AEROSOL DEPOSITION, GROWTH, AND DYNAMICS IN THE CONTINUOUS STIRRED TANK REACTOR

Thesis by

James Gleason Crump III

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosphy

California Institute of Technology

Pasadena, California

1983

(submitted September 17, 1982)

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. John Seinfeld, without whose encouragement and guidance this thesis would not have been written.

I am also grateful to Dr. Richard Flagan for his considerable advice, particularly on the experiments.

A special note of appreciation goes to Lenore Kerner, who so capably typed this manuscript and offered several helpful suggestions.

Finally, I am pleased to acknowledge the financial support provided by grants from the Exxon Education Foundation and General Motors Corporation and a Rockwell graduate fellowship.

ABSTRACT

This work examines three related topics in aerosol science. First, a continuous stirred tank reactor (CSTR) for studying the dynamics of chemically reacting aerosol systems is described. This apparatus is designed to allow aerosols to react under conditions of controlled temperature and relative humidity and is applied to the study of growth of aqueous manganese sulfate aerosols in a humid atmosphere containing sulfur dioxide. From experimental data the rate of conversion of sulfur dioxide to sulfuric acid in manganese sulfate aerosols is deduced.

Second, a new algorithm for inversion of aerosol size distribution data is presented. This algorithm is well suited to the ill-posed nature of the data inversion problem and is shown to give results superior to those obtained using conventional methods. This inversion technique is applied to the analysis of aerosol growth data.

Finally, the general steady state coagulation equations with particle sources and sinks are examined and shown to admit physically unrealistic solutions in some cases. General conditions are then given which insure the existence of physically acceptable solutions and these solutions are shown to have large particle tails that decay exponentially.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

INTRODUCTION

An understanding of the chemical and physical interactions which lead to evolution of particle size distributions in aerosol systems would have a strong impact on environment and possibly industrial applications. Unfortunately, at present there exists a lack of fundamental information concerning many of the processes occurring in such systems, as well as techniques to obtain such information.

Much of the interest in the evolution of chemically reacting aerosol systems has centered on predicting characteristics of plumes from coal-fired boilers. In such a plume, an aerosol may consist of aqueous drops containing heavy metal ions, oxides, alkali salts, and more or less insoluble matter, surrounded by an atmosphere containing sulfur dioxide. Because of the catalytic effect of the heavy metals (Bassett and Parker, 1951) sulfur dioxide is oxidized to sulfate in the aerosol phase. Much research has focussed on the rate and extent of this reaction (Haury, et al. 1978). In addition coagulation, particle deposition, and homogeneous gas phase reactions occur simultaneously, making such systems extremely complex.

Despite the complexity, some progress can be made by studying simpler aerosol systems containing only a few chemical components. In this way some of the features of plume behavior have been modeled (Bassett, et al. 1981, Freiberg, 1978). However, little has been done to study chemically reacting aerosols in a controlled environment where both chemical and physical behavior may be observed. Such studies are necessary to assess our understanding of the basic phenomenon and to fully integrate theories describing the various processes occurring in aerosol systems.

In Chapter 2 we describe such a system - a continuous stirred tank reactor (CSTR) for aerosol studies. In Chapter 3 we present some theoretical results on effluent size distributions from the CSTR resulting from the combined effects of coagulation, particle growth, and formation of new particles by homogeneous nucleation. Chapters 4 and 5 give theoretical results on wall loss rates in the CSTR, and in Chapters 10 and 11 we describe some theoretical and experimental results on particle growth due to oxidation of sulfur dioxide in manganese sulfate aerosols, concluding with a rate expression for the oxidation reaction based on the size distribution data.

Measurement techniques for aerosol size distributions are as yet only partly successful. Many of the techniques available yield data that are difficult to interpret in terms of size distributions. One of the advances that has occurred in the last few years is the realization that the mathematical problems involved in inversion of aerosol size distribution data are ill-posed (Twomey, 1975), and consequently are not amenable to solution by conventional techniques. Improved data inversion techniques, based on methods specifically adopted to ill-posed problems may be expected not only to improve data analysis of present measuring techniques, but also to enable a wide variety of techniques to be employed.

In Chapters 6 and 7 we present new algorithms for the inversion of aerosol size distribution data based on fundamental mathematical work of Tikhonov and Arsenin (1977) and Wahba (1977). In Chapter 8 we obtain calibration data for an optical particle sizing instrument enabling us to use these new methods to study size distribution changes during particle growth.

An area of aerosol science which has remained virtually untouched is that of mathematical analysis of the basic equations of aerosol dynamics. Although this area is never likely to be at the forefront of the aerosol dynamics field, it is important in the sense that rigorous mathematical results provide a means to assess the validity of numerical results on complex mathematical models. Moreover, questions of existence and uniqueness of solutions, although perhaps not intrinsically interesting, assume practical significance when they are answered negatively.

One of the first general existence results was given by McCleod (1962) for the unsteady discrete coagulation equation. Later, White (1980) proved a general global existence and uniqueness result for these equations under more restrictive conditions, and also showed that the number distributions resulting from solution of the equations must decay exponentially fast for large particle sizes. Later, White (1982) gave a proof of the exponential decay of the large particle tails for steady state source reinforced coagulating aerosols. In Chapter 9 we present an extension of this result, and show existence of solutions and exponential decay of the tails under more general conditions than those cited by White (1982). We also discuss the possible significance of physically unrealistic solutions.

REFERENCES

Bassett, H. and Parker, W. G., J. Chem. Soc., 1540 (1951).

Bassett, M., Gelbard, F., and Seinfeld, J. H. Atm. Env. 15, 2395 (1981).

Freiberg, J. Atm. Env. 12, 339 (1978).

Haury, G., Jorgan, S., and Hofmann, C. Atm. Env. 12, 281 (1978).

McCleod, J. B., Quart. J. Math. Oxford (2)13, 119 (1962).

Tikhonov, A. N. and Arsenin, V. Y., Solutions of Ill-Posed Problems, Halsted Press, Washington (1977).

Wahba, G., SIAM J. Numer. Anal. 14, 651 (1977).

White, W., Proc. Amer. Math. Soc. 80, 273 (1980).

White, W., J. Coll. Int. Sci. 87, 204 (1982).

CHAPTER 2

DESIGN OF CSTR SYSTEM

DESIGN OF CSTR SYSTEM

INTRODUCTION

This chapter presents a design and layout of a continuous stirred tank for aerosol studies. The goal in designing such a system was to allow studies of aerosol growth due to oxidation of sulfur dioxide to be carried out under controlled conditions of temperature, sulfur dioxide concentration, relative humidity, and particle size and concentration. At the same time, however, it was desired to keep the design flexible to allow for possible future work on other aerosol systems.

DESCRIPTION OF APPARATUS

Figure 1 shows a schematic of the apparatus. Aerosol is generated by atomization of an aqueous solution or dispersion in a device similar to that described by Liu and Lee (1975). Figure 2 shows a sketch of the atomizer. The aerosol generator is fed by a syringe pump and the feed solution is atomized by a stream of compressed air at 36 psig passing through a 0.0145" diameter orifice. By operating the syringe pump at a flow of 0.2 cm^3/min. nearly five hours of steady operation can be obtained using 60 ml syringes. Nearly steady operation can be maintained indefinitely by changing the syringe, which requires less than a minute.

From the atomizer the aerosol stream passes through a Kr-85 charge neutralizer, which reduces the charges on the aerosol to a bipolar Boltzmann equilibrium distribution, and then the gas passes to a 10-liter holding vessel to smooth fluctuations in particle concentrations.

Provision is made for dilution of the aerosol in this vessel by metered addition of air, and aerosol can also be removed to decrease flow

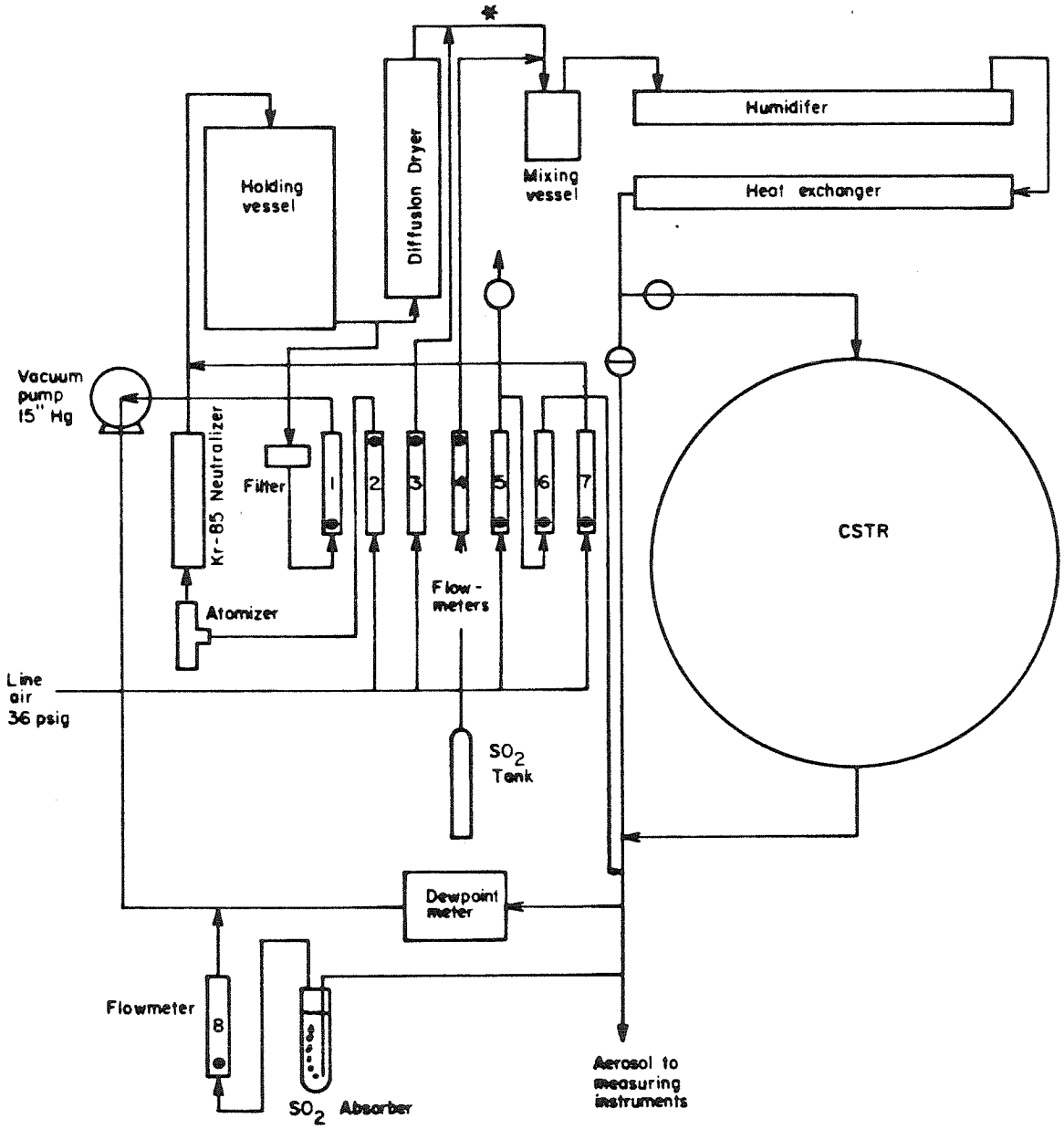


Figure 1. General layout of CSTR system.

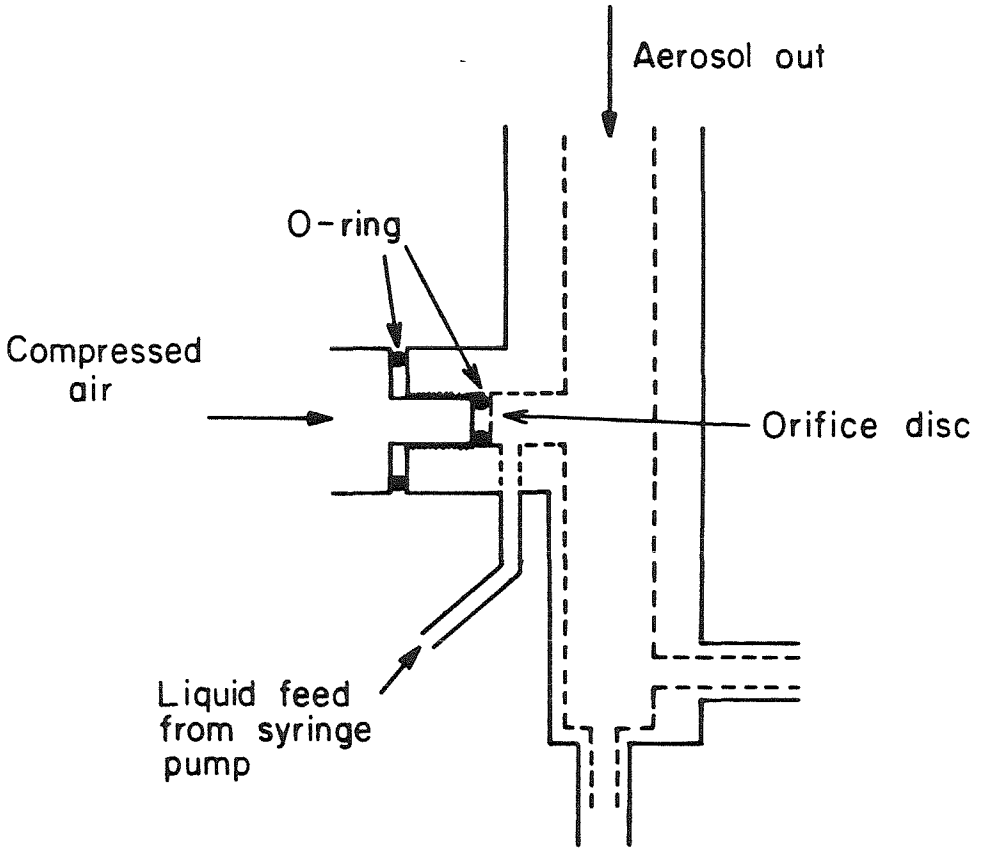


Figure 2. Atomizer for generating aerosol feed to system.

rate as the gas exists the holding vessel by means of a connection to a vacuum system through a filter and flowmeter.

The vacuum system is maintained at fifteen inches of mercury absolute pressure using a vacuum regulator. All connections to vacuum are to the same system.

Aerosol leaving the holding vessel is dried by passage through a three-foot tube surrounded by a concentric shell of Drierite. After drying, the aerosol may be further diluted by addition of air through the flowmeter. In experiments in which a monodisperse aerosol is used, a TSI model 3071 Electrical Mobility Classifier and a Kr-85 Charge neutralizer are attached to the system at this point, the monodisperse aerosol then passing to a one-liter mixing vessel where sulfur dioxide from a cylinder may be added.

From the mixing vessel the aerosol-gas mixture passes through a humidifier. This device consists of a pool of water maintained at a constant temperature by passage of temperature regulated water through about 40 feet of immersed copper coil. A sketch of the device is shown in Figure 3, and Appendix A details some design criteria. Aerosol-containing gas flowing over the water pool is humidified to saturation before passing through a heat exchanger, consisting of a water-jacketed tube through which temperature-regulated water passes.

The gas leaving the heat exchanger passes through a y-branch either to the reactor or directly to measuring instruments for measuring feed aerosol characteristics.

The reactor consists of a roughly spherical glass vessel of nominal 100 liters capacity. It is fitted with ports for entrance and exit of gas and temperature measurement. The entire vessel is supported in a

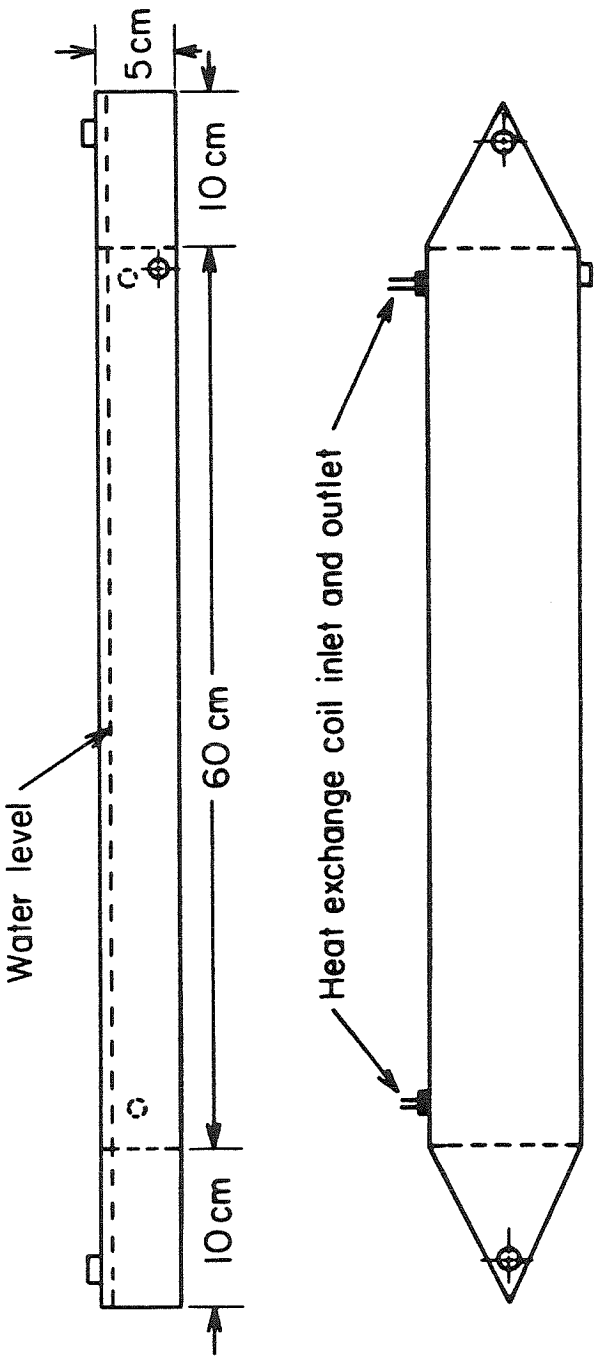


Figure 3. Aerosol humidifier

plywood frame, which completely encloses the upper half of the reactor, and which contains copper coils fed by water from the heat exchanger. The bottom of the reactor is insulated with glass wool.

Aerosol, either leaving the reactor or directly from the y-branch, is diluted, if necessary, and sampled for dewpoint using an EG&6 model 911 DEW-ALL digital humidity analyzer. This device also gives a reactor temperature readout from a thermocouple inserted in the reactor.

Sulfur dioxide concentration is measured by bubbling a known volume of gas through a solution of hydrogen peroxide and titrating the sulfuric acid formed with 0.01N sodium hydroxide.

The flowmeters, numbered 1-8 in Figure 1, are all rotameters obtained from Matheson Gas Products. The black dots on the flowmeters shown in Figure 1 show the locations of the flowmeter valves. Calibration was done at the pressure downstream of the valve if the valve is at the bottom and upstream if the valve is at the top.

TRACER EXPERIMENTS

The CSTR has no mechanical stirring apparatus, but depends on convective mixing due to the flow through the vessel to provide mixing. This offers the advantage that no extraneous surfaces are present which might interact with aerosol.

In order to assess the validity of the ideal mixing assumption, several tracer experiments using oxygen were carried out.

In such an experiment, nitrogen was first fed into the reactor at a flow of 2.09 l/min. and the decay in oxygen concentration was followed using a Beckman 755 oxygen analyzer. When the oxygen concentration had

fallen to approximately 10 percent of its ambient concentration, the measurements were stopped, and the nitrogen flow was increased to purge the vessel of oxygen. When the oxygen concentration became negligible (0.1%), air was admitted to the CSTR at 2.09 l/min., and again the oxygen concentration was monitored as a function of time.

Data from experiments in which nitrogen was fed to the reactor are plotted in Figure 4 as the logarithm of oxygen concentration versus time in minutes. Note that these points fall nearly on a straight line. Data from the experiments in which oxygen was fed into the reactor are plotted in Figure 5 as $\ln(1-c/20.9)$ versus time, where c is oxygen concentration in percent.

The vessel volume can then be calculated from the residence time determined from the slope of these plots and the known flowrates. Volumes given by the two kinds of experiments described above average 118 liters and differ by about 6 percent from each other. The value 118 liters obtained in these experiments is to be compared with a value of 115 liters obtained by estimates made from external measurements of the vessel.

Had there occurred significant channeling due to poor mixing in the vessel, a smaller effective volume would be expected to result from the tracer experimental data. Since this was not the case, it was concluded that the mixing was rapid enough relative to the residence time to be considered ideal.

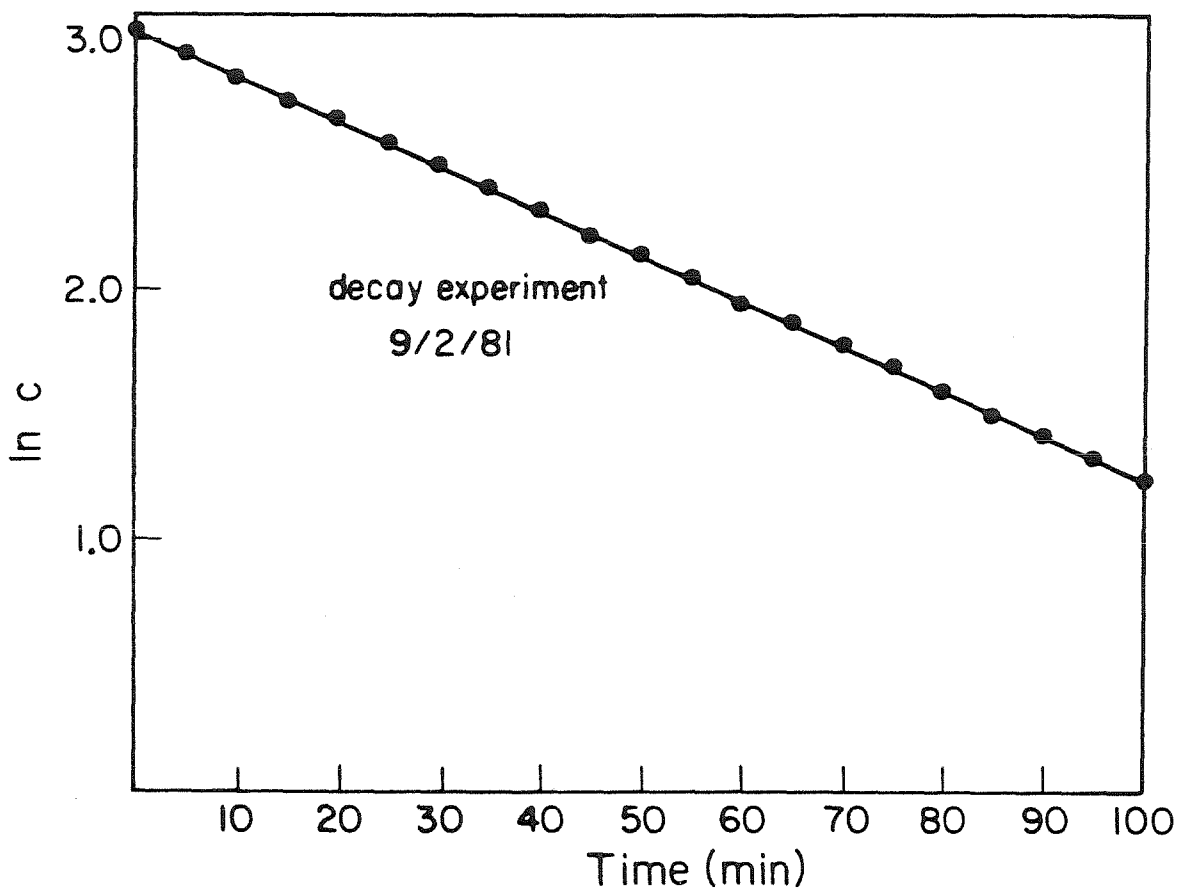


Figure 4. Logarithm of CSTR oxygen concentration (%) versus time.

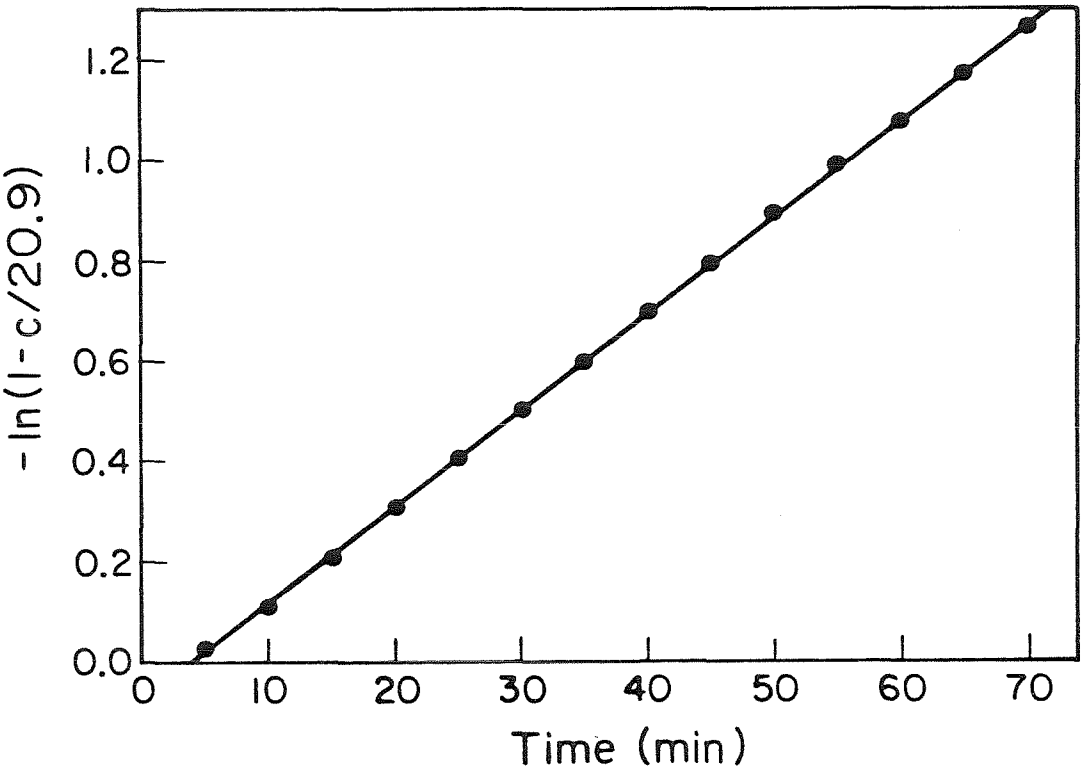


Figure 5. ln(1-c/20.9) versus time, where c is oxygen concentration in percent.

REFERENCES

Liu, B. Y. H. and K. W. Lee, Am. Ind. Hygiene Assoc. J., 36, 861 (1975).

APPENDIX

DESIGN OF AEROSOL HUMIDIFIER

Two important criteria determine the design of an aerosol humidifier. They are the efficient humidification of the gas and minimum loss of particles. That these objectives are compatible is suggested by the fact that particle Brownian diffusivities are several orders of magnitude smaller than gas diffusivities.

To treat the humidification problem, we assume fully developed laminar flow in the vapor space above the water pool and ignore the effect of the edge. Under these assumptions the following equation results if we ignore diffusion along the flow

$$0\ \frac{\partial^2 c}{\partial y^2} = 4v_o\left(\frac{y}{h} - \frac{y^2}{h^2}\right)\frac{\partial c}{\partial z} \tag{1}$$

where D is the water vapor diffusivity, c is water vapor concentration, h is thickness of the vapor space, y is distance measured from the water surface, and z distance downstream from the gas inlet.

The boundary conditions corresponding to the situation described are

$$c(0,z) = c_o \tag{2a}$$

$$\frac{\partial c}{\partial y}(h,z) = 0 \tag{2b}$$

$$c(y,0) = 0 \tag{2c}$$

In dimensionless form (1) and (2) assume the forms

$$\frac{\partial^2 u}{\partial \xi^2} = \frac{4h^2 v_o}{DL}(\xi - \xi^2)\frac{\partial u}{\partial s} \tag{3}$$

$$u(0,s) = 1 \tag{4a}$$

$$\frac{\partial u}{\partial \xi}(1,s) = 0 \tag{4b}$$

$$u(\xi,0) = 0 \tag{4c}$$

where

$$u = c/c_o \tag{5a}$$

$$\xi = y/h \tag{5b}$$

$$s = z/L \tag{5c}$$

The solution u is of the form

$$u(\xi,s) = 1 + \sum_{n=1}^{\lambda} c_n \psi_n(\xi)\, e^{-\frac{DL\lambda_n^2 s}{4h^2 v_o}} \tag{6}$$

where λ_n and ψ_n satisfy the eigenvalue problem

$$\psi_n'' + \lambda_n^2(\xi-\xi^2)\psi_n = 0 \tag{7a}$$

$$\psi_n(0) = 0 \tag{7b}$$

$$\psi_n'(1) = 0 \tag{7c}$$

and

$$c_n = -\int_0^1 (\xi-\xi^2)\psi_n(\xi)d\xi \Big/ \int_0^1 (\xi-\xi^2)\psi_n^2(\xi)d\xi \tag{8}$$

The quantity of interest here is the cup-mixed average concentration at the outlet, given by

$$\bar{u} = \int_0^1 (\xi-\xi^2)u(\xi,1)d\xi \Big/ \int_0^1 (\xi-\xi^2)\, d\xi \tag{9}$$

Numerical solution of (7) gives

$$\bar{u} = 1 - 0.899e^{-2.43\left(\frac{DLW}{hQ}\right)} - 0.0607e^{-23.5\left(\frac{DLW}{hQ}\right)} \tag{10}$$

where w is the vapor space width and Q the volume flow rate of gas.

The third term in (10) is negligible under conditions of near complete saturation, so can be omitted. To achieve humidification to within a fraction r of saturation, we require

$$0.899e^{-2.43\left(\frac{DLW}{hQ}\right)} < r$$

or equivalently

$$\frac{DLW}{hQ} \geq \frac{(\ln(0.899)-\ln r)}{2.43} = -\frac{(0.106+\ln r)}{2.43} \tag{11}$$

To treat particle loss one must account for both diffusional loss and gravitational sedimentation. As a first approximation, we assume these are independent and deal with sedimentation alone first.

Particles entering at height y_o above the water surface follow trajectories given by the following differential equations, provided they follow the flow in the horizontal direction

$$\frac{dx}{dt} = 4v_o\left(\frac{y}{h} - \frac{y^2}{h^2}\right) \tag{12a}$$

$$\frac{dy}{dt} = -v \tag{12b}$$

where v is the settling velocity of the particle. At t = 0 the particle positions satisfy

$$x(0) = 0 \tag{13a}$$

$$y(0) = y_o \tag{13b}$$

Since trajectories cannot cross, we need only determine the critical trajectory, $(x_c(t), y_c(t))$, for which $x_c(t) = L$ and $y_c(t) = 0$ for some t. Solving (12) with (13) for (x(t), y(t)) and setting

$$x(t) = L \tag{14a}$$

$$y(t) = 0 \tag{14b}$$

for the critical trajectory gives the relation

$$\left(\frac{y_o^2}{2h^2} - \frac{y_o^3}{3h^3}\right) = \frac{VWL}{6Q} \tag{15}$$

This defines y_o for the critical trajectory. So all particles entering above y_o escape, and those entering below are caught. The fraction entering below y_o is given by:

$$\int_0^{y_o} \left(\frac{y}{h} - \frac{y^2}{h^2}\right) dy \Big/ \int_0^1 \left(\frac{y}{n} - \frac{y^2}{h^2}\right) dy$$

$$= 6\left(\frac{y_o^2}{2h^2} - \frac{y_o^3}{3h^3}\right) = \frac{VWL}{Q}$$

Hence, to remove fewer than a fraction r of the particles requires that

$$\frac{VWL}{Q} < r \tag{16}$$

The loss by particle diffusion is expected to be small due to small particle diffusivities. The Equation (3) may be written

$$\frac{\partial^2 u}{\partial \xi^2} = \alpha(\xi - \xi^2)\frac{\partial u}{\partial s} \tag{17}$$

where $\alpha = \dfrac{4h^2 v_o}{DL} >> 1$. The boundary conditions in this case are

$$u(0,s) = u(1,s) = 0, \quad u(\xi,0) = 1 \tag{18}$$

We expect a concentration boundary layer, but little change in the bulk concentration, so we solve (17) and (18) near one wall using the approximate equation

$$\frac{\partial^2 u}{\partial \xi^2} = \alpha\xi\frac{\partial u}{\partial s} \tag{19}$$

with boundary conditions

$$u(0,s) = 0, \lim_{\xi\to\infty} u(\xi,s) = 1, u(\xi,0) = 1 \tag{20}$$

This can be solved by a similarity substitution and by applying it to both walls we obtain the cup-mixed average concentration at the outlet as

$$1 - \frac{6}{\Gamma(\frac{1}{3})}\left(\frac{9}{\alpha}\right)^{2/3} = 1 - 2.24\left(\frac{9}{\alpha}\right)^{2/3} \tag{21}$$

Hence for the loss fraction to be less than r we require

$$\frac{D_p LW}{hQ} \leq 0.2r^{3/2} \tag{22}$$

Using the values

$L = 60$ cm

$h = 1$ cm

$w = 10$ cm

humidification to within one percent of saturation will occur if, according to inequality 11

$$Q < 4.9 \quad \ell/\text{min} \tag{23}$$

Hence essentially complete humidification is obtained up to nearly 5 ℓ/min. flow. At 2 ℓ/min. saturation is achieved to within 0.002 percent. Under these conditions inequality (16) predicts that 5.4 percent of particles 1 μm in diameter will be lost by gravitational sedimentation, and 13 percent of 0.01 μm particles will be lost by diffusion.

The estimates can be expected to hold as long as flow remains laminar. At a flow of 5 ℓ/min. the Reynolds number for this system is given by

$$Re = \rho h \bar{v}/\mu = 53 \tag{24}$$

Consequently, laminar flow is expected.

CHAPTER 3

AEROSOL BEHAVIOR IN THE CONTINUOUS STIRRED TANK REACTOR

Published in AIChE Journal 26, 616 (1980).

Aerosol Behavior in The Continuous Stirred Tank Reactor

JAMES G. CRUMP

and

JOHN H. SEINFELD

Department of Chemical Engineering
California Institute of Technology
Pasadena, California 91125

The basic features of aerosol behavior in the CSTR are examined. Solutions are obtained for the steady state aerosol size distribution during simultaneous coagulation, particle growth by vapor condensation and new particle formation by nucleation. Explicit distributions are shown for the case of a monodisperse feed aerosol.

SCOPE

Studies of particle formation and evolution in combustion systems and in laboratory simulations of atmospheric chemistry sometimes involve the use of a CSTR. The interpretation of aerosol size distributions in a CSTR requires the development and solution of the general population balance equation applicable to that system. The phenomena that must be considered include coagulation, particle growth by vapor condensation and new particle formation by vapor nucleation. Because the general problem of aerosol behavior in the CSTR does not appear to have been studied previously, an examination of the qualitative features of the steady state size distributions that may be achieved is deemed an appropriate first step to a more in depth analysis. Of particular interest is the elucidation of the effects of varying residence time and the characteristic times for coagulation and growth by condensation on the size distributions attained.

0001-1541-80-3904-0610-$00.75.

CONCLUSIONS AND SIGNIFICANCE

A general solution for the steady state aerosol size distribution in a CSTR is obtained for an arbitrary feed aerosol size distribution and for kinetic coagulation coefficient and particle growth rate that are independent of particle size. The object of the analysis is to demonstrate the qualitative features of the size distribution, and, consequently, the simplified functional forms of the coagulation and growth coefficients were chosen so as to enable exact solution of the governing equations. The solution is explicitly illustrated for a monodisperse feed aerosol. Although the assumption of a constant coagulation coefficient is valid in certain instances, that of a size independent growth rate is not. Thus, the case of growth rate linearly proportional to particle volume is investigated as a perturbation to the constant growth case to determine the qualitative effect of the size dependent growth rate. Although simulation of a specific system will generally require numerical solution of the aerosol balance equation employing the proper detailed expressions for the coagulation and growth coefficients, the solutions presented here provide an indication of the expected behavior of the size distribution as residence time and other physical parameters are varied. In addition, the dimensionless groups governing aerosol behavior in a CSTR are defined. The groups indicate that for any system of this type, the relative magnitudes of the characteristic times for growth, coagulation and reactor residence will strongly influence the expected size distribution.

Understanding the processes associated with particle formation and evolution in combustion systems and in laboratory simulations of atmospheric chemistry requires consideration of aerosol behavior in typical chemical reactor configurations, particularly the continuous stirred tank reactor (CSTR) and the tubular flow reactor. The object of this work is to elucidate the basic features of aerosol behavior in the CSTR.

A variety of physical and chemical phenomena influence the size distribution of aerosols, including coagulation, growth by condensation of gases on the particles and formation of fresh particles by nucleation. In this work, we consider a general situation in which an aerosol of known size distribution is introduced into a CSTR together with a vapor species capable of transferring to the aerosol by condensation or of nucleating to form new particles. Thus, simultaneous coagulation, growth and new particle formation may occur in the reactor.

The equations governing the steady state aerosol size distribution and vapor concentration are presented and nondimensionalized. A general solution of these equations is then obtained for the case in which the kinetic coefficient of coagulation is independent of the sizes of the two particles and in which the rate of growth of an individual particle by condensation is independent of the size of the particle. The solution is explicitly illustrated for a monodisperse feed aerosol. Although the assumption of a constant coagulation coefficient is valid in certain instances, that of a size independent growth rate is generally not. Thus, the case of growth rate linearly proportional to particle volume and a monodisperse feed aerosol is then investigated to determine the effect of size dependent growth rate on the steady state aerosol size distribution.

EQUATIONS DESCRIBING GAS-AEROSOL BEHAVIOR IN A CSTR

Consider a CSTR operated at steady state in which an aerosol of known size distribution is fed into the reactor together with a gas, which condenses on the aerosol. The aerosol is characterized by its size distribution function $n(v)$, where $n(v)dv$ represents the number density of particles having volumes between v and $v + dv$. The molar concentration of the gas is denoted by c. Gas at concentration c_0 is continuously introduced into the reactor at volume flow rate q together with an aerosol of size distribution $n_0(v)$. The gas-aerosol system undergoes the following processes: condensation of the gas on the aerosol particles, homogeneous nucleation of the gas to produce new particles of a given volume v_0 and coagulation or coalescence of aerosol particles to form larger particles.

Let $S_0(c)$ be the rate of formation of particles by nucleation and $I(v) = cf(v)$ be the growth rate of a particle of volume v by condensation. The form $cf(v)$ assumes that condensation is irreversible, a good assumption for low vapor pressure gaseous species. Conservation of mass for aerosol and gas gives

$$q(c_0 - c) = 10^{-12}\left(\frac{\rho}{M}\right) Vc \int_0^\infty f(v)n(v)dv + 10^{-12}\left(\frac{\rho}{M}\right) Vv_0S_0(c) \quad (1)$$

$$q[n_0(v) - n(v)] + VS_0(c)\delta(v - v_0) - Vc\frac{d}{dv}[f(v)n(v)] + \frac{V}{2}\int_0^v \beta(v - \tilde{v}, \tilde{v})n(v - \tilde{v})n(\tilde{v})d\tilde{v} - Vn(v)\int_0^\infty \beta(v, \tilde{v})n(\tilde{v})d\tilde{v} = 0 \quad (2)$$

We note that the factors of 10^{-12} appear because particle volume is expressed in μm^3, whereas density has units of g cm^{-3}. It has been assumed that the total aerosol volume is small compared to that of the gas, so that the total volumetric flow rate for the mixture of gas and aerosol may be replaced by that of the gas alone, q. Unless the density of aerosol particles is extremely high, the right-hand side of Equation (1) will be negligible, and the gas concentration c will be the same as the feed gas concentration c_0. In any case, we will henceforth suppose that c is known and deal only with the aerosol balance, Equation (2). The boundary condition on (2) is

$$n(0) = 0 \quad (3)$$

All realistic distributions will also satisfy the condition $n(v) = 0$ as $v \to \infty$.

Because of the complex dependence of the coagulation coefficient $\beta(v, \tilde{v})$ and the condensation function $f(v)$ on the particle volume v, (2) must generally be solved numerically. In the present study, however, we are interested in elucidating the qualitative features of the solutions to (2) and for this reason will consider constant coagulation coefficient $\beta(v, \tilde{v}) = \beta$ and size independent condensation rate $f(v) = \sigma_0$. The validity of these assumptions has been discussed elsewhere (Ramabhadran et al., 1976). We note, in particular, that the assumption of constant β is physically realistic in the initial stages of Brownian coagulation of a monodisperse aerosol. However, the assumption of size independent growth is not in general obeyed. Usually, $f(v)$ can be expressed as $f(v) = \sigma_\gamma v^\gamma$, $0 \le \gamma \le 1$ (Ramabhadran et al., 1976).

With the above assumptions, (2) becomes

$$Vc\sigma_0 n'(v) = \frac{1}{2}V\beta n^* n(v) - V\beta n(v)\int_0^\infty n(\tilde{v})d\tilde{v} + q[n_0(v) - n(v)] + VS_0\delta(v - v_0) \quad (4)$$

where

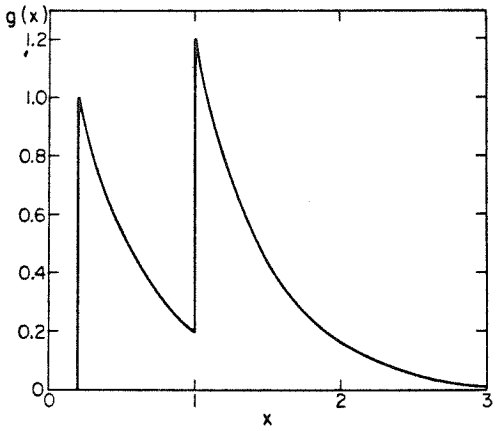


Figure 1. Dimensionless size distribution g(x) for monodisperse feed and no coagulation: $\kappa = 0$, $\alpha = 0.5$, $\omega = 0.5$, $x_0 = 0.2$.

$$n^*n(v) \equiv \int_0^r n(v - \bar{v})n(\bar{v})\, d\bar{v}$$

and $S_0 \equiv S_0(c)$.

We introduce the total number and volume densities of the aerosol

$$N = \int_0^\infty n(v)\, dv \qquad N_0 = \int_0^\infty n_0(v)\, dv$$

$$\bar{V} = \int_0^\infty vn(v)\, dv \qquad \bar{V}_0 = \int_0^\infty vn_0(v)\, dv$$

and the following dimensionless groups, $\alpha = \theta c\sigma_0 N_0/\bar{V}_0$, $\kappa = \beta\theta N$, $\omega = \theta S_0/N$, where the mean residence time $\theta = V/q$. α, κ and ω are ratios of residence time to characteristic condensation, coagulation and nucleation times, respectively.

Dimensionless particle volume and aerosol size distribution are defined as

$$x = N_0 v/\bar{V}_0$$

$$g(x) = \frac{\bar{V}_0}{N_0 N}\, n\left(\frac{\bar{V}_0 x}{N_0}\right)$$

Then, (4) becomes

$$\alpha g'(x) = \frac{1}{2}\,\kappa g^* g(x) - (1 + \kappa)g(x) + g_0(x) + \omega\delta(x - x_0) \qquad (5)$$

Note that the feed distribution $g_0(x)$ and the nucleation term $\omega\delta(x - x_0)$ play equivalent roles in (5); so, for simplicity, we combine them, setting $g_1(x) = g_0(x) + \omega\delta(x - x_0)$.

It is necessary to evaluate the steady state number density N. By integration of (4) from 0 to ∞ and solution of the resulting quadratic equation for N, one obtains

$$N = \frac{1}{\beta\theta}\left[\sqrt{1 + 2\beta\theta(N_0 + \theta S_0)} - 1\right] \qquad (6)$$

Subsequently we shall be interested in a monodisperse feed distribution $n_0(v) = N_0\delta(v - v_f)$. Using (6), we find that

$$g_0(x) = \left(1 + \frac{1}{2}\kappa - \omega\right)\delta(x - 1) \qquad (7)$$

STEADY STATE SOLUTIONS

In this section we examine solutions of the material balance (5). First we consider special cases in which one or more of the physical processes is unimportant.

No Coagulation; $\kappa = 0$

In the absence of coagulation, the aerosol fed to the reactor grows by condensation, and fresh aerosol forms by nucleation. There is no particle-particle interaction. Equation (5) can be easily solved to yield

$$g(x) = \frac{1}{\alpha}\left\{\int_0^x e^{-y/\alpha} g_0(x - y)\, dy + \omega U(x - x_0)e^{-(x - x_0)/\alpha}\right\} \qquad (8)$$

where U is the unit step function

$$U(x) = \begin{cases} 0 & x \le 0 \\ 1 & x > 0 \end{cases}$$

For a monodisperse feed, g_0 is given by (7), and

$$g(x) = \frac{1}{\alpha}\{(1 - \omega)U(x - 1)e^{-(x-1)/\alpha} + \omega U(x - x_0)e^{-(x-x_0)/\alpha}\} \qquad (9)$$

Figure 1 shows $g(x)$ for $0 < x_0 < 1$. Physically, we note that the size distribution $g(x)$ exhibits two peaks corresponding to the sizes of the feed aerosol ($x = 1$) and that freshly formed by nucleation ($x = x_0$). The distribution spreads toward larger sizes because of the condensation growth. The peak at $x = 1$ is higher than it would be in the absence of nucleation because some of the smaller particles have grown by condensation to sizes ≥ 1 and augment the feed aerosol. The degree of spreading of the distribution is controlled by the dimensionless parameter α, which depends upon the residence time θ.

No Nucleation; $\omega = 0$

When nucleation may be neglected, (5) becomes

$$\alpha g'(x) = \frac{1}{2}\,\kappa g^* g(x) - (1 + \kappa)g(x) + g_0(x) \qquad (10)$$

Equation (10) can be solved by Laplace transformation (see appendix) to yield the general solution for the aerosol size distribution

$$g(x) = \sum_{k=0}^{\infty} \frac{\kappa^k[\phi_k^* g_0^{*(k+1)}](x)}{2^k k!(k + 1)!\alpha^{2k+1}} \qquad (11)$$

where

$$\phi_k(x) = x^{2k}e^{-(1+\kappa)x/\alpha} \qquad (12)$$

and g_0^{*k} denotes the k fold convolution of g_0 defined by

$$g_0^{*k}(x) = g_0^* g_0^* \cdots {}^* g_0(x)$$

Again, if the aerosol in the feed is monodisperse, so that $g_0(x) = (1 + 1/2\kappa)\delta(x - 1)$, then

$$g_0^{*(k+1)}(x) = \left(1 + \frac{1}{2}\kappa\right)^{k+1}\delta[x - (k + 1)] \qquad (13)$$

and

$$\phi_k^* g_0^{*(k+1)} = \left(1 + \frac{1}{2}\kappa\right)^{k+1}[x - (k + 1)]^{2k} \times U[x - (k + 1)]e^{-(1+\kappa)[x-(k+1)]/\alpha} \qquad (14)$$

The resulting distribution is thus

$$g(x) = \sum_{k=0}^{\infty} \frac{\kappa^k(1 + 1/2\kappa)^{k+1}[x - (k + 1)]^{2k}U[x - (k + 1)]e^{-(1+\kappa)[x-(k+1)]/\alpha}}{2^k k!(k + 1)!\alpha^{2k+1}} \qquad (15)$$

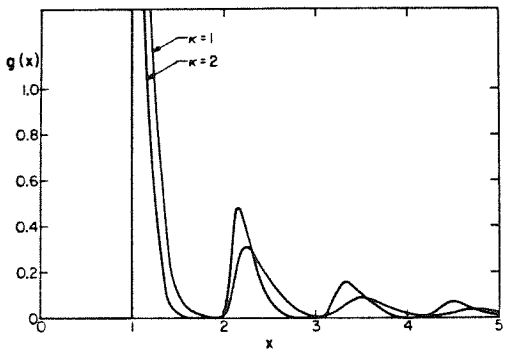


Figure 2. Dimensionless size distribution g(x) for monodisperse feed and no nucleation: $\kappa = 1$ and 2, $\alpha = 0.25$, $\omega = 0$.

To observe the physical significance of this solution, let us examine the case in which $\alpha << 1$, that is, when the characteristic time for condensation is large relative to the residence time. For small α, the function

$$U[x - (k+1)]\,[x - (k+1)]^{2k} e^{-(1+\kappa)(x-(k+1))/\alpha} / \alpha^{2k+1}$$

which appears in the k^{th} term of the series for $g(x)$ is sharply peaked near $x = k + 1$. The area beneath the peak is $(2k)!/(1+\kappa)^{2k+1}$, so that the above function behaves like $(2k)!\delta[x - (k+1)]/(1+\kappa)^{2k+1}$. It then follows that the distribution behaves like the solution of (5) when $\alpha = \omega = 0$:

$$g(x) = \sum_{k=0}^{\infty} \frac{\kappa^k (1 + 1/2\kappa)^{k+1} \delta[x - (k+1)]\,(2k)!}{2^k (k+1)!(1+\kappa)^{2k+1}} \qquad (16)$$

In fact, this solution consists simply of a series of spikes at $x = 1, 2, 3, \ldots$. Physically, this is plausible because coagulation is the only process by which the distribution evolves. Since the dimensionless feed distribution consists only of particles of size $x = 1$, only integer multiples of this size will result. If α is nonzero but still small, roughly the same distribution will be obtained, that is, peaked at the positive integers, but with some spreading toward larger sizes due to condensation. If α is sufficiently small, the peaks will be sharp and will not significantly overlap. Consequently, the distribution can be constructed by taking each term in the series individually. Thus, the peaks occur at approximately $x = (k+1) + 2k\alpha/(1+\kappa)$, $k = 0, 1, 2, \ldots$. The increased shift of the peaks at larger particle sizes results because, on the average, the larger particles have remained in the reactor longer and thus have had a longer time during which to grow. Figure 2 shows the distributions corresponding to $\alpha = 0.25$, and $\kappa = 1$ and 2.

General Solution

Now consider the case in which all processes occur. The general solution of (5) is given by (11), with g_0 replaced by $g_1 = g_0 + \omega\delta(x - x_0)$:

$$g(x) = \sum_{k=0}^{\infty} \frac{\kappa^k \{\phi_k^x[g_0 + \omega\delta(x - x_0)]\}^{*(k+1)}(x)}{2^k k!(k+1)!\alpha^{2k+1}} \qquad (17)$$

Once again, consider the monodisperse feed distribution (7), in which case the size distribution in the CSTR is

$$g(x) = \sum_{k=0}^{\infty} \frac{\kappa^k}{2^k k!(k+1)!\alpha^{2k+1}} \sum_{j=0}^{k+1} \binom{k+1}{j} \times \left(1 + \frac{1}{2}\kappa - \omega\right)^j \omega^{k+1-j} \times U\{x - [j + (k+1-j)x_0]\} \times \{x - [j + (k+1-j)x_0]\}^{2k} e^{-\frac{(1+\kappa)}{\alpha}(x - [j+(k+1-j)x_0])} \qquad (18)$$

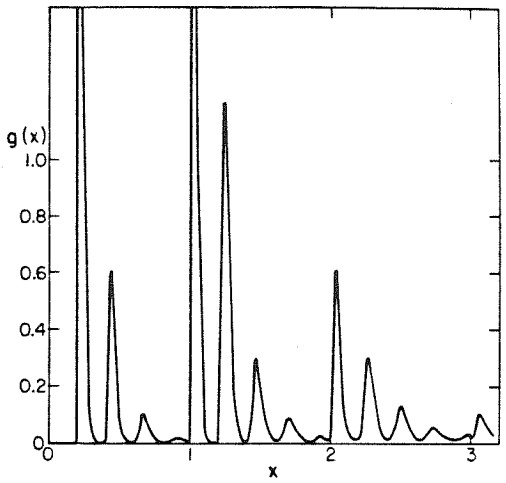


Figure 3. Dimensionless size distribution g(x) for monodisperse feed, including effects of coagulation, condensation and nucleation: $\kappa = 2$, $\alpha = 0.05$, $\omega = 1$, $x_0 = 0.2$.

As before, for small α (small condensation rate), the distribution approaches a sequence of delta functions

$$g(x) = \sum_{j=0}^{\infty} \sum_{k=j-1}^{\infty} \frac{\kappa^k (2k)!}{2^k k!(k+1)!(1+\kappa)^{2k+1}} \times \binom{k+1}{j} \left(1 + \frac{1}{2}\kappa - \omega\right)^j \omega^{k+1-j} \delta\{x - [j + (k+1-j)x_0]\} \qquad (19)$$

The peaks in the distribution (19) occur at all points $j + (k+1-j)x_0$, $k = 0, 1, 2, \ldots$, $j = 0, 1, 2, \ldots, k+1$. These points correspond to all possible sizes that can be formed by coagulation of particles initially of sizes 1 and x_0. For each value of k, the peaks at $j + (k+1-j)x_0$ represent the possible ways in which $k+1$ particles in a mixture of sizes 1 and x_0 may combine.

If $\alpha \neq 0$, the peaks are spread somewhat and shifted owing to condensation. Under the assumption that α is small enough so that the peaks do not overlap substantially, the peak at $x = j + (k+1-j)x_0$ is shifted to the right by $2k\alpha/(1+\kappa)$.

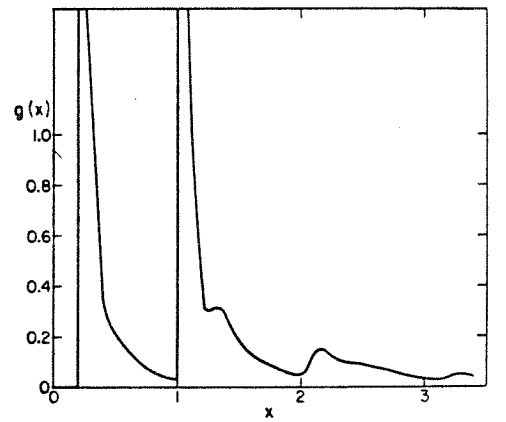


Figure 4. Dimensionless size distribution g(x) for monodisperse feed, including effects of coagulation, condensation and nucleation: $\kappa = 2$, $\alpha = 0.25$, $\omega = 1$, $x_0 = 0.2$.

It is interesting to note that for each k, the $k+1$ peaks at $j+(k+1-j)x_0$ for $j=0,1,\ldots,k+1$ are distributed binominally. This is precisely the distribution one would expect from random combination of particles of sizes 1 and x_0 originally in the ratio $N_0/\theta S_0 = (1 + 1/2\ \kappa-\omega)/\omega$. This phenomenon is illustrated in Figure 3, in which $N_0/\theta S_0=1$, and the peaks at 0.4, 1.2 and 2.0, corresponding to the three possible two-particle combinations, have the magnitude ratios 1:2:1. Similarly, the peaks near 0.7, 1.5, 2.3 and 3, corresponding to all three-particle combinations, have the ratios 1:3:3:1.

As α increases, eventually the peaks overlap, as shown in Figure 4. Only the peaks at x_0 and 1 retain their distinctiveness.

