INDICATORS OF PERFORMANCE OF DISPERSION MODELS AND THEIR REFERENCE VALUES

Eugene Genikhovich
Main Geophysical Observatory, St. Petersburg, Russia.

INTRODUCTION

Modern approaches to validation of dispersion models upon experimental data are usually based on calculation of different indicators of performance of the models (see for example Ohlesen, 1997). The aforementioned indicators can be considered as certain metrics invented for evaluating the measure of closeness of model predictions and measurements. The procedure of validation starts with generation of the model set of data called "predictions", which mimics a set of data of measurements called "measurements". Then one can use any kind of the concurrent analysis to evaluate the measure of closeness of these two sets or, in other words, to estimate the performance of the model in question.

The indicators of performance/closeness most often used in validation of dispersion models are defined by the following expression:

$$\mathrm{FBM} = \frac{< M > - < P >}{< M > + < P >}; \tag{1}$$

$$\mathrm{MFB} = < \frac{M - P}{M + P} >; \tag{2}$$

$$\mathrm{FAa} = \mathrm{Prob}\{\frac{P}{a} < M < aP\}; \tag{3}$$

$$\mathrm{NMSE} = \frac{< (M - P)^2 >}{< P > \cdot < M >}; \tag{4}$$

$$\mathrm{Corr} = \frac{< (P - < P >)(M - < M >) >}{\sqrt{< (P - < P >)^2 > \cdot < (M - < M >)^2 >}}. \tag{5}$$

Here FBM is the fractional bias of means values, MFB is the mean value of fractional biases, FAa characterizes an agreement within a factor of "a", NMSE is the normalized mean square error, Corr is the coefficient of linear correlation, brackets indicate the procedure of averaging, Prob means the probability of the event indicated in parentheses, M and P are measured and predicted concentrations, respectively.

When validating the dispersion model and estimating the indicators (1) – (7), one is interested in comparison their values with reference ones corresponding to a "good model". The following values of these indicators are expected in the case, when the data sets P and M are identical: FBM = MFB = NMSE = 0; Corr = 1; FAa = 1 (for a > 1). Such a case, however, is completely fictitious, and it is hard to expect that these values are applicable in any realistic situation. The goal of this paper is to present a more reliable set of reference values of these indicators and to analyze their efficiency. It will be done in connection with a problem of validation of a dispersion model, which formally predicts "actual" (corresponding to given meteorological conditions) centerline concentrations from a point source. Corresponding methods have been discussed in numerous publications including Irwin (1999), Irwin and Rosu (1998) and others.

"PERFECT MODELS" AND A METHODOLOGY OF THEIR VALIDATION

A dispersion model is represented by the following expression:

$$c = f(x_i, \omega_j, t_k), \tag{6}$$

where c means concentration of the pollutant ("tracer"), x_i are coordinates of the receptor point relative to the source, ω_j are governing meteorological parameters, and t_k are governing "technical" parameters of the source (like the emission rate, volume rate, stack height and diameter, effluent temperature and so on). In Eq. (6), f means a general functional dependence and could be a differential equation or even a physical instrument installed in the wind tunnel. For the sake of simplicity, however, it is assumed further on that f is an analytical expression. In fact Eq. (6) is usually constructed as a combination of several sub-models (e.g. transport and dispersion, plume rise and so on); all these models are assumed here to be deterministic.

In the process of validation, concentrations predicted with Eq. (6) should be compared with those measured in experiments. When validating this model, a whole interval of variations of the governing parameter ω_j, for example, should be divided (stratified) into a set of sub-intervals or gradations $\Delta_n\omega_j$ (a notation Δ_{nj} is also used in this paper). The width of these subintervals should satisfy to the following conditions:

$$|\alpha_j \Delta_{nj}| << 1, \tag{7}$$

where $\alpha_j = \partial(\ln f)/\partial\omega_j$. Correspondingly, the sample of the measured concentrations is stratified into a set of sub-samples corresponding to given subintervals of governing parameters and given distances from the source to the receptor points. Each combination of sub-intervals Δ_{nj} and distance from the source defines a certain "regime" (using terminology suggested by J. Irwin). It is assumed in this paper that the volume of each of sub-samples is large enough to make possible reliable statistical estimates.

