



1352-2310(95)00314-2

TRAJECTORY STATISTICS—A NEW METHOD TO ESTABLISH SOURCE-RECEPTOR RELATIONSHIPS OF AIR POLLUTANTS AND ITS APPLICATION TO THE TRANSPORT OF PARTICULATE SULFATE IN EUROPE

ANDREAS STOHL

Institute of Meteorology and Geophysics, University of Vienna, Hohe Warte 38, 1190 Wien, Austria

(First received 13 April 1995 and in final form 1 August 1995)

Abstract A new method of trajectory statistics was developed which allows the identification of source areas with higher spatial resolution than other methods. It uses ambient air pollutant concentration measurements at a receptor site and corresponding back trajectories arriving at that site. In a first step, each pollutant concentration is attributed to its trajectory. From this, a first guess “concentration field” is computed which shows potential source areas of the respective pollutant. In an iterative procedure, the concentrations are redistributed along their corresponding trajectories which continuously improves the “concentration field”. The new method was tested with a large set of back trajectories ending at 14 measurement sites of the EMEP network. It was used to identify potential source areas of particulate sulfate. The procedure successfully identified many source areas with a much higher resolution than what would have been achievable with existing methods. The locations of the source areas were compared with the locations of the emission maxima of an emission inventory and a good agreement was found, both qualitatively and quantitatively.

Key word index: Trajectories, trajectory statistics, sulfate, long-range transport, source-receptor relationship.

1. INTRODUCTION

Back trajectories have often been used in combination with air quality measurements to identify potential source areas of air pollutants and to determine their respective contribution at receptor sites. They were used, for instance, to estimate the impact of the release of radioactive material after the Chernobyl accident (Klug *et al.*, 1992).

Various statistical methods have been developed to identify source areas using long-term air pollution measurements and trajectory calculations. Miller (1981) classified a large number of trajectories according to transport speeds and transport sectors. These so-called “flow climatologies” can roughly identify “regions of influence” for a specific site, but individual source areas are difficult to detect with this method. Nevertheless, it was applied by many authors to group air and precipitation chemistry data (e.g. Henderson *et al.*, 1982; Colin *et al.*, 1989; Miller *et al.*, 1993).

Moody and Samson (1989) and Harris and Kahl (1990) replaced the transport sector classification of trajectories by an automated cluster analysis algorithm. Cluster analysis is a multivariate statistical technique that splits a data set into a number of groups. These groups differ from each other in terms of a distance measure which has to be specified.

Dorling *et al.* (1992a) presented a computationally efficient clustering algorithm which uses Euclidean distance between trajectories as a distance measure. Stohl and Scheifinger (1994) used this algorithm for a weather pattern classification which gives information on the direction and speed of the air flow. Although it is possible to relate air pollution data with trajectory clusters (e.g. Dorling *et al.*, 1992b), the identification of source areas is difficult, especially if several prominent source areas exist.

Ashbaugh (1983) and Ashbaugh *et al.* (1985) developed a method which was specifically designed to detect source areas. They calculated the accumulated residence times of all trajectories within cells of a grid which they superimposed to the domain of the trajectory computations. This was done separately for a subset of high pollution events (resulting in a partial residence time t) and for the whole data set (resulting in a residence time T). The spatial distribution of the ratio t/T provides information on source areas. Trajectories travelling over grid cells with high t/T are more frequently associated with high pollutant concentrations at the receptor site than such passing over grid cells with low t/T . Variations of Ashbaughs method were used by Stohl and Kromp-Kolb (1994), Zeng and Hopke (1989) and Cheng *et al.* (1993).

Ashbaughs method has deficiencies in the determination of the statistical significance of its outcome.

