	171	627
FLEXTRA (6 h) vs. FLEXTRA (3 h)	95	5	17	33	63	110	187	847

mum AVTD are more than 500 hPa for all three comparisons.

The question arises as to why LAGRANTO and TRAJKS agree much better with each other (significant on the 99% level for both horizontal and vertical positions according to a Student's t-test) than with FLEXTRA (Tables 2 and 3). Closer inspection of the model results revealed that the FLEXTRA trajectories fluctuated more in altitude than the other ones. To summarise this, consider the pressure ranges, i.e. the maximum fluctuations in height along a single trajectory, produced by the different models. The average pressure ranges of LAGRANTO, TRAJKS and FLEXTRA are 285, 290 and 311 hPa, respectively (Figure 2). Since the general motion within the warm conveyor belt was upward, FLEXTRA produced stronger ascent than the other two models. The differences are even larger for the trajectories with the strongest ascent: 566, 575 and 607 hPa for the 90th percentiles of the pressure ranges.

Obviously, the reason for the disagreement between FLEXTRA and the other models in terms of both AHTD and AVTD is the much stronger vertical motion produced by FLEXTRA. To get an indication whether this is realistic or not, we determined the number of trajectories that reached 238 hPa, the flight level of the aircraft (see Stohl & Trickl, 1999). We found that only one LAGRANTO, two TRAJKS and five FLEXTRA trajectories (nine in the case of three-hourly wind field data) reached that level. According to satellite images and the data from the aircraft which flew approximately 1000 km through moist, ozone-rich air exported from the North American boundary layer, the warm conveyor belt covered an extensive

Table 3. *Average, minimum and maximum values and percentiles (P10, P25, P50, P75, P90) of the AVTDs (in hPa) for all model comparisons of the three experiments. The average vertical distances travelled by the trajectories are indicated for every experiment.*

Comparison	Average	Minimum	P10	P25	P50	P75	P90	Maximum
Experiment 1, 48 hours			Average distance travelled: 272 hPa					
LAGRANTO vs. FLEXTRA	24.6	0.0	2.9	8.5	17.6	31.6	48.6	563.4
TRAJKS vs. FLEXTRA	20.8	0.1	2.5	6.5	14.9	26.4	42.0	562.5
LAGRANTO vs. TRAJKS	9.0	0.0	0.5	1.4	5.0	11.0	18.6	520.2
LAGRANTO vs. FLEXTRA (linear)	10.4	0.0	1.2	3.0	7.1	13.7	20.7	522.0
TRAJKS vs. FLEXTRA (linear)	12.9	0.0	1.5	3.8	9.3	16.6	23.6	521.3
FLEXTRA vs. FLEXTRA (linear)	18.8	0.1	1.4	4.5	11.3	24.1	38.8	564.5
TRAJKS (linear) vs. FLEXTRA (linear)	9.7	0.0	1.2	3.0	6.7	13.6	21.0	218.5
TRAJKS (linear) vs. LAGRANTO	1.7	0.0	0.1	0.2	0.5	1.0	2.0	521.3
FLEXTRA (6 h) vs. FLEXTRA (3 h)	40.4	0.1	4.1	11.4	27.5	48.8	81.2	589.6
Experiment 2, 48 hours			Average distance travelled: 71 hPa					
LAGRANTO vs. FLEXTRA	13.2	0.0	0.6	1.7	4.3	11.2	27.7	573.5
TRAJKS vs. FLEXTRA	10.9	0.0	0.6	1.6	3.9	10.1	24.9	587.4
LAGRANTO vs. TRAJKS	12.5	0.0	0.5	1.3	3.4	9.9	26.3	612.6
LAGRANTO vs. FLEXTRA (linear)	7.6	0.0	0.4	0.9	2.3	5.7	13.7	591.3
TRAJKS vs. FLEXTRA (linear)	10.1	0.0	0.5	1.3	3.3	8.6	23.3	663.8
FLEXTRA vs. FLEXTRA (linear)	8.2	0.0	0.4	1.0	2.7	7.1	18.7	536.5
FLEXTRA (6 h) vs. FLEXTRA (3 h)	35.1	0.0	2.4	6.3	14.7	33.3	80.9	713.8
Experiment 2, 120 hours			Average distance travelled: 115 hPa					
LAGRANTO vs. FLEXTRA	49.5	0.0	1.6	4.7	14.4	44.8	133.6	788.7
TRAJKS vs. FLEXTRA	47.7	0.0	1.7	4.7	13.7	43.3	126.0	806.4
LAGRANTO vs. TRAJKS	46.3	0.0	1.3	3.9	12.8	41.6	120.7	767.1
LAGRANTO vs. FLEXTRA (linear)	31.9	0.0	1.0	2.7	8.2	26.0	76.3	736.5
TRAJKS vs. FLEXTRA (linear)	43.8	0.0	1.4	4.0	12.0	38.1	113.3	776.4
FLEXTRA vs. FLEXTRA (linear)	38.8	0.0	1.1	3.1	9.7	33.7	100.3	811.6
FLEXTRA (6 h) vs. FLEXTRA (3 h)	89.4	0.0	5.0	13.2	33.5	94.0	266.1	839.6
Experiment 3, 48 hours			Average distance travelled: 67 hPa					
LAGRANTO vs. FLEXTRA	11.9	0.5	1.7	3.4	7.0	15.1	28.0	81.3
TRAJKS vs. FLEXTRA	12.2	0.5	1.7	3.8	8.5	14.8	25.7	88.8
LAGRANTO vs. TRAJKS	4.5	0.0	0.2	0.6	1.5	4.8	9.6	107.8
LAGRANTO vs. FLEXTRA (linear)	20.6	0.1	1.9	4.5	11.2	24.6	44.9	153.1
TRAJKS vs. FLEXTRA (linear)	19.4	0.2	2.2	4.5	10.7	25.7	41.0	143.5
FLEXTRA vs. FLEXTRA (linear)	20.3	0.1	1.1	4.4	10.3	26.8	50.2	150.9
FLEXTRA (6 h) vs. FLEXTRA (3 h)	20.8	0.1	1.2	5.1	9.4	27.1	51.4	167.7

