Pergamon

Atmospheric Environment Vol. 32, No. 24, pp. 4245–4264, 1998

Printed in Great Britain

PII: S1352–2310(98)00184–8

1352–2310/98 $19.00 + 0.00

VALIDATION OF THE LAGRANGIAN PARTICLE DISPERSION MODEL FLEXPART AGAINST LARGE-SCALE TRACER EXPERIMENT DATA

A. STOHL*,†, M. HITTENBERGER† and G. WOTAWA†

* Lehrstuhl für Bioklimatologie und Immissionsforschung, Ludwig-Maximilians-Universität München, Am Hochanger 13, 85354 Freising-Weihenstephan, Germany; and † Institute of Meteorology and Physics, University of Agricultural Sciences, Türkenschanzstr. 18, 1180 Vienna, Austria

Abstract—A comprehensive validation of FLEXPART, a recently developed Lagrangian particle dispersion model based on meteorological data from the European Centre for Medium-Range Weather Forecasts, is described in this paper. Measurement data from three large-scale tracer experiments, the Cross-Appalachian Tracer Experiment (CAPTEX), the Across North America Tracer Experiment (ANATEX) and the European Tracer Experiment (ETEX) are used for this purpose. The evaluation is based entirely on comparisons of model results and measurements paired in space and time. It is found that some of the statistical parameters often used for model validation are extremely sensitive to small measurement errors and should not be used in future studies. 40 cases of tracer dispersion are studied, allowing a validation of the model performance under a variety of different meteorological conditions. The model usually performs very well under undisturbed meteorological conditions, but it is less skilful in the presence of fronts. The two ETEX cases reveal the full range of the model's skill, with the first one being among the best cases studied, and the second one being, by far, the worst. The model performance in terms of the statistical parameters used stays rather constant with time over the periods (up to 117 h) studied here. It is shown that the method used to estimate the concentrations at the receptor locations has a significant effect on the evaluation results. The vertical wind component sometimes has a large influence on the model results, but on the average only a slight improvement over simulations which neglect the vertical wind can be demonstrated. Subgrid variability of mixing heights is important and must be accounted for. © 1998 Elsevier Science Ltd. All rights reserved

Key word index: CAPTEX, ANATEX, ETEX, tracer, Lagrangian particle dispersion model, trajectories.

1. INTRODUCTION

Although continental-scale tracer experiments provide unique opportunities to test long-range dispersion models, few such experiments have been conducted yet. Among the most interesting are CAPTEX (Ferber *et al*., 1986), ANATEX (Draxler *et al*., 1991) and ETEX (Archer *et al*., 1996; Nodop *et al*., 1998). These experiments are very different from each other, regarding scale, experimental design and meteorological conditions. During CAPTEX, the experiment on the smallest scale, seven tracer releases from two different locations were made. Samples were taken at distances of up to 1000 km from the release location, and until approximately 48 h after the release. During ANATEX, a total of 66 tracer releases was made from two locations. Samples were collected at distances of up to 3000 km, but both the spatial and the temporal resolution of the sampling network was rather coarse and it was often not possible to separate individual tracer plumes, since releases were made every 2.5 d. The most recent experiment was ETEX. Only two releases were made, but the tracer plumes were recorded across Europe for 90 h by a dense sampling network and at a high temporal resolution.

In this study, the experimental data are used to evaluate a recently developed dispersion model, to investigate the suitability of some statistical parameters found in the literature for model validation, to examine the influence of the concentration calculation procedure (grid cell estimates vs kernel methods) on the model results, to study the impact of the vertical wind velocity on the model results, and to test a parameterization of the subgrid variability of boundary layer heights. Identical statistical and graphical methods for all three data sets are applied, which allows a direct comparison of the model skill.

2. MODEL DESCRIPTION

FLEXPART, version 2.0*, is a recently developed Lagrangian particle dispersion model used as part of

‡ Corresponding author.

* The FLEXPART model source code and a description are available via the Internet from http://www.forst.uni-muenchen.de/LST/METEOR/Stohl/astohl.html.

Austria's emergency response system and for research (Stohl, 1997a). Results of applications to the first ETEX case were published by Stohl (1997b) and Stohl and Wotawa (1997) but this paper presents the first comprehensive evaluation of the model.

FLEXPART evolved from the trajectory model FLEXTRA that was described and validated by Stohl *et al.* (1995), Baumann and Stohl (1997), Stohl *et al.* (1997) and Stohl and Seibert (1998). Results of an evaluation of FLEXTRA using the constant level balloon flights performed during ETEX were reported by Stohl and Koffi (1998). FLEXPART is based on model level data of the T213 L31 numerical weather prediction model of the ECMWF (ECMWF, 1995). For the ETEX period, data were extracted from the ECMWF archives with 1° and 3 h resolution (analyses at 0, 6, 12, 18 UTC; 3 h forecasts at 3, 9, 15, 21 UTC). For the CAPTEX and ANATEX periods, ECMWF re-analysis data (Gibson *et al.*, 1996) were used. These data were assimilated with nearly the same model version that is now operational at ECMWF, but with lower horizontal resolution (T106 L31). Since forecasts were not available for the re-analysis dataset, only analyses with 1° and 6 h resolution were used.

The 31 ECMWF model levels were transformed to terrain-following Cartesian coordinates. An additional surface level was defined by the 10 m winds and the 2 m temperatures of the ECMWF model. For the CAPTEX and ANATEX cases, the profile method after Berkowicz and Prahm (1982) was applied to calculate friction velocity u_*, surface level heat flux $(\overline{w'\Theta'_v})_0$ and Obukhov length using data at 2 m, 10 m and at the first model level. For the ETEX cases, for which forecasts were available, accumulated flux data of the 3 and 6 h forecast fields were used instead, since boundary layer parameters derived from these data are more accurate than those calculated with the profile method (Wotawa and Stohl, 1997). Planetary boundary layer (PBL) heights were set to the height z where the Richardson number

$$\mathrm{Ri} = \frac{(g/\Theta_{v1})\ (\Theta_v - \Theta_{v1})\ (z - z_1)}{(u - u_1)^2 + (v - v_1)^2 + 100u_*^2} \qquad (1)$$

exceeded the value of 0.25 (Vogelezang and Holtslag, 1996). g is the acceleration due to gravity, Θ_{v1} and Θ_v are the virtual potential temperatures, z_1 is the height of the first model level, and (u_1, v_1), and (u, v) are the horizontal wind speed components at the first model level and at height z, respectively. Equation (1) holds only for stable and neutral conditions, but it can be extended (Vogelezang and Holtslag, 1996) to convective situations by replacing Θ_{v1} with

$$\Theta'_{v1} = \Theta_{v1} + 8.5\,\frac{(\overline{w'\Theta'_v})_0}{w_*} \qquad (2)$$

where w_* is the convective velocity scale. The second term on the right-hand side of equation (2) represents the excess temperature of rising thermals.

Spatial and temporal variations of PBL heights on scales not resolved by the ECMWF model play an important role in determining the thickness of the layer over which the tracer material is distributed. To elucidate this, we first consider unresolved temporal variations. Given a typical evolution of a convective mixed layer, the PBL height reaches its maximum value (say 1500 m) in the afternoon (for instance at 1700 LST), before a much shallower stable PBL forms. Now, if the meteorological data are available only at 1200 LST and at 1800 LST and the PBL heights at those times are, say, 1200 and 200 m, and some standard interpolation procedure is used to determine the PBL height at 1700 LST, the PBL height is significantly underestimated. In the above example, using linear interpolation, a value of 370 m would have been obtained instead of 1500 m. If a tracer release takes place shortly before the breakdown of the convective mixed layer, this leads to a serious overestimation of the surface concentrations (a factor of four in the above example). Even if the release occurs already in the morning hours, the thickness of the tracer cloud in the evening would be restricted to 1200 m in the above example, whereas the correct thickness would be 1500 m.

Similar arguments are valid for spatial variations of PBL heights. A major reason for spatial variability of PBL heights is complex topography. It increases mechanical mixing due to enhanced shear stress, induces gravity waves during stable conditions, and it provides elevated sources of sensible heat during daytime. All these complex effects can produce turbulence and can thereby locally increase the PBL heights above the elevated areas. In addition to topographical effects, significant variations of the PBL heights can also be caused by differences in landuse or soil wetness (Hubbe *et al.*, 1997). If a tracer cloud travels over such a patchy surface, its thickness would be determined by the maximum PBL height experienced along its path rather than by the average PBL height, which is obtained at the grid points of a coarse resolution meteorological model.

The best solution to this problem obviously is to use meteorological data with higher space and time resolution than the data from ECMWF. Unfortunately, such data are not as readily available, and if they are, it is not *a priori* clear whether the quality of these data (especially of the wind fields) is as high as the quality of the ECMWF data. Another solution would be to use a simple prognostic boundary layer model as an interpolation tool.

However, in this paper, we did not follow one of these solutions. Instead, we used a somewhat arbitrary parameterization to avoid a significant bias in the tracer cloud thickness which results in strong overestimations of the surface tracer concentrations. To account for spatial variations induced by topography, we added one standard deviation of the subgrid topography (available from ECMWF) to calculate an "envelope" PBL height. To account for temporal and

other spatial PBL height variations, we took the maximum PBL height of the eight grid points surrounding a particle's position in space and time, instead of interpolating the PBL heights. The effect of this is studied in Section 3.5.

In order to simulate the transport of the tracer material, FLEXPART calculates the trajectories of a large number of particles, each representing a parcel of tracer material. The trajectories are calculated according to

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(\mathbf{x}, t)\Delta t \tag{3}$$

with t being time, Δt the time increment, $\mathbf{x}$ the position vector, and $\mathbf{v} = \bar{\mathbf{v}} + \mathbf{v}'_{\mathrm{t}} + \mathbf{v}'_{\mathrm{m}}$ the wind vector that is composed of the grid scale wind $\bar{\mathbf{v}}$, the turbulent wind fluctuations $\mathbf{v}'_{\mathrm{t}}$ (Zannetti, 1992) and the mesoscale wind fluctuations $\mathbf{v}'_{\mathrm{m}}$.

The turbulent fluctuations are calculated according to the Langevin equation (Thomson, 1987)

$$\mathrm{d}v_{\mathrm{t}i} = a_i(\mathbf{x}, \mathbf{v}_{\mathrm{t}}, t)\,\mathrm{d}t + b_{ij}(\mathbf{x}, \mathbf{v}_{\mathrm{t}}, t)\,\mathrm{d}W_j \tag{4}$$

where a and b are functions of the position, the turbulent velocity and time. The $\mathrm{d}W_j$ are incremental components of a Wiener process with mean zero and variance $\mathrm{d}t$, which are uncorrelated with the other components and are uncorrelated in time. This form of the Langevin equation was described by Legg and Raupach (1982). Alternatively, the Langevin equation can be re-expressed in terms of $v_{\mathrm{t}i}/\sigma_{v_{\mathrm{t}i}}$, where $\sigma_{v_{\mathrm{t}i}}$ is the standard deviation of the turbulent wind, instead of $v_{\mathrm{t}i}$ (Wilson *et al.*, 1983). This form was shown by Thomson (1987) to fulfill the well-mixed criterion which states that "if a species of passive marked particles is initially mixed uniformly in position and velocity space in a turbulent flow, it will stay that way" (Rodean, 1996). Although the method proposed by Legg and Raupach (1982) violates this criterion in strongly inhomogeneous turbulence, their formulation was found to be practical, as numerical experiments have shown that it is robust against an increase in the integration time step used by the model. Therefore, the Legg and Raupach (1982) method is used when long time steps are allowed to save computation time (see later in this section); otherwise, the Wilson *et al.* (1983) method is used. The wind velocity standard deviations and the Lagrangian time scales were obtained by the parameterization of Hanna (1982). Boundary conditions at the surface and at the top of the PBL are specified following Wilson and Flesch (1993).

