

Parameter uncertainty in nonpoint source pollution modeling

by

Dipmani Kumar

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

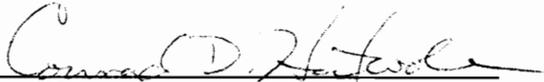
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Biological Systems Engineering

APPROVED:



Conrad D. Heatwole, Chairman



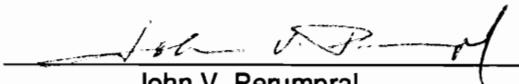
Frank E. Woeste



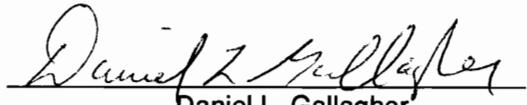
Theo A. Dillaha III



Saied Mostaghimi



John V. Perumpral



Daniel L. Gallagher

February 10, 1995
Blacksburg, Virginia

Parameter uncertainty in nonpoint source pollution modeling

by

Dipmani Kumar

C. D. Heatwole, Chairman

Biological Systems Engineering

(ABSTRACT)

The inherently non-deterministic nature of inputs in nonpoint source (NPS) pollution modeling and the lack of information in many cases results in considerable uncertainty in model input parameters, and therefore, model outputs. Uncertainty in input parameters has important implications for model use in simulation studies and for procedures used to validate and evaluate models. Several management-oriented NPS pollution models have been developed with low input requirements in mind and are generally promoted as tools for comparative analysis of alternative land-use scenarios, with an implied blocking of the effects of parameter uncertainty. However, the impact of input parameter uncertainty on comparative analysis has not been rigorously investigated.

A dynamic Monte Carlo simulation based procedure is presented for quantifying the impact of parameter uncertainty on both absolute and comparative predictions from NPS pollution models, with possible correlations among inputs considered. An application of the procedure in a simulation study employing a management-oriented continuous NPS pollution model, GLEAMS, indicated parameter uncertainty had important consequences for both absolute predictions and comparative analysis between scenarios.

Information on correlations among input variables in Monte Carlo studies is frequently not available, and is commonly determined subjectively. A novel interactive approach to generating

a correlation matrix from subjective information was developed. The procedure retains correlations between input variables with which a high degree of confidence is associated and allows use of these correlations to examine possible correlations between other variables.

Procedures were developed for defining probability distributions of important hydrologic inputs in management oriented NPS pollution models. A conceptual approach was used to derive runoff curve number distributions suitable for use in continuous models which adjust the input curve number based on continuous moisture accounting. New regression equations were developed for predicting saturated conductivity, field capacity, and wilting point using sand and clay content, porosity, and organic matter as regressors. Evaluation of the regression equations indicated acceptable predictions of field-scale variability in field capacity and wilting point but not of saturated conductivity. A comparison of point predictions from the new equations with previously developed equations reported in the literature showed improved prediction for all response variables.

Theoretically sound procedures for probabilistic evaluation of NPS pollution models were developed by analogy with the goodness-of-fit testing procedure. The procedures do not require the assumption of normality of prediction errors for application, and may be used to assess model performance for both absolute and comparative predictions. A probabilistic index of model performance that is simple to calculate and interpret is proposed. The procedures were illustrated by application to the GLEAMS model using data from a short-term field study that examined pesticide fate and transport in a Coastal Plains soil under alternative tillage treatments.

Acknowledgments

I would like to acknowledge the advice and support Dr. C. D. Heatwole provided during the course of this work. I would also like to express my appreciation to Dr. Heatwole for his patience and understanding in dealing with my idiosyncrasies. The input provided by the members of my examining committee is gratefully acknowledged. Dr. M. L. Wolfe kindly agreed to represent Dr. T. A. Dillaha while he was on sabbatical leave.

I will always remember the help, companionship, and empathy of all the graduate students in the department. In particular, Jeaan's help and encouragement tided me over a very difficult period when I reached the nadir of my tenure as a graduate student. Richard was always ready to listen to my ideas, and graciously explained what was wrong with most of them. Sebastian's willingness to help at all times is gratefully acknowledged. Faycal, Aida, and Sanjay very kindly indulged me when I wanted to complain about life and told me to get on with it anyway.

This work was supported in part by funds from the USEPA Office of Exploratory Research (Project R81-8511-010), and the Virginia Agricultural Experiment Station (Project 132524 and Southern Regional Project S-249).

Table of Contents

Chapter I. Introduction	1
Background and Justification.....	1
Objectives.....	4
Chapter II. Literature Review	6
Nonpoint source pollution simulation models	6
Uncertainty analysis	8
Characterizing input parameter uncertainty in NPS pollution simulation models	11
Interpreting model output.....	13
Characterizing input probability distributions.....	15
Parameter uncertainty in model evaluation.....	16
Chapter III. Quantifying the impact of parameter uncertainty on predictions from continuous models	18
Introduction.....	18
Implementing direct Monte Carlo simulation in batch mode	19
General purpose pre-processor module for Monte Carlo simulation	20
Direct Monte Carlo simulation in batch mode.....	21
Interactive assessment of correlation structures with subjective information.....	24
Interactive generation of a correlation matrix.....	25
Application to a nonpositive definite matrix	31
Simulation study	34
Procedures for model output analysis.....	35
Sensitivity analysis.....	39
Monte Carlo simulations	39
Results and discussion	43
Monte Carlo simulation with independent inputs.....	46
Monte Carlo simulation with correlated inputs	51
Sensitivity to input distribution shape	53
Summary and Conclusions.....	57
Chapter IV. Probability distributions for hydrologic inputs	59
Introduction.....	59
Runoff curve number distributions.....	61
Potential retention and runoff curve number distributions	62
Distributions resulting from factors other than antecedent moisture variation.	64
Application to conventional and conservation tillage systems.	68
Results and Discussion	70
Distributions of soil hydraulic properties.	75
Regression Analysis	76
Evaluation	92

Chapter V. Development of probabilistic model evaluation procedures	100
Introduction	100
A probabilistic approach to assessing model performance	101
Application	107
Field Data.....	107
Characterization of model inputs.	108
Results and Discussion.....	111
Chapter VI. Summary, Conclusions, and Recommendations	119
Summary and Conclusions	119
Recommendations for future work.....	124
Multivariate approaches to assessing model output	124
Determining the global sensitivity of input parameters	124
Incorporation of spatial structure in field scale models	125
Assessment of curve number methodology in continuous models.....	126
Improvement of regression equations.....	127
References	128
Appendix A General purpose Monte-Carlo pre-processor module	136
Listing	137
Example input file	145
Example output summary file.....	146
Appendix B Listing of routine for interactive assessment of correlation structures	147
Appendix C Listing of routines for approximating probability distributions of soil hydraulic properties	152
Runoff curve number distributions.....	153
Distributions for saturated conductivity, and moisture contents at 0.1 bar, 0.33 bar and 15 bar tensions.....	154
Vita	163

List of Figures

Figure 3.1	Flowchart of general procedure for implementing direct Monte Carlo simulation in batch mode.....	23
Figure 3.2	Flowchart of routine for interactive generation of a positive definite correlation matrix.....	29
Figure 3.3	Convergence of pesticide leaching over simulation period with increasing number of trials: (a) 50% exceedence probability levels (b) 10% exceedence probability levels.....	46
Figure 3.4	Variation in coefficient of skew over simulation period for various outputs with uniform, independent inputs.	52
Figure 3.5	Impact of input distribution shape on output inter-quartile ranges over simulation period: (a) runoff (b) percolation (c) sediment loss (d) surface pesticide loss (e) pesticide percolation.	56
Figure 4.1	Conceptual procedure for obtaining the S* distribution from the complete S distribution.....	67
Figure 4.2	Empirical distribution functions for potential retention conditioned on runoff.....	69
Figure 4.3	Histogram of generated values for complete CN distribution (conventional tillage).	72
Figure 4.4	Histogram of generated values for CN* distribution (conventional tillage).	72
Figure 4.5	Histogram of generated values for complete CN distribution (conservation tillage)..	73
Figure 4.6	Histogram of generated values for CN* distribution (conservation tillage).	73
Figure 4.7	Plots of ranked values for complete CN.....	74
Figure 4.8	Plots of ranked values for CN*.....	74
Figure 4.9	Diagnostic plots for $\ln K_s$ regression: (a) residual plot (b) rankit plot.	82
Figure 4.10	Diagnostic plots for $\theta_{0.1}$ regression: (a) residual plot (b) rankit plot.	83
Figure 4.11	Diagnostic plots for $\theta_{0.33}$ regression: (a) residual plot (b) rankit plot.	84
Figure 4.12	Diagnostic plots for θ_{15} regression: (a) residual plot (b) rankit plot.	85
Figure 4.13	Diagnostic plots for regression: (a) residual plot (b) rankit plot	88
Figure 4.14	Diagnostic plots for regression: (a) residual plot (b) rankit plot.	89
Figure 4.15	Diagnostic plots for regression: (a) residual plot (b) rankit plot.	90

Figure 4.16	Observed and predicted cumulative frequency plots for K_s : (a) trace-driven (b) simulation using input ranges.....	94
Figure 4.17	Observed and predicted cumulative frequency plots for $\theta_{0.1}$: (a) trace-driven (b) simulation using input ranges.....	95
Figure 4.18	Observed and predicted cumulative frequency plots for $\theta_{0.33}$: (a) trace- driven (b) simulation using input ranges	96
Figure 4.19	Observed and predicted cumulative frequency plots for θ_{15} : (a) trace-driven (b) simulation using input ranges.....	98
Figure 5.1	Goodness-of-fit approach to model evaluation.....	103
Figure 5.2	Obtaining the value of p_g from the model output distribution or output sampling distribution.	106
Figure 5.3	Predicted distribution and observed empirical distribution function (EDF) of Metolachlor mass in the soil profile on second sampling date.....	116
Figure 5.4	Model output sampling distribution (n=20) and observed value of median difference in Metolachlor mass between conservation tillage (CsT) and conventional tillage (CvT) on third sampling date.....	116
Figure 5.5	(a) Observed surface losses of Atrazine from conventionally tilled plot, and (b) scatter plot of indices of model performance over monitored period.	119

List of Tables

Table 3.1	DO correlation matrix (Song and Brown, 1990)	30
Table 3.2	Bounds for DO matrix.	30
Table 3.3	Capital investment correlation matrix (Ince and Buongiorno, 1991).....	32
Table 3.4	Results of initial assessment of capital investment matrix.	32
Table 3.5	Results of analysis for independence of annual output from GLEAMS,	38
Table 3.6	Results of association and sign tests on deterministic output for scenarios 1 and 2	38
Table 3.7	Uncertain parameters in GLEAMS and abbreviations.....	40
Table 3.8	Median relative sensitivity of uncertain parameters.	41
Table 3.9	Range for uncertain parameters assumed to be the same for scenarios 1 and 2.	42
Table 3.10	Ranges for parameters used to simulate differences between scenarios 1 and 2.....	42
Table 3.11	Correlations between selected input parameters.	44
Table 3.12	Results from deterministic simulation with statistics based on annual values over a 44 year simulation (Scenario 1).	47
Table 3.13	Differences between scenarios 1 and 2, deterministic simulations.....	47
Table 3.14	Results of Monte Carlo simulation with independent inputs (scenario 1).....	48
Table 3.15	Differences between scenarios 1 and 2, Monte Carlo simulations with independent inputs	48
Table 3.16	Results of Monte Carlo simulation with correlated inputs (scenario 1).	52
Table 3.17	Results of Monte Carlo simulations for differences between scenarios 2 and 1 with correlated inputs.	52
Table 4.1	Parameter estimates for rainfall and potential retention (S) distributions.	71
Table 4.2	Median of complete S distributions (S_{II}), corresponding S_{III} values with exceedence probabilities (p_e), and fraction of S variation explained by antecedent moisture (m).	71
Table 4.3	Summary of central tendency and dispersion measures for complete CN and CN* distributions.	71
Table 4.4	Number of observations and adjusted R^2 obtained with full models.....	81

Table 4.5	Number of predictors and adjusted R ² with selected model sub-sets.....	81
Table 4.6	Adjusted R ² and Durbin-Watson (D-W) statistic for final models.	87
Table 4.7	Input correlations estimated from the soil properties database (Rawls, 1980).....	87
Table 4.8	Comparison of point predictions from Rawls and Brakensiek (1985) regression equations and new regression equations.....	99
Table 5.1	Inputs considered random variables and abbreviations	109
Table 5.2	Distributions used for inputs in model evaluation.....	109
Table 5.3	Distributions and parameters common to conservation and conventional tillage simulations	112
Table 5.4	Distributions and parameters used to simulate differences between conservation and conventional tillage systems.....	113
Table 5.5	Input correlations used in simulations.....	114

Chapter I

Introduction

Remember when discoursing about water to adduce first experience and then reason.