ANALYSIS FOR SIZE DEPENDENT GROWTH RATE

In the previous section, it was assumed that the condensation growth rate of a particle was independent of its size; that is, $I(v) = \sigma_0 c$. Although this assumption enabled us to elucidate the general features of the CSTR size distribution, particle growth rates are, in general, a function of particle size. For a monodisperse feed distribution we have seen that the effect of condensation with a size independent growth rate is to spread and shift to the right the peaks in the size distribution resulting from coagulation. The greater spreading of the distribution at larger sizes was seen to be due to the longer residence time of the larger particles. With a size dependent growth rate of the form $I(v) = c\sigma_\gamma v^\gamma$, $0 < \gamma \leq 1$, we expect an even greater spreading of the distribution at larger sizes, but without a qualitative change in the nature of the distribution. In this section we consider the growth rate $I(v) = c\sigma_1 v$, which represents the maximum growth rate possible. The physical basis for such a growth rate has been discussed by Gelbard and Seinfeld (1979). In short, one obtains this expression when particle growth is chemical reaction controlled.

The aerosol material balance is

$$c\sigma_1 v n'(v) + c\sigma_1 n(v) = \frac{1}{2}\beta n^* n(v) - \beta n(v)\int_0^\infty n(\bar{v})d\bar{v} + \frac{1}{\theta}[n_0(v) - n(v)] \quad (20$$

where we have now dropped the nucleation term because of the equivalence of feed and nucleation sources demonstrated previously.

The equation is nondimensionalized as before with the exception that the condensation parameter α is now defined as

$$\alpha = \theta c \sigma_1$$

In dimensionless form, (20) is

$$\alpha x g'(x) + \alpha g(x) = \frac{1}{2}\kappa g^* g(x) - (1 + \kappa)g(x) + g_0(x) \quad (21)$$

Equation (21) is not amenable to solution by Laplace transformation. Therefore we will explore the behavior of g for small α. The important qualitative features of the solution will emerge in this case. We consider the monodisperse feed distribution $g_0(x) = (1 + 1/2\ \kappa)\delta(x-1)$.

When $\alpha=0$, the solution is, as we know, entirely discrete. For small $\alpha>0$, there will be some spreading introduced into the distribution, the spreading increasing with particle size. Because of the fundamental change in the character of the solutions from $\alpha=0$ to $\alpha\neq 0$, the analysis of the case of small α leads to a singular perturbation problem. This behavior is also reflected in the fact that the perturbing term $\alpha x g'(x)$ is not small when x is near a natural number, since there $g'(x)$ is very large.

Since nonzero values of g are expected to occur in small regions about each natural number k, we replace the term $\alpha x g'(x)$ in (21) by $\alpha k g'(x)$ for x near k. To begin, let us obtain an expression for the size distribution of the first peak, $x=1$. Let $\psi_1(x)$ represent this function. Note that no particles which have undergone coagulation will be in the size range of the first peak. Consequently, to obtain ψ_1, we may drop the term $1/2\ g^* g(x)$ in (21). The term $xg'(x)$ is replaced by $g'(x)$ since $x=1$, and we drop the term $\alpha g(x)$ since it is negligible compared to $(1+\kappa)g(x)$. Hence, ψ_1 is governed by

$$\alpha\psi_1'(x) = -(1+\kappa)\psi_1(x) + (1 + \frac{1}{2}\kappa)\delta(x-1) \quad (22)$$

Equation (22) may be solved to yield

$$\psi_1(x) = \frac{\left(1+\frac{1}{2}\kappa\right)}{\alpha} e^{-\frac{(1+\kappa)}{\alpha}(x-1)} U(x-1) \quad (23)$$

The function ψ_1 of (23), describing a sharp peak at $x=1$, is, in fact, identical to the first term in the series (15). This is physically plausible, since near $x=1$ the constant and linear growth terms are nearly the same.

If, however, we choose not to replace $xg'(x)$ by $g'(x)$, the equation governing ψ_1 is

$$\alpha x\psi_1'(x) = -(1+\kappa)\psi_1(x) + \left(1 + \frac{1}{2}\kappa\right)\delta(x-1) \quad (24)$$

the solution of which is

$$\psi_1(x) = \frac{\left(1 + \frac{1}{2}\kappa\right)}{\alpha} x^{-\frac{(1+\kappa)}{\alpha}} U(x-1) \quad (25)$$

The function in (23) is, however, an asymptotic approximation to this function, for if $0 \leq x-1 \geq \sqrt{\epsilon}$, then $x^{-1/\epsilon} \sim e^{-(x-1)/\epsilon}$ as $\epsilon \to 0^+$. Outside this asymptotic region, both functions are essentially zero. For the sake of simplicity, however, (23) is the preferred expression, so we shall employ it in subsequent calculations.

We now turn to the calculation of ψ_2, which is the distribution near the peak at $x=2$. We set $x=2$ in the coefficient of the first term, in (21). Now we must retain the coagulation term, because coagulation is the principal mode by which particles of size $x \cong 2$ are formed. Nevertheless, this term can be simplified since the full term $1/2\ \kappa(\psi_1+\psi_2)^*(\psi_1+\psi_2)$ may be approximated by $1/2\kappa\psi_1^*\psi_1$ near $x=2$. Hence, the governing equation for ψ_2 is

$$2\alpha\psi_2'(x) = \frac{1}{2}\kappa\psi_1^*\psi_1(x) - (1+\kappa)\psi_2(x) \quad (26)$$

the solution of which is

$$\psi_2(x) = U(x-2)\left\{\frac{\kappa\left(1 + \frac{1}{2}\kappa\right)^2}{\alpha(1+\kappa)^2} e^{-\frac{(1+\kappa)}{2\alpha}(x-2)} - \frac{\kappa\left(1 + \frac{1}{2}\kappa\right)^2}{\alpha(1+\kappa)^2} e^{-\frac{(1+\kappa)}{\alpha}(x-2)} - \frac{\kappa\left(1 + \frac{1}{2}\kappa\right)^2}{2\alpha^2(1+\kappa)}(x-2)e^{-\frac{(1+\kappa)}{\alpha}(x-2)}\right\} \quad (27)$$

The calculation of ψ_3 and subsequent functions proceeds in an analogous manner. In the case of ψ_3, the coagulation term contains only the convolution products of ψ_1 and ψ_2, since only these contribute significantly to the peak near $x=3$. The equation for ψ_3 is

$$3\alpha\psi_3'(x) = \kappa\psi_1^*\psi_2(x) - (1+\kappa)\psi_3(x) \quad (28)$$

the solution of which is

$$\psi_3(x) = U(x-3)\left\{\frac{27\kappa^2\left(1+\frac{1}{2}\kappa\right)^3}{16\alpha(1+\kappa)^4} e^{-\frac{(1+\kappa)}{3\alpha}(x-3)}\right.$$

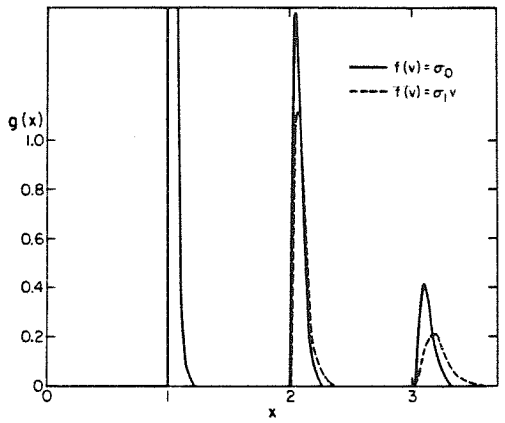


Figure 5. Comparison of dimensionless size distributions resulting from constant and linear condensation rates, $f(v) = \sigma_0$ and $f(v) = \sigma_1 v$, respectively, and monodisperse feed: $\kappa = 1$, $\alpha = 0.05$ in both cases.

$$+\frac{37\kappa^2\left(1+\frac{1}{2}\kappa\right)^3}{16\alpha(1+\kappa)^4}e^{-\frac{(1+\kappa)}{\alpha}(x-3)}-\frac{4\kappa^2\left(1+\frac{1}{2}\kappa\right)^3}{\alpha(1+\kappa)^4}e^{-\frac{(1+\kappa)}{2\alpha}(x-3)}$$

$$+\frac{7\kappa^2\left(1+\frac{1}{2}\kappa\right)^3}{8\alpha^2(1+\kappa)^3}(x-3)\,e^{-\frac{(1+\kappa)}{\alpha}(x-3)}$$

$$+\frac{\kappa^2\left(1+\frac{1}{2}\kappa\right)^3}{8\alpha^3(1+\kappa)^2}(x-3)^2e^{-\frac{(1+\kappa)}{\alpha}(x-3)}\Bigg\} \quad (29)$$

The sum $\psi_1 + \psi_2 + \psi_3$ represents the solution to (21) for the first three peaks in the distribution. The only limitation in this analysis is that the peaks must be sharp and the distributions about each peak must not overlap. As i increases, the peaks of the ψ_i tend to broaden, so that at a certain point this approach is no longer valid. Nevertheless, the first few peaks in the distribution provide an indication of the differences between the cases of size independent and size dependent particle growth rates. Figure 5 shows the distributions in these two cases. We see, as expected, that the effect of a linear size dependent growth rate is an acceleration of the broadening and a shift to the right of the peaks.

CONCLUSIONS

The qualitative features of aerosol formation and growth in a CSTR have been studied. Exact analytical solutions have been obtained to the aerosol balance equation in the case in which the kinetic coagulation coefficient and the particle growth rate by vapor condensation are independent of particle size. A perturbation solution has also been obtained for the case of linear volume dependent particle growth and a monodisperse feed aerosol.

The assumptions required to obtain these solutions are simplified, but they may nevertheless be useful in understanding the qualitative features of real systems, and in the case in which the feed is monodisperse and Brownian coagulation is occurring in the reactor, these solutions may, in fact, be reasonable approximations to the actual size distributions.

In addition, the important dimensionless groups governing aerosol behavior in a CSTR have been elucidated. These groups are ratios of characteristic times for coagulation, condensation and nucleation to the mean residence time and can be expected to play an important role in aerosol systems even when the simplifying assumptions employed here do not apply.

ACKNOWLEDGMENT

This work was supported by National Science Foundation grant PFR 76-04179.

APPENDIX: SOLUTION OF EQUATION (10)

The aerosol balance equation is

$$\alpha g'(x) = \frac{1}{2}\kappa g^* g(x) - (1+\kappa)g(x) + g_0(x) \quad (A1)$$

Application of the Laplace transform to (A1) gives

$$\alpha s G(s) = \frac{1}{2}\kappa G^2(s) - (1+\kappa)G(s) + G_0(s) \quad (A2)$$

where

$$G(s) \equiv \int_0^\infty e^{-sx} g(x)\,dx$$

Equation (A2) may be readily solved for $G(s)$, and we obtain

$$G(s) = (\xi - \sqrt{\xi^2 - 2\kappa G_0(s)})/\kappa \quad (A3)$$

where $\xi = (\alpha s + 1 + \kappa)$. Taylor expansion of $G(s)$ in powers of $G_0(s)$ gives

$$G(s) = \frac{G_0(s)}{\xi} + \sum_{k=0}^{\infty} \frac{(2k+1)!}{2^k k!(k+2)!}\frac{\kappa^{k+1}}{\xi^{2k+3}} G_0^{k+2}(s) \quad (A4)$$

If we denote by ϕ the function defined by

$$\phi_k(x) \equiv x^{2k} e^{-\frac{(1+\kappa)}{\alpha}x} \quad (A5)$$

then we see that

$$\phi_k = (2k)!\,\phi_0^{*(2k+1)} \quad (A6)$$

where

$$\phi_0(x) \equiv e^{-\frac{(1+\kappa)}{\alpha}x} \quad (A7)$$

and ϕ_0^{*k} denotes the k fold convolution of ϕ_0. Since under the Laplace transform, $\phi_0 \to \alpha/\xi$, we see that

$$\phi_0^{*(2k+2)} \to \frac{\alpha^{2k+3}}{\xi^{2k+3}}$$

or, equivalently, that

$$\frac{\phi_{k+1}}{(2k+2)!\alpha^{2k+3}} \to \frac{1}{\xi^{2k+3}}$$

Thus, termwise inversion of (A4) gives

$$g(x) = \frac{1}{\alpha}(g_0^* \phi_0)(x) + \sum_{k=1}^{\infty} \frac{\kappa^k[\phi_k^* g_0^{*(k+1)}](x)}{2^k k!(k+1)!\alpha^{2k+1}}$$

$$= \sum_{k=0}^{\infty} \frac{\kappa^k[\phi_k^* g_0^{*(k+1)}](x)}{2^k k!(k+1)!\alpha^{2k+1}} \quad (A8)$$

NOTATION

c = vapor concentration, g-mole cm^{-3}
c_0 = vapor concentration in feed, g-mole cm^{-3}
$f(v)$ = volume function appearing in $I(v)$, $\sigma_\gamma v^\gamma$, $\mu m^3 s^{-1}$ (g-mole cm^{-3})$^{-1}$
$g(x)$ = dimensionless size distribution function

$I(v)$ = rate of change of the volume of a particle by vapor condensation, $\mu m^3 s^{-1}$
M = molecular weight of the gas, g mole^{-1}
$n(v)$ = size distribution function μm^{-3} cm^{-3}
$n_0(v)$ = size distribution function of feed aerosol, μm^{-3} cm^{-3}
N = total aerosol number concentration, cm^{-3}
N_0 = total aerosol number concentration in feed, cm^{-3}
q = volumetric flow rate through CSTR, cm^{3}s^{-1}
S_0 = rate of formation of particles by nucleation, cm^{-3} s^{-1}
$U(x)$ = unit step function
v = particle volume, μm^3
v_f = volume of feed aerosol, μm^3
v_0 = volume of particles formed by nucleation, μm^3
V = CSTR volume, cm^3
$\bar{V}$ = total aerosol volume concentration, $\mu m^3 cm^{-3}$
$\bar{V}_0$ = total aerosol volume concentration in feed, $\mu m^3 cm^{-3}$
x = dimensionless particle volume, $N_0 v/\bar{V}_0$
x_0 = $N_0 v_0/\bar{V}_0$

Greek Letters

α = $\theta c_0 \sigma_0 N_0/\bar{V}_0$ or $\theta c_0 \sigma_1$
$\beta(v,\tilde{v})$ = coagulation coefficient, cm^{3}s^{-1}
γ = exponent in growth rate expression
$\delta(v)$ = Dirac delta function
$\delta(x)$ = $\bar{V}_0 \delta(v)/N_0$
θ = V/q
κ = $\beta\theta N$
ρ = liquid density of aerosol particles, g cm^{-3}
σ_γ = constant in expression for $f(v)$, $\mu m^{3-3\gamma}$ s^{-1}(g-mole cm^{-3})$^{-1}$
$\psi_i(x)$ = size distribution around i^{th} peak in the size distribution
ω = $\theta S_0/N$
$*$ = convolution operator

LITERATURE CITED

Gelbard, F., and J. H. Seinfeld, "Exact Solution of the General Dynamic Equation for Aerosol Growth by Condensation," *J. Colloid Interface Sci.*, **68**, 173 (1979).
Ramabhadran, T. E., T. W. Peterson, and J. H. Seinfeld, "Dynamics of Aerosol Coagulation and Condensation," *AIChE J.*, **22**, 840 (1976).

Manuscript received October 1, 1979, revision received February 13, and accepted February 29, 1980.

CHAPTER 4

TURBULENT DEPOSITION AND GRAVITATIONAL SEDIMENTATION OF AEROSOL IN A VESSEL OF ARBITRARY SHAPE

Published in Journal of Aerosol Science 12, 405(1981)

J. Aerosol Sci., Vol. 12, No. 5, pp. 405–415, 1981.
Printed in Great Britain.

0021–8502/81/050405–11 $02.00/0

TURBULENT DEPOSITION AND GRAVITATIONAL SEDIMENTATION OF AN AEROSOL IN A VESSEL OF ARBITRARY SHAPE

JAMES G. CRUMP and JOHN H. SEINFELD

Department of Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, U.S.A.

(*Received* 23 *June* 80 *and in revised form* 9 *March* 1981)

Abstract—A general formula for the rate of aerosol deposition due to turbulent diffusion, Brownian diffusion, and gravitational sedimentation in a turbulently mixed, enclosed vessel of arbitrary shape is derived. It is shown to reduce to the formula of Corner and Pendlebury *Proc. Phys. Soc.* **B64**, 645 (1951) for a cubical vessel. The result is essentially independent of the form of the eddy diffusivity near the wall and demonstrates that for vessels with non-vertical walls, sedimentation and diffusion are intimately coupled in determining the total deposition rate. The effect of inertia is estimated and used to assess the range of particle sizes for which the analysis is applicable.

INTRODUCTION

We consider the problem of predicting the rate of wall deposition of an aerosol in a turbulently mixed, enclosed vessel of arbitrary shape. Such predictions are of importance in understanding the losses of aerosol that occur in chemical reactors and other chambers. Losses of aerosol in an enclosed vessel result from deposition due to Brownian diffusion and turbulent transport to the walls and from gravitational settling. The classical result for this problem is that of Corner and Pendlebury (1951) who derived a formula for the aerosol deposition rate in a rectangular box.* In this work we consider the problem for a vessel of arbitrary shape. It will be shown that the result of Corner and Pendlebury is simply a special case of that obtained here. Because of the practical relevance of a spherical vessel, we first solve that problem completely before presenting the solution for the arbitrarily shaped enclosure. Both the spherical and rectangular cases, of course, fall out as special cases of the general result. The effect of vessel shape has not generally been accounted for in estimating aerosol deposition rates, and we will show that the vessel shape affects the coupling between Brownian and turbulent diffusion and sedimentation. Of particular interest is clarification of the assumptions on which the analysis rests.

The total aerosol deposition rate (particles s^{-1}) is computed from the product of the local deposition flux and the internal surface area of the vessel. The deposition coefficient per unit volume of the vessel, β, is then defined as the ratio of the deposition rate to the total number of particles in the vessel. It is frequently suggested that the deposition coefficient per unit volume should have the form

$$\beta = \frac{v}{h} + \frac{SD}{V\sigma}$$

where v is the terminal particle settling velocity, h is the vessel height, S and V the surface area and volume of the vessel, D the particle Brownian diffusivity, and σ the diffusion boundary layer thickness at the vessel wall. We show later that under certain circumstances β may be shown to reduce to this expression, but that in general the above simple expression for β is inadequate.

* The total loss rate, which results from both Brownian and turbulent transport and gravitational settling, is usually referred to simply as the deposition rate. In this context, therefore, the deposition rate will be taken as synonymous with the total wall loss rate. When discussing that contribution to the deposition rate arising from Brownian and turbulent transport, we will be careful to delineate such.

In the next section we consider the problem of deposition in a spherical vessel. In the section following we then generalize to a vessel of arbitrary shape.

TURBULENT DEPOSITION AND GRAVITATIONAL SEDIMENTATION OF AN AEROSOL IN A SPHERICAL VESSEL

We begin by considering the case of a sphere. The result for an arbitrary vessel is obtained by essentially the same reasoning, and is presented subsequently. The aerosol is assumed to be perfectly mixed in a turbulent core of the sphere, and only in a small (turbulent) boundary layer do gradients in concentration occur. These gradients give rise to deposition by both turbulent and Brownian diffusion. In addition, gravity imposes an overall downward drift on the aerosol particles, leading to enhanced deposition on the bottom of the vessel.

Let R denote the radius of the sphere, θ the polar angle, and suppose the turbulent boundary layer has thickness δ. It will turn out that the exact value of δ is of no consequence, but we must make two assumptions concerning it. The first is that $\delta/R \ll 1$. This is plausible since we assume at the outset that there is turbulent mixing in the vessel.

In the boundary layer we assume there exist three modes of transport of particles—turbulent diffusion, Brownian diffusion, and gravitational sedimentation. The flux due to molecular diffusion is given by $-D\nabla c$, where c is particle concentration and D is the Brownian diffusivity. That for turbulent diffusion is assumed to be represented by $-D_e\nabla c$, where D_e is an eddy diffusivity, which depends on position, as well as, possibly, particle size. We assume the average gas velocity near the walls is negligible. Then the effect of gravitational sedimentation is accounted for by the term $v\mathbf{k}\cdot\nabla c$, where v is the terminal settling velocity and $\mathbf{k}$ the unit normal vector in the vertical direction.

We assume a quasi-steady state in the boundary layer. That is, we take the concentration in the turbulent core to be constant. Under these assumptions the aerosol concentration in the boundary layer is governed by

$$\nabla\cdot[(D+D_e)\nabla c]-v\mathbf{k}\cdot\nabla c=0 \tag{1}$$

for $R-\delta<r<R$. The boundary conditions on equation (1) are

$$\begin{aligned} c &= 0 \qquad & r &= R \\ c &= c_0 \qquad & r &= R-\delta \end{aligned} \tag{2}$$

where c_0 is the bulk concentration. For the eddy diffusivity, we use Prandtl's mixing length expression near a wall, (Corner and Pendlebury, 1951) $D_e = k_e x^2$, where x denotes distance from the wall, and k_e may be evaluated from the turbulent energy dissipation rate. The eddy diffusivity for aerosol particles should also depend upon particle size, since larger particles do not exactly follow the flow. However, Hinze (1959), following Tchen (1947), has solved the equation of motion for a particle moving in a homogeneous turbulent flow field, and has shown that in the long time limit particle and fluid eddy diffusivities are equal. Experiments carried out by Kalinske and Pien (1943) and by Rouse (1939) for eddy diffusion and settling of sand in water indicate little variation of eddy diffusivity with particle size. An explanation for this phenomenon, advanced by Fuchs (1964, p. 263), is that although larger particles do not follow the fluid motion as closely as small particles, their motion is more persistent due to their larger mass. This effect tends to equalize the average distance moved by both large and small particles over sufficiently long times. Nevertheless, we anticipate inertial effects to become important for large particles or strong turbulence, and the magnitude of these are estimated later to assess the validity of the above theory.

We begin the analysis by making the assumption that $D \ll k_e\delta^2$. This assumption is merely the statement that at the outer edge of the boundary layer turbulent diffusion dominates Brownian diffusion.

In polar coordinates equation (1) becomes

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2(D+D_e)\frac{\partial c}{\partial r}\right)+\frac{(D+D_e)}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial c}{\partial\theta}\right)+v\cos\theta\frac{\partial c}{\partial r}-\frac{v\sin\theta}{r}\frac{\partial c}{\partial\theta}=0 \tag{3}$$

Setting $x = R - r$, equation (3) assumes the form

$$\frac{1}{(R-x)^2}\frac{\partial}{\partial x}\left((R-x)^2(D+D_e)\frac{\partial c}{\partial x}\right)+\frac{(D+D_e)}{(R-x)^2\sin\theta}\frac{\partial}{\partial\theta}$$
$$\times\left(\sin\theta\frac{\partial c}{\partial\theta}\right)-v\cos\theta\frac{\partial c}{\partial x}-\frac{v\sin\theta}{(R-x)}\frac{\partial c}{\partial\theta}=0. \tag{4}$$

Now we introduce dimensionless variables by setting $z = x\sqrt{k_e/D}$ and $u = c/c_0$, giving

$$\frac{1}{\left(1-z\sqrt{\frac{D}{k_eR^2}}\right)^2}\frac{\partial}{\partial z}\left[\left(1-z\sqrt{\frac{D}{k_eR^2}}\right)^2(1+z^2)\frac{\partial u}{\partial z}\right]+\frac{D}{k_eR^2}\frac{(1+z^2)}{\left(1-z\sqrt{\frac{D}{k_eR^2}}\right)^2\sin\theta}\frac{\partial}{\partial\theta}$$
$$\times\left(\sin\theta\frac{\partial u}{\partial\theta}\right)-\frac{v}{\sqrt{k_eD}}\cos\theta\frac{\partial u}{\partial z}-\frac{v}{k_eR}\frac{\sin\theta}{\left(1-z\sqrt{\frac{D}{k_eR^2}}\right)}\frac{\partial u}{\partial\theta}=0. \tag{5}$$

Now, in the boundary layer $x \le \delta$, so $z \le \delta\sqrt{k_e/D}$, and hence $z\sqrt{\frac{D}{k_eR^2}} \le (\delta\sqrt{k_e/D})$ $\times\sqrt{\frac{D}{k_eR^2}} = \delta/R \ll 1$, so we may ignore $z\sqrt{\frac{D}{k_eR^2}}$ in comparison to 1, leaving the simplified equation

$$\frac{\partial}{\partial z}\left[(1+z^2)\frac{\partial u}{\partial z}\right]+\frac{D}{k_eR^2}\frac{(1+z^2)}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right)-\frac{v\cos\theta}{\sqrt{k_eD}}\frac{\partial u}{\partial z}-\frac{v\sin\theta}{Rk_e}\frac{\partial u}{\partial\theta}=0. \tag{6}$$

At $z = 0$, $u = 0$, and at $z = \delta\sqrt{k_e/D}$, $u = 1$, so $\frac{\partial u}{\partial z}\sim\frac{1}{\delta}\sqrt{D/k_e}$. Since at $z = 0$ and $z = \delta\sqrt{k_e/D}$, u is constant, we expect $\left|\frac{\partial u}{\partial\theta}\right| \le O(1)$. Thus, the second term of equation (6) has order $D/k_eR^2 = (D/k_e\delta^2)(\delta/R)^2 \ll 1$, so is completely negligible. The third term has order $\left(v/\sqrt{k_eD}\right)\left(\frac{1}{\delta}\sqrt{D/k_e}\right) = v/\delta k_e$, while the fourth has order v/Rk_e. The ratio of the fourth to the third terms is of order $(v/Rk_e)/(v/\delta k_e) = \delta/R \ll 1$. Hence we drop the fourth term of equation (6) in favor of the third, leaving

$$\frac{\partial}{\partial z}\left[(1+z^2)\frac{\partial u}{\partial z}\right]-\frac{v\cos\theta}{\sqrt{k_eD}}\frac{\partial u}{\partial z}=0 \tag{7}$$

together with the boundary conditions

$$\begin{aligned} u(0,\theta) &= 0 \\ u(\delta\sqrt{k_e/D},\theta) &= 1. \end{aligned} \tag{8}$$

The solution of equations (7) and (8) is

$$u(z,\theta)=\frac{\exp\left[\frac{v\cos\theta}{\sqrt{k_eD}}\tan^{-1}z\right]-1}{\exp\left[\frac{v\cos\theta}{\sqrt{k_eD}}\tan^{-1}(\delta\sqrt{k_e/D})\right]-1}. \tag{9}$$

Recalling that $\delta^2k_e/D \gg 1$, it follows that $\delta\sqrt{k_e/D} \gg 1$, so that $\tan^{-1}(\delta\sqrt{k_e/D}) \cong \pi/2$.

Thus, we have

$$u(z,\theta)=\frac{\exp\left[\dfrac{v\cos\theta}{\sqrt{k_eD}}\tan^{-1}z\right]-1}{\exp\left[\dfrac{\pi v\cos\theta}{2\sqrt{k_eD}}\right]-1} \tag{10}$$

The deposition flux at the surface is given by

$$D\left.\frac{\partial c}{\partial x}\right|_{x=0}=c_0D\left.\frac{\partial u}{\partial x}\right|_{x=0}=\frac{c_0v\cos\theta}{\exp\left[\dfrac{\pi v\cos\theta}{2\sqrt{k_eD}}\right]-1} \tag{11}$$

and the total deposition rate is

$$2\pi R^2\int_0^{\pi}\frac{c_0v\cos\theta\sin\theta\,d\theta}{\exp\left[\dfrac{\pi v\cos\theta}{2\sqrt{k_eD}}\right]-1}. \tag{12}$$

The deposition coefficient per unit volume is

$$\beta=\frac{3}{2R}\int_0^{\pi}\frac{v\cos\theta\sin\theta\,d\theta}{\exp\left[\dfrac{\pi v\cos\theta}{2\sqrt{k_eD}}\right]-1}. \tag{13}$$

Setting $x=\dfrac{\pi v}{2\sqrt{k_eD}}$, equation (13) becomes

$$\beta=\frac{3\sqrt{k_eD}}{\pi R}\left[\frac{2}{x}\int_0^x\frac{t\,dt}{e^t-1}+\frac{1}{2}x\right] \tag{14}$$

where we have used the fact that

$$\frac{1}{x}\int_0^x\frac{t\,dt}{1-e^{-t}}=\frac{1}{x}\int_0^x\frac{t\,dt}{e^t-1}+\frac{1}{2}x.$$

Therefore, we may write the deposition coefficient β as

$$\beta=\frac{3\sqrt{k_eD}}{\pi R}\left[2D_1(x)+\frac{1}{2}x\right] \tag{15}$$

where D_1 denotes the Debye function (Abramowitz and Stegun, 1964),

$$D_1(x)=\frac{1}{x}\int_0^x\frac{t\,dt}{e^t-1}. \tag{16}$$

Figure 1 shows $\beta\pi R/3\sqrt{k_eD}$ as a function of x. Note that β divides into two terms

$$\beta=\frac{6\sqrt{k_eD}}{\pi R}D_1\left(\frac{\pi v}{2\sqrt{k_eD}}\right)+\frac{v}{4R/3}. \tag{17}$$

The second term in equation (17) corresponds to the gravitational sedimentation term in the formula $\beta=v/h+SD/V\sigma$, since for a sphere $4R/3$ is the average height. Note that the first term, however, also depends on v, so is not a purely diffusive contribution. If $x\gg 1$, since $D_1(x)\sim\dfrac{\pi^2}{6x}$ for large x, we have

$$\beta\sim\frac{2k_eD}{Rv}+\frac{v}{4R/3},\quad x\gg 1 \tag{18}$$

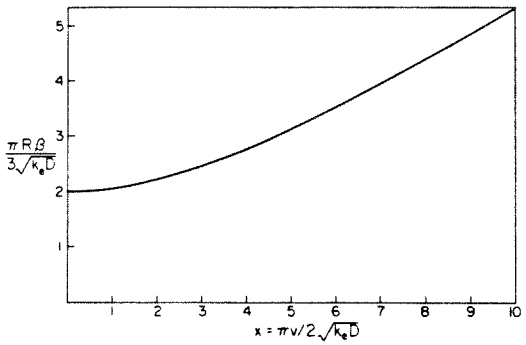


Fig. 1. Dimensionless deposition coefficient.

On the other hand, Corner and Pendlebury's result for a cube of side L,

$$\beta = \frac{1}{L}\left[\frac{8\sqrt{k_e D}}{\pi} + v\coth\left(\frac{x}{2}\right)\right] \tag{19}$$

has the property that for $x \gg 1$,

$$\beta \sim \frac{8\sqrt{k_e D}}{\pi L} + \frac{v}{L}. \tag{20}$$

As can be seen in equation (20), the sedimentation and diffusion terms actually are separate if $x \gg 1$ for an aerosol in a box. In this case the Fuchs formula holds provided we choose $\sigma = (3\pi/4)\sqrt{D/k_e}$. Hence, the diffusion boundary layer thickness must depend on particle size, as Harrison (1979) has shown experimentally.

In the case of a sphere equation (18) shows that diffusion and sedimentation terms do not separate for $x \gg 1$. Because of the inclined surface of the sphere, sedimentation is always coupled to the diffusion process, whereas this is not the case in a vessel having only vertical and horizontal sides.

The coefficient β depends on particle size, and in Fig. 2, β is shown as a function of particle

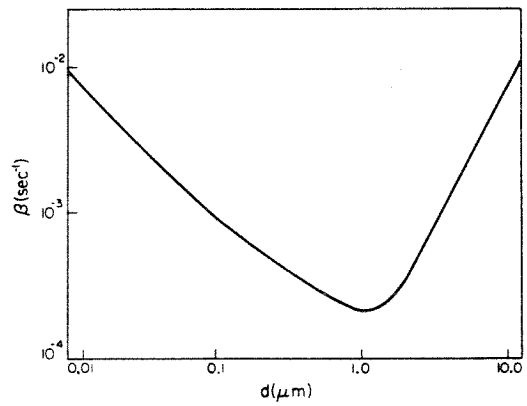


Fig. 2. Deposition coefficient as a function of particle size for $k_e = 36\,s^{-1}$, $R = 30\,cm$, and a water aerosol in air at room temperature.

size. In this figure k_e has been taken as $36\,s^{-1}$, corresponding to a value encountered in some of the experiments of Okuyama *et al.* (1977), in which a cylindrical stirred vessel was used. Also $R = 30\,cm$, and all other properties correspond to a water aerosol in air at room temperature. Note the minimum in β. Larger particles sediment rapidly, while smaller particles diffuse more rapidly. This does not, however, result in a cancellation of the size dependence of β since these two counteracting effects operate in distinct size regimes. In Figure 3 a hypothetical diffusion boundary layer thickness defined by the Fuchs formula is

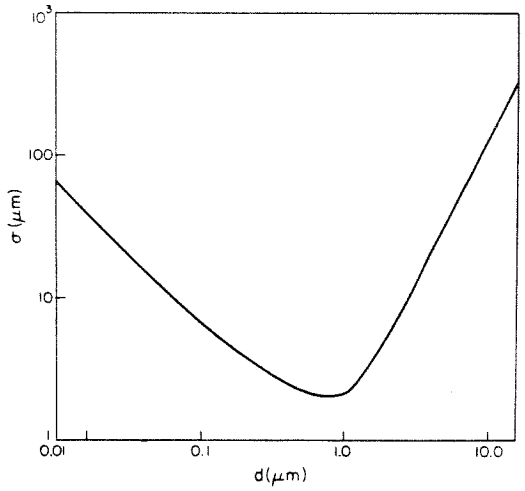


Fig. 3. Hypothetical Brownian diffusion boundary layer thickness calculated from the Fuchs formula and deposition coefficient of Fig. 2.

shown. Note that it exhibits behavior similar to that of β. Figure 4 shows the dependence of β on k_e.

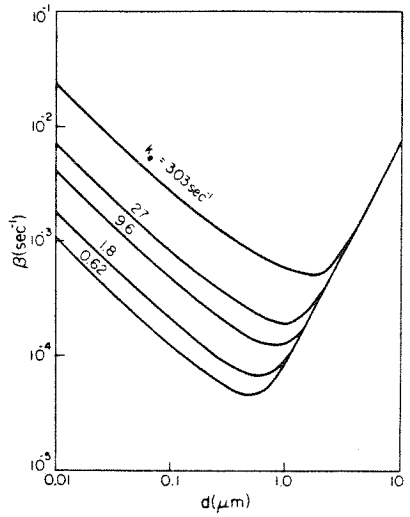


Fig. 4. Deposition coefficient as a function of particle size for various values of eddy diffusivity and same conditions as in Fig. 2.

EFFECT OF PARTICLE INERTIA

If the particles are large enough so that the average stopping distance s exceeds the turbulent boundary layer thickness, then the particles essentially coast through the boundary layer to the wall. In the foregoing analysis we have assumed that the principal mode of transport through the boundary layer is turbulent and Brownian diffusion. We can account for inertial effects by considering a model in which particles are essentially captured within a distance s of the wall, and transport of particles up to this point is accomplished by turbulent and Brownian diffusion. This model is crude, but will allow us to assess the validity of the non-inertial theory.

So, let $s \ll \delta$. We solve the transport equations as before but with the new boundary condition

$$c(s, \theta) = 0 \tag{21}$$

that gives the dimensionless concentration distribution,

$$u(z,\theta)=\frac{\exp\left(\frac{v\cos\theta}{\sqrt{k_eD}}\tan^{-1}z\right)-\exp\left(\frac{v\cos\theta}{\sqrt{k_eD}}\tan^{-1}(s\sqrt{k_e/D})\right)}{\exp\left(\frac{\pi v\cos\theta}{2\sqrt{k_eD}}\right)-\exp\left(\frac{v\cos\theta}{\sqrt{k_eD}}\tan^{-1}(s\sqrt{k_e/D})\right)}.$$

The flux is now given by

$$(D+k_es^2)\left.\frac{\partial c}{\partial x}\right|_{x=s}=\frac{c_0v\cos\theta\exp\left(\frac{v\cos\theta}{\sqrt{k_eD}}\tan^{-1}(s\sqrt{k_e/D})\right)}{\exp\left(\frac{\pi v\cos\theta}{2\sqrt{k_eD}}\right)-\exp\left(\frac{v\cos\theta}{\sqrt{k_eD}}\tan^{-1}(s\sqrt{k_e/D})\right)}. \tag{22}$$

This becomes equal to (11) provided the following conditions hold:

$$s\ll\delta \tag{23a}$$

$$\frac{v}{\sqrt{k_eD}}\tan^{-1}(s\sqrt{k_e/D})\ll 1. \tag{23b}$$

In general the particle size range over which the theory presented previously will hold depends on the degree of turbulence. As an example, we estimate the above quantities for two sets of conditions in a spherical vessel of radius 30 cm stirred by a six blade impeller, of blade length 10 cm and width 2 cm. For such a vessel Bates *et al.* (1966) give the power correlation

$$p=4\rho n^3d^5 \tag{24}$$

where ρ = fluid density, n = rotational speed of stirrer, and d = blade length. If we assume complete turbulent dissipation, we obtain the turbulent dissipation rate ε as

$$\varepsilon=4n^3d^5/V \tag{25}$$

where V = total vessel volume.

Following Okuyama *et al.* (1977), we set $k_e=0.4(2\varepsilon/16\nu)^{1/2}$ where ν is the kinematic viscosity of fluid. We estimate the stopping distance of particles using the Stokes drag with Cunningham correction and the velocity

$$u=30\left(\overline{\frac{du}{dx}}\right)=30(2\varepsilon/15\,\nu)^{1/2}. \tag{26}$$

That is, we take the size of the vessel as an upper bound on the characteristic length scale of the turbulence.

For the case of particles of 0.1 μm radius, unit density (g/cm^3), and a stirring speed of 500 r.p.m., the above equations yield

$$s\sqrt{k_e/D}=0.17 \tag{27}$$

$$\frac{v}{\sqrt{k_eD}}\tan^{-1}(s\sqrt{k_e/D})=1.4\times10^{-3} \tag{28}$$

Now, we know that $\delta\sqrt{k_e/D}\gg 1$, so (27) shows that $s\ll\delta$. Equation (28) shows that the other condition (23b) holds. In fact, since for particles of 0.1 μm radius we have $\frac{v}{\sqrt{k_eD}}=8.12\times10^{-3}$ and $\tan^{-1}x\le\frac{\pi}{2}$ for all x, inertia does not play an important role until $s\sqrt{k_e/D}$ becomes significantly larger than one. Since $s\sqrt{k_e/D}=1$ at a rotational speed of 1100 r.p.m., inertial effects are insignificant until the stirrer speed greatly exceeds 1100 r.p.m.

At lower levels of turbulence, the analysis is valid for larger particles. For example, at 100 r.p.m. with 0.5 μm diameter particles we obtain

$$s\sqrt{k_e/D} = 0.17 \tag{29}$$

$$\frac{v}{\sqrt{k_e D}} \tan^{-1}(s\sqrt{k_e/D}) = 0.18. \tag{30}$$

TURBULENT DEPOSITION AND GRAVITATIONAL SEDIMENTATION OF AN AEROSOL IN A VESSEL OF ARBITRARY SHAPE

We have just considered deposition and sedimentation in a sphere. However, the same analysis applies to any vessel, provided the boundary layer thickness δ is small relative to a characteristic dimension of the surface, and relative to the sum $\left(\frac{1}{r_1}+\frac{1}{r_2}\right)$, where r_1 and r_2 are the principal radii of curvature at a point; and again, provided $D \ll k_e\delta^2$. In this case, by choosing the x-coordinate to lie along the inward unit normal to the surface, and the other two coordinates to be tangent to the surface we obtain the same equations for the concentration in the boundary layer with the x variable chosen as given above. The term $\cos\theta$ for the case of the sphere is replaced by $\mathbf{n}(y)\cdot\mathbf{k}$, where $\mathbf{k}$ is the unit vector in the vertical direction, $\mathbf{n}(y)$ is the unit outward normal to the surface. The details of this calculation are shown in the Appendix. The result is that β is given by

$$\beta = \frac{1}{V}\int_{\Sigma} \frac{v\mathbf{n}(x)\cdot\mathbf{k}\,\mathrm{d}A(x)}{\exp\left[\dfrac{\pi v\mathbf{n}(x)\cdot\mathbf{k}}{2\sqrt{k_e D}}\right]-1} \tag{31}$$

where Σ denotes the surface and V the volume of the vessel, and $\mathrm{d}A(x)$ is the differential area element on Σ. If $\mathbf{n}(x)\cdot\mathbf{k} = 0$, the integrand is to be interpreted as the limit as $\mathbf{n}(x)\cdot\mathbf{k} \to 0$, which is easily seen to be $2\sqrt{k_e D}/\pi$.

The above formula is actually a generalization of the formula obtained by Corner and Pendlebury (1951) for a cube of side L. To see this, note that on the vertical sides of the cube, $\mathbf{n}(x)\cdot\mathbf{k} = 0$, so the integrand is $2\sqrt{k_e D}/\pi$, and the integral over the four vertical sides is $8L^2\sqrt{k_e D}/\pi$. On the top, $\mathbf{n}(x)\cdot\mathbf{k} = 1$, so the integral is $vL^2/[\exp(\pi v/2\sqrt{k_e D})-1]$. On the bottom $\mathbf{n}(x)\cdot\mathbf{k} = -1$, and the integral is thus, $-vL^2/[\exp(\pi v/2\sqrt{k_e D})-1]$. The total integral over the surface of the cube is then

$$L^2\left[\frac{8\sqrt{k_e D}}{\pi} + v\coth\left(\frac{\pi v}{4\sqrt{k_e D}}\right)\right].$$

Hence, dividing by the volume, L^3, of the cube gives

$$\beta = \frac{1}{L}\left[\frac{8\sqrt{k_e D}}{\pi} + v\coth\left(\frac{\pi v}{4\sqrt{k_e D}}\right)\right] \tag{32}$$

the formula of Corner and Pendlebury (1951).

The result may be expected to hold provided first that our approximation of homogeneous turbulence is adequate, and second, if the turbulence is sufficient to insure that the concentration is uniform except in a boundary layer whose size is small compared to the dimensions and radii of curvature of the vessel. The inequality $D \ll k_e\delta^2$ then follows, because in order for the concentration to remain fixed at the edge of the boundary layer, turbulence must intercede, since aerosol diffusion coefficients are usually quite small. Finally, we require that the inequalities (23a) and (23b) hold.

It is also worth noting that the particular form of D_e chosen is not too important. If we

choose $D_e = k_e x^3$ as suggested by Friedlander (1977) and assume $D \ll k_e \delta^3$ then similar analysis leads to

$$u(z,\theta) = \frac{\left[\frac{(1+z)^3}{(1+z^3)}\right]\frac{v\cos\theta}{6(k_eD^2)^{1/3}}\exp\left\{\frac{v\cos\theta}{\sqrt{3}(k_eD^2)^{1/3}}\left[\tan^{-1}\frac{2x-1}{\sqrt{3}}+\frac{\pi}{6}\right]\right\}-1}{\exp\left[\frac{2\pi v\cos\theta}{3\sqrt{3}(k_eD^2)^{1/3}}\right]-1} \tag{33}$$

where $z = x(k_e/D)^{1/3}$. Then the surface flux is given by

$$\frac{v\cos\theta}{\exp\left[\frac{2\pi v\cos\theta}{3\sqrt{3}(k_eD^2)^{1/3}}\right]-1} \tag{34}$$

which is very similar to our previous expression.

In fact, if we assume $D_e = k_e x^n$, the surface flux is given by

$$\frac{v\cos\theta}{\exp\left[\frac{\pi v\cos\theta}{(n\sin\frac{\pi}{n})\sqrt[n]{k_eD^{n-1}}}\right]} \tag{35}$$

and the loss coefficient, β, is given by

$$\beta = \frac{3n\sin\frac{\pi}{n}\sqrt[n]{k_eD^{n-1}}}{2\pi R}\left[2D_1(x)+\frac{1}{2}x\right] \tag{36}$$

where

$$x = \frac{\pi v}{(n\sin\frac{\pi}{n})\sqrt[n]{k_eD^{n-1}}}. \tag{37}$$

For vessels of arbitrary shape,

$$\beta = \frac{1}{V}\int_\Sigma \frac{v\mathbf{n}(x)\cdot\mathbf{k}\,dA(x)}{\exp\left[\frac{\pi v\mathbf{n}(x)\cdot\mathbf{k}}{(n\sin\frac{\pi}{n})\sqrt[n]{k_eD^{n-1}}}\right]-1}. \tag{38}$$

CONCLUSIONS

We have derived a formula for the rate of aerosol deposition on the walls of a turbulently mixed, enclosed vessel of arbitrary shape under the assumption of homogeneous turbulence near the walls. The result is shown to reduce to the previously obtained results of Corner and Pendlebury (1951) and Fuchs (1964) under appropriate conditions. We have also noted that the result is essentially independent of functional form of the eddy diffusivity near the wall. In vessels with non-vertical sides, sedimentation and diffusion are intimately coupled, and the simple formula $\beta = v/h + SD/V\sigma$ is inadequate. Finally, we have given conditions on the particle stopping distance under which the analysis can be expected to hold.

Acknowledgement—This research was supported in part by a grant from the Exxon Education Foundation.

REFERENCES

Abramowitz, M. and Stegun, I. A., Editors (1965) *Handbook of Mathematical Functions*, Dover, New York.
Bates, R. L., Fondy, P. L. and Fenic, J. G. (1966) in *Mixing, Theory and Practice*, (Edited by V. W. Uhl and J. B. Gray), Academic Press, New York.
Corner, J. and Pendlebury, E. D. (1951) *Proc. Phys. Soc.* **B64**, 645.
Friedlander, S. K. (1977) *Smoke, Dust and Haze*, John Wiley, New York.
Fuchs, N. A. (1964) *The Mechanics of Aerosols*, Pergamon Press, Oxford.
Harrison, A. W. (1979) *J. Colloid Interface Sci.* **69**, 563.
Hinze, J. O. (1959) *Turbulence*, McGraw-Hill, New York.
Kalinske, A. A. and Pien, C. L. (1943) *Trans. Am. Geophys. Union* **24**, Part II, 530.
Okuyama, K., Kousaka, Y., Kida, Y. and Toshida, T. (1977) *J. Chem. Engng Japan*, **10**, 142.
Rouse, H. (1939) in *Proceedings of the Fifth International Congress of Applied Mechanics*, John Wiley, New York 550.
Tchen, C. M. (1947) *Mean Value and Correlation Problems Connected with the Motion of Small Particles Suspended in a Turbulent Fluid*, Martinus Nijhoff, The Hague.

APPENDIX. DERIVATION OF EQUATION (21)

Let $y = (y_1, y_2, y_3)$ be (not necessarily orthogonal) curvilinear coordinates. Then the differential expression

$$\sum_j \frac{\partial}{\partial x_j}\left(a(x)\frac{\partial u}{\partial x_j}\right)$$

in rectangular coordinates becomes

$$\sum_{i,j} \frac{1}{\sqrt{g}}\frac{\partial}{\partial y_i}\left(a(y)\frac{\partial u}{\partial y_j} g^{ij}\sqrt{g}\right)$$

where $g = |\det(g^{ij})|$, and

$$g^{ij} = \sum_k \frac{\partial y_j}{\partial x_k}\frac{\partial y_i}{\partial x_k}.$$

In our case we assume we have a surface given by $\mathbf{x}' \equiv (x_1', x_2', x_3') = R\mathbf{f}(y_2, y_3)$, where y_2 and y_3 are dimensionless parameters and R is a characteristic length. We define curvilinear coordinates (y_1, y_2, y_3) by

$$(x_1', x_2', x_3') = R[\mathbf{f}(y_2, y_3) - y_1\mathbf{n}(y_2, y_3)]$$

where $\mathbf{n}(y_2, y_3)$ is the unit normal to the surface. Let $(x_1, x_2, x_3) = (x_1'/R, x_2'/R, x_3'/R)$. Then the differential equation we are to solve is

$$\frac{1}{R^2}\nabla\cdot[(D + D_e)\nabla u] - \frac{v\mathbf{k}}{R}\cdot\nabla u = 0 \tag{A.1}$$

where $D_e = k_e(y_1 R)^2$. The boundary conditions are

$$\begin{aligned} u &= 0 \qquad y_1 = 0 \\ u &= 1 \qquad y_1 = \delta/R \end{aligned} \tag{A.2}$$

The coordinate transformation we have chosen has the properties $g^{11} = 1$, $g^{1j} = 0$ for $j \neq 1$, and

$$\sqrt{g} = \left|\mathbf{n}\cdot\left(\frac{\partial \mathbf{f}}{\partial y_2} - y_1\frac{\partial \mathbf{n}}{\partial y_2}\right)\times\left(\frac{\partial \mathbf{f}}{\partial y_3} - y_1\frac{\partial \mathbf{n}}{\partial y_3}\right)\right|$$

Hence, in the y-coordinates we have:

$$\frac{1}{R^2\sqrt{g}}\frac{\partial}{\partial y_1}\left((D + k_e R^2 y_1^2)\frac{\partial u}{\partial y_1}\sqrt{g}\right) + \frac{(D + k_e R^2 y_1^2)}{R^2}\sum_{i,j\neq 1}\frac{1}{\sqrt{g}}\frac{\partial}{\partial y_i}\left(\frac{\partial u}{\partial y_j} g^{ij}\sqrt{g}\right)$$

$$-\frac{v}{R}\mathbf{n}(y)\cdot\mathbf{k}\frac{\partial u}{\partial y_1} - \frac{v}{R}\sum_{j\neq 1}\frac{\partial u}{\partial y_j}\frac{\partial y_j}{\partial x_3} = 0. \tag{A.3}$$

Now we set $z = y_1\sqrt{\dfrac{k_e R^2}{D}}$, giving

$$\frac{1}{\sqrt{g}}\frac{\partial}{\partial z}\left((1 + z^2)\frac{\partial u}{\partial z}\sqrt{g}\right) + \frac{D}{k_e R^2}(1 + z^2)\sum_{i,j\neq 1}\frac{1}{\sqrt{g}}\frac{\partial}{\partial y_i}\left(\frac{\partial u}{\partial y_j} g^{ij}\sqrt{g}\right)$$

$$-\frac{v}{\sqrt{k_e D}}\mathbf{n}(y)\cdot\mathbf{k}\frac{\partial u}{\partial z} - \frac{v}{Rk_e}\sum_{j\neq 1}\frac{\partial u}{\partial y_j}\frac{\partial y_j}{\partial x_3} = 0 \tag{A.4}$$

where $\mathbf{n}(y) \equiv \mathbf{n}(y_2, y_3)$ is the unit normal on the surface, and $\mathbf{k}$ is the unit vector in the vertical (x_3') direction. Now, making the assumption that $\dfrac{\delta}{R} \ll 1$ and $k_e\delta^2 \gg D$, as before, we find that the second sum above is negligible, as is the

fourth compared to the third, leaving

$$\frac{1}{\sqrt{g}}\frac{\partial}{\partial z}\left((1+z^2)\frac{\partial u}{\partial z}\sqrt{g}\right)-\frac{v}{\sqrt{k_eD}}\mathbf{n}(y)\cdot\mathbf{k}\frac{\partial u}{\partial z}=0. \tag{A.5}$$

Now,

$$\frac{1}{\sqrt{g}}\frac{\partial}{\partial z}\left((1+z^2)\frac{\partial u}{\partial z}\sqrt{g}\right)=\frac{\partial}{\partial z}\left((1+z^2)\frac{\partial u}{\partial z}\right)+(1+z^2)\frac{\partial u}{\partial z}\frac{1}{\sqrt{g}}\frac{\partial(\sqrt{g})}{\partial z}$$

and

$$\left|\frac{\partial u}{\partial z}\frac{1}{\sqrt{g}}\frac{\partial(\sqrt{g})}{\partial z}\right|\approx\frac{\delta}{R}\left|\frac{1}{\sqrt{g}}\frac{\partial(\sqrt{g})}{\partial y_1}\right|$$

$$\left|\frac{\partial(\sqrt{g})}{\partial y_1}\frac{1}{\sqrt{g}}\right|=\left|\mathbf{n}\cdot\left(\frac{\partial\mathbf{n}}{\partial y_2}\times\frac{\partial\mathbf{f}}{\partial y_3}\right)+\mathbf{n}\cdot\left(\frac{\partial\mathbf{f}}{\partial y_2}\times\frac{\partial\mathbf{n}}{\partial y_3}\right)\right|\Big/\left|\mathbf{n}\cdot\left(\frac{\partial\mathbf{f}}{\partial y_2}\times\frac{\partial\mathbf{f}}{\partial y_3}\right)\right|$$

near the surface. This is just twice the mean curvature or $\left(\frac{1}{r_1}+\frac{1}{r_2}\right)$, where r_1 and r_2 are the (dimensionless) principal radii of curvature. Thus,

$$\frac{\delta}{R}\left|\frac{1}{\sqrt{g}}\frac{\partial(\sqrt{g})}{\partial y_1}\right|=\left(\frac{\delta}{Rr_1}+\frac{\delta}{Rr_2}\right)$$

and Rr_1 and Rr_2 are the (dimensional) principal radii of curvature. We assume that δ is small compared with these and the resulting simplified equation is

$$\frac{\partial}{\partial z}\left((1+z^2)\frac{\partial u}{\partial z}\right)-\frac{v}{\sqrt{k_eD}}\mathbf{n}(y)\cdot\mathbf{k}\frac{\partial u}{\partial z}=0.$$

Now everything follows as before, and the formula for the total deposition rate is

$$c_0\int_\Sigma\frac{v\mathbf{n}(y)\cdot\mathbf{k}}{\exp\left[\dfrac{\pi v\mathbf{n}(y)\cdot\mathbf{k}}{2\sqrt{k_eD}}\right]-1}\,\mathrm{d}A(y).$$

CHAPTER 5

PARTICLE WALL LOSS RATES IN VESSELS

PARTICLE WALL LOSS RATES IN VESSELS

James G. Crump, Richard C. Flagan* and John H. Seinfeld
Department of Chemical Engineering
California Institute of Technology
Pasadena, California 91125

ABSTRACT

Aerosol particle wall loss rates were determined experimentally in a spherical continuous stirred tank reactor. The particle size and mixing rate dependences are shown to agree with the theoretical result of Crump and Seinfeld (1981), in which the particle loss coefficient β is related to particle diffusivity D, particle settling velocity v, the coefficient of the eddy diffusivity k_e and vessel radius R by

$$\beta = \frac{12 k_e D}{R\pi v} \int_0^{\frac{\pi v}{2\sqrt{k_e D}}} \frac{te^t}{e^t - 1}\, dt + \frac{v}{4R/3} .$$

For the vessel used in these experiments, k_e was found to be proportional to the 3/2 power of the volumetric flow rate, in accordance with theoretical expectations. Results of a similar nature may be expected to hold in vessels of arbitrary shape.

*Environmental Engineering Science

INTRODUCTION

Loss of particles to vessel walls occurs in all experimental situations involving aerosols, and in some cases can be an important factor in altering shapes of particle size distributions. The loss rate to the walls depends in general on vessel size and shape, mixing characteristics, and particle size. Consequently, for loss rate determinations to be useful in size distribution studies these dependences must be known.

In this paper we describe experiments designed to measure particle loss rates in a continuous stirred tank reactor and show that the size and mixing dependence agree with the theoretical result of Crump and Seinfeld (1981).

Wall loss rates are generally expressed through the wall loss coefficient β, defined by the equation

$$\frac{\partial n}{\partial t} = - \beta n \tag{1}$$

where $n(d,t)$ is the particle size distribution function in the vessel. This equation holds provided the aerosol in the vessel is well-mixed, except possibly in a small boundary layer near the wall, whose total volume is negligible compared to that of the vessel.

Most investigators have assumed the loss rate coefficient to be related to particle and vessel characteristics by

$$\beta = \frac{v}{h} + \frac{SD}{V\sigma} \tag{2}$$

where v is particle sedimentation velocity, h is vessel height, S is vessel surface area, D is particle diffusivity, V is vessel volume, and σ

is a diffusion boundary layer thickness. This expression is based on the assumption of complete convective mixing in the vessel except in a small layer near the wall where Brownian motion is the only diffusional transport mechanism. Gravitational sedimentation is assumed to be independent of diffusion, and the loss rates due to sedimentation and diffusion are summed to give the total loss rate in Eq. (2).