Mesoscale motions are neither resolved by the ECMWF data nor are they covered by the turbulence parameterization which is based on measurements that are representative for a temporal scale of less than an hour and correspondingly short length scales. This spectral gap must be filled, since mesoscale motions can significantly accelerate the growth of a dispersing plume (Gupta *et al.*, 1997). For this, we use a method that is similar to the one recently described by Maryon (1998), namely to solve an independent Langevin equation for the mesoscale wind velocity fluctuations ("meandering" in Maryon's terms), assuming that they are largely independent from the turbulent fluctuations covered by Hanna's (1982) parameterization. In contrast to Maryon (1998), who used time scales and velocity variances derived from a spectral analysis of a time series of wind measurements at a single station, we assume that the variance of the wind observed at the grid scale provides some information on its subgrid variance. The wind velocity standard deviation used for the mesoscale Langevin equation is set to half the standard deviation of the 16 grid points surrounding the particles' position in space and time. The corresponding time scale is taken as half the interval at which wind fields are available, assuming that the linear interpolation between the grid points can recover half the subgrid variability. According to Stohl *et al.* (1995), who compared wind field data at different resolution, this is not an unlikely assumption. This empirical approach does not describe actual mesoscale phenomena, but it is similar to the ensemble methods which are useful to assess trajectory accuracy (Kahl, 1996; Baumann and Stohl, 1997; Stohl, 1998). It provides an estimate of how the grid scale flow reacts to perturbations on smaller scales. It is not so important in the PBL, where the interaction of the vertical shear of the mean horizontal wind and the turbulent vertical motions dominates plume dispersion, but it is significant for long-range transport in the free troposphere.

FLEXPART can be used in two different modes: while in the numerically more accurate one the integration time steps are limited by the Lagrangian time scale, longer time steps are allowed in the computationally faster one. In the first mode, $\Delta t = T_{\mathrm{L}}/C_{\mathrm{L}_T}$, where $C_{\mathrm{L}_T} > 1$ must be prescribed, and $T_{\mathrm{L}} = \min(T_{\mathrm{L}_u}, T_{\mathrm{L}_v}, T_{\mathrm{L}_w})$, where T_{L_u}, T_{L_v} and T_{L_w} are the Lagrangian time scales for the three wind components. In the second mode, the time steps are only limited by the Courant–Friedrichs–Lewy criterion and another criterion that relates the maximum time step to the frequency of the input fields: $\Delta t = \Delta T/C_t$, where ΔT is the interval between wind fields and $C_t > 1$. This mode was used here because experiments have revealed that the results are not strongly affected by this simplification, although surface concentrations are somewhat underestimated (compare Stohl and Thomson, 1998). This is in agreement with Uliasz (1994) who found that even a complete neglect of the autocorrelations is sufficiently accurate in regional-scale applications. C_t was set to allow maximum time steps of 15 min.

The number of particles necessary to yield smooth concentration fields partly depends on the time having passed since their release. Shortly after the release, when the particles form a compact cloud, relatively few particles suffice to describe the dispersion. Later,

as the cloud grows, more particles are needed to obtain smooth concentration fields. To save computation time but maintain accuracy, particles are split into two (each receiving half of the mass of the original particle) after travel times of Δt_s, $2\Delta t_s$, $4\Delta t_s$, $8\Delta t_s$, and so on. Δt_s was set to 1 d for the ETEX cases, starting with 100,000 particles. No particle splitting was used for the CAPTEX (100,000 particles) and ANATEX (50,000 particles) cases.

Concentrations were calculated on a three-dimensional grid as well as at the measurement locations using kernel methods (Lorimer, 1986). An explanation of this is given in Section 3.4, where the sensitivity of the model results to the method used for calculating the concentrations is studied. Time-averaged concentrations C_T at time T are calculated by frequent (every 15 min) sampling during the period $[T - \Delta T, T]$. ΔT was set to 6, 24 and 3 h for the CAPTEX, ANATEX and ETEX data, respectively, according to the intervals at which measurement data were available.

The FLEXPART model also provides procedures to calculate radioactive decay, wet deposition, dry deposition, and gravitational settling of particulate matter (Stohl, 1997a), but they were not used in this study.

3. RESULTS

3.1. *Suitability of some statistical parameters for model evaluation*

Model evaluation was done both qualitatively by comparing predicted and observed concentration maps and quantitatively based on statistical parameters. The statistical evaluation was entirely based on measured and modeled values paired in space and time, the strictest analysis method. It was not clear from the beginning which statistical parameters are suitable for that purpose. We considered the following ones, partly based on the ATMES-II study (Mosca *et al.*, 1997):

- Fractional bias $\mathrm{FB} = 2B/(\bar{P} + \bar{M})$, where $B = (1/N)\sum_{i=1}^{N}(P_i - M_i)$ is the bias, N is the number of paired data points, P_i and M_i are the model predictions and the measurements, and $\bar{P}$ and $\bar{M}$ are the average predictions and measurements, respectively.
- Normalized Mean Square Error $\mathrm{NMSE} = (1/N)\sum_{i=1}^{N}(P_i - M_i)^2/\bar{P}\bar{M}$.
- Pearson correlation coefficient r.
- Spearman rank-order correlation coefficient r_s (Press *et al.*, 1990)
- Figure of Merit in Space (equivalent to the threat scores used by some authors) $\mathrm{FMS} = 100\,A_p \cap A_m / A_p \cup A_m$, where A_p and A_m are the predicted and measured subsets of concentrations above a significant level ($0.1\ \mathrm{ng\,m^{-3}}$ for the ETEX cases, and $2\ \mathrm{fl\,\ell^{-1}}$ for the CAPTEX and ANATEX cases). Sometimes, A_p and A_m designate areas; here, they simply represent the sets of model results and measurement data at the sampling sites.
- FA2 and FA5, the percentage of model predictions that agree within a factor of two and a factor of five, respectively, with the measurements.
- $\mathrm{FOEX} = 100(N_{(P_i > M_i)}/N - 0.5)$, where $N_{(P_i > M_i)}$ is the number of overpredictions. FOEX ranges between -50 and $+50\%$, and indicates whether overpredictions or underpredictions are more frequent.

The suitability of the above parameters was examined using the global analysis for the first ETEX case, assuming that all measured data are known with an accuracy of $0.03\ \mathrm{ng\,m^{-3}}$, which is approximately the magnitude of the tracer background concentration variability (Piringer *et al.*, 1997). A Monte Carlo (MC) global analysis was made, manipulating the original measurements by adding normally distributed random errors with a standard deviation of $0.03\ \mathrm{ng\,m^{-3}}$. Negative values were set to zero. Then, the MC analysis results were compared with those from an undisturbed global analysis. To ensure stability of the results, this experiment was repeated 20 times, but the outcome was similar in all repetitions.

It turned out that some of the statistical parameters are highly sensitive to measurement errors, whereas others are very robust (Table 1). Most sensitive are the FA2 and FA5 values, which are almost halved in the MC analysis. This is due to the great weight that is given by these parameters to the large number of zero or near-zero concentration values. For example, a small measurement error (or an error caused by the subtraction of the uncertain background concentration) can change a measured value that should be zero actually into a small positive value, which pushes a correct model estimate of zero outside the FA2 and FA5 ranges. Therefore, FA2 and FA5 should not be used for model evaluation.

Table 1. Sensitivity of the statistical parameters to small changes in measured concentration values

	FB	NMSE	r	r_s	FMS	FA2	FA5	FOEX
Reference analysis	0.01	14.9	0.59	0.66	0.53	0.57	0.66	0.09
MC analysis	− 0.01	13.7	0.59	0.47	0.54	0.34	0.46	− 0.09

Note: "Reference analysis" is the normal global analysis of the first ETEX case, "MC analysis" is the average of 20 global analyses using randomly manipulated measurements.

The same is true for r_s. Although it would be sensible to use a non-parametric correlation coefficient, r_s is not suited because of the large number of identical (i.e. zero or very small) concentration values. Any ranking of these identical values gives ill-determined ties. The analysis also indicates that FB is slightly sensitive to concentration changes. However, this is an artifact of the restriction of the manipulated measurements to positive values, which increases the average of the measured concentrations. In reality, FB will only be affected by systematic errors. The same is probably true for NMSE and FOEX.

Very robust are the Pearson correlation coefficient r and FMS which remain almost unchanged in the MC analysis (FMS only if the significant concentration level is well above the uncertainty of the measurements). Since only statistical quantities that are not sensitive to measurement errors are sensible for the model evaluation, the further examinations are based exclusively on FB, NMSE, FMS and r. FOEX is not used because it gives similar information as FB.

It must be pointed out that it is not strictly correct to compare NMSE values obtained from different observational data sets, since NMSE partly depends on the distribution of the measurements. Furthermore, NMSE gives a smaller value for an overpredicting model than for an underpredicting one of otherwise comparable performance (Poli and Cirillo, 1993). Therefore, NMSE is not ideal to evaluate model performance, but it is reported here because it is widely used for model validation.

3.2. *Model evaluation*

3.2.1. *CAPTEX*. CAPTEX was conducted during September and October 1983 in the northeastern United States and southeastern Canada (Ferber *et al*., 1986). There were seven 3 h releases, C1-C7, of approximately 200 kg perfluoro-monomethyl-cyclohexane (PMCH) from two different locations. Tracer samples of 3 and 6 h duration were collected at 84 sites. The 3 h samples were converted to 6 h averages to have a uniform dataset. Data are typically available until 36 to 48 h after the release and up to downwind distances of approximately 1000 km. Data from release C6 are not used because the amount of material released was smaller and few samples were collected.

Releases C1–C4 were made during daytime under anticyclonic conditions with well-developed convective boundary layers. The southwesterly winds carried the tracer right over the sampling network. C5 and C7 were made from the second location during the nighttime behind cold fronts. Winds were from a northwesterly direction.

For all six CAPTEX releases, reasonably good agreement between the model results and the measurements is observed (Fig. 1). The results for C1 and C2 are especially good. The model cannot catch the small-scale variations in the measured tracer concentrations because they are mostly below the resolution of the meteorological input data, but there is good agreement on the larger scale.

Less agreement is found for C3: the advection speed of the tracer is underestimated, and the concentrations are strongly overpredicted by the model. Both facts indicate that in reality much of the tracer was transported at higher levels. For C4, the predicted tracer cloud is transported too far south, off the coast. For C5, the measurements suggest a complicated small-scale plume structure which is not captured well by the model. Unresolved mesoscale flow patterns might have been important for this nighttime release. C7 is well predicted, with a slight overestimation of transport speed.

Figure 2 shows a scatter plot of observed and predicted concentrations for all CAPTEX releases. Most pairs are composed either of zero measured or of zero predicted concentrations resulting in an L-shaped plot. However, especially for higher concentrations the model shows some skill.

Table 2 contains results of the summary statistics of the global analyses of each release individually, and a global analysis of all the data together. For five releases, there is a (mostly small) negative fractional bias, whereas for C3, the model strongly overestimates the tracer concentrations. Totally, the biases cancel. The NMSE are all within 13 and 40, and the FMS range from 31 to 60%. r are all significantly (at the 99% confidence level) different from zero, and range between 0.32 and 0.70. Overall, these statistics indicate a satisfactory model performance.

CAPTEX data were used in many other studies (Draxler, 1987; Draxler and Stunder, 1988; Lee, 1987; Brost *et al*., 1988; Kao and Yamada, 1988; Shi *et al*., 1990) to evaluate dispersion models, but few evaluations were quantitative, and even fewer were based on paired statistics. Shi *et al*. (1990) modeled releases C1, C2 and C5 and reported FMS values for the same threshold level as was used here (2 fl ℓ^{-1}) of 28, 30, and 10%, respectively, which can be compared directly to the FLEXPART results of 48, 60 and 34%. Brost *et al*. (1988) gave FMS values which cannot be compared directly to those obtained in this study, since they used measured concentrations that were interpolated to a regular grid and used different thresholds. However, their values seem to be lower than ours. Draxler (1987) reported correlation coefficients for concentrations above 3 fl ℓ^{-1} that are lower than those calculated in this paper, but they cannot be directly compared, since we also used concentrations less than 3 fl ℓ^{-1} to calculate the correlation coefficients. Since all these studies are around 10 yr old, it is not surprising that FLEXPART shows more skill in predicting the tracer dispersion during CAPTEX. These models were simpler, and the ECMWF reanalyses are also an improvement upon older analyses. Nevertheless, the results of FLEXPART are encouraging.

