

С.И. Еременко

## Диффузионная батарея при мониторинге атмосферы: методика оценки погрешности данных

*Институт химической кинетики и горения СО РАН, г. Новосибирск*

Поступила в редакцию 7.06.99 г.

Предложена практическая методика обработки данных сетчатой диффузионной батареи, предназначенная для мониторинга атмосферного аэрозоля. Методика основана на анализе статистики флуктуаций концентрации и обеспечивает надежную оценку погрешности данных и как следствие – надежные спектры размеров частиц.

Данные получают как непрерывную серию измерений. Затем для каждого вектора данных в серии определяется оптимальная длина окружающей его выборки для дальнейшего усреднения. Такая выборка дает максимальный выигрыш точности за счет усреднения данных при уверенности в том, что погрешность определена достаточно надежно. Далее усредненные по таким выборкам данные используются для получения спектров размеров частиц.

### Введение

Расчет спектра размеров аэрозоля из данных, полученных при помощи диффузионной батареи, представляет собой некорректную обратную задачу, не имеющую единственного решения. Разные методы решения таких задач предлагают разные критерии выбора из множества возможных решений, удовлетворяющих исходным данным.

Мы ранее предложили для этой задачи метод усреднения по множеству решений [1], который основан на предположении, что точность исходных данных известна и решение получается как результат усреднения по выборке из множества решений, удовлетворяющих исходным данным, причем границы этого множества определяются ошибкой данных. Метод хорошо себя показал в лабораторных испытаниях, когда источником аэрозоля служил стабильный генератор [2]. Имитационные расчеты и эксперименты с бимодальным аэрозолем показали, что оптимальное разрешение достигается при отношении сигнал-шум около 50, при этом два монодисперсных пика в спектре размеров разрешаются, если отношение их диаметров 2,5 и больше. Дальнейшее улучшение точности данных слабо влияет на разрешение.

Диффузионный метод определения спектра размеров в диапазоне 3–200 нм представляется особенно перспективным для исследования атмосферного аэрозоля, для которого низкая разрешающая способность не является существенным ограничением. Однако практика нескольких лет измерений показала, что в этом случае необходим особый подход к оценке вариации данных, без которого не удастся получать надежные спектры размеров.

Высокая чувствительность решения обратной задачи указанного выше типа к оценке погрешности данных уже обсуждалась неоднократно (см., например, [3, 4]). Заниженная оценка погрешности данных ведет к неустойчивости решения и к появлению в спектре размеров ложных пиков. Завышение погрешности данных приводит к необоснованной потере разрешения, т. е. полученный спектр размеров получается сглаженным, а пики уширенными.

Таким образом, для корректного решения задачи необходима надежная оценка точности исходных данных. Это означает, что для получения достоверного спектра размеров частиц необходимо не только измерить значения концентрации частиц в каждом канале диффузионной батареи, но и правильно оценить ошибку этих значений. Очевидно, что для интерпретации данных уширение спектра размеров предпочтительнее появления ложной структуры, поэтому особенно тщательно следует избегать ошибок в сторону занижения погрешности. На практике этим часто пренебрегают, что, по-видимому, и приводит к распространенному мнению о ненадежности спектров размеров частиц, полученных при помощи диффузионной батареи.

Цель настоящей работы заключается в разработке практической методики сбора и обработки данных диффузионной батареи, позволяющей получать достоверные спектры размеров как для лабораторных, так и для атмосферных измерений.

### Влияние точности данных и ошибки в ее оценке на надежность спектра размеров

Чтобы оценить степень влияния ошибки в оценке точности данных на надежность решения, было проведено несколько серий вычислительных экспериментов, состоявших в следующем.

Задавали тот или иной спектр размеров частиц. Для этого спектра генерировали серию векторов данных с разными реализациями случайного шума, заданного величиной относительного стандартного отклонения. Спектр размеров восстанавливали при помощи алгоритма MSA, задавая разные значения относительного стандартного отклонения данных, тем самым симулируя ошибку оценки точности данных.

Результаты численных экспериментов показали следующее. При росте относительной ошибки данных от 1 до примерно 10% восстановленный спектр размеров уширяется, качественно сохраняя свой вид (рис. 1), при этом такие характеристики, как средний, среднегеометрический и

среднекубический диаметры изменяются незначительно (рис. 2). При относительной ошибке данных около 14% в спектре размеров появляется ложное «плечо» со стороны левой границы размеров и с этого момента восстановленный спектр теряет качественное согласие с исходным.