A flaw in this model is the evident inconsistency of the assumptions of complete convective mixing outside the boundary layer and complete absence of convection within the boundary layer. As a result, the quantity σ is really a hypothetical variable defined by Eq. (2), and dependent in general on particle size, vessel characteristics, and degree of mixing.

Several previous studies of particle wall loss in closed vessels have been carried out are are summarized in Table 1. The work of Van de Vate and that of Harrison merit some explanation, since they represent the only attempts to address the particle size dependence of the loss coefficient.

Using polystyrene latex spheres with diameters ranging from 0.09 μm to 1.3 μm, Van de Vate obtained excellent agreement of his experimental values of the loss coefficient with Eq. (2) by taking σ to be 0.85 mm, independent of particle diameter. However, he also assumed the density of polystyrene latex spheres to be 0.95 g/cm^3 instead of the accepted value of 1.05 g/cm^3. Moreover, Van de Vate asserted that the value he obtained for σ should hold universally in closed vessels since loss rate is only a weak function of the degree of convective mixing in the vessel.

Harrison, also using polystyrene latex spheres, determined the boundary layer thickness as a function of particle diameter to be of the form

$$\sigma = 3.7d^{-1.7} \qquad (3)$$

where both σ and d are in μm. This result is, however, doubtful since it predicts increasing diffusional contributions to the loss rate with increasing particle diameter, despite the fact that the diffusivity decreases with particle size, and diffusion due to mild convection is not expected to be size dependent. Possibly, charge effects contributed to Harrison's results.

In any case, in no previous work were the effects of vessel shape or degree of convective mixing quantitatively accounted for. In addition, quantitative loss rate data have typically been reported in terms of the hypothetical boundary layer thickness σ, whose dependence on these factors is unknown.

DEPENDENCE OF WALL LOSS RATE ON GEOMETRY AND CONVECTIVE MIXING

Crump and Seinfeld (1981) obtained an expression for the wall loss coefficient in a vessel of arbitrary shape as a surface integral of the form

$$\beta = \frac{1}{V} \int \frac{v\hat{n}(x) \cdot \hat{k} dA(x)}{\exp\left[\frac{\pi v \hat{n}(x) \cdot \hat{k}}{2\sqrt{k_e D}}\right] - 1} \tag{4}$$

where $\hat{n}(x)$ is the unit outward normal to the surface, and $\hat{k}$ is the unit vector in the vertical direction. For a spherical vessel, such as that used in the experiments to be described, Eq. (4) reduces to

$$\beta = \frac{6\sqrt{k_e D}}{R} D_1\left(\frac{\pi v}{2\sqrt{k_e D}}\right) + \frac{v}{4R/3} \tag{5}$$

where R is the sphere radius, k_e is the coefficient of the turbulent eddy diffusivity in the wall layer defined by

$$D_e = k_e y^2 \tag{6}$$

in which y is distance from the wall, and D_1 is the Debye function given by

$$D_1(x) = \frac{1}{x} \int_0^x \frac{te^t}{e^t-1} dt \tag{7}$$

The only unknown parameter in the expression for β is the coefficient of the eddy diffusivity k_e. Okuyama et al. (1977) suggest that

$$k_e \propto (\varepsilon/\nu)^{1/2} \tag{8}$$

where ε is the turbulent energy dissipation rate and ν is the kinematic viscosity of the fluid. This expression is essentially the Prandtl mixing length formula, where $(\varepsilon/\nu)^{1/2}$ is proportional to the rms velocity gradient in the boundary layer.

Using Eq. (8) and assuming a constant fraction of the incoming kinetic energy is dissipated in turbulence we obtain, for a vessel through which a volumetric flow rate Q exists,

$$k_e \propto Q^{3/2} \tag{9}$$

Convection due to thermal gradients has been ignored and can be expected to be important only at very low flow rates.

DESCRIPTION OF EXPERIMENTS

A continuous stirred tank reactor (CSTR) was used to measure particle loss rates in two sets of experiments conducted at two flow rates. The CSTR is a glass vessel, roughly spherical in shape, of approximately 118 liters capacity, fitted with inlets and outlets. Stirring was accomplished by means of convection due to flow through the vessel, and in several tracer experiments using oxygen, mixing was shown to be nearly ideal. In most of the experiments sodium chloride particles were used, but polystyrene latex particles were used in two.

The aerosol was produced by atomization of sodium chloride solutions or dilute suspensions of polystyrene latex, followed by passage through a Kr-85 charge neutralizer and diffusion drying tube. In the case of sodium chloride aerosols, the aerosol was passed through a TSI model 3071 electrostatic classifier to separate out a monodisperse fraction of desired size. The system was allowed five to six residence times to reach steady state, after which several measurements were made of both the CSTR output and the feed. Particle concentrations were measured using either a TSI model 3030 electrical aerosol analyzer for particles smaller than 0.2 μm diameter, or a Royco model 226 laser optical particle counter for larger particles.

Current readings from the electrical aerosol analyzer were converted to particle concentrations using the monodisperse sensitivities of Liu and Pui (Kapadia, 1980) according to

$$c = 1/S(d) \sum_j \Delta I_j \qquad (10)$$

where I_j is the current reading in channel j, and j ranges over all channels in which particles register a response. S(d) is the electrometer sensitivity of the instrument.

The loss coefficient β can be determined from the experimental data as follows. Consider a monodisperse distribution of particles of number concentration c_f fed to the reactor at a volumetric flow rate Q. The steady state balance on the number concentration c of particles of the feed size in the vessl is

$$0 = Q(c_f - c) - \beta Vc - KVc^2 \tag{11}$$

where K is one-half the value of the monodisperse coagulation coefficient for the feed size particle.

The standard deviation σ_β in the loss coefficient based on the measurements of c and c_f was approximated by

$$\sigma_\beta = \left[\left(\frac{\partial \beta}{\partial c}\right)^2 \sigma_c^2 + \left(\frac{\partial \beta}{\partial c_f}\right)^2 \sigma_{c_f}^2\right]^{1/2} \tag{12}$$

where σ_c and σ_{c_v} are standard deviations of the output and feed particle concentrations, respectively. The values of β and σ_β for the two sets of experimental conditions are shown in Tables 2 and 3.

DISCUSSION OF RESULTS

Figures 1 and 2 show the experimentally determined values of β as a function of particle diameter, together with the theoretical expression given by Eq. (5) with values of k_e chosen by least squares fitting of the data.

The data from the experiment in Table 2, for a flow rate of 3.8 ℓ/min, are shown in Figure 1. The value of k_e that produced the best fit of theory and experiment is 0.028 sec^{-1}. The data for 0.14 μm particles are not plotted since these were taken using the optical particle counter, which was subsequently found to be extremely insensitive to particles this small.

Figure 2 shows the predicted and observed values of β for the experiment in Table 3. The best fit value of k_e was found to be 0.068 sec^{-1},

and the flow rate through the CSTR for these experiments was 2.1 ℓ/min. Not shown is the data point for 0.34 μm particles, since it was considered unreliable having been obtained from optical particle counter measurements of aerosol which should have consisted of 0.51 μm particles. The points at 0.312 and 0.784 μm were obtained using polystyrene latex and correcting approximately for the density difference between polystyrene latex and sodium chloride. These points show good agreement with the theoretical curve.

The ratio of the two best fit values of k_e for the two sets of experiments is 0.068/0.028 = 2.43, and the ratio of the corresponding flowrates to the 3/2 power is $(3.8/2.1)^{3/2} = 2.43$, giving exact agreement with Eq. (9). One obtains the following relation between the cofficient k_e and the volumetric flow rate through the vessel,

$$k_e = 0.00918\ Q^{3/2} \tag{13}$$

where k_e is given in sec^{-1} if Q is expressed in ℓ/min.

The scatter in the data is more than can be accounted for by measurement error, and probably indicates some uncontrolled factor affecting deposition rate, possibly electrical charge effects. In any case, due to the scatter, the best that can be said concerning the validity of (5) and (9) is that they are consistent with the data, but not necessarily proved.

It is worth noting that the eddy diffusivity in Eq. (6) is not the only definition possible. Crump and Seinfeld (1981) consider an alternative form in which the eddy diffusivity near the wall is proportional to the cube of the distance from the wall. Although the cubic functionality results in a stronger size dependence of β for small particles than the square functionality, and possibly a better fit of the second set of data, scatter in the data prohibits any conclusive choice, so we have used the more classical form of Eq. (6).

In general, application of Eq. (4) to an arbitrary closed vessel will require experimental determination of the eddy diffusivity coefficient k_e unless this can be calculated a priori. Because Eq. (4) contains the particle size dependence of the loss rate, the determination of k_e need only be carried out for a limited particle size range, however. In addition, if the turbulent energy dissipation rate can be expressed as a function of easily measured system variables, as was done here, the loss rate dependence on mixing will also be accounted for by Eq. (4) through the dependence of k_3 on the turbulent dissipation rate given by Eq. (8).

CONCLUSIONS

Experimental results have been presented on particle loss rates in a continuous stirred tank reactor. The particle size and mixing rate dependence of the loss rate coefficient has been shown to agree with the theoretical formula of Crump and Seinfeld. Although the results presented here apply to a spherical vessel, similar results may be expected to hold in a vessel of arbitrary shape provided the more general expression for the wall loss coefficient is used. The results presented here demonstrate the utility of the theory in correlating aerosol wall loss data in experimental apparatus.

REFERENCES

Crump, J. G. and J. H. Seinfeld, J. Aerosol Sci. 12, 405-415 (1981).

Friedlander, S. K., Smoke, Dust and Haze, John Wiley and Sons, N.Y. (1977).

Harrison, A. W., J. Colloid and Interface Sci. 69, 563-570 (1979).

Kapadia, A., Thesis, University of Minnesota, (1980).

Longstroth, G. and T. Gillespie, Canadian J. Res. 25B, 455-471 (1947).

Lieberman, A. and J. Rosinski, J. Colloid Sci. 17, 814-822 (1962).

Okuyama, K., Y. Kousaka, Y. Kida, and T. Yoshida, J. Chem. Eng. Japan 10, 142-147 (1977).

Van de Vate, J. F., J. Colloid and Interface Sci. 41, 194-197 (1972).

Table 1. Previous Studies on Aerosol Wall Loss in Vessels

Investigator	Particles Used	Diffusion Boundary Layer Thickness σ	Comments and Conclusions
Langstroth and Gillespie (1947)	polydisperse NH_4Cl		Showed dependence of loss on mixing rate
Lieberman and Rosinski (1962)	polydisperse zinc-cadmium sulfide		Demonstrated strong effect of particle charging on loss rate
Van de Vate (1972)	monodisperse polystyrene latex (PSL)	σ=0.085 cm for particle diameters between 0.09 μm and 1.3 μm.	Used 0.95 g/cm^3 as PSL density instead of accepted value 1.05 g/cm^3
Okuyama, et al. (1977)	tobacco smoke		Measured loss and coagulation as function of stirring rate.
Harrison (1979)	monodisperse polystyrene latex	$\sigma=3.7d^{-1.7}$ for 0.234 μm < d < 2.02 μm ($\bar{\sigma}$ in μm)	

Table 2. Experimentally Determined Values of the Wall Loss Coefficient β and Estimated Standard Deviations. Residence Time in Vessel = 58 min.

Particle diameter (μm)	$\beta(sec^{-1})$	$\sigma_\beta(sec^{-1})$
0.14	$4.8x10^{-6}$	$5x10^{-6}$
0.34	$2.2x10^{-5}$	$2.8x10^{-6}$
0.075	$4.7x10^{-5}$	$6.0x10^{-6}$
0.042	$7.2x10^{-5}$	$1.5x10^{-5}$
0.024	$1.0x10^{-4}$	$1.0x10^{-5}$
0.34	$3.6x10^{-5}$	$5.0x10^{-6}$
0.51	$6.8x10^{-5}$	$3.9x10^{-6}$
0.13	$1.7x10^{-5}$	$2.6x10^{-6}$
0.21	$2.0x10^{-5}$	$2.5x10^{-6}$
0.21	$2.6x10^{-5}$	$3.0x10^{-6}$

Table 3. Experimentally Determined Values of the Wall Loss Coefficient β and Estimated Standard Deviations. Residence Time in Vessel = 31 min.

Particle diameter (μm)	$\beta(sec^{-1})$	$\sigma_\beta(sec^{-1})$
0.024	$1.8x10^{-4}$	$5.0x10^{-6}$
0.042	$8.4x10^{-5}$	$1.4x10^{-5}$
0.075	$3.8x10^{-5}$	$3.0x10^{-6}$
0.13	$2.8x10^{-5}$	$1.0x10^{-5}$
0.34	$1.5x10^{-5}$	$3.8x10^{-6}$
0.51	$4.8x10^{-5}$	$4.6x10^{-6}$
0.794 (PSL)*	$6.1x10^{-5}$	$5x10^{-6}$
0.794 (PSL)*	$1.26x10^{-4}$(corrected)	$1x10^{-5}$
0.312 (PSL)*	$1.8x10^{-5}$	$4.3x10^{-6}$
0.312 (PSL)*	$2.9x10^{-5}$(corrected)	$6.8x10^{-6}$

*In the second set of measurements two values of β were obtained using polystyrene latex and were corrected to the density of sodium chloride by the factor

$$(2D_1(x_1) + \frac{1}{2} x_1)/(2D_1(x_2) + \frac{1}{2} x_2)$$

where $x_i = \pi v_i/2\sqrt{k_e D}$ and D_1 is the Debye function.

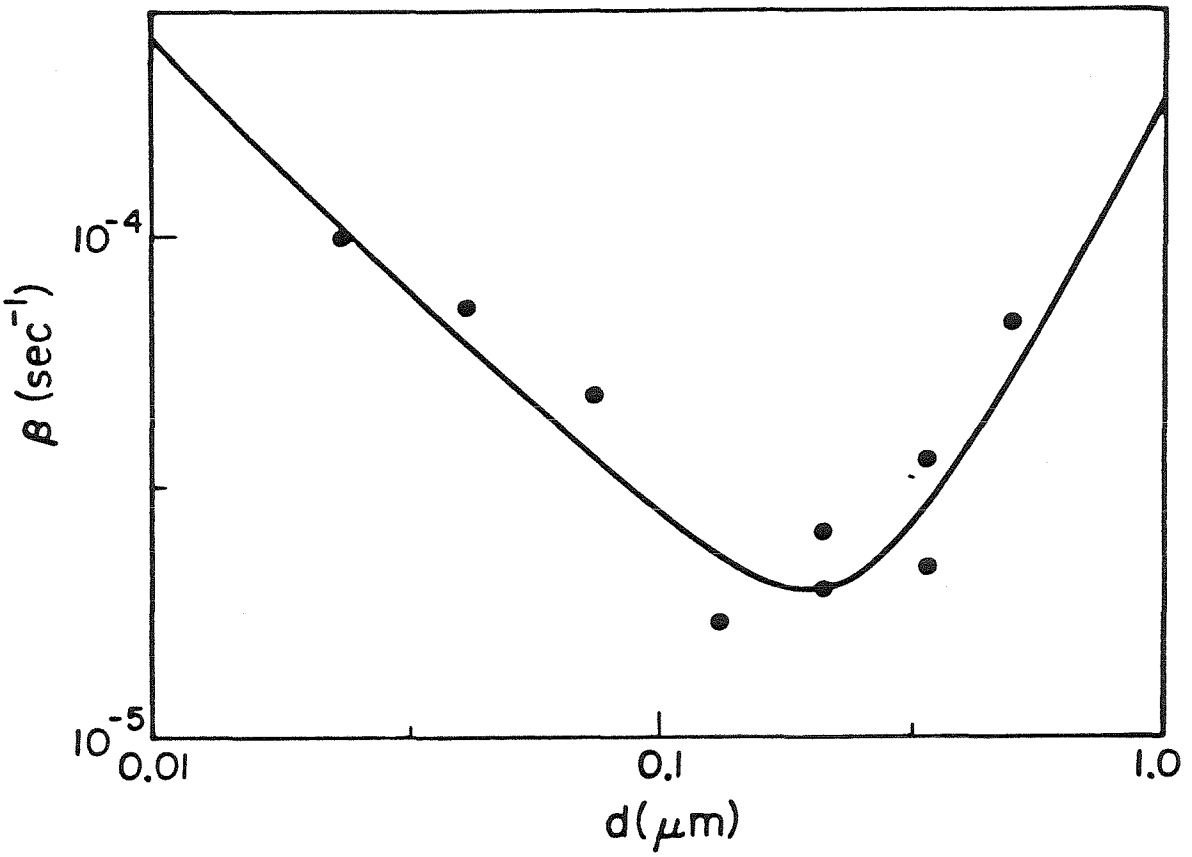


Figure 1. Loss coefficient values from data of Table II (points). Curve is the theoretical prediction of Crump and Seinfeld (1981) for the best fit value of k_e= 0.028 sec^{-1}.

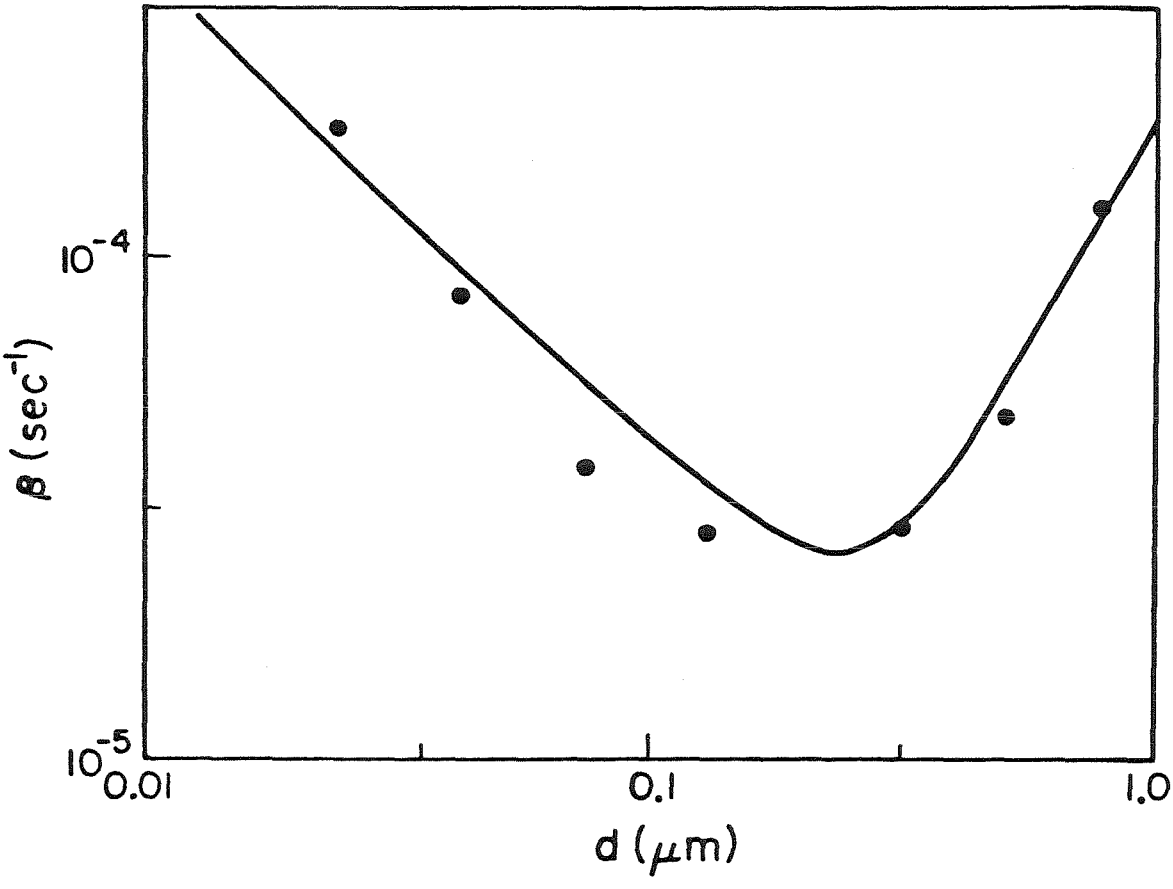


Figure 2. Loss coefficient values from data of Table III (points). The curve is the theoretical prediction of Crump and Seinfeld (1981) for the best fit value of k_e = 0.068 sec^{-1}.

CHAPTER 6

A NEW ALGORITHM FOR INVERSION OF AEROSOL SIZE DISTRIBUTION DATA

Published in Aerosol Science and Technology 1, 15(1982).

A New Algorithm for Inversion of Aerosol Size Distribution Data

James G. Crump and John H. Seinfeld
Department of Chemical Engineering, California Institute of Technology, Pasadena, CA 91125

The determination of an aerosol size distribution from conventional data is an ill-posed problem. The general characteristics of the aerosol size distribution inversion problem are discussed, and several existing methods are assessed. A new algorithm, in the spirit of Twomey's constrained linear inversion algorithm, employing the concept of generalized cross validation is developed. Extensive numerical tests on simulated data from a Marple impactor show superior performance of the method. Different versions of the algorithm are available based on different choices of the function spaces in which the assumed distributions lie. The new algorithm offers promise as the most accurate now available for inversion of aerosol data.

INTRODUCTION

The treatment of data obtained from aerosol size distribution measurements has long been a source of controversy, and, at present, several different techniques are in use for reconstructing size distributions from experimental data. Each of these techniques has certain deficiencies, and indeed, it is doubtful that any method can fulfill the ideal of perfect reconstruction of a size distribution from a limited set of data.

However, it seems a good deal can be done to improve data analysis. It is the purpose of this paper first to discuss the general characteristics of the problem and then to present a new method based on Wahba's generalized cross-validation approach. We compare this method theoretically and numerically with some other methods presently in use.

STATEMENT AND NATURE OF THE PROBLEM

For a number of aerosol size measurement instruments the instrument response varies linearly with the size distribution being measured. This is true, for example, of diffusion batteries, mobility measuring devices, inertial impactors, and optical measurement systems. Such a device can, at least in principle, be calibrated so that the response is a known linear function of the size distribution. The data inversion problem then amounts to determining which distribution caused a given measured response in an instrument the linear response curve of which is known. Typically, one measures a finite number of instrument responses, so that the problem can be stated as follows:

Find the size distribution f such that

$$L_i f = y_i, \qquad i = 1, 2, \ldots, n, \tag{1}$$

where f is the unknown size distribution, y_i the ith datum, and L_i the ith instrument response linear functional.

Strictly speaking, the data y_i are known only approximately, since they are the results of experimental measurements, and it will turn out that this is part of the difficulty associated with the solution of problem (1).

Problem (1) can also be expressed in the more general form

$$Tf = \mathbf{y}, \tag{2}$$

Aerosol Science and Technology 1:15–34 (1982)

0278-6826/82/010015-20$02.75

where T is the linear map defined on some vector space of functions by $Tf=(L_1f, \ldots, L_nf)$. At this point it is worthwhile to make some general comments on the terminology and properties of the general linear inverse problem (2).

For the moment suppose T is a linear map between normed vector spaces X and Y. We assume that y is known and f is to be found.

Problem (2) is well posed if it is uniquely solvable for every $\mathbf{y}$ and the solution f varies continuously with the data $\mathbf{y}$. Well-posedness is often characterized by the following three conditions:

a. For every $\mathbf{y}$ there is a solution f.
b. The solution f is unique.
c. The solution f is stable.

Condition c requires explanation. The solution f is said to be *stable* if, for any sequence of perturbations in $\mathbf{y}$ tending to zero, the corresponding sequence of perturbations in the solution f also tends to zero. In analogy to the terminology used in systems of linear equations, we say Eq. (2) is *overconstrained* if a fails and *underdetermined* if b fails.

In our case we are given a finite set of data and must determine a function from infinitely many possibilities. Hence the problem is underdetermined. We shall also see shortly that it is unstable.

Before continuing, however, another term deserves mention. In solving (2) numerically we would end up with an equation of form of (2), in which T is a matrix and $\mathbf{f}$ and $\mathbf{y}$ vectors. Even if a–c hold for this system, it may happen that small perturbations in the data $\mathbf{y}$ cause relatively large disturbances in the solution $\mathbf{f}$. In this case the problem is called *ill conditioned*. This manifests itself in an extreme insensitivity of the data to large perturbations in the solutions.

In general, the linear functionals L_i are linear integral operators of the Fredholm type; that is,

$$L_if=\int K_i(x)f(x)\,dx. \tag{3}$$

The vexing nature of the inverse problem can be seen by adding to some solution f of (1) the function $A \sin \omega x$. If A is large enough, then f is completely dominated by large amplitude oscillations. Then by choosing ω large enough, and from the fact that

$$\lim_{\omega \to \infty} \int K_i(x) \sin \omega x\, dx = 0,$$

it can be seen that this wildly oscillating function is also a solution to the problem, within any degree of accuracy with which we can measure the data. This behavior is not limited to the sine function but clearly holds for any trigonometric polynomial of high frequency, and consequently for any rapidly oscillating functions that are, in some sense, limits of high frequency trigonmetric polynomials. Consequently, almost any function that oscillates rapidly enough can be added to a solution of (1) without affecting the validity of (1). From this discussion it can be concluded that Eq. (1) itself is not sufficient to afford a solution of the inversion problem, but some additional information, more or less qualitative in nature, will need to be used in order to obtain acceptable solutions. This is a general characterisitc of ill-posed problems, and investigation of such problems has long focused on what criteria should be used (Tihonov, 1963a, b).

FUNCTIONALS USED FOR SIMULATION OF SIZE DISTRIBUTION INVERSION

Numerical simulations described in this paper were based on the Marple impactor (Marple and Liu, 1974). This is a particularly convenient device because, by using different flow rates, several sets of data can be obtained, thereby increasing the resolution obtainable from the data.

Typically, in impactor calibration the collection efficiency of each stage is measured as a function of some size parameter, often the Stokes number N_{Stk}, yielding efficiency curves $E_i(x)$, from which the kernel functions K_i are obtained by

$$K_1(x)=E_1(x),$$

$$K_i(x)=E_i(x)(1-E_{i-1}(x))\cdots(1-E_1(x)) \qquad (i=2, \ldots, n). \tag{4}$$

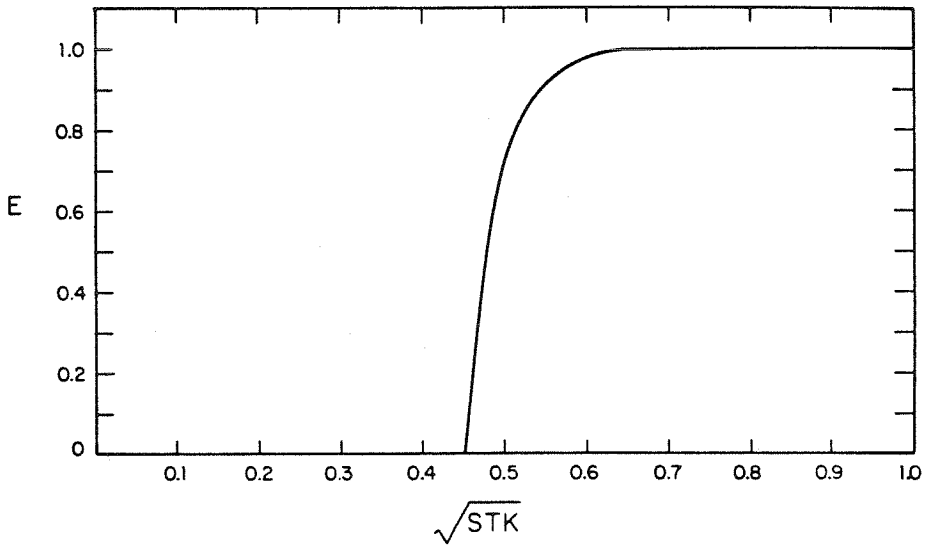
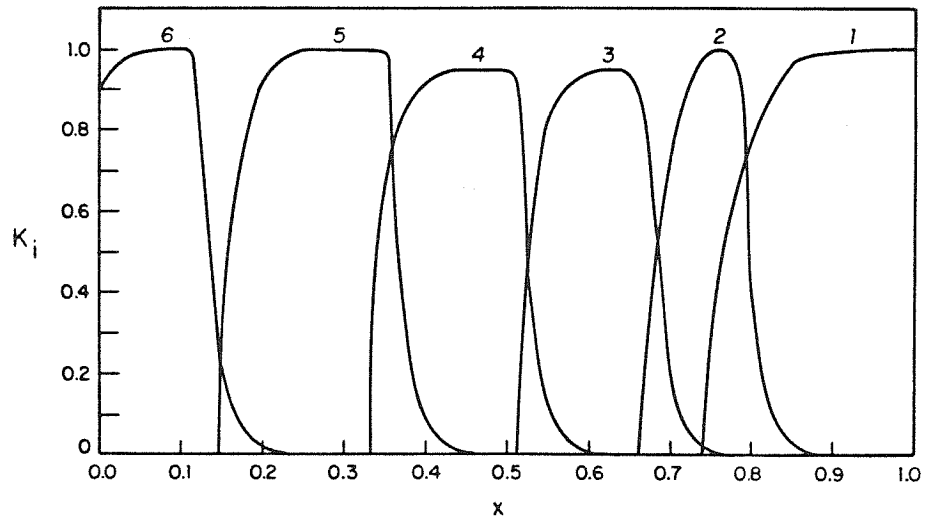


FIGURE 1. Analytical approximation to Marple's efficiency curve vs Stokes number.

In the simulations given here, six stages with round jets of diameters 0.8, 0.6, 0.4, 0.25, 0.15, and 0.1 cm are used. The efficiency curves $E_i(x)$ were calculated from a universal efficiency curve as a function of Stokes number. Marple's numerical work shows that for $N_{Re} > \sim 500$, all the stage efficiency curves fall approximately on the same universal curve when plotted as a function of Stokes number. We have thus an analytic approximation to this curve, which is shown in Figure 1. Figure 2 shows the kernel functions K_i plotted as a function of dimensionless size for a flow rate of 5 liter/min through the impactor. The particle density was assumed to be 1 g cm^{-3}, and all other properties were taken to be those of air at room temperature and pressure.

FIGURE 2. Kernel functions $K_i(x)$ vs dimensionless particle size $x = \ln(d/d_1)/\ln(d_2/d_1)$ where $d_1 = 0.6$ μm and $d_2 = 30$ μm.



REVIEW OF AEROSOL DATA INVERSION METHODS

In this section we discuss some of the methods of solving the inversion problem that have been proposed in the past. We concentrate especially on the Twomey nonlinear algorithm (Twomey, 1975) since this method seems to have gained wide popularity.

Histogram Method

The histogram method is one of the simplest and, for this reason alone, one of the better techniques for handling aerosol data. It consists simply of assuming a histogram distribution with a number of size ranges or channels equal to the number of measurements made. This is a natural approach for instruments such as the inertial impactor and the electrical mobility analyzer, since these instruments are designed to achieve fairly sharp size cutoffs, and therefore provide natural channel boundaries.

The disadvantage of the histogram method is that it provides no indication of the shape of the distribution within the channel boundaries and moreover does not produce a smooth distribution—a fact which is significant if the derivative of the size distribution is desired. In addition, even for well-designed impactors, or mobility analyzers, the channel cutoffs are not perfect. By using the histogram method, one must ignore detailed calibration data that may be available. Such data should be used whenever possible.

Constrained Linear Inversion

Constrained linear inversion was introduced into the aerosol measurement field by Twomey in 1965 to infer aerosol size distributions from nucleopore filter measurements taken at different flow rates, although it had previously been employed in other contexts (Phillips, 1962; Twomey, 1963). It is actually a version of the regularization method described by Tihonov (1963a, b).

The idea of the method was to choose a solution of f to problem (1) that is the smoothest in some sense, such as having the smallest derivative or second derivative. In order to take error into account in the data, Twomey relaxed the requirement that the distribution f actually be a solution to (1), requiring instead that it merely be "close" to a solution. More specifically, the method consisted of first assigning the size variable discrete values for m sections and letting f_i be the value of f in the ith section. Then the kernel functions $K_i(x)$ are similarly made discrete, so that K_{ij} is the value of K_i in the jth section. Then in matrix notation problem (1) assumes the form

$$K\mathbf{f}=\mathbf{y}. \tag{5}$$

In general, $m>n$, so that (5) has many solutions. In addition, due to error in the data, the real solution $\mathbf{f}$ more correctly ought to satisfy

$$\| K\mathbf{f}-\mathbf{y} \| < \epsilon, \tag{6}$$

where ϵ is some small positive number. Hence, Twomey reasoned that $\mathbf{f}$ should be the "smoothest" vector satisfying (6), with ϵ chosen a priori. The smooth criteria suggested by Twomey were typically of the form of minimizing first or second differences of the vector $\mathbf{f}$, or its variance. The difficulty was that there really was no a priori method to choose the parameter ϵ.

The new method that we shall present shortly bears resemblance to Twomey's constrained linear inversion method. The use of generalized cross validation provides a means to choose the smoothing parameter ϵ from the data.

Nonlinear Inversion

In 1975 Twomey introduced the nonlinear inversion algorithm as a substitute for the linear version, which appeared to give poor results for certain types of distributions that Twomey believed to be fairly typical of those encountered in the atmosphere (Twomey, 1975). It is difficult to determine from Twomey's paper why the linear method failed, since he does not give details of the numerical routines used.

In any event, the proposed nonlinear method seeks to find a reasonable solution to (1) by an

iterative scheme. Approximate solutions are defined by

$$f_k^1(x) = [1 + (r_{k-1}^n - 1)K_1(x)]f_{k-1}^n(x),$$

$$r_k^n = y_1 \Big/ \int K_1(x) f_k^n(x)\, dx,$$

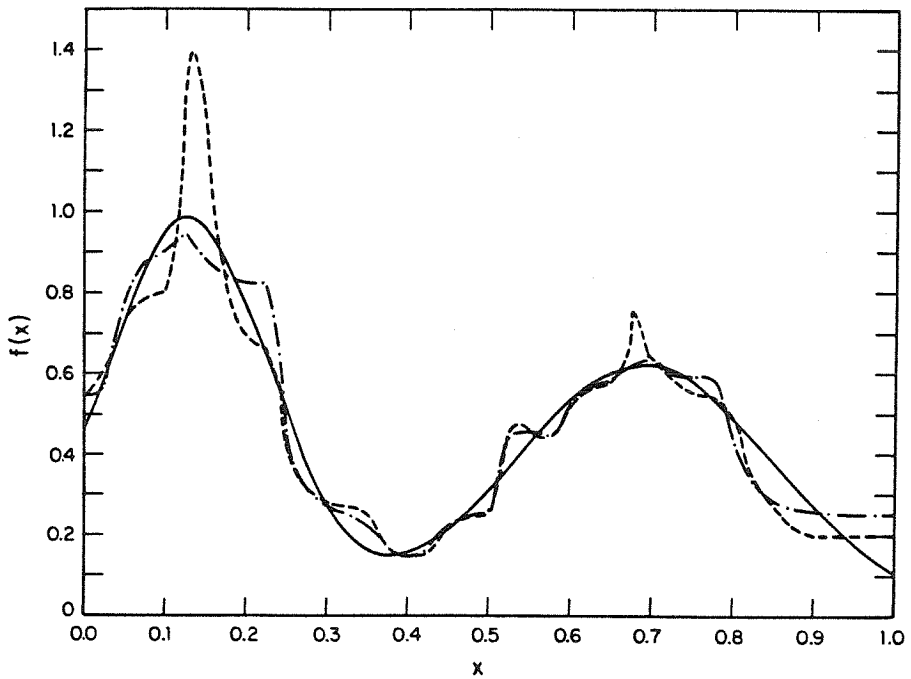
$$f_k^i = [1 + (r_k^{i-1} - 1)K_i(x)]f_k^{i-1}(x) \qquad (i = 2, \ldots, n), \tag{7}$$

$$r_k^{i-1} = y_i \Big/ \int K_i(x) f_k^{i-1}(x)\, dx \qquad (k = 1, 2, 3, \ldots).$$

Basically, the idea of the method is to start with an initial guess distribution, and then to correct repeatedly until the functions converge within some specified degree. One begins by correcting the distribution to improve the fit at the first data point, and then proceeds through the entire set. Then this process is repeated. It is clear from the formulas that the corrections will be in the right direction; i.e., if one of the iterates gives too large a value of y_i, then on the next pass through the iteration scheme it will be changed to decrease this value. Other than this, however, there is really no theoretical basis for using this iterative method to solve the inversion problem.

The main difficulty of the nonlinear inversion method is that it does not really address the problem of ill-posedness inherent in the solution of (1). It can be proved theoretically, and shown numerically, that the solution obtained by this method depends upon the initial guess. Figure 3 shows both a bimodal lognormal distribution and the results of applying Twomey's nonlinear inversion algorithm to data obtained by applying the kernel functions for the Marple impactor to the distribution. In one plot the initial guess is a constant distribution having value 1, while in the other a constant distribution of value 0.1 is used. Note the difference in the resulting distributions. Both were iterated until the error

FIGURE 3. Size distributions obtained by inversion of simulated impactor data. Curve 1 (———) is the true bimodal lognormal distribution. Curve 2 (—·—) is the corresponding inverted distribution obtained by Twomey nonlinear algorithm with initial guess $f(x) = 1$. Curve 3 (- - -) is the Twomey inverted distribution with initial guess $f(x) = 0.1$.



criterion,

$$\sum_{j=1}^{m} [f_k^n(x_j) - f_k^1(x_j)]^2 < 10^{-6}, \quad (8)$$

was satisfied. In the numerical computations $m = 80$; i.e., the size variable was assigned discrete values for 80 sections to carry out the quadratures. Figure 4 shows an extreme case in which the initial guess was a rapidly oscillating function taking on values that varied periodically from 1 to 0.1. Note that the oscillation has remained in the solution. Figure 5 shows a lognormal distribution having mean size 5 μm and geometric standard deviation 1.5, and the result of applying the nonlinear inversion algorithm to this distribution. An error criterion of 10^{-7} was used in this case. In all of these results, two sets of simulated impactor data were used, taken at 5 and 10 liter/min.

Note the presence of spurious lumps in the Twomey inverted distributions. This behavior is partly due to the fact that the algorithm corrects each guess by multiplying it by something that looks like the kernel functions. A glance at Figure 2 should convince the reader that these functions are not particularly smooth, thus producing near discontinuities in the final distributions.

It should also be noted that if the Twomey iterates actually do converge, they converge to a solution of the system (1), which, as we have seen, is underdetermined. In this case, we would be far better off simply to solve the linear equations obtained by discretizing (1) subject to the constraint that the solution be positive. (This constraint is implicit in the Twomey nonlinear algorithm.) However, we are still left with the difficulty of choosing one out of the infinitely many possible solutions to this problem. In actual practice the iterates do not seem to converge, or at least they do so extremely slowly. Hence, the resulting distribution is only an approximate solution to the system (1). Unfortunately, not much else can be said about this solution. There are in general many functions that approximately satisfy (1) but that bear

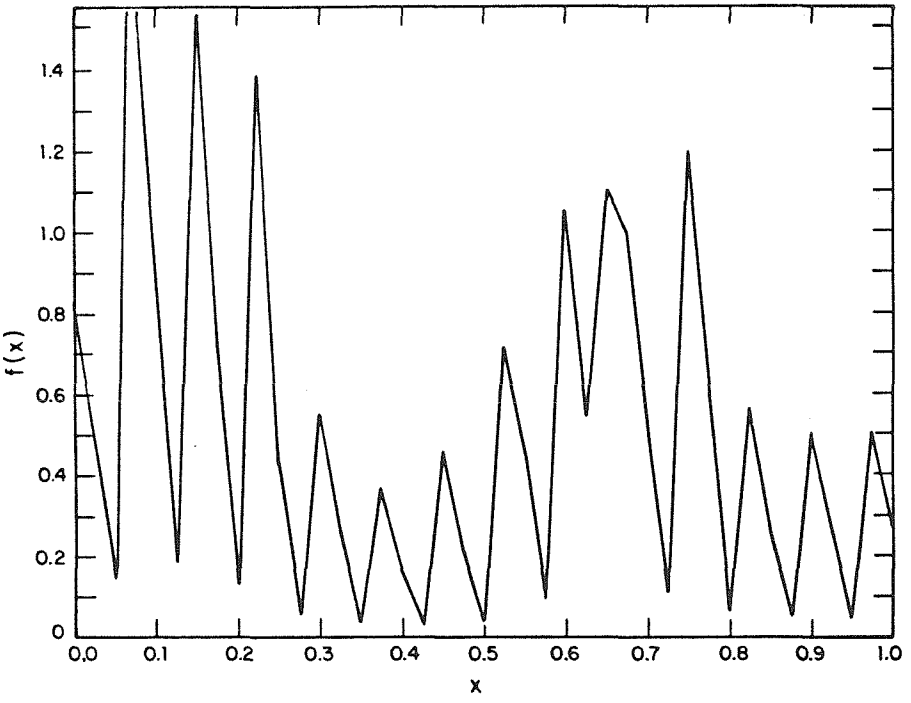


FIGURE 4. Size distribution obtained by Twomey nonlinear method for the bimodal lognormal distribution of Figure 3 using oscillator initial guess.

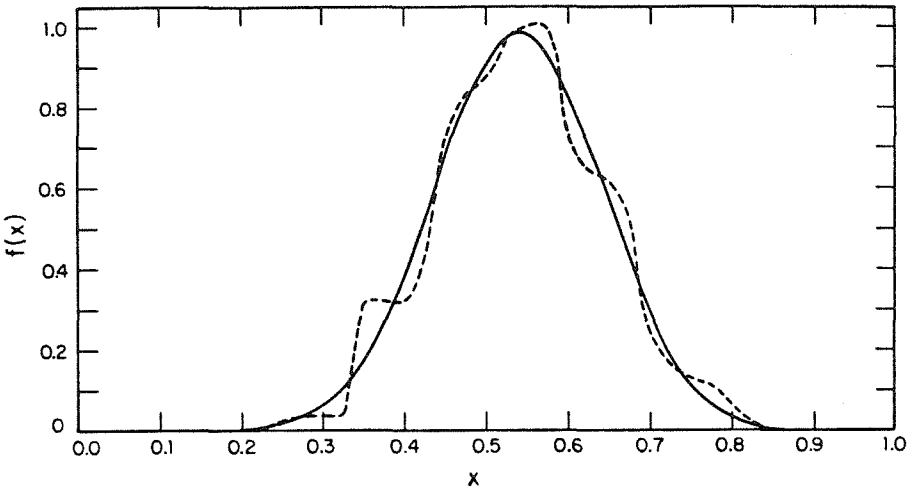


FIGURE 5. Size distributions obtained by inversion of simulated impactor data. Curve 1 (———) is the true lognormal distribution, and curve 2 (---) is the corresponding distribution obtained by Twomey nonlinear algorithm with initial guess $f(x) = 1$.

little resemblance to one another. For example, the solutions shown in Figures 3 and 4 are all approximate solutions to (1) for the case of the Marple impactor.

In short, the Twomey nonlinear algorithm does not confront the main problem associated with inversion of aerosol size distribution data, which is the ill-posed nature of the linear inverse problem.

Nonlinear Programming with Physical Constraints

Cooper and Spielman (1976) proposed nonlinear programming with physical constraints as a replacement to the constrained linear inversion method of Twomey. Their alternative was to replace problem (1) by the following optimization problem with constraints:

$$\text{Find } f \text{ to minimize } \sum_{i=1}^{n} (L_i f - y_i)^2$$

$$\text{subject to the constraints } f(x) \geq 0,$$

$$\int f(x)\, dx = 1. \quad (9)$$

In the usual aerosol applications the integral constraint is not appropriate, but even including it still leaves the problem badly underdetermined. To see this, let T be the linear operator from a vector space of reasonable functions into R^{n+1} defined by

$$(Tf)_i = L_i f, \qquad i = 1, 2, \ldots, n,$$
$$(Tf)_{n+1} = \int f(x)\, dx. \quad (10)$$

Then T has an infinite-dimensional kernel, $\ker T$. Suppose we have a strictly positive solution f to (9). Then, by adding any element of $\ker T$ to f, we still have a solution satisfying (9), except possibly for the positivity constraint. However, since f is strictly positive, we can multiply this element of $\ker T$ by a suitable small constant if necessary so that the new function is still nonnegative. (We are assuming continuity of these functions, which is permissible, since size distributions are usually subject to this requirement.) By increasing the magnitude of this constant until the sum of f and this function achieves a zero, we see that there is always a solution to (9) having a zero. Hence, the ill-posedness of (9) can be very severe.

Hence, the nonlinear programming method suffers from the same defect as Twomey's nonlinear algorithm, in that it does not choose rationally among many possible solutions. Other data inversion methods are used based on fitting bimodal and trimodal lognormal distri-

butions to the data. Since these procedures necessarily use a small number of parameters in fitting the data, they produce nice looking results but with a sacrifice of information content. Except in cases in which the distributions are known a priori to have certain form these methods are not suggested.

DATA INVERSION USING GENERALIZED CROSS VALIDATION

In this section we describe a new method for aerosol size distribution data inversion. As mentioned, the basic method is in the spirit of Twomey's linear inversion algorithm, and generalized cross validation is used to determine an appropriate value of the smoothing parameter from the data.

We assume for the moment that there are no errors in the data of problem (1) and that the problem is to find a distribution f that fits the data. Even though, as we have seen, f is not uniquely determined, the data do, of course, say something about f. Following Backus and Gilbert (1967, 1968), we ask which linear functionals L have the property that Lf is unambiguously determined from the data. The motivation for this inquiry is the fact that for any fixed x_0 in the domain of the size variable, the mapping e_{x_0} defined by $e_{x_0}(f)=f(x_0)$ is a linear functional. Hence, it may turn out that $e_{x_0}(f)$ can be determined or at least well approximated from the data. In this way we could obtain values of $f(x)$ for certain values of x.

In fact, it turns out that the only linear functionals L for which Lf is unambiguously determined from the data are of the form

$$L=\sum_{i=1}^{n} c_i L_i. \tag{11}$$

It is clear from (11) that if L is such a functional then Lf is determined from the data by the equations

$$Lf=\sum_{i=1}^{n} c_i L_i f=\sum_{i=1}^{n} c_i y_i. \tag{12}$$

Backus and Gilbert reasoned that among all possible linear combinations of the L_i there might be some that closely approximate e_{x_0} for some values of x_0. Equivalently, there may exist certain linear combinations of the kernel functions K_i that resemble a delta function at x_0.

Instead of looking for sharply peaked linear combinations of the K_i as did Backus and Gilbert, we shall take a slightly different approach. Backus and Gilbert used an optimization technique to find the most "deltalike" linear combinations of the kernels. One criterion was to choose constants $c_1, \ldots, c_n$ to minimize

$$\int_0^1 (x-x_0)^2 \left[\sum_{i=1}^{n} c_i K_i(x)\right]^2 dx$$

for some fixed x_0

The disadvantages of this method are that it is not computationally easy and does not guarantee that the inverted distribution will be close to the solution. We shall use another approach in which we define distances between linear functionals. The idea advanced here is that given x_0 we seek the linear functional L in the span of the L_i that is closest to the functional e_{x_0} in the space of linear functionals. This can be done fairly easily if we choose the space of acceptable distributions to be a Hilbert space with an inner product and norm related by

$$\|f\|^2=(f,f). \tag{13}$$

If L is a continuous linear functional on H, then the norm of L can be defined naturally as

$$\|L\|=\sup|Lf|, \qquad \|f\|\le 1 \tag{14}$$

Hence it follows that $|Lf|\le\|L\|\,\|f\|$ for all $f\in H$.

Now, from elementary Hilbert space theory (Rudin, 1973) it follows that every continuous linear functional on H can be represented in the sense that there is a function g in H such that

1. $\|g\|=\|L\|$ and
2. for every f in H, $Lf=(g,f)$. (15)

Using (15) we can now easily solve the minimization problem:

Find L in the span of the L_i

to minimize $\|L-e_{x_0}\|$. (16)

To do this let g_i be the function representing L_i in the sense of (15), and let g represent the unknown L, and g_{x_0} the evaluation functional e_{x_0}. Then by the fact that norms are preserved (and hence inner products) by replacing functionals by their representing functions, problem (16) is equivalent to the following:

Find g in the span of the g_i

$$\text{to minimize} \quad \|g - g_{x_0}\|. \tag{17}$$

This minimization is accomplished precisely when g is chosen so that $g - g_{x_0}$ is orthogonal to the span of the g_i. Hence if we let g be given by

$$g = \sum_{i=1}^{n} c_i g_i, \tag{18}$$

we find g by setting

$$(c_j g_j - g_{x_0}, g_i) = 0, \qquad i = 1, \ldots, n. \tag{19}$$

or equivalently,

$$\sum_{j=1}^{n} (g_i, g_j) c_j = (g_{x_0}, g_i) = g_i(x_0). \tag{20}$$

In matrix form, $A\mathbf{c} = \mathbf{g}_0$, where $\mathbf{g}_0$ is the vector with ith component $g_i(x_0)$. Clearly A is self-adjoint and positive, so its eigenvalues are all positive. This will be important later.

Solving Eq. (20) gives us the approximate value for $f(x_0)$

$$f(x_0) \cong \sum_{i=1}^{n} (A^{-1}\mathbf{y})_i g_i(x_0). \tag{21}$$

Through some algebraic manipulation it is readily seen that

$$\left| f(x_0) - \sum_{i=1}^{n} (A^{-1}\mathbf{y})_i g_i(x_0) \right|$$

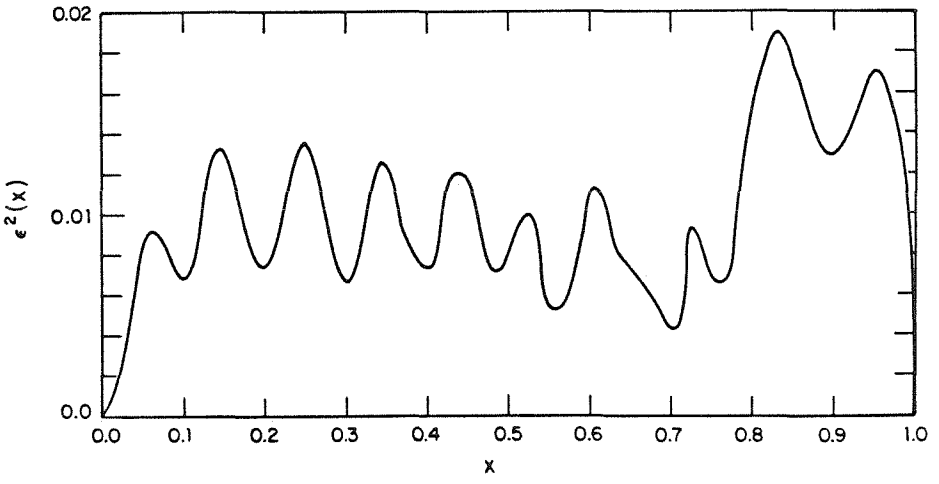
$$\leq (\|g_{x_0}\|^2 - \mathbf{c}^{\mathrm{T}} A \mathbf{c})^{1/2} \|f\|. \tag{22}$$

The factor multiplying $\|f\|$ is the relative resolution error $\epsilon(x_0)$, which is shown for the Marple impactor and two flow rates in Figure 6. The Hilbert space used in this case was $H_0^1(0, 1)$, which will be discussed later.

Note that the error is proportional to $\|f\|$, which is a priori unknown. It is in assuming that this quantity is not too large that we are restricting the solutions to the problem (1). It turns out that in these Hilbert spaces for which the functionals e_{x_0} are all continuous, the norm tends to measure in some sense the smoothness of solutions. Hence, by assuming the norm of f is not large we are in a sense assuming f to be reasonably smooth.

We have not yet been specific about the choice of function space for this problem. The only requirement is that it be a Hilbert space for which the evaluation functionals are all continuous. One possibility, as we mentioned previously, is the Sobolev space $H_0^1(0, 1)$, consisting of all absolutely continuous functions f on the interval (0, 1) that are square integrable, whose

FIGURE 6. Square of the error function ϵ.



derivatives are square integrable, and that satisfy $f(0)=f(1)=0$ (Adams, 1975). The inner product on this space is given by

$$(f, g)=\int_0^1 f'(x)g'(x)\,dx. \tag{23}$$

The evaluation functional e_{x_0} is represented in this space by the function g_{x_0} defined by

$$g_{x_0}=\begin{cases}x(1-x_0), & 0\le x\le x_0\le 1,\\ x_0(1-x), & 0\le x_0\le x\le 1.\end{cases}$$

Hence for every

$$f\in H_0^1(0, 1), f(x_0)=\int_0^1 g_{x_0}'(x)f'(x)\,dx.$$

The functions g_i corresponding to the linear functionals L_i, where

$$L_i(f)=\int_0^1 K_i(x)f(x)\,dx, \tag{24}$$

are defined by

$$g_i(x)=\int_0^x tK_i(t)\,dt+x\int_x^1 K_i(t)\,dt - x\int_0^1 tK_i(t)\,dt. \tag{25}$$

Another possibility is the space $H_0^1(0, 1]$, which is similar to the preceding except it has a zero boundary condition only at $x=0$. In this case the functions g_i are given by

$$g_i(x)=\int_0^x tK_i(t)\,dt+x\int_x^1 K_i(t)\,dt. \tag{26}$$

Both these spaces contain boundary conditions that may not always be appropriate. We shall discuss the more general case in a later section.

It turns out that if the inversion problem is solved in either one of these spaces, then it is equivalent to finding the distribution f of smallest norm that fits the data: that is, of all functions f such that $L_if=y_i$, $i=1, \ldots, n$, let the solution be the one that minimizes $\|f\|$.

Since in this particular case the norm of f is given by (23), the solution is actually the smoothest. Hence, we have been led in a more or less natural way to the idea of smoothing the data, as Twomey originally suggested.

In the next section we discuss the use of cross validation to invert data when errors are present.

Cross Validation and Errors

We have not yet considered the effect of errors on the inversion procedure. These effects tend to become more pronounced as the matrix A of inner products of representing functions becomes larger, since A tends to become ill conditioned. Then the direct inversion of A becomes difficult, and even when it can be accomplished, the results tend to amplify small errors in the data.

One way to avoid the difficulties associated with ill-conditioning of the matrix A is to determine the coefficient vector $\mathbf{c}$ by the formula

$$\mathbf{c}=(A+n\lambda I)^{-1}\mathbf{y} \tag{27}$$

for some positive λ. Since, as we have seen, the eigenvalues of A are all positive, this has the effect of washing out the smallest eigenvalues, thereby eliminating the worst effects of ill-conditioning. At the same time, however, we would like to choose λ small enough that (27) gives a solution reasonably faithful to the data.

It can be shown that the function $f=\Sigma_{i=1}^n c_ig_i$, where the coefficients are determined from (27) and the g_i are the representing functions for the functionals L_i as before, is also the solution to the following problem:

Find f in the Hilbert space H

$$\text{to minimize}\quad (1/n)\sum_{i=1}^n [L_if-y_i]^2+\lambda\|f\|^2. \tag{28}$$

This is exactly the idea that Twomey's linear inversion was based on. The difficulty of determining an appropriate value of λ is overcome by generalized cross validation (Craven and Wahba, 1979; Golub et al., 1979, Wahba 1975, 1977).

The rationale behind cross validation is that if λ is a good value of the smoothing parameter, then by omitting one data point y_k and solving

the resulting problem, using the value λ to obtain a solution $f_{k,\lambda}$, this solution should allow us to predict what the unused datum should be; i.e., $L_k f_{k,\lambda}$ should be close to y_k. Hence we choose that value of λ which makes, on the average, $|L_k f_{k,\lambda} - y_k|$ smallest. Specifically, we choose λ to minimize the function

$$V(\lambda) = (1/n) \sum_{k=1}^{n} (L_k f_{k,\lambda} - y_k)^2 w_k(\lambda). \tag{29}$$

The weights $w_k(\lambda)$ are chosen so that $V(\lambda)$ has the same minimizing value of λ as

$$T(\lambda) \cong (1/n) \sum_{i=1}^{n} (L_i f - L_i f_\lambda)^2,$$

where f is the actual solution and f_λ is the solution of (28) in the limit of large n.

