

# Diffuse battery application to atmospheric monitoring: the method of estimating the data errors

S.I. Eremenko

*Institute of Chemical Kinetics and Combustion,  
Siberian Branch of the Russian Academy of Sciences, Novosibirsk*

Received June 7, 1999

The method of data errors estimation intended for use in continuous monitoring of atmospheric aerosol with a diffusion battery is presented. Such estimation is necessary for reliable data inversion to size spectra. The technique is based on statistical analysis of particle concentration fluctuations. Data were obtained as continuous series of battery scans. For each scan the optimum number of the neighboring scans to be averaged was determined. Such a sample provides for maximum gain in the accuracy owing to data averaging provided that the errors have been determined quite reliably. The averaged data together with the error estimates were used for further data inversion.

## Introduction

Calculation of the aerosol size spectrum from the data obtained by means of a diffuse battery is the ill-posed inverse problem that has no a unique solution. Different methods for solving such problems suggest different criteria for selection of the solution satisfying the initial data from the set of possible solutions.

Earlier<sup>1</sup> we have suggested the method for solving this problem lying in averaging of the set of solutions. The method is based on the supposition that the accuracy of the initial data is known and the solution is obtained as a result of averaging over a sample from the set of solutions satisfying the initial data. The boundaries of this set are determined by the accuracy of the data used. The method has demonstrated its advantages in the laboratory tests, when the stable generator<sup>2</sup> was used as an aerosol source. The imitation calculations and experiments with bimodal aerosol showed that the optimal resolution is reached at the signal-to-noise ratio of 50, two monodisperse peaks are separated if the ratio of the corresponding particle diameters is 2.5 and greater. The further improvement of the accuracy of the data weakly affect the resolution.

The diffuse method for determining the size spectrum in the range from 3 to 200 nm shows good promise for investigating the atmospheric aerosols, for which the low resolution is not a significant limitation. However, the experience of some years of measurements has shown that the specific approach to estimation of the data is necessary in this case, without which one can hardly succeed in obtaining reliable size spectra.

High sensitivity of a solution of the inverse problem of the aforementioned type to estimation of the errors in the initial data was discussed many times (see, for example, Refs. 3 and 4). Underestimation of the accuracy of data leads to instability of the solution and appearance of false peaks in the size spectrum. Overestimation of the accuracy of data leads to the loss of the resolution that can not be justified, i.e., the obtained spectrum is smoothed and the peaks are widened.

Thus, the reliable estimate of the accuracy of the initial data is necessary for correct solving of the problem. This means that for obtaining the reliable particle size spectra it is necessary

not only to measure the particle number density in each channel of the diffuse battery, but also to correctly estimate the accuracy of these measurements. Obviously, to interpret the data, widening of the size spectrum is more preferable than appearance of artifacts, so one should be especially careful in avoiding underestimation of the measurement accuracy. In practice, one often ignores this fact, that evidently leads to the widespread opinion that the particle size spectra obtained by means of the diffuse battery are not reliable.

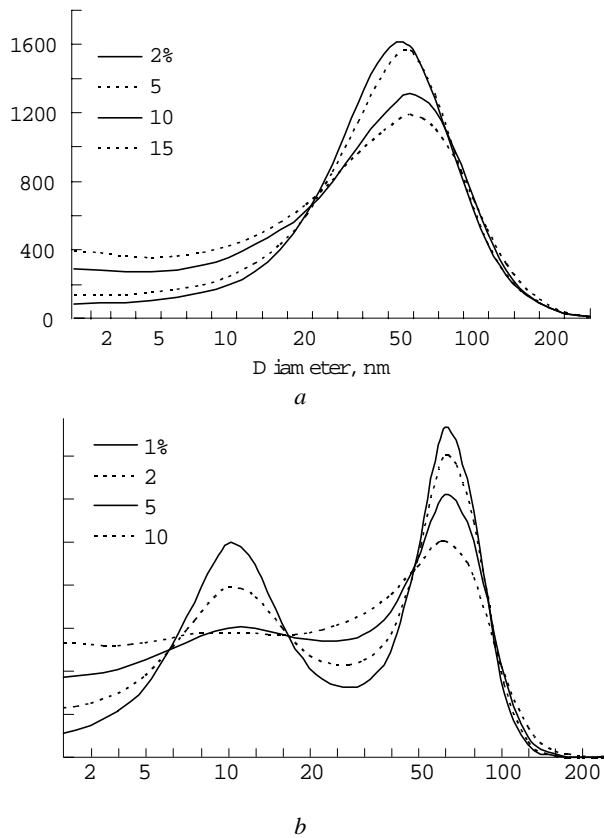
The purpose of this paper is to develop a practical technique for collecting and processing the data of diffuse battery, which will allow to obtain the reliable size spectra both in laboratory and atmospheric measurements.

