NUREG/CR-4342
SAND84-1307
RG
Printed September 1985

Uncertainty and Sensitivity Analysis of a Model for Multicomponent Aerosol Dynamics

J. C. Helton, R. L. Iman, J. D. Johnson, C. D. Leigh

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550
for the United States Department of Energy
under Contract DE-AC04-76DP00789

Prepared for
U. S. NUCLEAR REGULATORY COMMISSION

Available from

Superintendent of Documents
U.S. Government Printing Office
Post Office Box 37082
Washington, D.C. 20013-7982

and

National Technical Information Service
Springfield, VA 22161

NUREG/CR-4342
SAND84-1307
RG

UNCERTAINTY AND SENSITIVITY ANALYSIS OF A MODEL FOR MULTICOMPONENT AEROSOL DYNAMICS

J. C. Helton,[1] R. L. Iman,[2] J. D. Johnson,[3] C. D. Leigh[2]

Printed: September 1985

Sandia National Laboratories
Albuquerque, NM 87185
Operated by
Sandia Corporation
for the
U.S. Department of Energy

Prepared for
Division of Risk Analysis and Operations
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555
Under Memorandum of Understanding DOE 40-550-75
NRC Fin No. A-1339

---

1 Department of Mathematics
Arizona State University
Tempe, AZ 85287

2 Safety and Environmental Studies Division 6415
Sandia National Laboratories
Albuquerque, NM 87185

3 Science Applications International Corp.
Albuquerque, NM 87102

ABSTRACT

An uncertainty and sensitivity analysis of the MAEROS model for multicomponent aerosol dynamics is presented. Analysis techniques based on Latin hypercube sampling and regression analysis are used to study the behavior of a two component aerosol in a nuclear power plant containment for a transient accident with loss of AC power (i.e., a TMLB' accident). Conditional on assumed ranges and distributions for selected independent variables (e.g., initial distributions and mass loadings for each component, temperature, pressure, shape factors), estimates are made for distributions of model predictions and for the independent variables which influence these predictions. The analysis indicated that, for the situation under consideration, variables related to agglomeration (e.g., dynamic shape factor, material density, agglomeration shape factor, and turbulence dissipation rate) tended to dominate the observed variability. For comparison, an analysis based on differential techniques is also given. Further, a study of the effects on MAEROS predictions due to the number of particle size classes and the particle size class boundaries is presented. This analysis was performed as part of a project to develop a new system of computer codes (i.e., the MELCOR code system) for use in risk assessments for nuclear power plants.

Table of Contents

List of Figures

List of Tables

Acknowledgment

The authors wish to acknowledge the support provided by J. E. Brockmann, K. K. Murata, and J. L. Sprung of Sandia National Laboratories in the development of this analysis.

1. INTRODUCTION. The following presentation has two purposes. The first is to present and illustrate techniques for uncertainty and sensitivity analysis of complex models. The second is to present an uncertainty and sensitivity analysis of a problem involving a two-component aerosol in a dry reactor containment building. The uncertainty and sensitivity analysis techniques are based on Latin hypercube sampling and regression analysis. The behavior of the aerosol is predicted with the model contained in the MAEROS program (Gelbard, 1982).

This work arises from a project supported by the U.S. Nuclear Regulatory Commission at Sandia National Laboratories to develop a new system of computer codes (i.e., the MELCOR code system) for use in risk assessments for nuclear power plants (Sprung et al., 1983). Ultimately, this code system is intended as a replacement for various risk assessment codes which arose from the Reactor Safety Study (U.S. Nuclear Regulatory Commission, 1975). Among the project's charters is the requirement that the resultant code system be amenable to uncertainty and sensitivity analysis. This presentation derives from an effort to compile and compare different techniques for uncertainty and sensitivity analysis that might be suitable for use in conjunction with the MELCOR code system (Iman and Helton, 1985). The aerosol model implemented in the MAEROS program was picked for consideration because a modification of it will be incorporated into the MELCOR code system. Thus, it was felt that consideration of an analysis problem with this model would provide useful insight for the MELCOR code development effort with respect to both uncertainty and sensitivity analysis techniques and the behavior of an important part of the code system. Further, it was felt that the analysis techniques and results would be of interest beyond the MELCOR effort.

The paper is organized in the following manner. Section 2 presents the aerosol problem for analysis, and Section 3 describes the aerosol model implemented in MAEROS for noncondensing atmospheres. Then, Section 4 introduces the uncertainty and sensitivity analysis techniques, and Section 5 presents the results of the analysis. Section 6 describes how the uncertainty and sensitivity analysis techniques under consideration might be used to study aggregation problems, and Section 7 presents a differential sensitivity analysis for the problem under consideration. Finally, Section 8 contains a concluding discussion.

2. PROBLEM FOR ANALYSIS. The problem analyzed in this paper involves the behavior of aerosols suspended in the containment building of a nuclear power plant during a hypothetical severe core-melt accident. Such accidents are believed to progress through two major aerosol production phases. The first phase is before the reactor pressure vessel fails. The core temperature rises and aerosols containing volatile core components are produced. These in-vessel aerosols are leaked slowly into the reactor coolant system until the reactor pressure vessel fails catastrophically; at this time, the remaining aerosols are driven into the containment as a puff release. The second phase is after the reactor pressure vessel fails. The molten core interacts with the concrete basemat producing ex-vessel aerosols containing less volatile core materials. These core-concrete interactions produce a source of aerosols into the containment for several hours after vessel failure.

The scope of the problem is limited to the time period after vessel failure when the in-vessel aerosols have been released and the core-concrete interactions are providing an ex-vessel aerosol source into containment. Two aerosol components are considered: in-vessel aerosols and ex-vessel aerosols. The first component is assumed to be suspended at the start of the problem and the second component is introduced as a source. The distributions of particle size for both the initial aerosol suspended in the containment and the aerosol source are assumed to be lognormal with respect to particle diameter. This assumption is consistent with lognormal particle size distributions observed in experiments (Whitby, 1975).

A large, dry containment building is assumed. The geometric specifications are similar to those for the Zion nuclear power plant. Containment atmospheric conditions are similar to those predicted by computer analysis (Lipinski et al., 1985) for a transient accident with loss of AC power (i.e., a TMLB' accident). In particular, based on calculations for TMLB' accidents using the CONTAIN computer code (Bergeron et al., 1985), a super-heated containment atmosphere is assumed. The problem is not designed to simulate a TMLB' accident in the Zion power plant but rather to represent conditions possible during that type of accident.

Table 2.1 lists the independent variables considered in this investigation. In addition, the containment volume and the particle slip coefficient are taken to be 7.68E4 m^3 and 1.37, respectively. Each variable in Table 2.1 is assigned a range of possible input values based on conditions expected in a TMLB' accident. In most cases, defining the range of possible input values is hindered by a lack of analytical methods for defining the quantity and a lack of direct measurements of the quantity. The ranges given in Table 2.1 reflect information gathered from CONTAIN computer calculations, simple scoping calculations, and published experimental and analytical data. Several sets of experiments provide perspective with respect to the nature of aerosols likely to be found in nuclear reactor containments. Two of the most useful sets are the experiments conducted at Oak Ridge National Laboratory with sodium oxide and uranium oxide aerosols (Adams and Tobias, 1983) and the ABCOVE tests (Hilliard et al., 1983). Other relevant experimental programs are presented in Kress and Tobias (1980), Adams et al. (1980, 1981, 1982), Rehn (1980), Fermandjian et al. (1980), and Gieseke et al. (1984). Values for the mass median diameter of the size distribution, the standard deviation of the size distribution, the aerosol shape factors, and the particle material density are based on the observations reported in such experiments. Ranges for the initial aerosol mass loading and the aerosol source rate and timing are derived from calculations of in-vessel and ex-vessel aerosol production reported in a recent study on the reevaluation of the radiological source term (Gieseke et al., 1984). The input value ranges given in Table 2.1 for containment atmosphere parameters (i.e., temperature, pressure, molecular weight of atmospheric gas, ratio of thermal conductivity of atmospheric gas to that of suspended particles, and containment temperature gradients) are bounded by time-dependent variation for these parameters calculated for TMLB' accidents by the CONTAIN code. The diffusion boundary layer thickness values are based on the work of van de Vate (1980). However, boundary layer thicknesses as low as 1.E-5 m have been used (Lipinski et al., 1985).

Scoping calculations for the turbulent energy dissipation rate similar to those given in Appendix E of Lipinski et al. (1985) are the basis for the range chosen for turbulent energy dissipation in this study. Values given in Appendix E bound those

Table 2.1 Variables Used in MAEROS Analysis

Variable	Definition	Range	Distribution	Restrictions
GSD1	Geometric standard deviation for diameter of first component (unitless)	1.3 to 4.	Uniform	
MMD1	Mass median diameter for first component (m)	.2E-6 to 5.0E-6	Loguniform	.5 rank correlation with RM1
RM1	Total released mass for first component (kg)	100. to 1000.	Loguniform	.5 rank correlation with MMD1
GSD2	Geometric standard deviation for diameter of second component (unitless)	1.3-4.	Uniform	
MMD2	Mass median diameter for second component (m)	.05E-6 to 5.0E-6	Loguniform	.5 rank correlation with SR2
SR2	Total mass source rate for second component (kg/sec)	.1 to 1.	Loguniform	.5 rank correlation with MMD2
RD2	Release duration for second component (sec)	7.2E3 to 1.8E4	Uniform	
T	Temperature (K)	375. to 600.	Uniform	T < .001 P
P	Pressure (Pa)	3.75E5 to 8.E5	Triangular with apex at 5.8E-5	P > 1000. T
RCV	Ratio of ceiling area to volume (m^{-1}); RCV also assumed to be ratio of floor area to volume	.025-.075	Uniform	
RWV	Ratio of wall area to volume (m^{-1})	.08 to .24	Uniform	
χ	Dynamic shape factor (unitless)	1. to 3.	Uniform	$\chi \leq \gamma$
δ	Diffusion boundary layer thickness (m)	5.E-5 to 8.E-3	Loguniform	
ρ	Particle material density (kg/m^3)	2.E3 to 8.E3	Normal	
C_T	Constant associated with thermal accommodation coefficient (unitless)	1. to 3.	Uniform	
γ	Agglomeration shape factor (unitless)	1. to 3.	Uniform	$\gamma \geq \chi$
ST	Probability sticking factor (unitless)	.5 to 1.	Uniform	
∇T	Temperature gradient (K/m)	0. to 5.E-4	Uniform	
κ	Ratio of thermal conductivity of gas to that of particle (unitless)	.05 to 1.	triangular with apex at .5	
ϵ	Turbulence dissipation rate (m^2/sec^3)	.001 to .03	Uniform	
MW	Molecular weight of gas (kg/kg-mole)	20. to 40.	Uniform	

given here. Similarly, scoping estimates of the area available for settling in the reactor containment define the ranges chosen for the area to volume ratios in Table 2.1. The input ranges for the constant associated with the thermal accommodation coefficient and the probability sticking factor are strictly postulated.

For most of the parameters in Table 2.1, little is known about the distribution of values within the specified ranges. As a result, most of the parameters are given uniform or loguniform distributions. Loguniform distributions are used when the input range spans several orders of magnitude.

The following dependent variables are considered in this investigation: total suspended mass of each component as a function of time (kg), time to 90 percent and 99 percent deposition of the first component (sec), and time-integrated concentration of each component (kg sec/m^3). Total suspended mass of each component is related to the amount of radioactivity that can potentially reach the public. The suspended mass of a component at any given time can be released directly to the environment if the containment fails catastrophically. Since aerosols produced in-vessel generally contain the more volatile and more radiotoxic materials, the time for deposition of the first component affects the radiotoxicity of the materials suspended in the containment. The time-integrated concentration of a component provides information about the overall importance of a parameter in the accident sequence independent of time and about the amount of material that might be released through a leak path.

3. MODELING APPROACH. The modeling techniques used in this analysis for the evolution of a spatially homogeneous, multicomponent aerosol were developed by Gelbard and Seinfeld (1980). These techniques are based on dividing the particle size domain into a number of sections and requiring conservation of mass for each chemical component for the physical processes resulting in particle movement into, between and out of the sections. For this analysis, the physical processes considered are coagulation due to Brownian motion, gravity and turbulence, and deposition due to thermophoresis, gravity and diffusion. Due to the resulting simplicity in the differential equations describing particle evolution, a geometric constraint is imposed on particle mass at the section boundaries (i.e., $v_{i+1} \geq 2v_i$, where v_i is the largest particle mass allowed in section i). Specifically, when m sections and s components are considered, the following system of differential equations is assumed to describe aerosol behavior:

$$dQ_{\ell k}/dt = f_{\ell k}(Q,t) \tag{3.1}$$

$$= \sum_{i=1}^{\ell-2} \left({}^{2a}\beta_{i,\ell-1}\, Q_{\ell-1,k} - {}^{2a}\beta_{i\ell}\, Q_{\ell k} \right) Q_i \tag{3.2}$$

$$+ \sum_{i=1}^{\ell-2} \left({}^{1b}\beta_{i,\ell-1,\ell}\, Q_{\ell-1} + {}^{2b}\beta_{i\ell}\, Q_{\ell} \right) Q_{ik} \tag{3.3}$$

$$+\left(.5\,{}^{3}\beta_{\ell-1,\,\ell-1}\,Q_{\ell-1,k} - {}^{2a}\beta_{\ell-1,\ell}\,Q_{\ell k}\right)Q_{\ell-1} \qquad (3.4)$$

$$+ \;{}^{2b}\beta_{\ell-1,\,\ell}\,Q_{\ell}\,Q_{\ell-1,\,k} \qquad (3.5)$$

$$-\;.5\,{}^{3}\beta_{\ell\ell}\,Q_{\ell}\,Q_{\ell k} \qquad (3.6)$$

$$-\,Q_{\ell k}\sum_{i=\ell+1}^{m}{}^{4}\beta_{i\ell}\;Q_{i} \qquad (3.7)$$

$$-\;(\rho_{\ell 1}+\rho_{\ell 2}+\rho_{\ell 3})Q_{\ell k} \qquad (3.8)$$

$$+\;SR_{\ell k}\;. \qquad (3.9)$$

where $Q_{\ell k}$ denotes the concentration of component k in section ℓ (kg/m^3), Q_{ℓ} denotes the total concentration in section ℓ (kg/m^3), and the β's and ρ's denote coefficients which are defined in the next paragraph. The terms in (3.2) through (3.9) have the following correspondences:

(3.2), (3.3) - Coagulation of particles in section i with particles in sections $\ell-1$ and ℓ, where i varies from 1 to $\ell-2$.

(3.4), (3.5) - Coagulation of particles in sections ℓ and $\ell-1$.

(3.6) - Intra-sectional coagulation loss.

(3.7) - Removal from section due to coagulation with particles in higher sections.

(3.8) - Removal due to deposition on surfaces, where 1, 2 and 3 correspond to ceiling, walls and floor, respectively.

(3.9) - Source rate for component k into section ℓ.

The ρ's and β's in (3.2) through (3.8) are derived in Gelbard and Seinfeld (1980) and are defined in terms of single and double integrals. These integrals are given in Table 3.1. The variables u and v in Table 3.1 represent particle mass, and the x and y are functions of u and v. In the MAEROS computer program, this function f is defined by f(u) = ln(u) (Gelbard, 1982). Due to the derivation of the model, this is equivalent to u being uniformly distributed with respect to ln(u) over each section. A geometric constraint on particle mass is assumed for the derivation of the coagulation coefficients in Table 3.1. The function $\beta(u,v)$ in Table 3.1 is actually the sum of functions $\beta_B(u,v)$, $\beta_G(u,v)$ and $\beta_T(u,v)$ corresponding to coagulation due to Brownian motion, gravity and turbulence; these functions are defined in Table 3.2. The functions $\rho_i(u)$ in Table 3.1 are constructed from functions $\rho_T(u)$, $\rho_G(u)$ and $\rho_D(u)$ corresponding to deposition due to thermophoresis, gravity and diffusion; these functions are defined in Table 3.3.

Additional background on aerosol behavior in the context of nuclear reactor safety can be obtained from the review by Loyalka (1983).

4. APPROACH TO UNCERTAINTY AND SENSITIVITY ANALYSIS. A model can be thought of as a function $F(X_1,...,X_k)$ of independent variables $X_1,...,X_k$. If $X_1,...,X_k$ have ranges and distributions rather than fixed values, then $F(X_1,...,X_k)$ will also have a range and distribution. Uncertainty analysis involves determination of properties of the distribution of $F(X_1,...,X_k)$. Properties of interest include the distribution function, the expected value and the variance for $F(X_1,...,X_k)$. Sensitivity analysis involves determination of the importance of individual X_i in influencing $F(X_1,...,X_k)$. Measures of such influence include partial correlation coefficients, standardized regression coefficients and normalized partial derivatives.

