A TECHNIQUE FOR THE ASSESSMENT OF UNCERTAINTY IN WATER QUALITY MODELS USED FOR PUBLIC HEALTH RISK ANALYSIS

by

Mark A. Tumeo

and

Gerald T. Orlob

CALIFORNIA WATER RESOURCES CENTER

University of California

Contribution 201

ISSN 0575-4941

December 1990

A TECHNIQUE FOR THE ASSESSMENT OF UNCERTAINTY IN WATER QUALITY MODELS USED FOR PUBLIC HEALTH RISK ANALYSIS

By:
Mark A. Tumeo
Department of Civil Engineering
University of Alaska Fairbanks
Fairbanks, Alaska 99775

Gerald T. Orlob
Department of Civil Engineering
University of California Davis
Davis, California 95616

TECHNICAL COMPLETION REPORT
Project Number W-686

University of California Water Resources Center
Davis, California 95616

The research leading to this report was partially supported by the University of California Water Resources Center under grant number W-686.

This publication is furnished as a part of the information exchange program in water resources research and is published and distributed by the DIRECTOR'S OFFICE OF THE CALIFORNIA WATER RESOURCES CENTER. The Center sponsors projects in water resources and related research on the several campuses of the University of California with funds provided by the U.S. Department of Interior and from the State of California. Copies of this and other reports published by the Center may be obtained from:

WATER RESOURCES CENTER
UNIVERSITY OF CALIFORNIA
RUBIDOUX HALL
RIVERSIDE, CA 92521
(714) 787-4327

Copies of Center publications may be examined at either location of the Water Resources Center Archives: 410 O'Brien Hall, Berkeley Campus (415) 642-2666 or 2081 Engineering 1, Los Angeles Campus (213) 825-7734.

TABLE OF CONTENTS

TABLE OF CONTENTS
(CONTINUED)

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

(Keywords: Mathematical Models, Risk Analysis, Water Quality Modeling)

This paper reports on the development of a new mathematical technique to include stochasticity in environmental models used for environmental management and risk analysis. The technique is applied to analyze health risks associated with waterborne diseases. The results of the analysis compared favorably to the results of a Monte Carlo simulation of a similar system of equations.

The technique is based on the expansion of basic governing equations to include stochastic terms. The stochastic terms are then separated from the non-fluctuating terms, and the resulting set of equations are solved simultaneously. Solutions are used to calculate the moments of the output variables, which subsequently are used in conjunction with the Fokker-Planck Equation to produce an analytical solution for the probability density functions of the dependent variables.

The technique represents a new and potentially powerful tool for extending the capabilities of computer models in risk analysis and environmental quality management by providing analytical solutions for the probability density functions and associated moments of important environmental variables. It combines the positive characteristics of existing stochastic methods while eliminating many of their limitations and drawbacks. Because the method yields an analytic form for the variance of the output as a function of the variances of the input variables, it is possible to predict quantitatively the effects of stochastic variation in the input on the output. Sensitivity analysis, the traditional tool for analyzing variances, is extremely difficult to interpret when applied to more than one or two variables simultaneously. Through use of this new technique, it is possible to calculate explicitly the effects of variance from all the variables, in all the combinations which are present.

ACKNOWLEDGMENTS

The work described herein was partially funded by the California Water Resources Center at the University of California, Davis, under grant number W-686.

The authors would also like to acknowledge Dr. Lev Kavvas for his help and guidance in the area of stochastic differential equations and expectation mathematics and Ms. Blythe Hoyle for her editing comments. A special thanks is also extended to the office staff in Walker Hall, in the Department of Civil Engineering, Davis: Theresa Baciarini, Donna Douglas, Gina Gomez, Margaret Lozano-Williams, Pam Pickering, Hilda Pittman, Virginia Roy, Heidi Steiner, and Barbara Sullivan, for their logistic and emotional support during the completion of this report.

SECTION 1
INTRODUCTION

Computer models are becoming increasingly popular as management tools to aid the decision-maker in balancing the complex, interacting factors which must be considered in public health risk analysis. Dramatic advances in microcomputer technology that have placed small, high powered, machines in the hands of almost all decision-makers have greatly enhanced this trend (Heidtke et al. 1986). Inherent in the analysis of public health impacts is the concept that natural systems are a complex combination of deterministic and stochastic processes. However, most available environmental models are deterministic, that is, they require selection of specific values of input parameters such as reaction rates or growth rates and produce single-valued predictions of output variables in time and/or space. Hence, most existing environmental models are of limited value to decision makers because a purely deterministic structure is used to model processes that are intrinsically stochastic. In those cases where the public health is involved, it is essential that the probabilistic aspect of natural systems be included in the model that the decision maker uses in analyzing potential risk to the public. However, the application of existing stochastic methods to environmental models is sometimes difficult, if not impossible. In some cases, the results of a given method may not give enough information to truly be called a stochastic method. In those cases where stochastic information can be derived, the results may be difficult to interpret.

1.1 Objectives of Project

Because of the limitations in existing methods used to include stochasticity in computer models, a project was initiated at the University of California, Davis, aimed at developing a stochastic technique which would provide the mathematical rigor of stochastic differential equation approach (a mathematical technique to implicitly incorporate stochastic terms into the controlling differential equations) while retaining the straight forwardness and relative ease of use of Monte Carlo methods.

1.2 Characteristics of New Technique

In order to serve as a useful tool for modelers, the method should have three distinct characteristics:

1) It should provide analytical expressions for the moments and the probability density functions of key output variables.

2) The analytical expressions for the variance of the output variables (second moments), should give detailed, quantitative information about the individual and combined effects of input variances on variance in the outputs;

3) It should offer a tool through which, in certain modelling situations, an 'optimum' model could be selected based on the trade-off between the increased accuracy of adding more state variables versus the addition of inherent fluctuations which also accompanies such model extensions.

1.3 Organization of Report

The report is organized into five sections. The first section, the Introduction, provides a description of the justification and objectives of the project. Section 2 provides a detailed literature review and discussion of the background of stochastic modelling, with special emphasis on applications to environmental and risk analysis models. Section 3 gives details about the technical approach used in the development of the new technique and provides a 'step-by-step' guide in the application of the technique. Section 4 demonstrates the application of the new technique to health risk analysis. The principal conclusions, a discussion of the advantages and limitations of the new technique, and recommendations for future research are presented in Section 5.

SECTION 2
BACKGROUND OF STOCHASTIC METHODS IN ENVIRONMENTAL MODELLING

The first attempts to include natural variability in environmental models involved statistical extrapolation to predict DO values in streams after organic waste discharges (LeBosquet and Tsivoglou 1950). Bulmer (1957) extended statistical approaches to find general estimates of confidence limits. Statistical approaches however, are limited in capability to handle the complex variations present in natural systems. A more powerful approach is to encode the current state of knowledge about the variations affecting a process directly in a mathematical expression. Using this approach, the probability density function of the outcome becomes the "fundamental entity in the problem" (Moore and Brewer 1972). There are currently three "probabilistic" methods which are commonly used to include stochasticity in mathematical modelling: functional analysis, the Monte Carlo technique, and the stochastic differential equation approach.

2.1 FUNCTIONAL ANALYSIS

The title "functional analysis", as used in this report, refers to any method which involves the use of an assumed function to approximate the mean, variance and/or higher order moments of an output variable (Y) as a function of one or more input variables (X). This grouping encompasses several closely related methods identified in the literature as error analysis, uncertainty analysis, or confidence interval development. A representative example of a functional analysis method is 'First-Order Uncertainty Analysis' (Benjamin and Cornell 1970; Cornell 1972). This method involves the use of a Taylor Series Expansion or perturbation equation to represent the functional relationships of the state variables. It is then assumed that any random component in the system can be completely defined as a normal variation around a zero mean. The Taylor Series is then truncated after the first-order term, hence the name "First-Order Uncertainty Analysis". In the single variable case this procedure produces the following equation:

$$Y(t) = L[x(t)] + \left(\frac{dL}{dX}\ [X(t) - x(t)]\right) \qquad (2\text{-}1)$$

where: L[x(t)] is the operator of the system (linear or non-linear)
x(t) is the mean value
$\left(\frac{dL}{dX}\ [X(t) - x(t)]\right)$ is the convolution of the system function and the random process.

2.1-1 Applications in Environmental Modelling

The statistical approaches of LeBosquet and Tsivoglou (1950) and Bulmer (1957) appear to be the first attempts at applying the theories of functional analysis to environmental models. First-order analysis appears to have first been applied in environmental modelling by Burges and Lettenmaier (1975). Because first-order type analyses give an approximation for the effects of input variance on the fluctuation in output variables, they have been used extensively to examine the effects of error propagation on model output (O'Neill et.al. 1980; Gardner et al. 1980a and b). More recently, First-Order Uncertainty Analysis has been applied to a water quality model to provide an estimate of the sensitivity of output to selected input parameters (Brown 1986).

This approach has the advantage of only requiring estimates of the mean and variance (first and second moments) of the input parameters. In situations where limited information restricts analysis of the random components first-order analysis is a useful tool. In some cases, first-order analysis gives answers comparable to those obtained by more complex analysis procedures (Brown 1986). However, in more complicated situations, there are discrepancies between first-order analysis and nonlinearized methods such as a Monte Carlo technique (Scavia et al. 1981). Other functional analysis approaches, such as statistical estimation of moment generating equations (Karmeshu and Lara-Rosano 1987), point estimation techniques (Thorburn 1986), or numerical analysis (Dresnack and Dobbins 1968) are also usually limited to first and second moments.

2.1-2 Limitations of Method

Functional analysis techniques are in general, limited to generation of only the mean and variance of the output variable. Complete description of a distribution requires definition of all higher order moments. Thus, these methods do not entirely describe the distribution except in the

special case where the output is normally distributed. First-order error analysis can be extended by truncating the Taylor Series expansion after the second or third order, but carrying higher order terms complicates the mathematics. Some work has been done on relaxing the first-order approximation restrictions in other ways (Tung 1987) but little has yet been reported in the literature.