region (see Stohl & Trickl, 1999). Obviously, even FLEXTRA underestimates the extent to which boundary-layer air was lifted to the upper troposphere, but it captures the event somewhat better than the other two models. Use of wind fields with a three-hour time resolution seems to improve the calculation considerably.

4.2. Impact of the interpolation method

Our original hypothesis was that the differences between FLEXTRA on the one hand and LAGRANTO and TRAJKS on the other is caused by the different vertical coordinte systems. However, FLEXTRA also employs non-linear spatial interpolation, while both LAGRANTO and TRAJKS employ linear interpolation. To explore this, we repeated the FLEXTRA model runs with linear interpolation in space. Unexpectedly, this resulted in a much better agreement with the other two models. The average AHTD between LAGRANTO and FLEXTRA was reduced by more than a factor of three compared to the non-linear interpolation run, between TRAJKS and FLEXTRA by a factor of two (see Table 2). Similar reductions were obtained for the AVTD (Table 3). The better agreement among the models is reflected in a large reduction of the average pressure range of FLEXTRA from 311 hPa to 289 hPa (compared to 285 and 290 hPa for the other two models, see Figure 2). Accordingly, only one FLEXTRA trajectory reached the altitude of the aircraft. The linear interpolation obviously underestimates, even more strongly than the bicubic interpolation, the strong ascent within the warm conveyor belt as expected from the measurements. It is also found that the average horizontal travel distance of the trajectories is longer for the non-linear

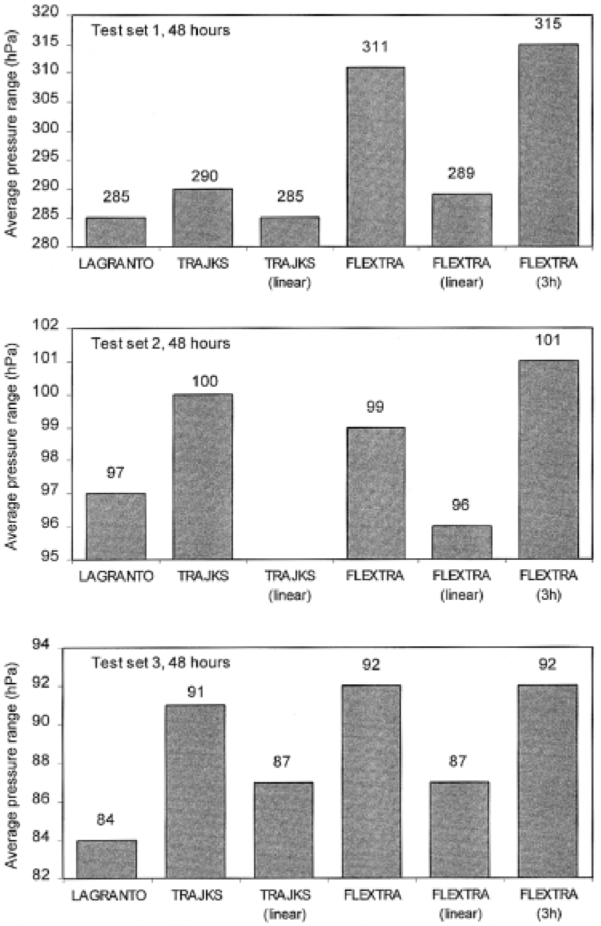

Figure 2. *Average pressure ranges (hPa) for the different trajectory models and interpolation methods after 48 h travel time for test set 1 (top), test set 2 (middle) and test set 3 (bottom).*

interpolation methods (not shown). Since TRAJKS normally uses quadratic time interpolation, we also repeated the TRAJKS computations, but this time with linear time interpolation. This resulted in even better agreement between the models (Tables 2 and 3).