Seibert *et al.* (1994) developed a better method to deal with uncertainties. Like Ashbaugh, they superimposed a grid to the domain of their trajectory computations. Then they calculated a logarithmic mean concentration for each grid cell according to the following formula:

$$\overline{C_{mn}} = \frac{1}{\sum_{l=1}^M \tau_{mnl}} \sum_{l=1}^M \log(c_l) \tau_{mnl} \quad (1)$$

where m, n are the indices of the horizontal grid, l is the index of the trajectory, M the total number of trajectories, c_l the concentration observed on arrival of trajectory l (corrected for the annual variation, see below) and τ_{mnl} the time spent in grid element (m, n) by trajectory l . Logarithmic concentrations were used because the concentrations followed an approximate log-normal distribution. A high $\overline{C_{mn}}$ means that, on average, air parcels passing over cell (m, n) result in high concentrations at the receptor site. The fields exhibit small-scale variations which are not necessarily statistically significant and should not be shown in a contour plot. Simple smoothing of the concentration field, however, is not justified, because this would also remove many significant structures. Therefore, Seibert *et al.* (1994) calculated a confidence interval (e.g., on the 95% confidence level) for the mean concentration of each grid cell assuming that the individual measurements are independent from each other and using t -statistics and the number of trajectories passing through each grid cell. Then they smoothed the concentration field with a 9-point filter, imposing the restriction that the values must be kept within their confidence interval. The smoothing was repeated until the change in the concentration field was less than a prespecified value. This procedure assures that significant variations are preserved while most of the insignificant ones are removed.

The Seibert *et al.* procedure does not extract all the information from the data, because the measured concentration is equally attributed to all segments of the related trajectory, whereas sources of air pollutants are often concentrated in "hot spots". The Seibert *et al.* procedure therefore underestimates the gradients of the "true" concentration fields. The method described here removes this discrepancy.

2. METHOD AND DATABASE

In order to understand the rationale of the method, let us imagine some trajectories which differ from each other, except for the fact that they all pass over one specific grid cell. Let all but one be "clean" trajectories, associated with low concentrations at the receptor site. Thus, no major pollutant sources are located along their paths and specifically not in the grid cell which they share with the one "polluted" trajectory. Therefore, the latter one must have taken up its pollutant load somewhere else along its path.

This information can be used for a redistribution of the concentrations along the trajectories.

First, the Seibert *et al.* method is used to compute a first guess concentration field $\overline{C_{mn}}$. Having done that, the redistribution starts. For the sake of simplicity, this process will be described only for a single trajectory l which is split into N_l segments with 1 h length each. Let c_l be the concentration measured upon arrival of trajectory l . Let furthermore X_{il} , $i = 1, N_l$, be the mean concentrations of the grid cells which are hit by segments $i = 1, N_l$ of trajectory l (this can be written as $X_{il} = 10^{\overline{C_{mn}^{(i)}}$ with (m, n) being the grid cell hit by trajectory segment i), and $\overline{X}_l = (\sum_{j=1}^{N_l} X_{jl})/N_l$ the average of the mean concentrations of the grid cells hit by the N_l segments of trajectory l . Then the redistributed concentrations along trajectory l are

$$c_{il} = c_l \frac{X_{il} N_l}{\sum_{j=1}^{N_l} X_{jl}} = c_l \frac{X_{il}}{\overline{X}_l}, \quad i = 1, N_l. \quad (2)$$

Equation (2) is now applied successively to all individual trajectories. After the redistribution is finished for all trajectories, a new concentration field can be computed with the equation

$$\overline{C_{mn}} = \frac{1}{\sum_{l=1}^M \sum_{i=1}^{N_l} \tau_{mnil}} \sum_{l=1}^M \sum_{i=1}^{N_l} \log(c_{il}) \tau_{mnil} \quad (3)$$

where τ_{mnil} is the residence time of segment i of trajectory l in grid cell (m, n) . Equation (3) is identical to equation (1), but the redistributed concentrations c_{il} are used instead of the measured concentrations c_l . Therefore, the summation is not only over the number of trajectories, but also over the number of segments of each trajectory. Insignificant features (on the 99% confidence level) of this field are again removed with the 9-point filter. With this new concentration field, a second redistribution of the concentrations along the trajectories is done, which in turn leads to a new concentration field. This procedure is repeated until the average difference between the concentration fields of two successive iterations is below 0.5%.

In this work, measurements with 24 h sampling time were used, but trajectories were available every 6 h. Therefore, the segments of four trajectories were combined into a single "composed trajectory" which allows a redistribution of the pollutant load between the individual trajectories of which the "composed trajectory" consists. This is an additional bonus of the new method, because not all trajectories might have crossed the same source areas. Correspondingly, the pollutant concentration at the receptor site might not have been constant during the 24 h period.