## The influence of accuracy of the initial data and the error in its estimation on the reliability of the size spectrum

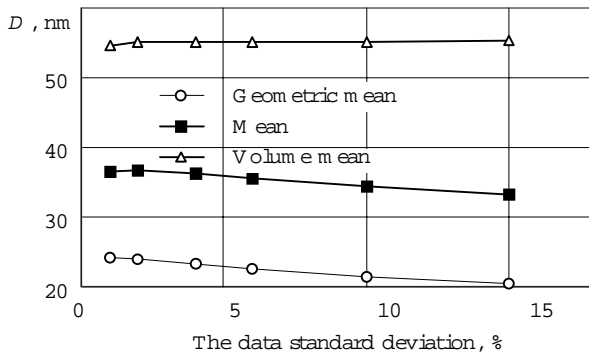
To assess the effect of the error in estimating the accuracy of data on the reliability of the solution, we have carried out a series of the following numerical experiments.

We have first preset one or another particle size spectrum. For this spectrum we generated the set of vectors with different realizations of random noise defined by the value of the relative rms error. The size spectrum was reconstructed by means of the MSA algorithm that assigns different values of the rms error in data, thereby simulating the error in the data accuracy.

Results of the numerical simulation have shown the following. When increasing the relative error of data from 1 to approximately 10%, the reconstructed size spectrum is widened but keeps its qualitative shape (Fig. 1). Such parameters as the mean, geometric mean, and root-mean-cube diameters change insignificantly (Fig. 2). The false "shoulder" appears at the left-hand side of the size spectrum at the relative error of 14% and starting from this point the reconstructed spectrum loses its similarity with the initial one.



**Fig. 1.** The mean solutions (particle size spectra reconstructed from the simulated readouts of the diffuse battery) as functions of the relative standard deviation of the data.



**Fig. 2.** Integral characteristics of the particle size spectra as functions of the relative standard deviation of the data. The size spectra shown in Fig. 1b are used.

Underestimating the accuracy when reconstructing the spectrum leads to its widening. Overestimating the accuracy does not significantly affect the size spectrum unless the preset error in data used in the solution is below 70% of the true one. Starting from this point, the reconstructed spectra noticeably lose the stability, and the false structures appear.

Thus, the estimates of the maximum permissible relative error in data (10–12%) and the maximum permissible error in the standard deviation of the data (30%) were obtained in the numerical experiments.

## Techniques for estimating the data variability

The data array obtained from the single scans of the battery does not make it possible to estimate the variation of data and so practically is not suitable for reconstruction of the size spectrum.

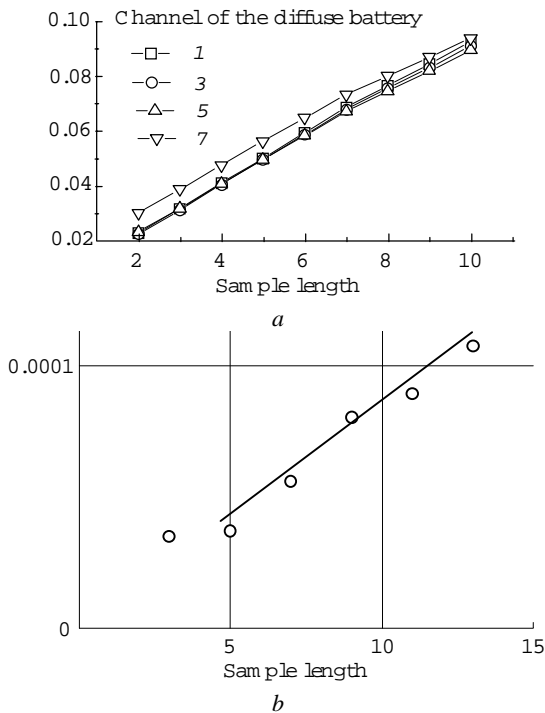
Sometimes the following expedient is used when analyzing single data array. One finds a solution that minimizes the Euclid norm of the vector of errors, and then use the obtained value as the estimate of the data variance. Such an estimate is permissible only in the case if the number of degrees of freedom for the data is significantly greater than that of the solution. Otherwise, such an estimate, as a rule, significantly underestimates the error of data that leads to the instability of the solution.