The analysis techniques used in this paper are based on Latin hypercube sampling and response surface construction. The techniques are intended for a model with specified inputs $X_1,...,X_k$ which have defined ranges, distributions and possibly correlations. The first step in an analysis is to generate a Latin hypercube sample from the X_i. This sample is then used in the generation of model predictions. Once generated, these predictions can be used in the estimation of distribution functions, expected values and variances. Further, these predictions can be analyzed to determine the relative importance of individual X_i in influencing their values. These techniques are now discussed in greater detail.

As originally described by McKay et al. (1979), Latin hypercube sampling operates in the following manner to generate a sample of size n from the k variables $X_1,...,X_k$. The range of each variable is divided into n nonoverlapping intervals of equal probability. One value from each interval is selected at random. The n values thus obtained for X_1 are paired at random with the n values obtained for X_2. These n pairs are combined in a random manner with the n values for X_3 to form n triples. The process is continued until a set of n k-tuples is formed. This resultant set of k-tuples is the Latin hypercube sample.

Due to the random pairing of variable values in the generation of the sample, there may exist spurious correlations between variable values within a Latin hypercube sample. This is most likely when n is small in relation to k. Such correlations

Table 3.1. Sectional Rate Coefficients with Geometric Constraint. This table is adapted from Tables 1 and 2 of Gelbard and Seinfeld (1980); the expressions given for ${}^3\beta$, ${}^4\beta$ and $\rho_{\ell,i}$ are valid even if the geometric constraint is not satisfied.

Symbol	Remarks	Coefficient[a]
${}^{1b}\beta_{i,\ell-1,\ell}$	$1 < \ell \le m$ $i < \ell - 1$	$\int_{x_{i-1}}^{x_i} \int_{f(v_{\ell-1}-v)}^{x_{\ell-1}} \frac{v\beta(u,v)}{uv(x_i-x_{i-1})(x_{\ell-1}-x_{\ell-2})}\,dydx$
${}^{2a}\beta_{i\ell}$	$1 < \ell \le m$ $1 \le i < \ell$	$\int_{x_{i-1}}^{x_i} \int_{f(v_\ell-v)}^{x_\ell} \frac{u\beta(u,v)}{uv(x_i-x_{i-1})(x_\ell-x_{\ell-1})}\,dydx$
${}^{2b}\beta_{i\ell}$	$1 < \ell \le m$ $1 \le i < \ell$	$\int_{x_{i-1}}^{x_i} \int_{x_{\ell-1}}^{f(v_\ell-v)} \frac{v\beta(u,v)}{uv(x_i-x_{i-1})(x_\ell-x_{\ell-1})}\,dydx$
${}^{3}\beta_{\ell\ell}$	$1 \le \ell \le m$	$\left\{ \int_{x_{\ell-1}}^{f(v_\ell-v_{\ell-1})} \int_{f(v_\ell-v)}^{x_\ell} \frac{(u+v)\beta(u,v)}{uv(x_\ell-x_{\ell-1})^2}\,dydx + \int_{f(v_\ell-v_{\ell-1})}^{x_\ell} \int_{x_{\ell-1}}^{x_\ell} \frac{(u+v)\beta(u,v)}{uv(x_\ell-x_{\ell-1})^2}\,dydx \right.$
${}^{4}\beta_{i\ell}$	$1 \le \ell < m$ $1 < i \le m$	$\int_{x_{i-1}}^{x_i} \int_{x_{\ell-1}}^{x_\ell} \frac{u\beta(u,v)}{uv(x_i-x_{i-1})(x_\ell-x_{\ell-1})}\,dydx$
$\rho_{\ell i}$	$1 \le \ell \le m$ $1 \le i \le 3$	$\int_{x_{\ell-1}}^{x_\ell} \frac{\rho_i(v)}{(x_\ell-x_{\ell-1})}\,dx$

[a] $x_i = f(v_i)$, $u = f^{-1}(y)$, $v = f^{-1}(x)$.

Table 3.2. Coagulation Functions (m^3/sec), where β_B ~ Brownian motion, β_G ~ Gravity and β_T ~ Turbulence (adapted from Table A.I of Gelbard (1982) and the associated MAEROS computer program).

$$\beta_B(u,v) = 2\pi (\mathcal{D}_u + \mathcal{D}_v)(D_u + D_v)\ \gamma / F_{uv}$$

$$\beta_G(u,v) = \pi \epsilon_{uv} (D_u + D_v)^2\ \gamma^2\ |V_{Tu} - V_{Tv}|\ ST/4$$

$$\beta_T(u,v) = (\beta_{T1}^2 + \beta_{T2}^2)^{1/2}\ ST$$

where

$\mathcal{D}_u$	$= kTB_u$	(diffusion constant, m^2/J sec)
D_u	$= (6u/\pi\rho)^{1/3}$	(particle diameter, m)
F_{uv}	$= (D_u + D_v)/(D_u + D_v + 2g_{uv}) + 8(\mathcal{D}_u + \mathcal{D}_v)/\bar{V}_{uv}(D_u + D_v)ST$	(non-isotopic gas correction)
k	$= 1.381 \times 10^{-23}$	(Boltzmann's Constant, J/°K)
B_u	$= C_u / 3\pi D_u \eta \chi$	(particle mobility, m^2/J sec)
C_u	$= 1. + Kn_u\ [1.37 + .4 \exp(-1.1/Kn_u)]$	(slip correction factor)
η	$= .0066164\ (.003661\ T)^{3/2}/(T + 114.)$	(gas viscosity, kg/m sec)
Kn_u	$= 2\lambda / D_u$	(Knudson number)
λ	$= (\eta/\rho_g)\ (1.89E\text{-}4\ MW/T)^{1/2}$	(mean free path, m)
ρ_g	$= 1.21E\text{-}4\ P\ MW/T$	(gas density, kg/m^3)
g_{uv}	$= (g_u^2 + g_v^2)^{1/2}$	
$\bar{V}_{uv}$	$= (V_u^2 + V_v^2)^{1/2}$	
g_u	$= [(D_u + \ell_u)^3 - (D_u^2 + \ell_u^2)]^{3/2}/(3D_u\ell_u) - D_u$	
ℓ_u	$= 8\mathcal{D}_u / \pi V_u$	
V_u	$= (8kT/\pi u)^{1/2}$	(mean velocity, m/sec)
ϵ_{uv}	$= 1.5\ [Min(D_u, D_v)/(D_u + D_v)]^2$	
V_{Tu}	$= \rho D_u^2 C_u g / 18\eta\chi$	(sedimentation velocity, m/sec)
g	$= 9.8$	(gravitational constant, m/sec^2)
β_{T1}	$= (\pi\epsilon\rho_g / 120\eta)^{1/2}\ \gamma^3\ (D_u + D_v)^3$	(turbulent shear kernel, m3/sec)
β_{T2}	$= .074\ |V_{Tu} - V_{Tv}|\ \gamma^2\ (D_u + D_v)^2\ (\rho_g \epsilon^3/\eta)^{1/4}$	(turbulent inertial kernel, m^3/sec)

and all other symbols are defined in Table 2.1.

Table 3.3. Deposition Functions (1/sec), where ρ_T ~ Thermophoresis, ρ_G ~ Gravity, and ρ_D ~ Diffusion (adapted from Table A.II of Gelbard (1982) and the associated MAEROS computer program). The actual deposition functions for ceiling, walls and floor are given by $\rho_1(u)$ = Max (0, $\rho_T-\rho_G+\rho_D$), $\rho_2(u)$ = Max (0, $\rho_T+\rho_D$), and $\rho_3(u)$ = Max (0, $\rho_T+\rho_G+\rho_D$), respectively.

$$\rho_T(u) = (A/V)\, 3\eta C_u (C_T Kn_u+\kappa)\, \nabla T$$

$$\cdot\{2x\rho_g T[1. + 3\cdot 1.37 Kn_u \ldots + 2(C_T Kn_u + \kappa)]\}^{-1}$$

$$\rho_G(u) = (A/V)\, V_{Tu}$$

$$\rho_D(u) = (A/V)\mathcal{D}_u/\delta$$

where A denotes the area of the particular surface under consideration, V denotes the volume of the containment, A/V = RCV for movements to the ceiling or floor, A/V = RWV for movements to the walls, and all other symbols are defined in Table 2.1 or Table 3.2.

---

can be avoided by modifying the generation of the sample through use of a technique introduced by Iman and Conover (1982a) and implemented by a computer program developed at Sandia National Laboratories (Iman and Shortencarier, 1984). This technique preserves the fundamental nature of Latin hypercube sampling but replaces the random pairing of variable values with a pairing that keeps all of the pairwise rank correlations among the k variables close to zero. The technique can also be used to induce a desired rank correlation structure among the variables.

Uncertainty analysis in conjunction with Latin hypercube sampling is now considered. In such an analysis, it is desired to estimate the distribution function, the expected value and the variance for the dependent variable under consideration. Due to the probabilistic nature of Latin hypercube sampling, it is possible to estimate these entities directly from the model predictions associated with the sample. For convenience in the following discussion, $(X_{1i},\ldots,X_{ki}, Y_i)$, $i = 1,\ldots,n$, is used to denote the individual sample elements and the corresponding predictions Y_i. As each sample element has probability $1/n$, an estimate of the cumulative distribution function for Y can be obtained directly from the values for the individual Y_i. Further, the expected value and variance for Y can be estimated by

$$\hat{E}(Y) = \sum_{i=1}^{n} Y_i/n \qquad (4.1)$$

and

$$\hat{V}(Y) = \sum_{i=1}^{n} [Y_i-\hat{E}(Y)]^2/n \ , \qquad (4.2)$$

respectively. As discussed by McKay et al. (1979), the preceding provides unbiased estimates for the distribution function and expected value associated with Y when Latin hypercube sampling as originally defined is used. A small bias may be introduced when the restricted sampling technique of Iman and Conover (1982a) is used. The estimator $\hat{V}(Y)$ of $V(Y)$ in (4.2) for variance is biased; the exact amount of bias is small but unknown. Iman and Conover (1980) have shown that

$$[(n-1)/n]V(Y) \leq E[\hat{V}(Y)] \leq V(Y) \tag{4.3}$$

when Y is a monotonic function of each of the X_i.

Sensitivity analysis in conjunction with Latin hypercube sampling is based on the construction of regression models. The observations $(X_{1i},\ldots,X_{ki},Y_i)$, $i=1,\ldots,n$, are used to construct models of the form

$$\hat{Y} = b_0 + \sum_q b_q Z_q \tag{4.4}$$

subject to the constraint that $\Sigma_i |Y_i-\hat{Y}_i|^2$ be minimized. In the preceding, b_0 and the b_q are constants, and each Z_q is a function of $X_1,\ldots,X_k$. One such commonly used model is

$$\hat{Y} = b_0 + \sum_{i=1}^{k} b_i X_i + \sum_{\substack{i,j=1 \\ i \leq j}}^{k} b_{ij} X_i X_j \quad , \tag{4.5}$$

where the X_i may be the original independent variables or some transform of them. Further, regression is often performed in a stepwise fashion in which a sequence of regression models is constructed by adding one independent variable at a time. In turn, a regression of the form appearing in (4.4) can be reformulated as

$$(\hat{Y}-\bar{Y})/\hat{s} = \sum_q (b_q \hat{s}_q/\hat{s}) \; (Z_q-\bar{Z}_q)/\hat{s}_q \quad , \tag{4.6}$$

where $\bar{Y} = \Sigma_i Y_i/n$, $\bar{Z}_q = \Sigma_i Z_{qi}/n$, $s = [\Sigma_i (Y_i-\bar{Y})^2/(n-1)]^{1/2}$ and $\hat{s}_q = [\Sigma_i (Z_{qi}-\bar{Z}_q)^2/(n-1)]^{1/2}$. The quotients $b_q \hat{s}_q/\hat{s}$ appearing in (4.6) are called standardized regression coefficients. A ranking of variable importance can be based on an ordering of their absolute values.

An important property of least-squares regression is that

$$\Sigma_i (Y_i-\bar{Y})^2 = \Sigma_i (\hat{Y}_i-\bar{Y})^2 + \Sigma_i (Y_i-\hat{Y}_i)^2 \quad . \tag{4.7}$$

The R^2 value for a regression falls between 0 and 1 and is defined by

$$R^2 = \sum_i (\hat{Y}_i - \bar{Y})^2 / \sum_i (Y_i - \bar{Y})^2 \quad . \tag{4.8}$$

As indicated by the equality in (4.7), the closeness of an R^2 value to 1 provides an indication of how successful the regression model is in accounting for the variation in Y.

For a regression model of the form

$$\hat{Y} = b_0 + b_1 Z \tag{4.9}$$

with an R^2 value of r^2, the number sign $(b_1)\ |r|$ is called the correlation coefficient between Y and Z, where $\text{sign}(b_1) = 1$ if $b_1 \geq 0$ and $\text{sign}(b_1) = -1$ if $b_1 < 0$. This number provides a measure of the linear relationship between these two variables. When more than one independent variable is under consideration, partial correlation coefficients are used to provide a measure of the linear relationships between Y and the individual independent variables. The partial correlation coefficient between Y and an individual variable Z_p is obtained from the use of a sequence of regression models. First, the following two regression models are constructed:

$$\hat{Y}' = a_0 + \sum_{q \neq p} a_q Z_q \quad \text{and} \quad \hat{Z}_p' = c_0 + \sum_{q \neq p} c_q Z_q \quad . \tag{4.10}$$

Then, the results of the two preceding regressions are used to define the new variables $Y - \hat{Y}'$ and $Z_p - \hat{Z}_p'$. By definition, the partial correlation coefficient between Y and Z_p is the correlation coefficient between $Y - \hat{Y}'$ and $Z_p - \hat{Z}_p'$. Thus, the partial correlation coefficient provides a measure of the linear relationship between Y and Z_p with the linear effects of the other variables removed.

When nonlinear relationships are involved, it is often more revealing to calculate partial correlation coefficients between variable ranks than between the actual values for the variables; such coefficients are known as partial rank correlation coefficients. Specifically, the smallest value of each variable is assigned the rank 1, the next largest value is assigned the rank 2, and so on up to the largest value which is assigned the rank n, where n denotes the number of observations. The partial correlations are then calculated on these ranks.

Stepwise regression, standardized regression coefficients and partial rank correlation coefficients will be used in the following sensitivity analysis. Additional discussion of these techniques can be found in Iman, Helton and Campbell (1981a, 1981b), Iman and Conover (1980, 1982b), and Iman and Helton (1985).

5. THE ANALYSIS. The analysis was performed for a Latin hypercube sample of size 50 generated from the 21 variables listed in Table 2.1. The construction technique described by Iman and Conover (1982a) was used to preclude spurious correlations within the sample and to induce the rank correlations between variables indicated in Table 2.1. The preceding construction requires a sample of size $n > k$, where k is the number of independent variables under consideration. Our experience has been that it is often sufficient to select $n \geq (4/3)k$, which would suggest a sample of size 28 for this analysis. However, as the model was not unduly expensive to evaluate, the decision was made to use the somewhat larger sample size of 50.

The Latin hypercube sample described in the preceding paragraph was generated without regard to the specified inequalities in Table 2.1. Once this sample was generated, the sample actually used in the analysis was obtained by considering the sample vectors individually and linearly mapping the range (and hence the particular sampled value) of each restricted variable (i.e., $T < .001\ P$ and $\chi \leq \gamma$) into its restricted range. Specifically, if the i^{th} sample vector contained the values of T_i, P_i, χ_i and γ_i for T, P, χ and γ, respectively, then the actual values P_i^* and γ_i^* used for P_i and γ_i in the analysis were defined by

$$P_i^* = 1000.\ T_i + (P_i - 3.75E5)(6.E5 - 1000.\ T_i)/4.25E5 \qquad (5.1)$$

and

$$\gamma_i^* = \chi_i + (\gamma_i - 1.)(3. - \chi_i)/2 \quad . \qquad (5.2)$$

Figures 5.1 through 5.5 provide summaries of the predictions associated with the resultant sample. Figures 5.1 and 5.2 display the total suspended mass of the two components as a function of time. Figure 5.3 contains the distribution functions for the first component for the times to 90 percent and 99 percent deposition. Finally, Figures 5.4 and 5.5 contain the distribution functions for the integrated concentrations from t = 0 sec to t = 72,000 sec for the first and second components, respectively. To obtain these predictions, 10 size sections were used, and the lower and upper section boundaries were taken to be .01E-6 m and 100.E-6 m, respectively. The analyses (including partial correlation and regression analyses shown later) were also repeated for 5, 15 and 20 sections; the results were essentially identical for 10, 15 and 20 sections and only slightly different for 5 sections. The effects of section definition are considered in greater detail in Section 6.