3.2.2. *ANATEX*. ANATEX is the experiment on the largest scale and with the greatest number of

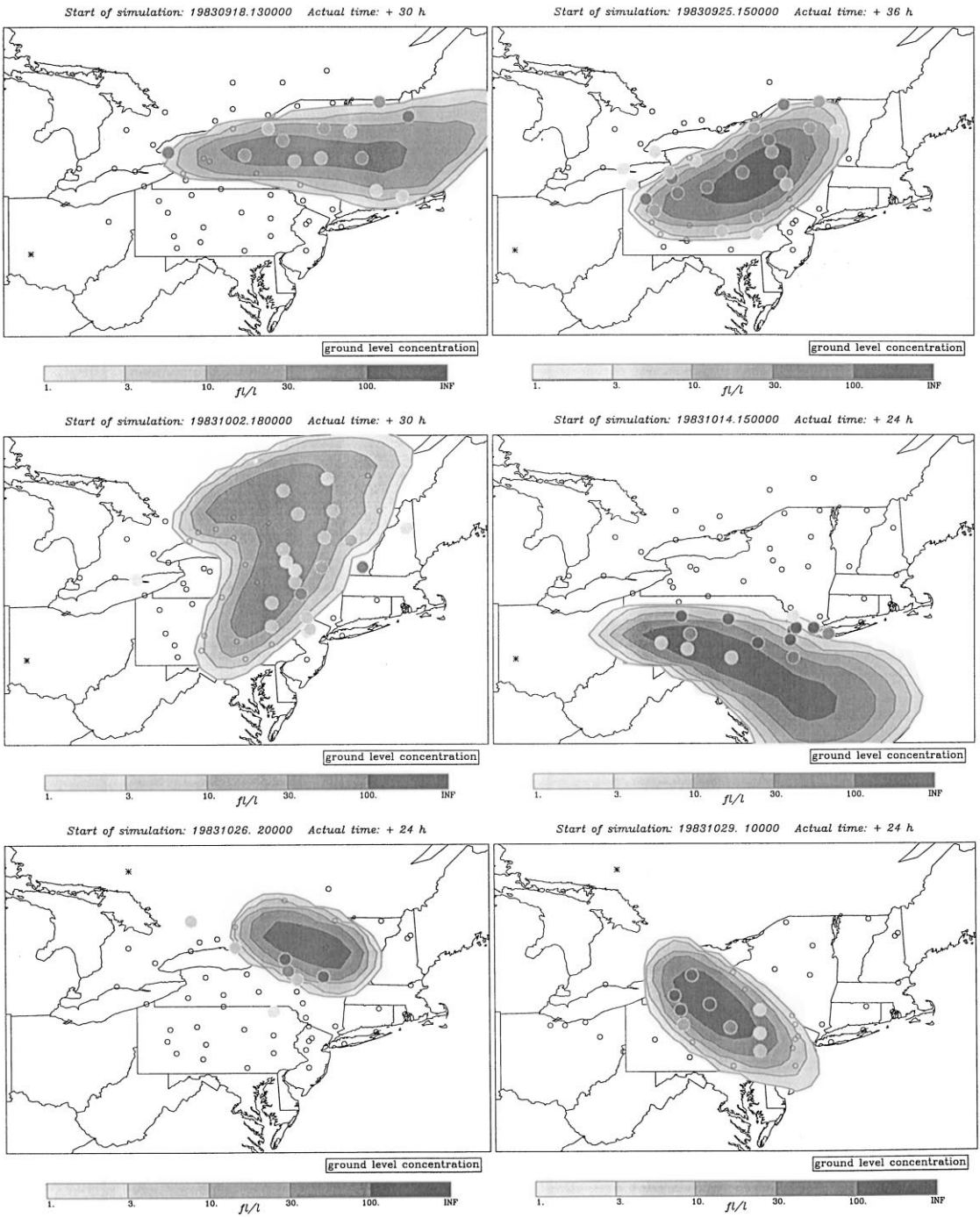

Fig. 1. Comparison of the FLEXPART model results with CAPTEX measurements for releases C1 (top, left), C2 (top, right), C3 (middle, left), C4 (middle, right), C5 (bottom, left) and C7 (bottom, right). Data are shown for the latest time for which a large number of measurements was available and before the plume left the investigation area. The data presented are 6 h averages at 26, 34, 29, 23, 22 and 20 h after the start of the six releases.

releases. 33 releases of different perfluorocarbons were made from each of two sites between 5 January 1987 and 29 March 1987. The tracers were routinely released in a 3 h period every 2.5 d, alternating between daytime and nighttime releases. *Orthocis*-perfluorodimethylcyclohexane (*oc*PDCH) was released from St. Cloud, perfluorotrimethylcyclohexane (PTCH) was released from Glasgow. In addition, PMCH was released every five days from St. Cloud to facilitate the separation of the plumes originating

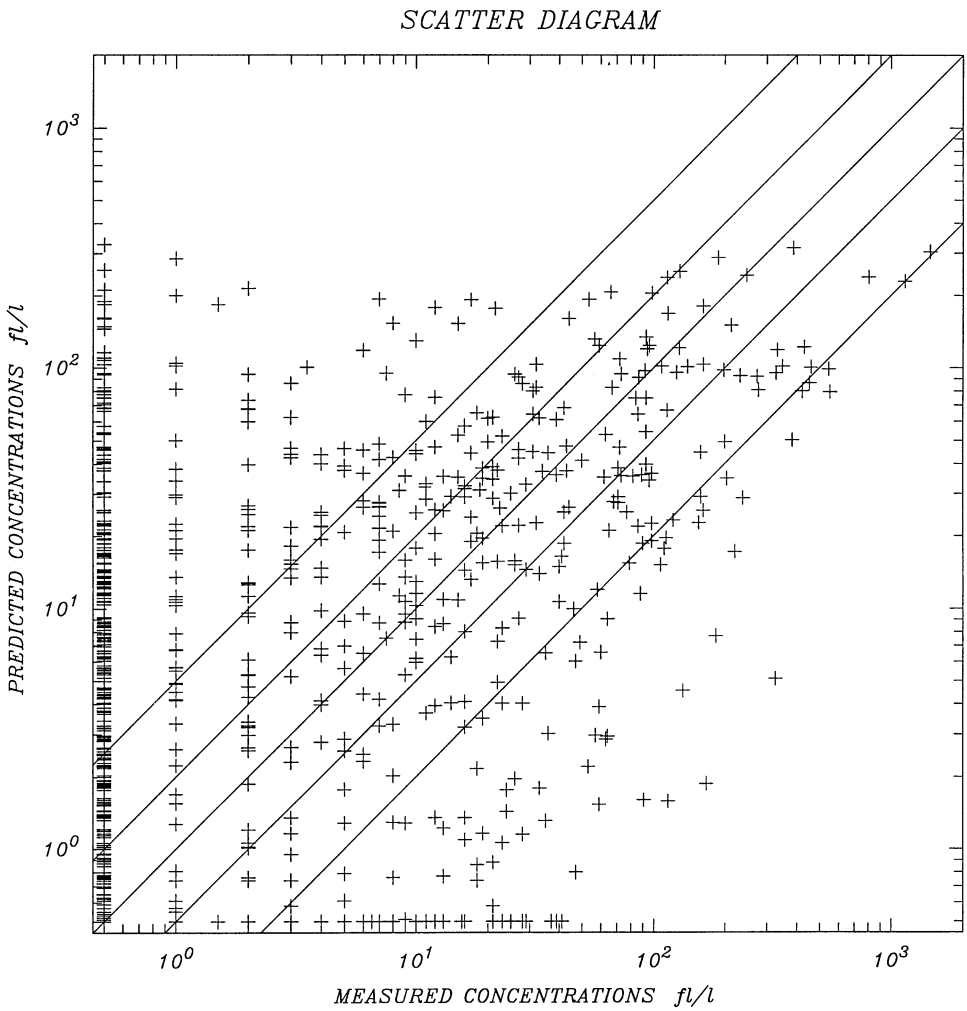

Fig. 2. Scatter plot between observed and predicted tracer concentrations for all CAPTEX releases. For better guidance, the FA2 and FA5 bands are shown. All concentrations less than 0.5 fl ℓ^{-1} are plotted at that value.

Table 2. Global analyses of the CAPTEX releases

Release	Number	FB	NMSE	FMS	r
C1	362	− 0.06	30.1	48	0.70*
C2	366	− 0.00	12.6	60	0.65*
C3	366	+ 1.02	24.5	31	0.32*
C4	312	− 0.13	38.1	44	0.41*
C5	316	− 0.55	40.0	34	0.58*
C7	206	− 0.23	13.6	42	0.39*
C1–C7	1928	0.00	22.6	46	0.49*

Note: Number gives the total number of data pairs used for the analysis (including zero values). r that are significantly (at the 99% confidence level) positive, are marked by "*".

from subsequent releases. Ground-level air samples of 24 h duration were taken at 77 sites (Draxler and Heffter, 1989; Draxler *et al.*, 1991). The spatial and temporal resolution of the data is rather low compared to CAPTEX and ETEX. We use these data to evaluate the FLEXPART model for times up to 117 h after the start of each release.

ANATEX data were used by many authors to evaluate models. They made qualitative comparisons (Fay *et al.*, 1995), comparisons of computed trajectories with tracer-derived trajectories (Haagenson *et al.*, 1990; Draxler, 1991), investigations of the plume arrival times at the different arcs of the sampling network (Clark and Cohn, 1990), evaluations of the footprint areas of the tracer plumes (Rodriguez *et al.*, 1995), and studied the concentration distribution functions (Clark and Cohn, 1990; Rodriguez *et al.*, 1995). Because of the problem of separating individual plumes, few attempts were made to pair modeled and measured concentrations in space and time. Since our method of evaluation necessitates such a pairing, we first developed a methodology to separate individual plumes. This was achieved by the following five steps:

First, *oc*PDCH concentrations greater than 1 fl ℓ^{-1} were checked for consistency with the PMCH concentrations (this was not done for lower concentrations because of the measurement uncertainty). For the daytime releases, when *oc*PDCH was released together with PMCH, a corrected *oc*PDCH concentration was calculated according to $C^c_{\mathrm{pdch}} = FC_{\mathrm{pmch}}$, where C_{pmch} is the PMCH concentration. F takes into account the different release amounts and the different molecular weights of *oc*PDCH and PMCH. In the rare cases where $C^c_{\mathrm{pdch}} > C_{\mathrm{pdch}}$, an average of these was used. For the nighttime releases, $C^c_{\mathrm{pdch}} = C_{\mathrm{pdch}} - FC_{\mathrm{pmch}}$.

In the second step, visual inspections of the measurement data were made. When it appeared that a plume could not be separated from another one, or when it was poorly documented by the measurements, it was discarded. The cases excluded from further study were mostly in agreement with those disregarded in other studies (e.g. Haagenson *et al*., 1990). After this step, 17 releases of *oc*PDCH and 15 releases of *oc*PTCH were left for further study.

In the third step, tracer concentrations that could not originate from the release considered were excluded. The exclusion criterion was defined by an upper limit for a plausible straight-line separation speed between the tracer material and the modeled tracer plume centroid. The maximum allowed separation speeds were 10, 7, 6, 5 and 4 ms^{-1} for the five days after each release. This step was done to remove data gathered, for instance, at a location 2000 km from the release site at the day of the release. Such data are not relevant for the release considered, but indicate the presence of the plume from a previous release. The relatively large thresholds for the separation speeds should ascertain that few data were excluded erroneously. This was verified by subjective inspections.

In the fourth step, all measurement points (including those with zero concentrations) that were covered by the footprint of the modeled plume of the following release were excluded. They were left to be explained by this following release. This step was only done for the PTCH data.

Finally, stations closer than 90 km to the release location were removed from the data set because the model cannot resolve the flow on such a small scale.

The number of releases during ANATEX is too large to discuss all cases individually. Instead, we will first focus on the overall results of the statistical evaluation (Table 3), before we will show a few interesting cases in more detail.