The most reliable estimate can be obtained if analyzing the series of  $N$  subsequently measured data vectors. One can think that if one selects quite big  $N$ , an arbitrarily accurate estimate of the variation of data can be obtained. However it is correct only in the case when the errors of the subsequent data vectors are independent and the whole sample is related to the same general totality. It is possible if the fluctuations of the number density of particles at the inlet of the device form a stationary random process, the autocorrelation time of which is less than the time of sampling of the data array.

One can suggest to use the dependence of the sample variance on the length of the series as a criterion of the fact that the sample satisfies these requirements. In the case of independent errors the variance does not depend on the length of the series. The reliable increase in the sample variance as the series length increases is an evidence of the significant contribution of the low-frequency fluctuations of the number density.

The dependences of the sample variance on the length of the series are shown in Fig. 3 for the atmospheric aerosol of suburban area and for the aerosol produced by the laboratory generator. It is seen that although the scale of variation of the data differs by an order of magnitude, the variance in both cases quickly increases as the length of the series increases. Practically this means that one can use no more that 2–3 neighbor measurements for estimating the variations for the atmospheric data presented.

This fact rises the problem, because the sample of 2–3 measurements does not make it possible to estimate the variance with a satisfactory accuracy. However, it is seen in Fig. 3 that the relative variation of the data is practically the same for all channels of the battery. The noticeably large values in the 7th channel are caused mainly by the fact that here the contribution of the statistical Poisson noise is noticeable. The latter is determined by the number of readouts during the exposure and can be easily estimated. The relative variation of the data shown in Fig. 3a minus the contribution of the Poisson noise is the same for all channels accurate to about 20%. Comparison of many data obtained in atmospheric measurements during last years shows that such a situation is typical for them.



**Fig. 3.** Relative variance of the data of the diffuse battery as a function of the sample length. Atmospheric measurements (a) and the laboratory generator (b).

If one accepts the hypothesis of the constancy of the relative variation of the number density in all channels of the battery, the sample of 2–3 neighbor measurements has enough degrees of freedom for estimating the variance accurate to 10–20% that is quite sufficient for our purposes. Thus, one can propose the following technique to be used for obtaining and processing the data of the diffuse battery.

### Technique for obtaining and processing the data

The best strategy of measurements is to make the single scanning duration as small as possible and to collect the series as long as possible. Practically, the scanning time is limited by the time in which the number density settles in the measurement volume of the counter after switching the channel (usually it is a few seconds). It is reasonable to set the time of exposure in each channel to be of the same order or 2–3 times longer. Finally we have a sequence of the data arrays  $\mathbf{b}_n$ . Let us suppose for simplicity that the contribution of the Poisson noise is negligible as compared with the number density fluctuations. Unless it is realized, it is easy to make all necessary corrections which we omit for simplicity of the description.

Let us choose the  $\mathbf{b}_n$  quite far from the beginning and the end of the series in order not to consider the boundary effects. Let us estimate the variance for the sample of  $2k + 1$  data arrays symmetrical relative to  $\mathbf{b}_n$ . Let us first take  $k = 1$  and estimate the relative standard deviation as follows.

Let us find the sample mean values for each  $j$ th channel of the battery

$$\bar{b}_j(k) = \sum_{i=n-k}^{n+k} \frac{b_{ij}}{2k + 1}$$

and estimate the relative sample variance

$$D_k = \frac{1}{m} \sum_{j=1}^m D_{kj} = \frac{1}{m} \sum_{j=1}^m \frac{1}{2k} \sum_{i=n-k}^{n+k} \frac{(b_{ij} - \bar{b}_j)^2}{(\bar{b}_j)^2},$$

where  $m$  is the number of the battery channels. The variance of the estimate  $D_{kj}$  is  $D_{kj}^2/k$  and, hence, the relative error of this series is

$$R(D_k) = \frac{1}{\sqrt{km(m-1)}}.$$

For example, it is approximately 13% at  $m = 8$  and  $k = 1$ .