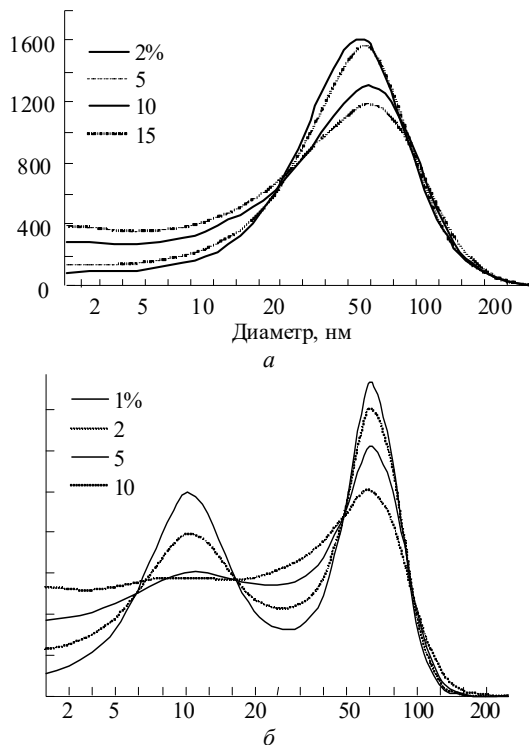


Рис. 1. Усредненные решения (восстановленные из симулированных проскоков диффузионной батареи спектры размеров частиц) в зависимости от относительного стандартного отклонения данных

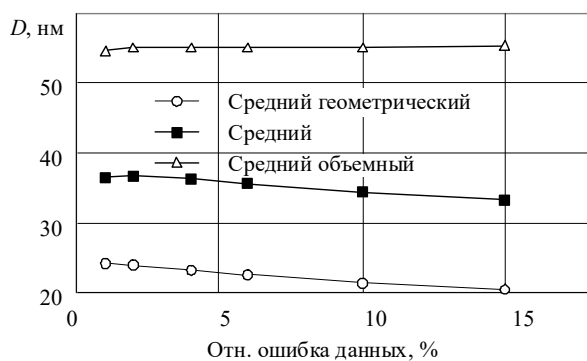


Рис. 2. Интегральные характеристики спектра размеров частиц в зависимости от относительного стандартного отклонения данных. Используются спектры размеров, приведенные на рис. 1, б

Занижение точности при восстановлении спектра ведет к его уширению. Завышение точности не отражается существенно на виде спектра размеров, пока заложенная при решении ошибка данных не становится меньше истинной на 70%. С этого момента восстановленные спектры размеров заметно теряют устойчивость и в них начинает появляться ложная структура.

Таким образом, из численных экспериментов получены оценки максимально допустимой относительной ошибки данных (10–12%) и максимально допустимой ошибки оценки стандартного отклонения данных – 30%.

## Способы оценки вариации данных

Вектор данных, полученный в результате однократного сканирования батареи, не дает возможности надежно оценить вариацию данных и поэтому практически непригоден для восстановления спектра размеров.

Иногда при анализе однократного вектора данных прибегают к следующему приему. Находят решение, минимизирующее евклидову норму вектора ошибок, после чего используют полученное значение как оценку дисперсии данных. Такая оценка допустима только в случае, когда число степеней свободы для данных значительно больше, чем для решения, в ином случае (в том числе для нашей задачи) такая оценка, как правило, значительно занижает погрешность данных, что приводит к неустойчивости решения.

Наиболее надежная оценка вариации данных получается путем анализа серии из  $N$  последовательно измеренных векторов данных. Казалось бы, выбирая  $N$  достаточно большим, можно получить сколько угодно точную оценку вариации данных. Однако это верно лишь в том случае, когда ошибки последовательных векторов данных независимы и всю выборку можно отнести к одной генеральной совокупности. Это возможно, если флуктуации концентрации частиц на входе прибора представляют собой стационарный случайный процесс, время автокорреляции которого меньше времени выборки вектора данных.

В качестве критерия того, что выборка удовлетворяет этим условиям, можно предложить зависимость выборочной дисперсии от длины серии. В случае независимых ошибок дисперсия не зависит от длины выборки. Достоверный рост выборочной дисперсии с ростом длины выборки указывает на существенный вклад низкочастотных флуктуаций концентрации.

На рис. 3 приведены зависимости выборочной дисперсии от длины выборки для атмосферного аэрозоля пригородной зоны и для аэрозоля от лабораторного генератора. Из рисунка видно, что хотя масштаб вариации данных различается на порядок, дисперсия в обоих случаях быстро возрастает с увеличением длины выборки. Практически это означает, что, например, для приведенных атмосферных данных для оценки вариации можно использовать не более 2–3 соседних измерений.