2.2 Monte Carlo Methods

Monte Carlo methods were developed in the 1940's as a numerical tool to solve complex sets of equations which were beyond the computational power available at the time (Ulam and von Neumann 1945; Metropolis and Ulam 1949). Originally, the method was used as a numerical technique to solve differential and integral equations (Meyer 1956). An excellent example of this application which elucidates the basic principles of Monte Carlo is given in a paper by Kahn (1950). Consider the problem of evaluating the following integral:

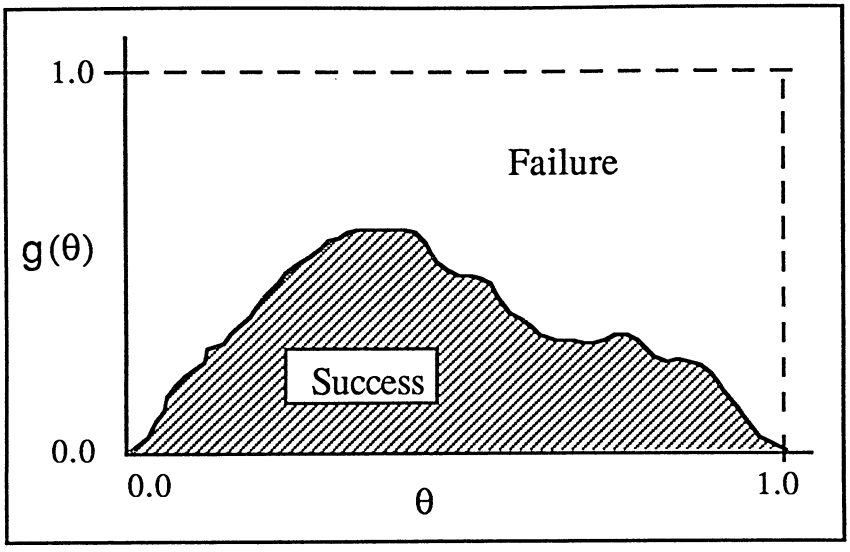
$$\int_0^\infty g(x)dx \qquad 0 < g(x) < 1 \tag{2-2}$$

The solution is to be determined using random sampling.with the following rules:

1) x and y are selected at random from a population uniformly distributed between 0 and 1; and
2) if y is less than g(x), the game is a success, otherwise it is a failure.

Figure 2-1 shows that the probability of success is equal to the area under the curve g(x). Therefore an estimate for the solution of equation 2-2 can be obtained by r/N, where r is the number of successes and N is the total number of trials. This example highlights the two basic traits of Monte Carlo methods: random sampling from a specified distribution and the use of that sample in a specified equation.

FIGURE 2-1
APPLICATION OF MONTE CARLO TECHNIQUE
(From Kahn 1950)



2.2-1 Applications in Environmental Modelling

Monte Carlo methods are premised on random sampling of a specified set. When used as a numerical technique in integration, the specified set is the set of possible solutions to the equation(s) over the range being examined. In environmental modelling, Monte Carlo methods are used to randomly select input values or equation parameters. Consider a model which gives a reaction rate as a function of temperature. The most common form of this is a modified form of the Arrhenius relationship:

$$K(T) = K_r \, A^{(T-T_r)} \qquad (2\text{-}3)$$

where: T = temperature of interest
T_r = reference temperature (usually 20°C)
A = constant
$K(T)$ = reaction rate at temperature T
K_r = reaction rate at reference temperature T_r

In natural systems, temperature is a stochastic variable. For this case, assume that the temperature varies normally around a mean ($\bar{T}$) with a standard deviation of σ_T. We wish to know the distribution of possible K values given this temperature variation. Application of Monte Carlo to this problem involves four steps:

1) A random number (ξ), normally distributed with a mean of zero and a standard deviation of one, is generated. The random number is normally distributed because the input variable is assumed to be normally distributed.

2) The random number is used to select an input temperature (T_i) using the following equation:

$$T_i = \bar{T} + \xi \sigma_T \qquad (2\text{-}4)$$

3) The reaction rate K(T) which results from temperature T_i is calculated using equation 2-3.

4) This process is repeated several times and the varying realizations of K(T) stored. After a sufficient number of repetitions, the probability density function of the resulting K(T) values can be computed.

Because the Monte Carlo technique provides a straight-forward way of including stochasticity in environmental models, it is probably the most widely used stochastic method. The procedure can easily be extended to several variables simultaneously. Input variables and or parameters can be linked by predicating the selection of one variable upon the random outcome of another (Tumeo and Orlob, 1986).

Monte Carlo techniques have been used to model environmental systems ranging in complexity from first-order decay of organic chemicals in a river (Kothandaraman and Ewing 1969; Burges and Lettenmaier 1975) to highly diversified biological communities (Tiwari and Hobbie 1976; Tiwari et al. 1976; Furness 1978). The method has also been used in ocean plume dispersion analysis (Orlob and Tumeo 1986; Orlob et al. 1987), lake models (Fedra 1979), predator-prey models (Gardner et al. 1980b), and ground water pollution models (Smith and Schwartz 1981a and b; Black and Freyburg 1987). Results of a Monte Carlo simulation are also the usual standard to which other methods are compared (van de Kramer 1983; Malone et.al. 1983; Brown 1986).

2.2-2 Limitations of Method

While the Monte Carlo technique is straight forward, it is not without its weaknesses and limitations. Random number generation, the basis of the method, is not an easy process, and can be troublesome and even unreliable, especially on smaller computers. Random number generators usually produce a set of uniformly distributed values between 0 and 1. If some other distribution for the input variable is required, then the distribution of the random number must be the transformed. Detailed discussion of the mathematics of random variable transformations may be found in Benjamin and Cornell (1970).

Another central issue in Monte Carlo techniques is determination of the number of times to run the model. Enough runs must be performed to obtain a statistically valid sample. The number required is usually a function of the type and complexity of the output distribution, which is not known a priori. In practice, a certain number of trials are executed (e.g. 300) and the resulting statistics calculated. The process is then repeated with a larger sampling (e.g. 600); new statistics calculated and compared to the previous run. If there is no 'significant' difference, then 300 runs are sufficient. If there is a significant difference, the number of runs are increased and the new statistics compared to the run of 600. This process is repeated until no 'significant' difference is found between successive runs. 'Statistical significance' can be determined by any one of several standard statistical tests for comparison of distributions (Benjamin and Cornell 1970).

2.3 Stochastic Differential Equation Technique

The development of the stochastic differential equation technique began around the turn of the century with Einstein's classic solution to the problem of Brownian motion (Einstein 1905).

$$\frac{\partial f(x,t)}{\partial t} = D \frac{\partial^2 f(x,t)}{\partial x^2} \tag{2-5}$$

where f(x,t) is the number of particles per unit volume and D is the coefficient of diffusion. Einstein derived the well known differential equation by considering the probability of a particle being at some point 'x' at time $t + \Delta t$, given its position at time t. This approach was unique in that it gave a boundary condition that arose from the probabilistic development of the problem:

$$\int_{-\infty}^{\infty} f(x,t)\, dx = n \tag{2-6}$$

where n is the total number of molecules in the system. Hence, the problem of diffusion outward from a point was totally mathematically defined. Einstein reported the solution as:

$$f(x,t) = \frac{(n)\exp(-x^2/4Dt)}{4\pi Dt} \tag{2-7a}$$

$$\text{where: } D = \frac{RT}{6\pi NkP} \tag{2-7b}$$

and: R = gas constant
T = temperature
P = radius of spherical molecule
N = number of molecules per mole
k = coefficient of viscosity

A similar development was derived independently by von Smoluchowski (1906) about the same time. However, von Smoluchowski's solution differed by a factor of 64/27. In later work, Langevin (1908) used Stoke's equation and statistical mechanics to rederive the work of von Smoluchowski and rectified the difference between Einstein's and von Smoluchowski's solutions. However, aside from simple extensions to analogous diffusion processes there were few attempts to use the techniques developed by Einstein, von Smoluchowski, and Langevin until the late 1940's. The main reason for the restricted application of stochastic differential equations before this time was the lack of mathematical procedures required to solve the equations. The only mathematics available to early researchers was Riemann Calculus. However, Riemann Calculus is not applicable to stochastic differential equations. Adequate mathematical grounding for work with stochastic differential equations was not available until the work of Ito (1944, 1946, 1951).

2.3-1 Applications in Environmental Modelling

The majority of work in the application of the theory of stochastic differential equations to environmental modelling has centered on the Streeter-Phelps equations (Streeter and Phelps 1925). The first well known work is probably that of Loucks (Loucks 1965; Loucks and Lynn 1966). In this work, the theory of Markov Chains, which had already been successfully applied in the analysis of hydrologic phenomena was extended to prediction of BOD and DO. Subsequent theoretical development of the application of stochastic differential equations to the Streeter-Phelps equations was carried out by researchers such as Thayer and Krutchkoff (1967), and Kothandaraman and Ewing (1969). More extensive stochastic treatment of the Streeter-Phelps equations was performed by Padgett and co-workers (Padgett 1975; Padgett and Durham 1976;

Padgett et al. 1977a and b; Padgett and Papadopoulos 1979). There has been only limited application of the technique beyond the Streeter-Phelps equations. Harris (1975) worked with stochastic development of composite water quality models. Tiwari and co-workers (1976) developed a set of stochastic differential equations for a complex ecological system, but did not attempt an analytical solution for the probability density functions of the state variable.

2.3-2 Limitations of Method

Stochastic differential equations continue to receive much attention in stochastic hydrology (Yevjevich 1987) and precipitation analysis (Kavvas et al. 1987), but the method does not appear to have been applied with great success in environmental modelling outside of treatment of the Streeter-Phelps equations.The main reason for the limited application of stochastic differential equations to date remains the complexity of the mathematics. As discussed above, Riemann Calculus is not applicable to the solution of stochastic differential equations. Instead, Ito Calculus or a similar stochastic calculus must be used. To date these calculi have only been developed for ordinary differential equations rather than for the partial differential equations employed to represent complex environmental systems. While it is possible to convert partial differential equations into ordinary differential equations, the result is usually a non-tractable set of non-linear equations. Hence, application of stochastic differential equations continues to be limited in environmental modelling.

2.4 STOCHASTIC MODELING IN HEALTH RISK ANALYSIS

2.4-1 DOSE RESPONSE DISTRIBUTIONS

Basic work in the area of stochastic modelling of health risks has been limited due to the inherent difficulties associated with predicting risk to a given population. The first step in modelling health response is determination of the dose of the 'contaminant' received. The 'contaminant' can be a pathogen, a disease causing chemical, or any other substance which may produce a response when encountered. The method of entry into the body may be oral (ingestion),

airborne (inhaled), or adsorption through the skin. The controlling equation most often used used to determine the dose, regardless of the route of entry, is:

$$D = V\ C \tag{2-8}$$

where: D = dose ingested (units depend on contaminant and route of entry)
V = volume ingested, inhaled, or adsorbed
C = mean concentration of contaminant
(units depend on contaminant and route of entry)

If the process is to be modeled stochastically, each variable in equation 2-8 must be prescribed to vary with a specified probability density function. However, calculation of the distribution of dose that a person may receive does not address the question of the risk of illness. The important question is actually the nature of the response of an individual once a 'contaminant' is introduced into one's system. This is commonly called the dose-response behavior of the individual.