Statistical properties of concentrations in the plume at given meteorological conditions were theoretically studied by Gifford (1959). He indicated, in particular, that centerline concentration is a stochastic variable. Using certain physical assumptions for short-term concentrations, valid mainly on average, Gifford derived an analytical expression for the probability density of centerline concentrations. Empirical PDFs of the centerline concentrations were studied by Irwin and Lee (1997). Processing the Kincaid data set (see Ohlesen, 1997), Genikhovich and Filatova (2001) found that, having stratified the sample of measure concentration accordingly to the distances from the source and meteorological conditions and having removed several outliers corresponding the lowest measured values, one can approximate PDFs of the centerline concentrations with the log-normal distribution which corresponds to the following probability density:

$$p(c) = \frac{1}{sc\sqrt{2\pi}} \exp\left(-\frac{\ln^2(c/m)}{2s^2}\right), \tag{8}$$

where ln (m) is the mean value (mathematical expectation) of logarithms of concentrations (in other words, m is the geometric mean of concentrations), and s is the standard deviation of these

logarithms of concentrations ("logarithmic standard deviation"). Such an approximation will be used in this paper too, but the results obtained can be easily reformulated for other PDFs.

Let us call the model "perfect", if, for each of sub-intervals of governing parameters, it generates a value of concentration, which is exactly equal to a certain statistical characteristics of the PDF of concentrations corresponding to these sub-intervals. It is assumed also in addition that perfect models "perfectly account" for the influence of governing technical parameters, which will be further omitted from Eq. (6).

This paper will be focused only on the cases when these statistical characteristics are either mean values of concentration in gradations or certain percentiles of their PDF. Using Eq. (8) one can easily derive the following expressions for these characteristics:

$$< c >= m \cdot e^{0.5s^2}; \tag{9}$$

$$\sigma_c^2 = m^2 e^{s^2} (e^{s^2} - 1); \tag{10}$$

$$c_p = m \cdot e^{Z_{1-2p} s\sqrt{2}}, \tag{11}$$

where <c> is the mean value of concentrations, σ_c^2 is their standard deviation, c_p is the p-th fractile of the PDF, Z_{1-2p} is a solution of the equation $erf(Z_{1-2p}) = 1-2p$, and erf is the error function.

Mathematical expectations (mean values) of the indicators of performance introduced by Eq. (1) – (3) can be directly calculated using Eq. (8). In particular,

$$FBM = \frac{e^{0.5s^2} - P/m}{e^{0.5s^2} + P/m}; \tag{12}$$

$$MFB = 1 - \frac{2}{\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{e^{-t^2}}{1 + \frac{m}{P} e^{ts\sqrt{2}}} dt; \tag{13}$$

$$FAa = 0.5[erf(\frac{\ln \frac{aP}{m}}{s\sqrt{2}}) - erf(\frac{\ln \frac{P}{am}}{s\sqrt{2}})], \tag{14}$$

where $erf(z) = 2\int_0^z \exp(-t^2) dt / \sqrt{\pi}$.

Assuming that the model predicts *constantly* the same value of concentration for the whole gradation, one can easily obtain that Corr = 0. In this case, substituting (8) into Eq. (4), one can obtain the following expression for NMSE:

$$NMSE = \frac{e^{s^2}(e^{s^2} - 1) + (e^{0.5s^2} - P/m)^2}{e^{0.5s^2} \cdot P/m}. \tag{15}$$

It should be noted, however, that in many validation exercises model predictions, M, are calculated with the use of the actual meteorological data accompanying the measured concentrations, P. It is obvious that indicators (1) – (3) should not be sensitive to this variation of the validation procedure as soon as conditions (7) are satisfied. It is not correct, however, when speaking about other indicators of performance. Corresponding effects on NMSE and Corr will be discussed further.

PERFORMANCE OF PERFECT MODELS

Accordingly to the aforementioned definition, if the model given by Eq. (6) is perfect, its prediction, P, is described either by Eq. (9) or by Eq. (11). Substituting these expressions into (12) – (15) one can evaluate the indicators of performance as functions of the only parameter "s" which characterize the variability of the measured centerline concentrations. Corresponding results are presented in Fig. 1 – 4. The curves on these figures correspond to the values of input parameters listed in Table 1. The integral in the right-hand side of Eq. (13) was calculated numerically.

