A PDF approach to processing the data of tracer experiments for validation of dispersion models

Eugene Genikhovich, Elena Filatova
Main Geophysical Observatory, 7 Karbyshev St., 194021 St. Petersburg, Russia

Introduction

A "standard" approach to validation of dispersion models for a point source is based on comparison of the data of tracer experiments with model predictions. Such a comparison is usually made on "individual" basis; in other words, one compares individual measurements at certain meteorological conditions with model predictions corresponding to the same conditions. Centerline concentrations, *i.e.*, measured arcwise maxima are mainly used in comparisons to eliminate the strong dependence of concentrations on the crosswind distance from the receptor point to the plume axis. Unfortunately, these data are still heavily contaminated with the noise, mainly due to meandering of the plume centerline as well as to stochastic deviations of the crosswind distribution from the Gaussian one. The most advanced approach to the data analysis suggested by Irwin and Rosu (1998) and Irwin (1999) is aimed to reduce this noise by means of proper grouping and processing the initial data set in order to eliminate the noisiest individual data and estimate conditional averages ("mathematical expectancies") of concentrations corresponding to certain gradations of governing parameters. In particular, the centerline concentration is estimated using the average of all concentration values measured along the arc at the receptor points within $\pm S_y$ around the point of maximum; here, S_y is the computed lateral dispersion of concentrations along the arc.

It seems, however, that such a procedure cannot completely "filter" the noise related to meandering the plume because measured values of concentrations along the arc as well as S_y are influenced by the meandering (see, for example, Gifford, 1959). As a result, for example, centerline concentrations are underestimated and lateral dispersions S_y are overestimated. Besides, when estimating rank statistics (like, for example, the second or twenty-fifth highest value of concentration) or upper percentiles with the use of the average values of concentrations in the gradations, one could expect a systematic bias (underestimation) in these estimates. That is why a different approach to the data processing is suggested in this paper.

Basic formalism

Let us consider the arcwise maximum pf measured concentrations, C', as a representative of the "real" centerline concentration, C, influenced by the meandering of the plume. A stochastic model of meandering can be written as follows:

$$C' = C \cdot \exp[-0.5\frac{(y-a)^2}{\sigma^2}], \quad (1)$$

where y is a stochastic variable which characterizes the displacement of the receptor point relative to the average position of the plume centerline, a, and has the dispersion σ^2. After taking the logarithm from both sides of Eq. (1), it can be re-written as

$$\ln C' = \ln C - 0.5\eta^2, \quad (2)$$

where η is the stochastic variable with the mean value and standard deviation equal to 0 and 1, correspondingly. Eq. (2) indicates that ln C' is the sum of two stochastic variables, ln C and $-0.5\eta^2$, which are assumed being independent. In such a case, the probability density $f_{\ln C'}(x)$ of the stochastic variable ln C' is the convolution of probability densities $f_{\ln C}$ and $f_{0.5\eta^2}$ of variables ln C and $-0.5\eta^2$:

$$f_{\ln C'}(x) = \int_{-\infty}^{\infty} f_{\ln C}(t) f_{-0.5\eta^2}(x-t)dt. \quad (3)$$

Taking the Fourier transformation from (3) and resolving the equation obtained, one could derive the following expression:

$$F_{\ln C}(\omega) = \frac{1}{\sqrt{2\pi}} \cdot \frac{F_{\ln C'}(\omega)}{F_{-0.5\eta^2}(\omega)}, \quad (4)$$

where $F_{\ln C}$, $F_{\ln C'}$ and $F_{-0.5\eta^2}$ are characteristic functions, *i.e.*, Fourier transformations of corresponding probability densities. As soon as the numerator and denominator in the right-hand side of Eq. (4) are known, the probability density of "real" centerline concentrations can be found from the inverse Fourier transformation:

$$f_{\ln C}(x) = \frac{1}{2\pi} \cdot \int_{-\infty}^{\infty} e^{i\omega x} \cdot \frac{F_{\ln C'}(\omega)}{F_{-0.5\eta^2}(\omega)} d\omega. \quad (5)$$

After integration between -8 and x it yields also an expression for the probability distribution function (PDF) of logarithms of the centerline concentrations, $P_{\ln C}(x)$. Both, $f_{\ln C}$ and $P_{\ln C}(x)$ can be used to estimate mean values, standard deviations and upper percentiles of the centerline concentrations.