Leonardo da Vinci

BACKGROUND AND JUSTIFICATION

Nonpoint source (NPS) or diffuse pollutant loadings originate from activities related to the use of land by man and from natural watershed processes (Novotny and Chesters, 1981). Essentially, water quality problems associated with NPS pollution result from the superposition of human activity on the hydrologic components of nature. Spooner et al. (1992) provide a comprehensive review of the fate and effects of NPS pollution from agricultural, forest, and urban sources. The recognition of NPS pollutants as major contributors to surface and groundwater quality degradation has led to an increased emphasis on their reduction and control. Best management practices (BMPs) are widely recognized as the most effective strategy to reduce the amount of NPS pollutants entering streams and groundwater (Duttweiler and Nicolson, 1983; Heatwole et al., 1991). In general, BMP effectiveness is site-specific, and depends on the site's soil, topographic, cropping, and climatological characteristics.

The most direct way of measuring NPS pollutant loadings is through watershed and field monitoring studies. However, monitoring is extremely time consuming and, because of the need for site-specific evaluations, is prohibitively expensive. D'Elia et al. (1989) noted that in some cases, data obtained in monitoring studies may become so compromised in one or more respects that the data are of little value. D'Elia et al. (1989) also point out that perhaps the greatest problem with monitoring is that it is so demanding that it often becomes a goal in itself

and the original purpose of monitoring is forgotten. A sound experimental design and rigorous quality assurance and quality control measures are necessary in any monitoring study to allow effective evaluation of the effectiveness of NPS pollution control measures (Spooner et al., 1985, 1988). The limitations of monitoring notwithstanding, good quality monitoring data are essential not only for providing information on the potential impact of alternative land management strategies, but also for the development and testing of NPS pollution simulation models.

Simulation models are important, cost-effective tools in the effort to reduce NPS pollutants because they can provide *a priori* assessment of alternative practices and can be used in many cases where monitoring is not feasible. However, modeling natural watershed processes is an extremely difficult task, and many questions and concerns about model application remain. Anderson and Woessner (1992) have noted that there are two prevalent opinions about simulation models:

1. Models are essential in performing complex analyses and in making informed predictions.
2. Models are worthless because they require too much data and therefore are too expensive to assemble and run. Furthermore, they can never be proved to be correct and suffer from a lack of scientific certainty.

The research in this dissertation is predicated on the first viewpoint, and will provide a framework for addressing concerns expressed in the second viewpoint.

The large degree of spatial variability in natural soils and the complex, dynamic nature of watershed processes make it extremely difficult to accurately model the fate and transport of NPS pollutants in a deterministic framework. Current NPS pollution models attempt to simulate natural watershed processes in a deterministic manner, employing effective input parameters

that are regarded as representative of a homogenous system equivalent to the natural heterogeneous system. Although such an approach is acceptable if data are available for determining effective input parameters, this is rarely the case in NPS pollution modeling.

The inherently non-deterministic nature of many model inputs and the lack of information in many cases results in considerable uncertainty in model input parameters, and therefore, model outputs. Haan (1989) notes however, that there is a reluctance to admit that model outputs are uncertain estimates, and emphasizes the importance of recognizing and incorporating uncertainty into the analysis of hydrologic systems. The increasing use of simulation models to assess NPS pollution impacts (Carsel et al., 1988b; Thomas et al., 1990; Phillips et al., 1993), and more recently their integration with economic models to evaluate regulatory policies for NPS pollution control (Heatwole et al., 1990; Weaver and Harper, 1993; Bouzhar et al., 1993) underscores the importance of directing efforts towards the development of procedures that will allow users to incorporate input parameter uncertainty in any analysis with NPS pollution simulation models.

Uncertainty in input parameters has important implications for both model use in simulation studies and procedures used to validate and evaluate models. The results of simulation studies that do not recognize and account for the impact of input parameter uncertainty cannot be used with a high degree of confidence, and model evaluation procedures which are based on determining effective parameters are not meaningful when the model is expected to be used under conditions of limited data availability.

Several management-oriented models (Carsel et al., 1985; Knisel et al., 1993) have been developed with low input requirements in mind and are generally promoted as tools for comparative analysis of alternative land-use scenarios, with an implied blocking of the effects of parameter uncertainty. However, the impact of input parameter uncertainty on comparative

analysis has not been rigorously investigated, and there are currently no model evaluation procedures that allow an assessment of model performance for comparative analysis. Thus, there is a need for a framework which allows model users to consider input parameter uncertainty in simulation studies as well as model evaluation procedures in an explicit manner for both absolute predictions and comparative analysis. Important considerations in developing this framework would be the use of rigorous, defensible procedures for model output analysis and an evaluation of approaches to approximating field-scale variability of critical hydrologic inputs in NPS pollution models from readily available information.

OBJECTIVES

The overall objective of this research is to develop probabilistic procedures for using and evaluating NPS pollution models, with the basic goal of improving the degree of confidence associated with the use of models in simulation studies. The methodology will be demonstrated using the pesticide component of a widely used management-oriented model, GLEAMS (Knisel et al., 1993). However, it is expected that the concepts and procedures developed through this research will have general applicability to other NPS pollution models. The specific objectives of the research are:

1. Develop procedures for quantifying the impact of parameter uncertainty on absolute and comparative predictions from continuous NPS pollution models, with possible correlations among inputs considered.
2. Develop and evaluate procedures for defining probability distributions of critical hydrologic inputs in NPS pollution models.

3. Develop probabilistic procedures for model evaluation which explicitly consider input data uncertainty and may be used to evaluate both absolute and comparative model predictions, and demonstrate the procedure using the GLEAMS model and data from a field study.

Chapter II

Literature Review

I can foretell the way of celestial bodies, but can say nothing of the movement of a small drop of water.

Galileo Galilei

NONPOINT SOURCE POLLUTION SIMULATION MODELS

The hydrologic, chemical, and biological processes occurring on a watershed are complex and dynamic in nature, and many soil, chemical, and crop properties exhibit tremendous spatial and temporal variability. Various approaches to modeling watershed processes have been employed and several models providing simulation at the field and watershed scale developed. Adiscott and Wagenet (1985) have distinguished between deterministic approaches, which presume that a system or process operates such that the occurrence of a given set of events leads to a uniquely definable outcome, and stochastic approaches, which presume the outcome to be uncertain and are structured to account for this uncertainty. The distinction is somewhat arbitrary, as deterministic models may in fact be used in a stochastic framework (Carsel et al., 1988a, 1988b). Wagenet and Rao (1990) remarked that although stochastic approaches are conceptually a step in the right direction given the uncertainty in model parameter values arising from the spatial heterogeneity of field soils, stochastic approaches have not yet been condensed into a form that allows practical use. Although stochastic approaches, both mechanistic (Destouni, 1993) and non-mechanistic (Jury et al., 1990), are being actively researched, all NPS pollution simulation models in general use at the current time are deterministic.

Deterministic NPS pollution simulation models can generally be classified as either screening or assessment models. Screening models are usually relatively simple and are intended to identify

problem areas within large basins (Khan and Liang, 1989). Assessment models are more complex and may be used to evaluate the effects of alternative land-use scenarios. Assessment models may be field-scale (Smith, 1992; Knisel et al., 1993) or watershed-scale (Johanson et al., 1981; Young et al., 1989). Field-scale models describe hydrologic processes within a single field or land resource unit, and are generally continuous simulation models, while watershed-scale models may be event-oriented (single storm predictions) or continuous. A distinction is also made between lumped parameter and distributed parameter models (Novotny and Chesters, 1981). Lumped parameter models represent the unit being modeled as an uniform system employing effective parameters which may or may not have physical significance. Distributed parameter models divide the modeled system into small units which are regarded as homogenous, with input parameters for the homogenous units generally having a physical basis. Field-scale models usually use lumped parameters, while watershed-scale models may use distributed or lumped parameters.

In terms of conceptualization and intended use, NPS assessment models can also be further classified as research or management models. Research models (Smith, 1992) attempt to represent basic hydrologic processes and pollutant dynamics in fundamental and mechanistic terms and have the capability to provide simulations of pollutant behavior under a wide range of conditions. However, research models can be quite data demanding, and frequently require calibration of input parameters. Management models (Knisel and Davis, 1993), generally developed for uncalibrated use, represent fundamental processes in a simplified manner, so that they are computationally efficient and can be used with readily available data for defining model inputs.

Research and management models correspond approximately to rate and capacity models referred to by Addiscott and Wagenet (1985). A rate model employs parameters that exhibit a large degree of spatial variability and is theoretically capable of simulating transient system

response, as the model is generally formulated using the governing differential equations of the system. Capacity models define amounts of change rather than rates of change, employing capacity parameters which are much less variable, with the model driven by the amount of rainfall and evapotranspiration. A possible advantage in using capacity models is that the impact of parameter uncertainty on model predictions may not be significant, particularly when compared to rate models. This is used to justify their use in situations where field data for site characterization and model calibration are not available. However, a recent study by Zhang et al. (1993) suggests that even with a capacity model, uncertainty in input parameters may have a significant impact on model response.

Management models are generally proposed as tools for comparative analysis rather than tools for making absolute predictions. Comparative analysis is the comparison of simulated responses to changes or differences in management systems, soils, climate, chemical properties, and other factors (Leonard and Knisel, 1990). In comparative analysis, the absolute value of model predictions is not of primary importance, but relative differences in predictions are used to guide management decisions. However, Leonard and Knisel (1990) have noted that even for relative comparisons, better methods are needed to translate uncertainties in input parameters to uncertainties in output.

UNCERTAINTY ANALYSIS

Uncertainty in input parameters may be attributed to several factors. There may be significant spatial and temporal variation (parameters may be inherently non-deterministic) and parameters may not be known with a high degree of accuracy e.g. literature based best estimates (information uncertainty). Efforts to account for this uncertainty in the past have commonly used a combination of sensitivity analysis and calibration procedures. Model sensitivity to parameters,

expressed as the relative rate of change in outputs of interest caused by a unit change in the input, provides a measure of the degree of variation that might be expected in model output due to uncertainty in an input parameter. Information on model sensitivity may then be used to establish required accuracy and precision in the model application as a function of data quantity and quality (Hern et al., 1985). The typical approach in modeling studies is to employ calibration of the most sensitive parameters. Formally, calibration, or "history matching" as it is referred to in the petroleum industry, may be defined as the process whereby model input parameters are modified until a reasonable comparison is made with observed data (Konikaw and Bredehoeft, 1992). Although the application of optimization techniques to calibration have been studied fairly extensively, finding an optimal set in high dimensional parameter space is difficult (Beven and Binley, 1992). The most frequently used calibration procedure, is in fact, the manual trial and error approach, involving selective improvement of initial parameter estimates by comparing observed and simulated data (Abbot et al., 1985).

Calibration is undesirable for NPS pollution modeling. The biggest reason, of course, is that data for calibration are very rarely available for agricultural watersheds. Calibration requires solving an inverse problem which is poorly defined and may be frequently prone to error. Beven and Binley (1992) have noted that a significant problem with traditional calibration procedures is the difficulty of distinguishing between multiple optima in the response surface. The existence of multiple optima may cause the exclusion of parameter sets distant from the current optimum which give more accurate results for a period of observations different from the one on which calibration was based.

Techniques for uncertainty analysis which allow consideration of parameter uncertainty in a more explicit manner include first order uncertainty analysis and Monte Carlo simulation. In first-order uncertainty analysis (FOUA) it is assumed that the mean and variance of a random variable are sufficient to characterize the variable. Uncertainties in model input parameters are propagated

through the model using the first order terms in a Taylor series expansion about the mean value of each non-deterministic input variable (Dettinger and Wilson, 1981). A significant advantage of FOUA is that if the assumptions it is based on are valid, relative contributions to output variance from each non-deterministic input can be easily quantified. However, the FOUA method is not applicable to non-linear models or when the coefficient of variation of inputs is large (Benjamin and Cornell, 1970).

Monte Carlo simulation (MCS) is a generally applicable procedure that is widely used in uncertainty analysis. With this method, a probability distribution is identified or assumed for each non-deterministic input parameter. Model simulations are then performed a predetermined number of times, with each simulation using a value for each uncertain input randomly selected from the assumed probability distribution. The result is a distribution of model output values which can be statistically analyzed to assess a prediction error that represents the combined effects of all uncertainties in the model (Song and Brown, 1990). MCS results include the effects of model non-linearity.

In MCS, it is extremely important to take into account correlations in input parameters, since ignoring correlations can lead to serious inaccuracies in the results of a simulation (Law and Kelton, 1991). Effective implementation of Monte Carlo simulation for models where the inputs are thought to be correlated can be difficult, as the introduction of correlation significantly increases the complexity of generating random variates (Ince and Buongiorno, 1991). It is generally not possible to specify the exact, full, joint (multivariate) distribution of the input vector (Law and Kelton, 1991). However, the distributions of individual components (the marginal distributions) of the input vector and the correlations between the individual components may be known from theoretical or experimental studies, or determined through subjective assessment.

A commonly used approach to generating random vectors with correlated components is to assume the marginal distributions of the components are all normal (Gaussian), since available multivariate-normal random number generators (IMSL, 1987), can be used to generate the required random vectors. With non-normal marginal distributions, transformations that result in normally distributed variables have been employed (Carsel and Parrish, 1988). This may result in the original marginal distributions and correlation structure not being accurately preserved. In addition, this approach cannot be used to generate correlated random variates from empirical distribution functions.