Table 1. Input parameters used in computations

Curve No	Modeled characteristics	p	Z_{2p-1}
0	Mean value	-	-
1	P - fractile	0.50	0
2	P - fractile	0.75	0.4769
3	P - fractile	0.90	0.9062
4	P - fractile	0.95	1.1631
5	P - fractile	0.98	1.6450
6	P - fractile	0.99	1.8214

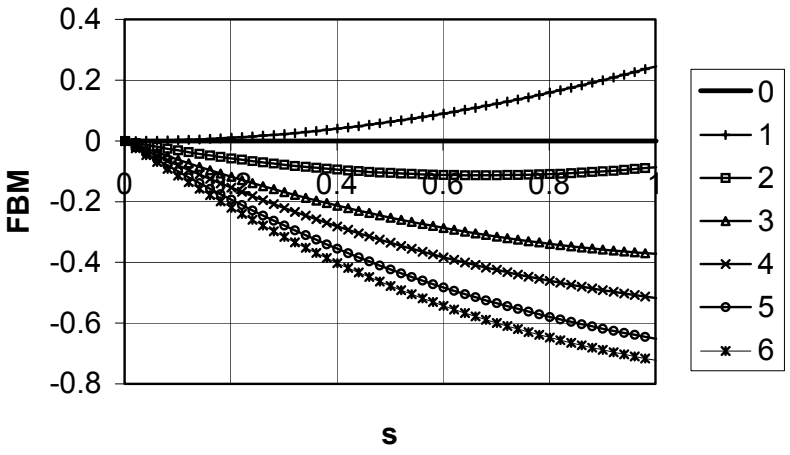

Figure 1. FBM as a function of s (see notations in Table 1)

When analyzing the results presented in Figure 1, it is obvious that FBM is equal to zero independently on s, if P is defined from Eq. (9). More informative is the fact that p = 0.75 gives

very small FBM values too and that FBM is "reasonably small", in a sense, when P matches fractiles corresponding to $0.5 \le p \le 0.9$.It is seen from Figure 2, however, that MBF, is equal to zero independently on s, if P matches the 50th percentile of the PDF of concentrations (in other words, if P is a geometric mean value of concentrations). The values of FBN seem to be acceptable, if P matches the same interval of fractiles or the mean value of concentrations.

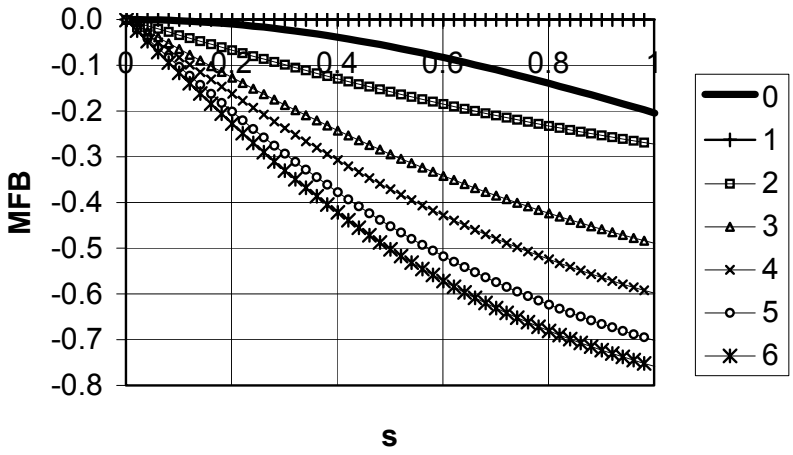

Figure 2. MFB as a function of s (see notations in Table 1)

The values of FAa calculated with a = 2 are presented in Figue 3. One can see here that P corresponding to the arithmetic mean value and to fractiles from p = 0.50 to 0.75 are again characterized with close values of this indicator. However, the model performs better, especially for large s, if P matches the geometric mean concentrations. As for NMSE presented on Figure 4, it demonstrates a non-monotonic behavior: errors corresponding to p = 0.75 are the smallest ones, if the values of s are small; if, however, s > 0.65, the NMSE indicator gives better model performance with P matching the 90th percentile.