Это обстоятельство порождает проблему, так как выборка из 2–3 значений не позволяет оценить дисперсию с удовлетворительной точностью. Однако из того же рис. 3 видно, что относительная вариация данных практически одинакова для всех каналов батареи. Заметно большие значения для 7-го канала обусловлены в основном тем, что для него становится заметным вклад статистического пуассоновского шума. Последний определяется числом отсчетов за время экспозиции и легко может быть оценен. Для данных рис. 3, а относительная вариация за вычетом вклада пуассоновского шума одна и та же для всех каналов с точностью около 20%. Сопоставление многих данных, полученных при атмосферных измерениях за последние годы, показывает, что такая картина типична для них.

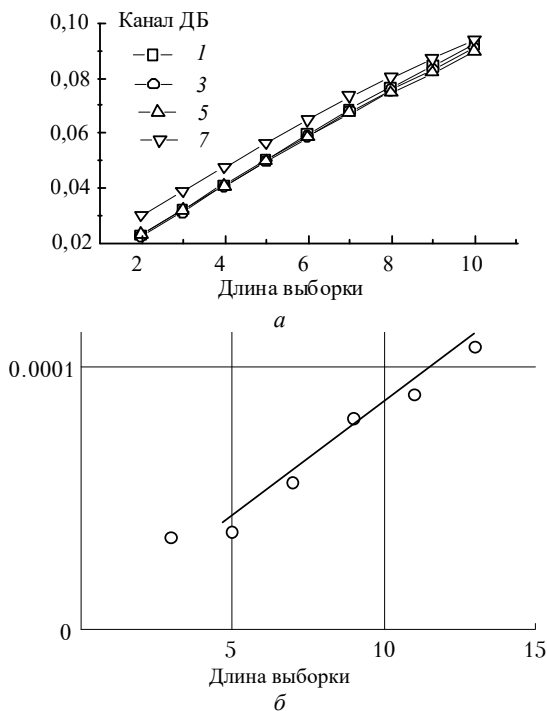


Рис. 3. Зависимость относительной дисперсии данных диффузионной батареи ДБ от длины выборки: *a* – атмосферные измерения, *б* – лабораторный генератор

Если принять гипотезу постоянства относительной вариации концентрации для всех каналов батареи, то выборка из 2–3 соседних измерений имеет уже достаточно степеней свободы для оценки дисперсии с точностью не менее 10–20%, что вполне достаточно для наших целей. Таким образом, можно предложить следующую технологию получения и обработки данных для диффузионной батареи.

#### 4. Методика получения и обработки данных

Наилучшая стратегия измерений – сделать время однократного сканирования как можно меньше и набрать серию данных как можно длиннее. Практически время сканирования батареи ограничено временем установления концентрации в измерительном объеме счетчика после переключения канала (обычно это несколько секунд). Разумно устанавливать время экспозиции в каждом канале того же порядка или в 2–3 раза больше. В итоге имеем последовательность векторов данных  $\mathbf{b}_n$ . Для простоты предположим, что вклад пуассоновского шума пренебрежимо мал по сравнению с флуктуациями концентрации. Если это не так, то легко внести необходимые поправки, которые опускаем для простоты изложения.

Выберем некоторый  $\mathbf{b}_n$  достаточно далеко от начала и конца серии, чтобы не рассматривать краевых эффектов. Оценим дисперсию по симметричной относительно  $\mathbf{b}_n$  выборке из  $2k + 1$  векторов данных. Возьмем вначале  $k = 1$  и оценим относительное стандартное отклонение следующим образом.

Найдем выборочные средние для каждого  $j$ -го канала батареи

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

и оценим относительную выборочную дисперсию

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

где  $m$  – число каналов батареи. Дисперсия оценки  $D_{kj}$  составляет  $D_{kj}^2/k$ , и, следовательно, относительная ошибка этой оценки

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

Например, для  $m = 8$  и  $k = 1$  это около 13%.

Теперь увеличим  $k$  на единицу и вновь сделаем оценку  $D_k$ . Будем повторять эту операцию до тех пор, пока полученная оценка относительного стандартного отклонения статистически значимо не превысит  $D_1$  на те 30%, которые мы сочли ранее предельно допустимой ошибкой. Если это произошло на  $p$ -м шаге, то примем предельную полуширину выборки  $k = p - 1$ . В качестве исходных данных для решения обратной задачи возьмем  $\bar{b}_j(k)$ , их стандартное отклонение оценим как

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

Особого подхода заслуживает случай, когда на первом же шаге оценка дисперсии возрастает существенно больше, чем на допустимую величину. В этом случае можно использовать в качестве данных для преобразования результат одного измерения без усреднения, а в качестве оценки его точности – оценку по двум его соседям (или по одному соседу, если это измерение – первое или последнее в серии). Такой подход оправдывается следующими соображениями.