Is seems to be reasonable to assume that ? in Eq. (2) is distributed in accordance with the Gaussian distribution. It can be shown that in this case

$$f_{-0.5\eta^2}(x) = \frac{e^x}{\sqrt{-\pi x}}, \text{ if } x < 0; \quad (6a)$$

$$f_{-0.5\eta^2}(x) = 0, \text{ if } x \geq 0. \quad (6b)$$

The Fourier transformation of (6a,b) yields:

$$F_{-0.5\eta^2}(\omega) = \frac{1}{\sqrt{2\pi}(1+\omega)^{1/4}} e^{0.5i \cdot \tan^{-1}(\omega)}, \quad (7)$$

where $\tan^{-1}$(?) indicates the arc tangent of ? .
Substituting Eq. (7) into (5), one can see that the formalism introduced in this paper is equivalent to filtering the characteristic function $F_{\ln C'}$ with the "spectral window" H(?) = 1/ $F_{-0.5\eta^2}$ (?) and that in this specific case H(?) = (1+ ?)$^{1/4}$exp(-0.5i·tan^{-1}?). As for $F_{\ln C'}$(?), it can be calculated numercally as the Fourier transformation of the empirical probability distribution of logarithms of measured centerline concentrations. It is possible also to fit this distribution with an analytical expression and use its analytical Fourier transformation. In particular, one can assume that ln C' are distributed in accordance with the normal distribution with mean (logarithmic) value and standard deviation equal to a' and s, correspondingly. In this case,

$$F_{\ln C'}(\omega) = \frac{1}{\sqrt{2\pi}} \cdot e^{-ia'\omega - 0.5s^2\omega^2}. \quad (8)$$

The integral obtained after substituting Eq. (7) and (8) into (5) was calculated numerically.

An example of application

The formalism described in this paper was applied to the Kincaid data set (see Olesen, 1997). All non-zero arcwise maxima were processed independently on the quality flag attached to these data, and their logarithms were calculated. They were stratified into subsets in accordance with the distance of the arc from the source. Additionally, they were sorted in accordance with Table 1 into two groups corresponding to different meteorological conditions (here, U_{10} is the wind speed at 10 m, u_* is the friction velocity and L is the Monin-Obukhov length scale).

Table 1 Groups and corresponding meteorological parameters.

Group number	U_{10}, m/s	u_*, m/s	L, m
1	1.4 ÷ 2.7	0.19 ÷ 0.32	-2.5 ÷ -12
2	2.4 ÷ 6.0	0.28 ÷ 0.57	- 10 ÷ -100

The volume of the individual subset of data N was varying between 15 and 50. These data were sorted into variational series in order of their increase, and the values of PDFs corresponding to k-th member of the series were estimated as k/(N+1). An example of the probability plot corresponding to the distance of 10 km from the source and Group #1 is shown on Fig. 1. The straight line on the plot corresponds to the best fit and is used to derive the parameters of the lognormal distribution. Similar graphs are plotted for all other distances and groups and used to reconstruct the "filtered" distributions of centerline concentrations. As an illustration of the effects of such a filtering, the ratios of the standard deviations $s_C/ s_{C'}$ of the filtered to non-filtered samples (PDFs) and corresponding rations of the 95th percentiles, C_{95}/C'_{95}, are given in Table 2 for the Group 1.

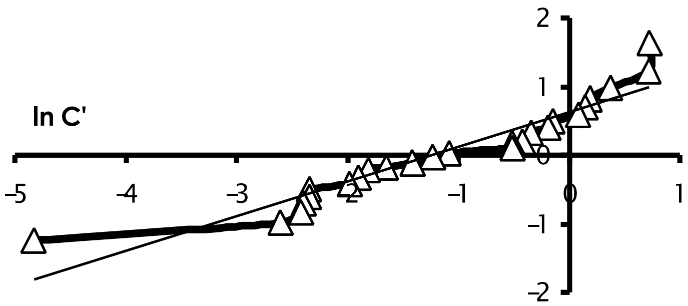

Figure 1 Probability plot corresponding to the 10 km distance from the source (straight line indicates the best-fit lognormal distribution).

Table 2 Ratios of filtered to non-filtered statistics.

Distance, km	0.5	1	2	3	5	7	10	15	20
$s_C/ s_{C'}$	0.89	0.91	0.89	0.89	0.83	0.83	0.90	0.86	0.86
C_{95}/C'_{95}	1.19	1.22	1.26	1.24	1.30	1.30	1.31	1.19	1.19

Discussion

The results presented in Table 2 show that the procedure of processing the data introduced in this paper works "in the right direction" decreasing the standard deviation and increasing the upper percentiles of the samples at hand. It is obvious that this procedure works faster if one would use an analytical approximation of the distribution of arcwise maxima corresponding to Eq. (8). It is possible, however, only if the lognormal distribution fits the empirical one reasonably well. It was found that the quality of such a fitting could be improved after removing from the sample from one to three lowest values of concentration (probably they were close to the sensitivity limit of the instruments used in the experiment). A longitudinal distribution of the average filtered centerline concentrations, <C> corresponding to Group 1 and constructed after removing one to three lowest values is shown on Fig. 2. It is compared there with the longitudinal distribution of the average non-filtered centerline concentrations, <C'>. Just as an illustration, the longitudinal distribution of the ground-level centerline concentrations, which are calculated with the use of the dispersion model, introduced by Genikhovich and Filatova (2001), is shown on this figure too.