Development of human dose-response relations is extremely difficult and complicated. The first difficulty is identification of the great number of possible contaminants and infectious agents. A second problem is the identification of the actual dose required which results in a response. As the ability to measure minute concentrations of contaminants in all media has improved, serious questions have arisen concerning the risk associated with chronic exposure to low concentrations of potential mutagens and carcinogens. Low concentrations of 'contaminants' can also cause acute responses. Very low concentrations of viruses or bacteria, even single pathogens, can cause infection (Hass 1983). Yet very low concentrations of most pathogens cannot be detected by many standard microbial tests. In some instances, indicator organisms are used to avoid problems in detection of other, more infectious organisms that may occur in extremely low concentrations. In water for example, the concentration of the indicator organism E. Coli is often used and related to the expected concentration of a more infectious pathogen (Kehr and Butterfield 1943), but the actual concentration of the pathogen may still be unknown.

It is generally impractical to set up experiments to test the response of humans to various infectious, mutagenic and carcinogenic agents. What little data do exist are derived from laboratory studies of other mammals (e.g. Plotkin and Katz 1965) or epidemiological studies of areas in which

humans were exposed to some known contamination source (e.g. Mosley 1965; McLean 1965; Geldreich 1972). Hass (1983) identified three common dose-response equations from data available in the literature:

Beta $$\theta = 1 - (1 + \frac{D}{b})^a \quad (2\text{-}9a)$$

Exponential $$\theta = 1 - e^{-rD} \quad (2\text{-}9b)$$

Log-Normal $$\theta = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(Z^2/2)\, dZ \quad (2\text{-}9c)$$

where: $$Z = \frac{(D - \mu)}{\sigma}$$

μ = geometric mean of lognormal distribution (empirical)

σ = standard deviation of lognormal distribution (empirical)

θ = probability of response (illness or disease)

D = dose received

a,b, and r = empirically derived constants.

Hald (1952) proposed a possible logistic dose-response equation of the form:

Logistic: $$\theta = \frac{1}{1 + \exp(-\{M + N \log(D)\})} \quad (2\text{-}9d)$$

where M and N are empirically derived constants and D and θ are as defined above.

Once a specific dose has been determined, the above equations can be used to calculate the associated probability of illness. Given the probability density function for dose received, a corresponding probability density function for the response can be generated. While this will allow calculation of the risk to one individual, it is still not sufficient to determine the risk to the total population. The number of people who become ill from a given exposure is a result of a convolution of the probability of the number of individuals within the population being exposed and the subsequent probability of response due to that exposure. Mathematically, the risk for a given population size may expressed as:

$$\text{Risk}(x) = \int_{-\infty}^{\infty} f(x|\theta)\ g(\theta)\ d\theta \qquad (2\text{-}10)$$

where $f(x|\theta)$ is the marginal distribution of illness for a set number of individuals ('x') given the probability of response (θ), and $g(\theta)$ is the probability density function for response.

2.4-2 Review of Cooper Model

Work in modelling of health risks from water borne diseases was first developed in a report by Cooper and co-workers (Cooper et al. 1983a). Interest in the project was based in part on the high incidence of illness and death from water-related diseases throughout the third world countries of Africa, Asia and Latin America. Table 2-1 shows the prevalence of major water related diseases in these areas.

TABLE 2-1
PREVALENCE OF MAJOR WATER-RELATED DISEASES IN AFRICA, ASIA, AND LATIN AMERICA
(Adapted from work by Walsh and Warren (1979))

INFECTION	PREVALENCE[a]	DISEASE[b]	DEATHS
Typhoid	1,000	500	25
Poliomyelitis	~80,000	15,000	1,200
Schistosomiasis	~200,000	20,000	500 to 1,000
Amebiasis	~400,000	1,500	30
Diarrhea	3 to 5 million	3 to 5 million	5,000 to 10,000
Ascariasis	0.8 to 1 million	1,000	1,000

[a] Estimated number of exposed individuals

[b] Number of individuals demonstrating symptoms of disease

The work of Cooper et al. resulted in a Monte Carlo model to analyze and quantify the risk of becoming ill or contracting an infectious disease from ingesting contaminated water. A modified version of equation 2-8 was used that incorporated a term (E) for the efficiency of any water treatment method which might be employed. Parameters in equation 2-8 were specified using a

normal distribution for volume ingested (V), a lognormal distribution for concentration of pathogens (C), and a uniform distribution for treatment efficiency (E_T). The model randomly selected a value for each of the variables from the specified distribution and calculated the dose ingested. This procedure was repeated several times and the probability density function of the dose ingested developed. To allow the calculation of risk, two additional assumptions were made:

1) The dose-response behavior was represented by a Beta distribution (equation 2-9a above); and

2) The marginal distribution of illness for a set number of individuals [$f(x|\theta)$] was assumed to be binomial and was modeled by:

$$f(x|\theta) = \binom{x}{n}\theta^x(1-\theta)^{n-x} \qquad (2\text{-}11)$$

where n is the total population exposed and x is the number of ill individuals.

The statistics calculated from the Monte Carlo simulation were used to estimate the parameters of the beta distribution. As a result, it was possible for the authors to analytically derive the following function for the risk of illness:

$$\text{Risk}(x) = \binom{x}{n}\frac{\beta(p+x,n+q-x)}{\beta(p,q)} \qquad (2\text{-}12)$$

where: $\beta(p,q) = \dfrac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$

$\Gamma(q)$ = gamma function
n = total population examined
x = ill population
Risk(x) = probability that x individuals will become ill
p,q = parameters of Beta distribution derived from Monte Carlo simulation of dose-response equation

SECTION 3
TECHNICAL APPROACH FOR A NEW STOCHASTIC TECHNIQUE

Because of the limitations in existing methods used to include stochasticity in computer models, a project was initiated at the University of California, Davis, aimed at developing a stochastic technique which would provide the mathematical rigor of stochastic differential equation approach (a mathematical technique to implicitly incorporate stochastic terms into the controlling differential equations) while retaining the straight forwardness and relative ease of use of Monte Carlo methods. The approach used to develop the new technique focused on the derivation of distribution functions for output parameters through a mathematical combination of stochastic variations in the input parameters. A two-tiered process is involved: first, mathematical expressions for the moments (i.e. the mean, variance, etc.) of the output variables are derived from the original equations which define the process; and second, these moments are used in conjunction with the Fokker-Planck equation (Fokker 1914; Planck 1917) to derive an analytic solution for the probability density functions of the selected state variables.

3.1 DEVELOPMENT OF TECHNIQUE

To present the development of the technique, consider a first-order differential equation of the form:

$$\frac{d\,C(t)}{dt} = -\,k\;C(t) \qquad (3\text{-}1a)$$

for which one solution is:

$$C(t) = C_0 e^{-k(t)} \qquad (3\text{-}1b)$$

where: $C(t)$ = concentration at time t (mg/l)
C_0 = initial concentration (time = 0)
k = decay rate (per day)

The basis of the derivation of the new technique is the realization that the parameters in the state equation have stochastic properties which must be carried forward in the solution process. In the

above example, the reaction rate constant k and the initial concentration C_0 are treated as stochastic variables with mean values and deviations about their respective means, yielding:

$$\frac{d[\overline{C(t)} + C(t)']}{dt} = -(k + k')[\overline{C(t)} + C(t)'] \qquad (3\text{-}2)$$

where: $\overline{C(t)}$ = mean concentration at time t (mg/l)
k = mean decay rate (per day)
k' = variation of k around mean
$C(t)'$ = variation of C around mean

The idea of writing the equation in terms of a mean and a 'deviation' is also used when developing a stochastic differential equation. However, in the stochastic equation, the 'deviation' is represented by a 'Gaussian White Noise' term (Gardiner 1985; Zielinski 1988). Terms in equation 3-2 which involve a variation (k' or C(t)') are separated from those terms which contain only mean values to give two equations.

$$\frac{d\,\overline{C(t)}}{dt} = -k\,\overline{C(t)} \qquad (3\text{-}3a)$$

$$\frac{d\,C(t)'}{dt} = -k'\,\overline{C(t)} - k\,C(t)' - k'C(t)' \qquad (3\text{-}3b)$$

Equation 3-3a is solved independently yielding:

$$\overline{C(t)} = C_0 \exp\{-kt\} \qquad (3\text{-}4a)$$

Substituting equation 3-4a for $\overline{C(t)}$, equation 3-3b is solved yielding:

$$C(t)' = (C_0 + C_0')\exp\{-kt\}\exp\{-k't\} - C_0\exp\{-kt\} \qquad (3\text{-}4b)$$

where: C_0 = mean initial concentration (time = 0)
C_0' = random variation in initial concentration

Equation 3-4b is a mathematical expression for the deviation of the process [$C(t)'$] around the mean [$\overline{C(t)}$] as a function of the random variations of the input parameters.

3.2 Calculation of Moments of a Distribution

Equations 3-4a and 3-4b can be used to determine the moments of the distribution of the output variable. The first moment of the process, the mean, is found directly from equation 3-4a. Higher order moments are by definition:

$$M[f(x)] = E\left\{ f(x) - E[f(x)] \right\}^n \tag{3-5}$$

Where E, the 'expectation operator', operating on f(x) is the mathematical operation which produces the expectation (mean) of the function f(x). The second moment, the variance, is found by setting n = 2. Discussion of the expectation operator can be found in textbooks on probability (e.g. Bain and Engelhardt 1987).

The actual value of the expectation depends on the type of distribution. If, for the example case above (equations 3-2 through 3-4), we assume that k' and C_0' are normally distributed with a mean of zero and a standard deviation of σ_K and σ_C, respectively, the variance of the concentration is found to be:

$$E[C'^2] = \sigma_{C_o}^2 \exp\{-2kt\} \exp(2t^2\sigma_k^2) + C_0^2 \exp\{-2kt\}\left(\exp(2t^2\sigma_k^2) - \exp(t^2\sigma_k^2)\right) \tag{3-6}$$

where σ_k is the standard deviation of k (per day), and σ_{C_o} is the standard deviation in initial concentration (mg/l).