Then let us increase  $k$  by a unit and estimate  $D_k$  again. Let us repeat this operation until the obtained estimate of the relative standard deviation becomes significantly greater than  $D_1$ , let it be the same 30%, which we assumed to be the maximum permissible error. If it has occurred at  $p$ th step, we accept the limit of the width of the sample of  $k = p - 1$ . Let us consider  $\bar{b}_j(k)$  as initial data for solving the inverse problem and estimate their standard deviation as

$$s_j = \bar{b}_j(k) \sqrt{D_k/[2k(2k + 1)]}.$$

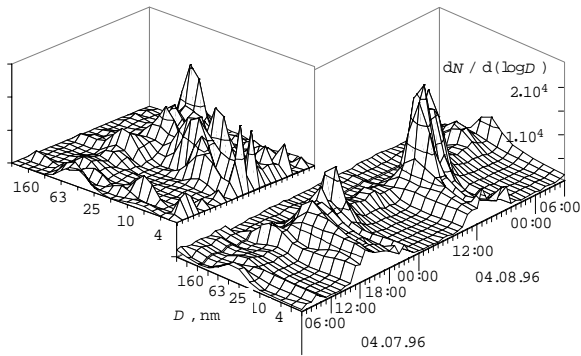
This case deserves special attention when the estimate of variance increases significantly greater than by the permissible value even at the first step. In this case one can use the result of one measurement without averaging as the initial data for the inversion, and the estimate from two neighbors (or one neighbor if the measurement is the first in the series) as the estimate of its accuracy. Such an approach is justified for the following reasons.

At a continuous scanning with the battery, the time interval between the data obtained in the same channel coincides with the scanning cycle duration. The increase of the variance as the sample length increases is caused by the fluctuations, the characteristic time of which is greater than the scanning time. Monotonic trend is the worst case. It is easy to show that in this case the difference between neighbor measurements is the upper estimate of the deviations within the limits of one scanning cycle. So, using such an estimate, we, in any case, do not overestimate the accuracy of data.

### Examples of application of the technique

An example of processing the data of continuous measurements of atmospheric aerosol by the technique described above is shown in Fig. 4. The data were obtained by means of the diffuse aerosol spectrometer (DSA)<sup>5</sup> developed at the Institute of Chemical Kinetics and Combustion, SB RAS. The result of processing the same data with estimating the error by usual Student formula is also shown here for a comparison. It is seen that in this case there are the irregular and obviously false peaks in the spectra. Application of the technique described above for estimating the variation of data leads to a more clear results. The diurnal behavior of ultrafine aerosol (< 100 nm) is well pronounced. Temporal variations of the size spectrum become more smooth.

Taking into account that we did not perform any temporal smoothing of the data, the latter fact seems to be a strong argument in favor of the reliability of the results obtained.



**Fig. 4.** Data of the continuous monitoring of the atmosphere by means of the diffuse battery. Each spectrum is obtained from the data averaged over 10 scanning cycles. The error in the data shown in the left-hand side plot is an estimate by the Student formula, and in the right-hand side plot the error is estimated by our technique.

### References

1. S. Eremenko and A. Ankilov, *J. Aerosol Sci.* **26**, Suppl. 1, S749–S750 (1995).
2. S.I. Eremenko, R. Caldow, A.M. Baklanov, M. Havlicek, and G. Sem, *J. Aerosol Sci.* **26**, Suppl. 1, S747–S748 (1995.)
3. J.K. Wolfenbarger and J.H. Seinfeld, *J. Aeros. Sci.* **24**, No. 2, 227–247 (1990).
4. J.J. Lloyd, C.J. Taylor, R.S. Lawson, and R.A. Shields, *J. Aerosol Sci.* **28**, No. 7, 1251–1264 (1997).
5. G.P. Reischl, A. Majerowicz, A. Ankilov, S. Eremenko, A. Baklanov, and R. Mavliev, *J. Aerosol Sci.* **22**, S325–S328 (1991).