Comparing only the linear versions of the trajectory models, the agreement between LAGRANTO and TRAJKS is again better than with FLEXTRA. Of the linear interpolation model versions, FLEXTRA produces the strongest ascent (289 hPa compared to 285 hPa for both LAGRANTO and TRAJKS, see Figure 2). It is likely that this is due to the different vertical coordinate systems and the different data retrieval algorithm used by FLEXTRA.

4.3. Results for the second test set

For the second test set, we calculated the summary statistics after 48 h and after 120 h travel time (Table 2). Because the trajectories started in the tropopause region, the average travel speed was higher and the travel distance longer than for the first test. Nevertheless, the average AHTD between LAGRANTO and FLEXTRA and between TRAJKS and FLEXTRA are smaller than for the first test: 138 km and 115 km (and are much smaller in relative terms: 4% and 3%, respectively) after 48 hours, whereas LAGRANTO and TRAJKS agree slightly less good than for the first test (129 km, 4%). Again, using linear interpolation with FLEXTRA reduces the disagreement with the other models still further (83 km and 101 km, respectively). After 120 hours, the AHTD are large for all comparisons (see Table 2): on average almost 700 km (10%), except for LAGRANTO and the linear FLEXTRA version.

The average vertical distance travelled after 48 hours (71 hPa) is, as expected, much smaller than for the first test (Table 3). Although the AVTD between the reference models are also smaller, they are rather high in relative terms (18%). Switching to linear interpolation reduces the AVTD between LAGRANTO and FLEXTRA by almost a factor of two, whereas there is much less reduction of the AVTD with TRAJKS. This is due to the quadratic time interpolation used in TRAJKS, which has no correspondence neither in FLEXTRA nor in LAGRANTO. The average pressure ranges are much smaller than for the first test set, but again, the range of FLEXTRA with non-linear interpolation is larger (99 hPa after 48 hours) than with linear interpolation (96 hPa) (Figure 2).

4.4. Results for the third test set

For the third test set, we report the summary statistics after 48 h travel time (Table 2). Because the trajectories started in the boundary layer, they travelled a shorter distance than for the first two test sets. Although AHTD are smaller in absolute terms, they are comparable to the previous results relative to the average travel distance: 5% for the comparison of FLEXTRA with the other two models and 2% between LAGRANTO and TRAJKS. Like for the first test set, LAGRANTO and TRAJKS agree much better with each other than with FLEXTRA. This is understandable, since both use pressure coordinates which, close to the surface, deviate significantly from the η coordinates. The average deviations between LAGRANTO and TRAJKS on the one hand and FLEXTRA on the other are reduced if LAGRANTO and TRAJKS trajectories hitting the surface are discarded (not shown). However, surprisingly, this time the linear FLEXTRA version agreed less with the other two models than the non-linear version. No explanation was found for this behaviour, but it is likely that close to the ground the deviations between the models are due to differences in the coordinate systems (and their interaction with interpolation) rather than by differences caused directly by the interpolation method. AVTD are comparable to the second test set, both in absolute and in relative terms (Table 3), except for the linear FLEXTRA version, which produced much larger deviations from all other models. As with the previous tests, the average pressure range (see Figure 2) of the F LEXTRA trajectories with non-linear interpolation (92 hPa) is larger than with linear interpolation (87 hPa). LAGRANTO again produced the weakest vertical motion (84 hPa).

5. Conclusions

We have compared the three trajectory models LAGRANTO, TRAJKS and FLEXTRA for three different situations which are typical for the application of trajectory models: low-level trajectories as often used to study air chemistry data, high-level trajectories as needed for interpretation of aircraft data, and trajectories in a strongly ascending airstream as required for the study of dynamical processes in the troposphere. The conclusions of this intercomparison are as follows.

(a) All three trajectory models have been implemented correctly and use the wind field information equally efficient. From an accuracy point of view the quality of the input data is more important than the choice among the three trajectory models.

(b) Deviations between the results from a single model using different interpolation schemes are the same magnitude as the deviations of different models.