A similar development is conceptually possible for any differential equation or set of differential equations. In addition, unlike other stochastic methods, it is not necessary to assume Gaussian distributions. Any distribution for which a second moment exists can be used in this technique. Complex random functions of time and space and may be included and may be mutually covariant if the form of the covariance matrix, or more exactly the expectation of the covariance matrix, is known. The method also provides a technique to derive explicit equations for all moments of the output variable as a function of the input variables. Higher order moments such as skew and kurtosis can be calculated using the appropriate power of "n" in equation 3-5 above.. For example, the third moment - or skew - can be found by setting n equal to three, yielding:

$$S[f(x)] = E\left\{ f(x) - E[f(x)] \right\}^3 \tag{3-7}$$

The probability density function of a variable is totally characterized by its moments. However, a given function may have an infinite number of moments, although in general, moments higher than the third or fourth rarely add to the description of the probability density function. In addition, calculation of higher order moments is increasingly complicated algebraically beyond the second or third moment. What is desired in addition to the moments is a general expression for the probability density functions of the state variables.

3.3 Calculation of Probability Density Functions

At this point, the technique draws upon the work of Kolmogoroff (1931), who demonstrated that the Fokker-Planck equation is a valid representation of the conditional probability of a process $P(\bar{x},t \mid \bar{x}_o,t_o)$, such that:

$$\frac{\partial P(\bar{x},t \mid \bar{x}_o,t_o)}{\partial t} = -\sum_{i=1}^{\infty} \frac{\partial}{\partial x}\{A_i(\bar{x},t)P(\bar{x},t \mid \bar{x}_o,t_o)$$

$$+\frac{1}{2}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{\partial^2}{\partial x^2}\{B_{ij}(\bar{x},t)P(\bar{x},t \mid \bar{x}_o,t_o)\} \qquad (3\text{-}8)$$

where $\bar{x}$ is a vector set of state variables, and $P(\bar{x},t \mid \bar{x}_o,t_o)$ is the probability of $\bar{x}$ at time t given the vector of initial conditions $\bar{x}_o$, at time t_o. $A_i(\bar{x},t)$ and $B_{ij}(\bar{x},t)$ can be physically interpreted as representing the rate of change in the general trend or mean of the process (the drift) and the rate of change in the variation of the process around the mean (the deviation), respectively. Kolmogoroff also demonstrated that the following conditions apply:

$$P(\bar{x},t \mid \bar{x},t_o) \quad \text{at } t_o = 0 \; = f(x,0) \qquad (3\text{-}9a)$$

$$\lim_{x \to \infty} P(\bar{x},t \mid \bar{x},t_o) = 0 \qquad (3\text{-}9b)$$

Solution of the Fokker-Planck equation under the initial and boundary conditions shown in equations 3-9a and 3-9b yields the conditional probability density function for the variables in question. If we assume that the characteristic moments of the distribution derived using the

techniques just discussed (see equations 3-4a and 3-6) can be used to represent the rates of change of the mean and variance in the Fokker-Planck equation, we then have a functional form for A_i ($\bar{x}$,t) and B_{ij} ($\bar{x}$,t).

$$A_i(\bar{x},t) = \frac{d(\text{Mean})}{dt} \tag{3-10a}$$

$$B_{ij}(\bar{x},t) = \frac{d(\text{Variance})}{dt} \tag{3-10b}$$

In the example case, the coefficients are functions of time only and the state variable vector (x) is one dimensional (concentration). Successful application of the new technique does not require this simplification. However, for ease of demonstration, a simple example case was selected. Equation 3-8 therefore reduces to:

$$\frac{\partial P(C(t),t \mid x_o,t_o)}{\partial t} = -A(t)\frac{\partial P(C(t),t \mid x_o,t_o)}{\partial x} + \frac{1}{2}B(t)\frac{\partial^2 P(C(t),t \mid x_o,t_o)}{\partial x^2} \tag{3-11}$$

Where x_o represents the initial condition of the state variables which affect the concentration, namely C_o and k. Equation 3-11 is an extended form of the general heat equation with probability of concentration substituted for temperature. Solutions for this equation under various conditions abound in the literature (Widder 1975; van Genuchten and Alves 1982; Mohsen and Baluch 1983; Risken 1984; Cannon 1984; Carslaw and Jaeger 1986). Under the conditions given in equations 3-9a and 3-9b and the assumptions presented in equations 3-10a and 3-10b, the solution to equation 3-11 is:

$$P(C(t),t \mid x_o,t_o) = \frac{1}{\sqrt{4\pi Z_2}}\exp\left(\frac{-(C(t)-Z_1)^2}{4Z_2}\right) \tag{3-12}$$

where: $Z_1 = \int_0^\infty A(t)\,dt = C_0 \exp\{-kt\}$

and $Z_2 = \frac{1}{2}\int_0^\infty B(t)\,dt$

$$= \frac{1}{2}\left(\sigma_{C_o}^2 \exp\{-2kt\}\exp(2t^2\sigma_k^2)\right) + \frac{1}{2}\left(C_0^2\left(\exp(2t^2\sigma_k^2) - \exp(t^2\sigma_k^2)\right)\right)$$

What is actually desired, is the one-time probability of any concentration [P(C,t)], which is by definition:

$$P(C,t) = \int_0^{C_{omax}} \int_0^{k_{max}} \frac{1}{\sqrt{4\pi Z_2}} \exp\left(\frac{-(C-Z_1)^2}{4Z_2}\right) \partial C_o \partial k \qquad (3\text{-}13)$$

In addition, P(C,t) must satisfy the normalization requirement:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(C,t)\, \partial C_o \partial k = 1 \qquad (3\text{-}14)$$

The solution of equation 3-14 yields the probability density function for concentration assuming a constant transport velocity, first-order decay, and a normal distribution for initial concentration and variation in the reaction rate. These simplifications are not implicitly necessary for the application of the new technique, but do allow a less complicated mathematical expression to be derived. This method has been applied successfully to the analysis of organic loading in a stream using the Streeter-Phelps Equations (Tumeo and Orlob 1989).

3.4 Quantification of Output Variance

Because the method yields an analytic form for the variance of the output as a function of the variances of the input variables (see equation 3-6), it is possible to use the technique to analyze the effects of variations in the input on the output. This has traditionally been done through sensitivity analysis. However, sensitivity analysis is extremely difficult to interpret when applied to more than one or two variables simultaneously. Through use of the technique presented herein, it is possible to calculate explicitly the effects of variance from all the variables, in all the combinations which are present. Initial investigations in this area has yielded some promise (Tumeo and Orlob 1987), but further research into this application of the technique is needed.

SECTION 4

EXAMPLE APPLICATION OF NEW TECHNIQUE TO HEALTH RISK ANALYSIS

To demonstrate the capabilities of the new technique, it was used to analyze the risk to a specified population of contracting disease from specific water borne bacteria. Results of the analysis were compared to a similar analysis using the Monte Carlo model developed by Cooper et al. (1983a). The comparison to a Monte Carlo technique serves to highlight the advantages and disadvantages of the new technique.

4.1 Derivation of Basic Equations

Both the Monte Carlo method and the new technique were applied to the same basic equation to estimate the dose of a pathogen ingested by drinking contaminated water (a modified form of equation 2-8):

$$D = V\,C\,E_T \tag{4-1}$$

where D is the dose ingested (number of pathogens), V is the volume ingested (ml), C is the mean concentration of pathogens (number/ml), and E_T is the mean efficiency of treatment process.[1]

To apply the new technique, the methodology described in equations 3-3 and 3-4 of Section 3 was followed and the following equations were derived:

$$\bar{\bar{D}} = \bar{\bar{V}}\,\bar{C}\,\bar{\bar{E_T}} \tag{4-2a}$$

$$D' = E_T\,C\,V' + V\,C\,E_T' + V\,E_T\,C' + C\,E_T'\,V' + E_T\,V'\,C' + V\,E_T'\,C' + V'\,E_T'\,C' \tag{4-2b}$$

where $\bar{D}$ is the mean dose of pathogens ingested (number), D' is the random fluctuation in dose around mean (number), $\bar{V}$ is the volume of water ingested (ml), V' is the random fluctuation in volume around mean (ml), $\bar{\bar{E_T}}$ is the mean efficiency of waste treatment, E_T' is the random fluctuation in efficiency around mean, $\bar{C}$ is the mean concentration of pathogen in discharge (#/ml), and C' is the random fluctuation in concentration around mean (#/ml).

---

[1] $E_T = (1-\eta)$ where η = removal efficiency.

Note that dispersion and die-away of the pathogen in the environment, or chemical decay in the case of a reactive contaminant, is not considered in this case. This assumption is not required for application of a Monte Carlo method or the new technique, but is used here to remain consistent with the assumptions of the model developed by Cooper et al.

4.1-1 Derivation of Moments of the Distribution

Following the approach outlined in Section 3, the first moment of the probability function of the dose is:

$$\text{Average Dose} = \overline{D} = \overline{V}\,\overline{C}\,\overline{E_T} \tag{4-3}$$

Derivation of the second moment of the probability density function of the dose involves squaring equation 4-2b and taking the expectation of each term. The final form of the second moment equation will depend on the types of distribution assumed for each variable. The derivation of the general equation for the second moment of the dose probability density function is shown in Appendix A.

4.1-2 Derivation of Probability Density Functions

4.1-2a *Density Function for Dose:* Following the development presented in Section 3, the solutions for the first and second moments of the probability density function for dose are used as the coefficients in the Fokker-Planck equation. Solution of this equation, integrated over all possible input parameter values, yields the probability density function for dose:

$$P(\text{Dose}) = \int_0^{E_{max}} \int_0^{V_{max}} \int_0^{C_{max}} \frac{1}{\sqrt{4\pi Z_{2dose}}} \exp\left(\frac{-(D-Z_{1dose})^2}{4Z_{2dose}}\right) \partial E \partial V \partial C \tag{4-4}$$

The Z_1 value for dose is given directly by equation 4-3. The Z_2 value of the process can be found by calculating the expectation of the square of the fluctuation term as shown in Appendix A.

4.1-2b *Density Function for Response:* The two models developed for this case study allowed the use of any of the four dose-response relations discussed in Section 2 (equations 2-9a through 2-9d). Empirical constants were selected based on the relationship used. Because there is a one-to-one mapping of dose to response, the probability of a given response (θ) is the same as the probability of the dose (D) which produces it. Therefore, the analytical solution for the probability density function of response (θ) is the same as the probability density function for dose (Eq. 4-4)

4.1-3 Determination of Input Distributions

All stochastic techniques require estimates of the characteristics of the input distributions. For the purpose of this study, the characteristics shown in Table 4-1 were used.