При непрерывном сканировании батареи смещение по времени между данными для одного и того же канала совпадает с длительностью цикла сканирования. Рост дисперсии при увеличении длины выборки обусловлен флуктуациями, характерное время которых больше времени цикла сканирования. Монотонный тренд представляет собой наихудший случай. Легко показать, что в этом случае различие между соседними измерениями служит верхней оценкой отклонений в пределах одного цикла сканирования. Поэтому, используя такую оценку, мы, во всяком случае, не завышаем точность данных.

#### Примеры применения методики

На рис. 4 приведен пример обработки данных непрерывных измерений атмосферного аэрозоля по описанной выше методике. Данные получены при помощи диффузионного спектрометра аэрозолей (ДСА), разработанного в ИХ-КиГ СО РАН [5]. Для сравнения на том же рисунке приведен результат обработки тех же данных с оценкой ошибки по обычной формуле Стьюдента. Видно, что в этом случае в спектрах присутствуют бессистемно расположенные, очевидно ложные пики. Применение для оценки вариации данных описанной выше методики приводит к значительно более вразумительным результатам. Отчетливо виден суточный

ход ультрадисперсного (<100 нм) аэрозоля, изменения спектра размеров со временем становятся плавными.

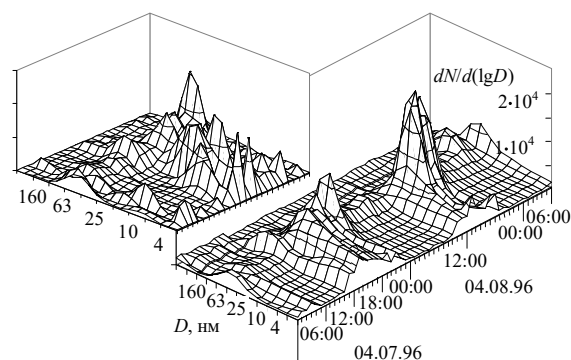


Рис. 4. Данные непрерывного мониторинга атмосферы диффузионной батареей. Каждый спектр размеров получен из данных, полученных усреднением по 10 последовательным циклам батареи. На левом графике ошибка данных оценивалась по Стюденту обычным методом, на правом – по предложенной нами методике

Последнее обстоятельство, учитывая, что никакого сглаживания данных по времени не проводилось, представляется сильным аргументом в пользу надежности полученных результатов.

1. Eremenko S. and Ankilov A. Conversion of the diffusion battery data to particle size distribution: Multiple Solutions Averaging algorithm (MSA) // J. Aerosol Sci. 1995. V. 26. Suppl. 1. P. S749–S750.
2. Eremenko S.I., Caldow R., Baklanov A.M., Havlicek M. and Sem G. Diffusion Battery Particle Sizing System Based on the MSA Data Conversion Algorithm: Experimental Examination // J. Aerosol Sci. 1995. V. 26. Suppl. 1. P. S747–S748.
3. Wolfenbarger J.K. & Seinfeld J.H. Inversion of aerosol size distribution data // J. Aerosol Sci. 1990. V. 24. N 2. P. 227–247.
4. Lloyd J.J., Taylor C.J., Lawson R.S. & Shields R.A. The use of the L-curve method in the inversion of diffusion battery data // J. Aerosol Sci. 1997. V. 28. N 7. P. 1251–1264.
5. Reischl G.P., Majerowicz A., Ankilov A., Eremenko S., Baklanov A., Mavliev R. Comparison of the Novosibirsk automated diffusion battery with Vienna electro mobility spectrometer // J. Aerosol Sci. 1991. V. 22. P. S325–S328.

#### S.I. Eremenko. Diffusion battery application to atmospheric monitoring: the method of data error estimation.

The method of data error estimation assigned for continuous atmospheric measurements with diffusion battery is presented. Such estimation is necessary for reliable data inversion to size spectra. The technique is based on the statistical analysis of particle concentration fluctuations.

Data were obtained as continuous series of battery scans. For each scan the optimum number of the surrounding scans to be averaged was determined. The number is maximum that still allows one to keep error estimation reliable enough. Averaged data together with error estimation values were used for further data inversion.