An important property of analyses involving Latin hypercube sampling is that it is possible to generate scatterplots in which a dependent variable appears on one axis and an independent variable appears on the other axis. Such a plot is given in Figure 5.6. Scatterplots can be very useful because they may reveal unexpected relationships between variables and thus provide guidance with respect to how an analysis might be performed. Latin hypercube sampling is particularly effective in providing such guidance because it forces the full range of each variable to be sampled. In this particular analysis, examination of scatterplots did not reveal any unexpected behavior. However, Iman and Helton (1985) give an example of an

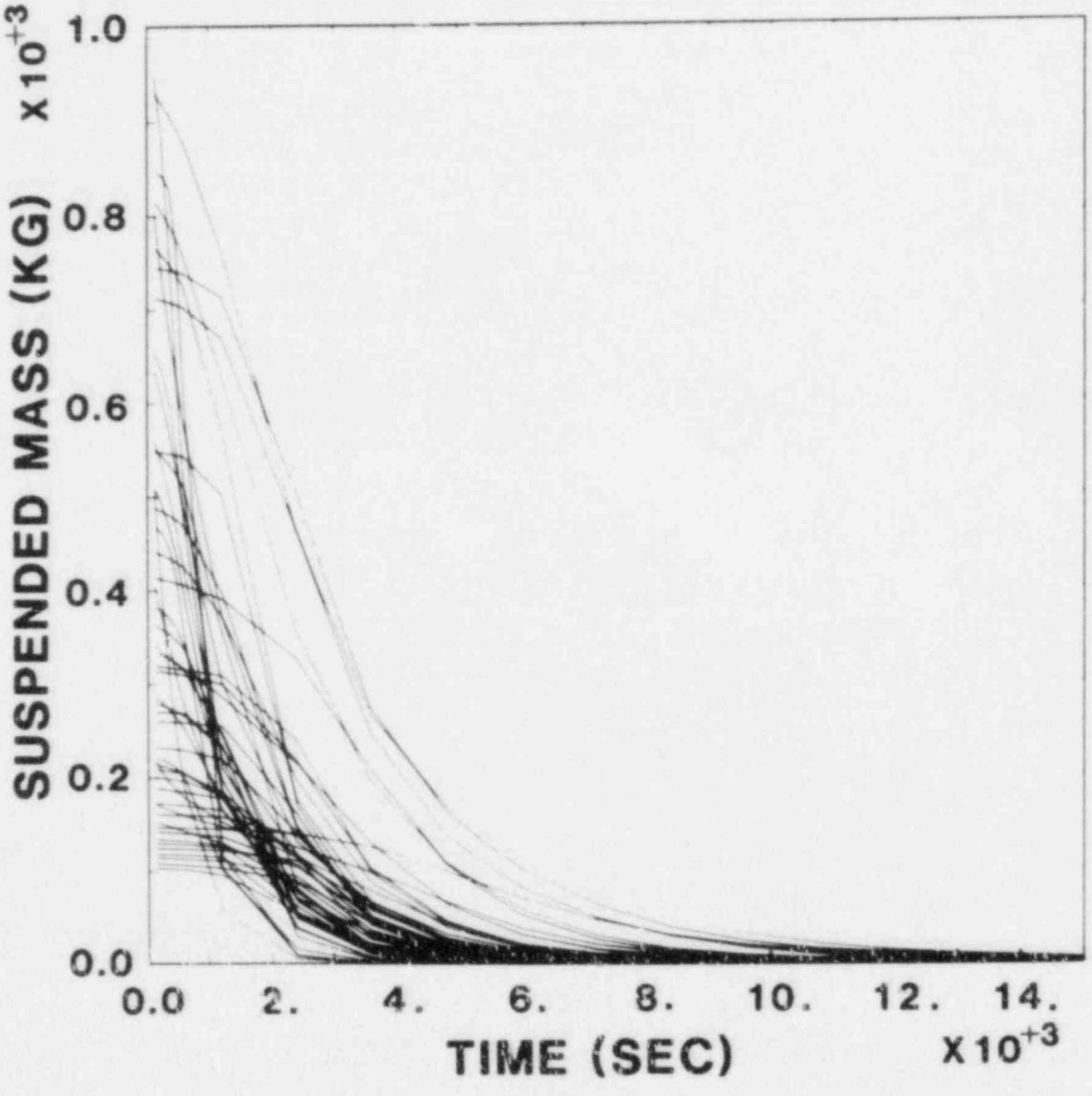

Figure 5.1. Total Suspended Mass (kg) of First Component as a Function of Time (sec). Each line represents the prediction associated with one sample vector.

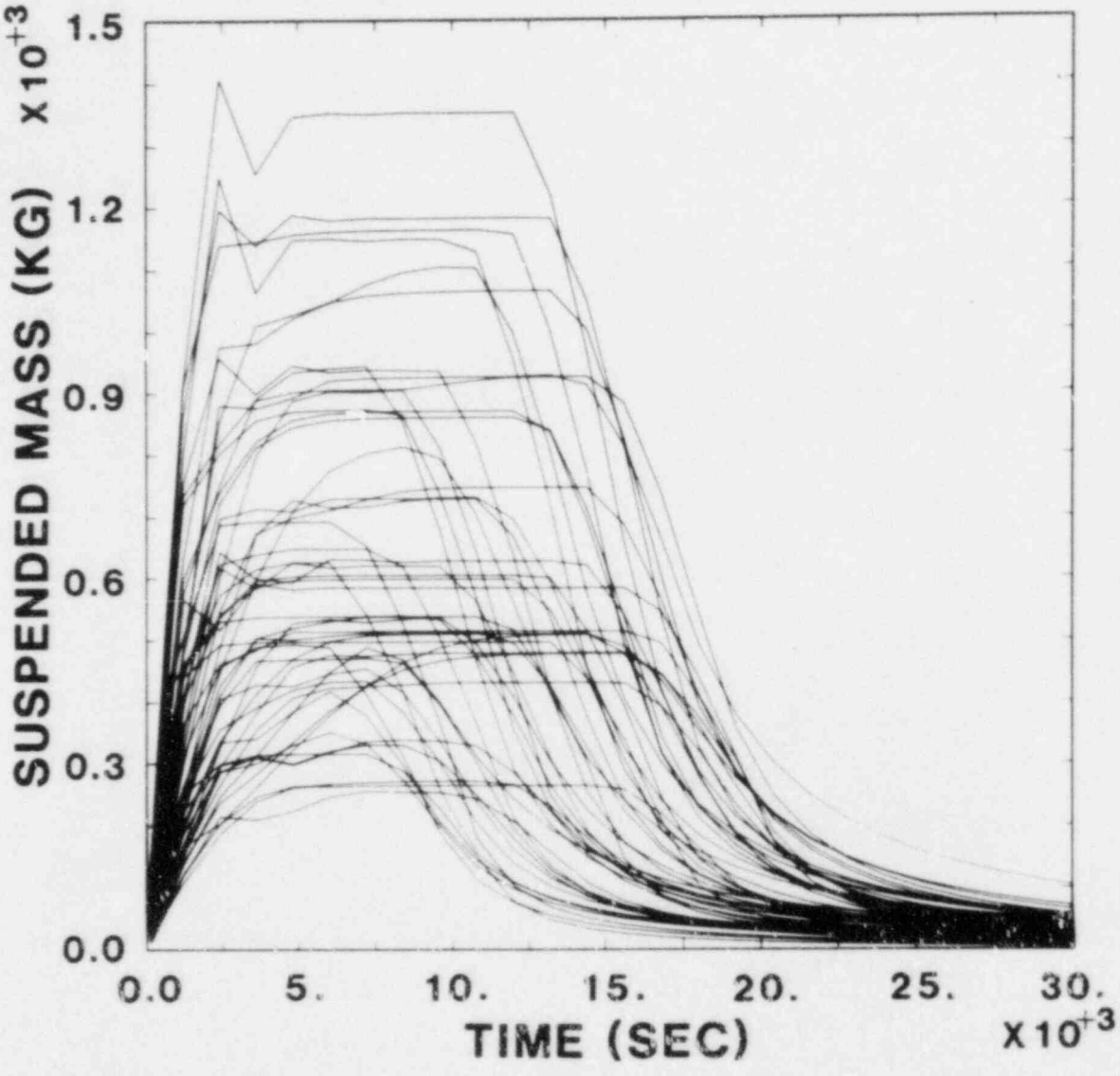

Figure 5.2. Total Suspended Mass (kg) of Second Component as a Function of Time (sec). Each line represents the prediction associated with one sample vector.

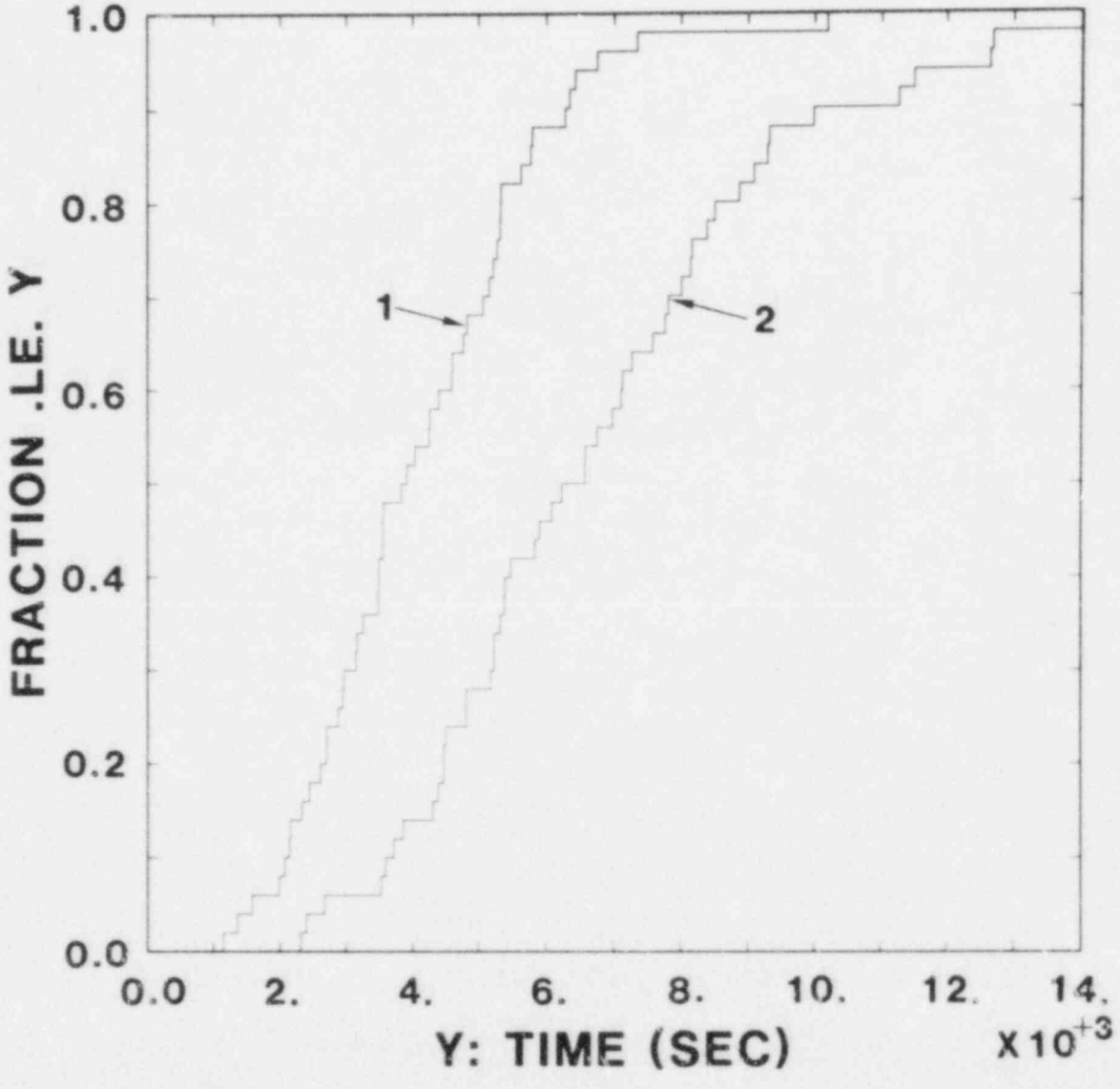

Figure 5.3. Distribution Functions for Time to 90 Percent and 99 Percent Deposition for the First Component (1 ~ Time to 90 percent deposition, 2 ~ Time to 99 percent deposition).

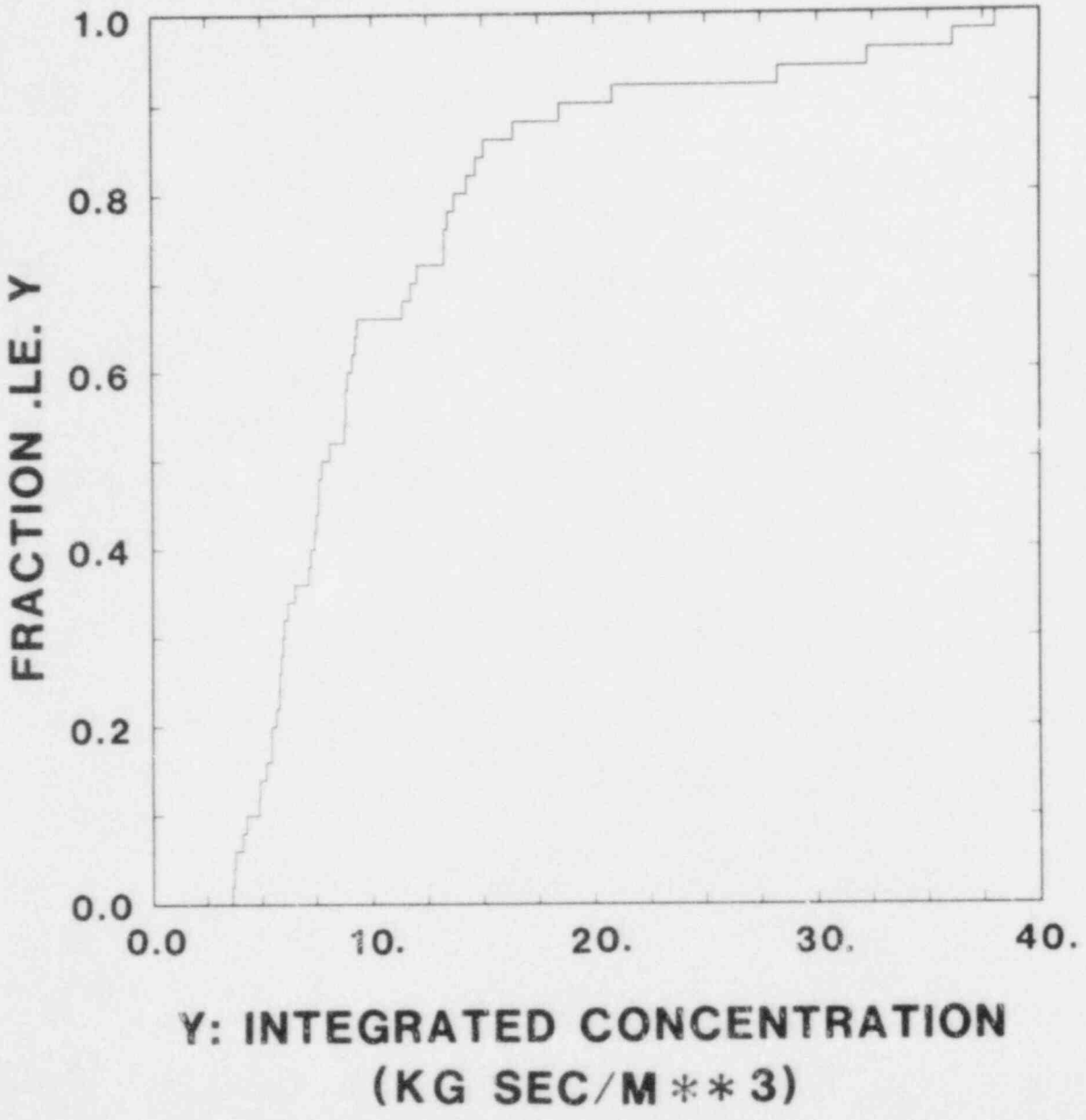

Figure 5.4. Distribution Function for Total Integrated Concentration (kg sec/m^3) of First Component.