We would have expected the model to compare better with the measurements for the *oc*PDCH releases than for the PTCH releases, since the individual *oc*PDCH plumes were better separable using the PMCH data. However, the agreement was better for PTCH than for *oc*PDCH, showing smaller NMSE and larger FMS and r on the average. We were not able to explain this controversial behavior, but it is possible that the PTCH plumes intermingled less than the *oc*PDCH plumes. The PTCH release location was more at the fringe of the sampling network, and plumes were more frequently carried out of it, leaving those plumes that were actually carried into the network more clearly separated. That would mean that there are still considerable ambiguities in the *oc*PDCH measurement data, even after all the attempts to remove them.

Table 3. Global analysis of the ANATEX releases

Release	Number	FB	NMSE	FMS	r
PDCH1	177	0.65	32.3	33	0.34*
PDCH3	99	0.77	12.5	42	0.83*
PDCH7	147	0.22	56.8	18	0.15
PDCH8	111	0.18	9.0	69	0.49*
PDCH11	187	0.27	38.7	50	0.43*
PDCH12	154	0.31	5.9	28	0.33*
PDCH15	203	0.67	21.5	22	0.44*
PDCH16	156	1.15	169.0	15	0.21*
PDCH20	157	0.45	66.3	10	0.00
PDCH22	112	− 0.51	5.7	27	0.39*
PDCH23	76	− 0.03	7.5	22	0.57*
PDCH24	115	− 0.85	44.4	50	0.92*
PDCH26	112	0.15	12.8	38	0.60*
PDCH27	165	1.12	38.2	18	0.24*
PDCH31	167	− 1.29	15.4	0	0.14
PDCH32	131	− 0.98	22.7	0	− 0.02
PDCH33	201	− 0.97	13.3	27	0.62*
PDCH1-33	2470	0.17	52.3	28	0.35*
PTCH1	146	− 0.54	21.1	67	0.46*
PTCH2	120	− 0.61	7.0	23	0.49*
PTCH3	186	− 0.40	10.1	46	0.68*
PTCH4	142	0.45	14.4	23	0.44*
PTCH7	148	− 0.94	6.7	100	0.91*
PTCH9	190	0.30	46.5	36	0.22*
PTCH10	136	− 0.44	4.1	34	0.52*
PTCH11	163	− 0.99	19.7	31	0.89*
PTCH12	160	− 0.35	4.8	10	0.24*
PTCH18	113	0.17	3.8	43	0.76*
PTCH22	70	− 0.81	7.0	33	0.65*
PTCH28	61	− 1.02	79.9	100	1.00*
PTCH30	103	0.56	31.3	12	0.79*
PTCH32	62	− 0.64	5.7	25	0.23
PTCH33	112	− 0.79	7.3	22	0.75*
PTCH1-33	1912	− 0.35	50.2	32	0.70*

Note: Number gives the total number of data pairs used for the analysis (including zero values). r that are significantly (at the 99% confidence level) positive, are marked by "*".

Many of the *oc*PDCH concentrations were, partly strongly, overestimated by the model. On the contrary, many PTCH concentrations were underestimated. Combining results for *oc*PDCH and PTCH, there is large scatter in FB, revealing no clear tendency of the model towards overprediction or underprediction. NMSE values also show large variation. Some cases (especially PTCH releases) have very small NMSE, indicating good model performance, but some also show very large NMSE values. FMS for the *oc*PDCH releases are nearly all low. The FMS exceeds 50% only for one case. For PTCH the results are better, with three cases of FMS > 50%.

For *oc*PDCH there are four cases with correlation coefficients not significantly (on the 99% confidence level) different from zero, and a few cases with high correlation. The correlations for PTCH are all but one significantly positive. A few are very high, indicating that the general transport patterns were well captured by the model.

Summarizing the results of the statistical evaluation, the model clearly showed skill to describe the tracer transport in some cases. In others, however, there is poor agreement with the measurements. Totally, the results are worse than for the CAPTEX data. This could be due to shortcomings of the model, intermingling of plumes, and the more complex meteorological conditions (see below) during ANATEX.

Figures 3–6 present four examples of the tracer transport. Figure 3 shows results for release PDCH8 for which the statistical evaluation indicated rather good model performance. There was small FB (0.18), small NMSE (9.0), very high FMS (69%) and a reasonable correlation (0.49). This favorable performance is reflected in the concentration maps. There was qualitative agreement with the measurements up to transport times of 105 h. The good model performance was observed during a period with rather simple meteorological conditions, where no fronts were present (Draxler, 1988).

As a sharp contrast, Fig. 4 shows release PDCH15 for which the statistical evaluation was much poorer with $\mathrm{FB} = 0.67$, $\mathrm{NMSE} = 21.5$, $\mathrm{FMS} = 22\%$ and $r = 0.44$. The modeled concentrations were much too high during the whole period. The meteorological situation during and after the tracer release was much more complex than in the previous example (Draxler, 1988). A cold front passed the release location shortly after the release. Later, a warm front and a second cold front passed by. It seems that vertical transport processes associated with these fronts that were not resolved by the ECMWF re-analyses carried most of the tracer material out of the PBL. This explains the large bias of the model. On the other hand, the transport of the material remaining in the PBL was reasonably well predicted by the model, since the modeled and measured surface tracer cloud positions corresponded in spite of the complex transport patterns (the tracer was first carried eastwards, later back westwards, northeastwards, and finally southwards).

Figure 5 shows release PTCH3, for which the model compared rather well with the measurements ($\mathrm{FB} = -0.40$, $\mathrm{NMSE} = 10.1$, $\mathrm{FMS} = 46\%$, $r = 0.68$). At 69–93 h, the tracer cloud was dispersed over a large area, indicated by both the measurements and the model results. The good model performance was again associated with the absence of fronts.

Figure 6 shows release PTCH32, a case that was poorly captured by the model ($\mathrm{FB} = -0.64$, $\mathrm{NMSE} = 5.7$, $\mathrm{FMS} = 25\%$, $r = 0.23$). The model produced a very elongated plume during the first 33 h after the release, whereas the observed plume was much broader and located eastwards of the modeled plume. The large negative bias is due to an erroneous plume position shortly after the release, when the modeled tracer cloud was advected outside of the sampling network. There were no fronts present, but the release site was located at the northwestern edge of a low pressure system. A slight shift of the position of this system in the ECMWF re-analyses might have sufficed to produce the erroneous transport patterns.

3.2.3. *ETEX*. During ETEX, two releases of 12 h duration each were made near Rennes in western France on 23 October and 14 November 1994. The experiment is described in detail elsewhere in this issue (Nodop *et al*., 1998). Tracer samples of 3 h duration are available up to 90 h after each release at 168 stations.

Figure 7 shows the temporal evolution of the tracer plume following the first ETEX release (E1). For the first 20 h after the start of the release, the model tends to overpredict the measurements, but the spatial position of the plume is captured very well. Up to 65 h after the release, the model results are in excellent agreement with the measurements. There is little bias, and the plume position is also correctly predicted. Probably the modeled plume travels somewhat faster eastwards than the actual plume but this is hard to judge because of a lack of measurements in eastern Europe. After 65 h, the plume is still predicted rather well, but the main part of the plume already starts to escape the sampling network. The low concentrations that are still measured in Central Europe cannot be found in the model results.

Comparing the statistical evaluations of the model performance (Table 4) with those for CAPTEX and ANATEX (Tables 2 and 3), it can be seen that the results for E1 are among the better ones produced by FLEXPART. It is important to keep in mind that the number of measurements was much larger during ETEX than during the other experiments, and the duration of each sample was 3 h rather than 6 or 24 h, which is more difficult to predict. Overall, the results for E1 come close to what one might call a perfect simulation for a model that neglects many physical processes (e.g. cloud venting).

E2 represents the other end of the spectrum of what can be achieved with the model (Fig. 8). The statistical evaluation reveals that this is, by far, the worst simulation of all done in this study (Table 4). This is true for all statistical parameters, but the outstanding feature is the large bias; FLEXPART overpredicted the observed concentrations by a factor of ten.

Although the spatial position of the tracer surface footprint is predicted rather well during the first 24 h (Fig. 8), the concentrations within the plume are highly overpredicted. During the following 24 h, the modeled tracer cloud is quickly carried out of the sampling network, but always stays in contact with the ground, while the measurements show that most of the tracer has disappeared. The poor model prediction was not caused by erroneous analyses of the

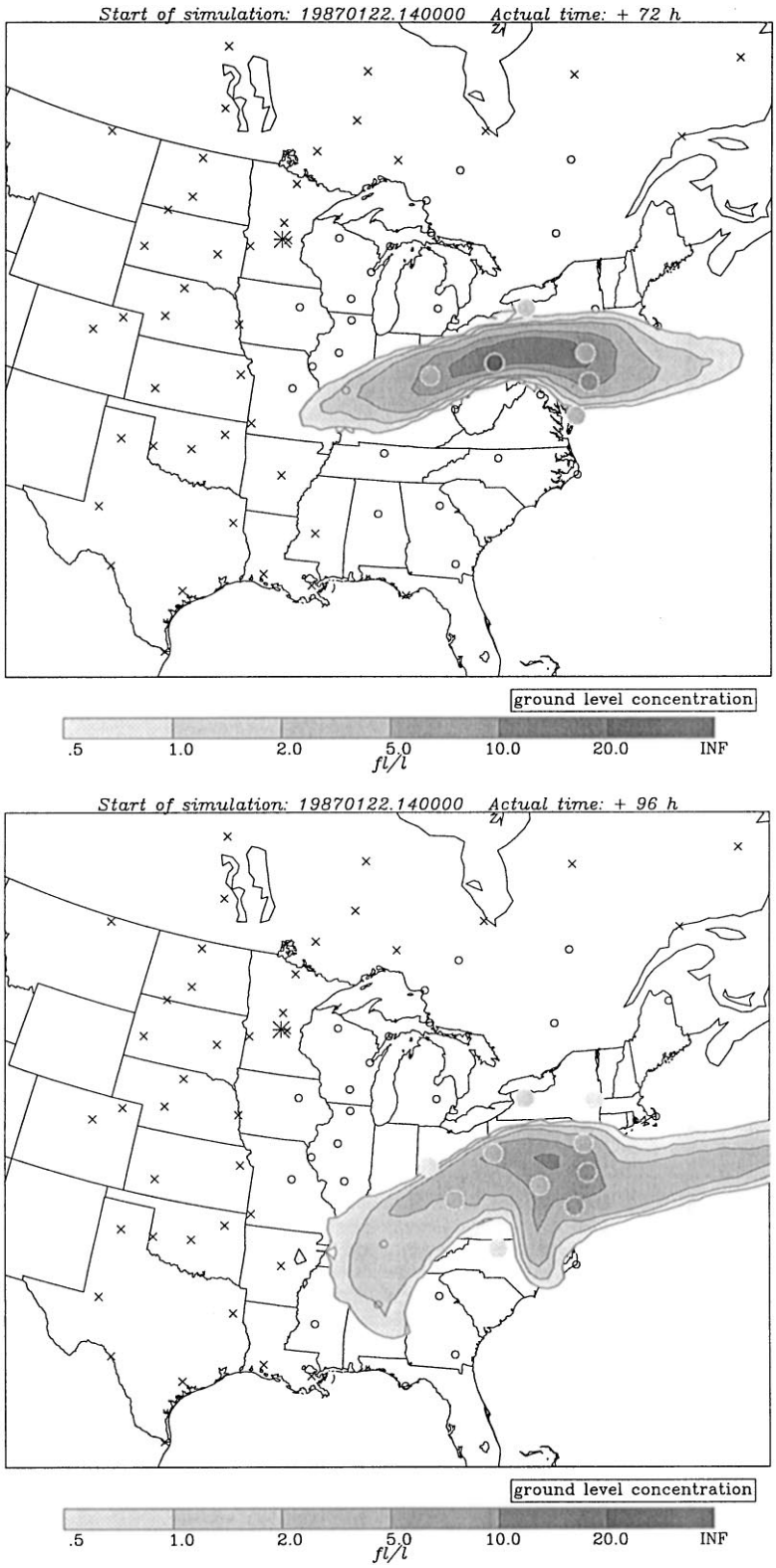

Fig. 3. Comparison of the modeled concentration fields with ANATEX measurements at 33–57 h (top) and 57–81 h (bottom) after the begin of release PDCH8 (23 January 1987, 5 UTC). '×' mark stations that were removed from the data set during one of the steps to remove ambiguities.