To test the capacity of the method, it was applied to identify source regions of particulate sulfate measured at several sites in Europe from 16 January to 31 December 1992. Particulate sulfate was chosen because it is subjected to long-range transport, and because reliable emission inventories exist for sulfur

dioxide (SO₂), from which most of the sulfate is produced. This is essential for a test of the new method, because the structure of the concentration field should be comparable to the structure of the emission field. The sulfate measurements of 13 sites were extracted from the EMEP database (Berg and Schaug, 1994) and data from Sevetijärvi were provided by Virkkula *et al.* (1995) (Table 1). The measurements had a temporal resolution of 24 h (Sevetijärvi 48 h). The criteria for the selection of these stations were good data quality and a uniform distribution over Europe.

The annual variation disturbs the identification of source areas and was therefore removed according to the formula:

$$c_l = c_{10} \frac{\bar{c}_g}{c_g} \quad (4)$$

where c_{10} is the concentration measured at a site, c_g is the 21 d running mean of the 14-station average concentration and \bar{c}_g its annual mean.

Ninety-six-hour back trajectories were calculated with the trajectory model FLEXTRA which was described in detail by Stohl *et al.* (1995). FLEXTRA was based on the analyzed and first guess model-level wind fields of the T213 L31 weather prediction model of the European Centre for Medium-Range Weather Forecasts (ECMWF, 1989). The wind fields had a horizontal resolution of 0.5° and a temporal resolution of 3 h. The model integrates with a flexible internal time step, but only hourly segments were extracted for this study. The trajectories used are a special type of mixed-layer trajectories for which the horizontal wind was averaged over a transport depth. This transport depth is determined from the mixed layer heights along the whole trajectory in an iterative procedure that was described by Stohl and Wotawa (1995).

A contour plot of \bar{C}_{mn} does not merely reflect the emission field of sulfur. It is also influenced by all processes which produce or remove sulfate. For example, if precipitation occurs along the path of a trajectory, sulfate may be efficiently removed by wet deposition. Regions with high average precipitation

may therefore be biased towards low \bar{C}_{mn} . This is also true for regions which are frequently crossed by trajectories prior to the regions with high average precipitation. These effects have to be taken into account when interpreting the results.

3. RESULTS

First, the method was applied to particulate sulfate measurements at the receptor station Kosetice. Figure 1 presents the concentration field \bar{C}_{mn} as computed with the Seibert *et al.* method, that is equivalent to the first guess field of the method described in this study. Note that the scale of Fig. 1 is logarithmic, the unit of the concentrations was $\mu\text{g m}^{-3}$. Results for those areas that are close to the border of the computation domain should be interpreted with caution, because relatively few trajectories have crossed these areas. Besides a general concentration gradient between the continent and the Atlantic, \bar{C}_{mn} does not reveal many structures.

Figure 2 shows the same result, but computed with the new method. In this case, \bar{C}_{mn} reveals detailed fine-scale structures. The concentration gradient between the continent and the Atlantic is much steeper than in Fig. 1. This indicates very low emissions from the sea as compared to the continent. Atlantic emissions are mainly of biogenic origin (in the form of dimethyl sulfide, DMS, which is transformed to sulfate), while European emissions are of anthropogenic origin (in the form of SO₂). The latter ones are clearly more important for receptor site Kosetice. Two pronounced concentration maxima exist. The highest concentrations are found in the so-called "Black Triangle", where the Czech Republic, Poland and the former East Germany meet. Transports over this region obviously cause the highest particulate sulfate concentrations in Kosetice. This is in excellent agreement with the EMEP SO₂ emission inventory for 1991 (Sandnes, 1993), which also shows the highest emissions in this area. A secondary maximum exists

Table 1. List of the measurement stations

Station	Country	Latitude	Longitude	Altitude (m)
Bilthoven	Netherlands	52.12	5.20	5
Eskdalemuir	U.K.	55.32	- 3.20	243
Illmitz	Austria	47.77	16.77	117
Irafoss	Iceland	64.08	- 21.85	61
Ispira	Italy	45.80	8.63	209
Kosetice	Czech Republic	49.58	15.08	633
La Cartuja	Spain	37.20	- 3.36	720
Preila	Lithuania	55.35	21.07	5
Utö	Finland	59.78	21.38	7
Sevetijärvi	Finland	69.58	28.83	130
Spitsbergen	Norway	78.90	11.88	474
Valentia	Ireland	51.93	- 10.25	9
Yarner Wood	U.K.	50.50	- 3.72	119
Witteveen	Netherlands	52.82	6.67	18