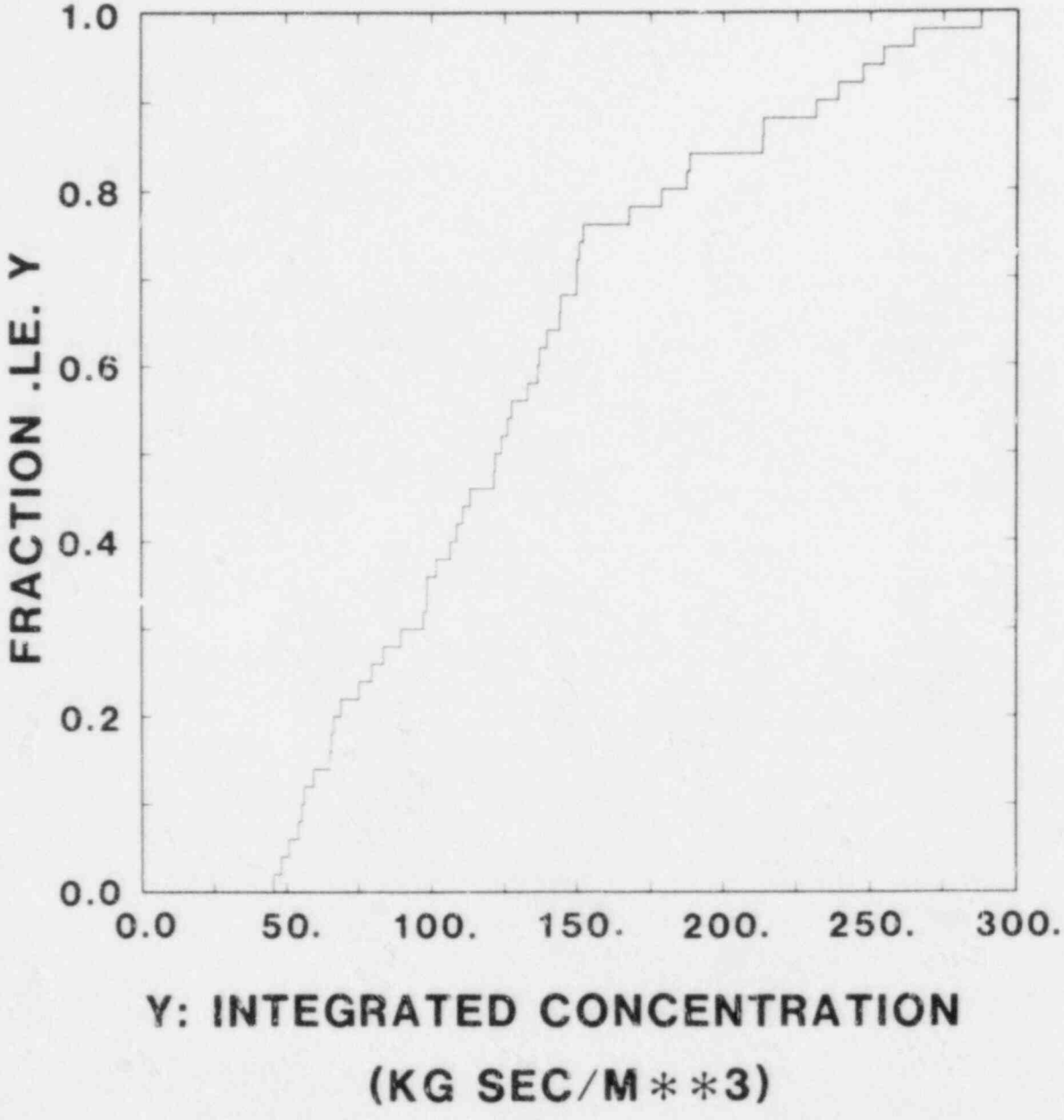

Figure 5.5. Distribution Function for Total Integrated Concentration (kg sec/m^3) of Second Component.

analysis in which examination of scatterplots provided important insights with respect to the behavior of the model under consideration.

As indicated earlier, uncertainty analysis is taken to include estimation of distribution functions, expected values and variances. Four such distribution functions are shown in Figures 5.3, 5.4 and 5.5. Further, expected values and variances can be estimated as indicated in (4.1) and (4.2). Such estimates are shown in Table 5.1 for the dependent variables in Figures 5.3, 5.4 and 5.5. If desired, estimates for distribution functions, expected values and variances could also have been made for the total suspended mass of each component at various times and also for other dependent variables such as mass median particle diameter, geometric standard deviation of particle diameter, deposition on various surfaces, and so on. The estimated distribution function provides a more useful and robust display of model variability than estimated expected value and variance. However, it should be recognized that such estimates are conditional on the distributions and restrictions assumed with respect to the independent variables (see Table 2.1).

Perhaps the most interesting aspect of the results shown in Figures 5.1 through 5.6 is not how much, but rather, how little variability is shown. A total of twenty-one independent variables was considered, with several variables being given ranges covering one or more orders of magnitude. Thus, it might be anticipated that the resultant dependent variables would vary over many orders of magnitude. In contrast, the observed variability was generally less than one order of magnitude. This is encouraging with respect to being able to model aerosol behavior in reactor containments.

Once the variability in model predictions is determined, the next step is to perform a sensitivity analysis to determine the importance of individual variables in influencing dependent variables of interest. For the dependent variables appearing in Figures 5.1 and 5.2 (i.e., time dependent suspended mass), this determination will be made on the basis of partial rank correlation coefficients. For the dependent variables appearing in Figures 5.3, 5.4 and 5.5 (i.e., times to 90 and 99 percent deposition and integrated concentrations), this determination will be made through the use of stepwise regression.

The analyses for total suspended mass are presented first. As these variables are functions of time, it is informative to give graphs of their partial rank correlation coefficients as functions of time. Partial rank correlations are used because they help to remove the effects of nonlinear relationships between independent and dependent variables. The correlations are presented in Figures 5.7 and 5.8. As a means of including only the more important variables, only those variables which have a correlation whose absolute value is greater than or equal to 0.75 at some point in time are included in these figures. Table 5.2 gives the calculated partial rank correlation coefficients for the total suspended mass of the first component at various points in time. This table also contains a ranking of variable importance for those variables with coefficients greater than or equal to 0.4 in absolute value; coefficients closer to zero suggest little significance for the associated variables. The preceding partial correlation calculations were performed with a computer program developed at Sandia National Laboratories (Iman et al., 1985). As indicated earlier, the partial rank correlation coefficient provides a measure of the tendency of two variables to increase or decrease together after a correction has been made to remove the effects of other variables.

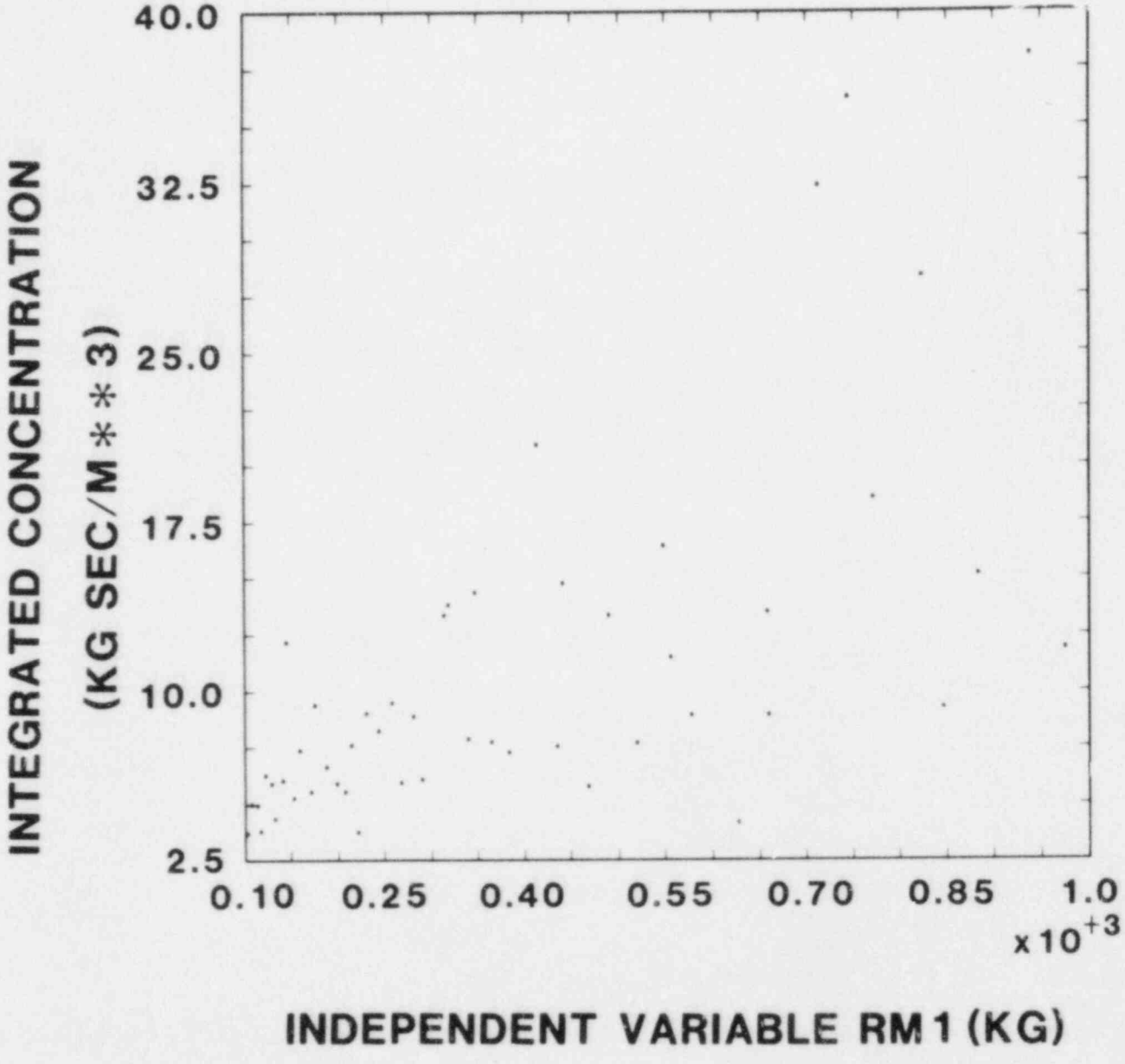

Figure 5.6. Scatterplot for Total Integrated Concentration (kg sec/m^3) of First Component and Amount (kg) of First Component Initially Present.

Table 5.1. Estimated Expected Values and Variances.

	Expected Value	Variance	Standard Deviation
Time to 90% Deposition of First Component (sec)	4110.	2.96E6	1720.
Time to 99% Deposition of First Component (sec)	6760.	7.70E6	2770.
Integrated Concentration for First Component ($kg\ sec/m^3$)	10.8	63.3	8.0
Integrated Concentration for Second Component ($kg\ sec/m^3$)	130.	3960.	62.9

As examination of Figure 5.7 and Table 5.2 shows, total suspended mass of the first component is dominated at the outset by RM1 (total released mass for first component). Thereafter, variables which affect agglomeration dominate. Specifically, SR2 (mass source rate for second component), γ (agglomeration shape factor) and ε (turbulence dissipation rate) have negative influences on the suspended mass of the first component while ρ (particle density) has a positive influence. Increasing SR2 increases the amount of material available for agglomeration while increasing γ and ε increases the rate constants for agglomeration. Examination of Table 3.2 suggests that the positive influence of ρ is due to its effect on decreasing particle diameter D_u. As shown in Figure 5.8, total suspended mass of the second component is initially dominated by SR2 (mass source rate for second component). The effect of release duration for the second component (RD2) can be seen by the corresponding changes in importance for SR2 and RD2. As for the first component, both γ and ε have strong negative influences. However, χ (dynamic shape factor) has a positive influence. Although not shown, ρ had a partial rank correlation coefficient of approximately 0.6. Thus, as with the first component, total suspended mass of the second component appears to be dominated by source rate and by factors affecting agglomeration.

The analyses for time to 90 percent and 99 percent deposition for the first component and for the integrated concentrations are now presented. These analyses were performed with stepwise regression. The set of candidate predictor variables consisted of the original 21 variables appearing in Table 2.1 plus their squares. For each regression, the logarithm of the dependent variable and of each logarithmically distributed independent variable is used instead of the original variable; this procedure helps to reduce the effects of nonlinearities. Further, each variable, or its logarithm as appropriate, is normalized to a new variable with mean zero and standard deviation one; this normalization helps to reduce correlations between

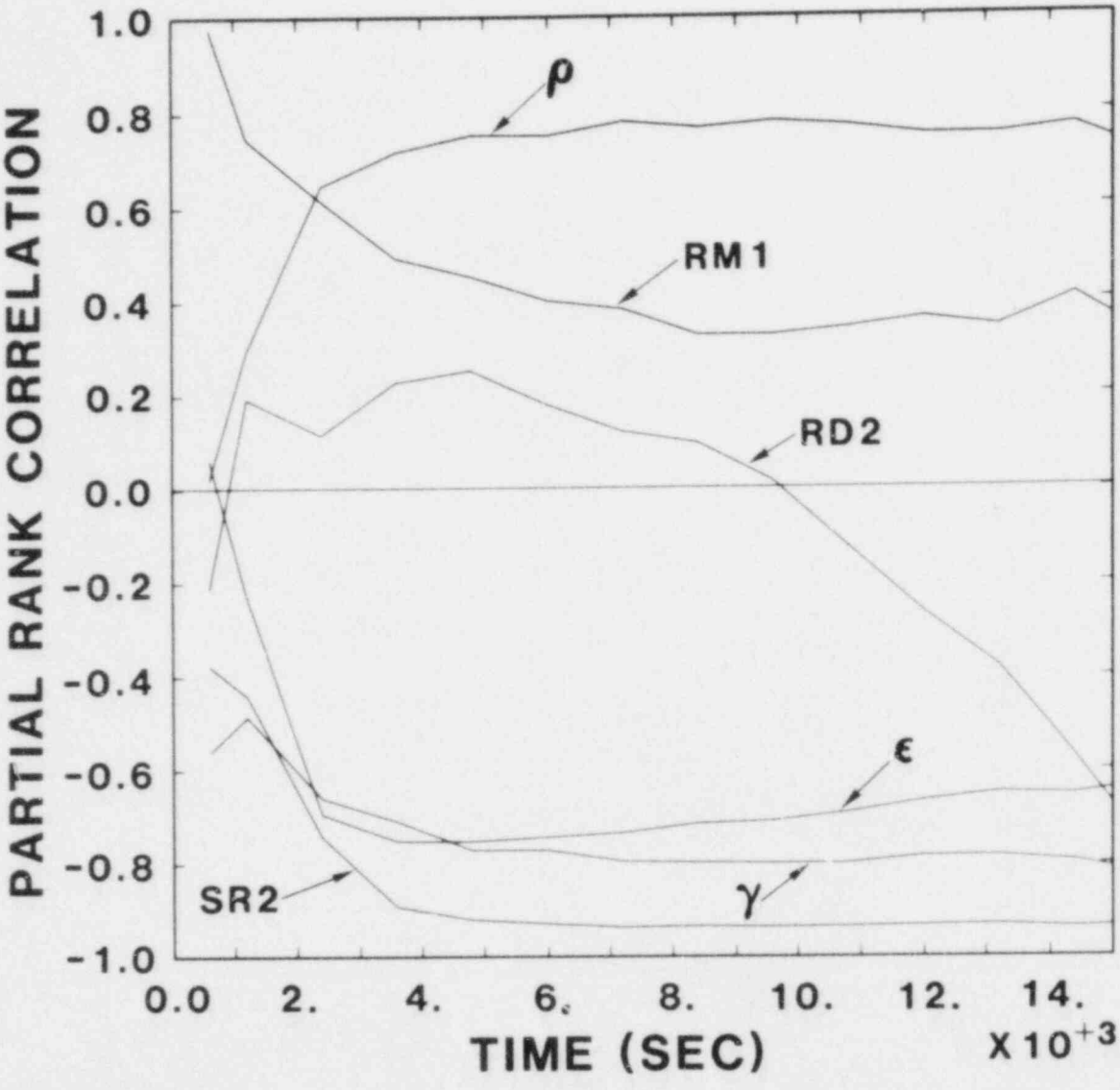

Figure 5.7. Partial Rank Correlation Coefficients for Total Suspended Mass of First Component.