Wahba (1977) shows under certain conditions that as the number of data points increases the cross-validation solution f_λ tends to f in the sense that

$$E\{\|f - f_\lambda\|\} \to 0 \tag{30}$$

where E denotes the expectation operator. In this analysis it is assumed that the errors in the data are normally distributed with mean zero and equal variance.

For computational purposes the formula given for $V(\lambda)$ is unwieldy. It can be shown (Craven and Wahba, 1979; Golub et al., 1979) that this formula is equivalent to

$$V(\lambda) = \frac{(1/n)\|(A + n\lambda I)^{-1}\mathbf{y}\|^2}{[(1/n)\,\mathrm{tr}\,(A + n\lambda I)^{-1}]^2}. \tag{31}$$

From (31) it is evident that $V(\lambda)$ is invariant under rotations of the matrix A and the data vector $\mathbf{y}$. In actual numerical work the most convenient method to evaluate (31) is to make use of the spectral decomposition of the matrix A. If $\hat{y}_i$ denotes the component of $\mathbf{y}$ in the direction of the ith normalized eigenvector and λ_i denotes the ith eigenvalue, then the formula for $V(\lambda)$ can be written in the equivalent form

$$V(\lambda) = n \sum_{i=1}^{n} \frac{\hat{y}_i^2}{(\lambda_i + n\lambda)^2} \Bigg/ \left[\sum_{i=1}^{n} \frac{1}{\lambda_i + n\lambda}\right]^2. \tag{32}$$

Once the spectral decomposition of A is known, the $\hat{y}_i$ can be evaluated. Then (32) can be easily evaluated in a minimum location routine without the need of any more algebra on the matrix A.

Numerical Tests of the Cross-Validation Algorithm

Numerical tests of the cross-validation algorithm were carried out both with and without simulated error added to the data. Log-normal distributions were used to produce simulated data using the kernel functions for the Marple impactor, and Simpson's rule for quadrature on a grid of 80 points over the size range was used. The representing functions were calculated using Eqs. (25) or (26) or equivalent ones in other Hilbert spaces. Table 1 summarizes the function spaces and representing functions considered. A was diagonalized by a Jacobi rotation method to give the spectral decomposition, and this was used in evaluating the cross-validation function to determine the proper value of λ by a global search technique. Once λ was determined the coefficients c_i were evaluated, and these coefficients multiplied by the appropriate representing functions were summed to give the final solution. Forty values of the distribution, evenly spaced over the size interval of interest, were calculated. Computationally, this method of solving the problem seems somewhat better than that used by Twomey in his linear inversion routine, since discrete values are assumed only for quadrature. All matrix algebra is done on relatively small systems whose dimension is equal to the number of data points. It is of interest to note that when no error was added to the data the cross-validation procedure always chose $\lambda = 0$, so the distributions were those that would have been obtained by the method discussed in the previous section.

Figures 7–13 show some of the results of the cross-validation inversion. Figures 7 and 8 give the results of inverting a unimodal and bimodal log-normal distribution, respectively. Also shown are the corresponding histograms obtained by assuming ideal impactor behavior. One set of impactor data was used, and no error was added.

TABLE 1. Function Spaces and Representing Functions Used in Inversion Algorithm

Function space	Representing functions	Comments
$H_0^1(0, 1)$ with norm $\|\|f\|\|^2 = \int_0^1 f'(x)^2\,dx$ and inner product $(f, g) = \int_0^1 f'(x)g'(x)\,dx$	$g_i(x) = \int_0^x tK_i(t)\,dt + x\int_x^1 K_i(t)\,dt - x\int_0^1 tK_i(t)\,dt,$ $g_i'(x) = \int_x^1 K_i(t)\,dt - \int_0^1 tK_i(t)\,dt$	Zero boundary conditions on solution at both ends of size interval. Solutions are continuous but derivatives need not be. Solution approximates actual distribution pointwise and derivative approximates derivative of actual distribution in mean square sense.
$H_0^2(0, 1)$ with norm $\|\|f\|\|^2 = \int_0^1 f''(x)^2\,dx$ and inner product $(f, g) = \int_0^1 f''(x)g''(x)\,dx$	$g_i(x) = \int_0^x (x - t)g_i''(t)\,dt,$ $g_i''(x) = \int_1^x (x - t)K_i(t)\,dt$	Zero boundary conditions on solution and its derivative at one end of size interval. Solution and derivative are continuous, and both are pointwise approximations to actual distribution and its derivative approximates second derivative of actual distribution in mean square sense.
$H_0^1(0, 1)$ with norm $\|\|f\|\|^2 = \int_0^1 f'(x)^2\,dx$ and inner product $(f, g) = \int_0^1 f'(x)g'(x)\,dx$	$g_i(x) = \int_0^x tK_i(t)\,dt + x\int_x^1 K_i(t)\,dt,$ $g_i'(x) = \int_x^1 K_i(t)\,dt$	Zero boundary condition on solution at one end of size interval. Otherwise similar to $H_0^1(0, 1)$.
$H^1(0, 1)$ with norm $\|\|f\|\|^2 = f(0)^2 + \int_0^1 f'(x)^2\,dx$ and inner product $(f, g) = f(0)g(0) + \int_0^1 f'(x)g'(x)\,dx$	$g_i(x) = \int_0^x tK_i(t)\,dt + x\int_x^1 K_i(t)\,dt + \int_0^1 K_i(t)\,dt,$ $g_i'(x) = \int_x^1 K_i(t)\,dt$	No boundary conditions. Otherwise similar to $H_0^1(0, 1)$.

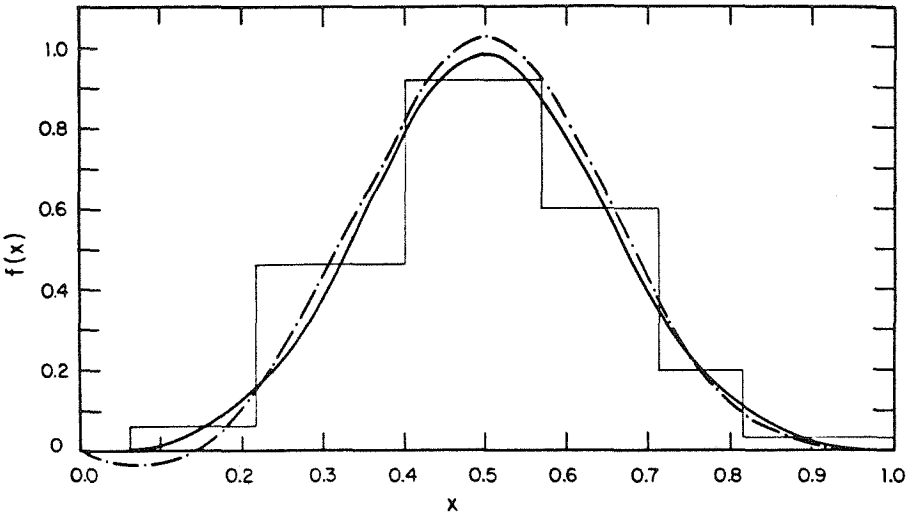
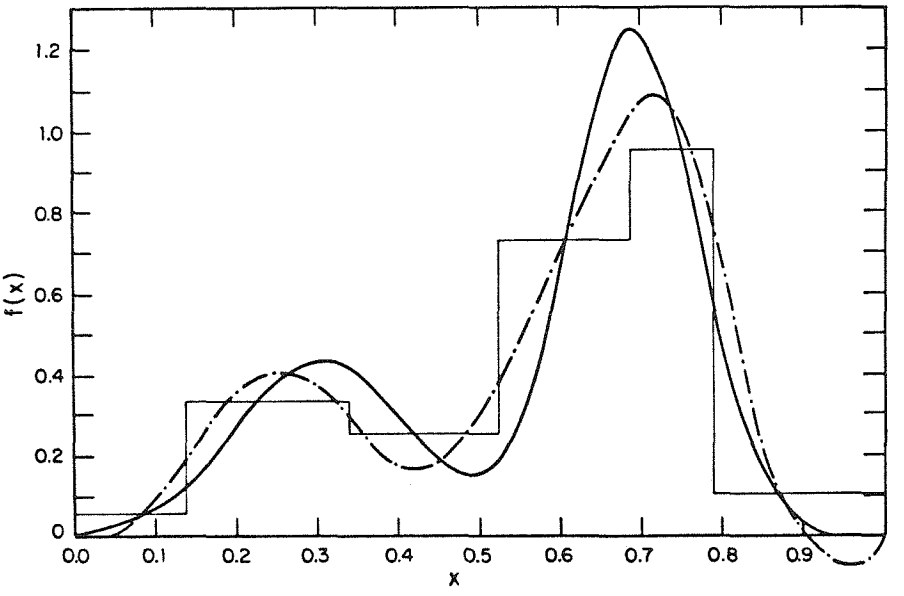


FIGURE 7. Size distribution obtained by inversion of simulated impactor data. Curve 1 (———) is the true lognormal distribution, and curve 2 (— · —) is the corresponding inverted distribution from cross validation in $H_0^1(0,1)$. Also shown is the histogram from 50% cutoffs.

Figures 9 and 10 show the same original distributions as Figures 7 and 8, but two sets of data were used in the inversion. Note the dramatic improvement in agreement with the data. It is interesting to note the negative values of the recovered distributions near the ends, which appears to occur because the impactor provides no information about the behavior of the distribution at the ends.

The second inverted distribution in Figure 10 shows the effect of adding error to the data. Uniformly distributed random error was added

FIGURE 8. Size distributions obtained by inversion of simulated impactor data. Curve 1 (———) is the true bimodal lognormal distribution, and curve 2 (—·—) is the corresponding inverted distribution from cross validation in $H_0^1(0,1)$. Histogram from 50% cutoffs is also shown.



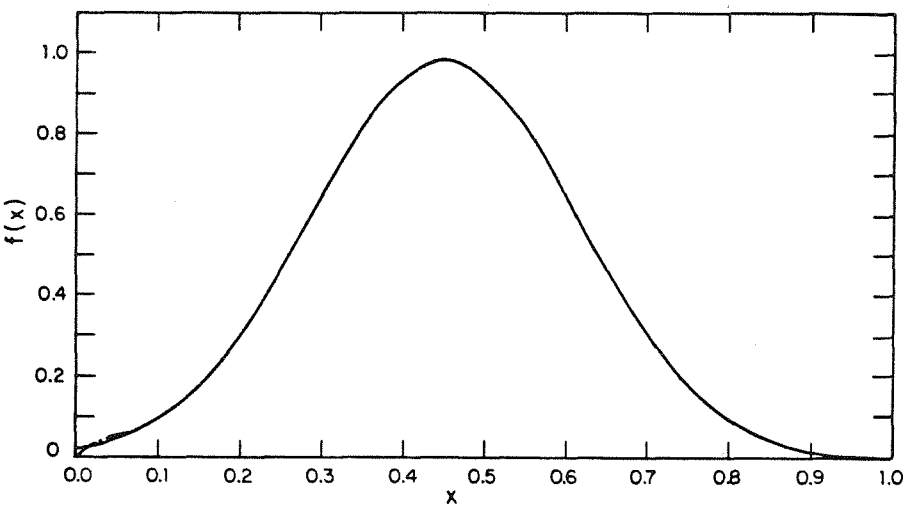
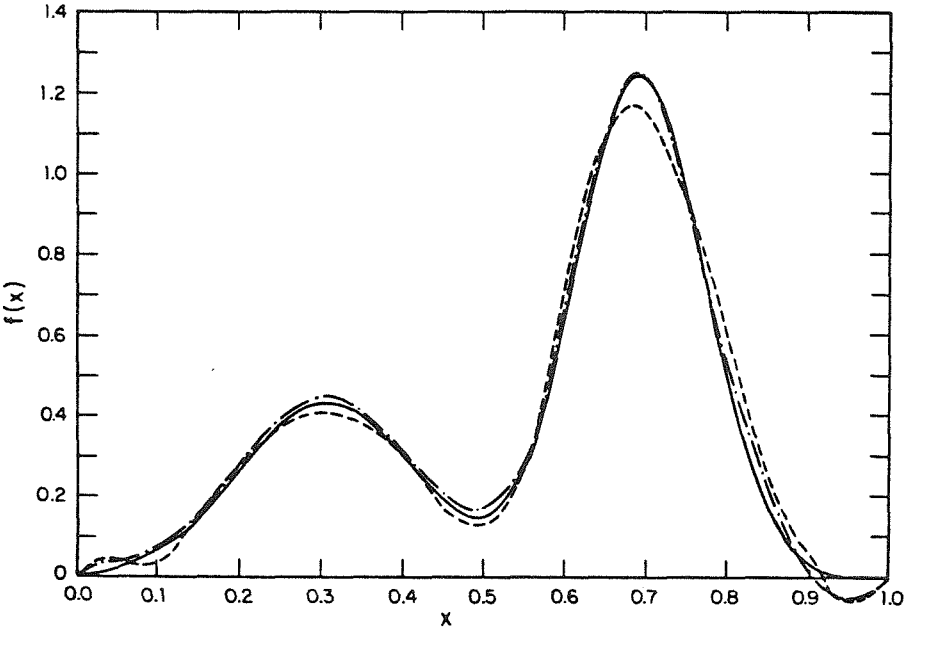


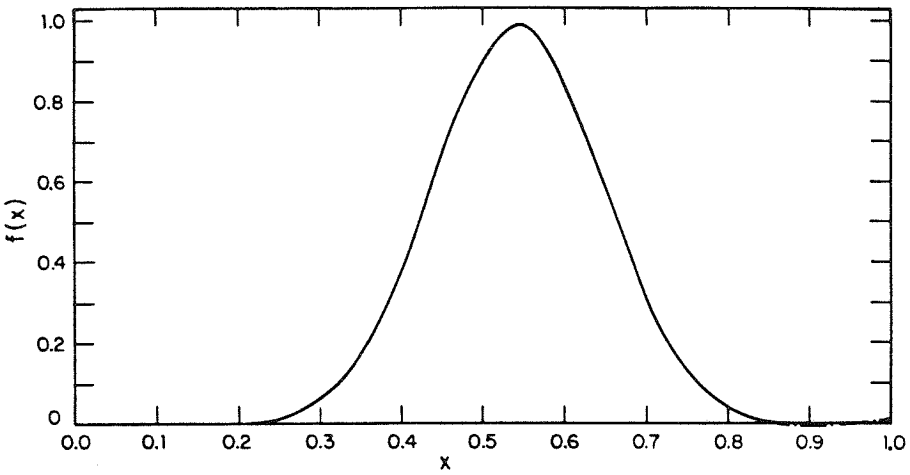
FIGURE 9. Size distributions obtained by inversion of simulated impactor data. Curve 1 (———) is the true lognormal distribution, and curve 2 (—·—) is the inverted distribution obtained by cross validation using two sets of data.

FIGURE 10. Size distributions obtained by inversion of simulated impactor data. Curve 1 (———) is the true bimodal lognormal distribution. Curve 2 (—·—) is the inverted distribution from cross validation. Curve 3 (– – –) is the distribution obtained by cross validation with error added to data (two data sets used).

to the data corresponding to a maximum of 3.3% of the maximum value of the data. Note that although the details of the recovered distribution change slightly, the agreement is still quite good.

Figures 11 and 12 show results of inversion in the space $H_0^2(0,1]$ using two sets of data. This space is similar to $H_0^1(0, 1]$ but satisfies the boundary conditions $f(0)=f'(0)=0$, and the norm is defined using the second derivative instead of the first. Note the Gibbs phenomenon in Figure 12. The distribution is really too





narrow for the instrument to measure accurately.

Figure 13 shows an inversion done in $H_0^1(0, 1]$.

FIGURE 11. Size distributions obtained by inversion of simulated impactor data. Curve 1 (———) is the true lognormal distribution, and curve 2 (— · —) is the inverted distribution from cross validation in $H_0^2(0,1)$ using two sets of data.

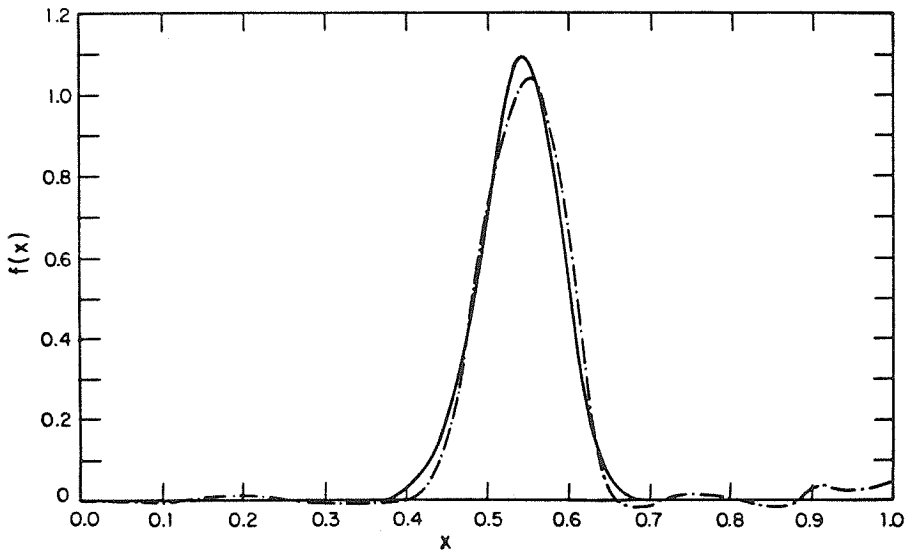
Hilbert Spaces Without Boundary Conditions

Up until now we have shown results only for spaces with certain boundary conditions. This is, in fact, overly restrictive, since there are many cases in which such conditions need not hold.

We could apply cross validation with no change to a space such as $H^1(0, 1)$, consisting of all absolutely continuous square integrable functions on (0, 1) whose derivatives are square integrable. The norm on this space is defined by

$$\|f\|^2 = \int_0^1 f(x)^2\,dx + \int_0^1 f'(x)^2\,dx. \qquad (33)$$

FIGURE 12. Size distributions obtained by inversion of simulated impactor data. Curve 1 (———) is the true lognormal distribution, and curve 2 (—·—) is the inverted distribution by cross validation in $H_0^2(0,1)$ using two data sets.



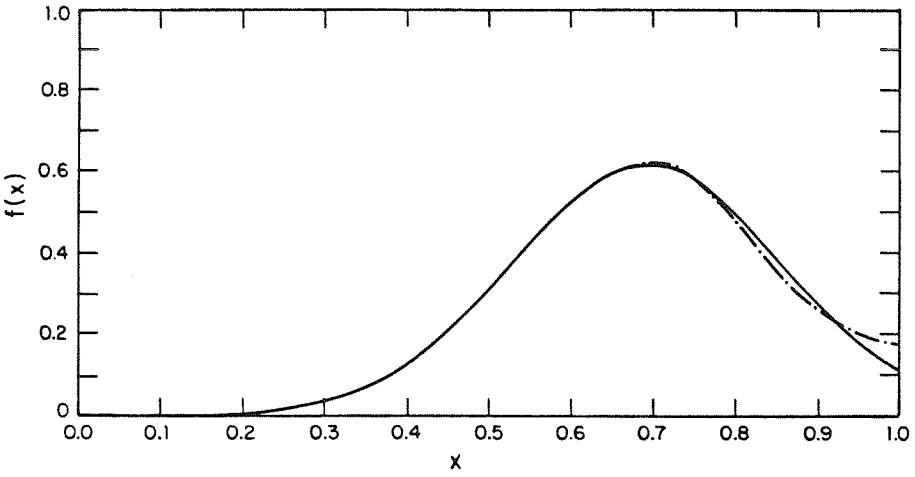


FIGURE 13. Size distributions obtained by inversion of simulated impactor data. Curve 1 (———) is the true lognormal distribution, and curve 2 (—·—) is the inverted distribution from cross validation in $H_0^1(0,1)$. using two data sets.

$H^1(0, 1)$ is a Hilbert space, so every continuous linear functional has a unique representing function. In this case the representing function for the functional L defined by

$$L(f) = \int_0^1 K(x)f(x)\,dx \tag{34}$$

is the solution g of the boundary value problem

$$g'' - g = -K(x), \qquad g'(0) = g'(1) = 0. \tag{35}$$

For the impactor these representing functions are flat near the ends of the dimensionless size interval and tend to peak gently somewhere in the middle near the 50% cutoff of the corresponding stage. The shape of these functions causes the matrix of inner products to be extremely ill conditioned. This happens because the functions are essentially equal to nonzero constants throughout a large part of the domain. As a result, the significant part of the inner product, the part containing the derivatives, is largely overshadowed by the other term, which plays no role in smoothing the function.

One way to alleviate this problem would be to subtract an appropriate constant from each representing function so that they would be normalized about zero in some sense. It turns out that this method works and is equivalent to a different approach, which we now outline.

Consider the space $H^1(0, 1)$ with the new norm

$$\|f\|^2 = f(0)^2 + \int_0^1 f'(x)^2\,dx \tag{36}$$

and the corresponding inner product. If we denote the old norm (33) by $\|\ \|^*$, then it can be shown that these two norms are equivalent in the sense that there exist strictly positive numbers k_1 and k_2 such that, for every f in $H^1(0, 1)$,

$$k_1\|f\|^* \le \|f\| \le k_2\|f\|^*. \tag{37}$$

Because of this equivalence good approximations in one norm are also good in the other. Instead of finding f to solve (28), we seek the function f_λ that minimizes

$$(1/n)\sum_{i=1}^{n}(L_i f - y_i)^2 + \lambda\int_0^1 f'(x)^2\,dx. \tag{38}$$

Note that we do not use the norm of f in the second term as in (28), because that would imply a restriction on the value of f at $x = 0$, which we do not wish to make.

Problem (38) can be reformulated in a way that makes it easy to solve by defining the operator P on $H^1(0, 1)$ as

$$Pf = f - f(0). \tag{39}$$

It is easy to show that P is self-adjoint and satisfies $P^2 = P$, so P is an orthogonal projection.

Hence (38) can be written as the following problem:

Find f

to minimize $(1/n)\sum_{i=1}^{n}(L_i f - y_i)^2 + \lambda \|Pf\|^2$. (40)

The solution to this problem is the solution to the equation

$$T^*Tf + n\lambda P^*Pf = T^*y, \tag{41}$$

where * denotes the adjoint of an operator, and T is an operator from $H^1(0, 1)$ into E^n defined by $Tf = (L_1 f, \cdots, L_n f)$. The solution to (41) is given in the Appendix, and we merely state it here:

$$f = f(0) + \sum_{i=1}^{n} c_i g_i, \tag{42}$$

where the g_i are the representing functions given in Table 1,

$$f(0) = \frac{(\mathbf{g}(0), (B + n\lambda I)^{-1}\mathbf{y})}{(\mathbf{g}(0), (B + n\lambda I)^{-1}\mathbf{g}(0))},$$

and $\mathbf{g}(0)$ is the vector with ith component $g_i(0)$ and the c_i are determined from

$$\mathbf{c} = (B + n\lambda I)^{-1}\mathbf{y} - f(0)(B + n\lambda I)^{-1}\mathbf{g}(0) \tag{43}$$

where B is the matrix whose ijth entry is

$$B_{ij} = \int_0^1 g_i'(x) g_j'(x)\, dx. \tag{44}$$

Note that this matrix will in general be better conditioned than the matrix of inner products of the g_i, with the inner product given by (36).

We have yet to determine the value of λ to be used in this inversion. This will be done by generalized cross validation. In general the cross-validation function (Golub et al., 1979) is defined by

$$V(\lambda) = \frac{n\|[I - F(\lambda)]\mathbf{y}\|^2}{\{\operatorname{tr}[I - F(\lambda)]\}^2}, \tag{45}$$

where $F(\lambda)$ is the matrix that maps the data vector $\mathbf{y}$ onto the vector $(L_1 f_\lambda, \ldots, L_n f_\lambda)$, and f_λ is the solution obtained using the parameter value λ.

To determine this matrix, first note that

$$\begin{aligned}(L_1 f_\lambda, \ldots, L_n f_\lambda) &= ((g_1, f_\lambda), \ldots, (g_n, f_\lambda)) \\ &= ((g_1, f_\lambda(0)), \ldots, (g_n, f_\lambda(0))) \\ &\quad + \left(\sum_{j=1}^{n}(g_1, g_j)c_j, \ldots, \sum_{j=1}^{n}(g_n, g_j)c_j\right) \\ &= f_\lambda(0)\mathbf{g}(0) + A\mathbf{c} = \mathbf{y} - n\lambda\mathbf{c}.\end{aligned} \tag{46}$$

The latter equality follows from the fact that f_λ satisfies (41). Thus, it follows that $F(\lambda)\mathbf{y} = \mathbf{y} - n\lambda\mathbf{c}$, so that $[I - F(\lambda)]\mathbf{y} = n\lambda\mathbf{c}$.

Let $\hat{y}_i$ and $\hat{g}_i(0)$ be the components of $\mathbf{y}$ and $\mathbf{g}(0)$, respectively, in the direction of the ith normalized eigenvector of B. Then we can write $V(\lambda)$ as (47), see below, where the λ_i are the eigenvalues of B.

Note that the matrix of inner products A is not used in the calculation of V, but the better-conditioned B is used instead.

Actually, there is no reason why the particular norm chosen must be used. We could have taken the usual norm on $H^1(0\,1)$ given by (33), and used the orthogonal projection operator P defined by

$$Pf = f - \int_0^1 f(x)\, dx \tag{48}$$

and carried out analogous reasoning to arrive at another expression for a cross-validation function.

Figure 14 shows the inverted distribution of the lognormal distribution shown in Figure 5. The result can be compared with that of the Twomey algorithm in Figure 5. No error was added to the data in either case.

$$V(\lambda) = \frac{n\sum_{i=1}^{n}\left[\frac{\hat{y}_i}{\lambda_i + n\lambda} - \frac{\hat{g}_i(0)}{\lambda_i + n\lambda}\left(\sum_{j=1}^{n}\frac{\hat{g}_j(0)\hat{y}_j}{\lambda_j + n\lambda} \Big/ \sum_{j=1}^{n}\frac{\hat{g}_j(0)^2}{\lambda_j + n\lambda}\right)\right]^2}{\left[\sum_{i=1}^{n}\frac{1}{\lambda_i + n\lambda} - \frac{\hat{g}_i(0)^2}{(\lambda_i + n\lambda)^2} \Big/ \sum_{j=1}^{n}\frac{\hat{g}_j(0)^2}{\lambda_j + n\lambda}\right]^2}. \tag{47}$$

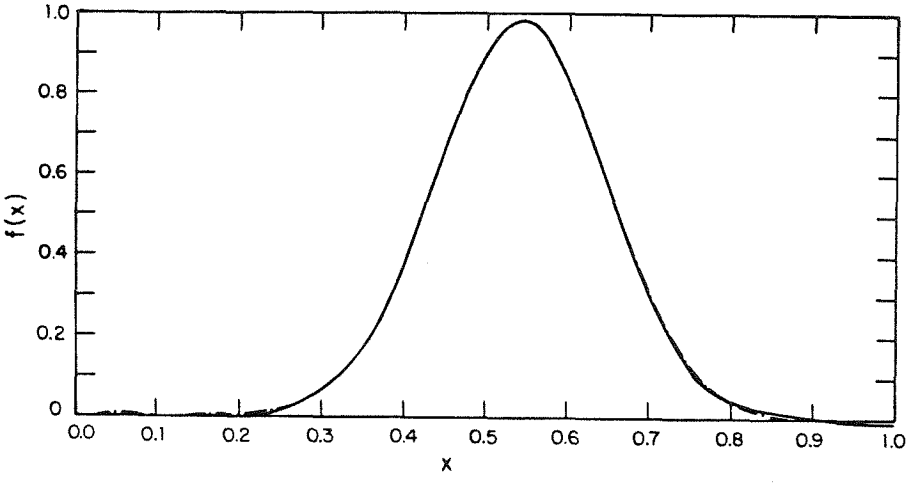


FIGURE 14. Size distributions obtained by inversion of simulated impactor data. Curve 1 (———) is the true lognormal distribution, and curve 2 (—·—) is the inverted distribution from cross validation in $H^1(0,1)$ using two data sets.

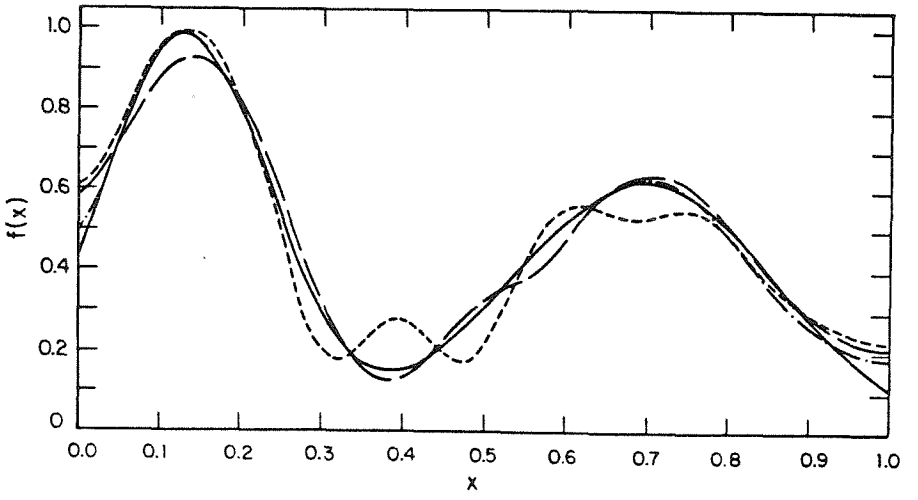
Figure 15 shows the inversion done in $H^1(0, 1)$ of the bimodal distribution shown in Figure 3. This inversion can be compared with those obtained by Twomey's nonlinear method shown in Figures 3 and 4.

Figure 15 also shows the cross-validation inversion of the same data but with uniformly distributed random error added to the data, corresponding to a maximum error of 3.5% of the maximum value of the data. It is of importance that the errors in the recovered distribution are roughly of the same order of magnitude as the errors in the data.

FIGURE 15. Size distributions obtained by inversion of simulated impactor data. Curve 1 (———) is the true bimodal lognormal distribution. Curve 2 (—·—) is the inverted distribution with no error added to data. Curves 3 (– – –) and 4 (—— ——) are inverted distributions with error added to data. Two sets of data were used in all areas.

Remarks

In the generalized cross validation as we have presented it here, the errors in the data are assumed to have equal variance. This need not



be assumed, however, and following Wahba and Wendelberger (1980) we can replace the datum y_i in the cross-validation function by y_i/σ_i, and the representing function g_i by g_i/σ_i, where σ_i is proportional to the standard deviation of the ith measurement. We often know the relative accuracy of the data but less often can estimate the absolute accuracy.

SUMMARY OF THE INVERSION ALGORITHM

Here we present a synopsis of the steps in the algorithm developed in the preceding sections.

1. Choose an appropriate function space. Table 1 lists those we have used in the construction of numerical examples and gives some properties of approximations in those spaces.
2. (a) For spaces with boundary conditions, evaluate the representing functions and the symmetric matrix A of inner products. See Table 1 for formulas of representing functions and definitions of inner products.
   (b) For $H^1(0, 1)$ (no boundary conditions), calculate representing functions g_i and the matrix B of inner products of the functions $g_i - g_i(0)$.

$$B_{ij} = (g_i - g_i(0),\ g_i - g_i(0)) = \int_0^1 g_i'(x) g_j'(x)\, dx.$$

3. Calculate the eigenvalues and eigenvectors of A or B. Check to make sure the eigenvalues are all positive.
4. Using the spectral decomposition obtained in step 3, evaluate and minimize the cross-validation function $V(\lambda)$ given by (32) for spaces with boundary conditions or (47) for $H^1(0, 1)$.
5. (a) For spaces with boundary conditions, determine the coefficient vector $\mathbf{c}$ by (27). Then the solution f is given by

$$f = \sum_{i=1}^{n} c_i g_i.$$

   (b) For $H^1(0, 1)$, the solution f is given by (42) and (43).

CONCLUSIONS

We have presented a general discussion of the problem of inversion of aerosol size distribution data and shown the shortcomings of presently available techniques. We have suggested the method of generalized cross validation as an alternative. This method is seen to be especially well suited to the ill-posed nature of the inversion problem and to perform better in numerical examples than the Twomey nonlinear inversion method.

APPENDIX
SOLUTION OF EQ. (41)

Equation (41) is

$$T^*Tf + n\lambda P^*Pf = T^*\mathbf{y} \tag{A.1}$$

where

$$Tf = (L_1 f, \ldots, L_n f) = ((g_1, f), \ldots, (g_n, f)).$$

Now, the adjoint T^* is defined by $(Tf, \mathbf{y}) = (f, T^*\mathbf{y})$, where this is to hold for all $f \in H^1(0, 1)$, and all $\mathbf{y} \in \mathbb{R}^n$. Thus

$$(Tf, \mathbf{y}) = \sum_{i=1}^{n} (g_i, f) y_i = \left(\sum_{i=1}^{n} y_i g_i, f \right) = (f, T^*\mathbf{y}),$$

where the latter inner product is taken in $H^1(0, 1)$. Hence $T^*\mathbf{y} = \sum_{i=1}^{n} y_i g_i$. Next we have

$$T^*Tf = \sum_{i=1}^{n} (g_i, f) g_i.$$

Hence Eq. (A.1) may be written

$$\sum_{i=1}^{n} (g_i, f) g_i + n\lambda Pf = \sum_{i=1}^{n} y_i g_i.$$

where we have used the fact that P is an orthogonal projection. Applying P to f gives

$$\sum_{i=1}^{n} (g_i, f) g_i + n\lambda f - n\lambda f(0) = \sum_{i=1}^{n} y_i g_i. \tag{A.2}$$

From this equation it follows that $f - f(0)$ lies in the linear span of the g_i. We may therefore write

$$f = f(0) + \sum_{i=1}^{n} c_i g_i.$$

Substituting into Eq. (A.2) gives

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(g_i g_j)c_j g_i + n\lambda\sum_{i=1}^{n}c_i g_i + f(0)\sum_{i=1}^{n}g_i(0)g_i = \sum_{i=1}^{n}y_i g_i. \tag{A.3}$$

Let $\mathbf{g}(0)$ be the n-dimensional vector whose ith component is $g_i(0)$, and A the nxn matrix with ijth entry (g_i, g_j). Then Eq. (A.3) becomes

$$(A+n\lambda I)\mathbf{c} = \mathbf{y} - f(0)\mathbf{g}(0). \tag{A.4}$$

Hence

$$\mathbf{c} = (A+n\lambda I)^{-1}\mathbf{y} - f(0)(A+n\lambda I)^{-1}\mathbf{g}(0). \tag{A.5}$$

$f(0)$ is unknown, but by taking the inner product (in $\mathbb{R}^n$) of (A.5) with $\mathbf{g}(0)$ we find

$$\sum_{i=1}^{n}c_i g_i(0) = (\mathbf{g}(0), (A+n\lambda I)^{-1}\mathbf{y}) - f(0)(\mathbf{g}(0), (A+n\lambda I)^{-1}\mathbf{g}(0)).$$

But since

$$f - f(0) = \sum_{i=1}^{n}c_i g_i,$$

we have

$$0 = f(0) - f(0) = \sum_{i=1}^{n}c_i g_i(0),$$

so it follows that

$$f(0) = \frac{(\mathbf{g}(0), (A+n\lambda I)^{-1}\mathbf{y})}{(\mathbf{g}(0), (A+n\lambda I)^{-1}\mathbf{g}(0))}.$$

Finally, note that, if B is the matrix whose ijth entry is $(g_i - g_i(0), g_j - g_j(0))$, then using

$$\sum_{i=1}^{n}c_i g_i(0) = 0$$

it is easily shown that

$$A\mathbf{c} = B\mathbf{c}. \tag{A.6}$$

Hence, the matrix A can be replaced by the better-behaved matrix B in the above calculations to give

$$f(0) = \frac{(\mathbf{g}(0), (B+n\lambda I)^{-1}\mathbf{y})}{(\mathbf{g}(0), (B+n\lambda I)^{-1}\mathbf{g}(0))},$$

$$\mathbf{c} = (B+n\lambda I)^{-1}\mathbf{y} - f(0)(B+n\lambda I)^{-1}\mathbf{g}(0),$$

$$f = f(0) + \sum_{i=1}^{n}c_i g_i.$$

This work was supported in part by a grant from the Exxon Education Foundation.

REFERENCES

Adams, R. A., (1975). *Sobolev Spaces. Academic, New York.*

Backus, G. E., and Gilbert, J. F. (1967). *Geophys. J. Roy. Astron. Soc.* 13:247–276.

Backus, G. E., and Gilbert, J. F. (1968). *Geophys. J. Roy. Astron. Soc.* 16:169–205.

Cooper, D. W., and Spielman, J. (1976). *Atmos. Environ.* 10:723–729.

Craven, P. and Wahba, G. (1979). *Numer. Math.* 31:377–403.

Golub, G. H., Heath, M., and Wahba, G. (1979). *Technometrics* 21:215–223.

Jackson, D. D. (1972). *Geophys. J. Roy. Astron. Soc.* 28:97–109.

Marple, V. A., and Liu, B. Y. H. (1974). *Environ. Sci. Technol.* 8:648–654.

Phillips, D. L. (1962). *J. Assoc. Comput. Mach.* 9:84–97.

Rudin, W. (1973). *Functional Analysis.* McGraw-Hill, New York.

Tihonov, A. N. (1963a). *Soviet Math.* 4:1035–1038.

Tihonov, A. N. (1963b). *Soviet Math.* 4:1624–1627.

Twomey, S. (1963). *J. Assoc. Comput. Mach.* 10:97–101.

Twomey, S. (1975). *J. Comput. Phys.* 18: 188–200.

Wahba, G. (1975). *Numer. Math.* 24: 383–393.

Wahba, G. (1977). *SIAM J. Numer. Anal.* 14:651–667.

Wahba, G., and Wendelberger, J. (1980). *Mon. Weather Rev.* 108:1122–1143.

Received 12 November 1981; accepted 26 November 1981

CHAPTER 7

FURTHER RESULTS ON INVERSION OF AEROSOL SIZE DISTRIBUTION DATA - HIGHER ORDER SOBOLEV SPACES AND CONSTRAINTS

Accepted for Publication in Aerosol Science and Technology

FURTHER RESULTS ON INVERSION OF AEROSOL SIZE DISTRIBUTION DATA - HIGHER ORDER SOBOLEV SPACES AND CONSTRAINTS

James G. Crump and John H. Seinfeld
Department of Chemical Engineering
California Institute of Technology
Pasadena, California 91125

ABSTRACT

The aerosol size distribution inversion algorithm of Crump and Seinfeld (1982), based on the concept of regularization with generalized cross-validation, is extended to Sobolev spaces of order m. The use of the cross-validation function for choice of an appropriate value of m in a particular application is discussed. An inversion algorithm that constrains the size distribution to be non-negative is introduced and shown to be of value for sharply peaked distributions.

INTRODUCTION

The aerosol size distribution inversion problem can be expressed as determining the size distribution function f(x) to satisfy

$$\Lambda_i f = y_i \qquad i = 1,2,\ldots, n \tag{1}$$

where Λ_i are known linear functionals and y_i are the measured data. Equation (1) usually has the concrete representation,

$$\int_0^1 k_i(x)f(x)\,dx = y_i \qquad i = 1,2,\ldots, n \tag{2}$$

where the k_i are kernel functions for the particular instrument. The size variable x has been taken to be dimensionless and normalized so that $0 \leq x \leq 1$.

Crump and Seinfeld (1982) presented an algorithm for inversion of aerosol size distribution data. In particular, they showed that if f is assumed to lie in a Hilbert space H, then f(x) can be approximated as

$$f(x) = \sum_{i=1}^{n} c_i \phi_i(x) \tag{3}$$

where the $\phi_i(x)$ are representing functions for the functionals Λ_i, and the coefficient vector c is determined by

$$c = (A + n\lambda I)^{-1} y \tag{4}$$

where A is the nxn matrix with ijth entry (ϕ_i,ϕ_j), the parentheses denoting the inner product in H. The positive parameter λ is determined by minimizing the cross-validation function $V(\lambda)$, defined by

$$V(\lambda) = \sum_{k=1}^{n} |\Lambda_k f_\lambda^k - y_k|^2 \, w_k(\lambda) \tag{5}$$

where f_λ^k denotes the solution of the same problem with the parameter value λ omitting the $\underline{k}$th datum, and $w_k(\lambda)$ are weights chosen to make the approximate solution converge to the actual solution under certain conditions (Golub et al., 1979).

The inversion algorithm of Crump and Seinfeld (1982) is based on the concept of regularization (to be explained shortly) and cross-validation and was shown to give superior performance when compared to other available inversion algorithms. In that work inversion formulas in the first-order Sobolev space $H^1(0,1)$ were presented. To achieve greater flexibility in fitting size distribution data it is useful to develop results for inversion in higher order Sobolev spaces $H^m(0,1)$, where $m > 1$. The properties of the inversion formulas in $H^m(0,1)$ depend on m, and, in particular, certain sets of data can be fit better by different choices of the order m. In this paper we develop inversion algorithms for the general Sobolev space $H^m(0,1)$, and discuss the choice of m for a particular application. Finally, although aerosol size distribution functions are inherently non-negative quantities, most inversion algorithms do not explicitly include a non-negative constraint. In some applications involving sharply peaked distributions it is useful to incorporate a non-negativity constraint into the inversion algorithm. Consequently, we present here an algorithm for constrained inversion and illustrate its use.

DATA INVERSION IN THE SOBOLEV SPACES $H^m(0,1)$

The space $H^m(0,1)$, for m a positive integer, is defined to be the collection of all real functions on the interval (0,1) having square integrable weak derivatives through order m (Adams, 1975). It may be shown that these functions have continuous derivatives through order (m-1), and that a norm may be defined on this space by

$$\|f\|^2 = \sum_{k=0}^{m-1} f^{(k)}(0)^2 + \int_0^1 f^{(m)}(x)^2 dx \tag{6}$$

where the superscripts denote the order of differentiation. The inner product of two functions f and g is defined as

$$(f,g) = \sum_{k=0}^{m-1} f^{(k)}(0)g^{(k)}(0) + \int_0^1 f^{(m)}(x)g^{(m)}(x)\, dx \tag{7}$$

The general data inversion problem in $H^m(0,1)$ may now be stated as follows. Find a function f in $H^m(0,1)$ minimizing

$$\frac{1}{n} \sum_{i=1}^{n} |\Lambda_i f - y_i|^2 + \lambda \int_0^1 f^{(m)}(x)^2 dx \tag{8}$$

where λ is to be determined by generalized cross-validation. The concept of adding the second term in (8) to the customary squared deviation first term is called regularization, in that the second term attempts to insure a degree of smoothness or regularity in the solution.

This problem can be conveniently treated by defining an orthogonal projection operator P on $H^m(0,1)$ by

$$Pf = f - \sum_{k=0}^{m-1} x^k f^{(k)}(0)/k! \tag{9}$$

Now if T denotes the linear map from $H^m(0,1)$ to $\mathscr{R}^n$ given by

$$Tf = (\Lambda_1 f, \ldots, \Lambda_n f) \tag{10}$$

The minimization functional (8) becomes

$$\frac{1}{n} \|Tf-y\|^2 + \lambda \|Pf\|^2 \tag{11}$$

where $\|\cdot\|$ refers to the usual norm on $\mathscr{R}^n$ in the first term, and to the norm defined by (6) on $H^m(0,1)$ in the second term.

Minimization of (11) is equivalent to solving the following equation for f,

$$T^*Tf + n\lambda P^*Pf = T^*y \tag{12}$$

where * denotes the operator adjoint. Since P is an orthogonal projection, $P^* = P$, and $P^*P = P^2 = P$. The adjoint of T is defined by

$$(Tf,y) = (f,T^*y) \tag{13}$$

for all $f \in H^m(0,1)$ and all $y \in \mathscr{R}^n$. It should be noted that the parentheses on the left in (13) refer to the inner product in $\mathscr{R}^n$, while those on the right refer to the inner product on $H^m(0,1)$. Since T is an operator from $H^m(0,1)$ to $\mathscr{R}^n$, T* is an operator from $\mathscr{R}^n$ to $H^m(0,1)$. It is well known that T* is uniquely defined (Rudin, 1973).

Let $\phi_1, \ldots, \phi_n$ be the representing functions for the linear functionals $\Lambda_1, \ldots, \Lambda_n$, respectively. Thus, for every function $f \in H^m(0,1)$

$$\Lambda_i f = (\phi_i, f) \qquad i = 1,2,\ldots, n \tag{14}$$

Equation (14) uniquely defines the representing functions (Rudin, 1973). The adjoint T* may be shown to be given by

$$T^*y = \sum_{i=1}^{n} y_i \phi_i \tag{15}$$

by substitution of this expression into (13).

Substituting (15) into (12) gives

$$\sum_{i=1}^{n} (\phi_i f)\phi_i + n\lambda \left[f - \sum_{k=0}^{m-1} x^k f^{(k)}(0)/k! \right] = \sum_{i=1}^{n} y_i \phi_i \tag{16}$$

Equation (16) implies that

$$f - \sum_{k=0}^{m-1} x^k f^{(k)}(0)/k!$$

lies in the linear span of the representing functions. Hence, there exist constants $c_1, \ldots, c_n$ such that

$$f - \sum_{k=0}^{m-1} x^k f^{(k)}(0)/k! = \sum_{i=1}^{n} c_i \phi_i \tag{17}$$

Substituting (17) into (16) and assuming linear independence of the ϕ_i, which is equivalent to linear independence of the Λ_i, yields, in matrix form,

$$(A + n\lambda I)c = y - \sum_{k=0}^{m-1} f^{(k)}(0)\phi^{(k)}(0) \tag{18}$$

where $A_{ij} = (\phi_i, \phi_j)$, and $\phi^{(k)}(0)$ is the n-vector with ith component $\phi_i^{(k)}(0)$.

Evaluating (17) and its derivatives through order m-1 at x = 0 gives

$$\sum_{i=1}^{n} c_i \phi_i^{(k)}(0) = 0 \qquad k = 0,1,\ldots, m-1 \tag{19}$$

Multiplying (18) by $(A + n\lambda I)^{-1}$ yields

$$c = (A + n\lambda I)^{-1}y - \sum_{k=0}^{n-1} f^{(k)}(0)(A + n\lambda I)^{-1}\phi^{(k)}(0) \tag{20}$$

Now, taking inner products in $\mathscr{R}^n$ of (20) with $\phi^{(j)}(0)$ for j = 0,1,..., m-1 gives, after using (19),

$$0 = \left(\phi^{(j)}(0),(A + n\lambda I)^{-1}y\right) - \sum_{k=0}^{m-1} f^{(k)}(0)\left(\phi^{(j)}(0),(A + n\lambda I)^{-1}\phi^{(k)}(0)\right) \tag{21}$$

If we let Φ be the mxn matrix with ijth element $\phi_j^{(i)}(0)$, then we have from (21)

$$\begin{bmatrix} f(0) \\ \vdots \\ f^{(m-1)}(0) \end{bmatrix} = [\Phi(A + n\lambda I)^{-1}\Phi^{+}]^{-1}\Phi(A + n\lambda I)^{-1}y \tag{22}$$

and

$$c = (A + n\lambda I)^{-1}y - (A + n\lambda I)^{-1}\Phi^{+}[\Phi(A + n\lambda I)^{-1}\Phi^{+}]^{-1}\Phi(A + n\lambda I)^{-1}y \tag{23}$$

To determine the cross-validation function $V(\lambda)$, we note that if $c = F(\lambda)y$ for some matrix $F(\lambda)$, then (Crump and Seinfeld, 1982)

$$V(\lambda) = \frac{n\|F(\lambda)y\|^2}{(\mathrm{tr}F(\lambda))^2} \tag{24}$$

Hence, $V(\lambda)$ is given by (24) with

$$F(\lambda) = (A + n\lambda I)^{-1} - (A + n\lambda I)^{-1}\Phi^{+}[\Phi(A + n\lambda I)^{-1}\Phi^{+}]^{-1}\Phi(A + n\lambda I)^{-1} \qquad (25)$$

By substitution into (14) and integration by parts, it may be verified that the representing functions ϕ_i are given by

$$\phi_i(x) = (-1)^m \int_0^x \int_0^{t_m} \cdots \int_0^{t_2} \psi_i^m(t_1)dt_1\, dt_2 \cdots dt_m$$

$$+ \sum_{k=0}^{m-1} (-1)^k x^k \int_0^1 \psi_i^k(t)dt/k! \qquad (26)$$

where

$$\psi_i^0(x) = k_i(x)$$

$$\psi_i^{k+1}(x) = \int_1^x \psi_i^k(t)dt \qquad (27)$$

and the $k_i(x)$ are the instrument kernel functions from (2).

Special Case of $H^2(0,1)$

In the special case of $m = 2$, the inverted distribution $f(x)$ is obtained from

$$f(x) = f(0) + xf'(0) + \sum_{i=1}^{n} c_i\phi_i(x) \qquad (28)$$

where

$$c = (M + n\lambda I)^{-1}\left(y - f(0)\phi(0) - f'(0)\phi'(0)\right) \qquad (29)$$

M is the nxn matrix with

$$M_{ij} = \int_0^1 \phi_i''(x)\phi_j''(x)\,dx \tag{30}$$

and

$$\phi_i(x) = \frac{x^3}{6}\int_0^x k_i(t)dt - \frac{x^2}{2}\int_0^x tk_i(t)dt + \frac{x}{2}\int_0^x t^2k_i(t)dt$$

$$- \frac{1}{6}\int_0^x t^3k_i(t)dt + (1 - \frac{x^3}{6})\int_0^1 k_i(t)dt$$

$$+ (x + \frac{x^2}{2})\int_0^1 tk_i(t)dt \tag{31}$$

$$\phi_i''(x) = x\int_0^x k_i(t)dt - \int_0^x tk_i(t)dt + \int_0^1 tk_i(t)dt$$

$$- x\int_0^1 k_i(t)dt \tag{32}$$

with

$$f(0) = \left\{ B\sum_{j=1}^{n} \frac{\hat{\phi}_j(0)\hat{y}_j}{\lambda_j+n\lambda} - C\sum_{j=1}^{n} \frac{\hat{\phi}_j'(0)\hat{y}_j}{\lambda_j+n\lambda} \right\} \Big/ E \tag{33}$$

$$f'(0) = \left\{ A\sum_{j=1}^{n} \frac{\hat{\phi}_j'(0)\hat{y}_j}{\lambda_j+n\lambda} - C\sum_{j=1}^{n} \frac{\hat{\phi}_j(0)\hat{y}_j}{\lambda_j+n\lambda} \right\} \Big/ E \tag{34}$$

The λ_j are the eigenvalues of M. $\hat{\phi}_j(0)$, $\hat{\phi}_j'(0)$, and $\hat{y}_j$ are the components of the n-vectors, $\phi(0)$, $\phi'(0)$ and y, respectively, in the direction of the jth eigenvector of M and

$$A = \sum_{j=1}^{n} \frac{\hat{\phi}_j(0)^2}{\lambda_j+n\lambda} \tag{35}$$

$$B = \sum_{j=1}^{n} \frac{\hat{\phi}_j'(0)^2}{\lambda_j+n\lambda} \tag{36}$$

$$C = \sum_{j=1}^{n} \frac{\hat{\phi}_j(0)\hat{\phi}_j'(0)}{\lambda_j+n\lambda} \tag{37}$$

with $E = AB-C^2$.

Finally, the scalar λ is determined by minimizing the cross-validation function

$$V(\lambda) = \frac{n \sum_{i=1}^{n} \left[\frac{\hat{y}_i}{\lambda_i+n\lambda} - \frac{\hat{\phi}_i(0)f(0)}{\lambda_i+n\lambda} - \frac{\hat{\phi}_i'(0)f'(0)}{\lambda_i+n\lambda} \right]^2}{\left\{ \sum_{i=1}^{n} \left[\frac{1}{\lambda_i+n\lambda} - \left(\frac{\hat{\phi}_i(0)^2 B}{(\lambda_i+n\lambda)^2} - \frac{2\hat{\phi}_i(0)\hat{\phi}_i'(0)C}{(\lambda_i+n\lambda)^2} + \frac{\hat{\phi}_i'(0)^2 A}{(\lambda_i+n\lambda)^2} \right) \Big/ E \right] \right\}^2} \tag{38}$$

CHOICE OF SPACE FOR INVERSION

We have developed above general results for data inversion in the Sobolev space $H^m(0,1)$. In solving a specific problem it is necessary to select the order m of the space to be used. In this section we discuss the factors involved in that choice. In short, the effect of choosing a larger value of m, i.e. higher Sobolev space, for inversion of data is to produce smoother inverted distributions.

To see this, consider a simple example. Let

$$k_i(x) = \sqrt{2} \sin \pi i x \tag{39}$$

Writing f(x) as a sine series,

$$f(x) = \sum_{i=1}^{\infty} c_i \sqrt{2} \sin \pi i x \tag{40}$$

the inversion problem assumes the form $c_i = y_i$, $i = 1,2,\ldots, n$. Since the data y_i inevitably contain noise, the simple solution of substituting y_i for c_i in (40) is unsatisfactory because some high frequency components will have larger amplitudes than they ought to. By using the foregoing theory for inversion in $H^m(0,1)$ for even m, we obtain the alternative expression,

$$c_i = \frac{y_i}{(1+n\lambda i^{2m})} \qquad i = 1,2,\ldots, n \tag{41}$$

If λ is small enough, the coefficients c_i behave like y_i for small i. However, for large i, $c_i \sim y_i/n\lambda i^{2m}$. Hence, high frequencies are removed from the inverted solution. Note that as m increases, the cutoff of high frequency components becomes sharper.

As a result of this discussion, we see that if a smooth inverted distribution is expected, a higher order Sobolev space is appropriate for inversion. On the other hand, a sharply peaked distribution is likely to be better inverted in a lower order space.

One way to choose the proper order m without resorting to purely intuitive criteria is to use the cross-validation function. Since, at its minimum, $V(\lambda)$ is an approximation to

$$\frac{1}{n} \sum_{i=1}^{n} (\Lambda_i f - \Lambda_i f^*)^2 + \sigma^2 \tag{42}$$

where f^* is the true solution, f is the inverted distribution, and σ is the standard deviation of the data (Golub et al., 1979), a good choice of m is that which gives the minimum value of $V(\lambda)$.

Our experience indicates that the inverted distribution is not too sensitive to the space used in inversion, although in higher order spaces the inversion tends to be more ill-conditioned. That is, the increased smoothness of the representing functions causes these functions to become more nearly linearly dependent, and hence increases the condition number of the Gram matrix A of Equation (20).

Figure 1 shows the results of inverting some data from the low pressure impactor in $H^1(0,1)$ and $H^2(0,1)$. Also shown is the histogram based on the 50 percent cutoffs. Note that the two solutions agree well within the size range where the impactor is most sensitive. The deviation below the lowest 50 percent cutoff is due to the different ways in which data are extrapolated in the two spaces. In this case the lower value of the cross-validation function $V(\lambda)$ occurred in $H^2(0,1)$, so this solution is expected to be better.