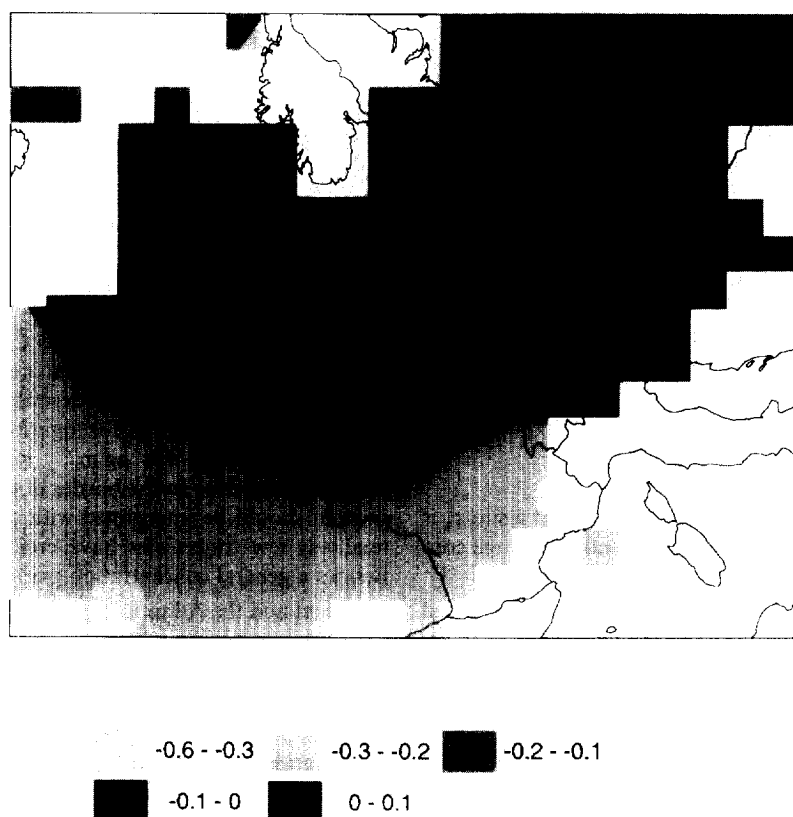


Fig. 1. Contour plot of the particulate sulfate concentration field \bar{C}_{mn} as computed with the Seibert *et al.* method for the receptor site Kosetice (black square). Grid cell size was $150 \times 150 \text{ km}^2$. Those cells which were crossed by less than 15 trajectories were left blank.

over England, again in good agreement with the emission inventory. The third maximum in western France is probably less realistic, although emissions in this area are higher than in most of the rest of France.

The trajectory statistics can be applied to all 14 stations simultaneously. In this case, trajectories ending at different receptor sites cross source areas from different directions. This enhances the redistribution process and thus improves the performance of the procedure. The different concentration levels at the individual measurement sites provide valuable additional information. For example, the maritime stations Irafoss and Spitsbergen have much lower annual mean concentrations than continental stations.

With the Seibert *et al.* method, values of \bar{C}_{mn} are much higher over the continent than over the sea (Fig. 3), but again, the Seibert *et al.* procedure does not reveal individual source areas. The new method performs much better (Fig. 4). A band of very high values of \bar{C}_{mn} stretches from England over The Netherlands and Germany to the "Black Triangle". This band can also be seen in the emission inventory, although the trajectory statistics in this case fail to detect the individual emission maxima over England and the Black Triangle. In Scandinavia, values of \bar{C}_{mn} are generally low, but a pronounced maximum exists over the Kola peninsula, again in very good agree-

ment with the emission inventory. Secondary maxima can be found over Poland, Spain and a small one over the Po Valley in northern Italy. The relatively high \bar{C}_{mn} over Spain are not represented in the emission inventory. A possible explanation could be the dry climate of Spain that prevents sulfate from being washed out which in turn leads to an overrepresentation in the concentration fields. On the other hand, the maximum over the Po Valley is more emphasized in the emission inventory than in the concentration field. The high frequency of inversions in the Po Valley could hinder the emissions from becoming subjected to long-range transport, leading to an underrepresentation in the concentration field. The concentration minimum over Eastern France seems to be realistic when compared to the emission inventory.