A generally applicable, efficient, and easily implemented modified multivariate normal procedure for generating correlated random variates was presented by Taylor and Bender (1989). The method preserves the exact marginal distribution of each variable and closely approximates the correlations between the variables. Han et al. (1991) empirically evaluated the modified multivariate normal procedure of Taylor and Bender (1989) to determine how well the approach worked for highly skewed distributions. They found that the method preserved the correlation between variables with excellent accuracy except for cases involving mixtures of positively and negatively skewed distributions in conjunction with high correlation. However, even with mixtures of positively and negatively skewed distributions in conjunction with high correlation, the difference between simulated and original correlations was less than 10%, which may not be of practical concern in many applications.

CHARACTERIZING INPUT PARAMETER UNCERTAINTY IN NPS POLLUTION SIMULATION MODELS

Loague et al. (1989, 1990) used FOUA to characterize the impact of data uncertainty on estimates of pesticide mobility made with simple screening indices. The simple pesticide mobility

indices were coupled with a digital soil database geographic information system to generate pesticide ranking maps. Loague et al. (1989) conclude that deterministically obtained pesticide rating maps with no indication of uncertainty have the potential to be misused by decision makers. They suggest that the effect of parameter uncertainty should be an important consideration for deterministic simulation models that are currently being used or suggested for making predictions of pesticide mobility.

Carsel et al. (1988a) developed a MCS procedure for making regional assessments of pesticide leaching using the PRZM (Carsel et al., 1985) management model. Their procedure used probability density functions for soil organic matter, field capacity, and permanent wilting point developed from approximately 3000 soils. Carsel et al. (1988b) carried out a MCS of PRZM linked to a simple groundwater solute transport model to provide regional assessments of pesticide residue loadings and movement in groundwater underneath and downgradient from treated fields. Carsel et al. (1988b) state that the MCS technique is useful for evaluating the potential for producing pesticide residues in drinking water wells before actual field applications are made and suggest that the technique can also be used to develop appropriate land management options, including restrictions on pesticide application.

Dean et al. (1989) developed the RUSTIC model which links an enhanced version of PRZM to a two-dimensional saturated zone model either directly or through a one-dimensional vadose zone model. RUSTIC includes a Monte Carlo shell and allows randomization of selected inputs to produce distributional outputs. However, the current status of RUSTIC is not clear. Users have reported problems in using the Monte Carlo shell and the saturated zone model component (Zacharias, 1992; Li, 1993).

Beven and Binley (1992) describe a unique MCS based procedure for calibration and uncertainty estimation of physically based distributed parameter model using general likelihood measures.

The procedure works with multiple sets of parameter values, assuming different sets of values may be equally likely, and allows updating of likelihood values, based on continuing input from new observation periods. Beven and Binley (1992) also suggest how the value of different types of data in providing uncertainty estimates may be assessed by examination of the effect of incorporating the data on the likelihood distribution.

Zhang et al. (1993) used MCS of a screening model, CMLS (Nofziger and Hornsby, 1986) to determine the effect of uncertainty in model parameters and rainfall variability on uncertainty of model response. They found that the magnitude of uncertainty in estimated solute travel time to a given depth resulting from parameter uncertainty was comparable to the uncertainty resulting from rainfall variability. Zhang et al. (1993) also apply the results of the uncertainty analysis in conjunction with sampling theory to determine the minimum number of samples required to define each of the input parameters to provide a stated level of uncertainty in model response.

INTERPRETING MODEL OUTPUT

An important consideration in developing a probabilistic framework for NPS pollution modeling is interpretation of model output. Currently, model output analysis tends to be *ad hoc*, ranging from simple graphical and quantitative techniques (Gassman and Anderson, 1989) to statistical analysis of long term simulations (Davis and Heatwole, 1990).

In using statistical procedures to analyze model output, it is extremely important to determine whether observations from a given data are independent. The runs up and down test (Gibbons, 1985) is useful for determining whether an ordered sequence of observations can be considered to be independent. If more than one data set is being analyzed, it is also necessary to establish whether the data sets can be regarded as independent from each other. The Spearman rank

correlation coefficient and the Kendall Tau coefficient may be used to test for independence between two data sets (Gibbons, 1985). When the null hypothesis of no association between data sets cannot be rejected, statistical procedures for independent samples should be employed in the analysis. If the null hypothesis is rejected, statistical procedures for paired or related samples should be used (Hollander and Wolfe, 1973; Gibbons, 1985). The statistical procedures generally involve testing for location (mean or median) and dispersion (variance) differences. If significant differences are indicated, estimators and confidence intervals at specified confidence levels for the location or dispersion difference may also be computed. Although parametric procedures, which assume data is derived from a normal distribution, are generally used in output analysis, the use of nonparametric procedures, which are distribution free, may be desirable in several instances. Parametric procedures are generally more powerful than nonparametric procedures when the underlying distribution can be reasonably regarded as normal. However, if this is not the case, parametric procedures may perform poorly, and nonparametric procedures should be employed. Stephens (1977) has suggested modified versions of commonly used empirical distribution function (EDF) goodness-of-fit tests to determine whether a set of independent, continuous observations can be regarded as being derived from some (unspecified) normal population.

When parameter uncertainty is taken into account, it is extremely important that consistent, defensible procedures are employed for output analysis. Well defined procedures for output analysis when parameter uncertainty is taken into account are essential before simulation models can be judiciously used to assess the effectiveness of proposed NPS pollution control measures. The procedures should answer basic questions such as whether parameter uncertainty affects the use of management models for comparative analysis. Although the effect of parameter uncertainty on absolute model predictions is well documented, there have been no attempts to quantify the impact of parameter uncertainty on relative model predictions.

CHARACTERIZING INPUT PROBABILITY DISTRIBUTIONS

Monte Carlo studies of NPS pollution simulation models require specification of the probability distributions of soil, crop, and chemical properties. Information on the nature of probability distributions of crop and chemical properties at the field-scale is scarce, and inputs for these properties are commonly assumed to be deterministic (Carsel et al., 1988b; Smith and Charbeneau, 1990). However, the variability of soil properties has been studied fairly extensively, and differing approaches have been used to approximate probability distributions of soil properties. Carsel et al. (1988a) fitted members of the Pearson and Johnson transformation systems to data from a large number of soils and obtained a multivariate normal distribution for field capacity, wilting point, and organic matter. Zhang et al. (1993) pooled observed data and values reported in the literature and modeled bulk density, field capacity, wilting point, and organic carbon content with a multivariate lognormal distribution.

The most desirable procedure for characterizing distributions of soil properties at the field-scale is, of course, through direct measurement. However, field and laboratory measurements of soil hydraulic properties are time consuming, tedious, and expensive. Although data to define distributions of soil hydraulic properties may be available in a few cases from detailed field studies, more commonly, this data is not available. Rawls and Brakensiek (1985) have reported ordinary least squares multiple linear regression equations derived from an extensive soils database for predicting soil hydraulic properties from soil texture, porosity, and organic matter content. Carsel and Parrish (1988) used the equations to develop a method for obtaining probability distributions for selected soil hydraulic properties for the 12 Soil Conservation Service (SCS) textural classes. However, the utility of the equations in providing acceptable predictions of field-scale variation in soil hydraulic properties has not been examined.

An important input in many management-oriented NPS pollution models is potential retention, or runoff curve number (CN), a variable that reflects the soil-cover complex. The CN, which is used to partition precipitation into runoff and infiltration, is fairly extensively documented for a large number of soil-cover complexes. Continuous hydrologic models commonly adjust the value of CN used for a particular storm based on the current soil water content (Williams et al., 1984). These models generally require input for the mean or average soil moisture condition CN. Haan et al. (1990) noted that the CN distribution is generally left skewed, and using a representative distribution illustrated that there is a high probability that the actual CN for a given storm may be ± 10 units away from the mean estimated CN.

PARAMETER UNCERTAINTY IN MODEL EVALUATION

The issue of input parameter uncertainty in model evaluation and validation studies has not been critically examined. Model evaluation studies (Pennell et al., 1990; Smith et al., 1991) have commonly focused on using subjective (graphical) criteria and objective error estimates (Loague and Green, 1991) to assess the predictive ability of a model. The error estimates are computed from differences in deterministic model output and the mean of observed values. This ignores both the model output variance resulting from parameter uncertainty and the variance in the observed data. Parrish and Smith (1990) developed the concept of a capability index for determining whether model predictions fall within a specified factor of true values. They related the concept to hypothesis testing using parametric procedures. Zacharias and Heatwole (1993a, 1993b) developed nonparametric analogs of the hypothesis testing procedure. Nonparametric procedures are particularly relevant for testing NPS simulation models because of the small observed sample sizes available from field studies and the fact that many variables of interest are not normally distributed. Although the capability index and hypothesis testing approach considers the variance in observed data, model output variance is still not taken into account.

Luis and McLaughlin (1992) have presented a stochastic approach to model validation. They make use of model output variance with model validation again regarded as a hypothesis testing problem, and they outline different tests for systematic biases, overall fit, and spatial structure. Incorporating the variance associated with model predictions is important in providing realistic assessments of model performance. Recently, Haan et al. (1993) have also suggested a model evaluation procedure which utilizes the model output variance to place confidence intervals on model predictions. They suggest that decisions regarding model acceptability can then be made based on the magnitude of measured data in comparison to the confidence interval and performance criteria placed on the model. However, viewing validation as a series of hypothesis tests or confidence interval determinations can complicate the task of interpreting the results of a model validation study. Measurements are typically available over a large spatio-temporal range and appropriate significance levels have to be selected. The hypothesis testing or confidence interval approach also cannot be directly employed to evaluate model performance for relative rather than absolute predictions. There is a need for a procedure which allows model evaluation for both absolute predictions and comparative analysis taking input parameter uncertainty into account, and provides an overall assessment of model performance that can be easily interpreted.

Chapter III

Quantifying the impact of parameter uncertainty on predictions from continuous models

....when you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of science, whatever the matter may be.

Lord Kelvin

INTRODUCTION

Previous Monte Carlo studies of continuous models have not examined the impact of parameter uncertainty in a dynamic framework. Carsel et al. (1988a, 1988b) employed random sampling of multi-year precipitation records to incorporate variations in the distribution and amount of rainfall, but did not consider the effect of parameter uncertainty and rainfall variation separately. Zhang et al. (1993) considered uncertainties resulting from variability in soil parameters and rainfall in predictions of time taken for the center of mass of a solute to travel 1 m. They used a stochastic generator to obtain different realizations of 40 year precipitation sequences, and quantified the separate and combined impacts of soil and rainfall variability on solute travel time. However, the model employed by Zhang et al. (1993) is generally regarded as a screening model rather than an assessment model (Nofziger and Hornsby, 1986), and only provides predictions of the location of the center of mass and relative amount of chemical remaining in the soil profile. The application of a chemical every year which would occur in practice cannot be represented and upward soil water and chemical movement in response to evaporative demand are not simulated by the model. In their study, Zhang et al. (1993) did not examine the dynamic variation of chemical loadings from a given land-use scenario for a given weather sequence. Another important consideration that has been overlooked in previous studies is the impact of parameter uncertainty on comparative predictions from alternative land-use scenarios. As we have noted before, although management models are promoted as being suitable for obtaining relative

rather than absolute predictions from alternative land-use scenarios, there have been no attempts to quantify the impact of parameter uncertainty on comparative analysis.

In this chapter, a flexible procedure that allows rapid development of Monte Carlo simulation overlays for deterministic computer models is presented. An interactive approach for generating correlation matrices from subjective information for Monte Carlo simulation with correlated inputs is developed. A dynamic Monte Carlo simulation study is carried out to quantify the impact of parameter uncertainty on absolute and relative pesticide losses under alternative land management systems employing the GLEAMS (Knisel, 1993) model. In the simulation study, rainfall inputs are fixed, and the impact of parameter uncertainty on model predictions determined for a given rainfall sequence. With most simulation studies in practice, good quality historical or synthetic precipitation data are generally available, while considerable uncertainty may exist in estimating model parameters. Rainfall variability is taken into account by employing long-term (44 year) simulation. This approach allows the explicit separation of the impact of parameter uncertainty and rainfall variability. A dynamic Monte Carlo simulation carried out in this manner results in different sequences of output for each realization of input parameters, all driven by the same rainfall input. Appropriate procedures for analyzing model output from a dynamic Monte Carlo simulation are presented.

IMPLEMENTING DIRECT MONTE CARLO SIMULATION IN BATCH MODE

Although some deterministic computer models, such as QUAL2E-UNCAS (EPA, 1987) and RUSTIC (EPA, 1989), provide modules which allow direct Monte Carlo simulation of the model, more frequently the Monte Carlo procedure must be implemented externally in some manner (Smith and Charbeneau, 1990; Zhang et al., 1993). The exact method of implementing Monte Carlo simulation varies, ranging from general-purpose model interfaces to *ad hoc* routines.