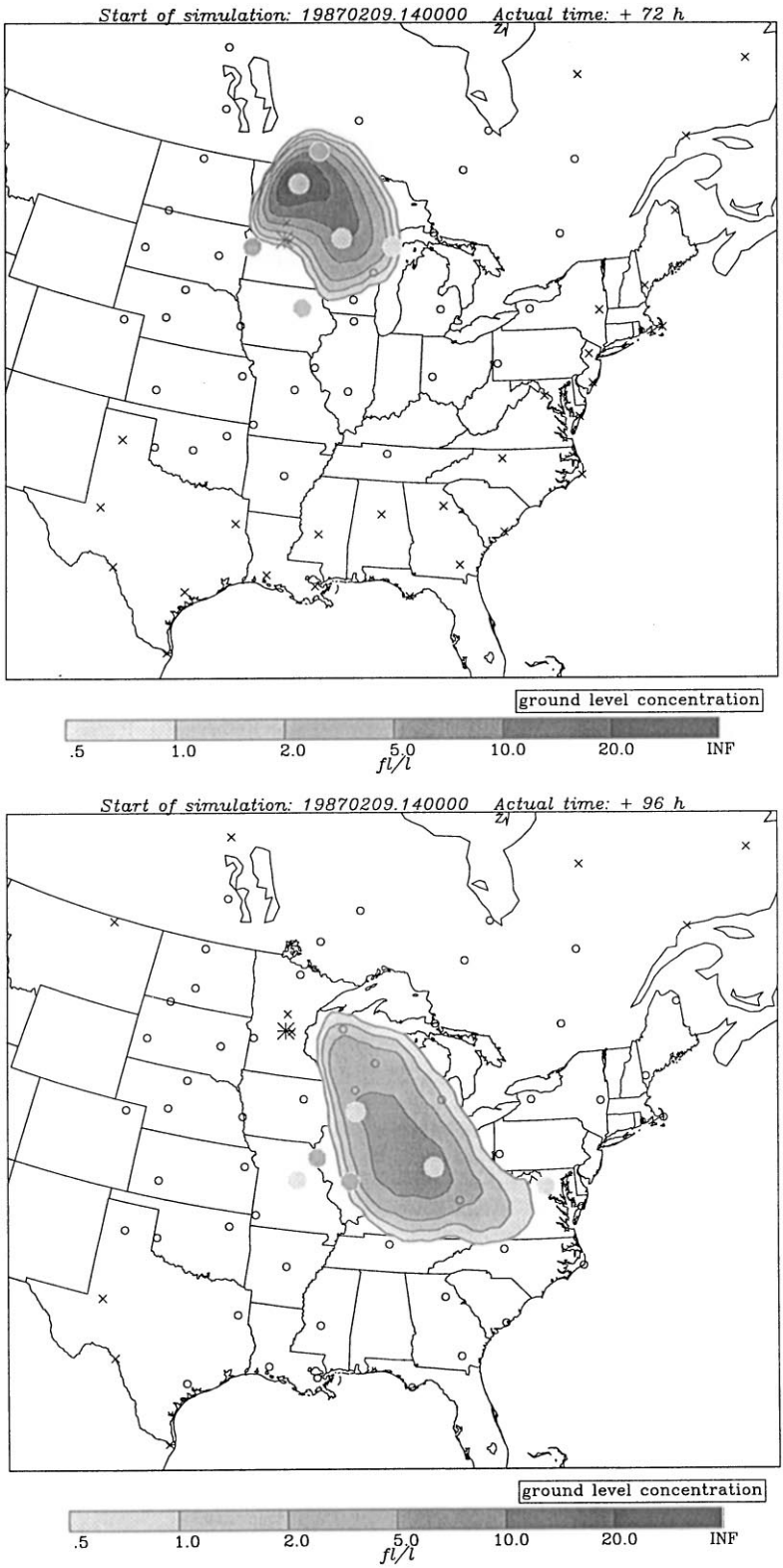

Fig. 4. Comparison of the modeled concentration fields with ANATEX measurements at 45–69 h (top) and 69–93 h (bottom) after the begin of release PDCH15 (9 February 1987, 17 UTC).

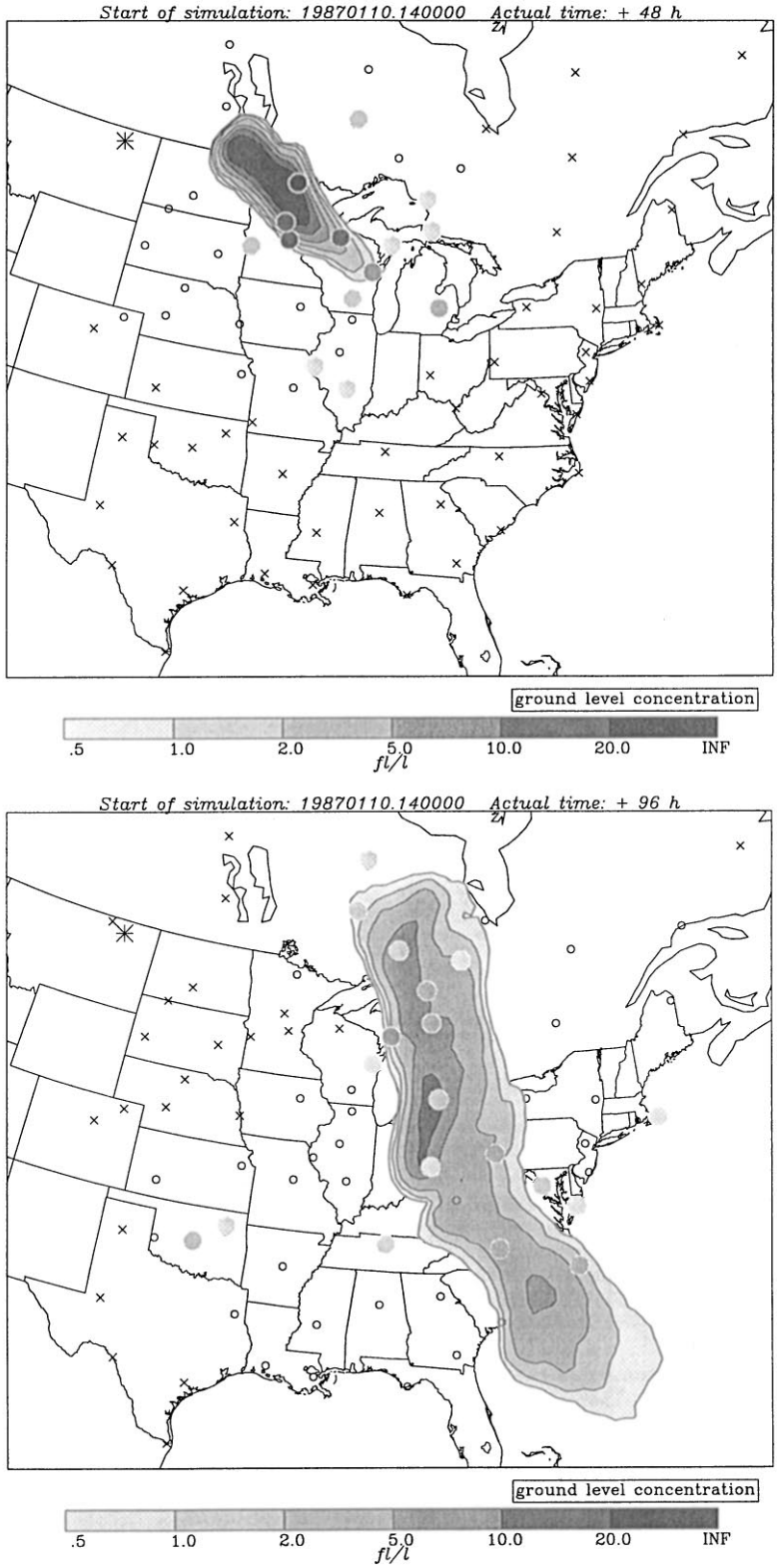

Fig. 5. Comparison of the modeled concentration fields with ANATEX measurements at 45–69 h (top) and 69–93 h (bottom) after the begin of release PTCH3 (10 January 1987, 17 UTC).

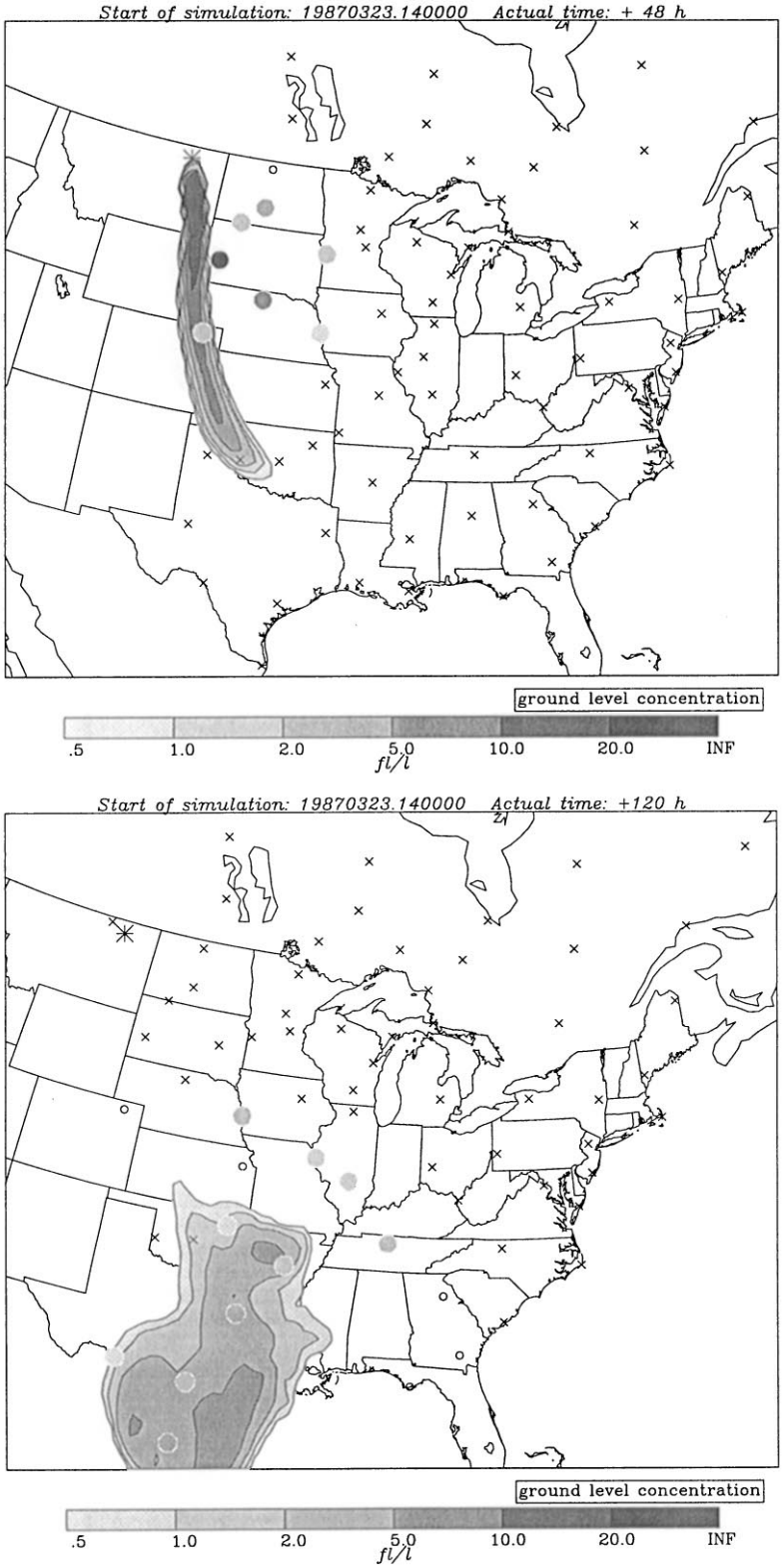

Fig. 6. Comparison of the modeled concentration fields with ANATEX measurements at 9–33 h (top) and 81–105 h (bottom) after the begin of release PTCH32 (24 March 1987, 5 UTC).