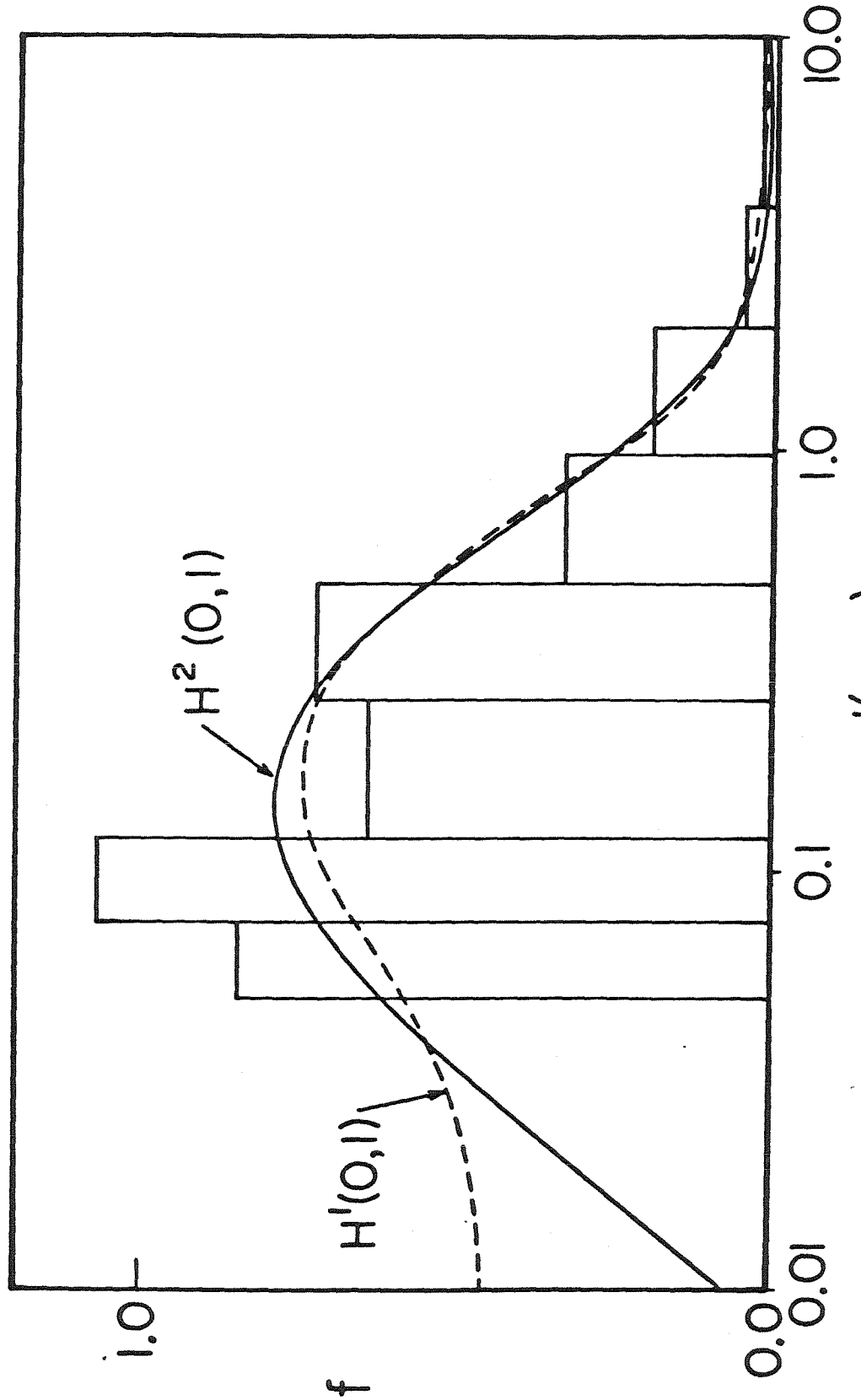


Figure 1. Inversion of data from an 8-stage low pressure impactor (Hering et al., 1979) using $H^1(0,1)$ and $H^2(0,1)$. The histogram based on 50 percent cutoffs for the impactor is also shown.

CONSTRAINED INVERSION

Since aerosol size distributions are necessarily nonnegative quantities, inversion with positivity constraints can be expected to yield better results in some cases than unconstrained inversion. This fact turns out to be particularly important when the inverted distribution is sharply peaked. In the unconstrained case, attempting to recover the peak often leads to the so-called Gibbs phenomenon, i.e. high frequency oscillations at the base of the peak. These oscillations appear to be suppressed when a positivity constraint is applied. In this section we describe an algorithm for constrained inversion of aerosol size distribution data.

The inversion problem is formulated as in (8), but with the added constraint, $f(x) \geq 0$. We reduce the problem to a finite-dimensional one by defining discrete approximations, $f_j = f(x_j)$, where $x_j = (j-1)/m$, $j = 1,2,\ldots, m+1$. Furthermore, we define a kernel matrix K, the elements of which are $K_{ij} = k_i(x_j)$, $i = 1,2,\ldots, n$, $j = 1,2,\ldots, m+1$. Now the problem can be expressed as follows. Find a vector $f \geq 0$ in $\mathcal{R}^{m+1}$ minimizing

$$\sum_{i=1}^{n}\left[\sum_{j=1}^{m+1} K_{ij} f_j \nu_j - y_i\right]^2 + n\lambda \sum_{j=1}^{m+1} (Df)_j^2 \nu_j \tag{43}$$

where the ν_j are coefficients for quadrature using Simpson's rule,

$$\begin{aligned} \nu_1 &= \nu_{m+1} = \frac{1}{3m} \\ \nu_{2i} &= \frac{4}{3m} \qquad i = 1,2,\ldots, m/2 \\ \nu_{2i+1} &= \frac{2}{3m} \qquad i = 1,2,\ldots, m/2-1 \end{aligned} \tag{44}$$

and D is a difference operator approximating a differential operator. For example, in $H^2(0,1)$, D is defined as

$$
\begin{aligned}
D_{1j} &= D_{m+1,j} = 0 && j = 1,2,\ldots, m+1 \\
D_{ii} &= -2m^2 && i = 2,3,\ldots, m \\
D_{i,i-1} &= D_{i,i+1} = m^2 && i = 2,3,\ldots, m \\
D_{ij} &= 0 && \text{otherwise}
\end{aligned} \tag{45}
$$

And in $H^1(0,1)$,

$$
\begin{aligned}
D_{i,i+1} &= m && i = 1,2,\ldots, m \\
D_{ii} &= -m && i = 1,2,\ldots, m \\
D_{ij} &= 0 && \text{otherwise}
\end{aligned} \tag{46}
$$

Letting $K'_{ij} = K_{ij}\nu_j$, $i = 1,2,\ldots, n$; $j = 1,2,\ldots, m+1$, and $D'_{ij} = D_{ij}\nu_i$, $i = 1,2,\ldots, m+1$; $j = 1,2,\ldots, m+1$, the problem becomes: Find $f \geq 0$ in $\mathscr{R}^{m+1}$ to minimize

$$
\left(f,(K'^{+}K' + n\lambda D^{+}D')f\right) - (f,2K^{+}y) + \|y\|^2 \tag{47}
$$

This is a quadratic programming problem that can be solved by standard procedures (Duffin et al., 1967). Using the Kuhn-Tucker theorem, the problem reduces to: Find $f \geq 0$ in $\mathscr{R}^{m+1}$ such that,

$$
Nf - d = g \tag{48}
$$

where $g \geq 0$ and $(f,g) = 0$, and where $N = K'^{+}K + n\lambda D^{+}D'$ and $d = K^{+}y$.

Figure 2 shows the result of constrained and unconstrained inversion of data produced from a narrow log normal distribution simulated for an eight-stage low pressure impactor. Note that the constrained inversion gives much better results in this case. Both inversions were done in $H^2(0,1)$. In carrying out the constrained inversion, the problem was solved using an algorithm given by Ravindran (1972) based on theory of Lemke (1965).

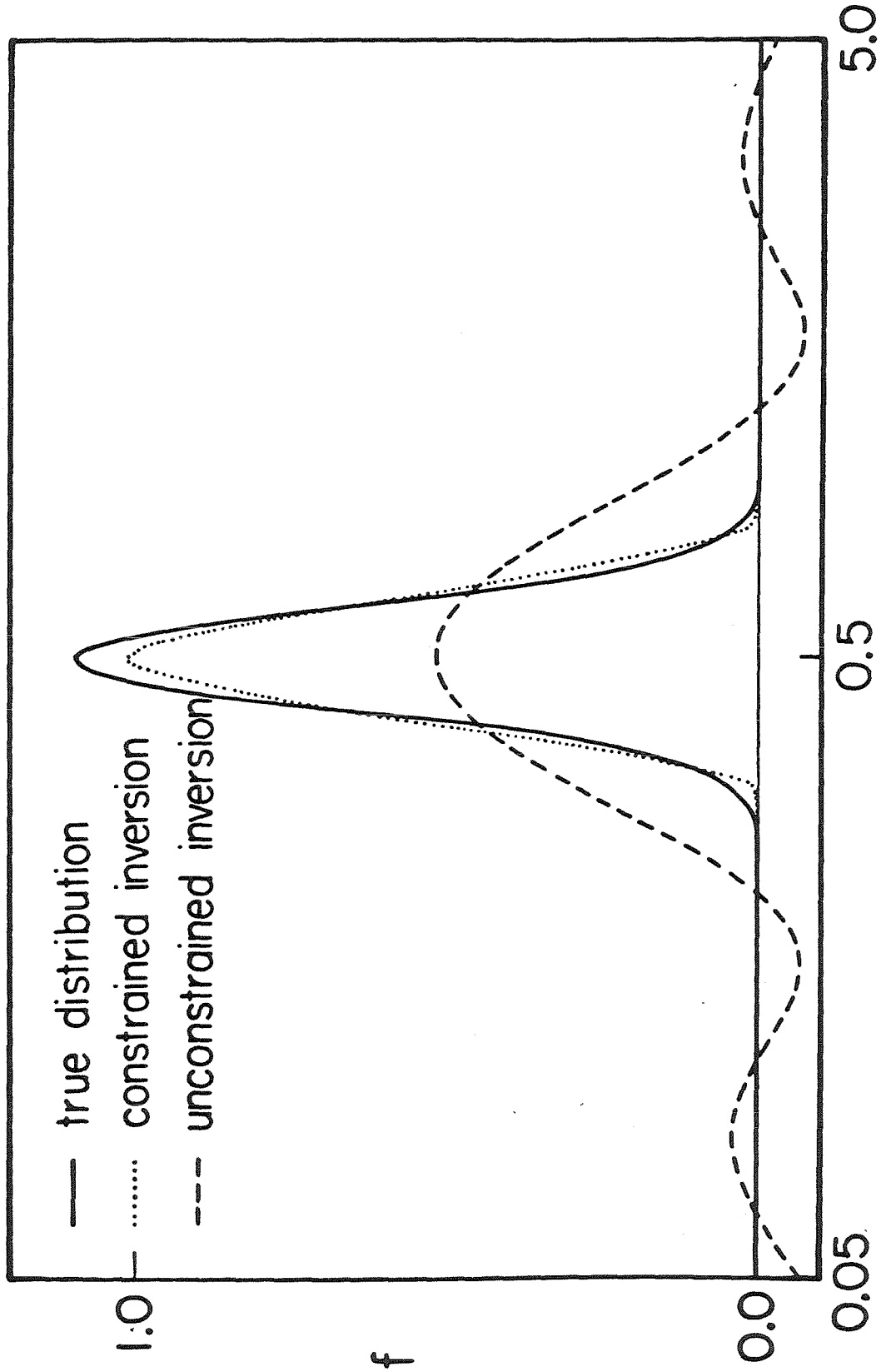


Figure 2. Constrained and unconstrained inversion of data produced from a narrow log normal size distribution simulated for an 8-stage low pressure impactor.

Unlike the unconstrained case, cross-validation is not applicable in the constrained inversion, so the parameter λ must be chosen by an alternate method. One such method is the method of the discrepancy (Tikhonov and Arsenin, 1977), by which f is obtained by solving (48) and λ is determined iteratively by the requirement that

$$\| K'f - y \| = \delta \tag{49}$$

where δ is a measure of the error in the data,

$$\delta = \left(\sum_{i=1}^{n} \sigma_i^2 \right)^{\frac{1}{2}} \tag{50}$$

where σ_i is the standard deviation of y_i.

CONCLUSIONS

Formulas for applying the method of regularization with generalized cross-validation to aerosol size distribution data inversion problems have been given for Sobolev spaces of order m. The use of the cross-validation function for choice of appropriate value of m has been noted, and the advantages of using a constrained inversion algorithm have been shown when the distribution is sharply peaked.

The inversion algorithms developed by Crump and Seinfeld (1982) as well as those given in the present paper have been incorporated into the two computer programs INVERSE and CINVERSE. These programs were reported by Crump (1982) in the computer program section of the journal and are available from the authors.

REFERENCES

Adams, R. A. (1975) Sobolev Spaces, Academic, New York.

Crump, J. G. (1982) INVERSE Aerosol Sci. and Tech., 1, XXX-XXX.

Crump, J. G. and Seinfeld, J. H. (1982) Aerosol Sci. and Tech., 1, 1-20.

Duffin, R. J., Peterson, E. E. and Zener, C. (1967) Geometric Programming - Theory and Application, John Wiley, New York.

Golub, G. H., Heath, M. and Wahba, G. (1979) Technometrics 21, 215-223.

Hering, S. V., Friedlander, S. K., Collins, J. J. and Richards, L. W. (1979) Environ. Sci. Tech. 13, 184-188.

Lemke, C. E. (1965) Management Science 11, 681-689.

Ravindran, A. (1972) Comm. ACM 15, 818-820.

Rudin, W. (1973) Functional Analysis, McGraw-Hill, New York.

Tikhonov, A. N. and Arsenin, V. Y. (1977) Solutions of Ill-Posed Problems, Halsted Press, Washington.

Wahba, G. (1977) SIAM J. Numer. Anal. 14, 651-667.

CHAPTER 8

CALIBRATION OF OPTICAL PARTICLE COUNTER

CALIBRATION OF OPTICAL PARTICLE COUNTER

INTRODUCTION

The size measuring instrument used in the aerosol growth studies was a Royco LAC 226 laser aerosol particle counter (OPC) having 15 channel resolution between 0.12 μm and 6.1 μm particle diameter and a channel for oversize particles.

Size resolution of the instrument is imperfect, and experimentally it is found that monodisperse aerosols typically give significant responses in several channels. Consequently, in order to obtain good size distribution measurements, calibration of the instrument and an appropriate data inversion technique are necessary.

In the following sections we describe the method of calibration, present the data obtained from experimental calibrations, and discuss the application of the data to inversion of measured size distribution data using the methods of Crump and Seinfeld (1982a, 1982b).

CALIBRATION METHOD

The goal of calibration is to determine kernel functions k_i for the ith channel, defined such that if f is the number distribution of particles entering the instrument, and y_i is the number concentration counted in the ith channel, then f, k_i, and y_i are related by

$$\int_0^1 k_i(x)f(x)dx = y_i \qquad (1)$$

where x is a dimensionless size variable. By using a monodisperse distribution, values of the kernel functions can be obtained experimentally, for if

$$f(x) = N\delta(x - x_o) \tag{2}$$

substitution into Equation (1) gives

$$Nk_i(x_o) = y_i \tag{3}$$

Consequently, if the total number concentration N and the measured response y_i are known for a monodisperse aerosol of dimensionless size x_o, then the values of the kernel function k_i at the same particle size can be found from Equation (3).

By repeating the same measurements for several particle sizes over the size range of interest, the values of the kernel functions over the particle size range may be obtained by interpolation.

EXPERIMENTAL PROCEDURE

Monodisperse particles of polystyrene latex (PSL) and sodium chloride ranging in diameter from 0.1 μm to 1.01 μm were used to calibrate the Royco instrument. Sodium chloride particles were used for smaller particles and PSL for larger sizes.

In the PSL experiments dilute suspensions of the monodisperse spheres were atomized in the device previously described, and the resulting aerosol was passed through a Kr-85 charge neutralizer before entering the particle counter. For each size measured, five separate measurements were taken, each lasting one minute. The channel responses were then averaged. The diameters of PSL particles used in these experiments were 0.126 μm, 0.234 μm, 0.312 μm, 0.357 μm, 0.500 μm, 0.600 μm, 0.721 μm, 0.792 μm, and 1.01 μm.

Several measurements of particle concentrations were also made simultaneously with a TSI model 3030 Electrical Aerosol Analyzer (EAA). These

measurements agree roughly with the OPC measurements for 0.357 μm particles, but in the case of 0.126 μm particles, the OPC yielded concentrations smaller than those of the EAA by a factor of between 40 and 60. This, likely, is due to the low signal to noise level of the OPC at this small particle size. For this particle size (0.126 μm diameter) it was assumed that that the EAA gave the correct concentration, and the data from the OPC were normalized by dividing by this value (see Equation (3)). For the other (larger) particle sizes, it was assumed that the OPC gave the correct number concentration. The results of these experiments are tabulated in Table I as the values of the kernel functions for the first six channels of the OPC at the particle sizes used in the experiments.

Several experiments were also done using monodisperse sodium chloride particles in order to obtain more values of the kernel functions near the lower end of the size range. The particle diameters used ranged from 0.10 μm to 0.13 μm.

In these experiments aerosol was generated by atomizing a dilute solution of sodium chloride. A monodisperse fraction of the desired size was then separated out by passing the aerosol through a TSI model 3071 Electrical Mobility Classifier. The resulting monodisperse aerosol was then passed through a Kr-85 charge neutralizer before making size measurements. Both OPC and EAA measurements were made, the latter being used to obtain values of the total concentration. As in the previous experiments, five separate measurements were made for each particle size used, and the resulting OPC responses and EAA measurements averaged. The values of the kernel functions for the first six channels obtained from these experiments are shown in Table II.

Table 1. OPC Calibration data for PSL

diameter (μm)	k_1	k_2	k_3	k_4	k_5	k_6
0.126	0.022	0.001	0.000	0.000	0.000	0.000
0.234	0.153	0.844	0.003	0.000	0.000	0.000
0.312	0.031	0.204	0.763	0.001	0.000	0.000
0.357	0.030	0.140	0.827	0.003	0.000	0.000
0.500	0.035	0.012	0.289	0.661	0.003	0.002
0.600	0.029	0.008	0.190	0.766	0.006	0.000
0.721	0.035	0.016	0.012	0.271	0.655	0.010
0.792	0.025	0.016	0.017	0.176	0.725	0.038
1.01	0.024	0.014	0.013	0.036	0.331	0.577

Table II. OPC Calibration Data for NaCl

diameter (μm)	k_1	k_2	k_3	k_4	k_5	k_6
0.10	0.063	0.010	0.000	0.000	0.000	0.000
0.11	0.092	0.016	0.001	0.000	0.000	0.000
0.12	0.114	0.024	0.001	0.000	0.000	0.000
0.13	0.135	0.045	0.002	0.000	0.000	0.000

DISCUSSION

The values of the kernel functions obtained for 0.126 μm PSL tend to disagree with the values obtained for NaCl with diameters 0.12 μm and 0.13 μm. This is probably due to shape irregularities of the NaCl particles and the difference in refractive indices of the two materials.

For the small particles, the NaCl measurements were chosen to be used in calibration since they appear very similar to measurements made with $MnSO_4$, the material used in growth experiments.

Ideally, the instrument should be calibrated with particles having the same light scattering characteristics as those to be measured, but the difficulty of producing monodisperse particles of the proper size and composition precludes this.

The values of the kernel functions obtained as described in the last section, omitting the results of 0.126 μm PSL, were interpolated at 41 equally logarithmically spaced diameters from 0.1 μm to 1.0 μm,* and these interpolated values were incorporated into the data inversion program CINVERSE (Crump and Seinfeld, 1982b, Crump 1982).

*The kernel function data were interpolated since the experimental data provided only a small number of points across the diameter interval 0.1 m to 1.0 m.

The technique used was to find a smooth interpolating function $k_i^*(x)$ for each kernel function $k_i(x)$, such that for each particle size $x_j (j=1,\ldots,n)$

$$k_i^*(x_j) = k_i(x_j)$$

and each k_i minimizes the functional

$$\int_0^1 k_i^{*''}(x)dx$$

Here the variable x is defined by

$$x = \log d_p + 1$$

and x varies between 0 and 1.

The algorithm used to carry out the interpolation was a modification of CINVERSE.

Figure 1 shows a plot of the six kernel functions together with the ideal rectangle kernel functions obtained from the channel size cutoffs supplied by the manufacturer. It can be seen from this figure that there is considerable deviation between the ideal curves and the experimental curves.

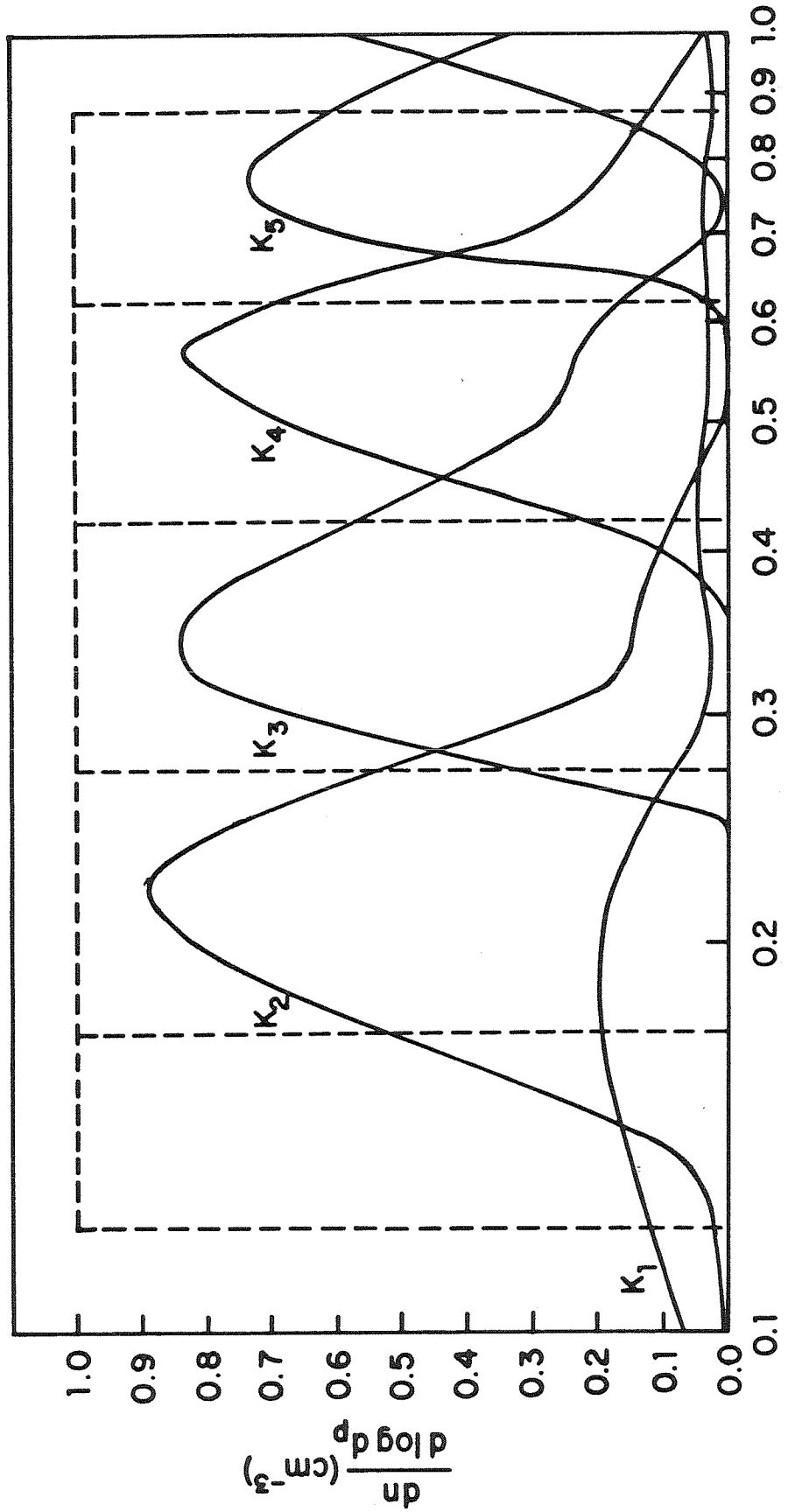


Figure 1. Kernel functions k_j for first six channels of OPC (solid curves). Dotted histogram represents ideal kernel functions assuming no cross-sensitivity and using manufacturer's diameter values for channel boundaries.

References

Crump, J. G. (1982) INVERSE Aerosol Sci. and Tech., 1, XXX-XXX

Crump, J. G. and Seinfeld, J. H. Aerosol Sci. and Tech., 1, 1-20

Crump, J. G. and Seinfeld, J. H. submitted to Aerosol Sci. and Tech.

CHAPTER 9

ON EXISTENCE OF STEADY STATE SOLUTIONS TO THE COAGULATION EQUATIONS

Accepted for Publication in the Journal of Colloid and Interface Science

ON EXISTENCE OF STEADY STATE SOLUTIONS TO THE COAGULATION EQUATIONS

James G. Crump and John H. Seinfeld
Department of Chemical Engineering
California Institute of Technology
Pasadena, California 91125

ABSTRACT

This work examines the general steady state coagulation equations with sources and sinks. These equations are shown to admit physically unacceptable solutions in some cases, and it is hypothesized that in such a case a gelation or precipitation phenomenon occurs. General conditions on the coagulation and loss coefficients are given that insure the existence of physically realistic solutions. The physically realistic solutions have tails that decay faster than any power of particle size.

1. INTRODUCTION

White (1) has considered certain properties of the steady state solutions of the population balance equation including sources, coagulation, and removal, i.e.

$$0 = a_k + \frac{1}{2} \sum_{i+j=k} b_{ij} x_i x_j - x_k \sum_j b_{kj} x_j - c_k x_k \qquad [1]$$

The two nonlinear terms describe the rate of change of concentration by binary agglomeration, a_k is the rate of introduction of particles of size k, and $-c_k x_k$ accounts for removal of particles of size k, such as by convection, sedimentation, or diffusion to the walls of the system.

For systems of physical interest the coefficients b_{ij} and c_k are unbounded, and this fact makes it difficult to analyze the behavior of solutions to Eq. [1]. In addition, as White has shown, it also gives rise to physically unrealistic solutions in some cases.

A solution to Eq. [1] is simply any nonnegative sequence x_k satisfying Eq. [1]. The existence of a solution implies that all the sums appearing in the equations are finite. However, the existence of a solution does not imply, contrary to widespread belief, that the solution conserves mass, that is, that the following equality holds

$$\sum_k k c_k x_k = \sum_k k a_k \qquad [2]$$

This equality may be "derived" by multiplying the kth equation in Eq. [1] by k and summing over all k. However, the rearrangements of the infinite sums required to obtain Eq. [2] are only valid provided $\sum_{i,j} i b_{ij} x_i x_j < \infty$; for then, all series appearing in the equations are absolutely convergent. We will see later, nevertheless, that the inequality

$$\sum_k kc_k x_k \leq \sum_k ka_k \qquad [3]$$

always holds regardless of b_{ij} so that even if mass is not strictly conserved as in Eq. [2], mass is always finite for any solution of Eq. [1].

Another property that solutions to Eq. [1] may possess is that of having finite moments of all orders, that is

$$\sum_{k=1}^{\infty} k^{\gamma} x_k < \infty \qquad \gamma = 0,1,2,... \qquad [4]$$

We will see in a later section that solutions need not have this property.

A physically acceptable solution to Eq. [1] must have the mass conservation property. The fact that Eq. [1] may have solutions which do not obey Eq. [2] was demonstrated by White (1), who proved the following:

1) If all $c_k = 0$, i.e., there is no sink for particles, then solutions to Eq. [1] may or may not exist depending on the form of the b_{ij}. Solutions which do exist do not conserve mass.
2) If all $c_k = 0$, and Eq. [1] has a solution, then some moment of the solution must be infinite.
3) If the removal and coagulation coefficients satisfy the relations, $c_j > cj^{\beta}$ and $b_{ij} \leq bi^{\beta}j^{\beta}$, with $c \neq 0$, then any solution of Eq. [1] has finite moments of all orders.

The assertion above that mass is not conserved in the absence of removal is hardly surprising. What is surprising is the existence of solutions to the steady state equations under such conditions. These solutions are not physically acceptable, and the second statement shows that one characteristic of such solutions of that some moment of them is always infinite. The third result states a condition under which this situation will not occur, which, roughly speaking,

is that the removal mechanism be at least as "strong" as the coagulation mechanism.

From these results White concluded that physically acceptable solutions of Eq. [1] cannot have power law tails, i.e., the solutions cannot have asymptotic forms $x_k \sim k^{-r}$ for large k, since in such a case, some moment would be infinite, contradicting the third result above.

However, White's results leave several questions unanswered:

1) How "strong" must the removal mechanism be relative to the coagulation mechanism in order that physically acceptable solutions to Eq. [1] exist?
2) Do solutions to Eq. [1] actually exist under general conditions when the coefficients b_{ij} and c_k are unbounded?
3) What is the significance of "unphysical" solutions to Eq. [1]?

As an example of the type of case not answered by White's work, consider Brownian coagulation of an aerosol in a vessel with removal occurring only by flow out of the vessel (convection). Then $b_{ij} \leq bi^{1/3}j^{1/3}$ and $c_k = 1$. The results of White do not include this case, but we will see that solutions to this system exist, conserve mass, and have finite moments of all orders.

White showed that in the absence of removal physically acceptable solutions do not exist. However, even when removal is present, physically acceptable solutions may fail to exist, as the following example of branched chain polymerization in a continuous stirred tank reactor indicates. In this case the kinetic coefficients b_{ij} are proportional to (i+2)(j+2) (2), and the removal, by convection, is constant.

Hence, if we assume only monomer is entering the reactor, Eqs. [1] take the form

$$x_k = \lambda\delta_{1k} + \frac{1}{2}\sum_{j=1}^{k-1} b_{k-j,j}x_{k-j}x_j - x_k\sum_{j=1}^{\infty} b_{kj}x_j \qquad [5]$$

where λ is a dimensionless feed concentration of monomer, and the coefficients $b_{ij} = (i+2)(j+2)$. Suppose that a mass conserving solution to Eq. [5] exists. Then the mass conservation equation holds, i.e.

$$\sum_k k x_k = \lambda \qquad [6]$$

Summation of Eq. [5] over k gives

$$\sum_k x_k = \lambda - \frac{1}{2} \sum_{i,j} b_{ij} x_i x_j \qquad [7]$$

$$= \lambda - \frac{1}{2} \sum_{i,j} (i+2)(j+2) x_i x_j$$

and from Eq. [7], we have, using Eq. [6],

$$\sum_k x_k \leq \lambda - \frac{1}{2} \lambda^2 \qquad [8]$$

It follows immediately from Eq. [8] that if $\lambda > 2$, $\sum_k x_k < 0$, so no physically acceptable solution exists.

Physically, this tells us that gelation has occurred in the reactor. That is, some amount of infinite polymer has formed. In fact, this does actually occur in cross-linked polymerizations, and the above coagulation kernel has been used to model the polymerization up to the gel point, and to estimate the time for gelation (2). Whether or not a phenomenon similar to gelation exists in aerosol systems seems to be unknown at the present time, although Smoluchowski (3) reports an experiment carried out by Paine in which colloidal particles being stirred coagulated negligibly during an induction period, then suddenly produced large clusters. This may have been due to gelation.

These last comments suggest that in some cases at least, the nonexistence of physically acceptable solutions to Eq. [1] may be an indication that precipitation similar to gelation in polymer systems has occurred in the system, and hence, although the solutions themselves do not have physical significance, they indicate a real physical phenomenon.

2. STATEMENT OF RESULTS

In the next sections we prove two results in answer to the questions raised by White. These two results give conditions under which well-behaved solutions to Eq. [1] exist. Then we also discuss the behavior of solutions which do not satisfy these conditions, and in so doing show what situations might give rise to gelation phenomena. We assume that $\sum k^{\gamma} a_k < \infty$ for all γ.

We will give a partial answer to the first question raised by proving the following assertion:

Assertion 1.

If the removal coefficients c_i and the coagulation coefficients b_{ij} satisfy the relations $c_i \geq ci^{\beta}$ for some $c > 0$, $\beta \geq 0$, and $b_{ij} \leq bi^{\alpha+\beta}j^{\alpha+\beta}$, where $\alpha < \frac{1}{2}$, then if solutions to Eq. [1] exist, they have finite moments of all orders, and consequently conserve mass.

It should be noted that the existence of a solution to Eq. [1] is not asserted here, but merely that if a solution exists, its moments are all finite under the given conditions. There are some indications that the exponent $\frac{1}{2}$ is the largest possible here, and we will give some arguments to this effect later. Nevertheless, proof is still lacking. Assertion 1 tells us that under the given conditions if solutions to Eq. [1] exist no gelation phenomena will occur, and thus, the assertion gives some indication of the conditions needed to observe such a phenomenon. It also should serve as a warning to those who solve Eq. [1]

numerically that if the above conditions are not met, the solutions obtained may not be meaningful.*

The second question will be answered under fairly general conditions, applicable to most aerosol systems, and it is given in the following assertion:

Assertion 2.

If the coefficients c_i and b_{ij} satisfy the same conditions given in Assertion 1 , except that we merely require that $\alpha < 1$, then there exists a non-negative solution to Eq. [1] having finite mass.

Assertion 2 is especially interesting in light of the fact that it guarantees the existence of solutions to Eq. [1], which do not satisfy the conditions of Assertion 1, and consequently may be unphysical. These solutions are those for which $\alpha > \frac{1}{2}$. No such solutions, however, have ever been found, to the knowledge of the authors. Nevertheless, we will present arguments which suggest that such solutions do indeed exist.

Before proving Assertions 1 and 2, we note that the substitution $x'_k = c_k x_k$ reduces Eqs. [1] to

$$0 = a'_k + \frac{1}{2} \sum_{i+j=k} b'_{ij} x'_i x'_j - x'_k \sum_j b'_{kj} x'_j - x'_k \qquad [9]$$

where $b'_{ij} = b_{ij}/c_i c_j$. From now on we omit the primes in Eq. [9] and deal only with this equation. Clearly if either Assertion 1 or 2 holds for Eq. [9], it also holds for Eq. [1]. Moreover, by the hypothesis of the two assertions, we may assume that the b_{ij} of Eq. [9] satisfy the relations $b_{ij} \leq bi^{\alpha} j^{\alpha}$, where $\alpha < \frac{1}{2}$ for proving Assertion 1 and $\alpha < 1$ for proving Assertion 2.

---

*Similar comments, may be expected to hold for continous, or mixed discrete-continuous forms of Eq. [1].

Proof of Assertion 1

Assertion 1 is basically an extension of White's result, and its proof is nearly identical to that of White. We make use of the following inequalities proved in Appendix A:

For $r \geq 2$, there is a number C_r such that for all $a,b \geq 1$

$$(a+b)^r \leq a^r + b^r + C_r a^{r-1} b^{r-1} \qquad [10]$$

For $1 \leq r \leq 2$ and all $a, b \geq 1$

$$(a+b)^r \leq a^r + b^r + 2a^{r/2} b^{r/2} \qquad [11]$$

Then we obtain the following generalizations of the inequality proved by White: If $\{x_k\}$ is a solution to Eq. [9], then

$$\sum_k k^p x_k \leq \sum_k k^p a_k + \frac{C_p}{2} \sum_i \sum_j i^{p-1} j^{p-1} b_{ij} x_i x_j \qquad (p \geq 2) \qquad [12a]$$

$$\sum_k k^p x_k \leq \sum_k k^p a_k + \frac{C_p}{2} \sum_i \sum_j i^{p/2} j^{p/2} b_{ij} x_i x_j \qquad (1 < p \leq 2) \qquad [12b]$$

$$\sum_k k x_k \leq \sum_k k a_k \qquad [12c]$$

where C_p is a constant depending only on p.

The proof of these inequalities is obtained by letting N tend to infinity in the following inequalities (where C_p is chosen according to Eq. [10] if $p \geq 2$, Eq. [11] if $1 < p \leq 2$, and $C_p = 0$ $p = 1$ and $r = p - 1$ for $p \geq 2$ and $p/2$ for $1 < p \leq 2$).

$$\sum_{k \leq N} k^p x_k - \sum_{k \leq N} k^p a_k$$

$$= \frac{1}{2} \sum_{k \leq N} k^p \sum_{i+j=k} b_{ij} x_i x_j - \sum_{k \leq N} k^p x_k \sum_j b_{kj} x_j$$

$$= \frac{1}{2} \sum_{k \le N} \sum_{i+j=k} (i+j)^p \, b_{ij} x_i x_j - \sum_{k \le N} \sum_j k^p b_{kj} x_k x_j$$

$$\le \frac{1}{2} \sum_{k \le N} \sum_{i+j=k} (i^p + j^p + C_p i^r j^r) b_{ij} x_i x_j - \sum_{k \le N} \sum_j k^p b_{kj} x_k x_j$$

$$= \frac{C_p}{2} \sum_{k \le N} \sum_{i+j=k} i^r j^r b_{ij} x_i x_j + \sum_{k \le N} \sum_{i+j=k} i^p b_{ij} x_i x_j - \sum_{k \le N} \sum_j k^p b_{kj} x_k x_j$$

$$= \frac{C_p}{2} \sum_{k \le N} \sum_{i+j=k} i^r j^r b_{ij} x_i x_j - \sum_{k \le N} \sum_{N-k<j} k^p b_{kj} x_k x_j$$

$$\le \frac{C_p}{2} \sum_{k \le N} \sum_{i+j=k} i^r j^r b_{ij} x_i x_j \, .$$

To prove Assertion 1, we show first that existence of a finite second moment implies that all moments are finite, under the hypothesis that $b_{ij} \le bi^\alpha j^\alpha$ with $\alpha < 1$. Let $\delta = 1 - \alpha > 0$. Set $\gamma_1 = 2 + \delta$. Since $\gamma_1 > 2$, inequality [12a] may be invoked to give

$$\sum_k k^{\gamma_1} x_k - \sum_k k^{\gamma_1} a_k \le \frac{bC\gamma_1}{2} \left(\sum_i i^{\gamma_1 + \alpha - 1} x_i \right)^2$$

$$= \frac{bC\gamma_1}{2} \left(\sum_i i^2 x_i \right)^2 < \infty$$

Hence, the $(2+\delta)$-th moment is finite. Now setting $\gamma_n = 2+n\delta$, and proceeding inductively, we see that all moments are finite.

Now assume $\alpha < \frac{1}{2}$. By what we just saw, it is only necessary to prove that the second moment is finite to prove that all moments are finite. This is shown by using the inequality [12b]

$$\sum_k k^2 x_k - \sum_k k^2 a_k \le \frac{bC_2}{2} \left(\sum_i i^{1+\alpha} x_i \right)^2 \quad .$$

Thus, the second moment is finite if the $(1+\alpha)$-th moment is. Now, repeating this argument shows that this moment is finite if the $(\frac{1}{2} + \alpha + \frac{1}{2}\alpha)$-th moment is finite, and after n repetitions, we have that the second moment is finite if the moment of order $\alpha \sum_{k=o}^{n} 2^{-k} + 2^{-n}$ is finite. Since $\alpha < \frac{1}{2}$, for some finite n, this quantity is less than 1. Since any solution to Eq. [9] has a finite first moment (by [12c]), this proves that the second, and consequently all, moments are finite. Then, since convergence of all the sums is guaranteed, mass conservation holds by elementary manipulation of Eq. [9]. This proves Assertion 1.

Proof of Assertion 2

To show existence of solutions to Eq. [9], we will assume $b_{ij} \leq bi^{\alpha}j^{\alpha}$ with $\alpha < 1$. We sketch the proof here, leaving technical details to Appendix B. The main idea is to examine truncated sets of Eqs. [9] containing only a finite number of variables and show that a sequence of solutions can be found that converges to a solution of the full problem.

The truncation is chosen to be mass conserving, and the Nth truncation is

$$0 = a_k - x_k + \frac{1}{2}\sum_{j=1}^{k-1} b_{k-j,j}x_{k-j}x_j - x_k \sum_{j=1}^{N-k} b_{kj}x_j \qquad [13]$$

$$k = 1,2,\ldots, N$$

If $\{x_k\}_{k=1}^{N}$ is a solution to Eq. [13], it is easy to show that $\sum_{k=1}^{N} kx_k = \sum_{k=1}^{N} ka_k \leq \sum_{k=1}^{\infty} ka_k$.

Using a degree theoretic argument given in Appendix B it can be shown that for every N there is a solution $\{x_k^N\}_{k=1}^{N}$ of Eq. [13]. If we extend the definition of the sequence $x^N \equiv \{x_k^N\}_{k=1}^{N}$ to $x_k^N = 0$ for $k > N$, then

$$\sum_{k=1}^{\infty} kx_k = \sum_{k=1}^{N} ka_k \leq \sum_{k=1}^{\infty} ka_k \equiv M,$$

so the approximate solutions lie in the set of nonnegative sequences $\{x_i\}_{i=1}^{\infty}$ such that $\sum_{i=1}^{\infty} ix_i \leq M$. This set is compact* in the set ℓ_1^{α} of sequences $\{x_i\}_{i=1}^{\infty}$ such that $\sum_{i=1}^{\infty} i^{\alpha}|x_i| < \infty$, so there is a sequence $\{x_k\}_{k=1}^{\infty}$ such that

$\lim_{j\to\infty} \sum_{k=1}^{\infty} k^{\alpha}|x_k^{N_j} - x_k| = 0_j$ where N_j is a subsequence of integers, and $N_j \to \infty$ as $j \to \infty$. Letting x and x^{N_j} denote $\{x_k\}_{k=1}^{\infty}$ and $\{x_k^{N_j}\}_{k=1}^{\infty}$, respectively, and if T denotes the operator on the right side of Eq. [9] and T^N the corresponding operator on the right of Eq. [13], then we have

$$T^N x^N = 0 \text{ for all } N \qquad [14]$$

since x^N is a solution to Eq. [13]. If we norm sequences $y \equiv \{y_i\}_{i=1}^{\infty}$ by $\|y\|_{\alpha} = \sum_{k=1}^{\infty} k^{\alpha}|y_k|$ and $\|y\| = \sum_{k=1}^{\infty} |y_k|$, then we have for the sequence x defined above,

$$\|Tx\| = \|Tx - Tx^{N_j} + Tx^{N_j} - T^{N_j}x^{N_j}\|$$

$$\leq \|Tx - Tx^{N_j}\| + \|Tx^{N_j} - T^{N_j}x^{N_j}\| . \qquad [15]$$

Since $\|x-x^{N_j}\|_{\alpha} \to 0$, and T is continuous, by an argument in Appendix B, the first term tends to zero. Then, since T^{N_j} uniformly approximates T on compact subsets of ℓ_1^{α}, and the x^{N_j} all lie in such a set, the latter term tends to zero. This proves $\|Tx\| = 0$, and hence $Tx = 0$, or equivalently Eq. [9] is satisfied by x. Therefore x is a solution of Eq. [9].

*A compact set is one such that every sequence of points in the set has a convergent subsequence whose limit is also in the set. See (4) p. 269.

3. DISCUSSION

Combining the results of Assertions 1 and 2 we obtain

Assertion 3

If the coefficients of Eqs. [1] satisfy the relations $c_i \geq ci^{\beta}$, $c > 0$, $\beta \geq 0$, and $b_{ij} \leq bi^{\alpha+\beta}j^{\alpha+\beta}$, where $\alpha < \frac{1}{2}$, then Eq. [1] has a nonnegative solution with finite moments of every order, and which conserves mass.

From this result we see that well-behaved steady state size distributions exist for an aerosol undergoing Brownian coagulation ($b_{ij} \leq bi^{1/3}j^{1/3}$) and gravitational sedimentation $\left(c_k \geq ck^{2/3}\right)$, or Brownian coagulation and convective removal ($c_k = c$).

By examining the proof of White's inequality [12c] we can see that if mass conservation fails, then $\sum_{i,j} ib_{ij}x_1'x_j = \infty$, which implies that the $(\alpha+1)$-th moment of the size distribution is infinite, and consequently, as White noted, the size distribution behaves asymptotically like a power law distribution.

To see this more concretely, set $p = 1$ in the proof of the inequalities [12(a,b,c)]. Then if mass is not conserved,

$$\lim_{N\to\infty} \sum_{k\leq N} \sum_{j>N-k} kb_{kj}x_k x_j = c \ .$$

If $b_{kj} \propto k^{\alpha}j^{\alpha}$, and we suppose $x_j \sim j^{-s}$ for some $s > 0$, we find

$$\lim_{N\to\infty} \sum_{k=1}^{N} \sum_{j=N-k+1}^{\infty} k^{1+\alpha-s}j^{\alpha-s} = c \ .$$

Approximating these sums by integrals gives

$$\int_1^{N-1} x^{1+\alpha-x} \int_{N-x+1}^{\infty} j^{\alpha-s}djdx \approx \int_1^{N-1} x^{1+\alpha-s}(N-x)^{1+\alpha-s}\, dx$$

(assuming $1 + \alpha - s < 0$). Let $f(x) = x^{1+\alpha-s}$. Then this last integral is roughly $f*f(N)$, where * denotes convolution, and $\lim_{N\to\infty} f*f(N) = c$. By the final value theorem, this means $\lim_{z\to 0} \hat{f}^2(z) = c$, where $\hat{f}$ is the Laplace transform of f. Hence $\hat{f}(z) \sim \frac{\sqrt{c}}{\sqrt{z}}$ for small z. Now this means that $f(x) \propto x^{-1/2}$ for large x, so the exponent $s = \frac{3}{2} + \alpha$, so $x_j \sim j^{-\frac{3}{2}-\alpha}$. It should be noted that the exact solutions given by White in the absence of removal for $b_{ij} = i^{\alpha}j^{\alpha}$, have exactly this asymptotic behavior. For this power law distribution to be consistent with the fact that solutions to Eq. [1] have finite first moments, we must require that $\alpha > \frac{1}{2}$. This is precisely the requirement that Assertion 1 not hold, and consequently is very suggestive that the exponent $\frac{1}{2}$ in that assertion is the best possible one.

Ziff (5) gives several solutions of the dynamic coagulation equation which support his belief that for b_{ij} such that $b_{ij} \geq i^{\alpha}$, with $\alpha > 1$, solutions exhibit gelation. This corresponds to our case in which $\alpha > \frac{1}{2}$, and hence tends to support the observations we made above. In aerosol systems it appears from these results that gelation phenomena would be rare, since gravitational sedimentation acts as a strong sink to counteract all of the usual coagulation mechanisms.

4. CONCLUSIONS

We have shown that if the particle removal coefficients satisfy $c_i \geq ci^{\beta}$ for $c > 0$, $\beta \geq 0$, and the coagulation kernel b_{ij} satisfies $b_{ij} \leq bi^{\alpha+\beta}j^{\alpha+\beta}$ with $\alpha < \frac{1}{2}$, then solutions to the steady state population balance exist, have finite moments of all orders, and conserve mass. It follows that such distributions cannot have power law tails. We have given evidence that the exponent $\frac{1}{2}$ is the largest possible, and that for $\alpha > \frac{1}{2}$, solutions have a power law behavior, do not conserve mass, and possibly indicate some behavior similar to gelation in polymerization systems.

REFERENCES

1. White, W. H., J. Colloid Interface Sci., 87, 204 (1982).

2. Stockmayer, W. H., J. Chem. Phys. 11, 45 (1943).

3. Smoluchowski, M. V., Z. Phys. Chem. 92, 129 (1917).

4. Edwards, R. E., Functional Analysis, Holt, Rinehart, Winston, N.Y., 1965.

5. Ziff, R. M., J. Stat. Phys. 23, 241 (1980).

APPENDIX A

(1) If $r \geq 2$, there is a number C_r such that for all $a, b \geq 1$

$$(a+b)^r \leq a^r + b^r + C_r\, a^{r-1}b^{r-1}$$

Proof:

Let n be the smallest integer $\geq r$. Then $n \geq 2$, and

$$(a+b)^n = a^n + b^n + \sum_{k=1}^{n-1} \binom{n}{k} a^k b^{n-k}$$

$$= a^n + b^n + nab^{n-1} + na^{n-1}b + \sum_{k=2}^{n-2} \binom{n}{k} a^k b^{n-k} \qquad [A.1]$$

Dividing by $(a+b)^{n-r}$ gives

$$(a+b)^r \leq a^r + b^r + nab^{r-1} + na^{r-1}b + \sum_{k=2}^{n-2} \binom{n}{k} \frac{a^k b^{n-k}}{(a+b)^{n-r}} \qquad [A.2]$$

Since $r \geq 2$, $nab^{r-1} \leq na^{r-1}b^{r-1}$, and $na^{r-1}b \leq na^{r-1}b^{r-1}$, and $a^k b^{n-k}/(a+b)^{n-r} \leq a^{r-1}b^{r-1}$. Thus $(a+b)^r \leq a^r + b^r + C_r\, a^{r-1}b^{r-1}$, where

$$C_r = \sum_{k=1}^{n-1} \binom{n}{k} = 2^n - 2 \qquad [A.3]$$

(2) If $1 \leq r \leq 2$

$$(a+b)^r \leq a^r + b^r + 2a^{r/2}b^{r/2}$$

for $a, b \geq 1$.

Proof:

$$(a+b)^2 = a^2 + b^2 + 2ab$$

Dividing by $(a+b)^{2-r}$ gives

$$(a+b)^r \leq \frac{a^2}{(a+b)^{2-r}} + \frac{b^2}{(a+b)^{2-r}} + 2\,\frac{a}{(a+b)^{1-\frac{1}{2}r}}\;\frac{b}{(a+b)^{1-\frac{1}{2}r}}$$

$$\leq a^r + b^r + 2a^{\frac{1}{2}r}b^{\frac{1}{2}r} \qquad [A.5]$$

APPENDIX B

Let ℓ_1 = the set of sequences $\{x_i\}_{i=1}^{\infty}$ such that $x \equiv \sum_{i=1}^{\infty} |x_i| < \infty$, and let ℓ_1^{α} equal the set of sequences such that $\|x\|_{\alpha} \equiv \sum_{i=1}^{\infty} i^{\alpha}|x_i| < \infty$. Define an operator T mapping ℓ_1^{α} into ℓ_1 by

$$(Tx)_i = a_i - x_i + \frac{1}{2}\sum_{j=1}^{i-1} b_{i-j,j}x_{i-j}x_j - x_i \sum_{j=1}^{\infty} b_{ij}x_j \qquad [B.1]$$

Here x denotes the sequence $\{x_i\}_{i=1}^{\infty}$. If $b_{ij} \leq bi^{\alpha}j^{\alpha}$, it is clear that the sequence $Tx \equiv \{(Tx)_i\}_{i=1}^{\infty}$ is in ℓ_1 if x is in ℓ_1^{α}. By elementary manipulations of Eq. [B.1], we obtain for x and y in ℓ_1^{α}

$$\|Tx - Ty\| \leq \|x-y\|_{\alpha} + \frac{3}{2} b(\|x\|_{\alpha} + \|y\|_{\alpha})(\|x - y\|_{\alpha}) \qquad [B.2]$$

Eq. [B.2] shows that T is continuous. In fact, it shows T is uniformly continuous on bounded sets in ℓ_1^{α}.

Next define a map T_N on $\mathscr{R}^N$ by

$$(T_N x)_i = a_i - x_i + \frac{1}{2}\sum_{j=1}^{i-1} b_{i-j,j}x_{i-j}x_j - x_i \sum_{j=1}^{N-i} b_{ij}x_j \qquad [B.3]$$

T_N can also be considered a map from ℓ_1^{α} into ℓ_1 by defining $(T_N x)_i = 0$ for $i > N$. Now,

$$\begin{aligned} (Tx - T_N x)_i &= \sum_{j=N-i+1}^{\infty} b_{ij}x_i x_j \qquad \text{for } i \leq N \\ &= (Tx); \qquad\qquad\qquad \text{for } i > N. \end{aligned} \qquad [B.4]$$

Hence we obtain

$$\|Tx - T_N x\| \leq 2b\|x\|_{\alpha} \sum_{j>N/2} i^{\alpha}|x_j| + \sum_{j>N} |(Tx)_i| \qquad [B.5]$$

It can be shown that compact sets in ℓ_1^α are those that are closed, bounded, and have the property that for any positive ε there is N such that every sequence x in the set satisfies $\sum_{i>N} i^\alpha |x_i| < \varepsilon$. From this, and the fact that images of compact sets under continuous functions are compact, it follows that $\|Tx - T_N x\|$ may be made arbitrarily small on compact sets in ℓ_1^α by choosing N large enough .

Incidentally, the above characterization of compact sets shows that the set of nonnegative sequences $x = \{x_i\}_{i=1}^\infty$ such that $\sum_{i=1}^\infty i x_i \leq M$ is compact in ℓ_1^α if $\alpha < 1$. To see this, note that the set is closed and bounded, and since $\alpha < 1$, given ε we can choose N large enough so that $MN^{\alpha-1} < \varepsilon$. Thus

$$\sum_{i>N} i^\alpha |x_i| = \sum_{i>N} i^{\alpha-1}\, i x_i \leq N^{\alpha-1} \sum_{i>N} i x_i \leq MN^{\alpha-1} < \varepsilon \qquad [B.6]$$

for all sequences x in the set.

Finally we must show that each of the truncated sets of equations can be solved to yield a nonnegative solution.

Let $D = \{x \in \mathscr{R}^N : 0 < x_i < M \text{ for } 1 \leq i \leq N\}$ where $M = \sum_{k=1}^\infty k a_k$. Now, viewing T_N as a map from $\mathscr{R}^N$ into $\mathscr{R}^N$ we define a homotopy H_t by

$$H_t(x)_i = a_i - x_i + t\left[\frac{1}{2}\sum_{j=1}^{i-1} b_{i-j,j} x_{i-j} x_j - x_i \sum_{j=1}^\infty b_{ij} x_j\right]$$

$$\text{for } x \in \bar{D} \text{ and } 0 \leq t \leq 1.$$

Note that $H_0 = a - x$ and $H_1 = T_N$. For the moment suppose $a_k > 0$ for all k. We will use the fact that the topological degree at 0 of H_t is independent of t provided no zeroes of H_t cross the boundary of D. That is, if H_0 has a zero in D, and by varying t no zeroes cross the boundary of D, then the parity of the number of zeroes of H_t cannot change, i.e., zeroes can appear and disappear only in pairs. The one-dimensional case illustrates this fact well.

So we need only show that $H_t(x) \neq 0$ if x is a boundary point of D. If x is a boundary point, then either $x_i = 0$ for some i or $x_i = M$ for some i. If $x_i = 0$, then

$$(H_t x)_i = a_i + \frac{T}{2} \sum_{j=1}^{i-1} b_{i-j,j} x_{i-j} x_j > 0, \qquad [B.7]$$

so $H_t x \neq 0$. If $x_i = M$ for some i, and $H_t x = 0$, then by what we just proved $x_j > 0$ for all j. By mass conservation, which is satisfied by any zero of H_t,

$$\sum_{j=1}^{N} j x_j = \sum_{j=1}^{N} j a_j \leq M = x_i < \sum_{j=1}^{N} j x_j \ , \qquad [B.8]$$

a contradiction. Hence, no zeroes of H_t cross the boundary of 0, and since H_0 has 1 zero in D, T_N has an odd number, so has at least one.

If some $a_k = 0$, then the result follows by approximating a_k by a sequence of nonzero numbers tending to zero and by a limiting argument a zero of T_N is obtained again.

CHAPTER 10

MANGANESE SULFATE AEROSOL GROWTH DUE TO CATALYTIC OXIDATION OF SULFUR DIOXIDE IN THE CSTR

MANGANESE SULFATE AEROSOL GROWTH DUE TO CATALYTIC OXIDATION OF SULFUR DIOXIDE IN THE CSTR

INTRODUCTION

In this section we predict the steady state aerosol size and composition in a CSTR in which sulfur dioxide is being absorbed by aqueous manganese sulfate aerosol and oxidized to sulfuric acid. The prediction considers the feed distribution, the sulfur dioxide concentration, the relative humidity, and the mean residence time.