In this study, direct Monte Carlo simulation of GLEAMS was implemented in batch mode in two stages. The required random variates were first generated using a general-purpose pre-processor module. Repeated model execution was performed using the batch processing facility in DOS, with *ad hoc* linking routines used to read vectors of random variates and generate new model input parameter files for each Monte Carlo trial. An output storage routine was used to read the desired output for each trial and write it to a file which was then appended to a main output file.

General purpose pre-processor module for Monte Carlo simulation

Software for generating independent random variates from probability distributions are widely available (MINITAB, 1993; IMSL, 1987). Routines for generating random variates from the multivariate-normal distribution are also available (IMSL, 1987). Software for generating correlated random variates with non-normal marginal distributions has recently become available (Market Engineering Co., 1992). However, the software requires the Excel spreadsheet program (Microsoft, 1992) to run. A microcomputer based, stand-alone, Monte Carlo pre-processor module was developed for generating correlated random variates. The module, coded in FORTRAN, employs a modified multivariate-normal (MMN) approach suggested by Taylor and Bender (1989) to generate correlated random variates from several commonly used continuous probability distributions. A listing of the pre-processor module code is given in Appendix A. The module makes extensive use of public-domain routines available from RANLIB (Brown and Lovato, 1991). Distributions considered are the uniform, exponential, normal, lognormal, weibull, gumbel, gamma, beta, and triangular. To generate non-normal random vectors with a given correlation structure using the MMN procedure, the multivariate normal distribution is first used to obtain standard normal vectors with the same correlation structure. The cumulative normal distribution function of each component of the standard normal vectors is then evaluated and the

resulting correlated vectors distributed uniformly over zero and one used to generate variates from the desired marginal distributions employing the inverse transform method (Law and Kelton, 1991). In the routine, distributions which do not have closed-form expressions for the inverse cumulative distribution function are evaluated by numerical integration. The MMN method preserves the exact marginal distribution of each variable and closely approximates the correlations between the variables.

The module also allows truncation of distributions which are not bounded on both ends by re-scaling variates to the desired interval before applying the inverse transform (Law and Kelton, 1991). The input file specifies the name of the output file, the number of variates to be generated, distribution types and relevant parameters, and if the variables are correlated, the input correlation matrix. The module provides summary information on the expected and actual means and variances of variables, and correlations if specified. Example input and summary output files are given in Appendix A.

Direct Monte Carlo simulation in batch mode

A flowchart of the general procedure used to implement direct Monte Carlo simulation in batch mode is shown in Figure 3.1. Control over repeated model execution is provided by a Monte Carlo controller routine, which initializes linking routines at the start of a run and terminates the run once the required number of trials have been performed. Linking routines read input from a base parameter file and the file of random inputs, generated by the pre-processor module. The linking routines create new input files for each trial by substituting appropriate random inputs for parameters to be randomized in the base file.

Batch execution was found to be a flexible procedure for implementing direct Monte Carlo simulation of the GLEAMS model. The procedure may be used with general-purpose model

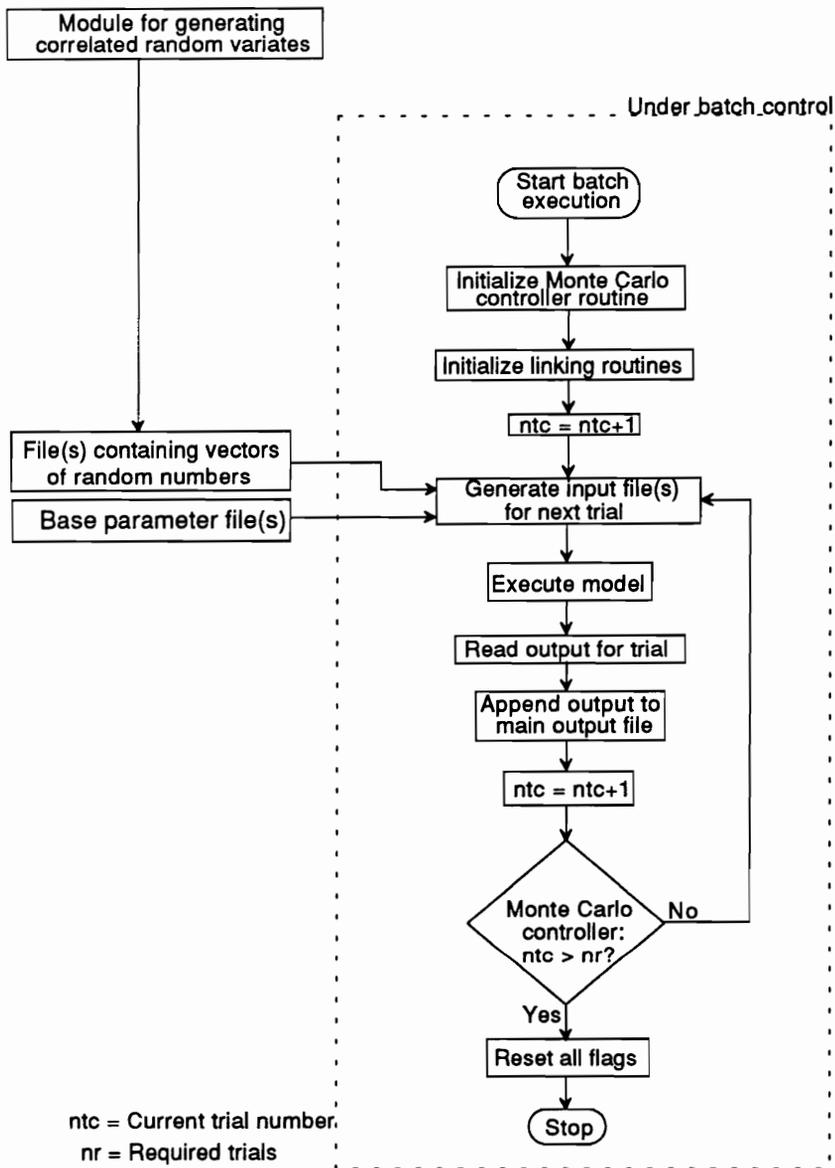


Figure 3.1 Flowchart of general procedure for implementing direct Monte Carlo simulation in batch mode.

interface linkages or *ad hoc* routines for randomization of a particular set of input parameters. Other than directly specifying model input and output file names, the procedure does not require changes to the source-code. This may be an important factor in insuring the integrity of the original model, as implementing routines for Monte Carlo simulation internally within a model usually requires extensive source-code changes and can potentially result in inadvertent modifications. Also, in some cases the source-code may not be available because of proprietary restrictions.

INTERACTIVE ASSESSMENT OF CORRELATION STRUCTURES WITH SUBJECTIVE INFORMATION

The presence of correlations in inputs implies that some combinations of inputs are more probable or less probable than they would be otherwise if inputs were assumed to be independent. In general, there is no *a priori* method of establishing the impact of correlated inputs on model output. This can be illustrated by examining the first-order approximation of model output variance with two correlated inputs:

$$\sigma_Y^2 = \left(\frac{\partial f}{\partial X_1} \right)^2 \sigma_{X_1}^2 + \left(\frac{\partial f}{\partial X_2} \right)^2 \sigma_{X_2}^2 + 2 \left(\frac{\partial f}{\partial X_1} \right) \left(\frac{\partial f}{\partial X_2} \right) \rho_{12} \sigma_{X_1} \sigma_{X_2} \quad (3.1)$$

where σ_Y^2 is the model output variance, f is the model function, $\sigma_{X_1}^2$ and $\sigma_{X_2}^2$ are the variances of the inputs X_1 and X_2 respectively, and ρ_{12} is the correlation between X_1 and X_2 . The third term in equation (3.1) represents the contribution of correlations in inputs to the output variance. Correlated inputs may result in an increase or a decrease in output variance, depending on the signs of the correlation between the inputs and the two partial derivatives.

Law and Kelton (1991) emphasized that correlations in input must be taken into account in any simulation study, and stated that ignoring input correlations can lead to serious inaccuracies in the results of a simulation. Song and Brown (1990), in a study of the Streeter-Phelps dissolved oxygen equation, found that the standard deviation of predictions with correlated inputs can be 20-40% larger than with independent inputs.

In general, to generate correlated variates, it is necessary for the input correlation matrix to be positive definite, which ensures that the matrix can be factorized. Information on correlations among input variables in Monte Carlo studies is sparse, and correlations are frequently determined using subjective considerations. Oren (1981) noted that a serious drawback of subjectively determining a correlation matrix through pairwise assessment of the relationships among the variables is that potential inconsistencies in the assessments can lead to a correlation matrix that is not positive definite. Ince and Buongiorno (1991) presented a simple procedure to test for and restore positive definiteness to an indefinite matrix, based on a recursive method for factorizing a positive definite matrix described by Scheuer and Stoller (1962). Their procedure is applicable to a specified covariance matrix, and requires increasing the input variance of variables by a small amount, keeping the covariance the same, so that the resulting matrix is positive definite. Hu and Olkin (1991) presented a numerical procedure based on the solution of a semi-infinite quadratic program that may be used to find a positive definite matrix closest to a given sample covariance matrix. These procedures are global in the sense that they generally result in changes to all entries in the observed or estimated covariance matrix. However, global procedures can potentially result in changes to input variances and correlations that are known or have been estimated with high confidence.

To overcome these limitations, an interactive procedure for generating a correlation matrix is developed that guarantees a positive definite matrix and allows use of known correlations and

subjective estimates for unknown correlations. Correlations that are known or estimated with a high degree of confidence are retained, and these correlations may be used to assess possible values for other unknown or poorly known correlations.

Interactive generation of a correlation matrix

Scheur and Stoller (1962) describe a recursive algorithm for factorizing a positive definite matrix, \mathbf{S} of order n so that $\mathbf{C}\mathbf{C}^T=\mathbf{S}$, where \mathbf{C} is the Cholesky factor of \mathbf{S} , as follows

$$c_{i1} = \sigma_{i1} / \sqrt{\sigma_{11}} \quad (1 \leq i \leq n) \quad (3.2.1)$$

$$c_{ii} = \sqrt{\sigma_{ii} - \sum_{k=1}^{i-1} c_{ik}^2} \quad (1 < i \leq n) \quad (3.2.2)$$

$$c_{ij} = \left[\sigma_{ij} - \sum_{k=1}^{j-1} c_{ik} c_{kj} \right] / c_{jj} \quad (1 < j < i \leq n) \quad (3.2.3)$$

$$c_{ij} = 0 \quad (i < j \leq n) \quad (3.2.4)$$

where the c_{ij} 's are elements of \mathbf{C} and the σ_{ij} 's are the elements of \mathbf{S} .

Let the elements of a correlation matrix of n variables, \mathbf{P} , be denoted by p_{ik} . Noting that the summation term in equation (3.2.2) must be less than one and that all diagonal terms in a correlation matrix are equal to one, equations (3.2.1) through (3.2.4) lead to the following condition for each row, i , of the correlation matrix that must be satisfied for positive definiteness

$$\sum_{k=1}^{i-1} \frac{\left[p_{ik} - \sum_{j=1}^{k-1} c_{ij} c_{jk} \right]^2}{1 - \sum_{j=1}^{k-1} c_{kj}^2} < 1 \quad (2 \leq i \leq n) \quad (3.3)$$

where the c_{ij} 's are the elements of the Cholesky factor of \mathbf{P} determined recursively with the starting relationship $c_{21} = p_{21}$. Equation (3.3) can then be used to compute the upper and lower limits of any $(i,k)^{\text{th}}$ entry in a correlation matrix, given the correlations of i and k with previously entered variables in the matrix.

Since the correlation matrix is symmetric, only the lower (or upper) triangular entries are needed. Equation (3.3) results in expressions for each row of the correlation matrix, and allows computation of the upper and lower bounds for any one correlation entry in the row, given all other correlations in that and preceding rows. This fact can be used to advantage in the sequential, row-wise generation of a positive definite correlation matrix by assigning values to correlations that are not known with a high degree of confidence, and examining the bounds imposed on the correlation value that is known with some confidence. If the bounds on the correlation value that is known with some degree of confidence are satisfactory, the subjectively assigned correlations may be accepted, otherwise new values need to be assigned and the evaluation repeated.

A simple example with three variables will serve to illustrate the utility of Equation (3.3).