TABLE 4-1
CHARACTERISTICS OF INPUT VARIANCE

	TYPE OF DISTRIBUTION					
	NORMAL		LOGNORMAL		UNIFORM	
VARIABLE	Mean	Variance	Log Mean	Log Variance	Mean	Range
Volume Ingested (ml)	10	1	2.3	0.0	10.0	9.0 - 11.0
Treatment Efficiency (%)	99.459	0.0025	NA	NA	99.495	99.0-99.99
Concentration (# of pathogens/ml)	172.0	9897	5.15	21.6	172.0	73.0 - 271

The range of the uniform distribution for treatment efficiency and the means and standard deviations for volume ingested and pathogen concentration were drawn from the work of Cooper et al. (1983a and b). While it is not required by either the Monte Carlo method or the new technique, the same characteristic variations were maintained for the different possible input distributions. In this example application, the pathogen selected was Salmonella typhi and a logistic dose-response equation (equation 2-9d) was used. The empirical constants (M = -7.9934; N = 1.9293) were taken from the work of Cooper and his co-researchers.

4.2 Comparison of Monte Carlo Method and the New Technique

4.2-1 Comparison of the Moments of the Dose Distribution

The first two moments of the dose distribution under the various combinations of inputs were calculated using the new technique and the Monte Carlo model. The various cases examined are shown in Table 4-2. The results of these calculations are shown in Table 4-3.

TABLE 4-2
CASES EXAMINED FOR MOMENT ANALYSIS

Case Number	TYPE OF DISTRIBUTION USED: Volume Ingested	Pathogen Concentration	Treatment Efficiency
1	Normal	Lognormal	Uniform
2	Normal	Normal	Uniform
3	Lognormal	Uniform	Normal
4	Lognormal	Lognormal	Normal
5	Uniform	Normal	Uniform
6	Normal	Normal	Normal
7	Uniform	Uniform	Uniform
8	Lognormal	Lognormal	Uniform

TABLE 4-3
COMPARISON OF MOMENTS OF DOSE DISTRIBUTION

Case	First Moment (Mean) Monte Carlo	First Moment (Mean) New Technique	Second Moment (Variance) Monte Carlo	Second Moment (Variance) New Technique
1	2365.9	342,000	4.88×10^{8}	2.42×10^{20}
2	9.21	8.69	62.1	58.5
3	9.47	9.28	75.6	3.53×10^{4}
4	2261.1	732,098	2.50×10^{7}	7.66×10^{23}
5	13.46	8.69	57.5	57.6
6	11.60	8.69	131	125.0
7	9.89	8.69	58.7	355.0
8	3938.0	683,380	1.65×10^{8}	7.14×10^{23}

At first glance, it appears that there is a discrepancy between the Monte Carlo method and the new technique. Upon inspection however, it can be seen that the only large differences are in those cases which involve a lognormal distribution, most notably, cases 1, 3, 4, and 8. However, in the comparison of the cumulative density functions for both dose and response for Case 1 (see section

4.2-2 below), the method yields approximately the same distributions as does the Monte Carlo technique with a shift in the distribution towards the larger doses (the probability of a larger dose is higher with the new technique).

This highlights the fact that the new technique is limited in its ability to model a given distribution by the accuracy of the moment equations. This discrepancy will not arise in those instances where the moment estimators accurately track the moments of the distribution, e.g., normal or uniform distributions. However, the moment estimators used to describe lognormal distributions are not as accurate. Aitchison (1957) and Johnson and Kotz (1970) report the following equations for lognormal moment estimators:

$$E[x] = \exp(\xi + 0.5\,\sigma) \tag{4-5a}$$

$$E[x^2] = \exp(2\xi + 2\sigma^2) \tag{4-5b}$$

where ξ and σ are parameters of the distribution. Johnson and Kotz report that the moment sequence represented by this set of equations is not unique to the lognormal distribution -"i.e., the distribution cannot be defined by its moments." These authors point out that estimating the mean and variance of a lognormal distribution becomes increasingly difficult as the variance increases. Finny (1941) reported that the ability of the estimator for the second moment of the lognormal distribution (equation 4-5b) to track the true variance, decreased as a function of σ^2. Inspection of the cases in which the deviations between the new technique and the Monte Carlo method are the greatest reveals that these are the cases with large variances on the lognormal distribution.

4.2-2 Comparison of Probability Density Functions

To further analyze the characteristics of the new technique, the scenario described in Table 4-4 was examined using the new technique and compared to the results of a Monte Carlo simulation of the same scenario. The characteristic values used for the input distributions are the same as those reported by Cooper et al. (1983a) for their example case.

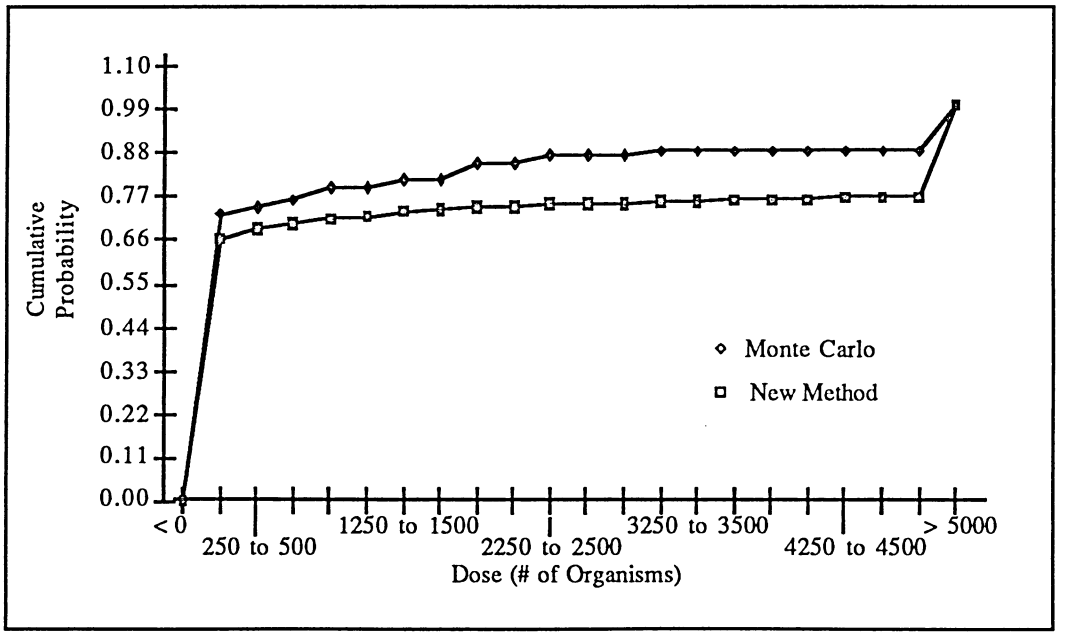
TABLE 4-4

SCENARIO EXAMINED FOR COMPARISON OF CUMULATIVE DENSITY FUNCTIONS

Variable	Type	Characteristic Values
Volume Ingested (ml)	Normal	Average = 10.000 Std Dev= 1.000
Pathogen (#/100 ml) Concentration	Lognormal	Average = 5.150 Std Dev = 4.600
Treatment Efficiency (%)	Uniform	Minimum = 99.000 Average = 99.495 Maximum = 99.990

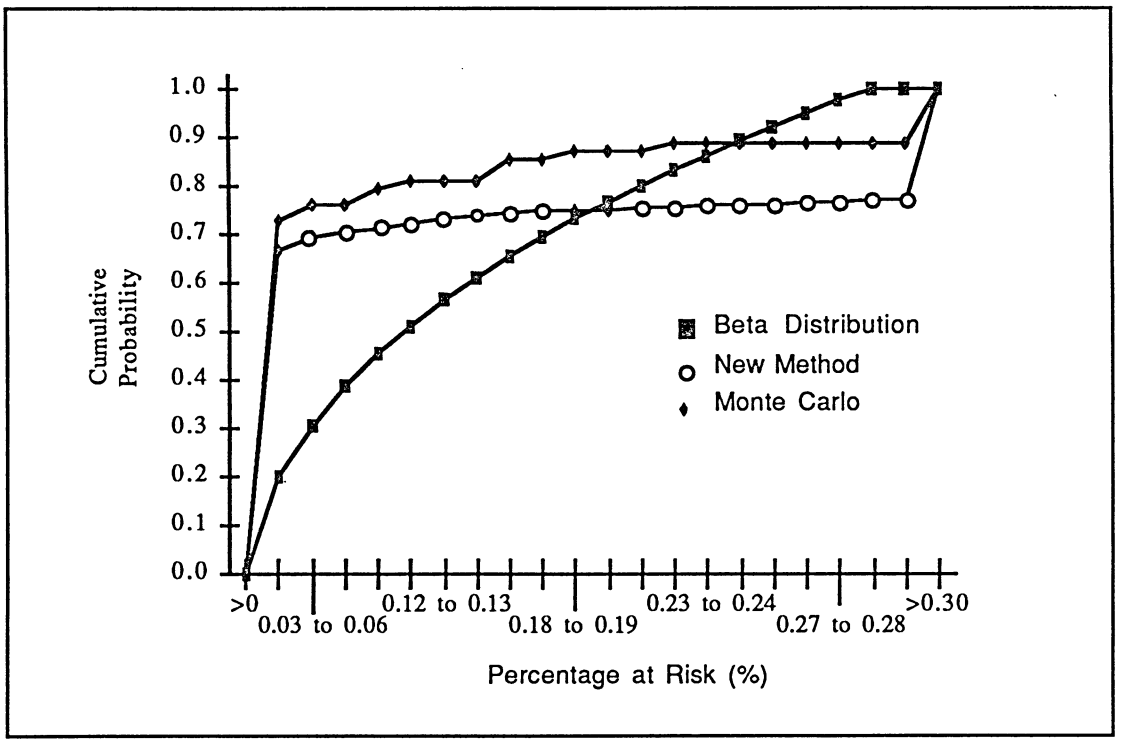
4.2-2a *Dose:* The cumulative density functions for dose predicted by the Monte Carlo model and the new technique are shown in Figures 4-1. The discrepancy in prediction of the moments of the lognormal distribution does not appear to radically affect the calculation of the density functions. The distributions for dose show that the new technique does predict higher probabilities for higher doses. Correspondingly, the cumulative probability density function is lower for the new technique in the lower dose ranges.

FIGURE 4-1
CUMULATIVE DENSITY FUNCTION FOR DOSE



4.2-2b *Response:* Figure 4-2 shows the cumulative distributions for response predicted by the Monte Carlo model, the new technique, and by a Beta distribution using the beta parameters estimated from the Monte Carlo run. As with the dose distributions, the new technique predicts higher probabilities for larger responses. It can also be seen that the assumption of a Beta distribution for the density function of response is inadequate. Consequently, the number of illnesses predicted using the Monte Carlo model and the new technique will not agree because the Monte Carlo model uses the Beta distribution assumption for the marginal distribution of illness given risk.