In the last part we present an algorithm which allows computation of the new size distribution resulting from a change in relative humidity. This is important in making comparisons between model predictions and experimental data. Although the analysis is presented for manganese sulfate aerosol, it is applicable in principle to the catalytic oxidation of SO_2 in aerosol particles of other species.

STEADY STATE SIZE DISTRIBUTION IN A CSTR

We begin by discussing the case of a monodisperse feed aerosol to the CSTR. In this case the steady state output size distribution $n(x_o,x)$ obeys the equation

$$0 = QN_o\delta(x-x_o) - Qn(x_o,x) - V\beta\Big(x,\rho(x_o,x)\Big)\, n(x_o,x) - V\frac{d}{dx}\Big(I(x_o,x)n(x_o,x)\Big) \qquad (1)$$

where x is a dimensionless size-related variable, x_o is the size of the feed aerosol, $\rho(x_o,x)$ is the particle density, and $I(x_o,x)$ is the particle growth rate $\frac{dx}{dt}$ and $\beta(x,\rho)$ is the wall loss coefficient. We have included

explicitly the dependence on x_o of the various quantities, since this dependence is important when the feed is polydisperse. Note that coagulation effects are ignored here. In general the growth rate $I(x_o,x)$ can be expected to be composition as well as size dependent. However, in the case of a monodisperse feed of aerosol in which all particles have identical composition, the composition of particles becomes a unique function of size, hence the solution of equation (1) directly yields both size and composition.

The solution (1) is

$$n(x_o,x) = N_o / \left(\tau I(x_o,x)\right)_{(x \geq x_o)} \exp\left[-\int_{x_o}^{x} \frac{1/\tau+\beta(t,\rho(x_o,t))}{I(x_o,t)}\,dt\right] \qquad (2)$$

where τ is the mean residence time of the CSTR.

In the case of a polydisperse feed, the feed distribution can be approximated as a linear combination of monodisperse feeds, and (1) solved for each. Then by linearity of (1), the solution is a linear combination of solutions for each feed size. Hence we need only consider the case of a monodisperse feed, for which (1) and (2) hold.

It follows from Equation (2) that to compute the distribution requires only that we know $I(x_o,x)$ given x_o,x, and a kinetic rate expression for formation of sulfuric acid. The wall loss coefficient β requires, in addition, the particle density. In the next section we describe the procedure for obtaining these quantities.

CALCULATION OF THE GROWTH RATE $I(x_o,x)$

Let us define x as

$$x = \ln(d_p/d_1)/\ln(d_2/d_1) \tag{3}$$

where d_1 and d_2 are lower and upper particle diameter limits, and d_p is particle diameter. We shall take d_1 = 0.1 μm and d_2 = 1.0 μm in subsequent analysis. From (3) and our choice of d_1 and d_2 we obtain

$$I(x_o,x) \equiv \frac{dx}{dt} = \frac{1}{3\ln 10 v_p}\frac{dv_p}{dt} \tag{4}$$

where v_p is particle volume, which is to be calculated from the role of production of sulfuric acid and the requirement of thermodynamic equilibrium. This latter requirement can be shown to be a very good approximation due to the small size of the particles of interest (See Appendix). In general the particles in the CSTR will consist of aqueous solutions of manganese sulfate, sulfuric acid, and sulfur dioxide along with the various ions in equilibrium with it. The concentration of sulfur dioxide and other reduced sulfur ions is governed by equilibrium between the gas and liquid phase, and in the presence of sulfuric acid is so small that its effect on the water activity of the mixture is negligible. Thus, the water activity can be expressed in terms of the concentrations of the manganese sulfate and sulfuric acid alone. Let c_a and c_m denote the acid and manganese salt concentrations, respectively. Equilibrium between the particles and the gas can be expressed in terms of that for water alone. The condition is

$$a_w(c_a,c_m) = RH \tag{5}$$

where a_w is the water activity of the mixture as a function of the concentrations, and RH denotes the ambient relative humidity. Given x_o and x, Equation (3) gives the corresponding volumes $v_{p,o}$ and v_p, and since feed particles contain only manganese sulfate the feed concentration $c_{m,o}$ is obtained by solving (5) with $c_a = 0$. Then mass conservation of the manganese gives

$$c_m = v_{p,o} c_{m,o} / v_p \tag{6}$$

Then solution of (5) yields c_a.

Letting $R(c_a,c_m)$ be the rate of acid formation as a function of the acid and manganese concentrations, we have, by definition of R

$$R(c_a,c_m) = \frac{1}{v_p} \frac{d}{dt} (v_p c_a) = \frac{dc_a}{dt} + \frac{c_a}{v_p} \frac{dv_p}{dt} \tag{7}$$

Differentiation of (5) with respect to t gives

$$\frac{\partial a_w}{\partial c_a} \frac{dc_a}{dt} + \frac{\partial a_w}{\partial c_m} \frac{dc_m}{dt} = 0 \tag{8}$$

But we also have

$$\frac{dc_m}{dt} = - \frac{c_{m,o} v_{p,o}}{v_p^2} \frac{dv_p}{dt} \tag{9}$$

Substituting (9) into (8), and using (7) gives

$$\frac{1}{v_p} \frac{dv_p}{dt} = R(c_a,c_m) \Big/ \left[c_m \left(\frac{\partial a_w}{\partial c_m} \Big/ \frac{\partial a_w}{\partial c_a} \right) + c_a \right] \tag{10}$$

And then $I(x_o,x)$ is obtained from (4).

Hence, the problem is essentially reduced to one of determining the water activity a_w as a function of the concentrations of sulfuric acid and manganese sulfates c_a and c_m. The solution of this problem is discussed in the following section.

EVALUATION OF WATER ACTIVITY OF H_2SO_4-$MnSO_4$ MIXTURES

Since water activity data on mixtures of H_2SO_4 and $MnSO_4$ in water are not available, the mixing rules devised by Kusik and Meissner (1978) have been used here to obtain a_w. Kusik and Meissner present a formula for the activity coefficients of ionic mixtures as functions of the activity coefficients of all pure ionic salts which enter into the mixture. For a solution of $MnSO_4$ and H_2SO_4, these are the only two pure ionic compounds needed. The Kusik and Meissner formula for the water activity of mixtures is, however, incorrect. Thus, the Gibbs-Duhem equation was used to obtain the following formula, consistent with the mixing rules of Kusik and Meissner,

$$\log a_w = \frac{9I_{H^+}I_{SO_4^=}}{2I_T^2}\log(a_w)_a + \frac{4I_{Mn^{++}}I_{SO_4^=}}{I_T^2}\log(a_w)_m \qquad (11)$$

$$+ r$$

where

$$r = 0.0156\left(\frac{9I_{H^+}I_{SO_4^=}}{4I_T} + \frac{I_{SO_4^=}I_{Mn^{++}}}{I_T}\right)$$

$$- 0.0156\,(I_{H^+} + 0.25\,I_{SO_4^=} + 0.25I_{Mn^{++}}) \qquad (12)$$

where I_{H^+}, $I_{Mn^{++}}$, $I_{SO_4^=}$, and I_T are the ionic strengths of H^+, Mn^{++}, $SO_4^=$, and the total ionic strength, respectively, assuming complete dissociation. In Equation (11) $(a_w)_a$ and $(a_w)_m$ are the water activities of pure sulfuric acid solution and manganese sulfate solution, respectively, evaluated at the total ionic strength of the mixture. In the evaluation of (11), data on water activity from Robinson and Stokes (1965) were used. As a result, the mixture

water activity is determined as a function of the acid and manganese molalities. Since the solutions in the aerosol are fairly concentrated, the molar concentrations differ appreciably from the molal concentrations, and the density must be known as a function of concentration to convert from one measure to another.

Figures 1 and 2 show density-concentration data for pure sulfuric acid and pure manganese sulfate solutions*. The partial mass volume is seen to be fairly constant over the entire region covered by the graphs, and the values taken from these were assumed to hold in mixed solutions as well. If $\bar{V}_i$ is the partial mass volume of the ith species (cm^3/g), M_i its molecular weight (g/mole), and c_i its molar concentration (mole/liter), then we obtain the relation

$$1000 = \sum_i \bar{V}_i M_i c_i \tag{13}$$

For the case of sulfuric acid, manganese sulfate, and water, (13) becomes

$$1000 = \bar{V}_a M_a c_a + \bar{V}_m M_m c_m + \bar{V}_w M_w c_w \tag{14}$$

we also have the relation that density in g/cm^3 is related to molar concentrations by

$$\rho = \frac{1}{1000} (M_a c_a + M_m c_m + M_w c_w) \tag{15}$$

Using (14) to eliminate c_w from (15) and the values of $\bar{V}_a$ and $\bar{V}_m$ obtained from Figures 1 and 2, we obtain

$$\rho = 1 + \frac{1}{1000} \left[(1 - \bar{V}_a) M_a c_a + (1 - \bar{V}_m) M_m c_m \right] \tag{16}$$

*CRC Handbook and International Critical Tables.

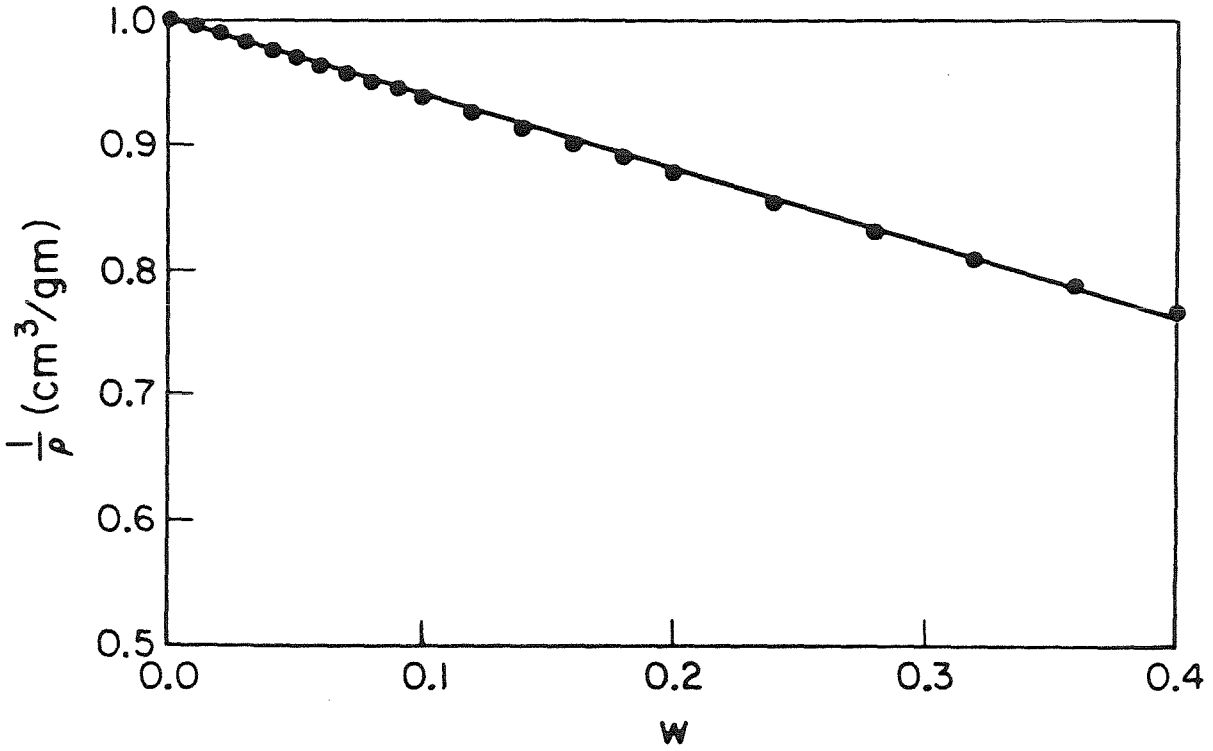


Figure 1. Density of aqueous sulfuric acid as function of weight fraction acid, together with linear approximation.

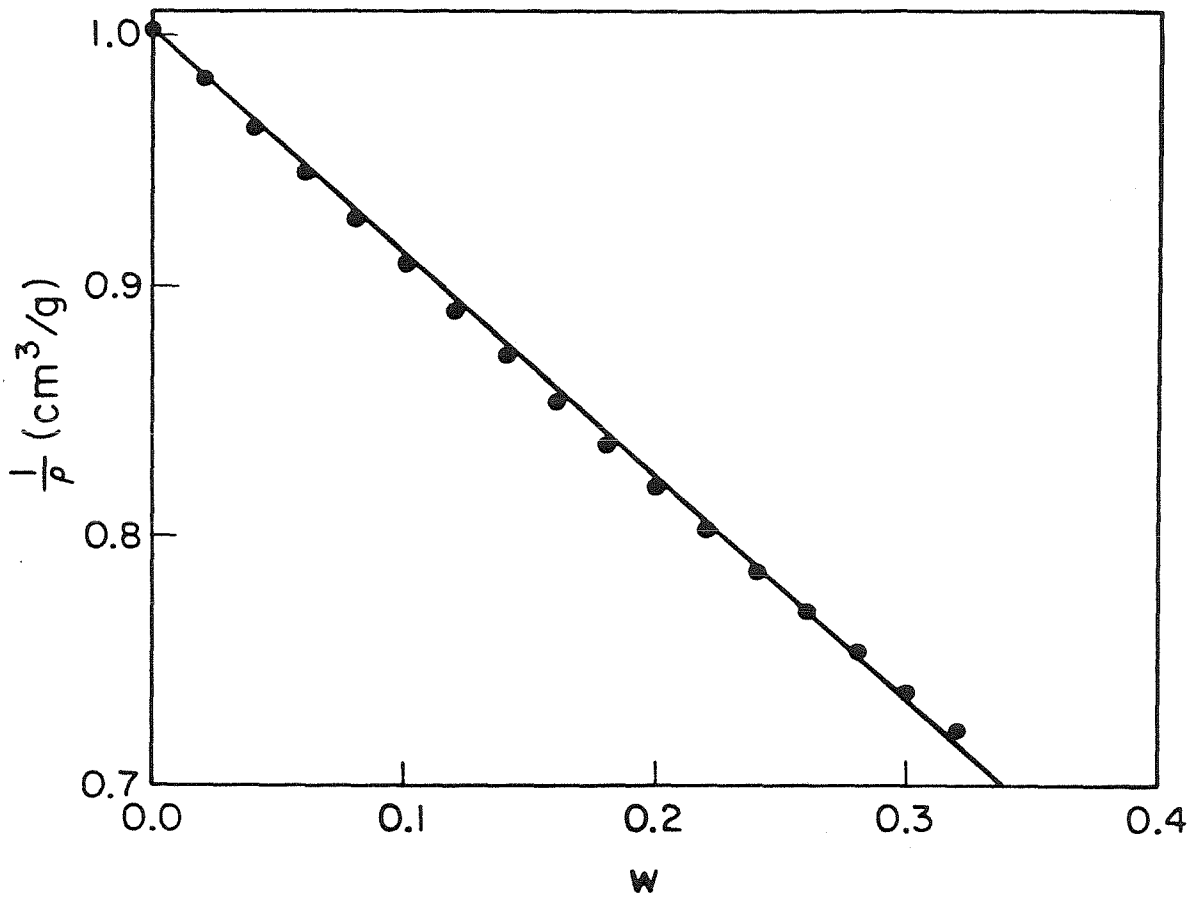


Figure 2. Density of aqueous manganese sulfate as function of weight fraction $MnSO_4$, together with linear approximation.

where $\bar{V}_a$ = 0.401 cm^3/g and $\bar{V}_m$ = 0.106 cm^3/g. Knowing ρ, then concentrations and molalities are related by

$$m_i = c_i/(\rho - M_i c_i/1000) \tag{17}$$

CHANGE OF SIZE DISTRIBUTION CAUSED BY CHANGE IN RELATIVE HUMIDITY

The size distributions in the CSTR are not necessarily the same as those seen by measuring instruments. This is due to the fact that relative humidity is a strong function of temperature, so that instruments operating above or below ambient temperature will measure at different relative humidities.

Thus, to compare predicted size distributions with data one must be able to convert to the conditions under which the measurement was made.

To do this one needs the size and composition distribution. Given a dimensioness size x, and acid and manganese concentrations, c_a and c_m, respectively, the factor λ by which drop volume will change at a new relative humidity is given by

$$RH = a_w(\lambda c_a, \lambda c_m) \tag{18}$$

where RH is the new relative humidity. This equation holds provided no manganese sulfate precipitates. The absence of a solid phase can be checked by comparing the activity product to the solubility product. That is, it must be true that

$$a_w a_{Mn^{++}} a_{SO_4^=} \leq SP \tag{19}$$

The activities are evaluated from Kusik and Meissner's activity coefficient for the mixture

$$\log \gamma^{\pm} = \left(\frac{I_{Mn^{++}} + I_{SO_4^{=}}}{I_T} \right) \log \gamma_{MnSO_4} + \left(\frac{9I_{H^+}}{2I_T} \right) \log \gamma_{H_2SO_4} \qquad (20)$$

where γ_{MnSO_4} and $\gamma_{H_2SO_4}$ are the pure solution manganese sulfate and sulfuric acid activity coefficients, respectively, and are evaluated at the total ionic strength of the mixture.

The water activity is included in (19) based on the assumption that the solid phase in equilibrium would be the monohydrate, which is the case in aqueous solutions of pure manganese sulfate above 27°C.

The solubility product, defined by

$$SP = a_w a_{Mn^{++}} a_{SO_4^{=}} \qquad (21)$$

for saturated solutions of pure manganese sulfate was calculated from solubility data given in the International Critical Tables. Figure 3 shows a plot of lnSP versus 1/T and from it one obtains the approximation

$$\ln SP = \frac{1500}{T} - 8.155 \qquad (22)$$

This equation was used subsequently.

In case (19) fails, the value of λ obtained is incorrect due to precipitation of $MnSO_4$, and the new concentrations in the drop are determined by

$$a_w(c_a', c_m') = RH \qquad (23)$$

$$a_w(c_a', c_m') a_{Mn^{++}}(c_a', c_m') a_{SO_4^{=}}(c_a', c_m') = SP \qquad (24)$$

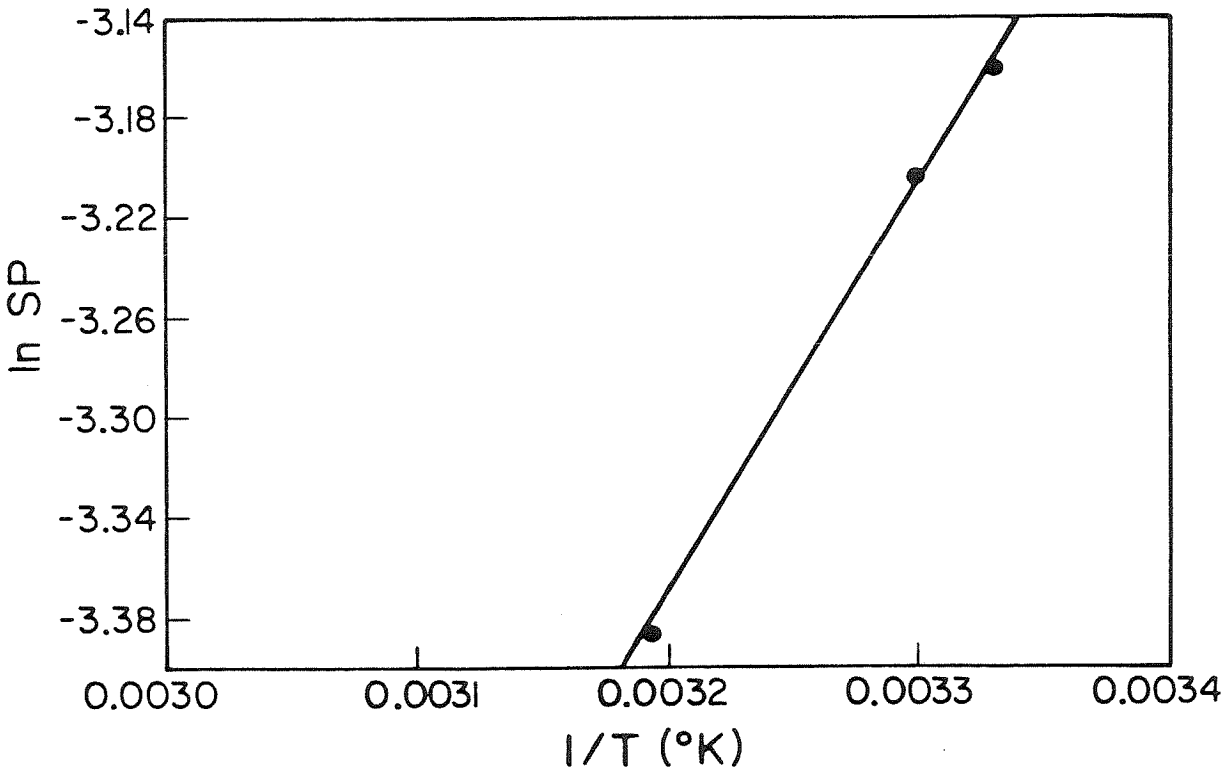


Figure 3. ln SP plotted as function of 1/T. Curve is plot of linear approximation of data.

These equations are solved for c_a' and c_m'. Then mass conservation of manganese and acid gives

$$vc_m = v_\ell c_m' + v_s \rho_m / M_m \tag{25}$$

$$vc_a = v_\ell c_a' \tag{26}$$

In these equations v is the original drop volume, ρ_m the density of the solid phase (which is supposed to be $MnSO_4 \cdot H_2O$), and M_m the molecular weight of the solid phase. v_ℓ is the volume of the liquid phase after the change in humidity, and v_s is the corresponding solid phase volume. Finally the total volume v' of the drop after the change in humidity is the sum of liquid and solid phase volumes

$$v' = v_\ell + v_s \tag{27}$$

Now λ is obtained by the relation

$$\lambda = v/v' \tag{28}$$

Having obtained λ by either solving (18) or from (28), the size distribution n' is obtained from the original distribution by the transformation

$$n'(x_o, x - \tfrac{1}{3} \log \lambda(x)) = n(x_o, x)/(1 - \tfrac{1}{3} \tfrac{d}{dx} \log \lambda(x)) \tag{29}$$

This result holds for a monodisperse feed. If the feed is polydisperse, the separate size distributions corresponding to each size in the feed are transformed according to (29) and the resulting distributions are summed to give the final transformed distribution.

REFERENCES

CRC Handbook of Chemistry and Physics, R. C. Weast, ed. The Chemical Rubber Co., Cleveland (1971).

Friedlander, S. K. Smoke, Dust and Haze, John Wiley and Sons, New York (1977).

International Critical Tables, E. W. Washburn, ed., McGraw-Hill, New York, (1928).

Kusik, C. L. and Meissner, H. P., AIChE Symposium Series 74, 14 (1978).

Robinson, R. A. and Stokes, R. H., Electrolyte Solutions, Butterworths, London (1965).

APPENDIX

EQUILIBRATION TIME FOR AEROSOL PARTICLES

We make an estimate for the time to reach thermodynamic equilibrium of aerosol particles of diameter less than 1 μm.

We assume the particles are large enough for continuum analysis to hold. Then the flux of water vapor to the particle surface is (Friedlander, 1977)

$$2\pi d_p D(p-p_d)/RT \text{ mole/sec} \qquad \text{(A1)}$$

where

D = Diffusivity of water vapor in air = $0.24(cm^2/sec)$

d_p = drop diamter (cm)

p = water partial pressure (atm.)

p_d = drop water vapor pressure (atm.)

T = absolute temperature (°K)

R = gas constant = 82.05 atm. cm^3/mol °K

The above flux must equal the quantity

$$\frac{d}{dt}\ \frac{\rho v_p}{M} \qquad \text{(A2)}$$

where v_p = particle volume (cm^3)

ρ = density of water (1.0 g/cm^3)

M = molecular weight of water = 18 g/mol

where we have assumed no volume change on mixing. This assumption is adequate for the purpose of estimation. Setting $v_p = \pi d_p^3/6$, and equating the flux expression (A1) to the growth expression (A2) yields the equation

$$\frac{d}{dt} \ln d_p = 4MD(\rho - p_d)/\rho RT\ d_p^2 \qquad (A3)$$

Now, $p - p_d = p_o^w(RH - a_w)$ where

p_o^w = vapor pressure of water (atm.)(=0.0313 atm. at 25°C)

RH = relative humidity

a_w = water activity of drop

Then

$$\frac{d}{dt} \ln d_p = 4MDp_o^w(RH - a_w)/\rho RT\ d_p^2 \qquad (A4)$$

We now consider the longest possible time for equilibrium to be established within one percent for a drop of diameter less than 1 μm. Before equilibrium is reached within one percent, we have

$$|RH - a_w| > 0.01 \qquad (A5)$$

So, from (A5) and substitution of appropriate values into Equation (A4), we obtain

$$\left| \frac{d}{dt} \ln d_p \right| > 22 \qquad (A6)$$

Typically drop diameters will not change by more than a factor of two when relative humidity changes, so using (A6), this gives, upon integration

$$\ln 2 > 22t \qquad (A7)$$

or

$$t < 0.03 \text{ sec} \qquad (A8)$$

CHAPTER 11

OXIDATION OF SULFUR DIOXIDE IN AQUEOUS MANGANESE SULFATE AEROSOLS

OXIDATION OF SULFUR DIOXIDE IN AQUEOUS MANGANESE SULFATE AEROSOLS

INTRODUCTION

The phenomenon of aerosol growth due to oxidation of sulfur dioxide has received considerable attention in recent years (Bassett et al., 1981; Freiberg, 1978; Cains and Carabine, 1978; Wadden et al., 1974), particularly in reference to plumes from coal-fired boilers. A central question in such studies concerns the role of heterogeneous liquid phase oxidation of sulfur dioxide, catalyzed by dissolved ash from the coal combustion aerosol, and its consequent effect on evolution of the aerosol size distribution.

To begin to understand these processes, simpler systems, containing a single metal salt, have been employed, notably manganese salts, which are known to be active catalysts for SO_2 oxidation (Bassett and Parker, 1951) and to occur in fly ash from coal combustion facilities (Haury, et al. 1978).

Much of the research on oxidation of SO_2 in Mn-catalyzed systems has been done in bulk solutions with the object of determining the kinetics of the reaction. In such experiments manganese concentrations are typically less than 10^{-3} M. Some work has also been done on aerosol systems, but in such systems one must distinguish between two very different situations. One is that of fog or rain droplets, which typically contain only traces of heavy metal contaminants (Hoffmann and Jacob, 1982), and the other is that of deliquesced metal salt particles in humid air, which frequently contain metal salts in concentrations of 1 M or higher. This latter situation appears more representative of plume aerosol, and in one study (Haury, et al., 1978) manganese sulfate aerosols were even shown to give

oxidation rates similar to those observed on fly ash aerosols. Perhaps the most important reason for studying oxidation of SO_2 in aerosol systems to assess our understanding of the interaction of physical and chemical processes fundamental to the design of plume models by comparing results of theoretical calculations with experimental observations made under controlled conditions.

A promising technique for studying aerosol growth is the continuous stirred tank reactor (CSTR). In a CSTR long residence times necessary to observe significant growth of aerosol are more easily achieved than in tubular reactors due to the excessive lengths required of the latter. In addition the CSTR operates at steady state, making possible the use of measurement techniques requiring long times, such as, perhaps, inertial impactors. This type of measurement would be necessary for composition distribution determination. In addition, a batch reactor for aerosol studies must be quite large in order to use conventional measurement techniques for particle size analysis, and the large size makes careful temperature control, and hence good control of relative humidity, difficult. Perhaps the most serious drawback of the CSTR is the distribution of residence times which results in polydisperse effluents even when the feed is monodisperse. However, the simplicity of modeling the CSTR compensates for this disadvantage to some extent since there is, in principle, no real difficulty in accounting for this polydispersity.

In the next section we discuss some of the more salient studies of the kinetics of oxidation of SO_2 in aqueous $MnSO_4$ solutions, and, based on the available data suggest a rate expression appropriate for aerosol systems. In the following section we discuss experiments using a CSTR

to measure particle growth rates and deduce a kinetic coefficient for the SO_2 oxidation based on the data.

PREVIOUS WORK ON OXIDATION OF SO_2 IN AQUEOUS $MnSO_4$ SOLUTIONS

The kinetics of oxidation of sulfur dioxide in aqueous solutions of manganese sulfate have been studied by numerous techniques, both in bulk liquid and in aerosol systems. Table I summarizes some of the work done to date.

Generally speaking catalyst concentrations in the previous studies are typically less than 10^{-3} M in bulk solution studies, and range upwards of 1 M in studies of aerosols, in which concentration is determined by thermodynamic equilibrium of water. This large difference no doubt accounts for considerable variance in the reported rates of reaction.

The most accurate kinetic data appear to have been obtained from bulk solutions, where variables can be accurately measured and controlled. Aerosol studies, on the other hand, have tended to be less fruitful in terms of yielding rate expressions. Several of the studies cited in Table I merit further comment.

The most complete and reliable kinetic data appear to be those of Martin (1982) and Coughanowr and Krause (1965), both of which indicate a second order dependence of the rate of SO_2 consumption on the manganese concentration and a zero order dependence on SO_2 concentration. Moreover, if the pH of the experiments of Coughanowr and Krause is taken as three, a reasonable estimate given the reported liquid phase SO_2 concentration of 1.9×10^{-3} M, their results agree with the rate expression of Martin. Coughanowr and Krause also report a weak dependence of the SO_2 consumption

Table I. Summary of Previous Studies of SO_2 Oxidation in Aqueous $MnSO_4$

Investigator	Type of Experiment	Results
Bassett and Parker (1951)	Bulk liquid phase oxidation with 70 percent SO_2 in O_2 using various metal salt catalysts.	Mn most active catalyst. Rate dependent on $MnSO_4$ and SO_2 concentrations. Poisoning effect due to natural rubber noted.
Johnstone and Coughanowr (1958)	Aqueous drops, 700-900 μm diameter containing 250-1000 ppm $MnSO_4$ in atmosphere of 30-200 ppm SO_2.	Rate = k_o, where $k_o = 1.07 \times 10^{-4}$ M sec^{-1} for $[MnSO_4]$ = 500 ppm $k_o = 2.87 \times 10^{-4}$ M sec^{-1} for $[MnSO_4]$ = 1000 ppm, $k_o = 4.33 \times 10^{-5}$ M sec^{-1} for $[MnSO_4]$ = 250 ppm
Coughanowr and Krause (1965)	Bulk phase study using dissolved SO_2 at $\sim 1.9 \times 10^{-3}$ M and $[MnSO_4]$ from 5 to 10,000 ppm.	Rate = $k_o[MnSO_4]^2$ for $[MnSO_4]$ < 100 ppm, where $k_o = 2 \times 10^{-8}$ M ppm^{-2} sec^{-1}. k_o increases slowly for $[MnSO_4]$ > 500 ppm. Inhibition of reaction by rubber noted.
Matteson, Stöber, Luther (1969)	Aerosol study using polydisperse $MnSO_4$ droplets 2-5 m diameter. Rh 95%, SO_2 concentration 2-140 ppm	$\text{Rate} = \frac{0.0037[MnSO_4][SO_2]}{40 + [SO_2]} - \frac{48[MnSO_4][H_2SO_4]^2}{40 + [SO_2]}$ $[MnSO_4]$, $[H_2SO_4]$ in M: $[SO_2]$ in ppm. No reaction for RH < 95%.
Wadden, Quon, Hulburt (1974)	Theoretical model of growing, coagulating aerosol using mechanism of Matteson, et al., and data of Coughanowr (1956)	$\text{Rate} = \frac{0.034[MnSO_4][SO_2]}{200 + [SO_2]} - \frac{0.071[MnSO_4][H_2SO_4]^2}{200 + [SO_2]}$ $[MnSO_4]$, $[H_2SO_4]$ in M; $[SO_2]$ in ppm.
Cheng, Corn, Frohliger (1971)	Submicron $MnSO_4$ particles affixed to teflon beads in flow reactor absorbed gaseous SO_2 fed to reactor. SO_2 loss measured.	Rate = $9.09 \times 10^{-11}[SO_2]$ moles (sec mg $MnSO_4$) in drops at RH = 95%. $[SO_2]$ in ppm.

Table I. Summary of Previous Studies of SO_2 Oxidation in Aqueous $MnSO_4$ (Cont)

Investigator	Type of Experiment	Results
Cains and Carabine (1978)	Aerosol phase in tube flow reactor. 63 ppm SO_2 and 95% RH used. Particle conc. $\cong 10^5$ cm^{-3}	$\text{Rate} = \frac{0.0338[SO_2][MnSO_4]}{[SO_2] + 187} - \frac{0.066[MnSO_4][H_2SO_4]^2}{[SO_2] + 187}$ M sec^{-1}; $[SO_2]$ in ppm; $[MnSO_4]$,$[H_2SO_4]$ in M
Himmelblau and Kaplan (1981)	Aerosol of 0.1 M $MnSO_4$ in $(NH_4)_2SO_4$ and SO_2 0.1 to 1.0 ppm in laminar flow reactor. Particle sizes of 1 to 10 μm diameter.	Rate = kc^{IV} assumed, where c^{IV} = concentration S(IV) species in drop, in M. k varied from 0.027 sec^{-1} to 0.066 sec^{-1}.
Martin (1982)	Bulk solution studies for 10^{-5} M < $[Mn^{++}]$ < 10^{-3} M. 10^{-3} M < [S(IV)] < 10^{-1} M	Rate = $4.7[H^+]^{-1}[Mn^{++}]^2$ M sec^{-1} for $0 \leq pH \leq 3$, 10^{-5} M < $[Mn^{++}]$ < 10^{-3} M, and 10^{-4} < [S(IV)] < 10^{-3} M. Rate = $25[H^+]^{-1}[Mn^{++}][S(IV)]$ for [S(IV)] < 10^{-6} M

rate on manganese concentration when the latter exceeds 3×10^{-3} M up to the maximum concentration used in their experiments of 7×10^{-2} M.

In an aerosol, the concentrations of manganese sulfate are much higher than this value. Based on osmotic coefficient data from Robinson and Stokes (1965) and solubility data from the International Critical Tables (1928), the $MnSO_4$ concentration of a drop suspended in an atmosphere at 98 percent relative humidity is 1.13 M. There does not appear to be any quantitative data on the rate of SO_2 oxidation in aqueous solutions at such high concentrations of $MnSO_4$, but the data of Coughanowr and Krause suggest a zero order rate dependence on $MnSO_4$ concentration at high concentrations.

Kaplan and Himmelblau (1981) present results from a flow reactor study in which aerosols, largely of deliquesced ammonium sulfate doped with smaller amounts of manganese sulfate were passed through a laminar flow reactor with a humidified atmosphere containing SO_2. The results are difficult to interpret in terms of the rate of SO_2 oxidation both because of scatter in the data and the complicating presence of ammonium sulfate.

One of the more widely cited contributions is that of Matteson, et al. (1969), which has become the basis of more recent work (Wadden et al. (1974), Cains and Carabine (1978)). However, the rate expression of Matteson et al. does not predict the second-order dependence of the SO_2 oxidation rate on $[Mn^{2+}]$ at low manganese concentration, contrary to claims of its authors. (See Appendix.)

In addition, based on data obtained in this work, the Matteson et al. expression substantailly underpredicts particle growth rates.

The work of Wadden et al. (1974) is included in Table I despite the fact that it does not represent primary experimental data, because it gives

a rate expression for oxidation of sulfur dioxide in manganese sulfate solutions which was specifically chosen for aerosol systems of the type considered in this work. This rate expression is based on the Matteson et al. mechanism, but uses data from Coughanowr (1956). Wadden et al. do not state how the rate coefficients were chosen, but the resulting expression is nearly identical to that of Cains and Carabine (1978), which, as is stated below, is incorrect on theoretical grounds. As a result, considerable doubt is thrown on the correctness of the rate expression of Wadden, et al. Moreover, comparison of predictions based on this rate law with data from the present work shows that the rate of Wadden, et al. predicts much larger growth rates than are observed.

Cains and Carabine (1978) performed experiments in which a polydisperse submicron aerosol was passed through a tubular reactor together with SO_2 and humidfied air, for residence times of up to twenty minutes. Particle sizes were then measured by a light scattering technique. Due to high particle concentrations ($\sim 10^5$ cm^{-3}) coagulation also occurred, precluding any more than a qualitative comparison of the data with calculations based on a theoretical rate expression. Again, the rate expression used was based on the mechanism of Matteson et al., but the rate constants were evaluated from other data. In fact, the data of Johnstone and Coughanowr (1958), in which the rate of oxidation of SO_2 showed a second order dependence on manganese concentration, was used to evaluate one constant, two were taken from Wadden et al., and the last was from Matteson et al. The use of the results of Johnstone and Coughanowr depends on the assertion of Matteson, et al. that the second order rate dependence is predicted by their mechanism. Since this assertion is incorrect, it is

unlikely that the expression of Cains and Carabine is correct. The expression, as was stated previously, is similar to that of Wadden, et al. and hence can be expected to predict much more rapid growth than was observed in the present work. Nevertheless, Cains and Carabine assert that calculations based on their rate expression are roughly in agreement with their experimental observations of the growth rate. This seems unlikely since the experimental growth rates obtained by them appear comparable to those obtained here. The calculations carried out by Cains and Carabine may be in error, as they state that the relative humidity of deliquescence of $MnSO_4$ is 94.3 percent, whereas, the correct value, according to osmotic coefficient (Robinson and Stokes, 1965) and solubility data (International Critical Tables, 1928) is near 84 percent.

Based on the foregoing discussion several conclusions on the rate of oxidation of SO_2 in aqueous solutions of $MnSO_4$ may tentatively be drawn:

(1) For $[Mn^{++}] \leq 10^{-3}$ M, 10^{-4} M $\leq [S(IV)] \leq 10^{-3}$ M, the rate dependence on $[Mn^{++}]$ is second order.

(2) The presence of acid inhibits the reaction (Coughanowr and Krause (1965), Matteson, et al. (1969), Martin (1982)). The dependence on $[H^+]$ may be inverse as Martin has shown.

(3) For 10^{-4} M $\leq [S(IV)] \leq 10^{-3}$ M, the rate is zero order in [S(IV)]. This corresponds to about 80-800 ppm SO_2 in the gas phase for an acid solution, or 5-50 ppm over a solution of pH 3.

(4) For $[S(IV)] \leq 10^{-6}$ M, the rate is first order in both [S(IV)] and $[Mn^{++}]$. This corresponds to less than 1 ppm SO_2 in the gas phase over an acid solution and 0.05 ppm over a solution of pH 3.

(5) For $[Mn^{++}] > \sim 3x10^{-3}$ M, the rate is probably zero order in $[Mn^{++}]$, based on extrapolation of data of Coughanowr and Krause (1965).

From these conclusions a reasonable rate expression for an aerosol system containing in excess of 10 ppm SO_2 in the gas phase and $[Mn^{2+}] > \sim 3x10^{-3}$ M might be

$$R = K/[H^+]$$

although at the lower SO_2 concentrations there may be a weak dependence on $[SO_2]$ as well. Using conclusion (5), we can estimate by means of Martin's kinetic expression that

$$K \sim 4.7x(3x10^{-3})^2 = 4.2x10^{-5}\ M^2\ sec^{-1}$$

This value can only be considered an order of magnitude estimate, however, due to the uncertainties involved in pinpointing a concentration of $MnSO_4$ at which the reaction becomes zero order in $[MnSO_4]$.

AEROSOL GROWTH EXPERIMENTS

Experimental Procedure

Before beginning an experimental run 1000 ppm SO_2 in air from a cylinder was bubbled through the water to be used in the humidifier until its SO_2 concentration reached a level in equilibrium with the desired gas phase concentration. The SO_2 concentration in solution was determined by oxidizing an aliquot with hydrogen peroxide and titrating the sulfuric acid with 0.01 N sodium hydroxide.

During the experiments at 97 percent relative humidity it was discovered that the SO_2 concentration to the feed was diminished by absorption in the humidifier, due apparently to reaction with the copper heat exchange coils. Nevertheless, the SO_2 concentration remained constant through most of the experiment. To control SO_2 concentration in subsequent runs, the aerosol was humidified before being mixed with the SO_2. This procedure resulted in lower humidities, but better control over the SO_2 concentration.

Before introducing aerosol into the system the sulfur dioxide-air mixture at the proper relative humidity was passed through the reactor for approximately three hours. After this time the aerosol flow was initiated by starting the atomizer syringe pump and the classifier was started, and after allowing time for the flows to stabilize, the reactor valve was opened and aerosol began flowing through the reactor. The system was allowed approximately five hours to reach steady state. In all experiments the mean reactor residence time was about an hour.

During this time measurements of the sulfur dioxide concentration were made periodically by drawing reactor effluent through an absorber containing a dilute solution (approximately 10 percent) of hydrogen peroxide at

a known flow rate for 20 minutes. The absorbing solution was then titrated for sulfuric acid using 0.01 N sodium hydroxide and neutralized bromthymol blue as an indicator, and the SO_2 concentration was then determined from the known gas volume and sulfur content, making corrections for the relative humidity of the gas.

Dewpoint measurements of the reactor effluent were made periodically by opening a valve to the dewpoint meter. No filter was used in the line to the meter since it was found that filters interfered with the dewpoint measurement by absorbing water. Because of the absence of the filter, the mirrored dew collecting surface of the meter was cleaned periodically by wiping with acetone. These cleanings did not appear to affect the results.

Two hours before measurements were to be made, the optical particle counter (OPC) was turned on to allow the sample chamber to reach a steady temperature. This temperature was somewhat higher than the reactor temperature, and the resulting change in relative humidity had a strong influence on measured particle sizes. Consequently, it was important for this temperature to remain constant during measurements.

Approximately one-half hour before measurements were made the OPC was connected to the reactor effluent stream to allow equilibration of the instrument sheath air. During this time the measured particle sizes appeared to increase as the relative humidity of the sheath air increased. After 15 to 20 minutes, the size distribution appeared constant.

Several measurements were then made of the reactor effluent aerosol, and simultaneous measurements of dewpoint and SO_2 concentration were carried out. Then the reactor bypass valve was opened and size measurements of the feed aerosol were made together with a dewpoint measurement.

It was not possible to measure directly the relative humidity within the sample chamber of the particle counter, but a temperature measurement was made by attaching a thermocouple junction to a brass fitting carrying the sheath air to the sample chamber. These measurements gave a temperature of 26.7°C, or roughly three degrees higher than the reactor temperature. Assuming no moisture is lost within the instrument, due to hygroscopicity of filters, for example, this temperature change corresponds to a drop in relative humidity of approximately 15 percent. Thus, the importance of correcting for changes in relative humidity occurring in the optical particle counter is evident.

The conditions for each of the experiments are tabulated in Table II, and actual averaged data are shown in Table III.

Data Inversion

The size distribution data for the experiments listed in Table II were determined using the constrained inversion algorithm of Crump and Seinfeld (1982). In carrying out this inversion it was found that the data from the first channel did not agree with the previous calibration data obtained using polystyrene latex spheres. The latter data indicated a peak response of the first channel to particles near 0.2 μm diameter, and although the experimental feed and effluent aerosol peaked near this size, only negligible response in the first channel was seen. This may have been due to a difference in refractive index of the manganese sulfate particles as compared to polystyrene latex, or it may have been due a sensitivity of the first channel response to changes in laser power output of the instrument, which had been noted during the course of calibration. To avoid interference by first channel data, a large standard deviation was used in inverting these data.

Table II. Summary of Experimental Conditions and Best Fit Rate Constants for Aerosol Growth Experiments

Experiment	SO_2 conc. (ppm)	RH(%)	Residence Time (min)	Rate constant k (M^2 sec^{-1})	Weighted RMS dev. of OPC and Simulated Data
1	50	97	61	7.6×10^{-5}	0.228
2	22	97	59	5.1×10^{-5}	0.194
3	14	95	60	8.4×10^{-5}	0.248
4	38	95	58	1.21×10^{-4}	0.222
5	0	94	58	--	--
6	92	91	58	--	--

Table III. Experimental Data

Experiment 1		CSTR(cm^{-3})	σ	Feed (cm^{-3})	σ
Res. time 61 min.	Ch. 1	9.86	0.64	20.3	0.7
CSTR temp. 24.2°C	2	17.5	5.3	24.9	2.9
CSTR dewpt. 23.7°C	3	61.6	5.3	24.3	2.4
Feed dewpt. 23.8°C	4	5.3	0.34	4.3	0.16
SO_2 conc. 50 ppm	5	2.9	0.1	3.1	0.2
	6	0.75	0.06	0.65	0.08
Experiment 2					
Res. time 59 min.	Ch. 1	9.05	0.62	14.9	1.0
CSTR temp. 24.0°C	2	177	3.4	230	4.9
CSTR dewpt. 23.5°C	3	76.9	3.1	62.5	4.3
Feed dewpt. 23.5°C	4	6.0	0.24	6.3	0.16
SO_2 conc. 22 ppm	5	3.1	0.1	3.9	0.21
	6	0.66	0.05	0.68	0.08
Experiment 3					
Res. time 60 min.	Ch. 1	10.2	0.44	14.5	1.1
CSTR temp. 24.2°C	2	172	4.6	216	5.0
CSTR dewpt. 23.4°C	3	65.1	3.9	61.0	6.0
Feed dewpt. 23.5°C	4	4.9	0.41	6.0	0.29
SO_2 conc. 14 ppm	5	2.9	0.11	3.3	0.22
	6	0.65	0.13	0.55	0.06
Experiment 4					
Res. time 58 min.	Ch. 1	4.38	0.26	7.9	0.27
CSTR temp. 24.1°C	2	125	2.51	178	2.25
CSTR dewpt. 23.3°C	3	98.2	2.27	89.9	1.74
Feed dewpt. 23.3°C	4	7.7	0.33	7.5	0.13
SO_2 conc. 38 ppm	5	3.4	0.06	4.1	0.11
	6	0.76	0.04	0.91	0.06
Experiment 5					
Res. time 58 min.	Ch. 1	7.02	0.28	10.1	0.26
CSTR temp. 23.9°C	2	171	2.3	205	1.6
CSTR dewpt. 22.8°C	3	93.9	2.0	94.7	0.9
Feed dewpt.	4	7.2	0.23	7.4	0.17
SO_2 conc. 0 ppm	5	3.8	0.1	4.2	0.13
	6	0.91	0.05	0.93	0.06

Table III. Experimental Data (Continued

Experiment 6		CSTR(cm^{-3})	σ	Feed (cm^{-3})	σ
Res. time 58 min.	Ch. 1	5.48	0.71	7.9	0.43
CSTR temp. 24.0°C	2	136	7.5	167	4.0
CSTR dewpt. 22.4°C	3	65.8	6.8	60.9	3.6
Feed dewpt. 22.6°C	4	5.0	0.5	4.42	0.16
SO_2 conc. 92 ppm	5	2.0	0.22	2.32	0.15
	6	0.24	0.04	0.26	0.04

Discussion of Results

The inverted feed and CSTR effluent number distributions corresponding to experiments 1-6 of Table II are shown in Figures 1-6, respectively.

The distributions of Figures 1-6 were measured at a relative humidity which could not be directly determined, as has been previously stated. However, temperature measurements on the sheath air inlet to the sample cell indicate, and the assumption that filters in the instrument do not affect the humidity lead to relative humidity estimates within the sample cell of between 82 percent and 84 percent. Osmotic coefficient data for manganese sulfate (Robinson and Stokes, 1965) along with solubility data (International Critical Tables, 1928) indicate that the relative humidity of deliquescence of manganese sulfate is between 83 percent and 84 percent at room temperature. The electrical mobility classifier was operated so as to produce dry particles of approximately 0.2 μm diameter. Since the measured feed and effluent distributions peak near this size, it was at first assumed that the particles in the OPC sample chamber were dry. To test this assumption a calculation was carried out.

Based on the assumption that relative humidity is 83 percent, corresponding to dry feed particles, data from experiment one were used to attempt to determine a constant k in the rate expression for SO_2 oxidation,

$$R = k/[H^+]. \qquad (1)$$

To do this, the feed distribution, assumed to be composed of dry particles, was converted to 97 percent RH, and this feed distribution was used in the steady state general dynamic equation for the aerosol size distribution in the CSTR to predict effluent size distributions. The latter were then

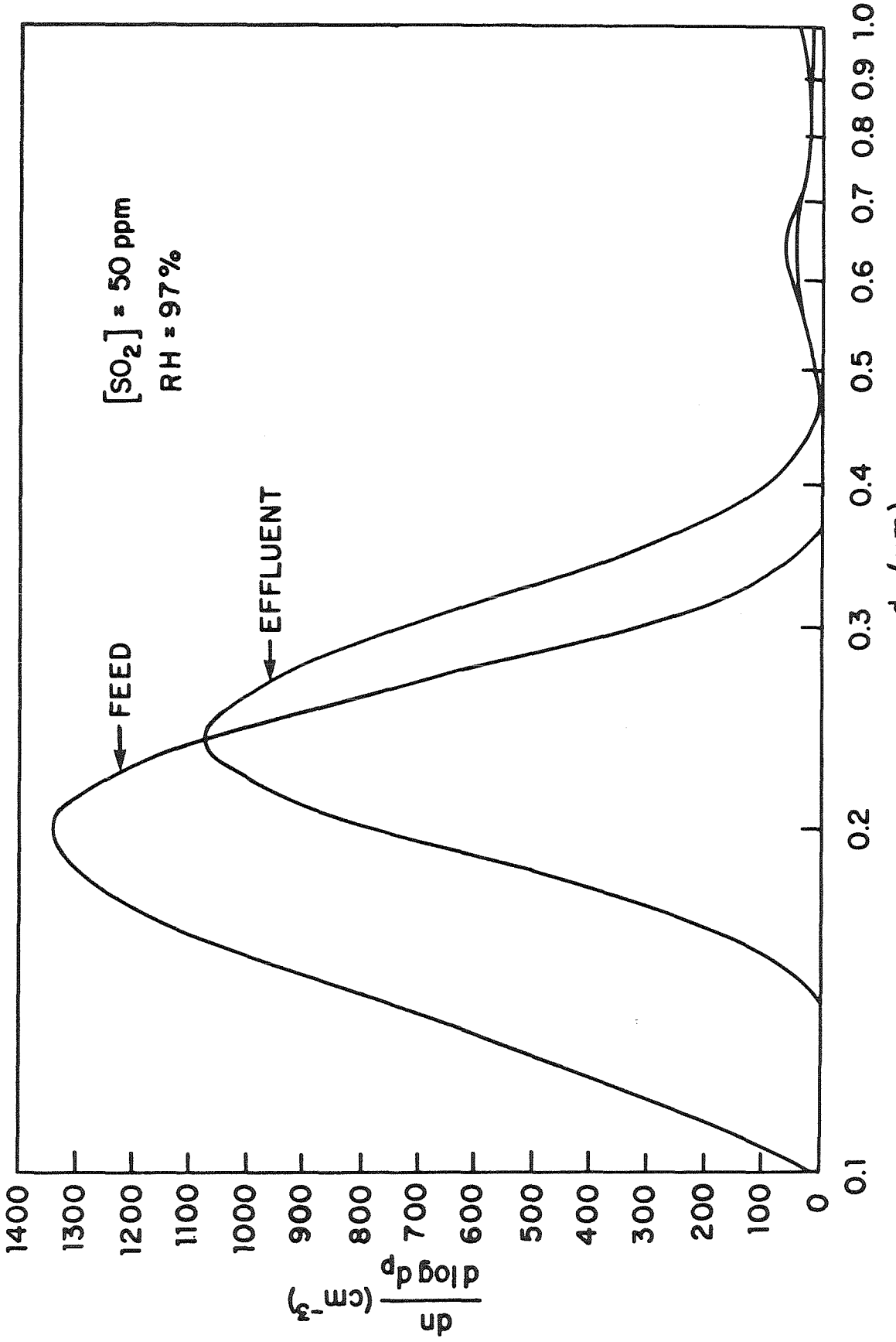


Figure 1. Measured feed and effluent number distributions for experiment 1.

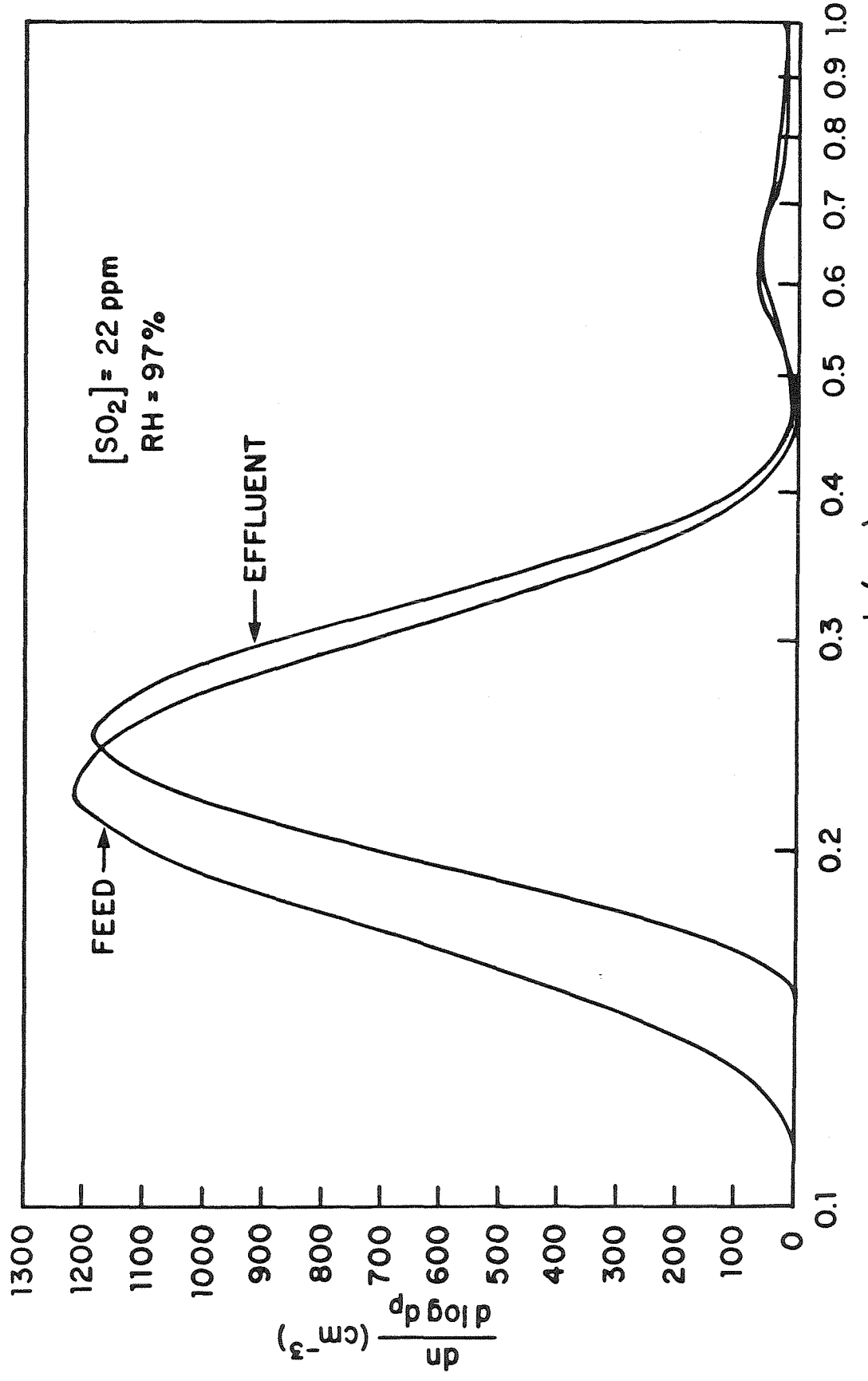


Figure 2. Measured feed and effluent number distributions for experiment 2.