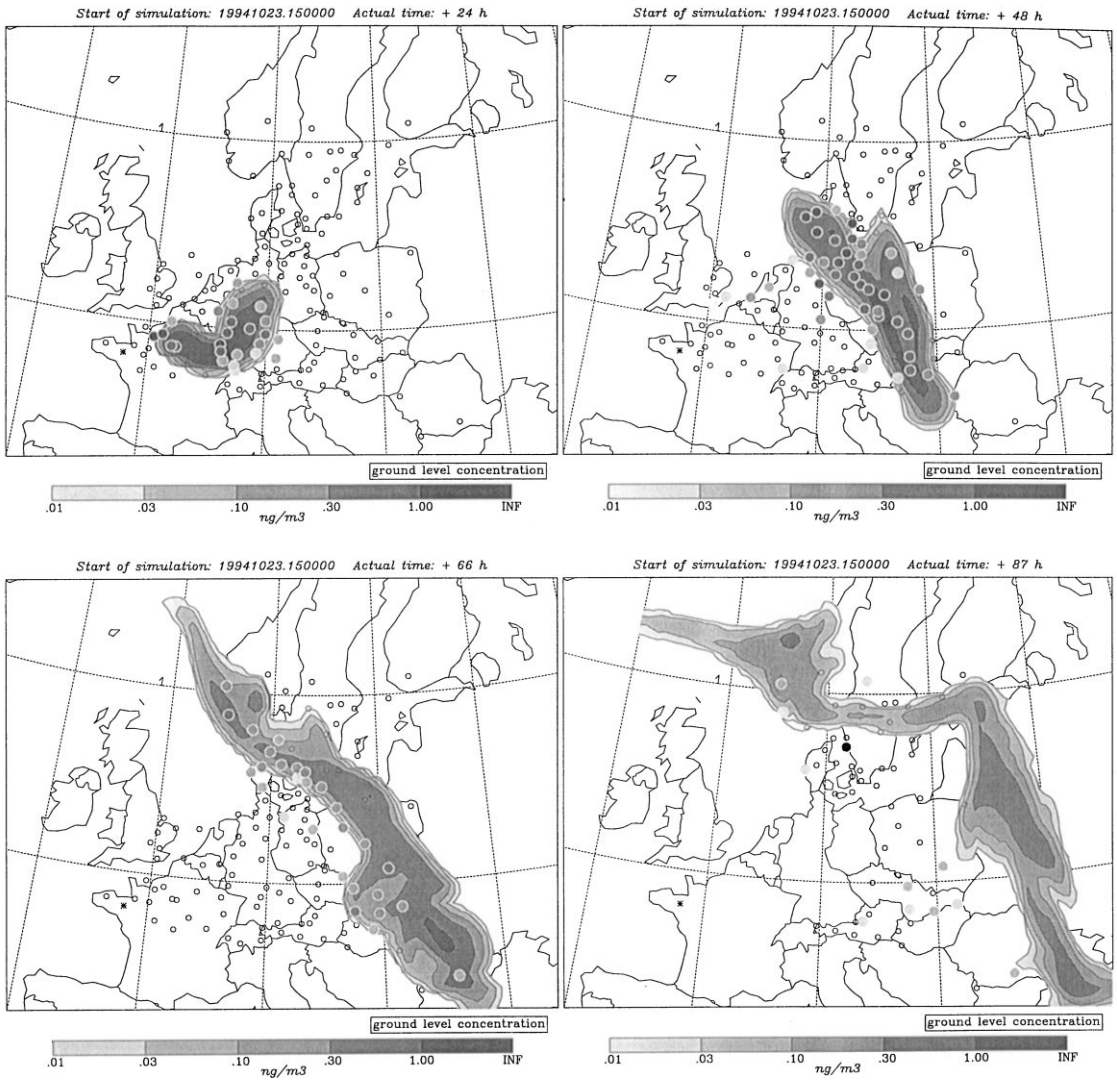

Fig. 7. Comparison of the FLEXPART model results with the measurements during E1. The plots refer to 23, 47, 65 and 86 h after the start of the release.

Table 4. Global analysis of the ETEX releases

Release	Number	FB	NMSE	FMS	*r*
E1	2991	0.01	14.9	53	0.59*
E2	2205	1.73	970.2	4	0.29*
E1–E2	5196	0.56	78.5	40	0.29*

Note: Number gives the total number of data pairs used for the analysis (including zero values). *r* that are significantly (at the 99% confidence level) positive, are marked by "*".

horizontal wind, since trajectories were predicted excellently for the same case. See Stohl and Koffi (1998) for a discussion.

3.3. *Dependence of the model performance on the travel time*

It was interesting to investigate how the model performance depends on the time passed since the beginning of a release. Therefore, the statistical parameters used in the global analyses were calculated separately for every sampling period. One would assume that the model performance deteriorates with time. Surprisingly, this seems not to be the case, as can be seen in Fig. 9, a scatterplot of *r* against travel time for all releases. Large and small values of *r* can be found shortly after the release as well as 100 h later. This is typical also for FB, NMSE and FMS.

An explanation of this behavior is as follows: Shortly after the release, the tracer cloud is small, and its structure is not well resolved either by the measurement network or by the model. Although a particle model is more accurate than a Eulerian model close to the source, its effective resolution also depends on the density of the meteorological input data. Therefore, low correlations can be expected at the begin of a simulation. Ideally, a phase with good agreement between model and measurements should follow,

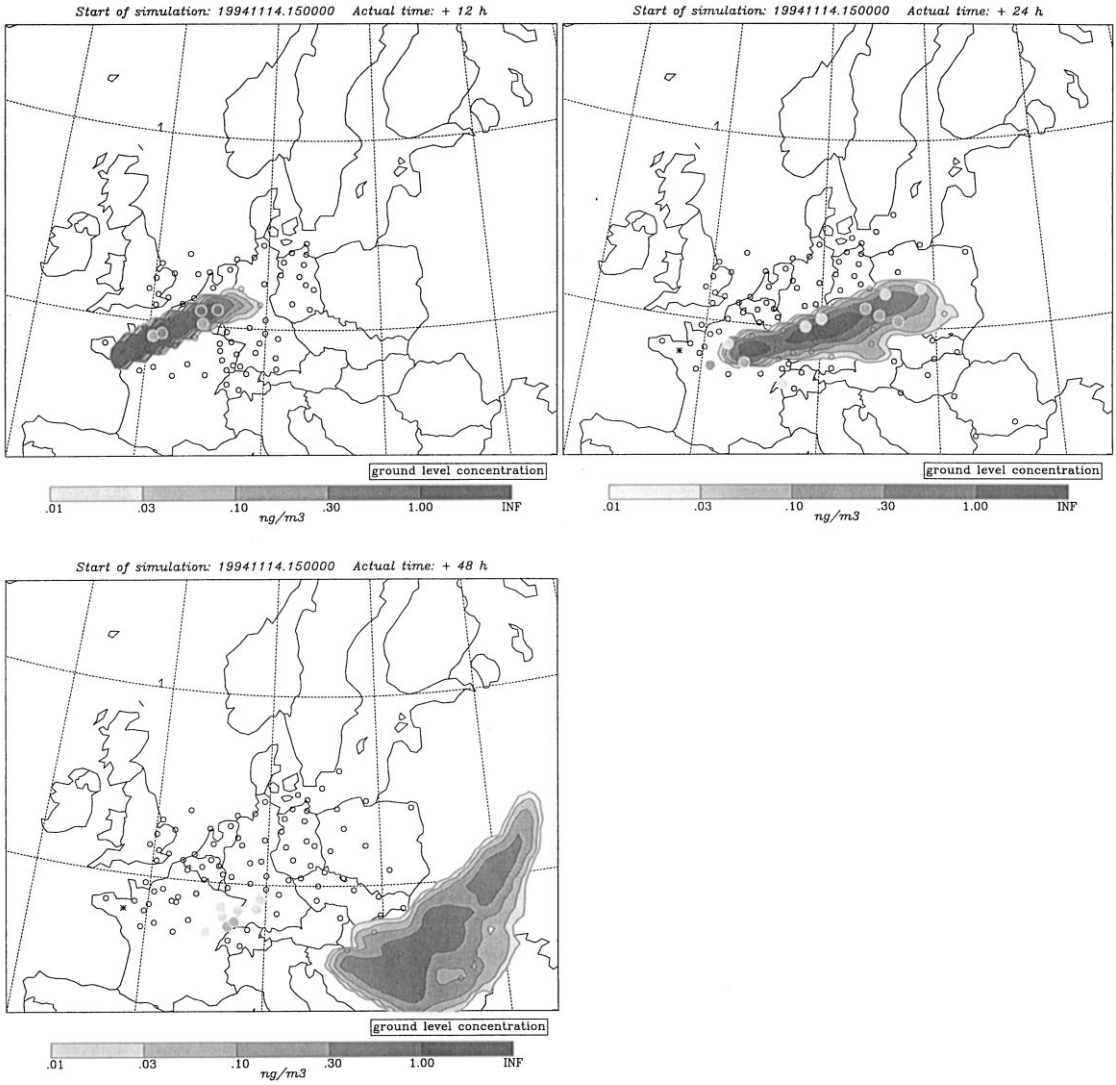

Fig. 8. Comparison of the FLEXPART model results with the measurements during E2. The plots refer to 12, 24 and 48 h after the start of the release.

before the model results start to deteriorate due to the accumulation of transport errors. These assumptions were confirmed by the inspection of many individual cases. However, since the duration of the above-mentioned phases varies with the meteorological conditions, no general pattern can be identified. We assume that a general deterioration of model performance might have been noticed for travel times longer than the maximum times used in this study.

3.4. *The effect of the concentration calculation procedure*

Four methods to calculate tracer concentrations at the receptor locations, given the spatial particle distributions, were tested. Normally, the FLEXPART model uses a parabolic kernel (Uliasz, 1994) to estimate tracer concentrations at the receptor locations. The bandwidths of this kernel are calculated according to

$$h = 20{,}000\ \mathrm{m} + 0.8\ \mathrm{m\,s^{-1}}\, t_t + 16500\ \mathrm{m}\sqrt{t_t/10{,}800\ \mathrm{s}}, \quad (5)$$

where t_t is the time since the release. This gives bandwidths of 20, 140 and 380 km after 0, 24 and 96 h travel time, respectively. Because this definition is based on rather small estimates of trajectory errors (Stohl, 1998), and because the estimated concentrations are most strongly affected by the particles closest to the receptor locations, results of a hypothetically perfect simulation are not spoiled by this definition of the bandwidths. But due to its smoothing effect, it can minimize the impact of small inaccuracies in the transport simulations. All concentration estimates for the sampling sites presented so far were obtained with this method.

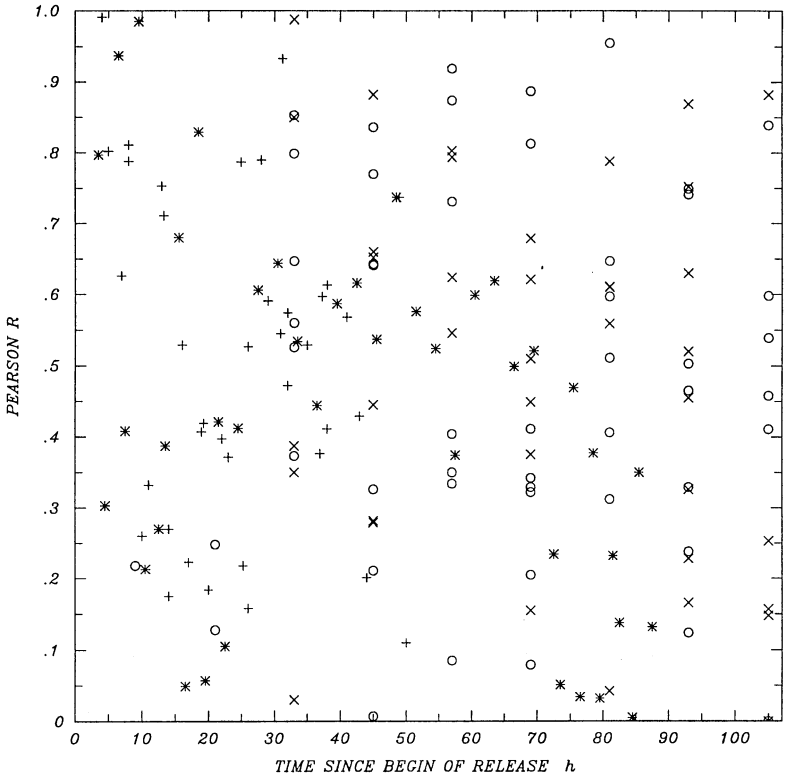

Fig. 9. Dependence of r on the travel time for the CAPTEX releases (+), the ANATEX *oc*PDCH releases (○), the ANATEX PTCH releases (×), and the ETEX releases (∗). Times are given as midpoint values of the sampling interval. Only values that were based on at least 20 measurements are shown.