For a more quantitative comparison between the emission inventory and the concentration fields, it is useful to compute the correlation coefficient between the yearly emission rates and the concentrations in the grid cells. As the emission field does not follow any standard distribution, the nonparametric Spearman rank-order correlation coefficient r_s (Press *et al.*, 1992) was used. The first 12 columns of the EMEP inventory were excluded from the comparison, because over the Atlantic most grid cells have very low

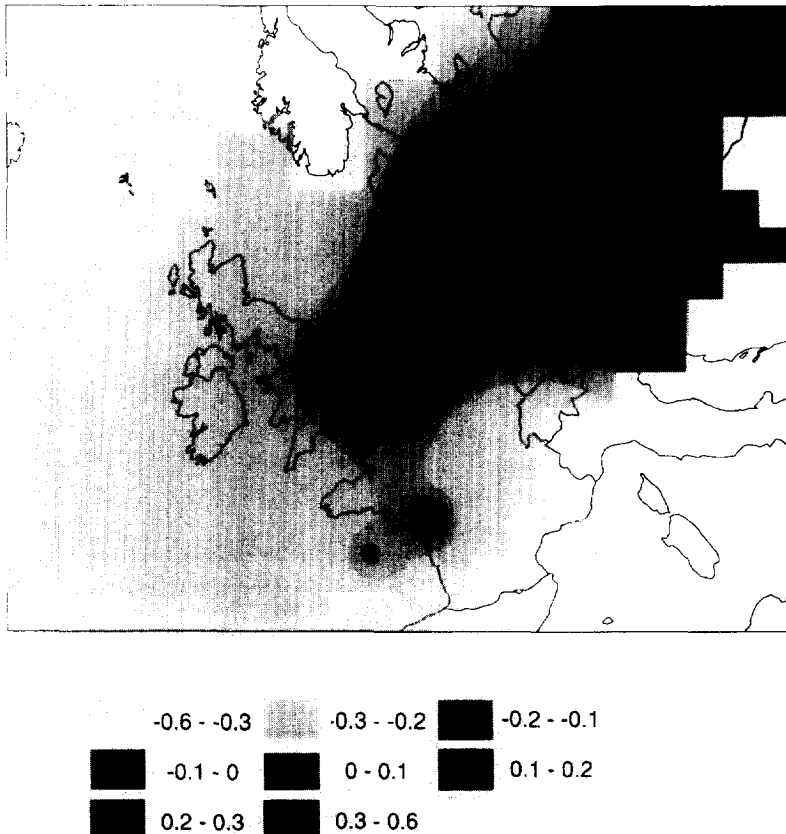


Fig. 2. Same as Fig. 1, but computed with the new method.

emissions and the trajectory statistics is not able to reproduce these small variations. In addition, the emissions in this area are largely of biogenic origin and vary considerably with space and time. The emission inventory was smoothed with two iterations of a 9-point operator to reduce the most high-frequency variations.

Table 2 shows the results for the Seibert *et al.* method and the one presented here. The last row in Table 2 gives the results for the combined statistics for all stations. In this case, only those grid cells crossed by more than 200 trajectories were used. First, it can be seen that, in general, r_s is higher with the redistribution method than with the Seibert *et al.* method (8 stations and the combined case). Only for three receptor sites did the Seibert *et al.* method perform slightly better. One of these (La Cartuja) has a very low r_s for both methods. This might indicate that this station is not representative for the investigation of long-range transport. For another one of the three stations (Ispra), the correlation coefficient is low. This is most probably due to the proximity of the Alps. In the Alps, trajectories have higher uncertainties than in other regions (see Stohl *et al.*, 1995) and precipitation events occur more frequently. In some cases, the redistribution method improved the results significantly, e.g. for

the stations Kosetice and Bilthoven and for the combined statistics.

4. CONCLUSIONS

A new method of trajectory statistics was developed and applied to the transport of particulate sulfate in Europe. Particulate sulfate was chosen for a test of this method, because for this species source regions as revealed by the trajectory statistics can be compared with the spatial distribution of sulfur emissions in the EMEP inventory. The agreement with the emission inventory was very satisfying, both qualitative and quantitative. Recently, Virkkula *et al.* (1995) applied the new method to identify source regions of the total sulfate, non-sea-salt sulfate, ammonium and sodium arriving at the Finnish station Sevetijärvi and could successfully identify the very different source regions of these species. Thus, trajectory statistics as presented here are a useful tool for the identification of source areas of air pollutants.