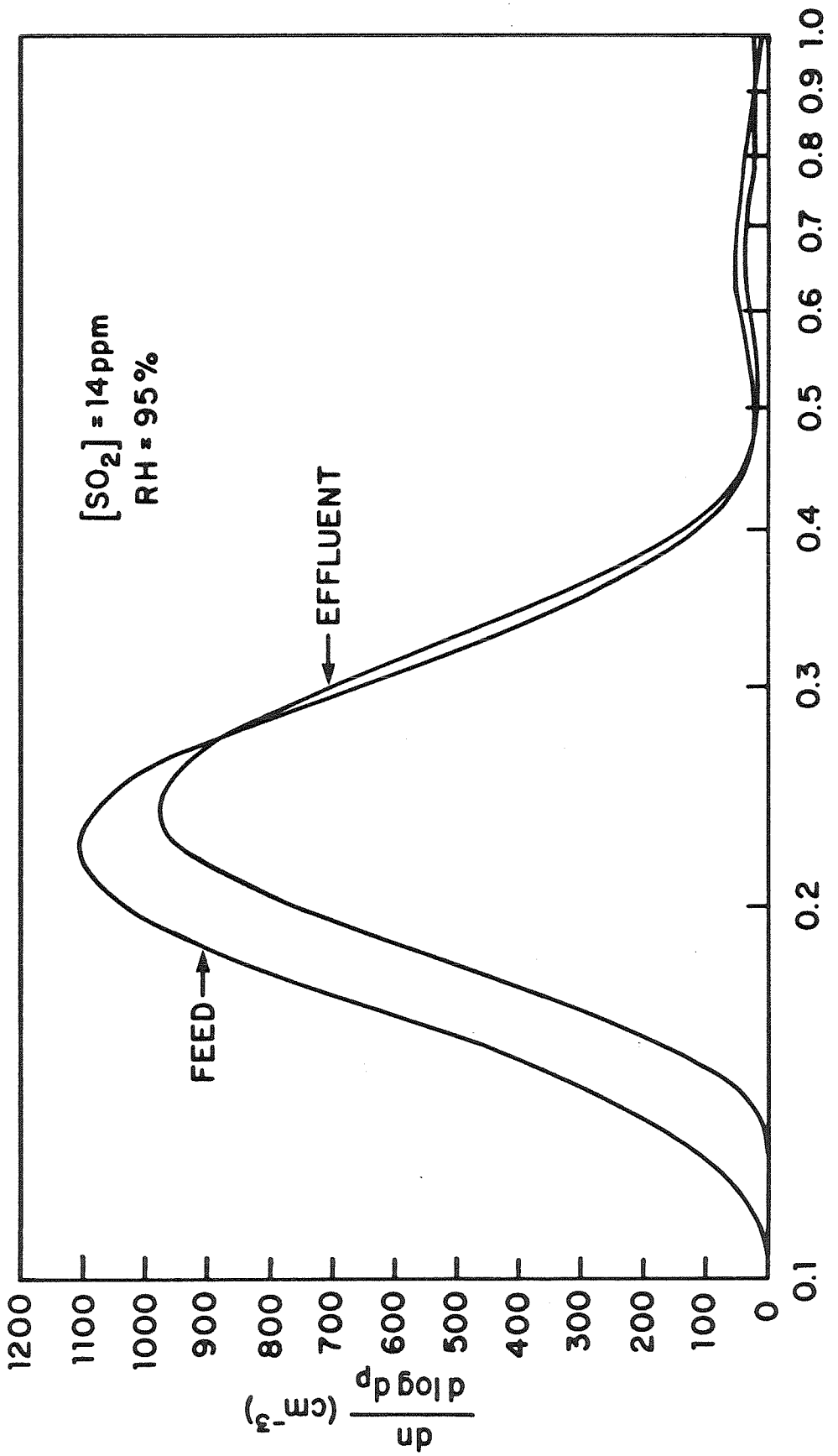


Figure 3. Measured feed and effluent number distributions for experiment 3.

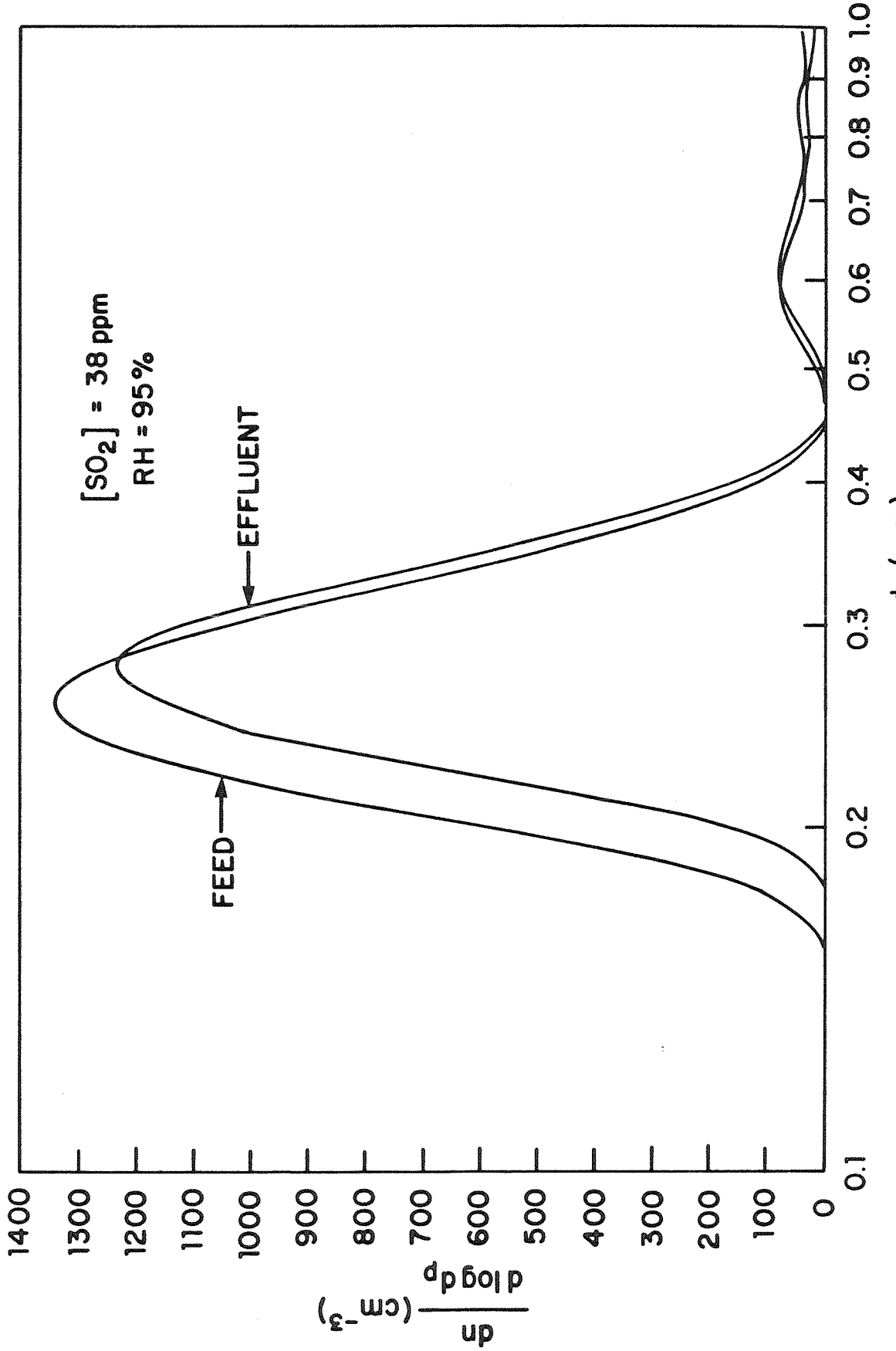


Figure 4. Measured feed and effluent number distributions for experiment 4.

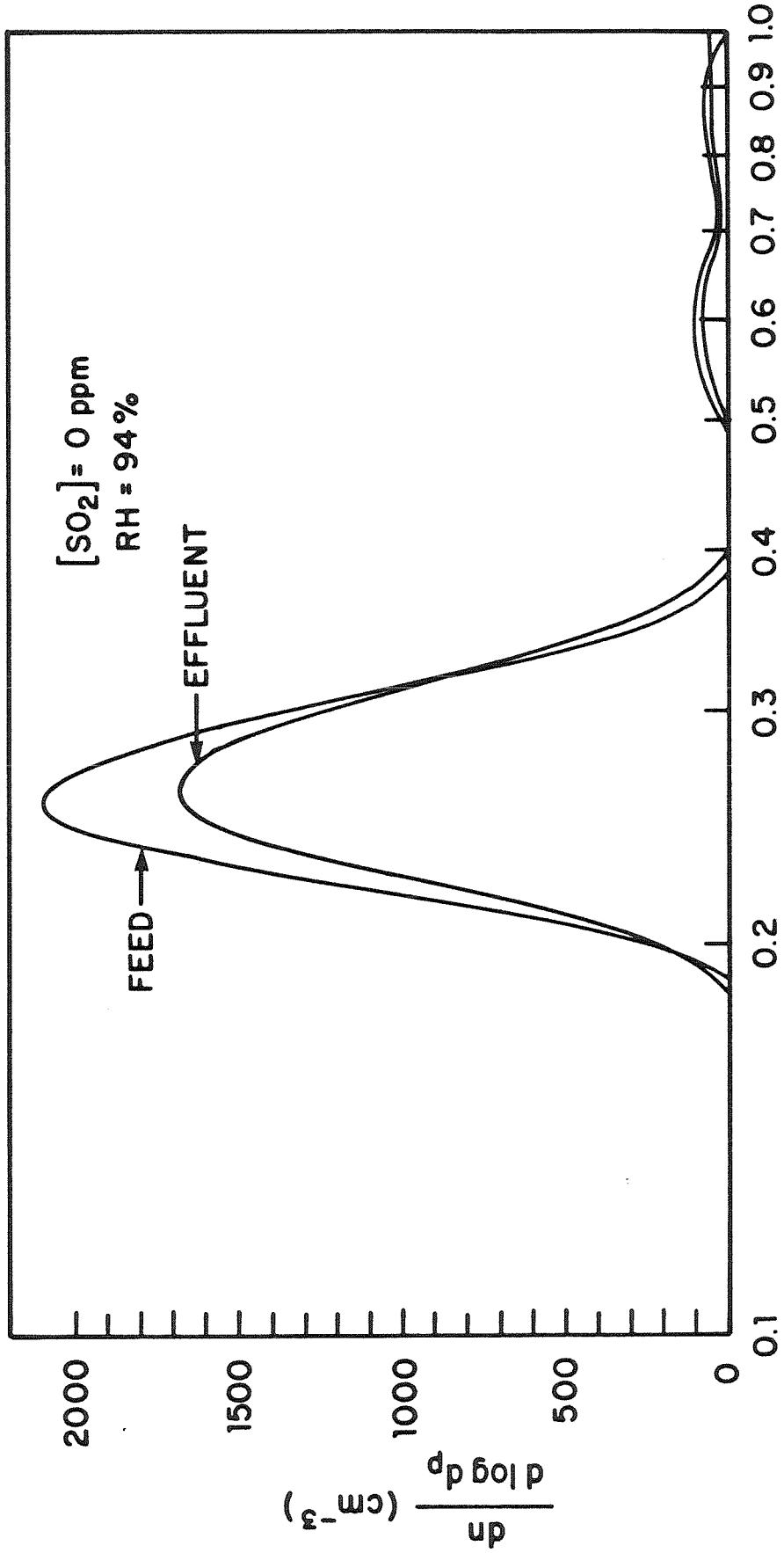


Figure 5. Measured feed and effluent number distributions for experiment 5

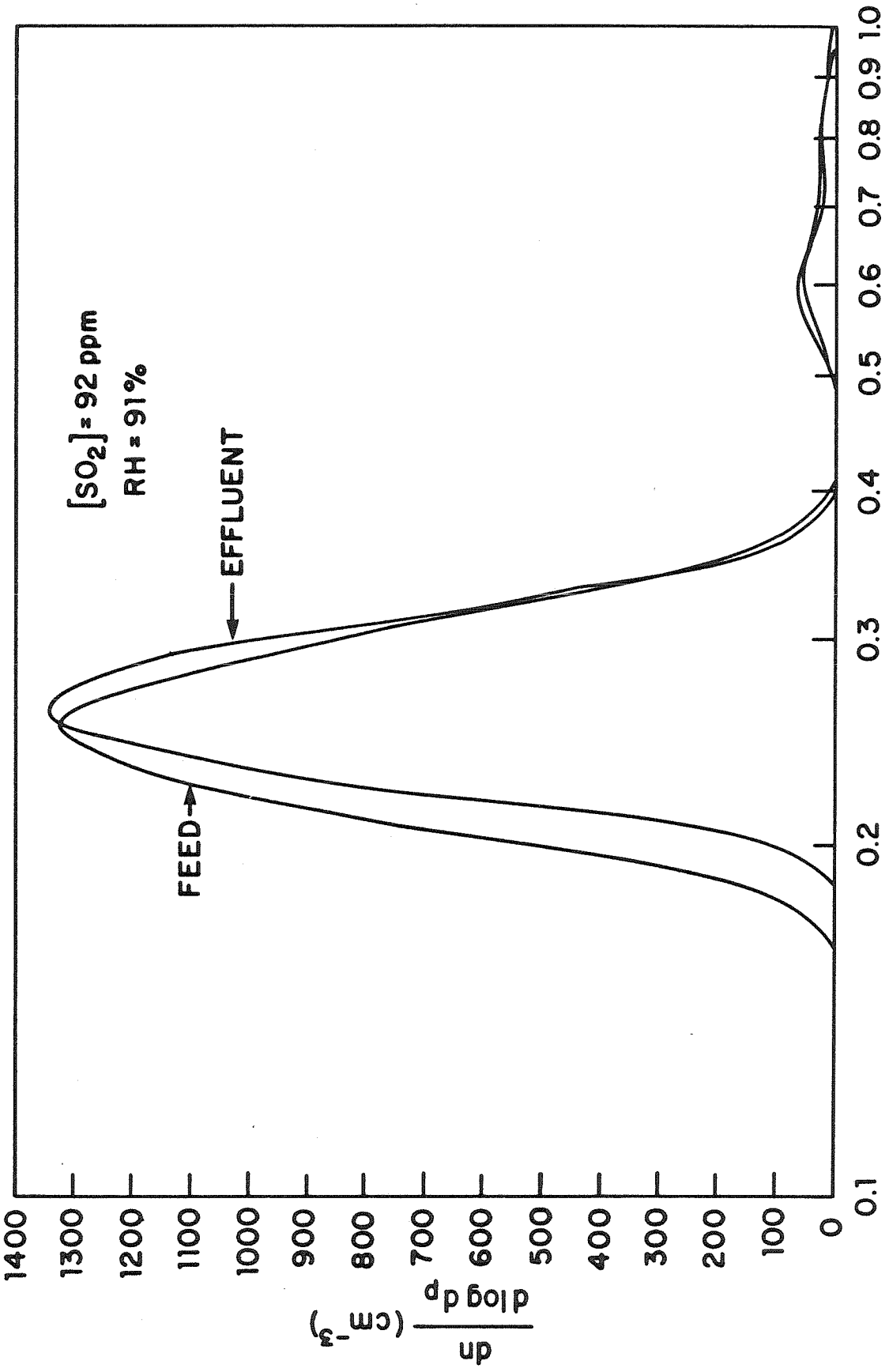


Figure 6. Measured feed and effluent number distributions for experiment 6.

reduced to 83 percent RH using the second algorithm discussed in the previous chapter. The resulting distribution for $k = 10^{-7}$ M^2 sec^{-1} is given in Figure 7 together with the measured inverted effluent distribution. The rate used in this simulation corresponds to a manganese sulfate concentration of approximately 1.5×10^{-4} M according to the result of Martin, and this value is well within the range of experimental values used by Martin. However, the $MnSO_4$ concentration of experiment one was close to 1.5 M, or some 10^4 times larger than concentration which gives the same rate in Martin's kinetic expression. This observation throws doubt on the validity of the rate used in this simulation, especially when one considers that when the relative humidity is assumed to be 84 percent, a value of k of close to 8×10^{-5} M^2 sec^{-1} is obtained. This rate corresponds to a concentration of 4.2×10^{-3} M in Martin's expression, which is well above the range of concentrations he studied. This latter result suggests that the particles observed are actually wet. In order to investigate this further, measurements of dry aerosol were made under the same conditions as those used in experiments 1-6. The Royco Optical Particle Counter and an Environment One Condensation Nuclei Counter were used to analyze the particles.

The optical particle counter indicated a smaller size than when the aerosol was humidified, most of the particles falling in channel 1 of the instrument. The total number concentration indicated by the optical particle counter was approximately 100 cm^{-3}, and the condensation nuclei counter gave a reading between 80 and 100 cm^{-3}.

This result tends to confirm the assumption that the humidified particles are wet during measurement in the OPC. The explanation of the

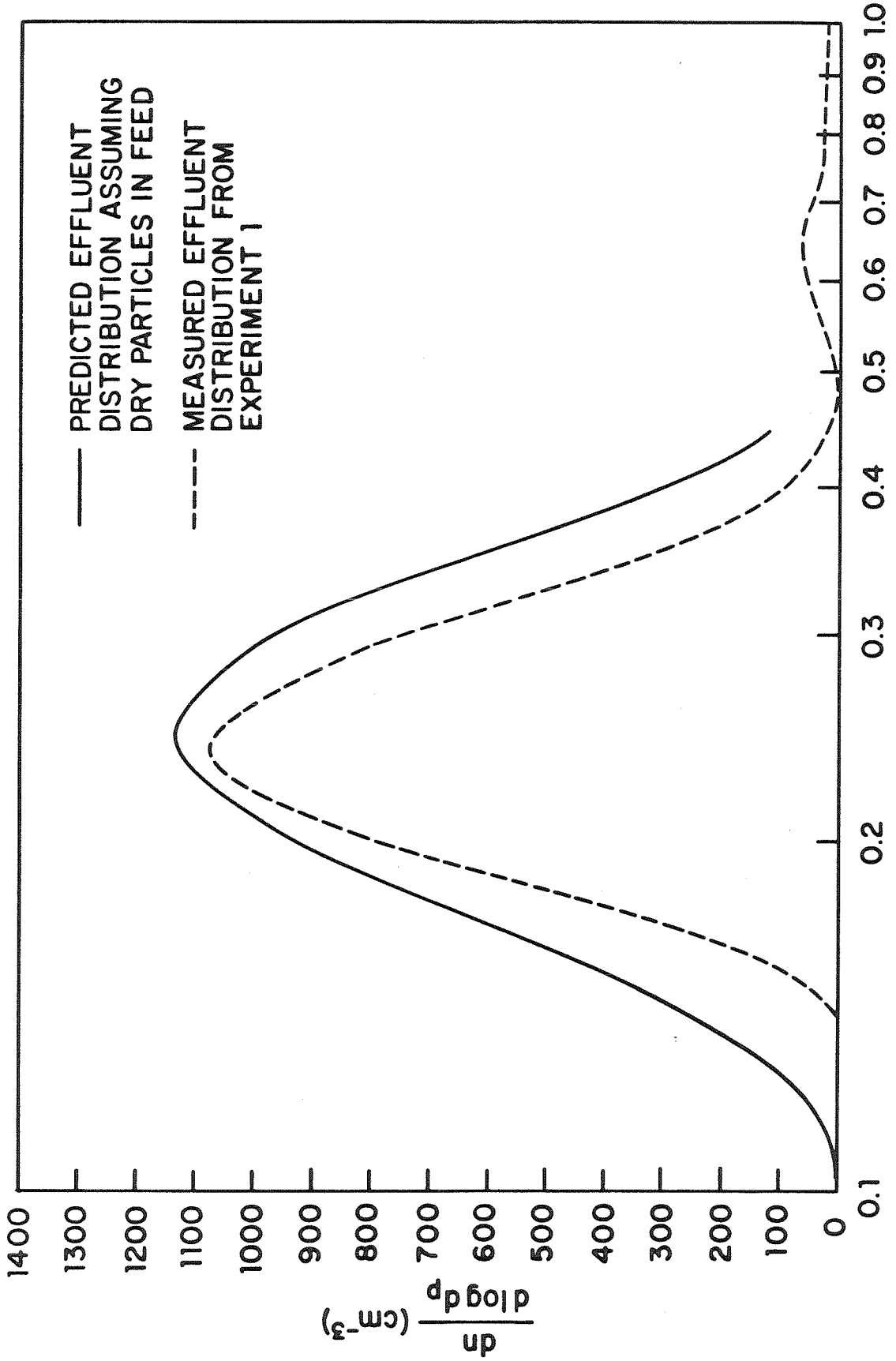


Figure 7. Measured number distribution from experiment 1 and predicted number distribution at 83 percent relative humidity based on feed distribution of experiment 1 using rate expression $R = 10^{-7}/[H^+]$.

extremely small rate constant obtained under the assumption of a dry aerosol can be seen from Fig. 8, which shows a plot of the logarithm of the factor by which particles shrink when the relative humidity is changed from 97 percent to either 83 percent or 84 percent as a function of the logarithm of the ratio of particle diameter to feed particle diameter to the CSTR. The two curves are extremely similar except at the smallest sizes, corresponding to the smallest acid concentrations. In this region the 83 percent curve rises rapidly with decreasing particle size because even minute amounts of acid are enough to prevent the particles from completely drying out. This phenomenon is further illustrated in Figure 9, which shows a lognormal distribution at both 83 percent and 84 percent relative humidity together with the output of the CSTR corresponding to the same feed distributions at 97 percent RH, one hour residence time, 50 ppm SO_2, and a rate of

$$R = 10^{-4}/[H^+] \text{ M sec}^{-1} \qquad (2)$$

Note the similarity of the two output distributions at 83 percent and 84 percent RH contrasted with the large difference between the input distributions at the two relative humidities.

From the foregoing discussion it can be concluded that the aerosol observed in the OPC was wet. Its relative humidity was probably in the range 82-84 percent. There is a possibility that supersaturation occurred, but it could not have occurred to a great degree given the above range of relative humidities possible and it appears that, provided the aerosol is not dry, small identical changes in the relative humidity of both feed and output should have a negligible effect on growth rate estimates.

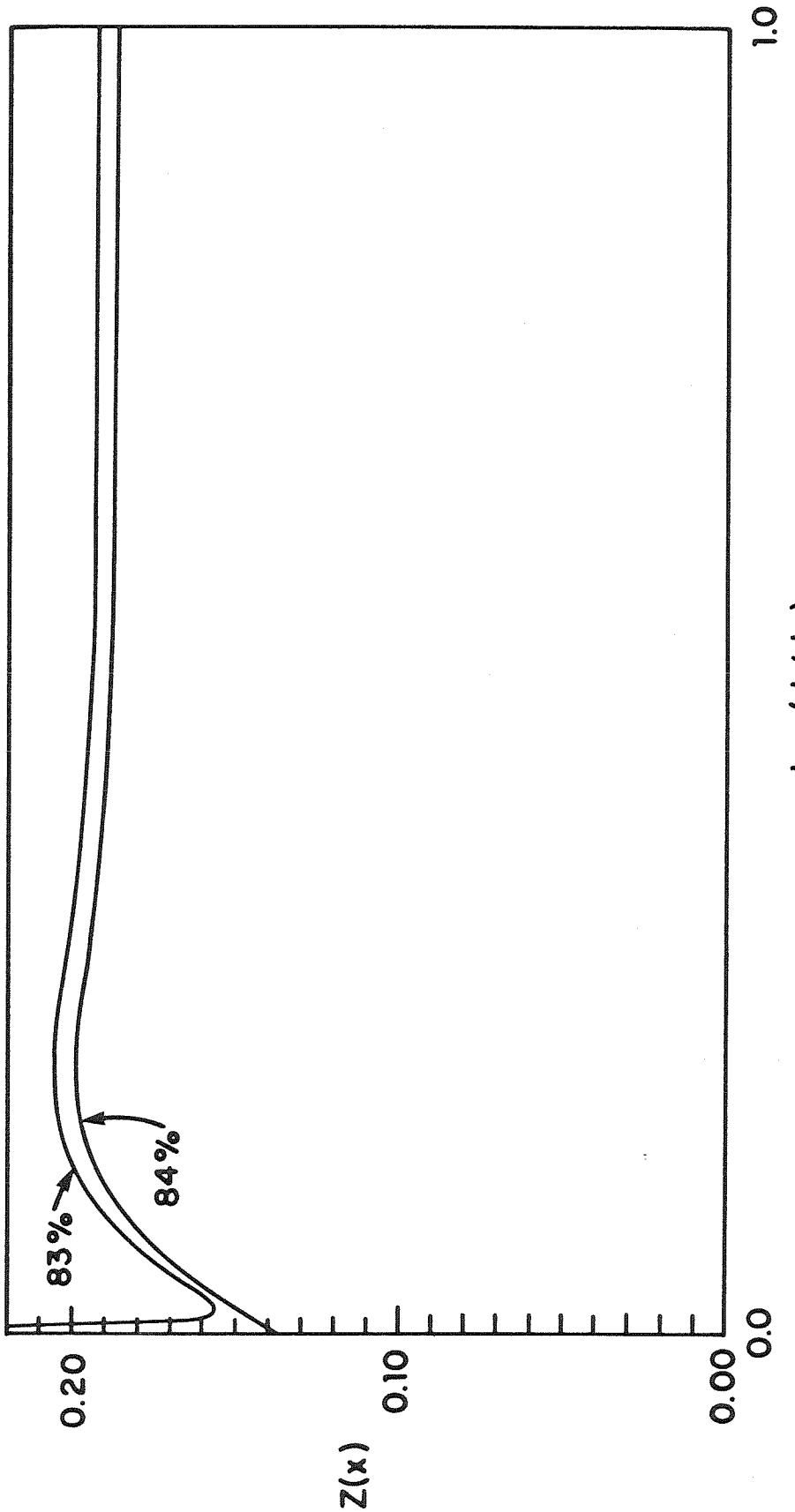


Figure 8. Plots of $\log(d_{97}/d_{RH})$ versus $\log(d_{97}/d_f)$, for RH = 83 percent and RH = 84 percent relative humidity, where d_{97} is particle diameter at 97 percent relative humidity, d_{RH} is particle diameter at relative humidity RH, and d_v is the feed particle diameter at 97 percent relative humidity.

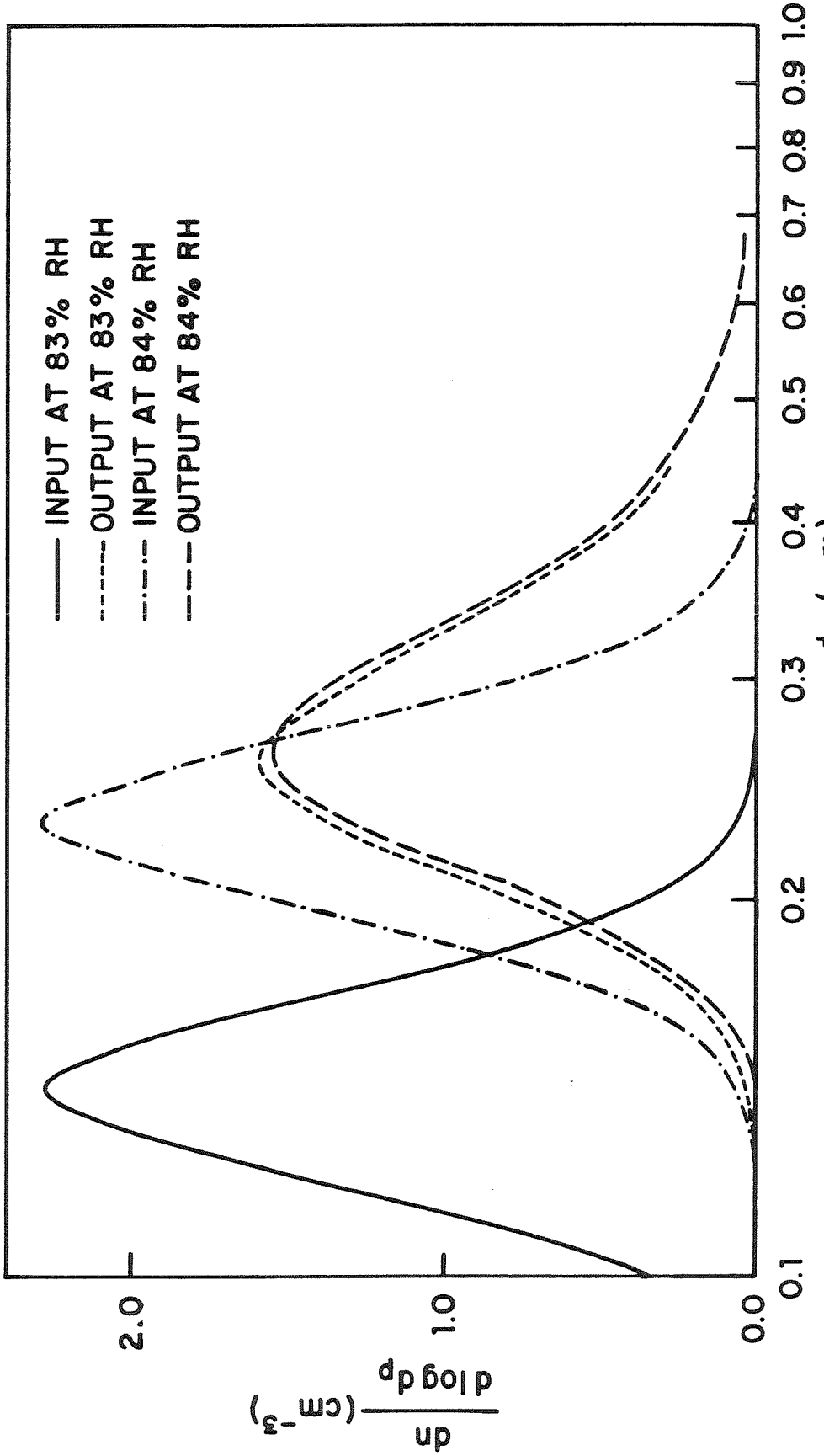


Figure 9. Lognormal feed distribution of pure $MnSO_4$ at 83 percent and 84 percent relative humidity and corresponding CSTR effluent distributions at the same relative humidities. CSTR conditions assumed were 50 ppm SO_2, 97 percent RH, 60 minutes residence time, and reaction rate $R = 10^{2.4}/[H^+]$

Consequently, in estimating reaction rates from the data, it was assumed that the aerosol in the OPC sample chamber had a relative humidity of 84 percent. The procedure used was then to convert the input distributions from 84 percent to the appropriate relative humidity for the experiment being modelled, and calculated output distributions for conditions of the experiment using the rate expression (1) for several values of the parameter k. These distributions were then reduced to 84 percent RH, and simulated values of the counts of channels 2, 3, and 4 of the OPC were computed using the kernel functions obtained by calibration. Finally, the value of k was chosen by a weighted least squares fit of the simulated OPC data to the measured data, where the weights were chosen to be proportional to the values of the measured data. This fitting was done for experiments 1-4. Experiment 5 exhibited negligible growth, since no SO_2 was present, and the growth in experiment 6 was very small due to the low relative humidity, so the results of these two experiments were not used in the evaluation of the rate constant. The results of the rate coefficient determination are shown in Table II. The mean value of k is 8.3×10^{-5} M^2 sec^{-1} with a standard deviation of 2.5×10^{-5} M^2 sec^{-1}. There does not appear to be any dependence of k on the SO_2 concentration, although scatter in the data, precludes making any stronger assertions. Figures 10-14 show simulated distributions corresponding to feed data of experiments 1-4, and 6, respectively using the mean value of k given above. Note that the location at the peak in the distribution is fairly accurately given by these distributions. In contrast, Figures 15 and 16 show distributions resulting from other kinetic expressions. Figure 15 shows the result of using the

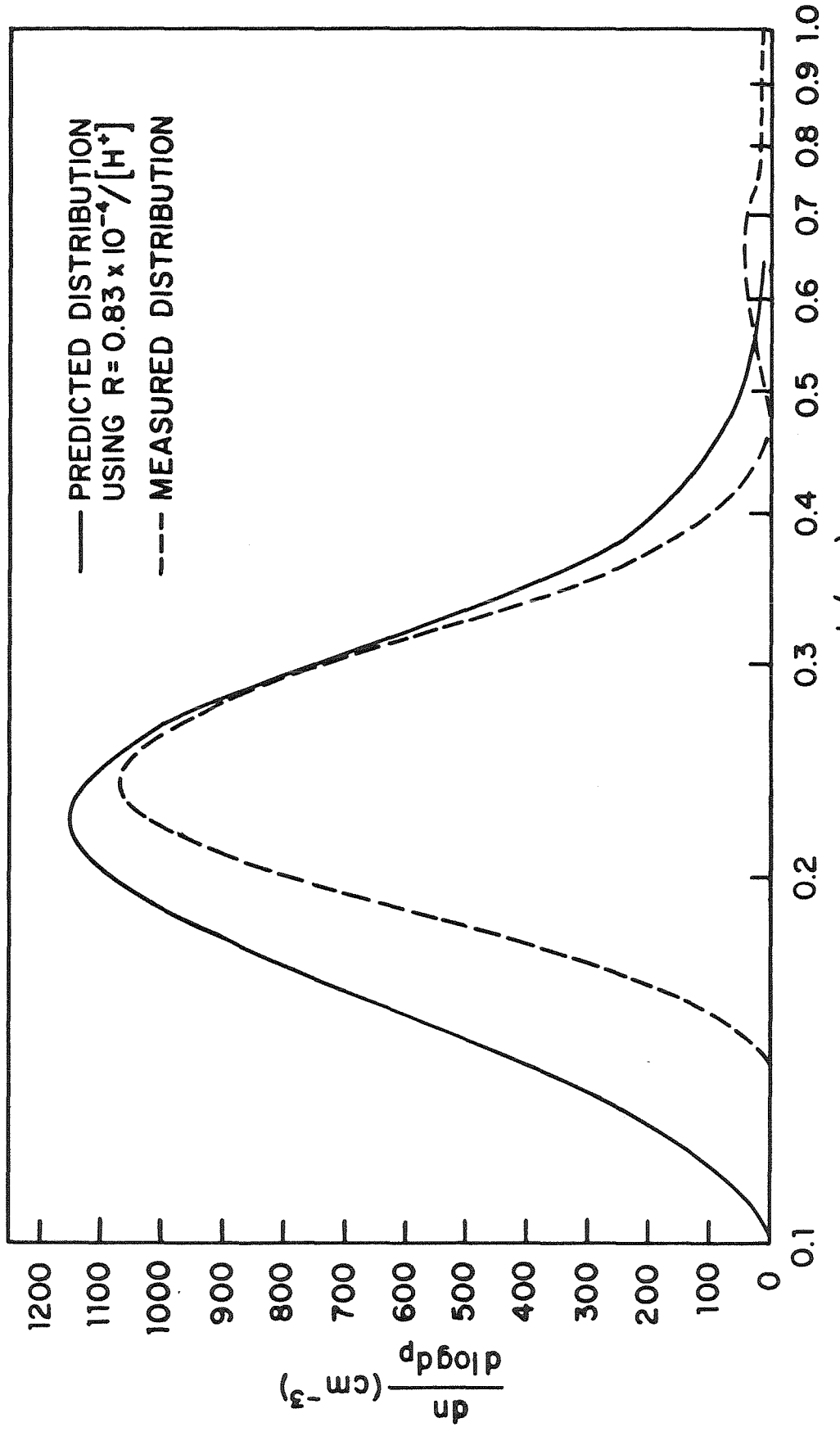


Figure 10. Predicted and measured effluent number distributions from experiment 1 based on rate $R = 0.83 \times 10^{-4}/[H^+]$.

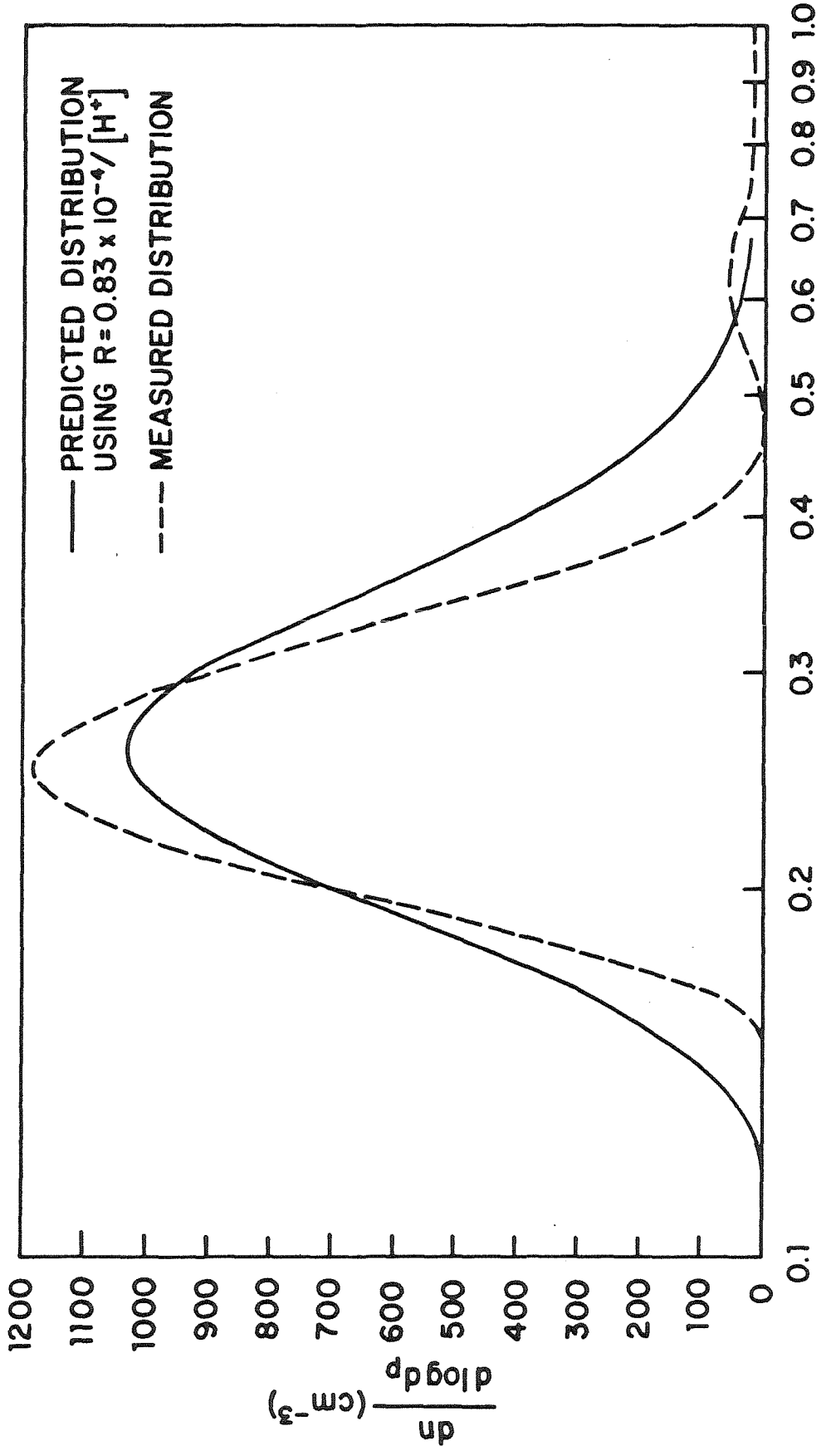


Figure 11. Predicted and measured effluent number distributions from experiment 2 based on rate R = $0.83 \times 10^{-4}/[H^+]$.

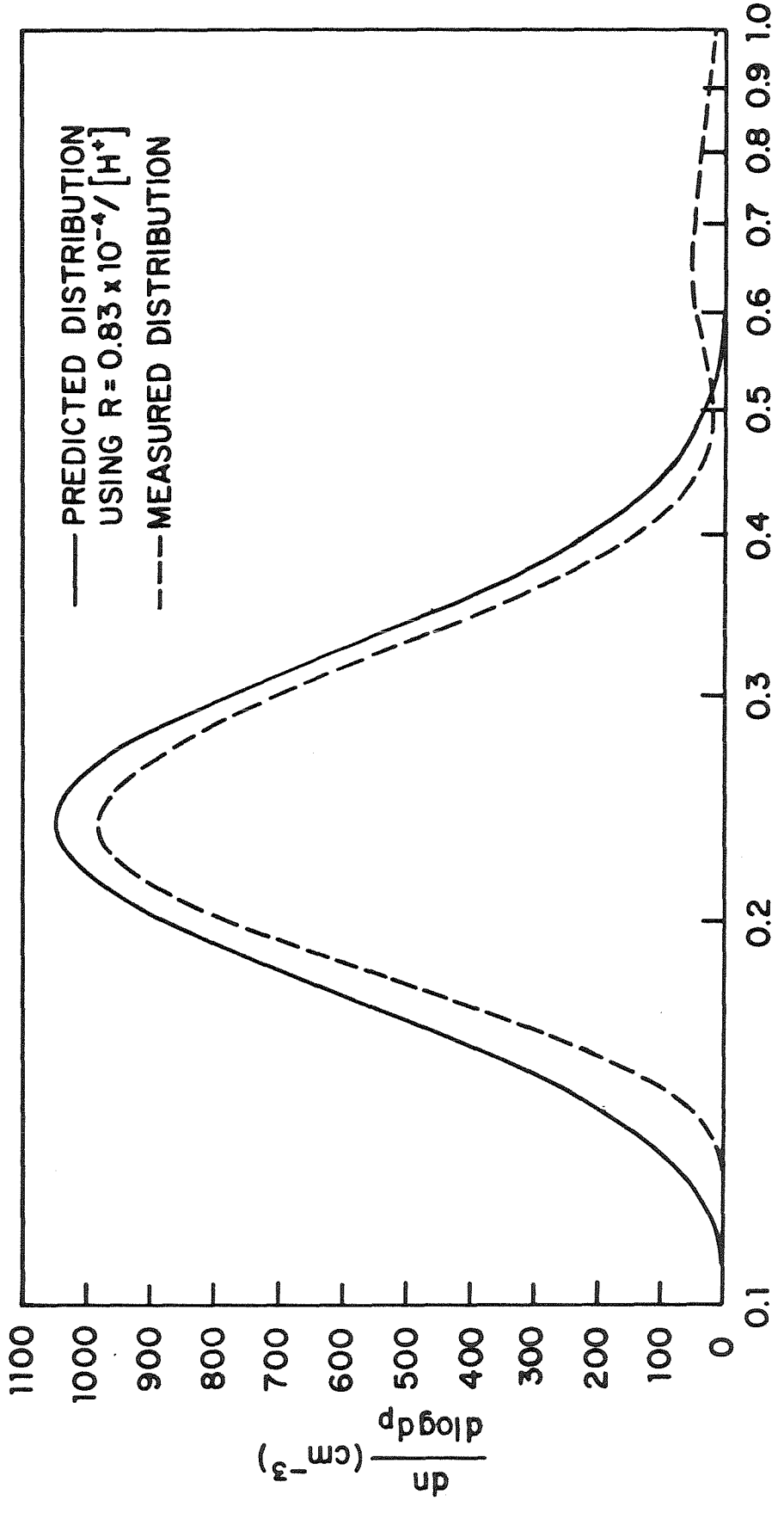


Figure 12. Predicted and measured effluent number distributions from experiment 3 based on rate R = $0.83 \times 10^{-4}/[H^+]$.

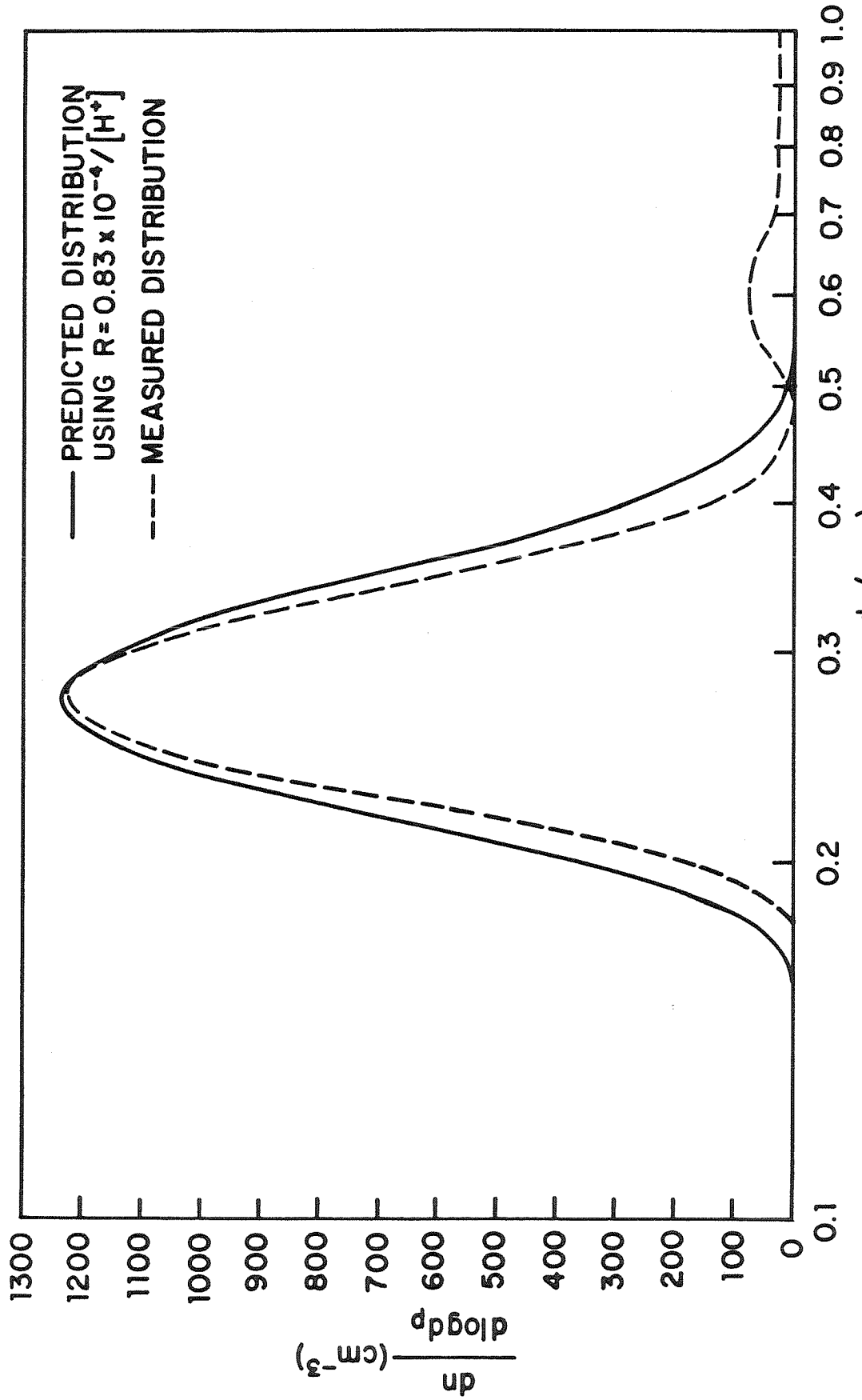


Figure 13. Predicted and measured effluent number distributions from experiment 4 based on rate R = $0.83 \times 10^{-4}/[H^+]$.

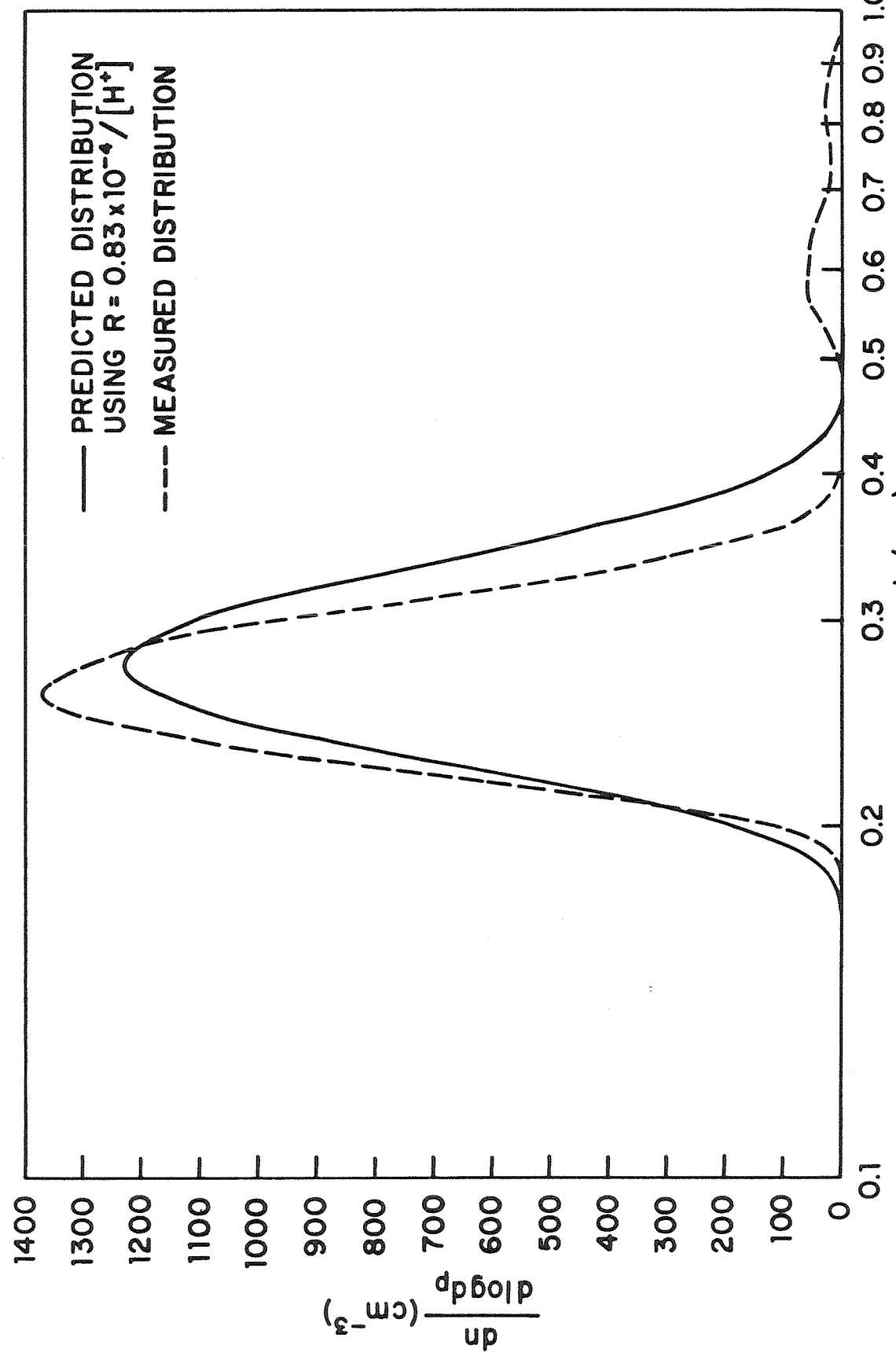


Figure 14. Predicted and measured effluent number distributions from experiment 6 based on rate R = $0.83 \times 10^{-4}/[H^+]$.

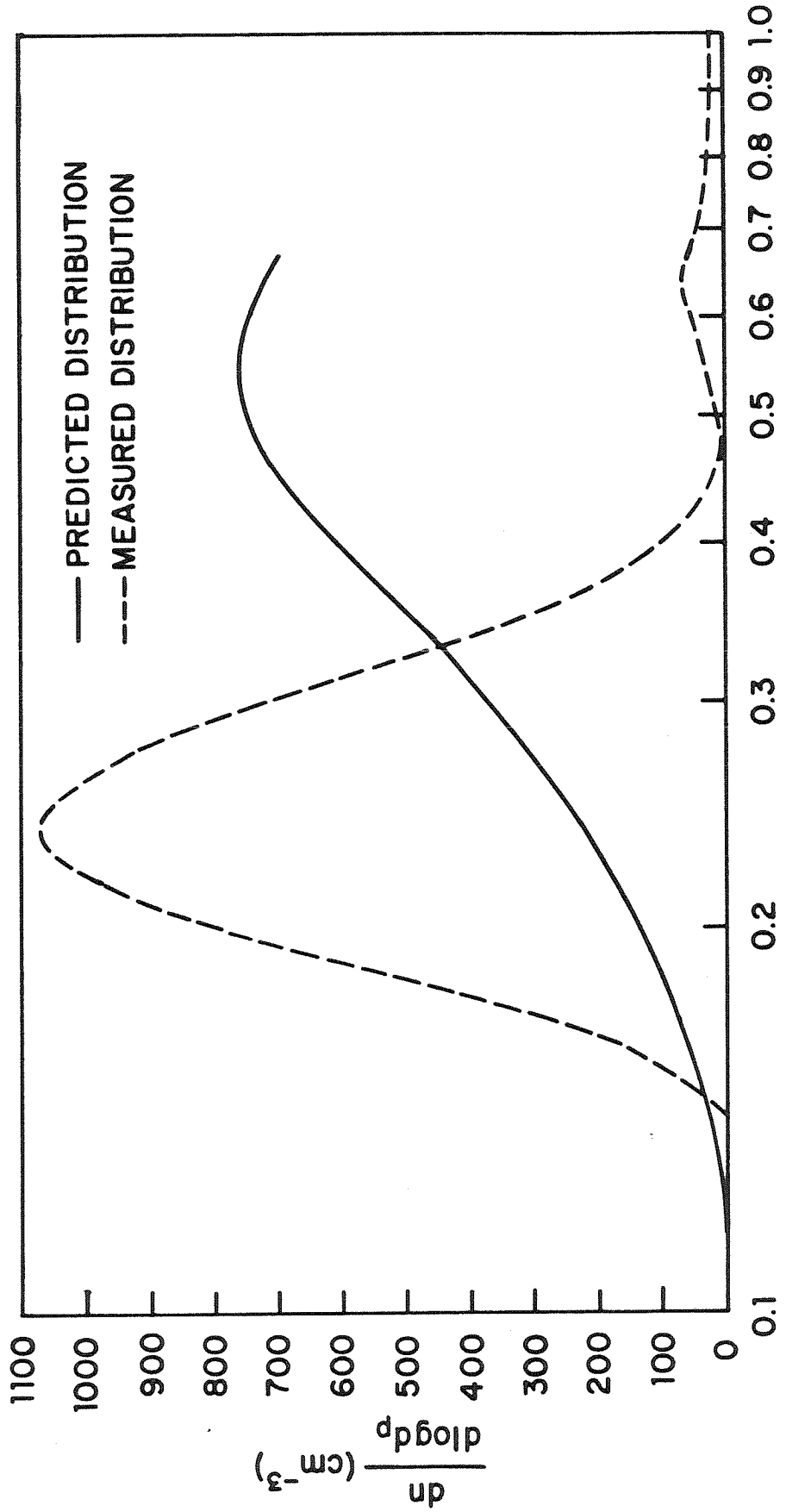


Figure 15. Predicted and measured effluent number distributions from experiment 1 using rate expression of Wadden et al. (1974).