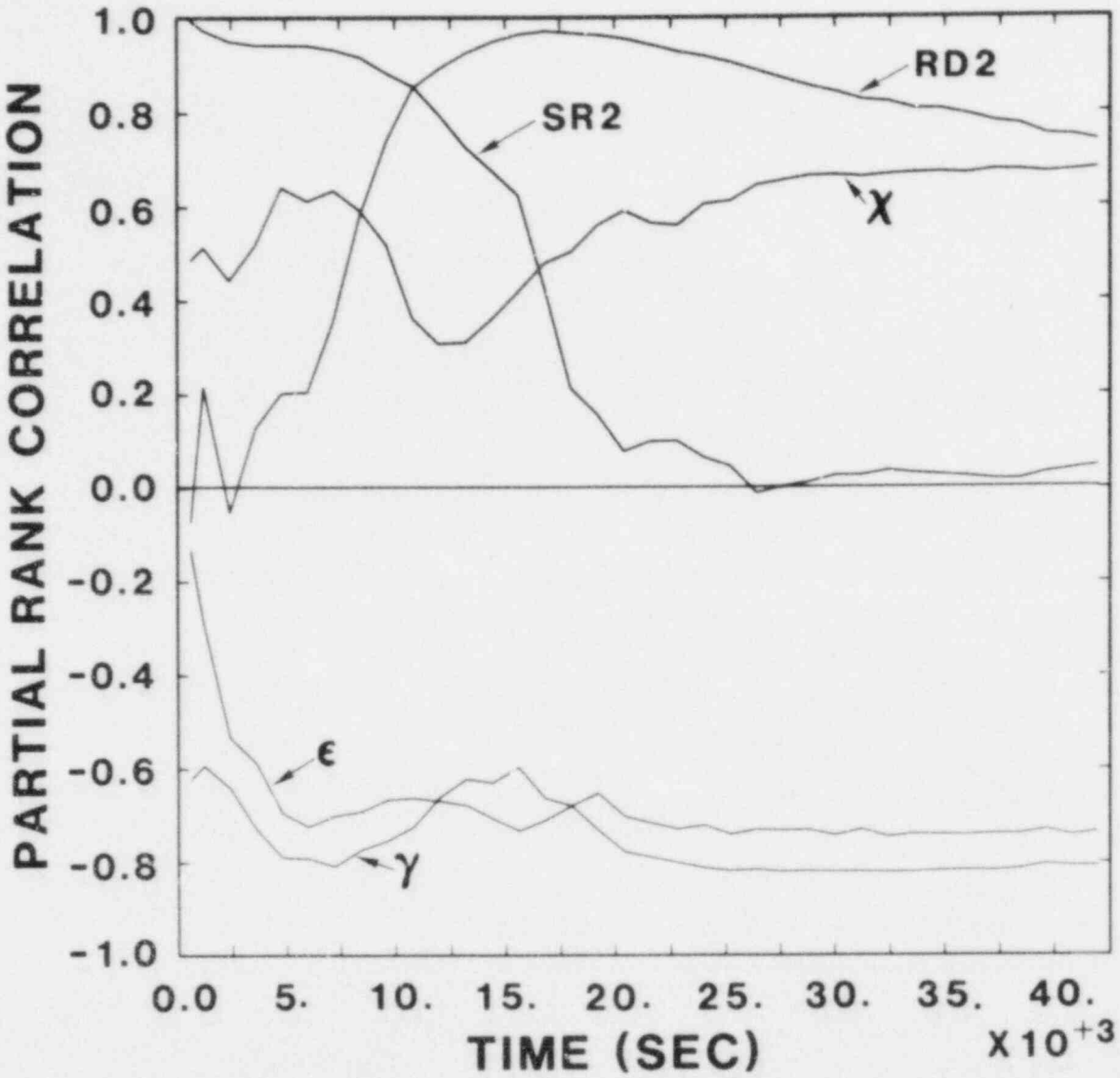

Figure 5.8. Partial Rank Correlation Coefficients for Total Suspended Mass of Second Component.

Table 5.2. Partial Rank Correlation Coefficients for Total Suspended Mass of First Component at Selected Times.

Var	20 min	40 min	60 min	120 min
GSD1	-.245	-.210	-.041	.013
MMD1	-.211	-.578 6	-.547 6	-.490 6
RM1	.773 1[a]	.630 5	.529 7	.414 8
GSD2	-.127	.297	.386	.444 7
MMD2	-.004	-.059	.128	.295
SR2	-.428 3	-.762 1	-.891 1	-.937 1
RD2	.142	.138	.221	.184
T	.196	.436 8	.368	.283
P	.033	-.174	-.126	-.116
RCV	-.219	.025	.001	-.006
RWV	-.142	-.174	-.151	-.114
χ	.358	.335	.351	.361
δ	-.020	.011	.135	.086
ρ	.248	.640 4	.714 4	.765 4
C_T	-.192	-.201	.055	.107
γ	-.507 2	-.671 3	-.720 3	-.786 2
ST	-.121	-.454 7	-.623 5	-.596 5
∇T	-.100	.107	.096	.133
κ	-.162	-.226	-.190	-.104
ϵ	-.254	-.712 2	-.765 2	-.767 3
MW	.005	-.099	-.176	-.275

[a] Variable rank based on absolute value of partial rank correlation coefficient

variables and their squares. The stepwise regression was then performed on these variables with a program available at Sandia National Laboratories (Iman et al., 1980); in this regression, a variable had to be significant at the .01 α-level to enter into a regression model and remain significant at the .05 α-level to stay in a regression model. This generated a sequence of regression models for each dependent variable. These models are summarized in Tables 5.3 and 5.4. A considerable amount of discretion was used in selecting the stopping point in each stepwise regression. This decision was made on the basis of the previously indicated α-levels, the leveling off of R^2 values and the behavior of PRESS values. The PRESS value (Allen, 1971) is used to indicate if a regression model is overfitting the data on which it is based.

The analyses for times to 90 percent and 99 percent deposition of the first component in Table 5.3 indicate that these times are dominated by agglomeration-related variables. For example, SR2 (source rate of second component) and γ (agglomeration shape factor) are the first two variables selected in both analyses and also have the largest standardized regression coefficients. Further, most of the other variables in the regression analyses are related to agglomeration. It is interesting to consider the effects indicated for RM1 (total released mass for first component) by the e analyses. First, as shown by the signs of the standardized regression coefficients, increasing RM1 decreases the times to 90 percent and 99

Table 5.3. Stepwise Regression Analyses for Time to 90 percent and 99 percent Deposition of First Component.

	Time 90% Deposition			Time 99% Deposition		
Step	Var [a]	SRC [b]	R^{2c}	Var [a]	SRC [b]	R^{2c}
1	SR2	-.612	.311	SR2	-.700	.405
2	γ	-.595	.503	γ	-.607	.615
3	RM1	-.325	.647	ϵ	-.335	.743
4	ϵ	-.347	.787	ρ	.257	.806
5	ρ	.254	.848	MMD1	-.140	.862
6	χ	.252	.901	ST	-.214	.917
7	ST	-.210	.944	χ	.247	.951
8	MMD1	-.168	.968	RM1	-.189	.974
9	ϵ^2	.106	.978	ϵ^2	.105	.984

[a] Variables listed in order in which they entered the regression.

[b] Standardized regression coefficients for variables in final regression model.

[c] R^2 value for regression model with entry of each variable.

percent deposition. This is because larger values for RM1 result in (1) larger particles (i.e., RM1 has a .5 rank correlation with MMD1), (2) more rapid agglomeration due to higher particle densities, and (3) larger required concentrations of the first component at the specified percentage deposition times. In contrast, if the dependent variables were times to reach specified concentration levels that were independent of RM1, then the resultant regression coefficients for RM1 would probably be positive. Second, the effects indicated for RM1 tend to decrease with time. For time to 90 percent deposition, RM1 is third and fourth in importance based on entry into the regression model and size of standardized regression coefficient in the final regression model, respectively; for time to 99 percent deposition, RM1 is eighth in importance on the basis of either of these criteria. As the final regression models have large R^2 values (i.e., .978 and .984), these models are quite successful in accounting for the observed variation. Thus, independent variables not in the final regressions had little effect on the deposition times. For example, little importance is indicated for the distributional assumptions associated with the two components (i.e., GSD1, MMD1, GSD2, MMD2).

Table 5.4 Stepwise Regression Analyses for Integrated Concentration of the First and Second Component. Table organization is the same as for Table 5.3.

	Integrated Conc. First Comp.			Integrated Conc. Second Comp.		
Step	Var	SRC	R^2	Var	SRC	R^2
1	RM1	.786	.481	SR2	.630	.318
2	SR2	-.433	.613	RD2	.431	.597
3	γ	-.445	.737	ϵ	-.331	.708
4	ϵ	-.272	.814	γ	-.485	.818
5	MMD1	-.257	.871	χ	.236	.867
6	χ	.231	.907	MMD2	-.232	.899
7	ρ	.163	.935	ST	-.193	.932
8	ST	-.157	.960	ρ	.171	.965
9	$MMD1^2$	-.123	.974	$MMD2^2$	-.107	.977
10				ϵ^2	.098	.986

The analyses for integrated concentration in Table 5.4 indicate that the integrated concentration of each component is dominated by the amount of that component introduced into the system and by factors which affect agglomeration. Specifically, RM1 (total released mass for first component) is indicated as the most important independent variable with respect to the integrated concentration of the first component while SR2 (source rate for the second component) and RD2 (release duration for the second component) are indicated as the most important variables with respect to the integrated concentration of the second component. After these initial selections, both regressions bring in variables related to agglomeration (e.g., γ, ε, χ, ρ, ST). Small negative effects are indicated for the mass median diameters associated with the two components (i.e., MMD1 and MMD2); this is because increasing the mass median diameter tends to move mass to large particles which settle more rapidly. The high R^2 values (i.e., .974 and .986) indicate that the final regression models are successful in accounting for most of the observed variation in the integrated concentrations.

Table 5.5 contains the results of an additional sequence of regressions which was performed for comparison with the partial rank correlation analysis results contained in Table 5.2. Specifically, a regression model containing all 21 independent variables was constructed for each of the dependent variables in Table 5.2. The resultant standardized regression coefficients are presented in Table 5.5; further, the table also contains a ranking of variable importance for those variables with coefficients greater than or equal to .2 in absolute value. As each regression model was required to include all 21 independent variables, the regression coefficients close to zero probably have little significance. The variable rankings in Tables 5.2 and 5.5 are similar but not the same. As the ranking criteria are different, the two rankings will not necessarily be the same. Further, partial rank correlation coefficients and standardized regression coefficients are not directly comparable.

6. THE AGGREGATION PROBLEM. An important problem in model development is the degree of resolution that should be built into the model. Often, more accuracy can be achieved by making the model more detailed. The tradeoff is that the additional detail may significantly increase the expense of evaluating the model. The analysis of this tradeoff between model resolution and model complexity is known as the aggregation problem (e.g., Aoki, 1971; Ijiri, 1971; Golikeri and Luss, 1974; Zeigler, 1976; Gardner et al., 1982; Gibbons et al., 1982). Sensitivity analysis provides a way to investigate the impact of increasing model complexity. This can be illustrated by adding the three variables listed in Table 6.1 to the 21 variables listed in Table 2.1 and then repeating the analyses given in Section 5.

For purposes of comparison, the analyses leading to Tables 5.3 and 5.4 were repeated for the 21 variables appearing in Table 2.1 and the additional 3 variables appearing in Table 6.1. The resultant regression analyses are presented in Tables 6.2 and 6.3. These results and other unpresented analyses (i.e., partial rank correlation plots and additional regression models) indicate that the uncertainty induced by the modeling parameters in Table 6.1 is small in comparison to the uncertainty induced by the physical parameters in Table 2.1. For example, the variables in Table 6.1 either do not appear in individual regressions in Tables 6.2 and 6.3 or appear only near the end of the regressions.

Table 5.5. Standardized Regression Coefficients for Total Suspended Mass of First Component at Selected Times.

Var	20 min	40 min	60 min	120 min
GSD1	-.101	-.065	-.015	.013
MMD1	-.183	-.147	-.055	.008
RM1	.699 1[a]	.133	-.016	-.106
GSD2	-.116	-.014	.019	.050
MMD2	-.035	-.096	-.023	.047
SR2	-.334 3	-.565 2	-.667 1	-.740 1
RD2	.017	-.009	.016	.023
T	.103	.172	.152	.139
P	-.008	-.035	-.035	-.026
RCV	-.086	-.032	-.013	-.017
RWV	-.036	-.034	-.026	-.030
χ	.284 4	.317 4	.294 4	.246 4
δ	.015	.010	.036	.023
ρ	.119	.228 5	.227 5	.229 5
C_T	-.076	-.016	.027	.033
γ	-.468 2	-.612 1	-.613 2	-.582 2
ST	-.074	-.204 6	-.220 6	-.204 6
∇T	-.010	.036	.033	.044
κ	-.118	-.060	-.047	-.037
ε	-.200 5	-.367 3	-.365 3	-.356 3
MW	.008	.009	-.007	-.035
	R^2 = .800	R^2 = .865	R^2 = .911	R^2 = .939

[a] Variable rank based on absolute value of standardized regression coefficient.

Table 6.1. Additional Variables for Analysis of Aggregation.

Variable	Definition	Range	Restrictions
SSB	Smallest section boundary (m)	.01E-6 to 1.E-6	Loguniform
LSB	Largest section boundary (m)	50.E-6 to 200.E-6	Uniform
NS	Number of sections (integer values only)	5 to 20	Uniform

Table 6.2. Stepwise Regression Analyses of Aggregation Effects for Time to 90 percent and 99 percent Deposition of First Component. Table organization is the same as for Table 5.3.

	Time 90% Deposition			Time 99% Deposition		
Step	Var	SRC	R^2	Var	SRC	R^2
1	SR2	-.635	.372	SR2	-.700	.489
2	γ	-.567	.530	γ	-.586	.648
3	MMD1	-.168	.647	ϵ	-.324	.733
4	ϵ	-.346	.748	MMD1	-.120	.796
5	ρ	.226	.807	ρ	.230	.854
6	ST	-.252	.861	ST	-.204	.901
7	RM1	-.271	.912	χ	.289	.934
8	χ	.299	.956	RM1	-.163	.949
9	SSB	-.089	.964	NS	-.147	.962
10				ρ^2	-.125	.973

Table 6.3. Stepwise Regression Analyses of Aggregation Effects for Integrated Concentration of the First and Second Component. Table organizaton is the same as for Table 5.3.

	Integrated Conc. First Comp.			Integrated Conc. Second Comp.		
Step	Var	SRC	R^2	Var	SRC	R^2
1	RM1	.776	.491	SR2	.661	.332
2	SR2	-.423	.625	RD2	.489	.531
3	γ	-.420	.738	ϵ	-.341	.651
4	ϵ	-.261	.812	γ	-.478	.772
5	MMD1	-.232	.867	MMD2	-.270	.830
6	ST	-.210	.905	ST	-.219	.875
7	χ	.244	.943	χ	.248	.920
8	ρ	.152	.965	ρ	.178	.940
9				$MMD2^2$	-.157	.961

Another way to investigate the effects of aggregation is by comparing distribution functions generated by Latin hypercube sampling. Specifically, the distribution functions are generated and compared for several different formulations of the model involving different levels of complexity. The fundamental question of interest is whether changing the complexity of the model has a significant effect on model predictions. As already indicated, it is possible that other uncertainties with respect to the system will negate any improvement in model predictions that might result from greater model complexity. The use of Latin hypercube sampling has the advantage that it forces the full range of each uncertain parameter to be considered.

The results of two such comparisons appear in Figures 6.1 and 6.2. These figures were generated with use of a Latin hypercube sample of size 50 from the variables in Table 2.1 and use of the MAEROS model with the indicated number of sections and section boundaries. The results contained in these figures and in other unpresented figures clearly indicate that the dependent variable uncertainty resulting from the variation defined in Table 2.1 far exceeds the variation caused by the section boundaries or number of sections used. If the results for 20 sections and boundaries of .01E-6 and 400.E-6 meters are taken as "correct," then the results for the computationally much less demanding formulations with 10 sections and boundaries of .1E-6 and 50.E-6 meters are almost the same. Indeed, even when the number of sections is reduced to 5, the difference in predictions is small in comparison to the differences caused by the variation of the variables in Table 2.1.

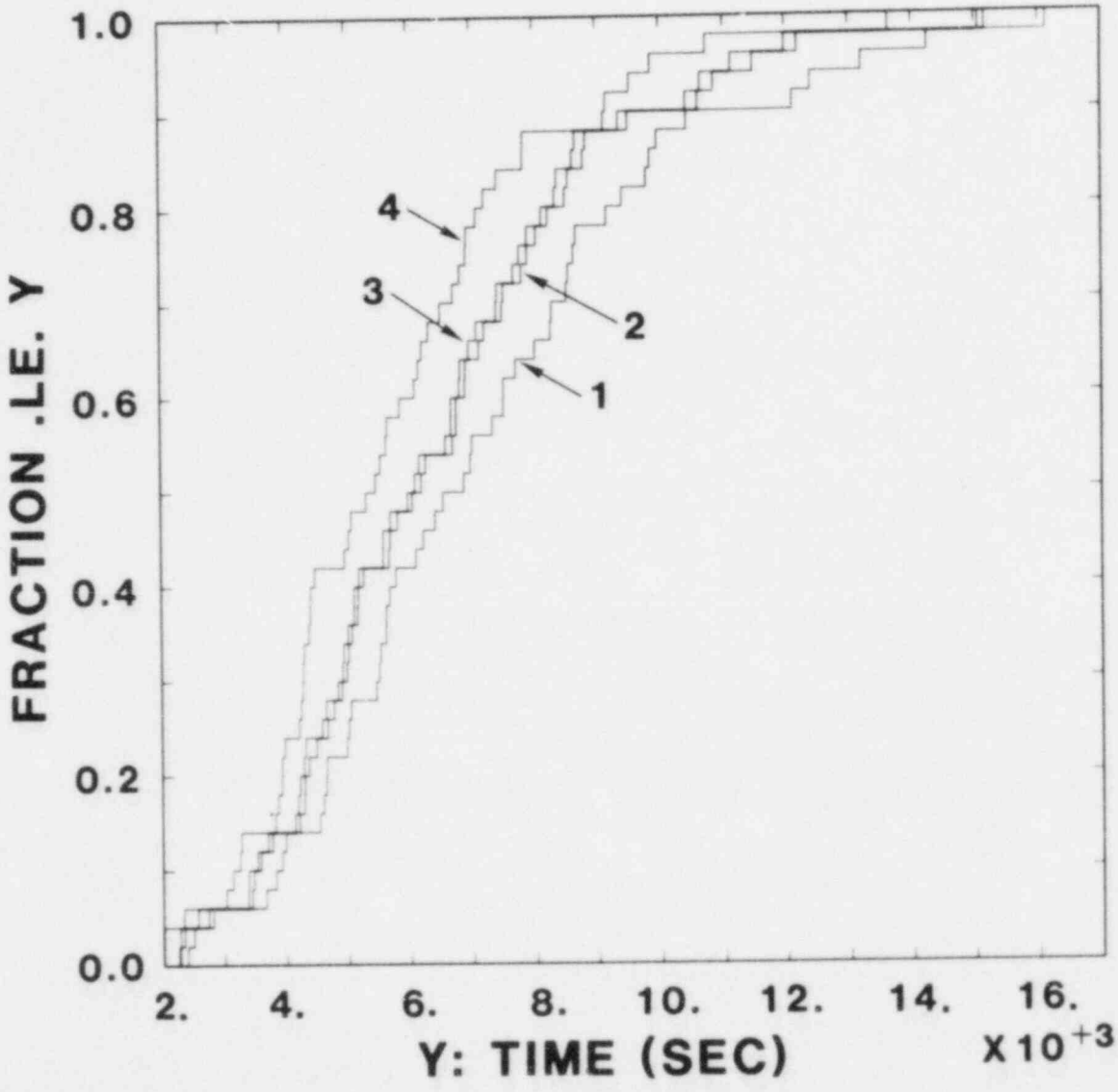

Figure 6.1. Distribution Functions for Time to 99 percent Deposition of First Component for Different Section Boundaries and Number of Sections (1 ~ .1E-6 to 50.E-6 m, 5 sections; 2 ~ .1E-6 to 50.E-6 m, 10 sections; 3 ~ .01E-6 to 400.E-6 m, 20 sections; 4 ~ 1.E-6 to 50.E-6 m, 15 sections).