(c) When linear spatial interpolation methods are used in all three models, average horizontal position deviations in the free atmosphere after 48 h travel time are 4% or less of the distance on the sphere between the starting and the ending point of the trajectories. Close to the surface, where differences in the vertical coordinate systems used are largest, average horizontal position deviations may be up to 10% of the travel distance. Vertical position deviations are some 5% for strong vertical motions, as in warm conveyor belts.

(d) The choice of the interpolation scheme is important. Non-linear space interpolation produces stronger vertical excursions of the trajectories, which, in the case study, was found to be in better agreement with observations. This result is in agreement with the interpolation experiments of Walmsley & Mailhot (1983) and Stohl *et al.* (1995), who have also argued for higher-order interpolation methods.

(e) Use of wind fields every three hours instead of the more usual six hours represents a significant improvement for trajectory calculations, producing stronger vertical motions.

(f) The FLEXTRA results using non-linear interpolation and wind fields with 6 hours time resolution for the tests 1 and 3 of this study are made available via the FLEXTRA internet site (http://www.fw.tum.de/LST/METEOR/stohl/). We encourage the use of this dataset for testing other models.

Acknowledgments

Andreas Stohl acknowledges that ECMWF and Deutscher Wetterdienst admitted access to ECMWF data within the special project 'Validation of trajectory calculations'. We thank D. Brunner and P. Siegmund from the Royal Netherlands Meteorological Institute for helpful discussions, and D. Jeker from ETH Zürich for supplying the LAGRANTO trajectories for the second test set. The comments of two anonymous reviewers and the editor, R. W. Riddaway, are gratefully acknowledged.

References

Jeker, D. P., Pfister, L., Thompson, A. M., Brunner, D., Boccippio, D., Pickering, K. E., Wernli, H., Kondo, Y. & Staehelin, J. Measurements of nitrogen oxides at the tropopause: Attribution to convection and correlation with lightning. *J. Geophys. Res.*, **105**: 3679–3700.

Kahl, J. D., Harris, J. M., Herbert, G. A. & Olson, M. P. (1989). Intercomparison of three long-range trajectory models applied to Arctic haze. *Tellus*, **41B**: 524–536.

Kuo, Y.-H., Skumanich, M., Haagenson, P. L. & Chang, J. S. (1985). The accuracy of trajectory models as revealed by the observing system simulation experiments. *Mon. Wea. Rev.*, **113**: 1852–1867.

Merrill, J. T., Bleck, R. & Avila, L. (1985) Modeling atmospheric transport to the Marshall Islands. *J. Geophys. Res.*, **90**: 12927–12936.

Petterssen, S. (1940). *Weather Analysis and Forecasting*. McGraw-Hill Book Company, New York, 221–223.

Pickering, K. E., Thompson, A. M., McNamara, D. P. & Schoeberl, M. R. (1994). An intercomparison of isentropic trajectories over the South Atlantic. *Mon. Wea. Rev.*, **122**: 864–879.

Pickering, K. E., Thompson, A. M., McNamara, D. P., Schoeberl, M. R., Fuelberg, H. E., Loring Jr., R. O., Watson, M. V., Fakhruzzaman, K. & Bachmeier, A. S. (1996). TRACE A trajectory intercomparison. 1. Effects of different input analyses. *J. Geophys. Res.*, **101**: 23903–23925.

Ritchie, H., Temperton, C., Simmons, A., Hortal, M., Davies, T., Dent, D. & Hamrud, M. (1995). Implementation of the semi-Lagrangian method in a high-resolution version of the ECMWF forecast model. *Mon. Wea. Rev.*, **123**: 489–514.

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

Scheele, M. P., Siegmund, P. C. & Velthoven, P. F. J. (1996). Sensitivity of trajectories to data resolution and its dependence on the starting point: in or outside a tropopause fold. *Meteorol. Appl.*, **3**: 267–273.

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

Simmons, A. J. & Burridge, D. M. (1981). An energy and momentum conserving finite difference scheme and hybrid vertical coordinates. *Mon. Wea. Rev.*, **109**: 758–766.

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

Stohl, A. & Seibert, P. (1998). Accuracy of trajectories as determined from the conservation of meteorological tracers. *Q. J. R. Meteorol. Soc.*, **125:** 1465–1484.

Stohl, A. & Trickl, T. (1999). A textbook example of long-range transport: Simultaneous observation of ozone maxima of stratospheric and North American origin in the free troposphere over Europe. *J. Geophys. Res.*, **104:** 30445–30462.a

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

Walmsley, J. L. & Mailhot, J. (1983). On the numerical accuracy of trajectory models for long-range transport of atmospheric pollutants. *Atmos. Ocean*, **21:** 14–39.

Wernli, H. & Davies, H. C. (1997). A Lagrangian-based analysis of extratropical cyclones. I: The method and some applications. *Q. J. R. Meteorol. Soc.*, **123:** 467–489.