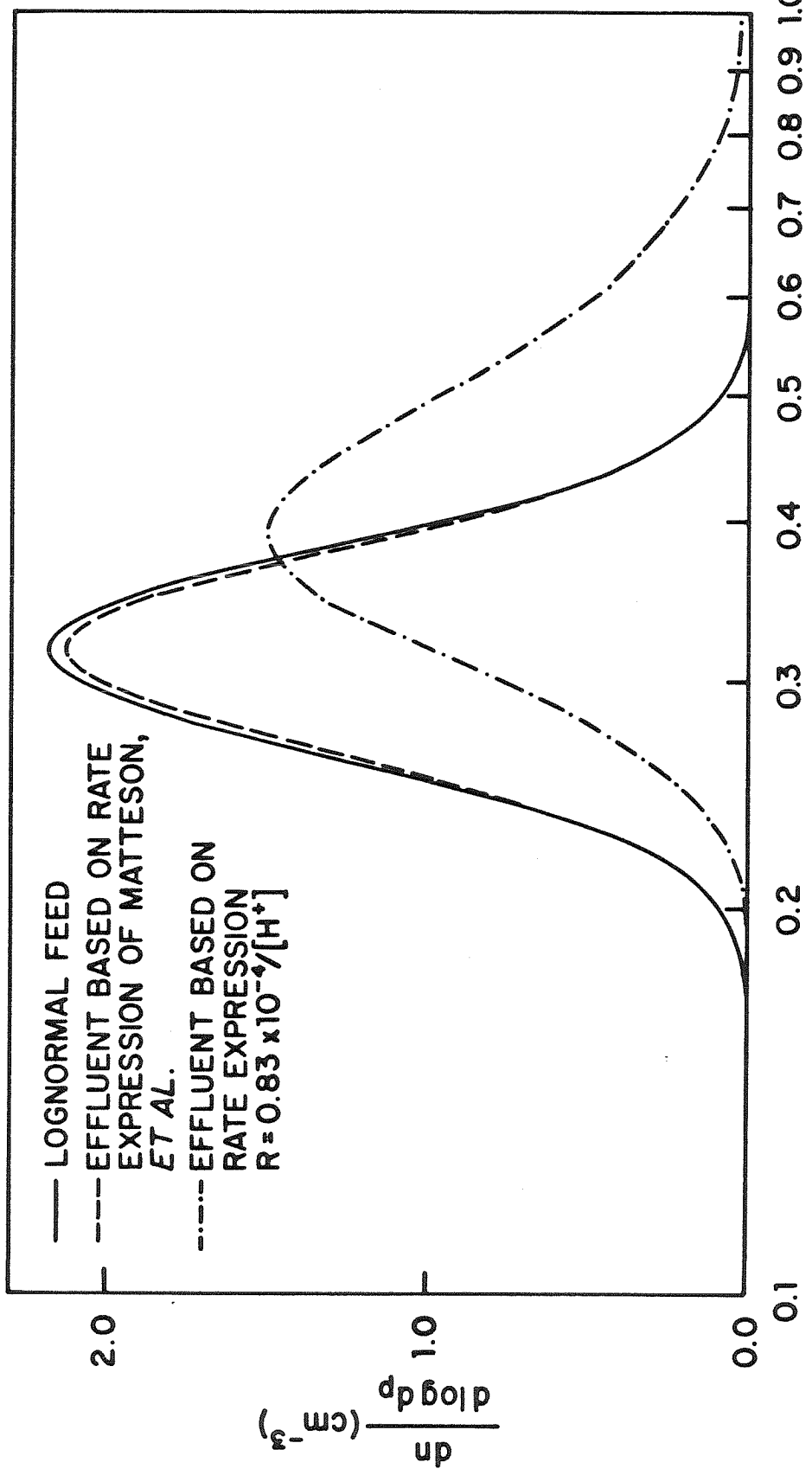


Figure 16. Solid curve is lognormal feed at 97 percent relative humidity. Dashed curve is predicted effluent distribution based on rate expression of Matteson, et al. and dotted-dashed curve is predicted effluent distribution using rate $R = 0.83 \times 10^{-4}/[H^+]$. CSTR conditions were assumed to be 50 ppm SO_2, 97 percent RH, 60 minutes residence time.

expression of Wadden et al. (1974) on the data of experiment 1. Note that the peak is much larger than either the experimental value, also shown, or the simulated value of Figure 10. Figure 16 shows the result of a similar calculation using the expression of Matteson et al. The input in this example is a lognormal distribution at 97 percent RH and an SO_2 concentration of 50 ppm. Note that practically no growth occurs. This is because at 50 ppm SO_2 the Matteson expression predicts a maximum H_2SO_4 concentration of 0.062 M, corresponding to a diameter increase of about two percent in Figure 17.

One important feature predicted by the rate expression of Equation (1) is the dependence of growth on relative humidity. Since higher relative humidities lead to more dilute particles and hence lower $[H^+]$ predicts more rapid growth at higher humidities. This is, in fact, observed in these experiments, as can be seen from Figures 1-6.

The range of concentrations for which the rate expression (1) holds is difficult to assess based on these experiments. Based on the assertion of Coughanowr and Krause (1965) that the manganese dependence of the rate is very weak above a concentration of approximately $3x10^{-3}$ M, a conservative estimate for validity of (1) might be at concentrations of manganese sulfate greater than 10^{-2} M. The range of gaseous SO_2 concentrations over which (1) holds is probably for SO_2 greater than 10 ppm as concluded previously.

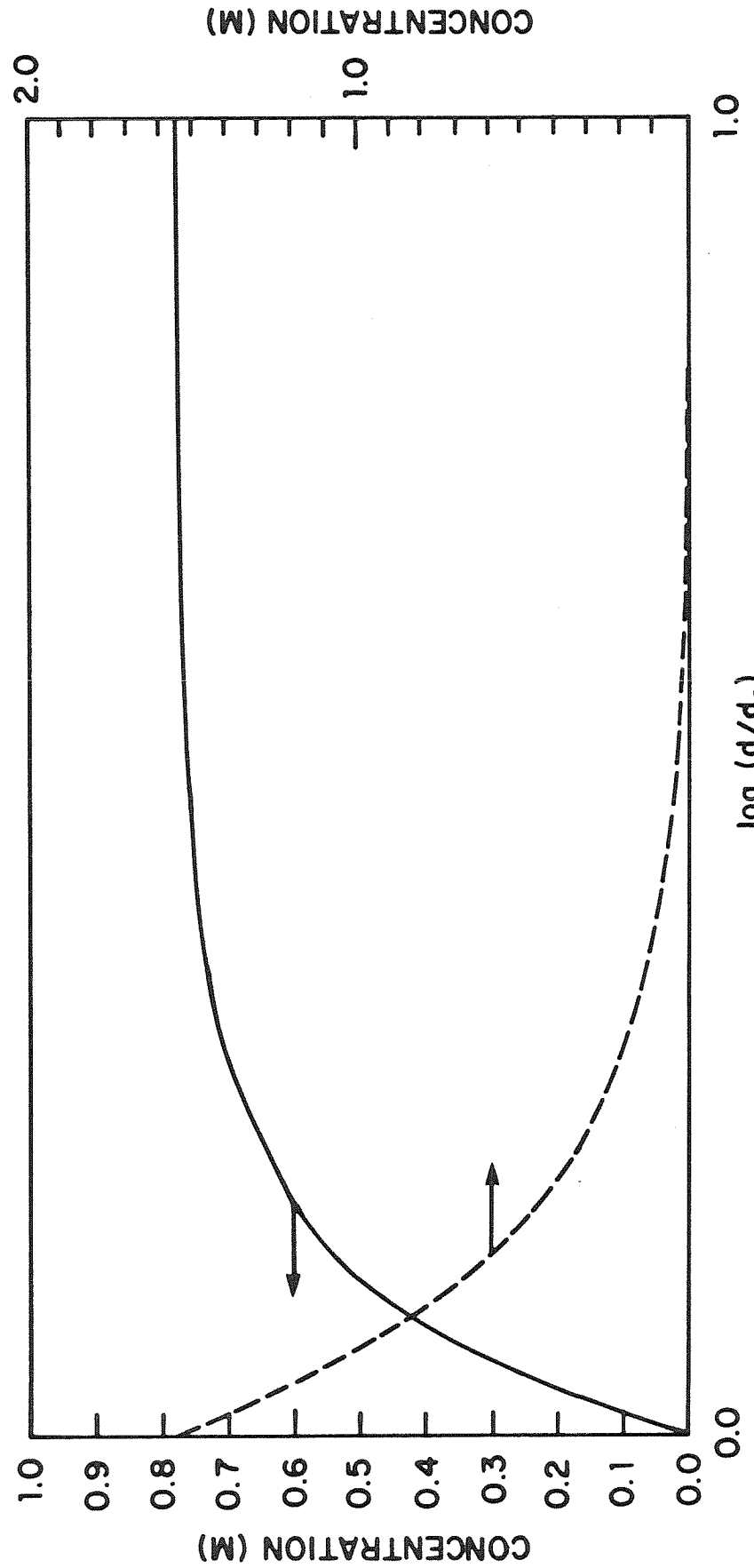


Figure 17. Sulfuric acid and manganese sulfate concentrations as function of the logarithm of ratio of particle diameter to feed particle diameter at 97 percent relative humidity.

CONCLUSIONS

Experiments in a continuous stirred tank reactor to measure growth rates of aqueous aerosols containing manganese sulfate in a humid atmosphere containing sulfur dioxide have been described. The results are consistent with the rate expression for oxidation of SO_2 in solutions of manganese sulfate

$$R = (8.3 \pm 2.5) \times 10^{-5}/[H^+] \text{ M sec}^{-1}$$

This expression is estimated to hold for $0.01 \leq [Mn^{++}]$ and 10 ppm $\leq [SO_2]$.

Because of the small amount of growth relative to the resolution of the optical particle counter used in measuring particle size distributions, these results must be considered somewhat tentative. In addition, the necessity to transform the size distribution due to differences in relative humidity in the measuring instrument and the reaction vessel introduces yet another element of uncertainty into the analysis.

In spite of these difficulties, however, the results were reasonably consistent and encourage further investigation.

REFERENCES

Bassett, H. and Parker, W. G. J. Chem. Soc., 1540 (1951).

Bassett, M., Gelbard, F., and Seinfeld, J. H. Atm. Env. 15, 2395 (1951).

Cains, P. W. and Carabine, M. D. J. Chem. Soc. Faraday Trans. I 11, 2689 (1978).

Cheng, R. T., Corn, M., and Frohliger, J. O. Atm. Env. 5, 987 (1971).

Coughanowr, D. R. Oxidation of SO_2 in Drops, Ph.D. Thesis. University of Illinois (1956).

Coughanowr, D. R. and Krause, F. E. Ind. Eng. Chem. Fund. 4, 61 (1965).

Crump, J. G. and Seinfeld, J. H. Aerosol Sci. and Tech. XX, XXX (1982).

Freiberg, J. Atm. Env. 12, 339 (1978).

Haury, G., Jordan, S., and Hofmann, C. Atm. Env. 12, 281 (1978).

Hoffmann, M. R. and Jacob, O. J. "Kinetics and Mechanism of the Catalytic Oxidation of Dissolved Sulfur Dioxide in Aqueous Solution: an Application to Nighttime Fogwater Chemistry" in Acid Precipitation: SO_2, NO, NO_2 Oxidation Mechanisms: Atmospheric Considerations. J. G. Calvert, ed. Ann Arbor Science Publishers, Ann Arbor (1982).

International Critical Tables, E. L. Washburn, ed., McGraw-Hill, New York (1928).

Johstone, H. F. and Coughanowr, D. R. Ind. Eng. Chem. 50, 1169 (1958).

Kaplan, D. J. and Himmelblau, D. M. Atm. Env. 15, 763 (1981).

Martin, L. R. manuscript (1982).

Matteson, M. J., Stober, W., and Luther, H. Ind. Eng. Chem. Fund. 8, 677 (1969).

Robinson, R. A. and Stokes, R. H. Electrolyte Solutions, Butterworths, London (1965).

Wadden, R. A., Quon, J. E., and Hulburt, H. M. Atm. Env. 8, 1009 (1974).

APPENDIX

The mechanism of Matteson, et al. for the manganese catalyzed liquid phase oxidation of SO_2 is given in the following reaction sequence:

$$SO_2 + Mn^{++} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} Mn\cdot SO_2^{++} \tag{A1}$$

$$2Mn\cdot SO_2^{++} + O_2 \underset{k_4}{\overset{k_3}{\rightleftharpoons}} [(Mn\ SO_2^{++})_2\cdot O_2] \tag{A2}$$

$$\underset{k_5'}{\overset{k_5}{\rightleftharpoons}} 2Mn\cdot SO_3^{2+} \tag{A3}$$

$$Mn\cdot SO_3^{++} + H_2O \underset{k_7}{\overset{k_6}{\rightleftharpoons}} Mn^{++} + HSO_4^- + H^+ \tag{A4}$$

The following variables are defined:

$B_o = [Mn^{++}]_o$

$R = [SO_2]$

$J = [HSO_4^-] = [H^+]$

$D = [Mn\cdot SO_2^{++}]$

$E = 2[(Mn\cdot SO_2^{++})_2\cdot O_2]$ (Matteson, et al. incorrectly give $\frac{1}{2}[(Mn\cdot SO_2^{++})_2\cdot O_2]$)

$F = [(Mn\cdot SO_3)^{++}]$

$X = D + E + F$

$B = [Mn^{++}] = B_o - X$

Using the above reaction sequence and the steady state approximation for the complexes gives the equations:

$$0 = \frac{dD}{dt} = k_1R(B_o-X) - k_2D - 2k_3[O_2]D^2 + 2k_4E^2 \tag{A5}$$

$$0 = \frac{dE}{dt} = 2k_3[O_2]D^2 - 2k_4E^2 - 2k_5E^2 + 2k_5'F^2 \tag{A6}$$

$$0 = \frac{dF}{dt} = 2k_5E^2 - k_6[H_2O]F - 2k_5'F^2 + 2k_7(B_o-X)J^2 \qquad (A7)$$

Adding Equations (A5), (A6), and (A7) gives

$$0 = \frac{dX}{dt} = k_1R(B_o-X) - k_2D - k_6[H_2O]F + 2k_7(B_o-X)J^2 \qquad (A8)$$

We make the further assumptions, following Matteson, et al. that the ratios D/X, E/X, and F/X are constant and write

$$k_6[H_2O]F = k_6'X \qquad (A9)$$

$$k_2D = k_2'X \qquad (A10)$$

Introducing these into Equation (A8) yields:

$$0 = k_1R(B_o-X) - k_2'X - k_6'X + 2k_7(B_o-X)J^2 \qquad (A11)$$

By stoichiometry, at constant total sulfur we have

$$R = R_o - X - J \qquad (A12)$$

Substitution of this into (A11) gives

$$0 = k_1(R_o-X-J)(B_o-X) - k_2'X - k_6'X + 2k_7(B_o-X)J^2 \qquad (A13)$$

Finally, we consider the initial stages at reaction where $J \approx 0$, and have the equation

$$0 = k_1(R_o-X)(B_o - X) - k_2'X - k_6'X \qquad (A14)$$

Solving for X, ignoring the X^2 term gives

$$X = \frac{k_1 R_o B_o}{k_1 R_o + k_1 B_o + k_2' + k_6'} = \frac{R_o B_o}{R_o + B_o + k_s} \tag{A15}$$

where $k_s = (k_2' + k_6')/k_1$. The loss rate of SO_2 is given by

$$\frac{dR}{dt} = - k_1 (R_o - X)(B_o - X) + k_2' X$$

$$= - \frac{k_1 (B_o^2 + k_s B_o)(R_o - X)}{R_o + B_o + k_s} + \frac{k_2' R_o B_o}{R_o + B_o + k_s} \tag{A16}$$

If we assume that $B_o << R_o$, and that R_o is large enough so that $k_s << R_o$, Equation (A16) assumes the form

$$\frac{dR}{dt} = - k_1 B_o^2 - (k_2' + k_6') B_o + k_2' B_o$$

$$= - k_1 B_o^2 - k_6' B_o \tag{A17}$$

Then, empirically $k_6' << k_1$, so provided B_o is not too small, we have

$$\frac{dR}{dt} = - k_1 B_o^2 \tag{A18}$$

This is the derivation of Matteson, et al. and based on this, Cains and Carabine determined k_1 by fitting (A18) to data of Johnstone and Coughanowr. However, when the assumption that $B_o << R_o$ is made, the solution of Equation (A14) is no longer valid, for the X^2 term is not small relative to the other terms. This can be seen by solving (A14) exactly, and noting that the solution is

$$X = \frac{R_o + B_o + k_s - \sqrt{(R_o + B_o + k_s)^2 - 4 R_o B_o}}{2} \tag{A19}$$

Ignoring k_s relative to R_o gives

$$X = \frac{R_o+B_o}{2} - \frac{(R_o+B_o)}{2}\sqrt{1 - \frac{4R_oB_o}{(R_o+B_o)^2}}$$

$$= \frac{R_o+B_o}{2} - \frac{R_o+B_o}{2}\left(1 - \frac{2R_oB_o}{(R_o+B_o)^2}\right)$$

$$= \frac{R_oB_o}{R_o+B_o} = B_o \qquad \text{(A21)}$$

Hence $X = B_o$, so the X^2 term is of the same magnitude as the B_oX term, so cannot be ignored. Now, using (A14), and the loss rate expression for SO_2

$$\frac{dR}{dt} = - k_1(R_o-X)(B_o-X) + k_2'X \qquad \text{(A22)}$$

we see that

$$\frac{dR}{dt} = - k_6'X = - k_6'B_o \qquad \text{(A23)}$$

This expression also follows directly from the solution of Equation (A11) by ignoring the J^2 term. This then gives

$$X = \frac{RB_o}{R+k_s} \qquad \text{(A24)}$$

and

$$\frac{dR}{dt} = - k_1 R(B_o - X) + k_2' X$$

$$= - \frac{k_1 k_s R B_o}{R + k_s} + \frac{k_2' R B_o}{R + k_s}$$

$$= - \frac{k_6' R B_o}{R + k_s} \qquad \text{(A25)}$$

And if $k_s << R$, we have

$$\frac{dR}{dt} = - k_6' B_o \qquad \text{(A26)}$$

Hence, no second-order dependence follows from the mechanism of Matteson, et al. and because of this fact, one must doubt the validity of the kinetic expression, and certainly the rederivation of rate coefficients by Carabine and Cains.

CHAPTER 12

RECOMMENDATIONS FOR FUTURE RESEARCH

RECOMMENDATIONS FOR FUTURE RESEARCH

The results obtained from the aerosol growth studies offer encouragement for the use of the CSTR in studying aerosol dynamics. It still appears that the best chemical kinetic data can be obtained by bulk studies, but the CSTR provides a useful technique by which these data can be evaluated in aerosol systems.

There are two principal difficulties in carrying out aerosol experiments, however; one is the control of temperature and relative humidity. The reaction vessel used in these experiments is insulated, and an attempt was made to control the temperature by placing heat exchange coils in the enclosure housing the vessel. This was successful in minimizing temperature variations during the very long experiments. Temperature variations seldom exceeded 0.2°C during a run. However, humidities above 97 percent are difficult to produce. Such humidities would have increased the rate of growth enabling more accurate measurement of the growth rate.

The second difficulty is in size distribution measurement. Although several techniques for size measurement are available, most are not suitable for aerosols which are sensitive to changes in relative humidity. The optical particle counter appeared to be the best suited, although it, too, was not entirely satisfactory, due to changes in relative humidity caused by its operating temperature. Although this problem can be corrected in principle, as was done in this work, it would be preferable to make direct measurements of the size distribution. This suggests the development of *in situ* optical measurement techniques. Although such techniques generally have larger lower size limits and poorer resolution, it seems that this

might be compensated by use of more sophisticated data inversion algorithms, such as those described in this work, together with larger data sets.

With regard to experimental systems worth investigating, an interesting one might be the manganese-iron (III) catalyzed oxidation of SO_2, since this is known to exhibit synergism (Martin, 1982) with higher rates of oxidation than either manganese or iron data alone account for. Hence, even with the optical particle counter available now, better resolution could be expected as a result of larger growth. Such a system is naturally more complicated, but should be amenable to the same kind of thermodynamic analysis employed here.

REFERENCES

Martin, L. R. manuscript (1982).

APPENDIX

MANUAL FOR AEROSOL SIZE DISTRIBUTION DATA INVERSION PROGRAMS INVERSE AND CINVERSE

INTRODUCTION

The programs INVERSE and CINVERSE are designed to implement the size distribution data inversion algorithms presented by Crump and Seinfeld (1982a, 1982b) on an interactive computer system. In this manual we discuss the structure of the programs to familiarize the user with their features. The use of the programs will be illustrated by some sample calculations.

DESCRIPTION OF ALGORITHMS

Both INVERSE and CINVERSE are based on regularization algorithms (Tikhonov and Arsenin, 1977). The general data inversion problem has the following equivalent forms (Crump and Seinfeld, 1982a)

$$\Lambda_i f = y_i \quad (i = 1,\ldots, n) \tag{1a}$$

$$\int_0^1 k_i(x)f(x)dx = y_i \qquad (i = 1,\ldots, n) \tag{1b}$$

where the Λ_i are linear functionals determined by instrument calibration, f is the unknown size distribution, and the y_i are measured data. The kernel functions k_i are determined by the functionals Λ_i. The above system of equations comprises an ill-posed mathematical problem, and consequently special methods are needed to solve it. INVERSE utilizes the regularization algorithm discussed by Crump and Seinfeld (1982a) to find the distribution f minimizing

$$\frac{1}{n}\sum_{n=1}^{n}(\Lambda_i f - y_i)^2/\sigma_i^2 + \lambda\|f\|^2 \tag{2}$$

where the σ_i are measured standard deviations in the data, and $\|\cdot\|$ is a seminorm of the form

$$\|f\|^2 = \int_0^1 f'(x)^2 dx \text{ or } \|f\|^2 = \int_0^1 f''(x)^2 dx \qquad (3)$$

The positive parameter λ, called the smoothing or regularization parameter, is chosen by minimizing the cross-validation function $V(\lambda)$ (Crump and Seinfeld, 1982a).

CINVERSE uses the same minimization functional of (2), but with the additional constraint on the distribution f that it be positive. Because of the constraint, the algorithm used is different, and the value of the regularization parameter λ must be chosen in a different way. CINVERSE uses the method of the discrepancy (Tikhonov and Arsenin, 1977) by which λ is chosen so that

$$\left[\frac{1}{n}\sum_{i=1}^{n}(\Lambda_i f - y_i)^2/\sigma_i^2\right]^{1/2} = 1 \qquad (4)$$

The values of σ_i^2 used in CINVERSE are defined to include errors in calibration data, and the latter are assumed accurate to within 10 percent. Thus σ_i^2 is given as

$$\sigma_i^2 = \sigma_i'^2 + 0.01 y_i^2 \qquad (5)$$

where σ_i' is the actual measurement standard deviation and the second term accounts for calibration error.

Both INVERSE and CINVERSE allow the regularization parameter to be chosen arbitrarily if desired, and if standard deviations are not known, all data are weighted equally ($\sigma_i = 1$ for all i).

LIMITATIONS

Because they are ill-posed, aerosol size distribution data inversion problems suffer from inherent limitations of the accuracy within which

solutions can be determined. Contrary to frequently expressed opinion, more powerful algorithms cannot overcome this difficulty. Consequently, some problems may be difficult to solve regardless of the algorithm.

Having made this caveat, we mention several sources of error which can interfere with the solution of ill-posed problems. Quadrature errors are important if aerosol size distributions show very sharp peaks. INVERSE and CINVERSE use Simpson's rule on a grid of up to 100 and 50 points, respectively. This is adequate in most cases, but can occasionally be a significant source of error. Calibration error is likely to be the most important source of error. Aerosol sizing instruments are difficult to calibrate accurately, and influences of particle shape, surface irregularities, etc. are almost always neglected. Consequently, calibration accuracy probably seldom exceeds 10 percent. Finally, the data themselves are imperfect, and this reduces somewhat the degree to which the solution can be determined.

PROGRAM INPUTS

The main inputs required by INVERSE and CINVERSE are calibration data for size measuring instruments. Data for the Low Pressure Impactor (Hering, et al., 1979) and the Electrical Aerosol Analyzer (Kapadia, 1980) are supplied in the form of subroutines CAL1 and CAL2. Dummy subroutines CAL3, CAL4, and CAL5 are provided and referenced in the main program. These are intended to be replaced by procedures to calculate the values of the kernel functions at discrete points of a size range selected by the user when the program is run.

To illustrate the use of these subroutines we consider a hypothetical six-channel optical particle counter. We suppose that measurements of instrument response have been made for 25 particle diameters. The diameters at

which the measurements have been made will be stored as the vector DIAM, and the responses in the six channels will be stored as the 6X25 array OPC. We will incorporate these data into the subroutine CAL3. The first three statements of the subroutine will be

```
      SUBROUTINE CAL3(NS,L,FK,D1,D2)
      DOUBLE PRECISION FK,D1,D2
      DIMENSION FK(25,100),DIAM(25),OPC(6,25)
```

In CINVERSE the dimensions for the array FK should be 50 X 50 instead of 25 X 100. The subroutine CAL3 is to evaluate the kernel functions k_i at L equally logarithmically spaced diameters, beginning at D1 and ending at D2. The value of the ith kernel function at the jth diameter is to be assigned the name FK(NS+I,J). Finally, the value of NS is to be incremented by 6, the number of channels of the instrument. Therefore, following the above statements we have

```
          DATA DIAM/ ..................................... /
          DATA OPC/ ...................................... /
          DO 40 J=1,L
          X=(J-1.)/(L-1.)
          D=D2**X/D1**(X-1.)
          DO 40 I=1,6
          IF(D.LT.DIAM(1)) GO TO 10
          GO TO 20
10        Z=OPC(I,1)
          GO TO 40
20        CONTINUE
          IF(D.GT.DIAM(25)) GO TO 30
          GO TO 35
30        Z=OPC(I,25)
35        CALL INTPLT(D,OPC,DIA,Z,I)
40        FK (I+NS,J)=Z
          NS=NS+6
          RETURN
          END
          SUBROUTINE INTPLT(D,OPC,DIA,Z,I)
          DIMENSION OPC(6,25),DIA(25)
          DO 20 K=1,50
          IF(D-DIAM(K))10,20,20
10        KP=K-1
20        CONTINUE
```

```
30        Z=OPC(I,KP)+(ALOG(D/DIAM(KP)))*(OPC(I,KP+1)-OPC(I,KP))/
      C   ALOG(DIAM(KP+1)/DIAM(KP))
          RETURN
          END
```

The DATA statements could, of course, be replaced by statements to read the data from a file. The subroutine INTPLT is a linear interpolation procedure.

In addition to the above, some minor changes in the main program are necessary. The DATA statement following the DIMENSION statement at the beginning of the program must be changed to include the number of channels in the new instrument. In this case it would now read

```
DATA ND/9,8,6,0,0,0,0,0,0,0/
```

Also the FORMAT statement 10 should be augmented to include the new instrument name.

Other inputs required include the number of intervals for quadrature, upper and lower diameter limits, and several convergence criteria, all of which are entered during execution following prompts.

USER INSTRUCTIONS FOR RUNNING PROGRAMS AND PROGRAM DESCRIPTIONS

INVERSE

Table I lists the important variables in this program, and Table II summarizes the subroutines used.

To illustrate the use of the program we go through an example. We suppose we have low pressure impactor data for the 8 stages of the impactor. The following is what would be seen on a terminal during execution of the program INVERSE. Lines marked # indicate user responses to prompts, and an asterisk indicates a comment.

TABLE I

Variable Name	Type	Meaning
APHI	25x25 DBLE PREC ARRAY	Gram matrix
APHIS	25x25 DBLE PREC ARRY	APHIS(I,J)=APHI(I,J)/SIGMA(I)/SIGM(J)
D1	DBLE PREC	Lower diameter limit in microns
D2	DBLE PREC	Upper diameter limit in microns
E	25x1 DBLE PREC ARRAY	Eigenvalues of Gram matrix
F	25x1 DBLE PREC ARRAY	Solution distribution
FK	25x100 DBLE PREC ARRAY	Kernel functions at equally logarithmically spaced points from D1 to D2
FKS	25x100 DBLE PREC ARRAY	Normalized kernel functions FKS(I,J)=FK(I,J)/SIGMA(I)
L	INTEGER	Number of points at which kernel functions are evaluated L = 2*M+1
M	INTEGER	Number of interval for quadrature of representing functions
M1	INTEGER	Number of points at which representing functions are evaluated M1=M+1
NDP	INTEGER	Number of data values
PHI	25x50 DBLE PREC ARRAY	Representing functions at equally logarithmically spaced points from D1 to D2
SIGMA	25x1 DBLE PREC ARRAY	Standard deviations of data
U	25x25 DBLE PREC ARRAY	Rows of U are eigenvectors of Gram matrix
V	25x1 DBLE PREC ARRAY	$V(I) = \int_0^1 tk_i(t)dt$
VS	25x1 DBLE PREC ARRAY	Normalized V VS(I)=V(I)/SIGMA(I)
Y	25x1 DBLE PREC ARRAY	Data
YS	25x1 DBLE PREC ARRAY	Normalized data YS(I)=Y(I)/SIGMA(I)

TABLE II

Subroutine	Inputs	Outputs	Purpose	Subordinate Subroutines and functions	References
CAL1	NS,L, D1,D2	FK	To calculate EAA kernel functions.	INTERP,S	Kapadia (1980)
CAL2	NS,L, D1,D2	FK	To calculate LPI kernel functions.	XLPI,XK	Hering, et al. (1974)
REP1	NDP,L, FK	PHI, APHI	To calculate representing functions and Gram matrix for inversion in $H^1(0,1)$		Crump and Seinfeld (1982a)
REP2	NDP,L FK	PHI, APHI, V	To calculate representing functions and Gram matrix for inversion in $H^2(0,1)$		Crump and Seinfeld (1982b)
SPECTR	NDP, APHIS	E,U	To diagonalize Gram matrix	ROTATN, JACOBI	Schwarz, Rutishauser and Stiefel (1973)
INVER1	NDP,M,E, U,PHIS, YS	F	To carry out inversion in $H^1(0,1)$	V1	Crump and Seinfeld (1982a)
INVER2	NDP,M,E, U,PHIS, VS,YS	F	To carry out inversion in $H^2(0,1)$	V2	Crump and Seinfeld (1982b)

```
          $RUN INVERSE
          THE FOLLOWING INSTRUMENTS MAY BE USED FOR DATA INVERSION
          ENTER DESIRED NUMBERS
          1 EAA
          2 LPI
          ENTER NUMBER OF INSTRUMENT
#         2
          ENTER NUMBER OF INSTRUMENT
#
*         Note that nothing was entered here, since only the LPI is
          being used.  A carriage return was executed, the blank being
          interpreted as a 0.

          ENTER NUMBER OF INTERVALS FOR QUADRATURE
          (MUST BE EVEN AND LESS THAN 50)
#         40
          ENTER LOWER DIAMETER IN MICRONS IN F FORMAT
#         0.5
          ENTER UPPER DIAMETER IN MICRONS IN F FORMAT
#         5.0
          THIS SUBROUTINE CALCULATES KERNEL FUNCTIONS
          FOR THE LOW PRESSURE IMPACTOR
          ENTER PARTICLE DENSITY IN G/CC
#         1.0
          ENTER 1 TO INVERT IN FIRST ORDER SOBOLEV SPACE
          ENTER 2 TO INVERT IN SECOND ORDER SPACE
#         2
*         The choice of Sobolev space is discussed by Crump and
          Seinfeld (1982b).  The results are generally not very different
          in the two spaces.

          ENTER 1 IF YOU INTEND TO ENTER RELATIVE
          STANDARD DEVIATIONS OF DATA; OTHERWISE HIT RETURN
#
*         If no standard deviations are entered, as is the case here,
          they are all set equal to 1.  In either case, the data and
          kernel functions are all normalized by dividing by the
          standard deviations.

          ENTER ERROR TOLERANCE FOR DIAGONALIZATION
          OF GRAM MATRIX IN D FORMAT
#         0.1D-16
*         The prompt is requesting a convergence criterion for the
          matrix diagonalization, done by a subroutine called SPECTR.
          Generally 10^-17 is a good choice, but occasionally it causes
          underflow, in which case a larger value is needed.

          THE EIGENVALUES OF THE GRAM MATRIX
          E(1)=0.3774946382269633D-06
          E(2)=0.1021145290821161D-01
```

```
          E(3)=0.2257521983729988D-04
          E(4)=0.8265088024713232D-06
          E(5)=0.2341875256301490D-03
          E(6)=0.4259721967781903D-05
          E(7)=0.1046034223045154D-06
          E(8)=0.2371007196315154D-08
          IF YOU WISH TO ALTER AN EIGENVALUE, ENTER
          CORRESPONDING NUMBER: OTHERWISE HIT RETURN
#
*         The purpose of the last message is to prevent problems later,
          since occasionally one of the smallest eigenvalues is nega-
          tive.  This is due to rounding error, and in such a case
          the negative value should be replaced by a small positive
          number.  This has practically no effect on the data inversion,
          since the highest frequencies, corresponding to the smallest
          eigenvalues, are excluded from the solution.

          ENTER DATUM Y(1)
#         0.026D00
          ENTER DATUM Y(2)
#         0.506D00
          ENTER DATUM Y(3)
#         0.3D01
          ENTER DATUM Y(4)
#         0.606D01
          ENTER DATUM Y(5)
#         0.685D01
          ENTER DATUM Y(6)
#         0.391D01
          ENTER DATUM Y(7)
#         0.104D01
          ENTER DATUM Y(8)
#         0.216D00

*         The data here are simulated from a lognormal distribution.
          They are, therefore, extremely accurate, having only quad-
          rature errors.

          IF YOU WISH TO SPECIFY CROSS VALIDATION PARAMETER
          ENTER 0; OTHERWISE ENTER 1.
#         1
*         We can either let the cross-validation algorithm choose the
          regularization parameter, or we can choose it arbitrarily.  In
          this case we allow the algorithm to choose it.

          ENTER TOLERANCE FOR SMOOTHING PARAMETER IN D FORMAT
#         0.1D-07

*         This is a tolerance for accuracy of determination of the
          smoothing parameter.  The appropriate value depends on the
          magnitude of the parameter, and hence on the problem being
          solved.
```

```
          ENTER VALUE OF STEP SIZE IN D FORMAT
#         0.1D-01
*         This parameter gives the initial step size for the search
          routine used to find the minimum in the cross-validation
          function

          X=0.0000000D+00   V(X)=0.1705829D-02
          INPUT DATA        CALCULATED DATA
          0.260D-01            0.260D-01
          0.506D+00            0.506D+00
          0.300D+01            0.300D+01
          0.606D+01            0.606D+01
          0.685D+01            0.685D+01
          0.391D+01            0.391D+01
          0.104D+01            0.104D+01
          0.216D+00            0.216D+00
          TO ENTER NEW DATA FOR INVERSION USING SAME
          STANDARD DEVIATIONS HIT RETURN;
          TO ENTER NEW STANDARD DEVIATIONS AS WELL, ENTER 1;
          TO TERMINATE ENTER 2
#         2
          FORTRAN STOP
          $
```

The output shown here is, first, the value of the cross-validation parameter, which in this case is 0. Also given is the minimum value of the cross-validation function V. In addition, the input data are reproduced along with corresponding data calculated from the recovered distribution. The program then prompts the user for more data, either using the same standard deviations, or with new ones, as the user wishes. Otherwise the run terminates.

The final size distribution in this case has been written into a file, and is reproduced in Table III. The values of the distribution are values of $dM/dlogd_p$, where M is mass, and d_p is particle diameter in microns. Note that the first value of the distribution is negative, although negligibly so. This frequently occurs near the ends of distributions, and is more severe in the case of a sharply peaked distribution. Figure 1 shows a plot of the

TABLE III

SIZE DIST.	DIAM(MICRONS)
-0.211D-01	0.050
0.195D+00	0.056
0.417D+00	0.063
0.659D+00	0.071
0.951D+00	0.079
0.134D+01	0.089
0.189D+01	0.100
0.267D+01	0.112
0.377D+01	0.126
0.524D+01	0.141
0.712D+01	0.158
0.939D+01	0.177
0.120D+02	0.199
0.148D+02	0.223
0.177D+02	0.251
0.206D+02	0.281
0.232D+02	0.315
0.254D+02	0.354
0.271D+02	0.397
0.281D+02	0.446
0.285D+02	0.500
0.281D+02	0.561
0.271D+02	0.629
0.254D+02	0.706
0.232D+02	0.792
0.206D+02	0.889
0.177D+02	0.998
0.148D+02	1.119
0.120D+02	1.256
0.938D+01	1.409
0.714D+01	1.581
0.528D+01	1.774
0.382D+01	1.991
0.271D+01	2.233
0.190D+01	2.506
0.132D+01	2.812
0.909D+00	3.155
0.613D+00	3.540
0.386D+00	3.972
0.192D+00	4.456
0.806D-02	5.00

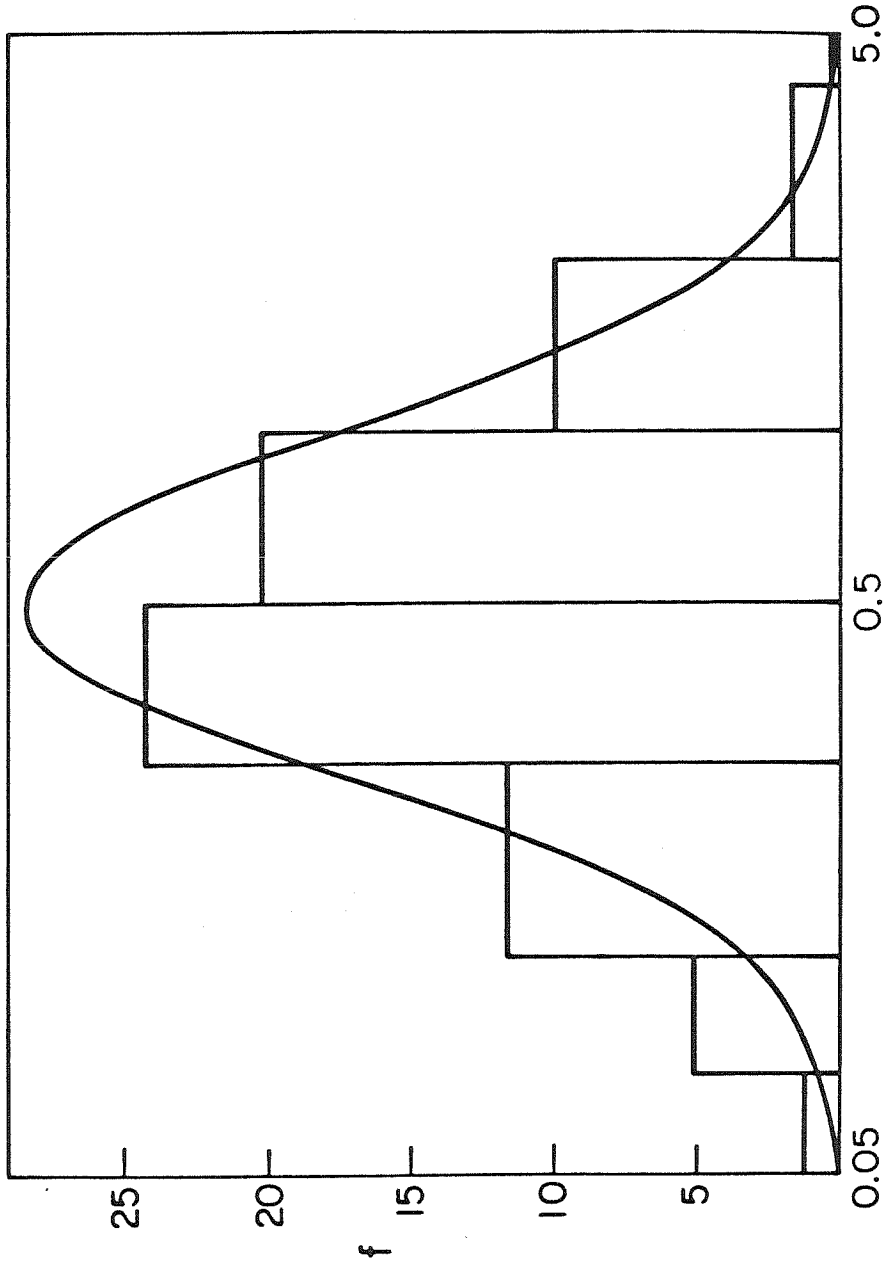


Figure 1. Inverted distribution obtained from simulated data from a lognormal distribution using program INVERSE. Histogram is distribution based on 50 percent cutoffs.

output distribution along with the histogram obtained by assuming ideal size cuts at the 50 percent cutoff particle diameter for each state.

CINVERSE

This program is structured in a similar way to INVERSE. Table IV lists the main variables and Table V gives the subroutines used in this program.

The use of the program is illustrated by means of an example. Here we use some actual data for ambient sulfur obtained with the Low Pressure Impactor. We assume the data are accurate within 10 percent.

```
          $RUN CINVERSE
          THE FOLLOWING INSTRUMENTS MAY BE USED FOR DATA INVERSION
          ENTER DESIRED NUMBERS
          1 EAA
          2 LIP
          ENTER NUMBER OF INSTRUMENT
#         2
          ENTER NUMBER OF INSTRUMENT
#
          ENTER NUMBER OF INTERVALS FOR QUADRATURE
          (MUST BE EVEN AND LESS THAN 50)
#         40
          ENTER LOWER DIAMETER IN MICRONS IN F FORMAT
#         0.05
          ENTER UPPER DIAMETER IN MICRONS IN F FORMAT
#         5.0
          THIS SUBROUTINE CALCULATES KERNEL FUNCTIONS
          FOR THE LOW PRESSURE IMPACTOR
          ENTER PARTICLE DENSITY IN G/CC
#         1.0

*         In this case the actual density of the aerosol is unknown.
          However, for the impactor, the calibration is in terms of the
          aerodynamic diameter, and a conversion is done to Stokes
          diameter.  Since these two are the same for unit density,
          the effect of entering unity for the density is to give the
          output distribution in terms of aerodynamic rather than
          Stokes diameter.

          ENTER 1 to INVERT IN FIRST ORDER SOBOLEV SPACE
          ENTER 2 to INVERT IN SECOND ORDER SPACE
#         2
          ENTER 1 IF YOU INTEND TO ENTER STANDARD
          DEVIATIONS OF DATA; OTHERWISE HIT RETURN.
#         1
```

TABLE IV

Variable name	Type	Meaning
D	50x50 DBLE PREC ARRAY	Difference operator approximation to first or second order differential operator
DP	50x50 DBLE PREC ARRAY	Weighted product of difference operator with itself. DP(I,J)= $\sum_K$ D(K,I)*V(K)*D(K,J)
D1	DBLE PREC	Lower diameter limit in microns
D2	DBLE PREC	Upper diameter limit in microns
F	50x1 DBLE PREC ARRAY	Solution distribution
FK	50x50 DBLE PREC ARRAY	Kernel functions at equally spaced logarithmic intervals from D1 to D2. Also weighted version of same, with weights of vector V. FK(I,J) = FK(I,J)* V(J)
FKS	50x50 DBLE PREC ARRAY	Normalized weighted kernel functions FKS(I,J)=FK(I,J)/S(I)
FM	50x50 DBLE PREC ARRAY	FM(I,J)= $\sum_K$ FKS(K,I)*FKS(K,J)
M	INTEGER	Number of intervals for quadrature
M1	INTEGER	Number of points at which kernel functions are evaluated. M1=M+1
NDP	INTEGER	Number of data values
S	50x1 DBLE PREC ARRAY	S(I)=(SIGMA(I)**2+0.01*Y(I)**2)$^{\frac{1}{2}}$
SIGMA	50x1 DBLE PREC ARRAY	Standard deviations of data
V	50x1 DBLE PREC ARRAY	Weights for quadrature by Simpson's rule
Y	50x1 DBLE PREC ARRAY	Data
YS	50x1 DBLE PREC ARRAY	Normalized data. YS(I)=Y(I)/S(I)

TABLE V

Subroutine	Inputs	Outputs	Purpose	Subordinate Subroutines and functions	References
CAL1	NS,M1, D1,D2	FK	To calculate EAA kernel functions	INTERP,S	Kapadia (1980)
CAL2	NS,M1, D1,D2	FK	To calculate LPI kernel functions.	XLPI,XK	Hering, et al., (1979)
REP1	M	D	To calculate first difference matrix for inversion in $H^1(0,1)$		
REP2	M	D	To calculate second difference matrix for inversion in $H^2(0,1)$		
LEMKE	NDP,M1 FM,FKS, DP,YS	F	To carry out inversion using Lemke algorithm	MATRIX, INITIA, NEWBAS, SORT, PIVOT, PRINT	Lemke (1965), Ravindran (1972), Crump and Seinfeld (1982b)

```
          ENTER SIGMA(1)
#         1.D00
          ENTER SIGMA(2)
#         1.D00
          ENTER SIGMA(3)
#         1.7D00
          ENTER SIGMA(4)
#         6.3D
          ENTER SIGMA(5)
#         1.2D01
          ENTER SIGMA(6)
#         1.0D01
          ENTER SIGMA(7)
#         4.4D00
          ENTER SIGMA(8)
#         1.7D00
          ENTER DATUM Y(1)
#         0.0D00
          ENTER DATUM Y(2)
#         1.1D01
          ENTER DATUM Y(3)
#         1.7D01
          ENTER DATUM Y(4)
#         6.3D01
          ENTER DATUM Y(5)
#         1.21D02
          ENTER DATUM Y(6)
#         1.01D02
          ENTER DATUM Y(7)
#         4.4D01
          ENTER DATUM Y(8)
#         1.7D01
          ENTER SMOOTHING PARAMETER IN D FORMAT
#         0.1D-03
          COMPLEMENTARY SOLUTION
          ITERATION No. 48
          XLO=0.1000000D-03     YO=0.2119790D+01
          TO CONTINUE FINDING SMOOTHING PARAMETER HIT RETURN;
          OTHERWISE ENTER 1
#
          ENTER SMOOTHING PARAMETER IN D FORMAT
#         0.1D-04
          COMPLEMENTARY SOLUTION
          ITERATION No. 60
          XLO=0.1000000D-04    YO=0.2040626D+01
          TO CONTINUE FINDING SMOOTHING PARAMETER HIT RETURN;
          OTHERWISE ENTER 1
#
          THE NEW VALUE OF THE SMOOTHING PARAMETER IS
          0.1669745D-30
          IF THIS IS SATISFACTORY, HIT RETURN; OTHERWISE TO
          ENTER ANOTHER VALUE HIT 1.
```

```
#       1
        ENTER VALUE OF SMOOTHING PARAMETER IN D FORMAT
#       0.1D-07
        COMPLEMENTARY SOLUTION
        ITERATION NO. 41
        XLO=0.1000000D-07   YO=-0.508077D+00
        TO CONTINUE FINDING SMOOTHING PARAMETER HIT RETURN;
        OTHERWISE ENTER 1.
#
        THE NEW VALUE OF THE SMOOTHING PARAMETER IS
        0.1608610D-07
        IF THIS IS SATISFACTORY, HIT RETURN; OTHERWISE TO
        ENTER ANOTHER VALUE HIT 1.
#
        COMPLEMENTARY SOLUTION
        ITERATION No. 41
        XLO=0.1608610D-07  YO=0.1800260D-01
        TO CONTINUE FINDING SMOOTHING PARAMETER HIT RETURN;
        OTHERWISE ENTER 1.
#
        THE NEW VALUE OF THE SMOOTHING PARAMETER IS
        0.1529093D-07
        IF THIS IS SATISFACTORY HIT RETURN; OTHERWISE
        TO ENTER ANOTHER VALUE HIT 1.
#
        COMPLEMENTARY SOLUTION
        ITERATION No. 41
        XLO=0.1529093D-07  YO=0.1760074D-02
        TO CONTINUE FINDING SMOOTHING PARAMETER HIT RETURN;
        OTHERWISE ENTER 1.
#       1
        INPUT DATA          CALCULATED DATA
        0.000D+00             0.109D+01
        0.110D+02             0.993D+01
        0.170D+02             0.195D+02
        0.630D+02             0.530D+02
        0.121D+03             0.905D+02
        0.101D+03             0.919D+02
        0.440D+02             0.473D+02
        0.170D+02             0.171D+02
        TO ENTER NEW DATA FOR INVERSION USING SAME
        STANDARD DEVIATIONS HIT RETURN;
        TO ENTER NEW STANDARD DEVIATIONS AS WELL, ENTER 1;
        TO TERMINATE ENTER 2
#       2
         FORTRAN STOP
        $
```

After the last data value has been entered control transfers to the subroutine LEMKE, which contains a quadratic programming routine (Ravindran, 1972) as well as a user-interactive iterative routine for choosing the smoothing parameter so as to satisfy Equation (4). In this case the initial guess was chosen to be 0.1×10^{-3}. LEMKE then indicates a solution has been found after 48 iterations. It then repeats the value of the smoothing parameter used in find the solution, as XLO, along with the value of YO, which is the difference between the left and right sides of Equation (4). The value of YO, therefore, is 0 when the smoothing parameter is chosen correctly. In this case it is not, so we proceed by entering another guess. Since the difference YO was positive, we need to use a smaller value of the smoothing parameter, so 0.1×10^{-4} is entered. This gives a value of YO which is only slightly smaller than the previous one, so we suspect that a much smaller value of the parameter is needed. After a carriage return LEMKE gives us a corrected guess for the smoothing parameter, found by a secant method from the previous two values. Because our first two guesses were poor, the corrected value is very small, so we choose to attempt a better guess and enter 0.1×10^{-7}. This gives a negative value of YO, but it is close to 0, so we now simply hit carriage returns until the value of YO becomes negligible.

As in INVERSE the final solution is used to compute values of the data, and these are displayed together with the actual data. The resulting output is shown in Table VI. Figure 2 shows a plot of this distribution together with the histogram from the 50 percent stage cutoffs.

TABLE VI

SIZE DIST	DIAM(MICRONS)
0.713D+01	0.050
0.384D+02	0.056
0.696D+02	0.063
0.101D+03	0.071
0.132D+03	0.079
0.162D+03	0.089
0.192D+03	0.100
0.220D+03	0.112
0.247D+03	0.126
0.272D+03	0.141
0.295D+03	0.158
0.313D+03	0.177
0.330D+03	0.199
0.341D+03	0.223
0.349D+03	0.251
0.350D+03	0.281
0.348D+03	0.315
0.340D+03	0.354
0.328D+03	0.397
0.310D+03	0.446
0.289D+03	0.500
0.263D+03	0.561
0.236D+03	0.629
0.205D+03	0.706
0.174D+03	0.792
0.143D+03	0.889
0.114D+03	0.998
0.896D+02	1.119
0.685D+02	1.256
0.548D+02	1.409
0.449D+02	1.581
0.413D+02	1.774
0.398D+02	1.991
0.401D+02	2.233
0.401D+02	2.506
0.373D+02	2.812
0.328D+02	3.155
0.252D+02	3.540
0.167D+02	3.972
0.811D+01	4.456
0.000D+00	5.00

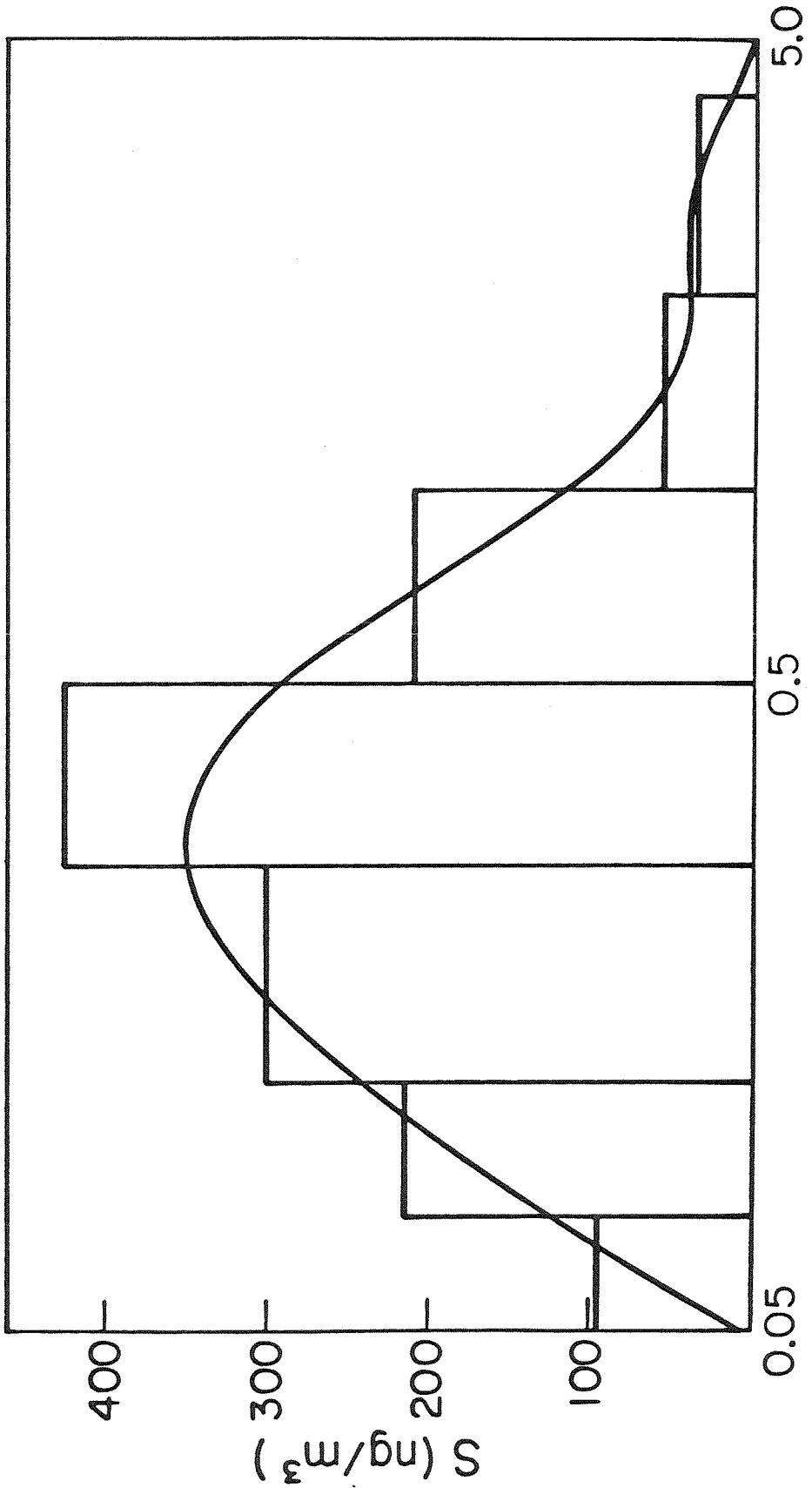


Figure 2. Inverted sulfur distribution from low pressure impactor data using program CINVERSE. Histogram is distribution based on 50 percent cutoffs.

REFERENCES

Crump, J. G. and Seinfeld, J. H. Aerosol Sci. and Tech., 1, 1 (1982a).

Crump, J. G. and Seinfeld, J. H. Submitted Aerosol Sci. and Tech. (1982b).

Hering, S. V., Friedlander, S. K. Collins, J. J. and Richards, L. W., Env. Sci. Tech. 13, 184 (1979).

Kapadia, A., Ph.D. Thesis, University of Minn. (1980).

Lemke, C. E. Management Sci. 11, 681 (1965).

Ravindran, A. Comm. ACM 15, 818 (1972).

Schwarz, H. R., Rutishauser, H. and Stiefel, E., Numerical Analysis of Symmetric Matrices, Prentice-Hall, Englewood Cliffs, N.J. (1973).

Tikhonov, A. N. and Arsenin, V. Y., Solutions of Ill-Posed Problems, Halsted Press, Washington (1977).