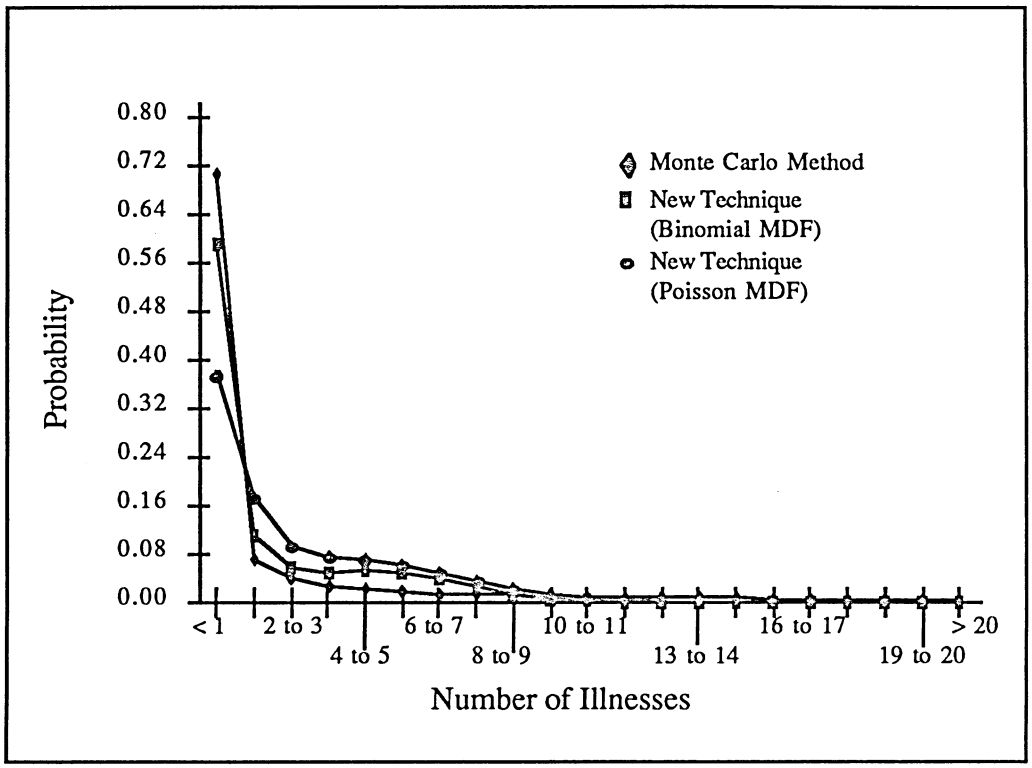
FIGURE 4-2
CUMULATIVE DENSITY FUNCTIONS FOR RESPONSE



4.2-2c *Risk of Illness:* The ultimate goal of both methods is to project the probability of illness to members of an exposed population. Figure 4-3 shows the risk of illness calculated by the two approaches. It could be argued that the new technique gives better predictions since Figure 4-2 indicates the new technique more closely models the density function of the response calculated using the Monte Carlo values. Recall that in the Monte Carlo approach the values of 'θ' [percentage responding] are used only to estimate the values of the beta parameters. However, Figure 4-2

indicates that the Beta distribution assumption may not adequate represent the response to illness. The new technique is not subject to the same limiting assumption of a Beta distribution for the density function of response. A distinct advantage of using the new technique in conjunction with a numerical integration scheme is that different marginal distributions are easily examined. Figure 4-3 shows the risk of illness under two different assumptions as to the form of the marginal distribution ($f[x|\theta]$), binomial and Poisson, as produced by the new technique.

FIGURE 4-3
PROBABILITY OF ILLNESS



4.3 Quantification of Output Variation

The new technique allows the derivation of an analytic representation of the variance of the output variable as a function of the variances of the input variables. It therefore allows the quantitative determination of the individual and compound contributions of variance in the input parameters to the total variance. In the example case, the total variance in the pathogen dose received is a function of the variances in the three input parameters, volume ingested, pathogen concentration, and

treatment efficiency. The equation for variance in dose received derived in Appendix A can be split into seven parts, each of which reveals specific information on the sources of variance:

1) Variance due to volume ingested only (4-6a)

$$VONLY = E_T^2C^2E[V'^2] - E_T^2C^2E[V']^2$$

2) Variance due to treatment efficiency only (4-6b)

$$EONLY = V^2C^2E[E_T'^2] - V^2C^2E[E_T']^2$$

3) Variance due to pathogen concentration only (4-6c)

$$CONLY = E_T^2V^2E[C'^2] - E_T^2V^2E[C']^2$$

4) Variance due to convolutions

4a) volume ingested and pathogen concentration (4-6d)

$$\begin{aligned} VANDC = {} & E_T^2E[V'^2]E[C'^2] + 2E_{TT}^2CVE[V']E[C'] \\ & + 2E_T^2CE[V'^2]E[C'] + 2E_T^2VE[C'^2]E[V'] \\ & + E_T^2E[V']^2E[C']^2 - 2E_T^2CVE[V']E[C'] \\ & - 2E_T^2CE[V']^2E[C'] - 2E_T^2VE[C']^2E[V'] \end{aligned}$$

4b) volume ingested and treatment efficiency (4-6e)

$$\begin{aligned} VANDE = {} & C^2E[E_T'^2]E[V'^2] + 2C^2E_TVE[V']E[E_T'] \\ & + 2E_TC^2E[V'^2]E[E_T'] + 2C^2VE[E_T'^2]E[V'] \\ & - C^2E[E_T']^2E[V']^2 - 2C^2E_TVE[V']E[E_T'] \\ & - 2E_TC^2E[V']^2E[E_T'] - 2C^2VE[E_T']^2E[V'] \end{aligned}$$

4c) treatment efficiency and pathogen concentration (4-6f)

$$\begin{aligned} CANDE = {} & V^2E[E_T'^2]E[C'^2] + 2V^2CE_TE[C']E[E_T'] \\ & + 2V^2CE[E_T'^2]E[C'] + 2E_TV^2E[C'^2]E[V'] \\ & - V^2E[E_T']^2E[C']^2 - 2V^2CE_TE[C']E[E_T'] \\ & - 2V^2CE[E_T']^2E[C'] - 2E_TV^2E[C']^2E[V'] \end{aligned}$$

4d) Variance due to all three (4-6g)

$$\begin{aligned} ALL3 = {} & E[E_T'^2]E[C']E[V'^2] + 4VCE[E_T'^2]E[C']E[V'] \\ & + 4E_TVE[C'^2]E[E_T']E[V'] + 2CE[C']E[E_T'^2]E[V'^2] \\ & + 4E_TCE[C']E[E_T']E[V'^2] + 2VE[V']E[E_T'^2]E[C'^2] \\ & + 2E_TE[E_T']E[C'^2]E[V'^2] + 6VCE_TE[E_T']E[C']E[V'] \\ & + E[E_T']^2E[C']^2E[V']^2 - 4VCE[E_T']^2E[C']E[V'] \\ & - 4E_TVE[C']^2E[E_T']E[V'] - 2CE[C']E[E_T']^2E[V']^2 \\ & - 4E_TCE[V']^2E[E_T']E[C'] - 2VE[V']E[E_T']^2E[C']^2 \\ & - 2E_TE[E_T']E[C']^2E[V']^2 - 6VCE_TE[E_{TT}']E[C']E[V'] \end{aligned}$$

Equations 4-6a through 4-6g can be used to calculate the percentage of the total variance which results from the variance of each input variable, both individually and in all possible combinations with other input variables. To demonstrate this application of the method, equations 4-6a through 4-6g were used to examine the components of variation present in the example case presented in Table 4-4. From Table 4-3, Case 1, it can be seen that this scenario yields a mean dose of 342,000 organisms (log mean = 5.53) and a variance of 2.42×10^{20} (log standard deviation = 10.19). Table 4-5 shows the percentage of the total variance caused by each variable alone and in combination with the other input variables.

TABLE 4-5
COMPONENTS OF VARIATION IN DOSE (by Percentage)
(Total Variance in Number of Organisms Ingested = 2.42×10^{20})

	Volume Ingested	Pathogen Concentration	Treatment Efficiency
Volume Ingested	0.00%	0.75%	0.00%
Pathogen Concentration		74.99%	24.02%
Treatment Efficiency			0.00%

Variation due to all three 0.24%

Table 4-5 shows that the major contributor to the variance in the dose is the variance in the pathogen concentration. Almost 75% of the total variance is caused by the variance in pathogen concentration alone. The remaining 25% results from terms which contain some convolution of the pathogen concentration, most notably, the convolution of the pathogen concentration and treatment efficiency. This type of information could be used to calculate, a priori, a permissible variance in the measurement of pathogen concentration which will assure an output within prespecified confidence limits. Using the above approach, the new technique provides a tool by which the sensitivity of the model to all the input variables can be analytically and simultaneously quantified, without the multiple runs required in traditional sensitivity analysis.

SECTION 5
CONCLUSION

5.1 Advantages and Limitations of New Technique

The technique presented in this report is a potentially powerful tool for extending the capabilities of computer models in risk analysis and environmental quality management. It combines the positive characteristics of existing stochastic methods while eliminating many of their limitations and drawbacks and is capable of providing analytical solutions for the probability density functions and associated moments of important environmental variables.

The new technique allows the derivation of explicit equations for all moments of the output variable as a function of the input variables. A development similar to the one described in Section 3 is conceptually possible for any analytically tractable differential equation or set of differential equations. Unlike other stochastic methods, it is not necessary to assume Gaussian distributions. Any distribution for which a second moment exists can be used as an input distribution in this technique. Complex random functions of time and space and may be included and may be mutually covariant if the form of the covariance matrix, or more exactly, the expectation of the covariance matrix is known. Higher order moments such as skew and kurtosis can be calculated using the appropriate power of "n" in equation 3-5. However, calculation of higher order moments becomes increasingly complicated algebraically beyond the second or third moment.

The restriction that the original differential equations be analytically tractable may present the most severe limitation to the new technique. It is believed however, that a combination of the new technique and various numerical solution techniques will yield satisfactory estimates of the probability density functions for the parameters of interest. Current research is focusing on the application of the method to cases where an analytic solution of the differential equations is not possible. Researchers at the University of Alaska Fairbanks are attempting to use the technique in conjunction with QUAL2E, a stream quality model. This involves work with highly non-linear sets of differential equations which do not have analytical solutions.

Because the method yields an analytic form for the variance of the output as a function of the variances of the input variables, it is possible to predict quantitatively the effects on model response of variations in the input. Sensitivity analysis, the traditional tool for analyzing variances, is extremely difficult to interpret when applied to more than one or two variables simultaneously. However, through use of the new technique, it is now possible to calculate explicitly and simultaneously the effects of variance from all sources, in all the combinations which are present , thus avoiding the difficulties and limitations of traditional sensitivity analysis.