When analyzing the basic assumptions introduced in this paper, one should notice that the normal distribution of ? is, strictly speaking, applicable only in the case where only one instrument is located on the arc and always placed "downwind" from the source. This assumption still works, if the receptor spacing is larger or comparable with the width of the plume. Far from the source, however, where the plume is "wide", the receptor spacing effect should be taking into account. Correcting the PDF for ? could do it, and corresponding results will be published later.

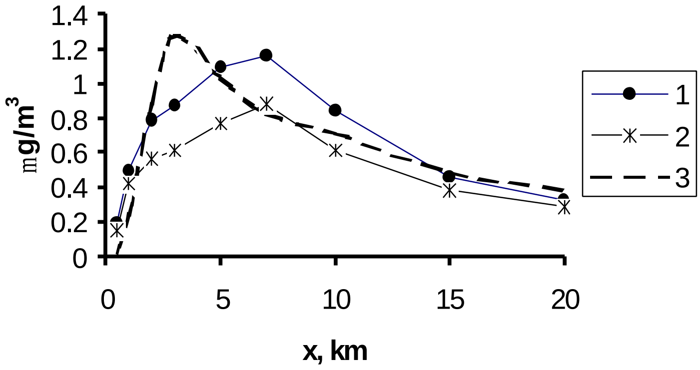

Figure 2 Longitudinal distribution of the ground-level centerline concentrations estimated with the proposed method (1), directly from the sample and with a dispersion model (3).

Conclusion

Initially, validation of the Gaussian-type dispersion models had been based on the direct comparison of measured and calculated concentrations paired in both, space and time. It was found soon, however, that errors in such a comparison were extremely high. The following development of the validation methods was aimed to avoid any direct comparison of concentration and switch to comparison of certain statistics. For majorant dispersion models, a corresponding method based on comparison with high percentiles of the PDF of concentrations was introduced in 1960th in the works of the Main Geophysical Observatory, St. Petersburg, Russia (for references see Genikhovich, 1998). For models, like Gaussian ones, predicting "actual" concentrations at given meteorological conditions a corresponding method based on comparison of averages was introduced by Irwin (1999) and Irwin and Rosu (1998).

In order to demonstrate the essence of the proposed technique we do not discuss in this paper the possibilities for combining it with Irwin's ideas. It is obvious, however, that such a combination could be constructed in rather a straightforward manner. As a result, a more sophisticated method based on the PDF approach and applicable to dispersion models predicting actual concentrations will be constructed. Even at the present state, however, the proposed technique allows to filter the meandering effects form the data of tracer experiments and, as a result, reduce their dispersion and increase their upper percentiles.

Acknowledgements

This work was partially supported by the Russian Foundation for Basic Research under the grant # 98-05 65606.

References

1. Irwin, J.S., Rosu, M.R. (1998) 'Comments on a Draft Practice for Statistical Evaluation of Atmospheric Dispersion Models', Proceedings of 10th Joint Conf. Applications of Air Pollut. Meteorology, Amer. Met. Soc., Boston, USA, pp. 6 – 10.
2. Irwin, J.S. (1999) 'Effects of Concentration Fluctuations on Statistical Evaluations of Centerline Concentrations Estimated by Atmospheric Dispersion Models', Proceedings of 6th Conf. Harmonization Atmospheric Dispersion Models, p. 38 - 47 (to be published in Int. J. Environment. Pollution)
3. Gifford, F. (1959) 'Statistical plume model'. In: Atmospheric Diffusion and Air Pollution (Ed. F.N. Frenkiel, P.A. Sheppard), Acad. Press, NY, pp. 143 – 164 (translated into Russian in 1962).
4. Feller, W. (1966) 'An Introduction to Probability Theory and Its Applications', J. Wiley, NY, 752 p. (translated into Russian in 1967)
5. Olesen, H.R. (1997) 'Data Sets and Protocol for Model Validation', Int. J. Environment. Pollution, vol. 5, Nos. 4-6, p. 693-701
6. Genikhovich, E.L., Filatova E.N. (2001) 'A unified multiple-source dispersion model', Trudy GGO (Proc. Main. Geophys. Observatory) (in press)
7. Genikhovich, E.L. (1998) 'Russian Regulatory Diffusion Models: Status, Results of Validation and International Intercomparisons', In: Air Pollution in the Ural Mountains. Environmental, Health and Policy Aspects (Ed. I. Linkov & R. Wilson), Kluwer Academic Publishers, Dorderecht, p. 75 – 80.