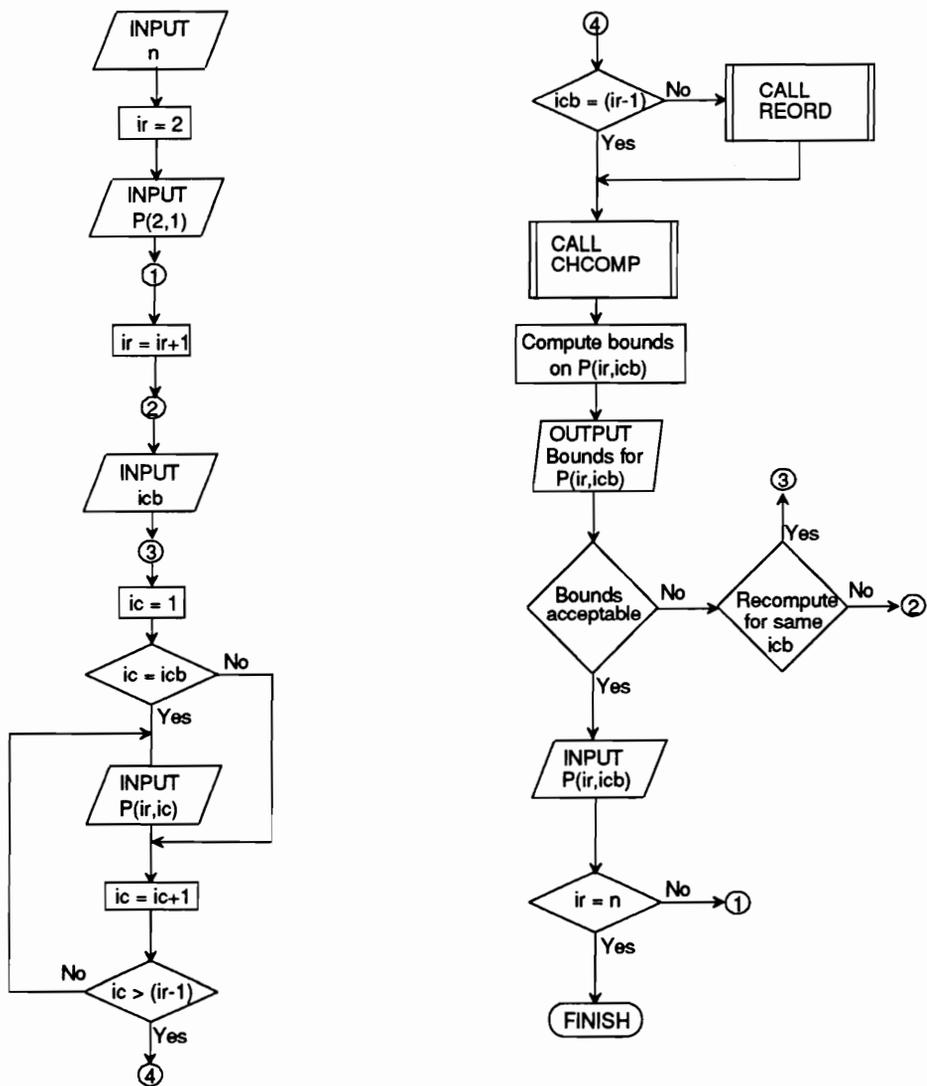
Expanding equation (3.3) with $i=3$ results in

$$p_{31}^2 + \frac{(p_{32} - p_{31}p_{21})^2}{1 - p_{21}^2} < 1 \quad (3.4)$$

Assuming p_{21} and p_{31} have been assigned values of 0.9 and -0.9 respectively, equation (3.4) results in p_{32} being bounded within the open interval (-1,-0.62). On the other hand, if values for both p_{21} and p_{31} are equal to 0.9, the bounds on p_{32} change to (0.62,1).

Equation (3.3) results in an expression that is quadratic in all p_{ik} entries. The form of the expression allows quick solution for the $(i,i-1)^{\text{th}}$ entry. However, as the number of variables increase, direct solution for other entries rapidly becomes laborious. This problem can be easily circumvented by using a dummy correlation matrix, in which variables are re-ordered so that bounds are always computed for the $(i,i-1)^{\text{th}}$ entry.

The procedures for computing the bounds for any entry in a row of a correlation matrix, given all other entries in that and preceding rows (considering only lower triangular values), was implemented in an interactive computer routine. A flowchart of the routine, which allows interactive, row-wise generation of a positive definite correlation matrix is shown in Figure 3.2. A listing of the routine is given in Appendix B. The routine initially requires input on the number of variables and the p_{21} correlation. Starting with row 3, the column number for which bounds are to be computed is entered. The routine then prompts for all other correlations in the row. If the correlations entered do not violate the condition required for positive definiteness, bounds on the requested correlation using equation (3.4) are displayed, otherwise the routine re-prompts for the correlations. Equation (3) is solved as an equality, with the right hand equal to $1-\epsilon$, where ϵ is a small positive number, thus insuring that the computed bounds represent a closed interval. The FORTRAN compiler used (Microsoft, 1991) provides an intrinsic function, EPSILON, which returns the smallest increment that when subtracted from one, produces a number smaller than one. The value returned by EPSILON in single precision is 1.192093×10^{-7} . Before displaying the computed bounds to the user, 0.0005 is subtracted from and added to the upper and lower bounds respectively. This prevents the entry of correlation values outside the computed bounds



n = Number of variables
 ir = Current row number
 $P(i,j)$ = Correlation matrix (order n)
 icb = Column number of current row for which bounds are to be computed
 ic = Current column number

REORD: Subroutine for reordering variables so that $icb = (ir-1)$.
 CHCOMP: Subroutine for computing elements of the Cholesky factor of P .

Figure 3.2 Flowchart of routine for interactive generation of a positive definite correlation matrix.

due to round-off error, since the bounds are displayed to three significant digits. If the bounds displayed are satisfactory, a value contained within the bounds may be selected and entered. Values outside the computed bounds are not accepted. If the bounds are not satisfactory, values for other correlations in the row may be reassigned and the bounds on the requested correlation computed again, or bounds on another entry in the row computed. This procedure is repeated for all rows in the matrix. Whenever the conditions for positive definiteness are violated, or the bounds computed for an entry are not satisfactory, the routine allows all correlation values to be re-entered.

The routine was verified by using it to compute bounds on the largest entry in the rows (starting with the third row) of a subjectively estimated positive definite correlation matrix reported in the literature. Table 3.1 shows a positive definite correlation matrix with nine variables used by Song and Brown (1990) in a study of the Streeter-Phelps dissolved oxygen (DO) equation, referred to hereafter as the DO matrix. The bounds obtained for the largest entry in each row (starting with the third row) of the DO matrix are compared with the values used by Song and Brown (1990) in Table 3.2.

For a positive definite matrix, the value used for any entry in the matrix must fall within the bounds computed for that entry. As can be seen from Table 3.2, the value used for the largest entry in all rows of the DO matrix fall in the bounds computed for that entry. This data set provides one test for verification of the routine. Additional verification of the routine was carried out by using it to generate hypothetical correlation matrices and independently checking the generated matrices for positive definiteness with a Cholesky factorization routine, CHFAC, available from IMSL (1987).

Song and Brown (1990) commented that the values used in the DO matrix were obtained after adjusting intuitive initial estimates based on purely qualitative criteria. The quality of the

Table 3.1 DO correlation matrix (Song and Brown, 1990).

1.0								
0.0	1.0							
0.3	0.0	1.0						
0.6	0.0	0.3	1.0					
0.3	0.0	0.6	0.8	1.0				
0.0	0.0	0.0	0.2	0.1	1.0			
0.0	0.0	0.0	0.0	0.0	0.0	1.0		
0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	
0.0	0.0	0.0	0.3	0.2	0.3	0.0	0.0	1.0

Table 3.2 Bounds for DO matrix

Row	Column	Value used	Computed bounds
3	1	0.300	[-0.999,0.999]
4	1	0.600	[-0.819,0.999]
5	4	0.800	[-0.376,0.871]
6	4	0.200	[-0.275,0.452]
7	1	0.000	[-0.602,0.602]
8	1	0.000	[-0.602,0.602]
9	6	0.300	[-0.723,0.990]

correlation estimates in the DO matrix could, ostensibly, be improved if the interactive approach developed here was used to generate the correlation matrix. The interactive approach makes it possible to assess several correlation structures which are considered feasible, and allows rational selection of correlations about which little is known, using information on correlations which are known with some confidence. This is illustrated in an example application of the procedure to a subjectively determined nonpositive definite correlation matrix.

Application to a nonpositive definite matrix

Ince and Buongiorno (1991) reported a Monte Carlo study of the economic performance of a new pulping and paper making technology. In the study, possible correlations between ten input variables were considered, and obtained from expert opinion. The resulting correlation matrix, shown in Table 3.3, is not positive definite. The correlation matrix was modified using the interactive procedure developed here.

As the first step, bounds were obtained for the largest (absolute value) entry in each row of the correlation matrix. If the computed bounds for a particular entry contained the value reported by Ince and Buongiorno (1991) for that entry, this was the value entered. If the value reported for an entry did not lie within the computed bounds (which would be the case for at least one row in a nonpositive definite matrix), the bound limit closest to the value reported was used for the entry.

The results of the initial assessment are shown in Table 3.4. Two correlation values (p_{65} and p_{75}) were found to be outside the computed bounds. The bounds for the p_{65} correlation indicated that the value reported needed to be changed to -0.268 to restore positive definiteness. With this change in the p_{65} correlation, the bounds for the p_{75} correlation showed that the reported value should be increased to 0.763 to restore positive definiteness. Selecting a limiting value for a given correlation entry will generally result in minimizing the width of bounds computed for

Table 3.3 Capital investment correlation matrix (Ince and Buongiorno, 1991).

1.0										
0.4	1.0									
0.7	0.0	1.0								
0.0	0.0	0.0	1.0							
0.0	0.0	0.0	-0.7	1.0						
0.0	0.0	0.0	-0.5	-0.5	1.0					
0.0	0.0	0.0	-0.7	0.7	0.0	1.0				
0.0	0.3	0.0	0.0	0.0	0.0	0.0	1.0			
0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.5	1.0		
0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.5	0.5	1.0	

Table 3.4 Results of initial assessment of capital investment matrix.

Row	Column	Reported value	Computed bounds	Value used
3	1	0.700	[-0.916,0.916]	0.700
4	1	0.000	[-0.591,0.591]	0.000
5	4	-0.700	[-0.999,0.999]	-0.700
6	5	-0.500*	[-0.268,0.968]	-0.268
7	5	0.700*	[0.763,0.794]	0.763
8	2	0.300	[-0.828,0.828]	0.300
9	8	0.500	[-0.737,0.999]	0.500
10	2	0.800	[-0.414,0.814]	0.800

*Reported value is not contained within the computed bounds.

correlation values in subsequent rows. For example, if the value entered for the p_{65} correlation is changed from -0.268 to 0.350, the bounds for the p_{75} correlation change from [0.763,0.794] to [0.070,0.910]. Thus, when carrying out an initial assessment in practice, it may be desirable to re-order variables in the correlation matrix so that correlations estimated with higher confidence appear in earlier rows.

The initial assessment shows how the procedure developed here may be used to make minimum changes to restore positive definiteness to a subjectively determined correlation matrix. However, the main advantage of the procedure is that correlations estimated with a high degree of confidence can be preserved, and further, these correlations may be used to examine possible values for other correlations which may not be known with high confidence. For the correlation matrix under consideration, we assume, for the purpose of illustrating this point, that the p_{75} correlation has been estimated with higher confidence than the p_{65} correlation. This would imply that the p_{65} correlation could be reduced beyond the required minimum change, and that it would be desirable to retain the original p_{75} correlation. When the p_{65} correlation was gradually reduced beyond the minimum required change (with other correlation values in rows 6 and 7 kept the same), it was found that using a value of -0.25 for the correlation resulted in bounds of [0.669,0.871] for p_{75} . When values of -0.25 and 0.7 were used for the p_{65} and p_{75} correlations with all other correlations kept the same, bounds for the p_{82} , p_{98} , and $p_{10,2}$ correlations did not change from those obtained during the initial assessment. Thus, if it were desired not to change the original p_{75} correlation, the only change necessary is to reduce the original estimate for the p_{65} correlation by about 50% (from -0.5 to -0.25).

The example application shows how the interactive nature of the procedure developed here may be used to advantage in assessing possible correlation structures from subjective information. The procedure does not require changes to correlations that are known or estimated with high confidence, and allows the use of these correlations to assess possible values of correlations

that are not known. The method allows interactive generation of a correlation matrix that is guaranteed to be positive definite.

SIMULATION STUDY

GLEAMS was used in a dynamic Monte Carlo framework to simulate annual runoff, percolation, sediment and pesticide losses from two alternative land-use practices in Rockingham County, Virginia. The two scenarios considered were a corn cropping system under conservation tillage with a winter rye crop planted after disking (Scenario 1), and a conventionally tilled corn cropping system with no winter cover crop (Scenario 2). These scenarios are typical of actual production systems in Rockingham County (Zhu, 1994).

An important distinction between this simulation study and previous Monte Carlo studies is that climatic inputs are fixed for all trials, with rainfall variability taken into account by employing long-term simulation. A 44 year historic rainfall record (1949-1992) obtained from the Virginia Water Resources Research Center (VWRRRC, 1993) was used. Mean monthly maximum and minimum temperatures used were based on a 26 year (1951-1975) record as reported in the Soil Survey of Rockingham County (Hockman et al., 1982). Mean monthly radiation values used were those given for Richmond, Virginia given in the GLEAMS user manual.