A uniform kernel with a constant bandwidth of 0.5° latitude and longitude was used to estimate grid cell concentrations to produce the concentration maps. The concentration in a grid cell is calculated by

$$c = \frac{1}{V} \sum_{i=1}^{N} (m_i f_i) \tag{6}$$

with V being the grid cell volume, m_i the mass of particle i, N the total number of particles, and f_i the fraction of particle i attributed to the respective grid cell. f_i is calculated using a uniform kernel (Fig. 10). For the purpose of this comparison, this kernel is also applied to estimate concentrations directly at the receptor locations.

As a third method, the concentrations were obtained from the grid cells which covered the receptors. Refining this method, the concentrations were estimated by linear interpolation between the grid cells around the receptor location. The latter two methods are used in many Lagrangian or Eulerian models to estimate concentrations at discrete locations.

Global NMSE and r are presented for the comparison of the different methods. For CAPTEX and ETEX, global statistics for the individual releases as well as for all the releases together are shown, while for ANATEX, for the sake of brevity, only global

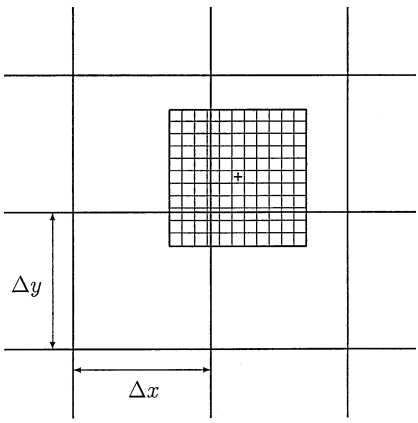

Fig. 10. Illustration of the uniform kernel used to calculate gridded concentration fields. The mass of a particle that is located at the center of the shaded rectangle with side lengths (Δx, Δy) is attributed to four cells according to the fraction of the shaded area that falls within each cell.

statistics for all *oc*PDCH and all PTCH releases combined are shown (Table 5). It can be seen that in many cases large differences of NMSE and r calculated with the different methods occur. For instance, for the ANATEX PTCH releases a rather high correlation of 0.70 and NMSE = 50.2 were obtained using the

Table 5. NMSE and r for the four different methods applied to estimate concentrations at the measurement sites

	Parabolic kernel		Uniform kernel		Cell estimate (interpolated)		Cell estimate (not interpolated)	
Release	NMSE	r	NMSE	r	NMSE	r	NMSE	r
C1	30.1	0.70	32.1	0.68	29.7	0.74	30.2	0.73
C2	12.6	0.65	14.4	0.61	14.1	0.61	16.4	0.55
C3	24.5	0.32	26.8	0.34	27.7	0.33	31.9	0.36
C4	38.1	0.41	39.4	0.41	40.1	0.39	35.0	0.49
C5	40.0	0.58	48.2	0.46	50.6	0.45	50.7	0.61
C7	13.6	0.39	16.9	0.31	17.1	0.31	19.1	0.27
C1–C7	22.6	0.49	25.8	0.46	25.8	0.46	27.1	0.46
PDCH1-33	52.3	0.35	86.0	0.18	91.0	0.15	59.5	0.16
PTCH1-33	50.2	0.70	85.3	0.38	97.9	0.15	94.4	0.18
E1	14.9	0.59	17.5	0.51	16.2	0.53	18.9	0.53
E2	970.2	0.29	491.4	0.29	483.8	0.37	450.6	0.51
E1–E2	78.5	0.29	46.4	0.33	45.0	0.34	44.1	0.41

parabolic kernel. Using the interpolated grid cell estimate, a rather poor correlation ($r = 0.15$) and NMSE = 97.9 were calculated. The choice of the concentration calculation procedure thus has a large effect on the model evaluation statistics.

There is no single method which was best in all the cases considered, but the parabolic kernel method often gave better results than the others. Only for E2, the parabolic kernel produced clearly worse results than the other methods, but in this case the model performance was the poorest of all 40 releases considered in this paper, whatever method was used. For the remaining releases, the other three methods usually gave worse results than the parabolic kernel. However, although the differences between these methods can also be quite large in individual cases, on average they performed rather similar.

3.5. *The effect of the subgrid scale parameterization of the PBL height*

Before the discussion on the subgrid-scale parameterization of the PBL heights is started, one must be aware of a deficiency of current particle models. To our knowledge, all particle models applied until recently for long-range dispersion calculations assume constant density flows, whereas in reality, density near the ground is higher than at the top of the PBL. This leads to an underestimation of ground-level concentrations, an overestimation of the concentrations at the top of the PBL, and in the presence of a vertical wind shear also to an inaccurate simulation of transport. The problem was recently solved by Stohl and Thomson (1998), who developed a density correction for particle models, which, however, was not available at the time of this study. Using this correction, surface concentrations for the CAPTEX cases increased by up to 8%. Thus, one would expect that a model without the density correction tends to underpredict surface concentrations.

In contrast to that, we have found that a previous version of FLEXPART tended to overpredict the ground-level concentrations. Since the method used to calculate PBL heights was shown to give good results in comparison with measurements (Vogelezang and Holtslag, 1996), we attributed these overpredictions to subgrid-scale variations of the PBL heights. Therefore, we increased the calculated PBL heights by adding one standard deviation of the subgrid topography, and by taking the maximum PBL height around a particle position as described in Section 2. However, this is only based on qualitative arguments, and it was not clear from the beginning whether it actually improves the results. Therefore, we compared the results of the reference model to calculations for which the subgrid parameterization for the PBL heights was not used.

Table 6 shows the results of this comparison. It can be noticed that the reference model version is neither biased towards overprediction nor towards underprediction. On the other hand, neglecting subgrid variations of the PBL heights leads to systematic overpredictions in most cases. Of the 40 cases studied here, 28 were overpredicted, and only 12 were (most of them slightly) underpredicted. Also in terms of the NMSE, the model performance was improved for most cases except for many of the ANATEX PTCH releases, although the NMSE favors an overpredicting model. The FMS values produced by the two model versions were very similar in most cases, but they also point towards a slight improvement due to the subgrid parameterization. The same is true for r. Visual comparisons confirmed that the concentration patterns were not very sensitive to increasing the PBL heights, but the general concentration level was reduced.

3.6. *The influence of the vertical wind velocity*

In recent years, it has been shown by some authors that it is beneficial to use the grid-scale vertical wind

Table 6. Comparison of the reference simulations with simulations for which subgrid variations of the PBL heights were not accounted for, and with simulations for which the vertical wind was set to zero

	Reference				No subgrid PBL variations				Zero vertical wind			
Release	FB	NMSE	FMS	r	FB	NMSE	FMS	r	FB	NMSE	FMS	r
C1	− 0.06	30.1	48	0.70	0.10	25.2	44	0.70	− 0.07	34.8	45	0.63
C2	− 0.00	12.6	60	0.65	0.02	14.0	62	0.56	− 0.06	12.5	60	0.67
C3	1.02	24.5	31	0.32	1.16	54.7	31	0.26	0.93	22.6	30	0.31
C4	− 0.13	38.1	44	0.41	0.09	31.5	59	0.42	− 0.76	72.8	40	0.45
C5	− 0.55	40.0	34	0.58	− 0.21	31.9	29	0.48	− 0.26	30.7	38	0.52
C7	− 0.23	13.6	42	0.39	0.12	14.8	40	0.26	− 0.51	18.8	45	0.34
C1–C7	0.00	22.6	46	0.49	0.19	24.3	47	0.38	− 0.13	25.5	45	0.50
PDCH1-33	0.17	52.3	28	0.35	0.54	65.5	25	0.41	0.34	47.3	22	0.33
PTCH1-33	− 0.35	50.2	32	0.70	0.17	22.1	32	0.76	− 0.04	29.6	33	0.71
E1	0.01	14.9	53	0.59	0.74	58.4	53	0.48	− 0.02	16.6	50	0.59
E2	1.73	970.2	4	0.29	1.84	1134.7	3	0.24	1.81	1045.4	5	0.29
E1–E2	0.56	78.5	40	0.29	1.12	124.0	37	0.36	0.74	98.2	38	0.26

for trajectory calculations although it is much less accurate than the horizontal wind (for a discussion see Stohl (1998)). To investigate the influence of the vertical wind on the model results, all simulations were repeated switching it off (Table 6). All in all, the results are not strongly affected by the neglect of the vertical wind, but the results can be quite sensitive in individual cases. For instance, the plume of release PTCH7 is completely lifted off the ground when the vertical wind is used in the simulation. This is confirmed by the measurements, as practically no concentrations above the background were observed 72 h after the release. On the other hand, if the vertical wind is switched off in the simulations, ground-level concentrations up to more than 10 fl ℓ^{-1} are calculated at the same time.

From the global analyses it becomes not completely clear whether it is, on the average, beneficial to use the vertical wind or not, although a tendency towards improvement may be noticed. Partly, these inconclusive results can be explained by the large interpolation errors of the vertical wind that occur when coarse resolution fields such as the ECMWF re-analyses are used (Stohl *et al*., 1995). For both ETEX cases, for which the resolution of the meteorological fields was higher, results were (though not strongly) improved by including the vertical wind.

4. CONCLUSIONS

FLEXPART is a recently developed Lagrangian particle dispersion model. In this study, it was applied to calculate the long-range dispersion of inert tracer gases. 40 releases during CAPTEX, ANATEX and ETEX for which tracer gas measurements were available at many locations, were considered. The statistical evaluation of the model performance was based on calculated and measured concentrations paired in space and time. A Monte-Carlo analysis where random errors were added to the measurement data to represent typical uncertainties revealed that FA2 and FA5 which are often used for model validation are very sensitive to uncertainties in the measurements. Therefore, they should not be used for model validation in future studies. We also found the non-parametric Spearman rank correlation coefficient to be sensitive to measurement uncertainties. Very robust are the Pearson correlation coefficient, the NMSE, FOEX, fractional bias, and the FMS.

It was found that the model usually performs very well under undisturbed meteorological conditions, but shows less skill in the presence of fronts. The analyses of the model results showed that FLEXPART captured most CAPTEX releases well because the meteorological conditions were relatively simple. For many of the ANATEX releases, the model results were relatively poor because of the complex meteorological conditions. ETEX showed the full range of what can be obtained with the model. After the first ETEX release, which took place behind a cold front, the tracer was advected in a strong westerly flow, but did not encounter regions with strong vertical motions. For this case, FLEXPART produced excellent results. During the second experiment, the tracer was released immediately ahead of a cold front, and was subsequently advected almost parallel to this front for some hours. The model failed completely in this case, giving an overestimation of the ground level concentrations by a factor of ten. The poor model performance in the presence of fronts can be explained by two factors: first, the resolution of the meteorological fields used is too coarse to represent the variability of vertical motions in the vicinity of fronts. Secondly, convective clouds behind cold fronts can vent tracer material out of the PBL. This process is presently not accounted for in the model. If, in the future, a parameterization of cloud venting can be found, the model

results may improve. Any such parameterization should be checked with the aircraft measurements that were undertaken during the experiments.

The subgrid variability of PBL heights is important for the results of the dispersion model. If this variability is neglected, the model systematically overestimates the ground-level concentrations. The parameterization of subgrid PBL variability that we used is based on qualitative arguments. A more quantitative and more physical parameterization should be found, which could also be inspired by inspecting the aircraft measurements.