Another important advantage of the new technique is that it allows derivation of an explicit equation for the probability density function of the output. Because the method can be generally applied, there are no limitations as to the distribution chosen to represent the variation of the independent variables outside the requirement that a second moment exist for the distribution. The accuracy of the resulting probability distribution will be limited by the accuracy of the moment estimators used. This limitation is highlighted in the example application presented in Section 4. This problem will not arise in those instances where the moment estimators accurately track the moments of the distribution, e.g., normal or uniform distributions.

5.2 Future Work

The technique presents many possible areas for future work. Among other areas, research is needed to explore: 1) the extent to which the method can be applied to existing water quality and public health risk analysis models; 2) the application of the method to cases where an analytic solution of the differential equations is not possible also needs further exploration; and 3) the application of the technique to the analysis of variance.

Another potential use of the method that results from the derivation of an analytic solution for variance is for optimization of the number of state variables used to model a given process. This idea has been discussed very little in the literature. Costanza and Sklar (1983) developed a technique to compare "articulation" (the number of state variables) to "accuracy" (the percent deviation between the model and a data set). It may be possible to use the method developed in this study to examine the trade-offs between the number of state variables used to model a process and

the incremental increase in variance in the response which results from the addition of each new state variable. By using the analytic solution for variance, it would be possible to calculate the increase in the variance of the prediction induced by the addition of state variables. Additional terms generally add variance to the prediction, even though the intent of adding terms may be to produce a more accurate model. At some point, the increase in accuracy expected by the addition of another state variable will be negated by the increase in variance which that state variable induces. To date, this idea has not been formally tested.

REFERENCES

Aitchison, J., and J. A. C. Brown. 1957. *The Lognormal Distribution*, Cambridge University Press, Cambridge, Massachusetts.

Bain, L. J., and M. Engelhardt. 1987. *Introduction to Probability and Mathematical Statistics,* Duxbury Press, Boston, Massachusetts.

Benjamin, J. R., and C. A. Cornell,. 1970. *Probability, Statistics and Decision for Engineers,* McGraw-Hill Book Company, New York, New York.

Black, T.C., and D.L. Freyburg. 1987. Stochastic modelling of vertically averaged concentration uncertainty in a perfectly stratified aquifer, *Water Resources Research*, **23**(6): 997-1004.

Brown, L. C. 1986. Comparison of uncertainty methodologies using QUAL2E. Abstract in *Abstracts of Technical Papers: Proceedings of the 59th Annual Conference of the Water Pollution Control Federation*, October 5-9, Los Angeles, California, Water Pollution Control Federation, Alexandria, Virginia.

Bulmer, M.G. 1957. Approximate confidence limits for components of variance. *Biometrika*, **44**: 159-167.

Burges, S. J., and D. Lettenmaier 1975. Probabilistic methods in stream quality management. *Water Resources Bulletin*, **11**(1): 115-131.

Cannon, J. R. 1984. *The One-Dimensional Heat Equation.* Vol. 23 of *The Encyclopedia of Mathematics and Its Applications,* G. Rota, editor, Addison-Wesley Publishing Co., Menlo Park, California

Carslaw, H.S. and J.C. Jaeger. 1986. *Conduction of Heat in Solids* (second edition). Clarendon Press, Oxford, England.

Cooper, R.C., et al. 1983a. Assessment of risk associated with water-related infectious agents. In volume I of *Infectious Agent Risk Assessment Water Quality Project,* Sanitary Engineering and Environmental Health Research Laboratory, University of California, Berkeley, California.

Cooper, R.C., et al. 1983b. Factors associated with the transmission of waterborne infectious disease. In volume II of *Infectious Agent Risk Assessment Water Quality Project,* Sanitary Engineering and Environmental Health Research Laboratory, University of California, Berkeley, California.

Cornell, C. A 1972. First-Order analysis of model and parameter uncertainty. Paper presented at the International Symposium on Uncertainties in Hydrologic and Water Resource Systems, Tucson, Arizona, C. C. Kisel and L. Duckstein, ed., University of Arizona, Tucson.

Costanza, R., and F. Sklar. 1983. Mathematical models of freshwater wetland and shallow water ecosystems: An articulated review. Paper prepared for SCOPE International Conference on Freshwater Wetlands and Shallow Water Bodies, Tallinn, USSR, 3-13 August, 1983.

Dresnack, R., and W. E. Dobbins. 1968. Numerical analysis of BOD and DO profiles. *Journal of Sanitary Engineering*, **94**(SA5), pp. 789-807.

Einstein, A. 1905. "On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat. *Annalen der Physik*, Series 4, **17**(4): 549-578.

Fedra, K. 1979. A stochastic approach to model uncertainty: A lake modelling example. Working Paper No. WP-79-63, International Institute for Applied Systems Analysis, Vienna, Austria.

Finney, D.J. 1941. On the distribution of a variate whose logarithm is normally distributed. *Journal of the Royal Statistics Society*, Series B, **7**: 155-161.

Fokker, A.D. 1914. Die mittlere Energie rotierender elektrischer Dipole in Strahlungsfeld. *Annalen der Physik,* Series 4, **43**(5): 810-820.

Furness, R.W. 1978. Energy requirements of seabird communities, A bioenergenics model. *Journal of Animal Ecology*, **47**: 39-53.

Galambos, Janos, and Samuel Kotz. 1978. *Characterization of Probability Distributions* . No. 675 in series *Lecture Notes in Mathematics*, A. Dols and B. Eckmann, eds., Springer-Verlag, New York, New York.

Gardner, R.H., et al. 1980a. Application of error analysis to a marsh hydrology model. *Water Resources Research*, **16**(4): 659-664.

Gardner, R.H., et al. 1980b. Comparative error analysis of six predator-prey models. Ecology, **61**(2): 323-332.

Gardiner, C.W. 1985. *Handbook of Stochastic Methods* (second edition). Volume 13 in Springer Series in Synergetics, Springer-Verlag, New York, New York.

Geldreich, Edwin E. 1972. Waterborne pathogens. In *Water Pollution Microbiology,*. R. Mitchell ed., 207-244. John Wiley & Sons, Inc., New York, New York.

Hald, A. 1952. *Statistical Theory with Engineering Applications*, John Wiley & Sons, Inc., New York, New York.

Harris, C.J. 1975. River quality models near effluent outfalls. In *Mathematical Modelling of Turbulent Diffusion in the Environment,* C.J. Harris, ed., 459-496. Academic Press, New York, New York.

Hass, C.N. 1983. Estimation of risk due to low doses of microorganisms: A comparison of alternative methodologies. *American Journal of Epidemiology*, **118**(4): 573-582.

Heidtke, T.M., M.T. Auer and R.P. Canale. 1986. Microcomputers and water quality models: Access for decision-makers. *J. Water Pollution Control Federation*, **58**(10): 960-966.

Ito, K. 1944. Stochastic Integral. *Proceedings of the Imperial Academy* (Tokyo), **20**: 120-122.

Ito, K. 1946. On a stochastic integral equation. *Proceedings of the Imperial Academy* (Tokyo), **26**: 32-35.

Ito K. 1951. On stochastic differential equations. *Mem. American Mathematical Society*, **4**: 1-51.

Johnson, Norman L. and Samuel Kotz. 1970. *Continuous Univariate Distributions*. Volumes 2 and 3 in *Distributions in Statistics*. John Wiley & Sons, Inc., New York, New York.

Kahn, Herman. 1950. Random sampling (Monte Carlo) techniques in neutron attenuation problems - Part I. *Nucleonics*, **6**(5): 27-33.

Karmeshu and F. Lara-Rosano. 1987. Modelling data uncertainty in growth forecasts. *Applied Mathematical Modelling*, **11**(1): 62-68.

Kavvas, M. L., M. N. Saquib, and P. S. Puri. 1987. On a stochastic description of time-space behavior of extra-tropical cyclonic precipitation fields. *Stochastic Hydrology and Hydraulics*, **1**(1): 37-52.

Kehr, R. W., and C. T. Butterfield. 1943. Notes on the relation between coliforms and enteric pathogens. *Public Health Reports*, **58**(15): 589-607.

Kolmogoroff, A. 1931. Uber die analytischen Methoden in der Wahrscheinlich-keitsrechnung. *Mathmatika Annalen*, **104**: 415-458.

Kothandaraman, V., and B. B. Ewing. 1969. A probabilistic analysis of dissolved oxygen-biochemical oxygen demand relationship in streams. *J. Water Pollution Control Federation*, **41**(2), Part II: R73-R90.

LeBosquet Jr., M., and E.C. Tsivoglou. 1950. Simplified dissolved oxygen computations. *Sewage and Industrial Wastes*, **22**(8): 1054.

Langevin, Paul. 1908. Sur la theorie du mouvement Brownien. *Comptes Rendus de l'Academie des Sciences* (Paris), **146**: 530-533.

Loucks, D. P. 1965. *A Probabilistic Analysis of Wastewater Treatment Systems*. Cornell University Water Resources Center, Ithica, New York.

Loucks, D.P. and W.R. Lynn. 1966. Probabilistic models for predicting stream quality. *Water Resources Research*, **2**(3): 593-605.

Malone, R. et al. 1983. Comparison of techniques for assessing effects of loading uncertainty upon a long term phosphorous model. *Applied Mathematical Modelling*, **7**(1): 11-18.

McLean, Donald M. 1965. Transmission of viral infections by recreational water. In *Transmission of Viruses by the Water Route*. Gerald Berg, ed., 25-36. Interscience Publishers Inc., New York, New York.

Metropolis, N. and S. Ulam. 1949. The Monte Carlo method. *Journal of the American Statistics Association*, **44**(247): 335-341.

Meyer, H. A., ed. 1956. *Symposium on Monte Carlo Methods*. Proceedings of a symposium held at the University of Florida, 16-17 March, 1954, John Wiley & Sons, New York, New York.

Mohsen, M. F. N., and M. H. Baluch. 1983. An analytical solution to the diffusion-convection equation over a finite domain. *Applied Mathematical Modelling*, **7**(9): 285-287.

Moore, S. F., and J. W. Brewer. 1972. Environmental Control Systems: Treatment of Uncertainty in Models and Data. In *Proceedings of the International Symposium on Uncertainties in Hydrologic and Water Resource Systems* (Vol. 3), Tucson, Arizona, 11-14 December, 1972, C. C. Kisel and L. Duckstein, ed., University of Arizona, Tucson.