Trajectory statistics can be applied to single stations as well as to a combination of several stations. In the latter case, source regions can be identified over a larger spatial domain and with greater accuracy.

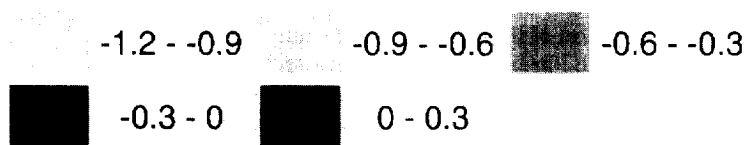
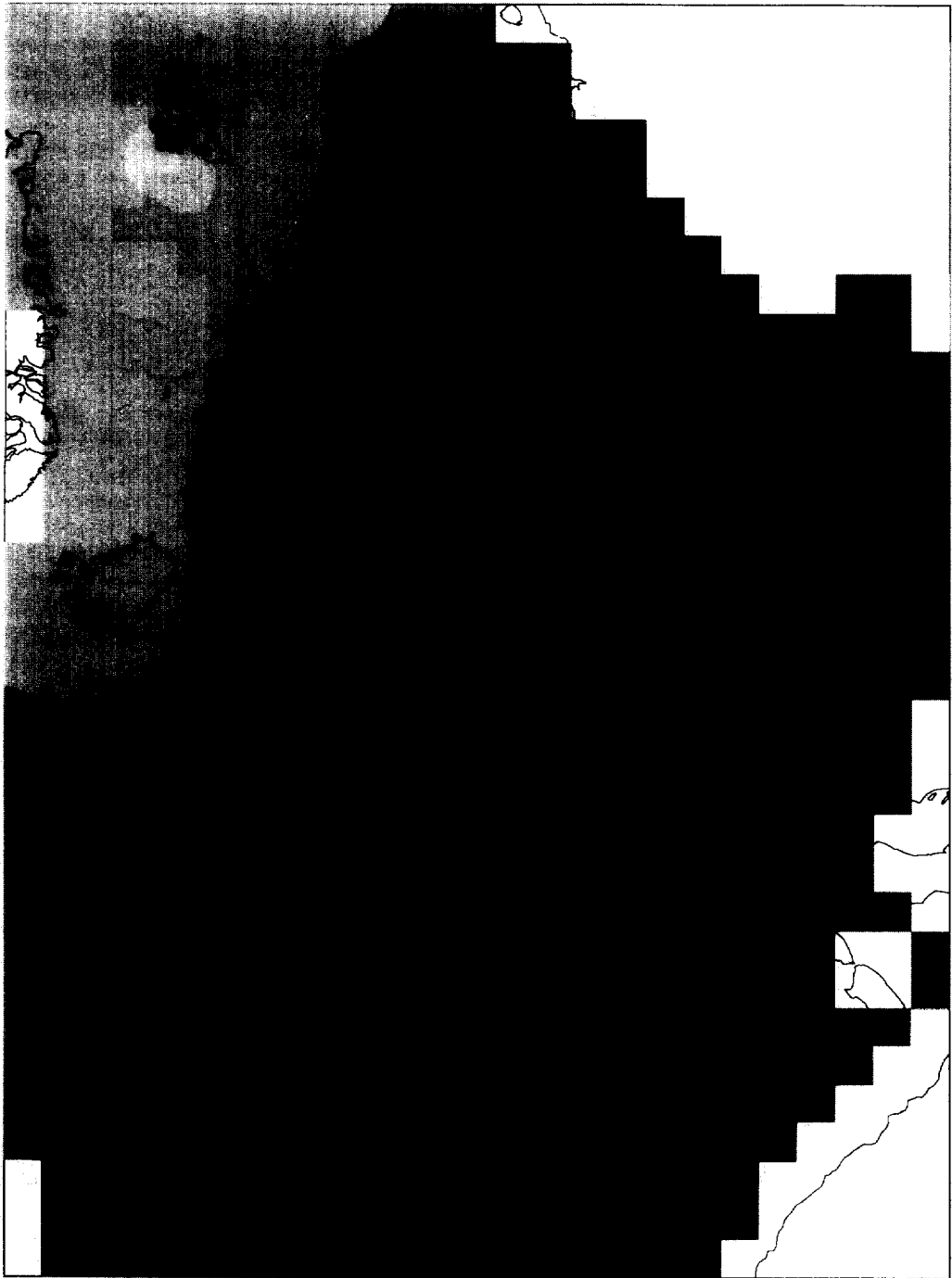