DISCUSSION

If model predictions are calculated with the use of the actual meteorological data accompanying the measured concentrations, indicators NMSE and Corr should be determined differently. It is obvious that $<(M-P)^2>$ could be presented in the following form:

$$< (M - P)^2 > = \sigma_M^2 - 2\sigma_M \sigma_P \mathrm{Corr} + \sigma_P^2 + (< M > - < P >)^2 , \qquad (16)$$

where σ_M and σ_P are standard deviations of measured and predicted concentrations. As it is seen from Eq. (16), NMSE could be expressed in the general case via Corr and other first and second moments of measured and predicted concentrations. They could be evaluated using the Taylor expansion of Eq. (6):

$$c = c_0 \cdot [1 + \sum_{(j)} \alpha_j (\omega_j - \omega_j^0)], \qquad (17)$$

where the point ω^0_j is located inside Δ_{nj} and $c_0 = f(x_i, \omega^0_j)$. It should be noted here that c_0 is a non-stochastic value. After averaging, Eq. (17) yields and expression for the mean value of the calculated concentration, $<c> = <P>$. It follows from this expression, in particular, that $c_0 = <c>$, if

$$\omega^0_j = <\omega_j>. \tag{18}$$

This condition is assumed to be satisfied in the following considerations.

Using Eq. (17) and (18) one can derive the following formulas:

$$\sigma^2_P = <P>^2 \cdot \sum_{(j,k)} \alpha_j \alpha_k \sigma_j \sigma_k R_{jk}\,; \tag{19}$$

$$\sigma_M \sigma_P \mathrm{Corr} = <P> \cdot \sum_{(j)} \alpha_j \sigma_M \sigma_j \rho_j\,, \tag{20}$$

where σ_j is the standard deviation of the meteorological parameter ω_j, R_{jk} is the coefficient of linear correlations between ω_j and ω_k (certainly $R_{ii} = 1$), and ρ_j is the coefficient of linear correlations between ω_j and measured concentrations. After substituting Eq. (19) and (20) into (16), one can see that NMSE and Corr reflect properties of the atmosphere rather than those of the model. It could be demonstrated on the standard Gaussian model. When the stability category is given, concentrations here depend only on one governing parameter that is the wind speed, U. It follows from (19) and (20) here that $\mathrm{Corr} = \pm\rho_{UC}$ where ρ_{UC} is the coefficient of correlation between measured concentrations and wind speeds.

The results presented on Figures 1 to 4 can be used in practical applications, if the value of s is found. It could be determined, for example, either by fitting the empirical PDF with the lognormal distribution or by calculating the second central moment of logarithms of measured concentrations. If the coefficient of variation of concentrations (the ratio of the standard deviation to the mean value), V, is known, one can find s using the relationship $V^2 = \exp(s^2) - 1$. Genikhovich and Filatova (2001) found that Kincaid data correspond to s varying in the range from 0.6 to 1.2. The procedure of filtration of PDF suggested by these authors could reduce s as much as in 20%. If such a procedure is not applied, however, the values of indicators of performance corresponding to s = 1 could be roughly considered as reference values when validating dispersion models.

It should be noted that Eq. (12) – (15) correspond to mathematical expectations of the indicators considered. In fact, however, they are estimated from samples of limited volume, and, as a result, one could expect a certain statistical scatter of these estimates around corresponding mathematical expectations. This scatter could be characterized by dispersions of indicators (1) – (4), which, in turn, can be easily estimated as second central moments using the same probability density (8). Finally, it should be pointed out that the aforementioned indicators are not self-consistent because their optimal values correspond to tuning the model to different statistical characteristics of the measured data set.

CONCLUSION

The aforementioned results indicate that, due to the inherent high variability of data of measurements, the indicators of performance of dispersion models predicting "actual" (i.e. individual) values of concentrations cannot be better than their reference To put it bluntly, they

indicate that individual values of axial concentrations, due to their turbulent nature, can be estimated only with a rather large error even if the deterministic model in use is perfect. This error is of order of 100% for the data sets used in the evaluation practice. If the errors in the input data are taken into account, the resulting errors in concentrations could be even large. Thus, the data sets appear to be very noisy when compared with model predictions.