The following steps were employed in the simulation study:

1. The validity of proposed statistical tests on output from a continuous simulation model was established.

2. A sensitivity analysis was carried out to determine which inputs were to be regarded as uncertain. This was deemed necessary because model outputs may become sensitive to particular parameters only at some threshold value.

3. Monte Carlo simulations were carried out to quantify the impact of parameter uncertainty on absolute and comparative model predictions.

Procedures for model output analysis

It will be instructive at this point to consider the nature of output resulting from a continuous simulation model in both a deterministic and a direct Monte Carlo framework. Deterministic output from a continuous field-scale NPS pollution model is generally a time-series of pollutant loadings. An output vector, \mathbf{O} , may be written as $[y_1, y_2, \dots, y_m]^T$ where the y_i 's represent pollutant loadings of interest at a specified time step. For example, if below root-zone pesticide leaching for a given management scenario was being examined, the y_i 's could be annual percolation of pesticide mass in g/ha, and m would be the number of years of simulation. The output vector in a continuous deterministic simulation reflects weather variability. However, \mathbf{O} really only represents an 'average' level of expected loadings, given the uncertainty in soil, crop, and chemical inputs which have been assumed to be constants. In a given year, we would expect the deterministic loading level to be exceeded approximately half the time. If uncertain inputs are characterized by probability distributions, a MCS of the continuous simulation model will result in an output matrix, \mathbf{O}_M , as follows

$$\begin{bmatrix} y_{11} & \cdots & \cdots & y_{1m} \\ y_{21} & \cdots & \cdots & y_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & \cdots & \cdots & y_{nm} \end{bmatrix} \quad (3.5)$$

where n is the number of runs, or Monte Carlo trials. The column entries represent output variability due to parameter uncertainty. Each row in \mathbf{O}_M is a different realization of \mathbf{O} obtained from different sets of possible soil, crop, and chemical inputs.

A possible approach to re-ordering the output matrix to allow a more intuitive interpretation would be to rank values in each column to obtain pollutant loading exceedence probabilities resulting from parameter uncertainty. Then, to obtain the dynamic loading at any given exceedence probability, the appropriate values from the ranked columns are extracted. However, this approach can only be justified if the values in any given row of the matrix are independent. In addition, independence of values in each row is required before any statistical tests can be performed on the output data. Hence, it is necessary to determine whether there is any serial correlation in the time-series of annual output from the model. To determine whether annual output from GLEAMS could be regarded as independent, the runs up and down test with a two-sided alternative (Gibbons, 1985) was carried out on output from deterministic simulations of the scenarios previously identified. Output from the first two years was discarded to allow for 'warm-up'. Lag-1 serial correlations were also computed for all outputs. Inputs in the deterministic simulations were literature derived best estimates.

The Monongahela fine sandy loam (0-7%) soil phase was selected as a representative soil type from map sheets in the Soil Survey of Rockingham County (SSRC). The Monongahela series is a hydrologic group C soil and is classified as a fine-loamy, mixed, mesic, typic Fragiudult. Soil textural and water retention properties were estimated from information given in the SSRC and regression equations reported by Rawls and Brakensiek (1982). Organic matter content at different depths was estimated using the surface organic matter content reported in the SSRC and a relationship suggested by Zhang et al. (1993). A 1 ha field area with 4% slope and length to width ratio of 2.0 was assumed. An effective rooting depth of 76cm was used, with the fraction of soil water available at the beginning of simulation taken to be 0.5. The model was modified to

allow runoff curve number to be input by season, which was estimated from values given in the GLEAMS manual. Inputs for the erosion component were estimated from the SSRC, Wischmeier and Smith (1978), and Engman (1986). For both scenarios, only losses of the pesticide Metolachlor were considered. The recommended application rate of Metolachlor for both scenarios (VCE, 1992) is the same (2.1 kg/ha active ingredient). Chemical properties for Metolachlor were obtained from a pesticide database compiled by Wauchope et al. (1992). Parameters used to represent differences between the scenarios were runoff curve number, the soil loss ratio (crop factor), Manning's roughness factor, and relative fractions of pesticide applied to the soil and crop residue.

The results of the analysis to assess whether annual output from GLEAMS could be regarded as independent are given in Table 3.5. The p-values obtained from the runs up and down tests clearly indicate that the null hypothesis of a random sequence of values cannot be rejected for any of the outputs. This is also supported by the lag-1 correlation values. The absolute values of the correlations are all considerably smaller than e^{-1} , the cutoff value below which Gutjhar (1988) has suggested observations can be regarded as independent. Thus, it was concluded that the re-ordering procedure for the output matrix O_M suggested previously is valid.

When analyzing output from two or more alternative scenarios, it is necessary to determine whether the output data are independent from each other. An association test employing Spearman's rank correlation coefficient with a two-sided alternative (Gibbons, 1985) was carried out on output from the two scenarios. The p-values obtained from the association tests (Table 3.6) strongly reject the null hypothesis of no correlation. It appears, therefore, when analyzing output from two or more alternative scenarios, it is important to use statistical procedures for paired or related samples. Gibbons (1985) notes that in an analysis of paired observations, it is technically improper and ordinarily disadvantageous to disregard the pairing. Gibbons (1985) further notes that the most convenient way to take advantage of the pairing is to subtract the two

Table 3.5. Results of analysis for independence of annual output from GLEAMS.

Output variable	Scenario	Number of Runs	p-value	lag-1 correlation
Runoff	1	26	0.533	-0.023
Percolation	1	30	0.383	0.158
Sediment loss	1	32	0.105	-0.137
Surface pesticide losses	1	28	0.901	-0.014
Sub-surface pesticide losses	1	28	0.901	0.090
Runoff	2	26	0.533	-0.025
Percolation	2	28	0.901	-0.013
Sediment loss	2	26	0.533	0.170
Surface pesticide losses	2	26	0.533	-0.013
Sub-surface pesticide losses	2	30	0.383	0.146

Table 3.6 Results of association and sign tests on deterministic output for scenarios 1 and 2.

Output variable	Rank Correlation	p-value for association test
Runoff	0.948	< 0.001
Percolation	0.923	< 0.001
Sediment loss	0.843	< 0.001
Surface pesticide losses	0.866	< 0.001
Sub-surface pesticide losses	0.939	< 0.001

measurements within each pair and use only the values of these differences for the analysis. Appropriate parametric and nonparametric tests for analysis of paired data are the paired-t and sign tests respectively (Gibbons, 1985).

Sensitivity analysis

GLEAMS parameters that were identified as uncertain from physical considerations are listed in Table 3.7. The relative sensitivity of model outputs of interest to these parameters was determined, using parameter estimates for scenario 1 as base values. The median relative sensitivity of outputs to parameters over the simulation period is given in Table 3.8.

If the median relative sensitivity of all outputs to a parameter was zero, the sensitivity of all outputs when the parameter was equal to its assumed maximum and minimum value were computed. If these sensitivities were also zero, the parameter value was fixed at its best-estimate value for the purpose of Monte Carlo simulations. Based on this criteria RC, BST, all SATK, CLAY(2), CLAY(3), SILT(2), SILT(3), and H2OSOL values were not randomized.

Monte Carlo simulations

The uncertainty in parameters identified in the sensitivity analysis was characterized by the range reported for the parameters in the literature. Table 3.9 shows the range for uncertain parameters which were assumed to be the same for scenarios 1 and 2. The ranges for parameters used to simulate differences between the scenarios is given in Table 3.10.

Three Monte Carlo simulations were carried out. The first simulation was based on the premise that prior to the introduction of any qualitative or quantitative information to a modeling exercise,

Table 3.7 Uncertain parameters in GLEAMS and abbreviations.

Parameter description	Abbreviation
Effective saturated hydraulic conductivity of soil horizon below root zone	RC
Fraction of plant available water in the soil when simulation begins	BST
Soil evaporation parameter	CONA
Porosity by horizon	POR()
Field capacity by horizon	FC()
Immobile water content by horizon	BR15()
Saturated hydraulic conductivity by horizon	SATK()
Organic matter content by horizon	OM()
Clay content by horizon	CLAY()
Silt content by horizon	SILT()
Runoff curve number by season	CN()
Soil erodibility	KSOIL
Soil loss ratios for rotation	CFACT()
Manning's roughness factors for rotation	NFACT()
Pesticide solubility	H2OSOL
Pesticide partitioning coefficient	KOC
Pesticide foliar half-life	HAFLIF
Pesticide soil half-life	SOLLIF
Pesticide wash-off fraction	WSHFRC
Coefficient for plant pesticide uptake	COFUP
Fraction of pesticide applied to crop residue	FOLFRC
Fraction of pesticide applied to soil	SOLFRC

() indicates multiple values may be input by horizon or crop rotation.

Table 3.8 Median relative sensitivity of uncertain parameters.

Parameter	Median relative sensitivity				
	Runoff	Percolation	Sediment yield	Surface pesticide losses	Subsurface pesticide losses
RC	0.000	0.000	0.000	0.000	0.000
BST	0.000	0.000	0.000	0.000	0.000
CONA	-0.481	-2.302	-0.250	-0.262	-3.907
POR(1)	-1.559	0.628	-0.824	-2.750	2.444
POR(2)	-0.532	0.271	-0.252	-0.672	2.010
POR(3)	-0.036	0.022	-0.016	-0.080	0.528
FC(1)	1.400	-1.616	0.821	2.144	-8.435
FC(2)	0.432	-1.706	0.210	0.565	-3.201
FC(3)	0.024	-0.671	0.010	0.059	-0.702
BR15(1)	-0.283	0.566	-0.138	-0.511	1.444
BR15(2)	-0.076	0.786	-0.028	-0.157	0.691
BR15(3)	-0.002	0.304	-0.001	-0.016	0.153
SATK(1)	0.000	0.000	0.000	0.000	0.000
SATK(2)	0.000	0.000	0.000	0.000	0.000
SATK(3)	0.000	0.000	0.000	0.000	0.000
OM(1)	0.000	0.000	0.000	1.569	-1.268
OM(2)	0.000	0.000	0.000	0.000	-1.837
OM(3)	0.000	0.000	0.000	0.000	-0.609
CLAY(1)	0.000	0.000	-0.096	0.000	0.000
CLAY(2)	0.000	0.000	0.000	0.000	0.000
CLAY(3)	0.000	0.000	0.000	0.000	0.000
SILT(1)	0.000	0.000	0.888	0.000	0.000
SILT(2)	0.000	0.000	0.000	0.000	0.000
SILT(3)	0.000	0.000	0.000	0.000	0.000
CN(1)	2.979	-0.728	3.914	0.313	-4.233
CN(2)	4.243	-0.724	1.695	8.263	-4.914
KSOIL	0.000	0.000	0.997	0.003	0.000
CFACT(1)	-0.170	-0.175	0.042	0.000	-0.185
CFACT(2)	0.000	0.000	0.036	0.000	0.000
CFACT(3)	0.000	0.000	0.000	0.000	0.000
CFACT(4)	0.000	0.000	0.237	0.002	0.000
CFACT(5)	0.000	0.000	0.000	0.000	0.000
CFACT(6)	0.000	0.000	0.066	0.000	0.000
CFACT(7)	0.000	0.000	0.138	0.000	0.000
NFACT(1)	0.000	0.000	-0.003	0.000	0.000
NFACT(2)	0.000	0.000	-0.010	-0.001	0.000
H2OSOL	0.000	0.000	0.000	0.000	0.000
HAFLIF	0.000	0.000	0.000	0.000	0.074
KOC	0.000	0.000	0.000	0.000	-4.337
SOLLIF	0.000	0.000	0.000	0.000	0.459
WSHFRC	0.000	0.000	0.000	0.000	-0.330
COFUP	0.000	0.000	0.000	0.511	2.722
FOLFRC	0.000	0.000	0.000	0.422	0.402
SOLFRC	0.000	0.000	0.000	0.557	0.552

Table 3.9 Range for uncertain parameters assumed to be the same for scenarios 1 and 2.

Parameter	Minimum	Maximum
CONA (mm day ^{-1/2})	3.300	4.500
POR(1) *	0.472	0.547
POR(2)	0.434	0.509
POR(3)	0.396	0.509
FC(1)	0.270	0.385
FC(2)	0.262	0.398
FC(3)	0.288	0.398
BR15(1)	0.076	0.188
BR15(2)	0.116	0.219
BR15(3)	0.116	0.214
% OM(1)	0.000	1.693
% OM(2)	0.000	1.135
% OM(3)	0.000	0.813
% CLAY(1)	10.00	27.00
% SILT(1)	18.00	80.00
KSOIL (Mg/ha)	0.157	1.165
HAFILF (days)	0.600	10.00
KOC (ml/g)	99.0	307.0
SOLLIF (days)	12.00	180.0
WSHFRC	0.550	0.860
COFUP	0.000	1.000

* No units indicate a parameter is dimensionless

Table 3.10 Ranges for parameters used to simulate differences between scenarios 1 and 2.