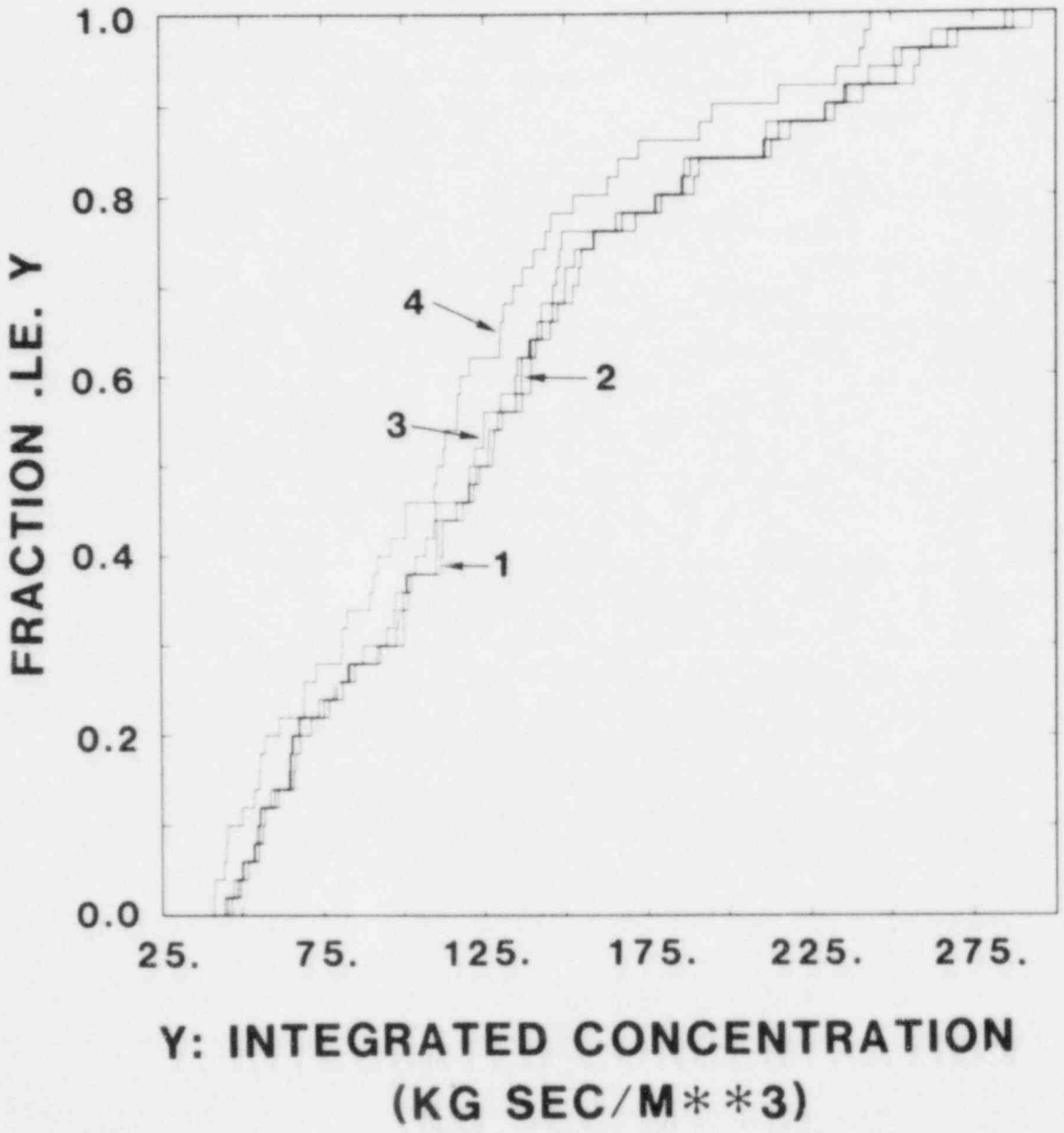

Figure 6.2. Distribution Functions for Integrated Concentration of Second Component for Different Section Boundaries and Number of Sections (1 ~ .1E-6 to 50.E-6 m, 5 sections; 2 ~ .1E-6 to 50.E-6 m, 10 sections; 3 ~ .01E-6 to 400.E-6 m, 20 sections; 4 ~ 1.E-6 to 50.E-6 m, 15 sections).

Another way to present the comparisons given in Figures 6.1 and 6.2 is by generating curves from pairs (x,y), where x is the "correct" model prediction associated with the most detailed model formulation and y is the model prediction associated with a less detailed formulation. If the predictions for the two formulations are the same and the same scale is used on both axes, then the associated curve will be a straight line through the origin with a slope of 1. In contrast, curves above and below this line indicate consistent over and under prediction, respectively.

7. DIFFERENTIAL ANALYSIS. Thus far, this presentation has used techniques based on Latin hypercube sampling and regression analysis. An alternate approach to sensitivity analysis is based on the partial differentiation of model predictions with respect to the independent variables under consideration. Once these partial derivatives are obtained, they can be used in ranking variable importance. Specifically, the model prediction Y of interest is approximated by a Taylor series of the form

$$Y(X) \doteq Y(X_0) + \sum_i \{\partial Y(X_0)/\partial X_i\} \quad \{X_i - X_{i0}\} \quad , \tag{7.1}$$

where X denotes a vector of variable values X_i and X_0 denotes a vector of "base case" variable values X_{i0}. Higher order Taylor series approximations are also possible.

The partial derivatives appearing in (7.1) are dependent on the units in which the X_i are expressed, and thus, are not by themselves very useful in ranking variable importance. Typically, the partial derivatives are normalized to

$$\{\partial Y(X_0)/\partial X_i\} \quad \{X_{i0}/Y(X_0)\} \tag{7.2}$$

or

$$\{\partial Y(X_0)/\partial X_i\} \quad \{S(X_i)/S(Y)\} \quad , \tag{7.3}$$

where S is used to denote standard deviation. An ordering of the absolute values of either of these normalizations can be used to rank variable importance. The normalization in (7.2) leads to a ranking of variable importance based on the effects of equal fractional changes of the base case values X_{i0}. The normalization in (7.3) leads to a ranking of variable importance based on the effects of perturbing the X_i from X_{i0} by equal fractions of the standard deviation $S(X_i)$. Thus, a ranking based on (7.3) incorporates effects due to distributions assigned to the X_i while a ranking based on (7.2) does not. Conceptually, the normalization in (7.3) is equivalent to a standardized regression coefficient.

If (7.2) and (7.3) are to be used to rank variable importance, it is first necessary to obtain the required partial derivatives. If the function Y appearing in (7.1) is relatively simple, then it may be possible to generate these derivatives analytically or by simple differencing schemes. Often, Y is too complex to permit such simple approaches. When Y arises in the solution of a system of equations, a common approach is to introduce a second system of equations that involves the desired

partial derivatives as unknown functions and then solve both systems simultaneously. It is this latter approach which was taken to generate partial derivatives for the following example.

As previously indicated, the model under consideration is a system of differential equations of the form

$$dQ_{\ell k}(X,t)/dt = f_{\ell k}\,[Q(X,t),X,t],\ Q_{\ell k}(0) = Q_{\ell k}(X,0) \quad , \tag{7.4}$$

where Q is the vector of unknown functions and X is the vector of variables whose effects are under investigation (i.e., the variables listed in Table 2.1). It is desired to estimate $\partial Q_{\ell k}(X_0,t)/\partial X_i$ for each element X_i of X, where X_0 denotes a vector of base case values for the X_i. This can be done by formulating a new problem which permits a simultaneous solution for both $Q_{\ell k}$ and $\partial Q_{\ell k}/\partial X_i$. This new problem is obtained by differentiating the expression in (7.4) with respect to X_i, which yields

$$\begin{aligned} \partial[dQ_{\ell k}(X,t)/dt]/\partial X_i &= \partial\{f_{\ell k}[Q(X,t),X,t]\}/\partial X_i \\ \partial[Q_{\ell k}(0)]/\partial X_i &= \partial[Q_{\ell k}(X,0)]/\partial X_i \quad . \end{aligned} \tag{7.5}$$

In turn, the preceding system can be reformulated as

$$\begin{aligned} d[\partial Q_{\ell k}(X,t)/\partial X_i]/dt &= \partial\{f_{\ell k}[Q(X,t),X,t]\}/\partial X_i \\ &\quad + \textstyle\sum_{pq} \partial\{f_{\ell k}[Q(X,t),X,t]\}/\partial Q_{pq} \cdot \partial Q_{pq}/\partial X_i \\ \partial[Q_{\ell k}(0)]/\partial X_i &= \partial[Q_{\ell k}(X,0)]/\partial X_i \quad . \end{aligned} \tag{7.6}$$

The systems in (7.4) and (7.6) can now be solved simultaneously to obtain $Q_{\ell k}$ and $\partial Q_{\ell k}/\partial X_i$.

The approach outlined in the preceding paragraph was used to obtain the partial derivatives which are presented in this section. This approach has been widely used and additional background is available elsewhere (e.g., Tomovic, 1963; Tomovic and Vukobratovic, 1972; Frank, 1978). As it is necessary to solve the systems in (7.4) and (7.6) for each X_i, generating partial derivatives in this manner can require a large amount of computation. As will be briefly indicated later, there exist procedures which have been developed to reduce the amount of required computation.

The partial derivatives of the integrated concentration of the first component appear in Table 7.1. These partial derivatives and all others in this presentation were calculated at the expected values of the variables appearing in Table 2.1. Columns 1 and 2 present normalizations of the derivatives as shown in (7.2) and (7.3), respectively. As the integer rankings listed in these columns indicate, these normalizations produce similar, but not identical, rankings of variable importance.

For comparison, Columns 3, 4 and 5 of Table 7.1 present rankings based on standardized regression coefficients; such coefficients are analogous to the normalized partial derivatives appearing in Column 2. Each regression involved the construction of a regression model with the 21 variables in Table 2.1 and the previously discussed Latin hypercube sample of size 50. No transformations were used for either the independent or dependent variables for the regression in Column 3. If integrated concentration were truly a linear function of the X_i, then the derivatives in Column 2 and the regression coefficients in Column 3 should be the same. The variables for the regression in Column 4 were normalized as discussed in Section 5. Finally, the variables for the regression in Column 5 were rank transformed. That is, each variable value was assigned a rank from 1 to 50 based on its relative size, and then the regression was performed on these ranks. Such rank regressions often provide a useful way to study nonlinear, monotonic relationships between variables within the framework of linear regression analysis (Iman and Conover, 1979).

Overall, the variable rankings in Columns 2, 3, 4 and 5 of Table 7.1 are similar. This is especially true for the more important variables (e.g., compare the variables ranked 1 through 7 in importance). There is less consistency for some of the less important variables. The partial derivatives and the indicated rankings in Columns 1 and 2 should be correct for at least small perturbations from the basecase values. However, the effects due to many of the less important variables are so small that they are lost in the concurrent variation of many more important variables, and hence, are not properly indicated in a regression model. In this regard, it is perhaps useful to consider the stepwise regression for integrated concentration of the first component presented in Table 5.4. There, a regression model with $R^2 = .907$ is produced with the use of five variables, while a regression model with $R^2 = .960$ is produced with the use of seven variables. In contrast, as shown in Column 4 of Table 7.1, use of all 21 variables produces a model with only a slightly improved R^2 value (i.e., $R^2 = .973$).

The partial derivatives of the total suspended mass of the first component at times 20 min., 40 min., 60 min. and 120 min. after the start of the problem appear in Tables 7.2 and 7.3. The derivatives in Table 7.2 have been normalized as shown (7.2), while those in Table 7.3 have been normalized as in (7.3). As for the derivatives in Table 7.1, the derivatives in Tables 7.2 and 7.3 indicate the effects of perturbations away from the expected values for the independent variables. The variable rankings in Tables 7.2 and 7.3 can be compared with those in Tables 5.2 and 5.5 obtained with partial rank correlation analysis and regression analysis, respectively. The variable rankings are similar but not the same. This is to be expected as the ranking criteria are not the same.

This presentation contains only a small number of the regressions and Taylor series generated in this analysis. For some variables, rankings based on partial derivatives normalized as in (7.3) and on standardized regression coefficients were similar; for other variables, they were different. If problems associated with the generation of the results are ignored, then which ranking is preferred depends on the type of information that is desired. If a ranking based on small perturbations of all variables is desired, then the ranking based on partial differentiation is probably better. However, if it is desired to assess the effects of the concurrent variation of a number of variables over possibly wide ranges, then rankings based on standardized regression coefficients and partial correlation coefficients are probably better.

Table 7.1. Normalized Partial Derivatives and Standardized Regression Coefficients for Integrated Concentration of First Component.

Var	1[a]		2[b]		3[c]		4[d]		5[e]	
GSD1	-.126	10[f]	-.065	11	-.086	8	-.054	11	-.045	13
MMD1	-.210	9	-.310	4	-.329	4	-.261	4	-.244	5
RM1	.754	2	.832	1	.794	1	.811	1	.830	1
GSD2	-.028	14	-.014	14	.008	19	-.011	19	.056	12
MMD2	-.054	13	-.108	8	.039	14	-.029	13	-.010	21
SR2	-.362	6	-.399	3	-.431	2	-.411	3	-.345	3
RD2	-.321E-3	20	-.137E-3	20	.338E-3	21	-.023	15	-.020	19
T	.486	4	.069	10	.028	16	.063	9	.085	9
P	-.114	12	-.030	13	.006	20	-.060	10	-.060	11
RCV	-.021	15	-.010	15	-.074	9	-.021	16	.012	20
RWV	-.010	17	-.005	16	-.030	15	-.014	18	-.029	17
χ	.476	5	.237	5	.073	10	.223	6	.261	4
δ	.604E-6	21	.509E-6	21	-.058	12	.004	20	-.037	15
ρ	.286	7	.096	9	.184	6	.185	7	.168	8
C_T	-.668E-3	19	-.333E-3	19	-.062	11	-.028	14	-.030	16
γ	-1.324	1	-.503	2	-.344	3	-.461	2	-.454	2
ST	-.557	3	-.186	7	-.086	7	-.167	8	-.173	7
∇T	-.014	16	-.565E-3	18	.016	17	-.016	17	-.074	10
κ	-.006	18	-.004	17	-.055	13	-.038	12	-.027	18
ϵ	-.240	8	-.217	6	-.236	5	-.254	5	-.179	6
MW	-.117	11	-.039	12	.011	18	.002	21	-.040	14

[a] $(\partial Y/\partial X_i)(X_i/Y)$.

[b] $(\partial Y/\partial X_i)(S(X_i)/S(Y))$ where S denotes standard deviation.

[c] Standardized regression coefficients with no normalizations or transformations. $R^2 = .905$.

[d] Standardized regression coefficients with (1) logarithms taken for Y and all log uniformly distributed variables, and (2) all variables normalized to mean 0 and standard deviation 1. $R^2 = .973$.

[e] Standardized regression coefficients with (1) all variables rank transformed, and (2) all variables normalized to mean 0 and standard deviation 1. $R^2 = .901$.

[f] Variable rank.