Mosley, J. W. 1965. Transmission of viral diseases by drinking water. In *Transmission of Viruses by the Water Route*. Gerald Berg, ed., 5-25. Interscience Publishers Inc., New York, New York.

O'Neill, R.V., R.H. Gardner and J.B. Mankin. 1980. Analysis of parameter error in a nonlinear model. *Ecological Modelling*, **8**: 297-331.

Orlob, G.T. and M.A. Tumeo. 1986. Monte Carlo simulation of coastline contamination by an ocean discharge of wastewater. In *Proceedings of the International Symposium on Buoyant Flow*. (Athens, Greece) G. Noutsopoulos, ed., 469-480. Frame Publishing Co., Athens, Greece.

Orlob, G.T., M.A. Tumeo and N. Kolivodiakos. 1987. Monte Carlo simulation applied to ocean outfall analysis. Paper presented at the California Water Pollution Control Association Conference, April 22-24, San Diego, California.

Padgett, W.J. 1975. A stochastic model for stream pollution. *Mathematical Bioscience*, **25**: 309-317.

Padgett, W.J., and S.D. Durham. 1976. A random differential equation arising in stream pollution. *Journal Dynamic Systems Measurement Control*, (Transactions ASME, Series G), **98**: 32-36.

Padgett, W.J., G. Schultz, and C.P. Tsokos. 1977a. A random differential equation approach to the probability distribution of BOD and DO in streams. *SIAM J. Applied Mathematics*, **32**(2): 467-483.

Padgett, W.J., G. Schultz, and C.P. Tsokos. 1977b. A stochastic model for BOD and DO in streams when pollutants are discharges over a continuous stretch. *International Journal of Environmental Studies*, **11**: 45-55.

Padgett, W.J. and A.S. Papadopoulos. 1979. Stochastic models for prediction of BOD and DO in streams. *Ecological Modelling*, **6**: 289-303.

Planck, M. 1917. Uber einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie. *Sitz. Preuss Akad. Wiss. Phys. Math. Kl.*, **325**: 324-341.

Plotkin, S. A. and M. Katz. 1965. Minimal infective doses of viruses for man by the oral route. In *Transmission of Viruses by the Water Route*. Gerald Berg, ed., 151-166. Interscience Publishers Inc., New York, New York.

Risken, Hannes. 1984. *The Fokker-Planck Equation: Methods of Solution and Application*, Volume 18 in Springer Series in Synergetics, Springer-Verlag, New York, New York.

Scavia, D. et al. 1981. Variance estimates for a dynamic eutrophication model of Saginaw Bay, Lake Huron. *Water Resources Research*, **17**(4): 1115-1124.

Smith, L., and F.W. Schwartz. 1981a. Mass transport, part 2: Analysis of uncertainty in prediction. *Water Resources Research*, **17**(2): 351-369.

Smith, L., and F.W. Schwartz. 1981b. Mass transport, part 3: Role of hydraulic conductivity data in prediction. *Water Resources Research*, **17**(5): 1463-1479.

Streeter, H.W., and E.B. Phelps,. 1925. A study of the pollution and natural purification of the Ohio River. Public Health Bulletin No. 146, U.S. Public Health Service.

Thayer, R. P. and R. G. Krutchkoff. 1967. Model for BOD and DO in streams. *Journal of Sanitary Engineering*, **93**(SA3): 59-71.

Thorburn, D. 1986. A Bayesian approach to density estimation. *Biometrika*, **73**(1): 65-75.

Tumeo, M. A. and G. T. Orlob. 1989. An analytic technique for stochastic analysis in environmental models. *Water Resources Research*, **25**(12): 2417-2422.

Tumeo, M. A., and G. T. Orlob,. 1987. Identification of components of uncertainty in environmental models. In *Advances in Environmental Modelling* A. Marani, ed., 599-608. Elsevier publishers, Amsterdam, The Netherlands.

Tumeo, M. A., and G. T. Orlob. 1986. The use of Monte Carlo methods in natural resource management models. In *Proceedings of the 19th Annual Simulation Symposium*,. March 12-14, 1986 (Tampa, Florida), D.L. Kimbler, ed., 73-84. IEEE Computer Press Society, Washington D.C.

Tung, Yeou-Koung. 1987. First-order reliability analysis. Report of the ASCE Technical Committee on Probabilistic Approaches to Hydraulics at the Annual Conference of the ASCE Hydraulic Division, August 3-7, Williamsburg, Virginia.

Tiwari, J. L. and J. E. Hobbie. 1976. Random differential equations as models of ecosystems: Monte Carlo simulation approach. *Mathematical Bioscience*, **28**: 25-44.

Tiwari, J. L. et al. 1976. Some stochastic differential equation models of aquatic ecosystems. *Ecological Modelling*, **4**(1): 3-27.

Ulam, S.M., and John von Neumann. 1945. Random ergodic theorems. *Bulletin American Mathematical Society* (Abstract 51-9-165), **51**(9): 660.

van de Kramer, S. 1983. Monte Carlo simulation and first-order error analysis: Two possible methods to cope with uncertainties in water quality modelling. Working Paper No. WP-83-9, International Institute for Applied Systems Analysis, Vienna, Austria.

van Genuchten, M. T. and W.J. Alves. 1982. Analytical solutions of the one-dimensional convective-dispersive solute transport equation. USDA Technical Bulletin No. 1661, Agricultural Research Service, USDA, Washington, D.C.

von Smoluchowski, M. 1906. Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Annalen der Physik* (Series 4), **21**(14): 756-780.

Widder, D.V. 1975. *The Heat Equation*, Academic Press, Inc., New York, New York.

Yevjevich, V. 1987. Stochastic models in hydrology. *Stochastic Hydrology and Hydraulics*, **1**(1): 17-36.

Zielinski, P.A. 1988. Stochastic dissolved oxygen model. *Journal of Environmental Engineering*, **114**(1): 74-90.

APPENDIX A

DERIVATION OF THE SECOND MOMENT OF THE PROBABILITY DENSITY FUNCTION FOR DOSE FOR THE EXAMPLE CASE IN SECTION 4

Second Moment (Variance of the Dose) = σ_d^2

$$\sigma_d^2 = E_T^2\, C^2\, E[V'^2] + V^2\, C^2\, E[E_T'^2] + E_T^2\, V^2\, E[C'^2] + E_T^2\, E[V'^2]\, E[C'^2]$$
$$+ 2\, E_T^2\, C\, V\, E[V']\, E[C'] + 2\, E_T^2\, C\, E[V'^2]\, E[C'] + 2\, E_T^2\, V\, E[C'^2]\, E[V']$$
$$+ C^2\, E[E_T'^2]\, E[V'^2] + 2\, C^2\, E_T\, V\, E[V']\, E[E_T'] + 2\, E_T\, C^2\, E[V'^2]\, E[E_T']$$
$$+ 2C^2\, V\, E[E_T'^2]\, E[V'] + V^2\, E[E_T'^2]\, E[C'^2] + 2V^2\, CE_T\, E[C']\, E[E_T'] + 2V^2\, C\, E[E_T'^2]\, E[C']$$
$$+ 2\, E_T\, V^2\, E[C'^2]\, E[V'] + E[E_T'^2]\, E[C'\, E[V'^2] + 4\, V\, C\, E[E_T'^2]\, E[C']\, E[V']$$
$$+ 4\, E_T\, V\, E[C'^2]\, E[E_T']\, E[V'] + 2\, C\, E[C']\, E[E_T'^2]\, E[V'^2] + 4\, E_T\, C\, E[C']\, E[E_T']\, E[V'^2]$$
$$+ 2\, V\, E[V']\, E[E_T'^2]\, E[C'^2] + 2\, E_T\, E[E_T']\, E[C'^2]\, E[V'^2] + 6\, V\, C\, E_T\, E[E_T']\, E[C']\, E[V']$$
$$- E_T^2\, C^2\, E[V']^2 - V^2\, C^2\, E[E_T']^2 - E_T^2\, V^2\, E[C']^2 - E_T^2\, E[V']^2\, E[C']^2 - 2E_T^2\, CVE[V']\, E[C']$$
$$- 2\, E_T^2\, C\, E[V']^2\, E[C'] - 2\, E_T^2\, V\, E[C']^2\, E[V'] - C^2\, E[E']^2\, E[V']^2 - 2\, C^2\, E_T\, V\, E[V']\, E[E_T']$$
$$- 2\, E_T\, C^2\, E[V']^2\, E[E_T'] - 2\, C^2\, V\, E[E_T']^2\, E[V'] - V^2\, E[E_T']^2\, E[C']^2 - 2\, V^2\, C\, E_T\, E[C']\, E[E_T']$$
$$- 2V^2 CE[E_T']^2 E[C'] - 2E_T V^2\, E[C']^2\, E[V'] - E[E_T']^2\, E[C']^2\, E[V']^2 - 4VCE[E_T']^2 E[C'] E[V']$$
$$- 4\, E_T\, V\, E[C']^2\, E[E_T']\, E[V'] - 2\, C\, E[C']\, E[E_T']^2\, E[V']^2 - 4\, E_T\, C\, E[V']^2\, E[E_T']\, E[C']$$
$$- 2\, V\, E[V']\, E[E_T']^2\, E[C']^2 - 2\, E_T\, E[E_T']\, E[C']^2\, E[V']^2 - 6\, V\, C\, E_T\, E[E_T']\, E[C']\, E[V']$$

where: $E[V']$ = expectation of V' $E[V'^2]$ = expectation of V'^2
$E[E_T']$ = expectation of E' $E[E_T'^2]$ = expectation of $E_T'^2$
$E[C']$ = expectation of C' $E[C'^2]$ = expectation of C'^2

The expectations of the terms shown in the above equation are dependent on the type of distribution chosen. There are three possible options for input distributions: normal, lognormal, or uniform. The type of distribution selected will determine the final form of the equation. Given the distribution selected, the expectation of the variable and the expectation of the variable squared are defined as follows (from Johnson and Kotz, 1970):

1) Normal: (assuming normal around 0 with a standard deviation of σ)

$$E[x] = 0 \qquad \text{(A-1a)}$$
$$E[x^2] = \sigma^2 \qquad \text{(A-1b)}$$

2) Lognormal: (assuming a geometric mean of 0 and geometric standard deviation of σ)

$$E[x] = \exp(0.5\sigma^2) \qquad \text{(A-2a)}$$
$$E[x^2] = \exp(2\sigma^2) \qquad \text{(A-2b)}$$

3) Uniform: (assuming mean of 0 and a range of plus or minus h)

$$E[x] = 0 \qquad \text{(A-3a)}$$
$$E[x^2] = \frac{h^2}{3} \qquad \text{(A-3b)}$$