Parameter	Scenario 1		Scenario 2	
	Minimum	Maximum	Minimum	Maximum
CN(1)	80.00	86.00	89.00	92.00
CN(2)	80.00	87.00	83.00	88.00
CFAC(1)	0.270	0.330	0.423	0.517
CFAC(2)	0.117	0.143	0.549	0.671
CFAC(3)	0.108	0.132	0.459	0.561
CFAC(4)	0.108	0.132	0.378	0.462
CFAC(5)	0.081	0.099	0.234	0.286
CFAC(6)	0.504	0.616	0.423	0.517
CFAC(7)	0.387	0.473	-	-
NFAC(1)	0.140	0.530	0.140	0.530
NFAC(2)	0.160	0.470	-	-
FOLFRC	0.500	0.650	-	-
SOLFRC	0.200	0.340	0.700	0.990

any parameter set combination that predicts the variables of interest must be considered equally likely as a simulator of the system (Beven and Binley, 1992). Therefore, in the first simulation independent inputs uniformly distributed over their ranges were used. In the second simulation, possible correlations among inputs were considered. Correlations reported by Zhang et al. (1993) and other correlations assigned from subjective considerations were used to generate an input correlation matrix employing the interactive procedure developed earlier. The final correlations used are summarized in Table 3.11. In the third simulation, the model sensitivity to input distribution shape was assessed by using skewed beta distributions for inputs. The absolute value of the coefficient of skew was 0.64, and shape parameters were selected so that all inputs had the same inter-quartile range as the first simulation.

Results and discussion

In general, the number of trials in a MCS will depend on the desired length of confidence interval (at a given confidence level) for the output parameter of interest. In the dynamic MCS carried out here, the objective was not to estimate the mean or variance of outputs in a given year at some pre-specified confidence level. Rather the dynamic variation in outputs at given exceedence probabilities were obtained using the re-ordering procedure for the MCS output matrix outlined earlier. Outputs were obtained at exceedence probabilities of 50% and 10%, which were taken to correspond to 'average' and 'critical' conditions respectively.

Figures 3.3a and 3.3b show the change in the 50% and 10% exceedence probability (EP) pesticide leaching levels over the length of simulation, plotted against the number of trials. It can be seen that the leaching levels appear to have stabilized after approximately 1600 trials. Therefore, 2000 trials were assumed to be sufficient to characterize the 50% and 10% EPs of model outputs.

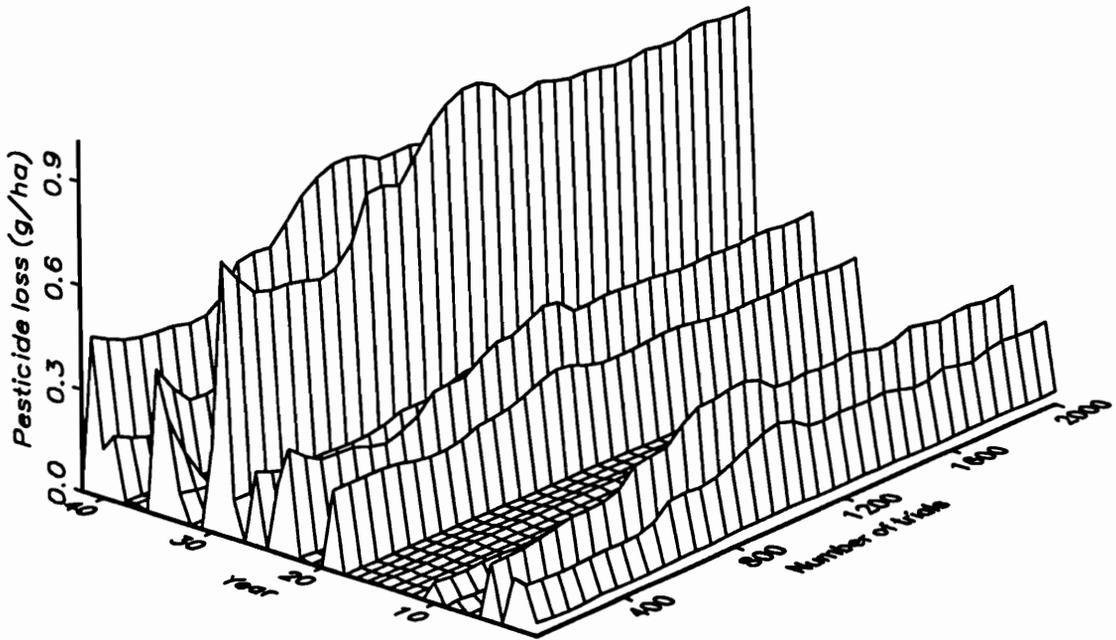
Table 3.11 Correlations between selected input parameters.

Input variables	Correlation
POR(i)/FC(i)	0.2
POR(i)/BR15(i)	0.2
POR(i)/OM(i)	0.3
FC(i)/BR15(i)	0.8
FC(i)/OM(i)	0.4
BR15(i)/OM(i)	0.6
FC(j)/FC(j+1)	0.1
BR15(j)/BR15(j+1)	0.1
OM(j)/OM(j+1)	0.2
CLAY(1)/FC(1)	0.7
SILT(1)/FC(1)	0.2
CLAY(1)/BR15(1)	0.7
SILT(1)/BR15(1)	0.2
CLAY(1)/OM(1)	0.2
SILT(1)/OM(1)	0.2
CFACT(k)/CFACT(k+1)	0.3
NFACT(1)/NFACT(2)	0.3
SOLFRC/FOLFRC	0.5

i, j = Horizon numbers ($i = 1$ to 3 , $j = 1$ to 2)

k = Crop rotation period (1 to 7 for scenario 1, 1 to 6 for scenario 2)

(a)



(b)

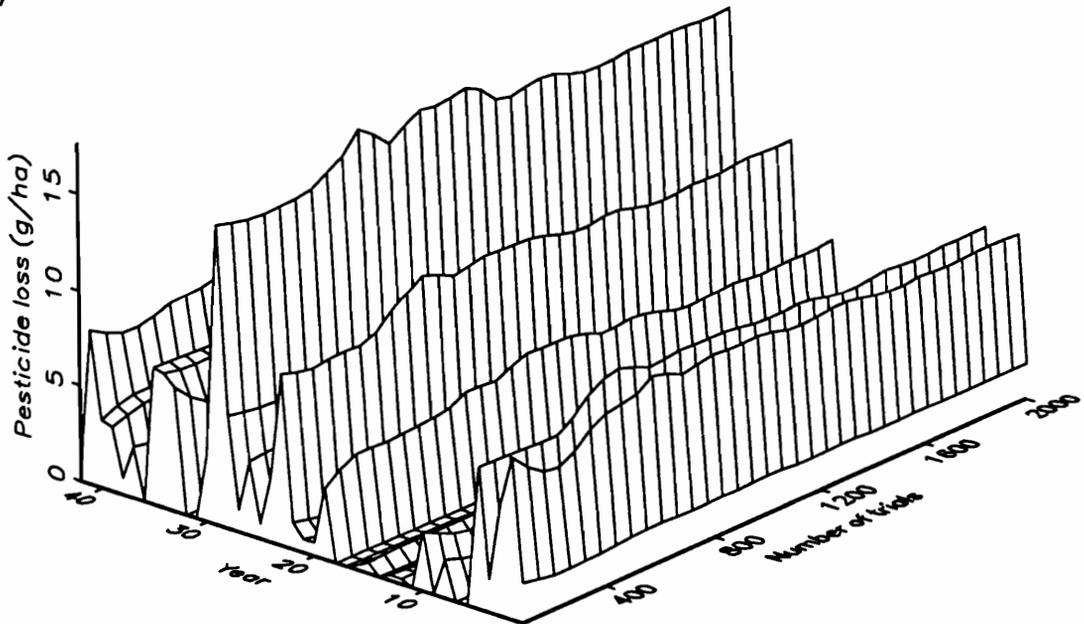


Figure 3.3 Convergence of pesticide leaching over simulation period with increasing number of trials: (a) 50% exceedence probability levels (b) 10% exceedence probability levels.

It has been previously established that annual output from GLEAMS can be regarded as independent, and therefore it is appropriate to compute summary statistics for annual output. Tables 3.12 and 3.13 show the results of deterministic simulations that provide a basis for comparison with the MCS assuming uniform, independent inputs. The value of uncertain inputs in the deterministic simulations were taken to be the mid-point of the ranges given in Tables 3.9 and 3.10. The results for differences between scenarios were obtained by subtracting scenario 1 outputs from scenario 2 outputs. It can be seen that the mean and medians are different for all outputs, with the largest differences evident for water and pesticide percolation. This indicates that outputs follow a skewed distribution. Therefore, the median and inter-quartile range (IQR) would be preferred to using the mean and standard deviation as measures of central tendency and dispersion. In addition, statistical tests for significant differences between the two scenarios which require that observations come from a symmetric distribution would not be valid. The nonparametric sign test (Gibbons, 1985) which does not require the assumption of symmetry was used to test for significant differences between the two scenarios. The sign test p-values for the deterministic simulation are also shown in Table 3.13. Significant differences are indicated for all outputs at the 0.05 level.

Monte Carlo simulation with independent inputs

The results of MCS assuming uniformly distributed, independent inputs for scenario 1 is shown in Table 3.14. It can be seen that the medians and IQRs for runoff, percolation, and sediment loss at the 50% EP correspond approximately to the deterministic values, indicating a linear response of these output variables over the range of inputs. However, both the medians and IQRs for surface and sub-surface pesticide losses at the 50% EP are fairly different from the deterministic values, indicating non-linearity in the response of these outputs over the range of inputs. The IQRs for all outputs at the 10% EP are higher than those at the 50% EP. Intuitively, this can be explained by saying that 'critical conditions', corresponding to the 10% EP outputs, would exhibit

Table 3.12 Results from deterministic simulation with statistics based on annual values over a 44 year simulation (Scenario 1).

Output variable	Mean	Median	Standard deviation	Inter-quartile range
Runoff (cm)	3.206	3.010	1.732	2.410
Percolation (cm)	3.711	2.562	3.875	5.921
Sediment loss (Mg/ha)	0.273	0.214	0.239	0.198
Surface pesticide losses (g/ha)	2.959	2.101	3.469	2.849
Pesticide percolation (g/ha)	0.1275	0.0223	0.2057	0.1871

Table 3.13 Differences between scenarios 2 and 1, deterministic simulations.

Output variable	Mean	Median	Standard deviation	Inter-quartile range	Sign test p-value
Runoff (cm)	3.365	3.131	1.298	1.526	< 0.001
Percolation (cm)	-1.392	-0.529	2.026	2.321	0.002
Sediment loss (Mg/ha)	0.553	0.566	0.223	0.207	< 0.001
Surface pesticide losses (g/ha)	4.963	4.379	3.786	2.439	< 0.001
Pesticide percolation (g/ha)	-0.0599	-0.0043	0.1099	0.1019	0.002

Table 3.14 Results of Monte Carlo simulation with independent inputs (scenario 1).

Output variable	50% exceedence probability		10% exceedence probability	
	Median	Inter-quartile range	Median	Inter-quartile range
Runoff (cm)	3.157	2.501	4.449	3.522
Percolation (cm)	2.548	5.930	5.342	8.330
Sediment loss (Mg/ha)	0.214	0.202	0.404	0.376
Surface pesticide losses (g/ha)	1.423	1.927	5.680	6.520
Pesticide percolation (g/ha)	0.0145	0.235	1.465	5.657

Table 3.15 Differences between scenarios 1 and 2, Monte Carlo simulations with independent inputs.

Output variable	50% exceedence probability			10% exceedence probability		
	Median	Inter-quartile range	Sign test p-value	Median	Inter-quartile range	Sign test p-value
Runoff (cm)	3.045	1.504	< 0.001	3.430	1.986	< 0.001
Percolation (cm)	-0.661	2.683	< 0.001	-0.721	3.233	0.003
Sediment loss (Mg/ha)	0.568	0.271	< 0.001	1.055	0.538	< 0.001
Surface pesticide losses (g/ha)	2.985	2.944	< 0.001	8.430	6.540	< 0.001
Pesticide percolation (g/ha)	-0.0042	0.1415	< 0.001	-0.241	1.636	0.0240

greater variability than 'average conditions', corresponding to the 50% EP outputs. In terms of the dynamic MCS model output, this implies that in addition to location shifts in the output distribution from year to year, there are also changes in the distribution shape and skewness. Intuitively, changes in skewness should impact the 10% EP output to a much greater degree than the 50% EP values. This was assessed by examining the variation in the coefficient of skew for distributions of various outputs over the simulation period (Figure 3.4). It can be seen that the variation in the coefficient of skew for pesticide percolation is much larger than the variation for runoff. This is reflected in the much larger relative increase in the IQR at the 10% EP for pesticide percolation compared to runoff. The relative increase in the IQR at the 10% EP for surface pesticide losses is intermediate, corresponding to the intermediate variation in coefficient of skew.