Table 7.2. Normalized Partial Derivatives $(\partial Y/\partial X_i)(X_i/Y)$ for Total Suspended Mass of First Component.

Var	20 min		40 min		60 min		120 min	
GSD1	-.164	3[a]	-.180	12	-.853E-01	13	-.168E-01	15
MMD1	-.160	5	-.349	9	-.350	11	-.459	10
RM1	.916	1	.510	5	.376	8	.399	11
GSD2	-.195E-01	12	-.934E-01	14	-.228E-01	15	.156	12
MMD2	-.169E-01	13	-.131	13	-.125	12	-.149E-01	17
SR2	-.319E-01	11	-.483	6	-1.16	5	-2.86	3
RD2	0.	21	0.	21	0.	21	.0	21
T	.128	6	.797	3	1.33	3	2.48	4
P	-.883E-02	17	-.182	11	-.358	10	-.734	9
RCV	-.623E-01	8	-.358E-01	15	.814E-01	14	.133	13
RWV	-.112E-01	15	-.141E-01	17	-.113E-01	17	-.165E-01	16
χ	.162	4	.777	4	1.23	4	2.24	5
δ	.738E-06	20	.806E-06	20	.554E-06	20	.295E-06	20
ρ	.344E-01	10	.450	7	.884	6	1.79	6
C_T	-.772E-03	19	-.906E-03	19	-.703E-03	19	-.698E-03	19
γ	-.251	2	-2.11	1	-3.88	1	-7.63	1
ST	-.109	7	-.901	2	-1.62	2	-3.08	2
∇T	-.155E-01	14	-.193E-01	16	-.153E-01	16	-.212E-01	14
κ	-.654E-02	18	-.821E-02	18	-.657E-02	18	-.961E-02	18
ϵ	-.500E-01	9	-.387	8	-.699	7	-1.29	7
MW	-.939E-02	16	-.186	10	-.367	9	-.754	8
	TSM[b] = 343 kg		TSM = 216 kg		TSM = 87.4 kg		TSM = 4.7 kg	

[a] Variable rank.

[b] Total suspended mass of first component.

Table 7.3. Normalized Partial Derivatives $(\partial Y/\partial X_i)(S(X_i)/S(Y))$ for Total Suspended Mass of First Component.

Var	20 min		40 min		60 min		120 min	
GSD1	-.917E-01	4[a]	-.775E-01	11	-.209E-01	13	-.201E-02	15
MMD1	-.256	2	-.430	4	-.245	6	-.156	6
RM1	1.09	1	.468	2	.196	7	.101	8
GSD2	-.109E-01	13	-.402E-01	13	-.560E-02	15	.186E-01	12
MMD2	-.368E-01	9	-.220	8	-.119	9	-.693E-02	14
SR2	-.380E-01	8	-.444	3	-.606	2	-.728	1
RD2	0.	21	0.	21	0.	21	0.	21
T	.197E-01	11	.943E-01	10	.901E-01	10	.811E-01	9
P	-.251E-02	17	-.400E-01	14	-.447E-01	12	-.446E-01	11
RCV	-.335E-01	10	-.148E-01	15	.191E-01	14	.153E-01	13
RWV	-.606E-02	14	-.586E-02	16	-.268E-02	16	-.189E-02	16
χ	.874E-01	5	.322	5	.290	4	.257	4
δ	.671E-06	20	.565E-06	20	.221E-06	20	.574E-07	20
ρ	.124E-01	12	.125	9	.140	8	.138	7
C_T	-.414E-03	19	-.375E-03	19	-.165E-03	19	-.800E-04	19
γ	-.103	3	-.669	1	-.699	1	-.668	2
ST	-.393E-01	7	-.250	7	-.257	5	-.237	5
∇T	-.822E-03	18	-.625E-03	18	-.197E-03	18	.867E-04	18
κ	-.474E-02	15	-.460E-02	17	-.209E-02	17	-.149E-02	17
ε	-.487E-01	6	-.291	6	-.298	3	-.269	3
MW	-.336E-02	16	-.516E-01	12	-.578E-01	11	-.577E-01	10

[a] Variable rank.

A problem which cannot be avoided in practice is that of cost, both human and computational. The differential analysis presented in this section was far more demanding, in terms of human and computer time, to implement than the analysis based on Latin hypercube sampling. First, it was necessary to write a computer program which would define and solve the two systems in (7.4) and (7.6). This turned out to be a very tedious and numerically difficult task. Second, once these systems were defined, it was necessary to solve them simultaneously for each of the 21 independent variables under consideration. In contrast, implementation of the analysis based on Latin hypercube sampling was much easier. Already available programs were used to generate the sample and to perform the regression and correlation analyses. Further, the basic model defined by (7.4) could be used in an essentially unchanged form by placing it within a DO loop which supplied the individual sample vectors to it and then recorded the desired dependent variables on a file for later analysis.

It is possible to reduce the amount of computational time required in a differential analysis by the use of specialized numerical procedures to avoid having to solve the systems in (7.4) and (7.6) once for each independent variable under consideration. These procedures are known as adjoint methods (e.g., Koda et al., 1979; Cacuci et al., 1980; Cacuci , 1981a and 1981b; Piepho et al., 1981; Hall et al., 1982; Cacuci et al., 1983) and Green's function methods (e.g., Hwang et al., 1978; Dougherty et al., 1979; Demiralp and Rabitz, 1981a and 1981b; Dacol and Rabitz, 1983). These procedures are conceptually similar and require the solution of a differential equation involving the Jacobian matrix for the original system (i.e., equation (7.4)). Once this second system has been solved, it can be used repeatedly in numerical integrations to evaluate the desired partial derivatives. However, implementation of these techniques is numerically more complex than simultaneously formulating and solving the systems in (7.4) and (7.6); thus, they are also more demanding of human time. Hence, there is a trade-off. It may be possible to save computational time but at the cost of greater human time in the implementation of the analysis. At the outset of this effort, the intent was to use adjoint methods to calculate the desired partial derivatives. Ultimately, it was decided that it would be more cost effective to repetitively solve the systems in (7.4) and (7.6) than to implement an adjoint analysis.

8. DISCUSSION. An approach to uncertainty and sensitivity analysis based on Latin hypercube sampling and regression analysis has been presented and illustrated. The basic idea is (1) to select a set of potentially important variables that define the input for the model, (2) to choose ranges and distributions for the variables, (3) to sample from the variables according to their assigned ranges and distributions, (4) to generate input values for the model from the sampled values of the variables, (5) to produce model output with the generated input, (6) to estimate distribution functions and population parameters directly from the model predictions, and (7) to assess the relationships between the original variables and model output by regression techniques.

The preceding approach has a number of desirable attributes. Latin hypercube sampling allows treatment of multivariate input involving arbitrary distributions and correlations. Further, the entire range of each variable is considered. This is important as different independent variables may be important at different times or

with respect to different dependent variables. In contrast, random sampling does not cover the range of each variable as efficiently, and fractional factorial sampling involves only a few values for each variable and is not amenable to the inclusion of distributions and correlations. Output generated by Latin hypercube sampling can be used to obtain scatter plots of the values of individual dependent variables versus individual independent variables; such plots can provide guidance with respect to individual variable importance and possibly with respect to the conduct of the analysis.

Uncertainty analyses can be performed through the direct estimation of distribution functions. Direct estimates for expected values and standard deviations are also possible; however, as model predictions are typically nonnormal and the preceding quantities are not robust measures of model behavior, expected values and standard deviations are not as useful in describing model behavior as are estimated distribution functions. Further, even if the distributions assigned to individual variables are questionable, distribution functions provide guidance with respect to potential ranges for dependent variables of interest. Sensitivity analyses can be performed with partial correlation analysis and regression analysis. Further, various data transformations can be used to facilitate both types of analysis. Regression analyses are probably best performed by stepwise regression; then, the order in which variables enter a regression, the changes in R^2 values with the addition of successive variables, and the standardized regression coefficients in the final regression model can be used to gain insight with respect to variable importance. Further, the R^2 value for the final regression provides insight with respect to the amount of variation in model predictions which has been accounted for.

An important aspect of this approach to uncertainty and sensitivity analysis is that it does not require extensive modification of the underlying model; rather, the model is placed within a DO loop, the sampled input parameters are supplied to it, and then the predictions of interest are stored for later analysis. Both the program that generates the Latin hypercube sample and the programs that are used in the analysis of the resultant predictions are independent of the actual model under consideration.

For comparison, a sensitivity analysis based on differential techniques was also presented. In such an analysis, model predictions of interest are expanded as Taylor series and then these series are used in assessing model behavior. However, such series are based on the behavior of model predictions at a specific set of values for the independent variables. Such behavior may or may not extend far beyond the point at which the analysis is performed. Thus, while differential analyses can give very good information for small perturbations of variable values, care must be taken in their use if larger perturbations are involved. This is particularly true if uncertainty analyses are performed with Taylor series approximations for model predictions through the use of Monte Carlo methods or variance propagation methods. Further, computation of the necessary partial derivatives may require a significant outlay of human and computational effort.

It is important to recognize the impacts that distribution assumptions can have on the outcome of an analysis. Estimated distribution functions are conditional on the distributions used for the independent variables. However, a valuable property of Latin hypercube sampling when dealing with computer models which are expensive to evaluate is that it is possible to recalculate estimated distribution

functions to reflect different distribution assumptions for the independent variables without rerunning the model (Iman and Conover, 1980). Other parts of an analysis are also affected by distribution assumptions. The points selected in a Latin hypercube sample ultimately derive from the distribution assumptions for the independent variables. Thus, the raw data on which partial correlation and regression analyses are based are influenced by distribution assumptions. This is probably not a serious problem unless the sampling procedure and the assumed distributions result in important areas of model behavior being omitted. Distribution assumptions do not enter directly into the calculation of partial correlation coefficients but they do influence the value of standardized regression coefficients due to the use of expected values and standard deviations for the independent variables in the associated normalization procedure. When rank regression is used, this effect is removed as all independent variables have the same expected value and standard deviation.

It might appear that analyses based on fractional factorial designs do not involve assumptions about variable distributions in the generation of a sample. In reality, such analyses make a fundamental assumption about the importance of variable distributions: that it is necessary to evaluate a model at only a limited number of values for a variable to assess the impact of this variable on model predictions. Similarly, it might appear that analyses based on partial derivatives normalized with respect to base case values are distribution independent. However, this is not entirely true as some type of decision with respect to the distribution of the independent variables must be made to select the base case point at which the partial derivatives are calculated. Unless the underlying model is linear with respect to the independent variables, its partial derivatives can change significantly with different base case selections. Thus, although the approach to uncertainty and sensitivity analysis presented in this paper may appear to make heavy use of distribution assumptions for the independent variables, careful consideration of other approaches to uncertainty and sensitivity analysis will reveal that they also contain distribution assumptions.

Analyses of the type described in this paper can have a number of uses, several of which are now indicated. Insights into the uncertainty in model predictions and the variables which cause this uncertainty can be developed. Such insights might lead to the conclusion that the uncertainty in model predictions was acceptable or, if such uncertainty was not acceptable, guidance might be provided as to the research needed to reduce the uncertainty. Further, insights into the operation of the models and the variables which influence this behavior can be gained. In turn, such insights may lead to ways to improve or simplify model operation. As a final example, uncertainty and sensitivity analysis provide a powerful tool for model quality assurance. Due to the extensive variation of all variables and the capability of regression techniques to reveal model behavior, the presented techniques provide a systematic way to search for errors in model development. In the following paragraphs, some of the insights gained from the analysis of aerosol behavior in a TMLB' accident are given. Additional discussion of these insights is available in Leigh and Helton (1985).

The time dependent behavior of each of the two aerosol components is shown in Section 5. Initially, the containment volume is loaded with particles consisting only of the first component; particles containing the second component are introduced as a source. The particles grow through agglomeration and deposit onto the

containment surfaces. As a result, the suspended mass of the first component is large at first, decreases as the accident progresses, and is eventually depleted. The suspended mass of the second component increases until it reaches a steady state plateau where the addition and depletion rates are equal and then decreases after its source is shut off. The time for depletion of the first component decreases when the source rate of the second component is increased, and since the first component is generally very radiotoxic, the presence of a source of less radiotoxic particles can help to reduce the radiological consequences of a releases from containment.

The range of predictions obtained from the MAEROS model for this scenario is shown in Section 5. The results of the regression analysis indicate that the initial mass loading of aerosol material, the source rate of aerosols, the dynamic and agglomeration shape factors, the turbulent energy dissipation rate, and the particle material density are the largest causes of uncertainty in the model predictions. The source rate and initial mass loading cause uncertainty at early times. As the calculation progresses, these parameters become less important and the parameters affecting the agglomeration rate constants become important (i.e., the shape factors, the turbulent energy dissipation rate and the material density). These results indicate that two physical processes control the depletion rate of the suspended material: aerosol agglomeration and resultant gravitational settling of large particles. Deposition by thermophoresis is not significant.

The ranges of justifiable values for the agglomeration and dynamic shape factors and the turbulent energy dissipation rate may be larger than used in this study (Lipinski et al., 1985). The shape factors are used in the analysis of experimental data and in various codes both as true physical parameters and as fitting or "fudge" parameters with values ranging from 1 to 10. The turbulent energy dissipation rate generally ranges from .001 m^3/sec to .02 m^3/sec. Uncertainty in its value reflects the uncertainty in the hydraulic conditions expected in a control volume. Since these parameters appear to account for a large part of the uncertainty in the predictions, obtaining more accurate values for these parameters is important and efforts to narrow the ranges for these three parameters are justified.

The effects of varying the largest and smallest section boundaries and the number of sections are shown in Section 6. The uncertainties introduced by varying these parameters are insignificant when compared to the overall model uncertainty. As a result, as long as large uncertainties in the parameters in Table 2.1 exist, it is probably adequate to use a particle size range from .1E-6 m to 50.E-6 m and approximately 5 to 10 size sections when modeling aerosol behavior in a dry control volume with a sectional approach like that used in the MAEROS model.

References

Adams, R. E., Han, J. T., Kress, T. S., and Silberburg, M. (1980). "Behavior of Sodium Oxide, Uranium Oxide and Mixed Sodium Oxide - Uranium Oxide Aerosols in a Large Vessel," in Proceedings of the CSNI Specialists Meetings on Nuclear Aerosols in Reactor Safety, NUREG/CR-1724, ORNL/NUREG/TM-404, CSNI-45, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37830, pp. 499-518.

Adams, R. E., Kress, T. S., and Tobias, M. L. (1981). Sodium Oxide and Uranium Oxide Aerosol Experiments: NSPP Tests 106-108 and Tests 204-207, Data Record Report, NUREG/CR-1767, ORNL/NUREG/TM-408, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37839.

Adams, R. E., Kress, T. S., and Tobias, M. L. (1982). Uranium Oxide and Sodium Oxide Aerosol Experiments: NSPP Mixed-Oxide Tests 303-307, Data Record Report, NUREG/CR-2697, ORNL/TM-835, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37830.

Adams, R. E., and Tobias, M. L. (1983). Aerosol Release and Transport Program Quarterly Progress Report for July - September 1982, NUREG/CR-2809, Vol. 3, ORNL/TM-8397/V3, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37830.

Allen, D. M. (1971). The Prediction Sum of Squares as a Criterion for Selecting Predictor Variables, Report No. 23, Department of Statistics, University of Kentucky, Lexington, Kentucky.

Aoki, M. (1971). "Aggregation," in Optimization Methods for Large Scale Systems (D. A. Wismer, ed.), McGraw Hill, New York, New York, pp. 191-232.

Bergeron, K. D., Clauser, M. J., Harrison, B. D., Murata, K. K., Rexroth, P. E., Schelling, F. J., Sciacca, F. W., Senglaub, M. E., Shire, P. R., and Trebilcock, W. (1985). User's Manual for CONTAIN, A Computer Code for Severe Nuclear Reactor Accident Containment Analysis, NUREG/CR-4085, SAND84-1204, Sandia National Laboratories, Albuquerque, New Mexico, 87185.

Cacuci, D. G. (1981a). "Sensitivity Theory for Nonlinear Systems. I. Nonlinear Functional Analysis Approach," J. Math. Phys. 22: 2794-2802.

Cacuci, D. G. (1981b). "Sensitivity Theory for Nonlinear Systems. II. Extensions to Additional Classes of Responses," J. Math. Phys. 22: 2803-2812.

Cacuci, D. G., Maudlin, P. J., and Parks, C. U. (1983). "Adjoint Sensitivity Analysis of Extremum-Type Responses in Reactor Safety," Nucl. Sci. Eng. 83: 112-135.

Cacuci, D. G., Weber, C. F., Oblow, E. M., and Marable, J. H. (1980). "Sensitivity Theory for General Systems of Nonlinear Equations," Nucl. Sci. Eng. 75: 88-110.

Dacol, D. K., and Rabitz, H. (1983). "Arbitrary Order Functional Sensitivity Densities for Reaction-Diffusion Systems," J. Chem. Phys. 78: 4905-4914.