Fig. 3. Combined contour plot of the particulate sulfate concentration field $\overline{C_{mn}}$ as computed with the Seibert *et al.* method for all receptor sites (black squares). Grid cell size was $150 \times 150 \text{ km}^2$. Those cells which were crossed by less than 40 trajectories were left blank.

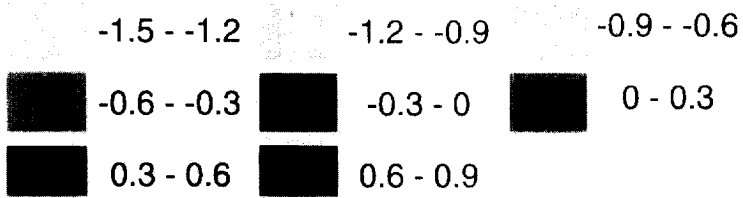


Fig. 4. Same as Fig. 3, but computed with the new method.

Table 2. Spearman rank-order correlation coefficient between the emission rates and the concentrations. The first column is the station name (last row: combined concentration field), the second column is the number of grid cells which were used to compute the correlation. Results are given separately for the Seibert *et al.* method ($r_s(1)$) and the redistribution method ($r_s(2)$)

Station	#	$r_s(1)$	$r_s(2)$
Bilthoven	198	0.68	0.84
Eskdalemuir	125	0.75	0.79
Illmitz	287	0.33	0.49
Ispra	196	0.55	0.50
Kosetice	290	0.56	0.84
La Cartuja	158	0.22	0.18
Preila	309	0.71	0.69
Utö	255	0.81	0.81
Sevettijärvi	185	0.52	0.55
Valentia	101	0.81	0.87
Yarner Wood	135	0.54	0.61
Witteveen	218	0.69	0.78
All stations combined	214	0.81	0.87

In the future, trajectory statistics could be used to validate emission inventories of species with very ill-known sources (e.g. heavy metals), if a sufficient number of sites with continuous ambient air concentration measurements of these substances exists. To improve emission inventories, several extensions of the method seem possible. It would be interesting to use an existing inventory as the first guess field and to improve this field with the redistribution method. Another approach would be to use results of an air quality simulation model and to trace back the bias between modelled and measured concentrations. With this method, areas that are problematic for the model could be revealed. Although these problems could have many reasons, a likely explanation could be an over- or underestimation of emissions.

The potential of trajectory statistics is not yet fully exploited. Several significant improvements seem possible. First, the statistics should account for the uncertainties of the trajectories (see, e.g., Stohl *et al.*, 1995). This could be done using ensembles of trajectories instead of single trajectories. Second, and even more important, the method should be extended to account for physical or simple chemical processes occurring along the trajectories. Of prime importance would be to allow for wet deposition along the trajectories.

Acknowledgements—I thank J. Schaugh for giving me access to the NILU database. A. Virkkula for providing the Sevettijärvi data, P. Seibert for a thorough review of the manuscript, H. Kromp-Kolb and G. Wotawa for numerous helpful discussions and all researchers who submitted their data to the NILU database.