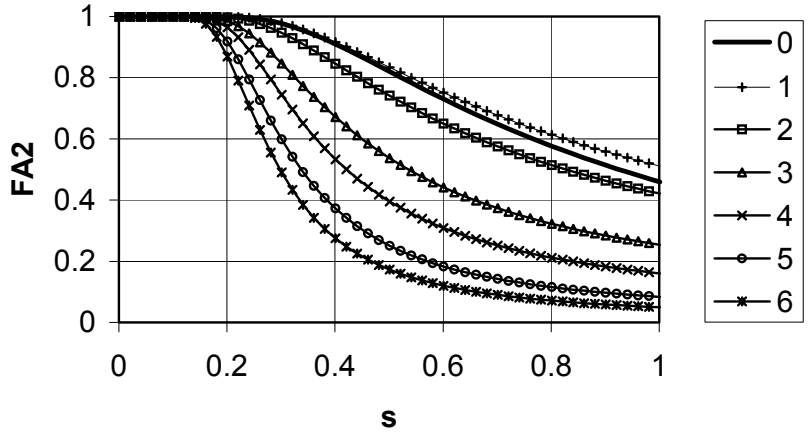

Figure 3. FA2 as a function of s (see notations in Table 1)

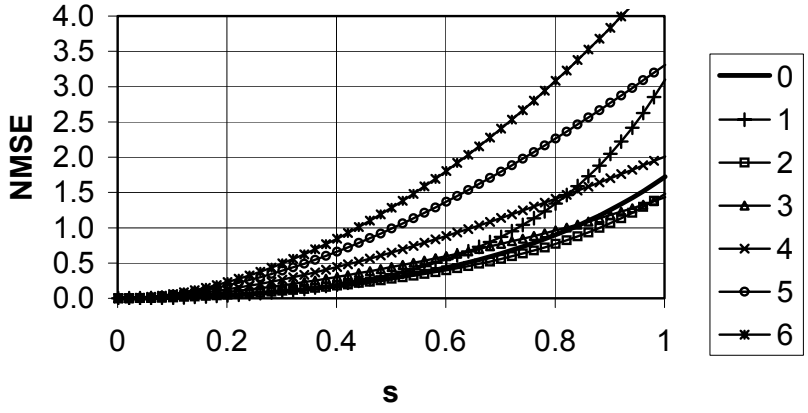

Figure 4. NMSE as a function of s (see notations in Table 1)

In such a situation only those functional dependencies included in Eq. (6) could be validated which result in significant variations of concentrations. This statement can be reversed in the following way: the values of concentrations normalized with corresponding emission rates, which are included in the data set used in validation purposes, should vary significantly (roughly speaking, the whole spectrum of regimes should cover several orders of magnitude of concentrations). When validating the probability distribution functions, the level of the noise in the measurements could be reduced, if the volume of the sample "corresponding to the evaluated

percentile" is big enough. It is not the case, however, for high percentiles, which obviously correspond to rare events. Here, one can find beneficial to work with dispersion models predicting directly upper percentiles. This type of models was discussed in this paper too, but it is obvious that they are "skewed" in the average sense and should be characterized using different indicators of performance. Additional discussion of this problem can be found, for example, in Genikhovich, Schiermeier (1995).

ACKNOWLEDGEMENTS

The author wishes to acknowledge Dr John Irwin for his friendly comments and helpful discussion.

REFERENCES

Genikhovich, E., Filatova, E., (2001) A PDF approach to processing the data of tracer experiments for validation of dispersion models. Proc. 7th Int. Conf. On Harmonization Within Atmospheric Dispersion Modelling for Regulatory Purposes, Belgirate, Italy. EC JRC, p. 33 – 36

Genikhovich E.L. & Schiermeier F.A., 1995 Comparison of United States and Russian complex terrain diffusion models developed for regulatory applications. *Atmospheric Environment*, v.29, No. 17, 2375 - 2385

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).

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*)

Irwin, J.S., Lee, R., (1997) 'Comparative evaluation of two air quality models: within-regime evaluation statistic'. *Int. J. Environment. Pollution*, vol. 8, Nos 3-6, pp. 346-355

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.

Olesen, H.R., (1997) 'Data sets and protocol for model validation', *Int. J. Environment. Pollution*, vol. 5, Nos. 4-6, p. 693-701