Demiralp, M., and Rabitz, H. (1981a). "Chemical Kinetic Functional Sensitivity Analysis: Elementary Sensitivities," J. Chem. Phys. 74: 3362-3375.

Demiralp, M., and Rabitz, H. (1981b). "Chemical Kinetic Functional Sensitivity Analysis: Derived Sensitivities and General Applications," J. Chem. Phys. 75: 1810-1819.

Dougherty, E. P., Hwang, J.-T., and Rabitz, H. (1979). "Further Developments and Applications of the Green's Function Method of Sensitivity Analysis in Chemical Kinetics," J. Chem. Phys. 71: 1794-1808.

Fermandjian, J., Malet, J. C., Casselman, C., Duverger de Cuy, G., Boulaud, D., and Madelaine, G. (1980). "Interpretation of the Behavior of Aerosols Generated by a Sodium Pool Fire," in Proceedings of the CSNI Specialists Meetings on Nuclear Aerosols in Reactor Safety, NUREG/CR-1724, ORNL/NUREG/TM-404, CSNI-45,Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37830, pp. 482-498.

Frank, P. M. (1978). Introduction to System Sensitivity Theory, Academic Press, New York, New York.

Gardner, R. H., Cole, W. G., and O'Neill, R. V. (1982). "Robust Analysis of Aggregation Error," Ecology 63: 1771-1779.

Gelbard, F. (1982). MAEROS User Manual, NUREG/CR-1391, SAND80-0822, Sandia National Laboratories, Albuquerque, New Mexico 87185.

Gelbard, F., and Seinfeld, J. H. (1980). "Simulation of Multicomponent Aerosol Dynamics," J. Colloid Interface Sci., 78: 485-501.

Gibbons, J. C., Wolsky, A. M., and Tolley, G. (1982). "Approximate Aggregation and Error in Input-Output Models," Resources and Energy 4: 203-230.

Gieseke, J. A., Cybulskis, P., Denning, R. S., Kuhlman, M. R., and Lee, K. W. (1984). Radionuclide Release under Specific LWR Accident Conditions, Vols. I-VII, BMI-2104, Battelle Columbus Laboratories, Columbus, Ohio, 43201.

Golikeri, S. V., and Luss, D. (1974). "Aggregation of Many Coupled Consecutive First Order Reactions," Chem. Eng. Sci. 29: 845-855.

Hall, M. C. G., Cacuci, D. G., and Schlesinger, M. E. (1982). "Sensitivity Analysis of a Radiative-Convective Model by the Adjoint Method," J. Atmos. Sci. 39: 2038-2050.

Hilliard, R. K., McCormack, J. D., and Postma, A. K. (1983). Results and Code Predictions for ABCOVE Aerosol Code Validation--Test AB5, HEDL-TME 83-16, UC-79T, Ti, Tp, Hanford Engineering Development Laboratory, Richland, Washington, 99352.

Hwang, J.-T., Dougherty, E. P., Rabitz, S., and Rabitz, H. (1978). "The Green's Function Method of Sensitivity Analysis in Chemical Kinetics," J. Chem. Phys. 69: 5180-5191.

Ijiri, Y. (1971). "Fundamental Queries in Aggregation Theory," J. Amer. Stat. Soc. 66: 766-782.

Iman, R. L., and Conover, W. J. (1979). "The Use of the Rank Transform in Regression," Technometrics 21: 499-509.

Iman, R. L., and Conover, W. J. (1980). "Small Sample Sensitivity Analysis Techniques for Computer Models, with an Application to Risk Assessment," Communications in Statistics A9: 1749-1842. Rejoinder to Comments, 1863-1874.

Iman, R. L., and Conover, W. J. (1982a). "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables," Communications in Statistics B11: 311-334.

Iman, R. L., and Conover, W. J. (1982b). Sensitivity Analysis Techniques: Self-Teaching Curriculum, NUREG/CR-2350, SAND81-1978, Sandia National Laboratories, Albuquerque, New Mexico 87185.

Iman, R. L., Davenport, J. M., Frost, E. L., and Shortencarier, M. (1980). Stepwise Regression with PRESS and Rank Regression (Program User's Guide), SAND79-1472, Sandia National Laboratories, Albuquerque, New Mexico, 87185.

Iman, R. L., and Helton, J. C. (1985). A Comparison of Uncertainty and Sensitivity Analysis Techniques for Computer Models, NUREG/CR-3904, SAND84-1461, Sandia National Laboratories, Albuquerque, New Mexico 87185.

Iman, R. L., Helton, J. C., and Campbell, J. E. (1981a). "An Approach to Sensitivity Analysis of Computer Models, Part 1. Introduction, Input Variable Selection and Preliminary Variable Assessment," J. Quality Tech. 13: 174-183.

Iman, R. L., Helton, J. C., and Campbell, J. E. (1981b). "An Approach to Sensitivity Analysis of Computer Models, Part 2. Ranking of Input Variables, Response Surface Validation, Distribution Effect and Technique Synopsis," J. Quality Tech. 13: 232-240.

Iman, R. L., and Shortencarier, M. J. (1984). A FORTRAN 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models, NUREG/CR-3624, SAND83-2365, Sandia National Laboratories, Albuquerque, New Mexico 87185.

Iman, R. L., Shortencarier, M. J., and Johnson, J.D. (1985). A FORTRAN 77 Program and User's Guide for the Calculation of Partial Correlation and Standardized Regression Coefficients, NUREG/CR-4122, SAND85-0044, Sandia National Laboratories, Albuquerque, New Mexico 87185.

Koda, M., Dogru, A. H., and Seinfeld, J. H. (1979). "Sensitivity Analysis of Partial Differential Equations with Application to Reaction and Diffusion Processes," J. Comp. Phys. 30: 259-282.

Kress, T. S., and Tobias, M. L. (1980). LMFBR Aerosol Release and Transport Program Quarterly Progress Report for January-March 1980, NUREG/CR-1790, ORNL/NUREG/TM-416, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37830.

Leigh, C. D., and Helton, J. C. (1985). "A Sensitivity Study of Aerosol Agglomeration and Deposition Using the MAEROS Model," Proceedings of the ANS Topical Meeting on Fission Product Behavior and Source Term Research, Snowbird, UT, July 15-18, 1984 (to appear).

Lipinski, R. J., Bradley, D. R. Brockmann, J. E., Griesmeyer, J. M. Leigh, C. D., Murata, K. K., Powers, D. A., Rivard, J. B., Taig, A. R. Tills, J., and Williams, D. C. (1985). Uncertainty in Radionuclide Release under Specific LWR Conditions, Vols. I and II, SAND84-0410, Sandia National Laboratories, Albuquerque, New Mexico, 87185.

Loyalka, S. K. (1983). "Mechanics of Aerosols in Nuclear Reactor Safety: A Review," Prog. Nucl. Energy 12:1-56.

McKay, M. D., Conover, W. J., and Beckman, R. J. (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," Technometrics 21: 239-245.

Piepho, M. G., Cady, K. B., and Kenton, M. A. (1981). An Importance Theory for Nonlinear Ordinary Differential Equations, HEDL-TME 81-24, Hanford Engineering Development Laboratory, Richland, Washington, 99352.

Rehn, B. (1980). "Direct Measurement of the Coagulation Shape Factor of Aerosol Particles," in Proceedings of the CSNI Specialists Meeting on Nuclear Aerosols in Reactor Safety, NUREG/CR-1724, ORNL/NUREG/TM-404, CSNI-45, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37830, pp. 139-150.

Sprung, J. L., Aldrich, D. C., Alpert, D. J., and Weigand, G. G. (1983). "Overview of the MELCOR Risk Code Development Program," in Proceedings of the International Meeting on Light Water Reactor Severe Accident Evaluation, pp. TS-10.1-1 to TS-10.1-8 , Cambridge, MA, August 28-September 1, 1983.

Tomovic, R., (1963). Sensitivity Analysis of Dynamic Systems, McGraw-Hill, New York, New York.

Tomovic, R., and Vukobratovic, M. (1972). General Sensitivity Theory, Elsevier, New York, New York.

U.S. Nuclear Regulatory Commission (1975). Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants, WASH-1400 (NUREG-75/014), Washington, D.C.

van de Vate, J. F., and ten Brink, H. M. (1980). "The Boundary Layer for Diffusive Aerosol Deposition onto Walls," in Proceedings of the CSNI Specialists Meeting on Nuclear Aerosols in Reactor Safety, NUREG/CR-1724, ORNL/NUREG/TM-404, CSNI-45, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37830, pp. 162-170.

Whitby, K. T. (1975). Modeling of Atmospheric Aerosol Size Distributions, Particle Technology Laboratory Report No. 253, Dept. of Mechanical Engineering, University of Minnesota, Minneapolis, Minnesota.

Zeigler, B. P. (1976). "The Aggregation Problem," in Systems Analysis and Simulation in Ecology (B. C. Patten, ed.), Academic Press, New York, New York, Vol. 4, pp. 299-311.

U. S. Government Printing Office
Receiving Branch (Attn: NRC Stock)
8610 Cherry Lane
Laurel, MD 20707
250 copies for RG

USNRC (5)
Division of Risk Analysis and Operations
Washington, DC 20555
M.Cunningham
J. Johnson
L. Lancaster
T. Margulies
D. Rasmuson

Los Alamos National Laboratory (3)
Group S1, MS 606
Los Alamos, NM 87545
H. F. Martz
M. E. Johnson
R. A. Waller

Frank Abbey
United Kingdom Atomic Energy Authority
Safety & Reliability Directorate
Wigshaw Lane
Culcheth
Warrington WA3 4NE
UNITED KINGDOM

David C. Aldrich
SAIC
1710 Goodridge Drive
P.O. Box 1303
McLean, VA 22102

George Apostolakis
Mechanics and Structures Dept.
UCLA
Los Angeles, CA 90024

Nathaniel F. Barr
U.S. Department of Energy
Washington, DC 20545

Steve Bartell
Environmental Sciences Division
Oak Ridge National Laboratory
P.O. Box X
Oak Ridge, TN 37830

Paul Baybutt
Battelle Columbus Laboratories
505 King Avenue
Columbus, OH 43201

Anton Bayer
Kernforschnungszentrum Karlsruhe
Postfach 3640
D-7500 Karlsruhe 1
WEST GERMANY

Thomas A. Bishop
Battelle Laboratories
505 King Avenue
Columbus, OH 43201

Klaus Burkart
Institut fur Neutronenphysik und
Reaktortechnik
Kernforschungszentrum Karlsruhe
Postfach 3640
D-7500 Karlsruhe 1
WEST GERMANY

D. G. Cacuci
Oak Ridge National Laboratory
Engineering Physics Division
P.O. Box X
Oak Ridge, TN 37830

James E. Campbell
Intera Environmental Consultants Inc.
11999 Katy Freeway
Houston, TX 77079

W. J. Conover
College of Business Administration
Texas Tech University
Lubbock, TX 79409

Michael Cullingford
P.O. Box 200
A-1400
Vienna
AUSTRIA

Pamela Doctor
Battelle Northwest
P.O. Box 999
Richland, WA 99352

Darryl Downing
Computer Sciences
Building 2029, P.O. Box X
ORNL
Oak Ridge, TN 37830

Daniel Egan
Office of Radiation Program s(ANR-460)
U. S. Environmental Protection Agency
Washington, DC 20460

Joachim Ehrhardt
Kernforschungszentrum Karlsruhe
Postfach 3640
D-7500 Karlsruhe l
WEST GERMANY

R. H. Gardner
Environmental Sciences Division
ORNL
Oak Ridge, TN 37830

Jim Gieseke
Battelle Laboratories
505 King Avenue
Columbus, OH 43201

William V. Harper
Performance Analysis Department
Battelle Laboratories
505 King Avenue
Columbus, OH 43201

Max Henrion
Dept. of Engineering and Public Policy
Carnegie-Mellon University
Pittsburgh, PA 15213

Edward Hofer
Gesellschaft für Reaktorsicherheit
D-8046 Garching
FEDERAL REPUBLIC OF GERMANY

F. O. Hoffman
Health and Safety Research Division
ORNL
Oak Ridge, TN 37830

Stephen C. Hora
College of Business Administration
Texas Tech University
Lubbock, TX 79409

Frank W. Horsch
Kernforschungszentrum Karlsruhe
Postfach 3640
D-7500 Karlsruhe l
WEST GERMANY

H. Jordan
Battelle Laboratories
505 King Avenue
Columbus, OH 43201

Geoffrey D. Kaiser
Consulting Division
NUS Corporation
910 Clopper Road
Gaithersburg, MD 20878

G. Neale Kelly
NII/HSE
Silkhouse Court
Tithebarn Street
Liverpool L2 2LZ
UNITED KINGDOM

Tom Kress
Bldg. 9108, MS-2
Oak Ridge National Laboratory
Oak Ridge, TN 37830

Greg McRae
Dept. of Chemical Engineering
Carnegie-Mellon University
Pittsburgh, PA 15213

M. Granger Morgan
Dept. of Engineering and Public Policy
Carnegie-Mellon University
Pittsburgh, PA 15213

E. M. Oblow
Oak Ridge National Laboratory
Engineering Physics Division
P.O. Box X
Oak Ridge, TN 37830

Derek J. Pike
Department of Statistics
The University of Reading
Whiteknights
Reading RG6 2AN
ENGLAND

Trevor Pratt
Brookhaven National Laboratory
Upton, NY 11973

Robert Ritzman
Electric Power Research Institute
3412 Hillview Avenue
Palo Alto, CA 94303

Ulf Tveten (Technical Secretary)
Institute for Energy Technology
Postboks 40
N-2007 Kjeller
NORWAY

W. E. Vesely
Risk, Safety, and Reliability Section
Battelle Columbus Laboratories
505 King Avenue
Columbus, Ohio 43201

Eric R. Ziegel
Standard Oil Company (Indiana)
Amoco Research Center
P.O. Box 400
Naperville, IL 60566

2564 G. W. Smith
3141 C. M. Ostrander (5)
3151 W. L. Garner (3)
6312 M. S. Tierney
6400 A. W. Snyder
6410 J. W. Hickman
6412 M. P. Bohn
6415 D. J. Alpert
6415 J. M. Griesmeyer
6415 F. E. Haskin
6415 J. C. Helton (23)
6415 R. L. Iman
6415 J. D. Johnson
6415 C. D. Leigh
6415 M. J. Shortencarier
6415 J. L. Sprung
6422 J. E. Brockmann
6422 J. E. Gronager
6431 R. M. Cranwell
6444 J. M. McGlaun
6444 S. W. Webb
6449 K. D. Bergeron
6449 K. K. Murata
8024 M. A. Pound

NRC FORM 335 (2-84) NRCM 1102, 3201, 3202
U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

SEE INSTRUCTIONS ON THE REVERSE

1. REPORT NUMBER (Assigned by TIDC and Vol. No. if any)
NUREG/CR-4342
SAND84-1307

2. TITLE AND SUBTITLE
Uncertainty and Sensitivity Analysis of a Model for Multicomponent Aerosol Dynamics

3. LEAVE BLANK

4. DATE REPORT COMPLETED
MONTH: September | YEAR: 1985

5. AUTHOR(S)
J. C. Helton, R. L. Iman, J. D. Johnson, C. Q. Leigh

6. DATE REPORT ISSUED
MONTH: September | YEAR: 1985

7. PERFORMING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185

8. PROJECT/TASK/WORK UNIT NUMBER

9. FIN OR GRANT NUMBER
A-1339

10. SPONSORING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)
Division of Risk Analysis
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555

11a. TYPE OF REPORT

b. PERIOD COVERED (Inclusive dates)

12. SUPPLEMENTARY NOTES

13. ABSTRACT (200 words or less)

An uncertainty and sensitivity analysis of the MAEROS model for multicomponent aerosol dynamics is presented. Analysis techniques based on Latin hypercube sampling and regression analysis are used to study the behavior of a two component aerosol in a nuclear power plant containment for a transient accident with loss of AC power (i.e., a TMLB' accident). Conditional on assumed ranges and distributions for selected independent variables (e.g., initial distributions and mass loadings for each component, temperature, pressure, shape factors), estimates are made for distributions of model predictions and for the independent variables which influence these predictions. The analysis indicated that, for the situation under consideration, variables related to agglomeration (e.g., dynamic shape factor, material density, agglomeration shape factor, and turbulence dissipation rate) tended to dominate the observed variability. For comparison, an analysis based on differential techniques is also given. Further, a study of the effects on MAEROS predictions due to the number of particle size classes and the particle size class boundaries is presented. This analysis was performed as part of a project to develop a new system of computer codes (i.e., the MELCOR code system) for use in risk assessments for nuclear power plants.

14. DOCUMENT ANALYSIS - a. KEYWORDS/DESCRIPTORS

b. IDENTIFIERS/OPEN ENDED TERMS

15. AVAILABILITY STATEMENT

16. SECURITY CLASSIFICATION
(This page) Unclassified
(This report) Unclassified

17. NUMBER OF PAGES

18. PRICE