REFERENCES

- Ashbaugh L. L. (1983) A statistical trajectory technique for determining air pollution source regions. *J. Air Pollut. Control. Ass.* **33**, 1096–1098.
- Ashbaugh L. L., Malm W. C. and Sadeh W. Z. (1985) A residence time probability analysis of sulfur concentrations at Grand Canyon National Park. *Atmospheric Environment* **19**, 1263–1270.
- Berg T. and Schaugh J. (eds.) (1994) EMEP Workshop on the Accuracy of Measurements. EMEP/CCC-Report 1/93, Oslo, Norway.
- Cheng M.-D., Hopke P. K. and Zeng Y. (1993) A receptor-oriented methodology for determining source regions of particulate sulfate at Dorset, Ontario. *J. geophys. Res.* **98**, 16,839–16,849.
- Colin J. L., Renard D., Lescoat V., Jaffrezo J. L., Gros M. J. and Strauss B. (1989) Relationship between rain and snow acidity and air mass trajectory in Eastern France. *Atmospheric Environment* **23**, 1487–1498.
- Dorling S. R., Davies T. D. and Pierce C. E. (1992a) Cluster analysis: a technique for estimating the synoptic meteorological controls on air and precipitation chemistry—method and applications. *Atmospheric Environment* **26A**, 2575–2581.
- Dorling S. R., Davies T. D. and Pierce C. E. (1992b) Cluster analysis: a technique for estimating the synoptic meteorological controls on air and precipitation chemistry—results from Eskdalemuir, south Scotland. *Atmospheric Environment* **26A**, 2583–2602.
- ECMWF (1989) *Research Manual 2. ECMWF Forecast Model—Adiabatic Part, 2nd Edn.* European Center for Medium Range Weather Forecasts, Reading.
- Harris J. M. and Kahl J. D. (1990) A descriptive atmospheric transport climatology for the Mauna Loa Observatory, using clustered trajectories. *J. geophys. Res.* **95**, 13,651–13,667.
- Henderson R. G. and Weingartner K. (1982) Trajectory analysis of MAP3S precipitation chemistry data at Ithaca, NY. *Atmospheric Environment* **16**, 1657–1665.
- Klug, Graziani G., Grippa G., Pierce D. and Tassone C. W. (eds.) (1992) *Evaluation of Long Range Atmospheric Transport Models Using Environmental Radioactivity Data from the Chernobyl Release*, 366 pp. Elsevier, Amsterdam.
- Miller J. M. (1981) A five-year climatology of back trajectories from the Mauna Loa observatory, Hawaii. *Atmospheric Environment* **15**, 1553–1558.
- Miller J. M., Moody J. L., Harris J. M. and Gaudry A. (1993) A 10-year trajectory flow climatology for Amsterdam island, 1980–1989. *Atmospheric Environment* **27A**, 1909–1916.
- Moody J. L. and Samson P. J. (1989) The influence of atmospheric transport on precipitation chemistry at two sites in the midwestern United States. *Atmospheric Environment* **23**, 2117–2132.
- Press W. H., Teukolsky S. A., Vetterling W. T. and Flannery B. P. (1992) *Numerical Recipes in FORTRAN. The Art of Scientific Computing*, 2nd Edn. Cambridge University Press, New York.
- Sandnes H. (1993) Calculated budgets for airborne acidifying components in Europe, 1985, 1987, 1988, 1989, 1990, 1991 and 1992. EMEP/MSC-W Report 1/93.
- Seibert P., Kromp-Kolb H., Baltensperger U., Jost D. T., Schwikowski M., Kasper A. and Puxbaum H. (1994) Trajectory analysis of aerosol measurements at high alpine sites. In *Transport and Transformation of Pollutants in the Troposphere* (edited by Borrell P.M., Borrell P., Cvitaš T. and Seiler W.), pp. 689–693. Academic Publishing, Den Haag.
- Stohl A. and Kromp-Kolb H. (1994b) Origin of ozone in Vienna and surroundings, Austria. *Atmospheric Environment* **28**, 1255–1266.
- Stohl A. and Scheifinger H. (1994) A weather pattern classification by trajectory clustering. *Meteorol. Z. N.F.* **6**, 333–336.
- Stohl A. and Wotawa G. (1995) A method for computing single trajectories representing boundary layer transport. *Atmospheric Environment* (in press).

- 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. *J. Appl. Met.* **34**, 2149–2165.
- Virkkula A., Mäkinen M., Hillamo R. and Pakkanen T. (1994) Inorganic ions in atmospheric aerosol in the Finnish Arctic. In 4th Int. Aerosol Conf. (edited by Flagan R.), pp. 418–419. Los Angeles, U.S.A.
- Virkkula A., Mäkinen M., Hillamo R. and Stohl A. (1995) Atmospheric aerosol in the Finnish Arctic: particle number concentrations, chemical characteristics, and source analysis. *Wat. Air Soil Pollut.* (accepted).
- Zeng Y. and Hopke P. K. (1989) A study of the sources of acid precipitation in Ontario, Canada. *Atmospheric Environment* **23**, 1499–1509.