In individual cases, it was possible to demonstrate the importance of considering the vertical wind in the model calculations. On the average, however, the model results did not benefit strongly from including the vertical wind. This is most probably due to the coarse resolution of the input data, especially for CAPTEX and ANATEX, which do not sufficiently resolve the highly variable fields of the vertical wind.

The results of the statistical evaluations showed that, on average, there is no clear dependence of the model performance on the time passed since the release. Very good as well as very poor model performance can be found shortly after the release and also five days later. A likely explanation for this is that in the initial phase the plume is not sufficiently resolved by the meteorological input data. Therefore, large errors can occur. In addition, the statistical results shortly after a release may be biased towards larger errors, because the plume is undersampled by the measurements. Subsequently the evaluation improves, before, due to the accumulation of transport errors, it deteriorates again. However, the results show that the model often shows considerable skill in predicting the measurements even five days after the release.

Another conclusion of this study is that the results of the model evaluation are very sensitive to the choice of the method by which the modeled concentrations at the receptor sites are obtained. We found that a parabolic kernel with bandwidths that increase with the travel time of the particles gave the best results. The other three methods, a uniform kernel and two methods which used the gridded concentration fields performed similar on average, but may nevertheless give very different results in individual cases.

Acknowledgements—A. Stohl and M. Hittenberger had the luck to spend their compulsory military services developing and evaluating the FLEXPART model. We are, therefore, thankful to the NBC school of the Austrian Armed Forces, specifically to Mjr. Januschke who was always supportive of the project. The development of the deposition routines for FLEXPART was supported by the Central Institute for Meteorology and Geodynamics (Dr. U. Pechinger). We are thankful to ECMWF to give us access and support (Mr. N. Kreitz). Dr. P. Seibert is acknowledged for numerous helpful discussions. Dr. R. Draxler supplied us with the ANATEX data and manuals and answered questions concerning the data. Dr. K. Nodop did the same concerning the ETEX data. We are strongly indebted to all the people who did the valuable measurements during CAPTEX, ANATEX and ETEX. An anonymous reviewer gave helpful comments.

REFERENCES

Archer, G., Girardi, F., Graziani, G., Klug, W., Mosca, S. and Nodop, K. (1996) The European long range tracer experiment (ETEX). Preliminary evaluation of model intercomparison exercise. In *Air Pollution Modeling and its Application XI*, eds S. E. Gryning and F. A. Schiermeier, Vol. 21, pp. 181–190. Plenum Press, New York.

Baumann, K. and Stohl, A. (1997) Validation of a long-range trajectory model using gas balloon tracks from the Gordon Bennett Cup 95. *Journal of Applied Meteorology* **36**, 711–720.

Berkowicz, R. and Prahm, L. P. (1982) Evaluation of the profile method for estimation of surface fluxes of momentum and heat. *Atmospheric Environment* **16**, 2809–2819.

Brost, R. A., Haagenson, P. L. and Kuo Y. -H. (1988) Eulerian simulation of tracer distribution during CAPTEX. *Journal of Applied Meteorology* **27**, 579–593.

Clark, T. L. and Cohn, R. D. (1990) The Across North America Tracer Experiment (ANATEX) Model Evaluation Study. EPA/600/3-90/051, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Draxler, R. R. (1987) Sensitivity of a trajectory model to the spatial and temporal resolution of the meteorological data during CAPTEX. *Journal of Applied Meteorology* **26**, 1577–1588.

Draxler, R. R. (1988) Across North America Tracer Experiment (ANATEX) : weather maps and tracer concentrations. NOAA Techn. Memo. ERL ARL-165. National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.

Draxler, R. R. (1991) The accuracy of trajectories during ANATEX calculated using dynamic model analyses versus rawinsonde observations. *Journal of Applied Meteorology* **30**, 1446–1467.

Draxler, R. R., Dietz, R., Lagomarsino, R. J. and Start, G. (1991) Across North America Tracer Experiment (ANATEX) sampling and analysis. *Atmospheric Environment* **25A**, 2815–2836.

Draxler, R. R. and Heffter, J. L., eds. (1989) Across North America Tracer Experiment (ANATEX) Vol. I: Description, ground-level sampling at primary sites and meteorology. NOAA Technical Memo. ERL ARL-167. National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.

Draxler, R. R. and Stunder, B. J. B. (1988) Modeling the CAPTEX vertical tracer concentration profiles. *Journal of Applied Meteorology* **27**, 617–625.

ECMWF (1995) *User Guide to ECMWF Products* 2.1. Meteorological Bulletin M3.2. ECMWF, Reading, UK.

Fay, B., Glaab, H., Jacobsen, I. and Schrodin, R. (1995) Evaluation of Eulerian and Lagrangian atmospheric transport models at the Deutscher Wetterdienst using ANATEX surface tracer data. *Atmospheric Environment* **29**, 2485–2497.

Ferber, G. J., Heffter, J. L., Draxler, R. R., Lagomarsino, R. J., Thomas, F. L., Dietz, R. N. and Benkovitz, C. M. (1986) Cross-Appalachian Tracer Experiment (CAPTEX 83). Final Report NOAA Technical Memo. ERL ARL-142, Air Resources Laboratory, NOAA Environmental Research Laboratories, Silver Spring, MD 20910, 60 p.

Gibson, R., Kållberg, P. and Uppala, S. (1996) The ECMWF Re-Analysis (ERA) project. *ECMWF Newsletter* **73**, 7–17.

Gupta, S., McNider, R. T., Trainer, M., Zamora, R. J., Knupp, K. and Singh, M. P. (1997) Nocturnal wind structure and plume growth rates due to inertial oscillations. *Journal of Applied Meteorology* **36**, 1050–1063.

Haagenson, P. L., Gao, K. and Kuo, Y. -H. (1990) Evaluation of meteorological analyses, simulations, and long-range transport using ANATEX surface tracer data. *Journal of Applied Meteorology* **29**, 1268–1283.

Hanna, S. R. (1982) Applications in air pollution modeling. In: *Atmospheric Turbulence and Air Pollution Modelling*, eds. F. T. M. Nieuwstadt and van Dop, H. Reidel, Dordrecht, Holland.

Hubbe, J. M., Doran, J. C., Liljegren, J. C. and Shaw, W. J. (1997) Observations of spatial variations of boundary layer structure over the southern Great Plains cloud and radiation testbed. *Journal of Applied Meteorology* **36**, 1221–1231.

Kao, C. -Y. J. and Yamada, T. (1988) Use of the CAPTEX data for evaluations of a long-range transport numerical model with a four-dimensional data assimilation technique. *Monthly Weather Review* **116**, 293–306.

Kahl, J. D. W. (1996) On the prediction of trajectory model error. *Atmospheric Environment* **30**, 2945–2957.

Lee, I. Y. (1987) Numerical simulations of cross-Appalachian transport and diffusion. *Boundary-Layer Meteorology* **39**, 53–66.

Legg, B. J. and Raupach, M. R. (1982) Markov-chain simulation of particle dispersion in inhomogeneous flows: the mean drift velocity induced by a gradient in Eulerian velocity variance. *Boundary-Layer Meteorology* **24**, 3–13.

Lorimer, G. S. (1986) The kernel method for air quality modelling — I. Mathemical foundation. *Atmospheric Environment* **20**, 1447–1452.

Maryon, R. H. (1998) Determining cross-wind variance for low frequency wind meander. *Atmospheric Environment* **32**, 115–121.

Mosca, S., Graziani, G., Klug, W., Bellasio, R. and Bianconi, R. (1997) ATMES-II — Evaluation of long-range dispersion models using 1st ETEX release data. Vol. I. Draft Report.

Nodop, K., Connolly, R. and Girardi, F. (1998) The European Tracer Experiment and perfluorocarbon measurements in Europe from two surface releases. *Atmospheric Environment* (this issue).

Piringer, M., Baumann, K., Rötzer, H., Riesing, J. and Nodop, K. (1997) Results on perfluorocarbon background concentrations in Austria. *Atmospheric Environment* **31**, 515–527.

Poli, A. A. and Cirillo, M. C. (1993) On the use of the normalized mean square error in evaluating dispersion model performance. *Atmospheric Environment* **27A**, 2427–2434.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W.T. (1990) *Numerical Recipes in FORTRAN. The Art of Scientific Computing*. Cambridge University Press, New York, 702 pp.

Rodean, H. (1996) Stochastic Lagrangian Models of Turbulent Diffusion, *Meteorological Monographs*, Vol. **26** (48). American Meteorological Society, Boston, U.S.A.

Rodriguez, D., Walker, H., Klepikova, N., Kostrikov, A. and Zhuk, Y. (1995) Evaluation of two pollutant dispersion models over continental scales. *Atmospheric Environment* **29**, 799–812.

Shi, B., Kahl, J., Christidis, Z. D. and Samson, P. J. (1990) Simulation of the three-dimensional distribution of tracer during the Cross-Appalachian Tracer Experiment. *Journal of Geophysical Research* **95**, 3693–3703.

Stohl, A. (1997a) *The FLEXPART Particle Diffusion Model – Scientific Description and Technical Documentation*. Report available from the author. 47p.

Stohl, A. (1997b) The FLEXPART model — performance evaluation using tracer data from the first ETEX release. *ATMES-II — Evaluation of Long-Range Dispersion Models Using 1st ETEX Release Data. Vol. II — Models description*. Draft Report.

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

Stohl, A., Baumann, K., Wotawa, G., Langer, M., Neininger, B., Piringer, M. and Formayer, H. (1997) Diagnostic downscaling of large scale wind fields to compute local scale trajectories. *Journal of Applied Meteorology* **36**, 931–942.

Stohl, A. and Koffi, E. (1998) Evaluation of trajectories calculated from ECMWF data against constant level balloon flights during ETEX. *Atmospheric Environment* **32**, 4151–4156.

Stohl, A. and Seibert, P. (1998) Accuracy of trajectories as determined from the conservation of meteorological tracers. *Qarterly Journal of the Royal Meteorological Society* **125**, 1465–1484.

Stohl, A. and Thomson, D. J. (1998) A density correction for Lagrangian particle dispersion models. *Boundary-Layer Meteorology* (in press).

Stohl, A. and Wotawa, G. (1997) Validation of the Lagrangian particle model FLEXPART using ETEX data. In: Nodop, K. (editor) : *ETEX Symposium on Long-Range Atmospheric Transport, Model Verification and Emergency Response*, European Commission EUR 17346, pp. 167–170.

Stohl, A., Wotawa, G., Seibert, P. and 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. *Journal of Applied Meteorology* **34**, 2149–2165.

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

Uliasz, M. (1994) Lagrangian particle dispersion modeling in mesoscale applications. In *Environmental Modeling*, ed. P. Zannetti, Vol. II, pp. 71–101. Computational Mechanics Publications, Southampton, U.K.

Vogelezang, D. H. P. and Holtslag, A. A. M. (1996) : Evaluation and model impacts of alternative boundary-layer height formulations. *Boundary-Layer Meteorology* **81**, 245–269.

Wilson, D. J. and Flesch, T. K. (1993) Flow boundaries in random-flight dispersion models: enforcing the well-mixed condition. *Journal of Applied Meteorology* **32**, 1695–1707.

Wilson, J. D., Legg, B. J. and Thomson, D. J. (1983) Calculation of particle trajectories in the presence of a gradient in turbulent-velocity scale. *Boundary-Layer Meteorology* **27**, 163–169.

Wotawa, G. and Stohl, A. (1997) Boundary layer heights and surface fluxes of momentum and heat derived from ECMWF data for use in pollutant dispersion models — problems with data accuracy. In *The Determination of the Mixing Height – Current Progress and Problems*, eds. S.-E., Gryning, F. Beyrich and E. Batchvarova, EUR-ASAP Workshop Proceedings, 1–3 October 1997, Risø National Laboratory, Denmark.

Zannetti, P. (1992) Particle modeling and its application for simulating air pollution phenomena. In *Environmental Modelling*, eds P. Melli and P. Zannetti, pp. 211–241, Computational Mechanics Publications, Southampton, U.K.