The results of the MCS for differences between scenarios, obtained as before, by pairing outputs is shown in Table 3.15. As stated in the introduction to this chapter, an important aspect of the MCS carried out here is to quantify the impact of parameter uncertainty on comparative analysis or relative model predictions. The following rationale was used to assess the impact of parameter uncertainty on relative model predictions. Consider the probability distribution of any output of interest resulting from parameter uncertainty for the first scenario in any given year. Suitable changes are made to the inputs to reflect another scenario, and the output distribution for the same year obtained with the same climatic inputs. If the output distribution for the second scenario exhibits only a location shift with respect to the output distribution for the first scenario, with no change in shape or skewness, the difference between the distributions at any EP will be the same. Thus, the impact of parameter uncertainty on relative model predictions was assessed by comparing estimators for output differences at different EPs. It was inferred that if parameter uncertainty did not impact the model, estimators for output differences at different exceedence probabilities would be similar. A more liberal interpretation of the contention that parameter

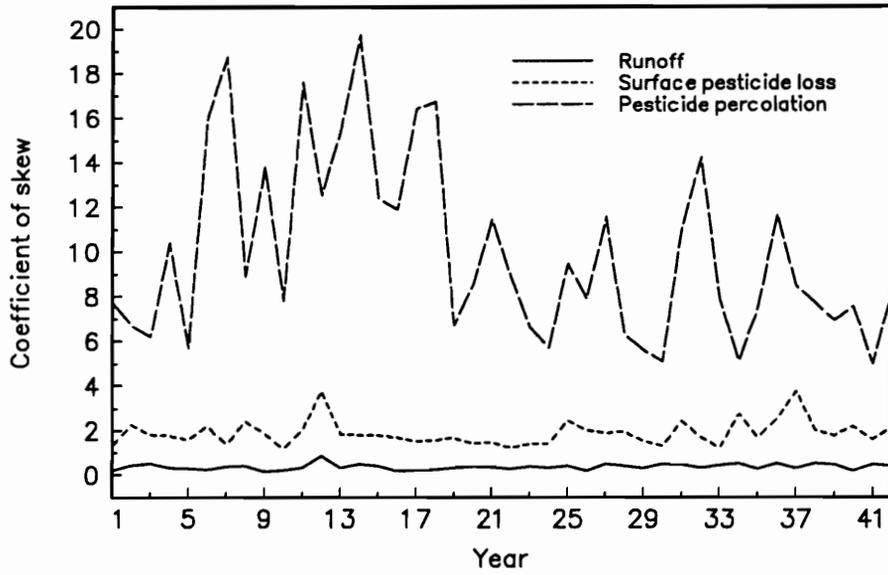


Figure 3.4 Variation in coefficient of skew over simulation period for various outputs with uniform, independent inputs.

uncertainty does not impact model use for comparative analysis is to infer that statistical tests for significant differences at different exceedence probabilities will yield the same results. As can be seen from Table 3.15, the median estimators for runoff and percolation differences at the 50% EP and 10% EP are reasonably close. The median estimators for other output variables, however, are very different at the 50% and 10% EPs. Thus, while relative predictions of runoff and percolation from the model do not seem to be greatly impacted by parameter uncertainty, all other outputs considered are affected. If the sign test p-values are compared at the 50% and 10% EPs, it can be seen that significant differences are indicated for all outputs at the 0.05 level. It should be noted though that the larger p-values for the sign test at the 10% EP compared to values at the 50% EP show that the null hypothesis of no differences is not rejected as strongly when values at 10% EP are considered.

Monte Carlo simulation with correlated inputs

The results of the MCS for scenario 1 with uniformly distributed, correlated inputs is shown in Table 3.16. A comparison with results obtained with independent inputs reveals that there is some change in the location and dispersion of all outputs at both the 50% and 10% EPs, with changes somewhat greater at the 10% EP. It can be seen that the magnitude and direction of change is not consistent, and depends on the output variable being considered. The largest changes are evident in water and pesticide percolation at the 10% EP, with increases in both the median and IQR.

The results of the MCS for differences between scenarios with correlated inputs is shown in Table 3.17. Here again, a comparison against results with independent inputs reveals some change in location and dispersion for all outputs. The most notable change is in pesticide percolation at the 10% EP, which indicates differences are not significant at the 0.05 level.

Table 3.16 Results of Monte Carlo simulation with correlated inputs (scenario 1).

Output variable	50% exceedence probability		10% exceedence probability	
	Median	Inter-quartile range	Median	Inter-quartile range
Runoff (cm)	3.148	2.479	4.384	3.395
Percolation (cm)	2.584	5.982	4.425	8.096
Sediment loss (Mg/ha)	0.210	0.192	0.412	0.384
Surface pesticide losses (g/ha)	1.348	1.850	5.340	6.120
Pesticide percolation (g/ha)	0.0199	0.2470	1.210	6.447

Table 3.17 Results of Monte Carlo simulations for differences between scenarios 1 and 2 with correlated inputs.

Output variable	50% exceedence probability			10% exceedence probability		
	Median	Inter-quartile range	Sign test p-value	Median	Inter-quartile range	Sign test p-value
Runoff (cm)	3.126	1.515	< 0.001	3.371	1.871	< 0.001
Percolation (cm)	-0.668	2.734	< 0.001	-0.649	3.265	0.004
Sediment loss (Mg/ha)	0.553	0.2667	< 0.001	1.056	0.565	< 0.001
Surface pesticide losses (g/ha)	2.628	2.988	< 0.001	8.469	6.910	< 0.001
Pesticide percolation (g/ha)	-0.0073	0.1648	< 0.001	-0.1557	1.646	0.0576

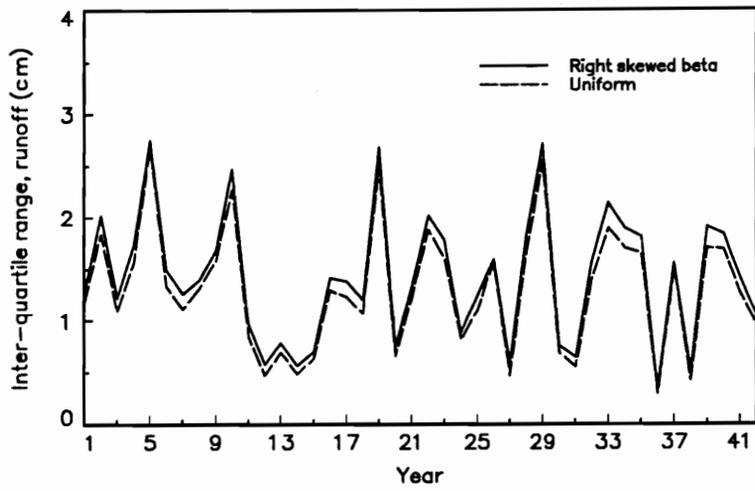
Sensitivity to input distribution shape

A MCS was carried out to assess model sensitivity to input distribution shape. Independent inputs were generated from beta distributions with a coefficient of skew equal to 0.64. Although the range used for the beta distributions was the same as the range for the uniform distributions employed in the first MCS, the results cannot be directly compared as the input medians are different. However, the coefficient of skew of the beta distributions was chosen so that the IQRs of the beta distributions were the same as the IQRs of uniform distributions used in the first MCS. Therefore, the model sensitivity to input distribution shape was assessed by comparing the IQRs of output distributions obtained with right skewed independent beta inputs and independent uniform inputs.

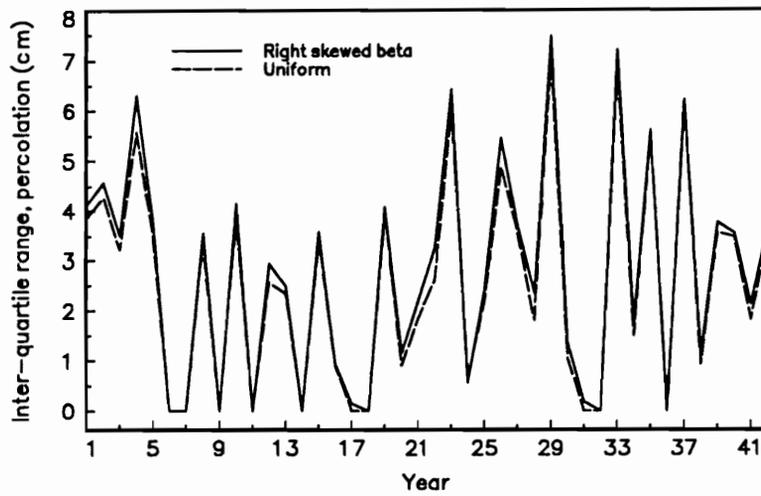
The variation in output IQRs over the simulation period with right-skewed beta inputs and uniform inputs for different outputs is shown in Figure 3.5. It can be seen that while the impact of input distribution shape on output IQRs is minor for runoff and percolation, the IQRs of other outputs are significantly affected. Surface pesticide losses and pesticide percolation show the greatest sensitivity to input distribution shape.

The impact of input distribution shape on comparative model predictions was assessed by comparing sign test p-values as here again it is not possible to compare median estimators and IQRs for differences directly. The sign test p-values obtained for differences between scenarios are shown in Table 3.18. The p-values for runoff, sediment loss, and surface pesticide losses show no change compared to p-values obtained using uniform inputs. However, p-values for water and pesticide percolation at the 10% EP indicate no significant differences between the scenarios. At the 50% EP, though significant differences are indicated for water percolation at the 0.05 level, differences in pesticide percolation are also not significant.

(a)



(b)



(c)

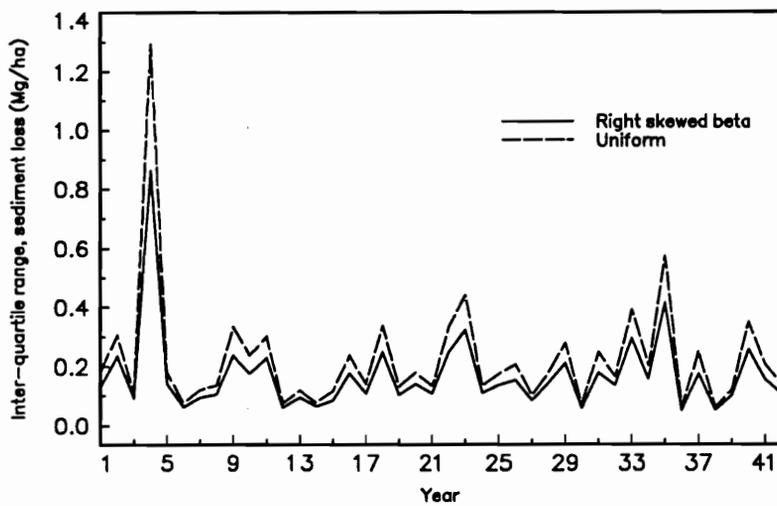


Figure 3.5 Impact of input distribution shape on output inter-quartile ranges over simulation period: (a) runoff (b) percolation (c) sediment loss.

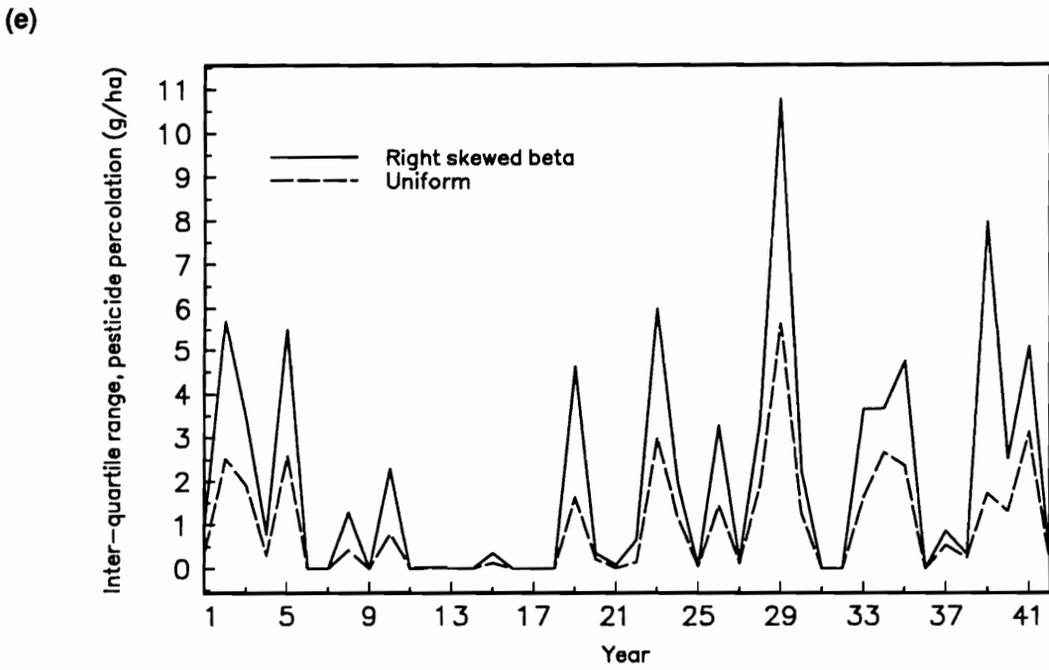
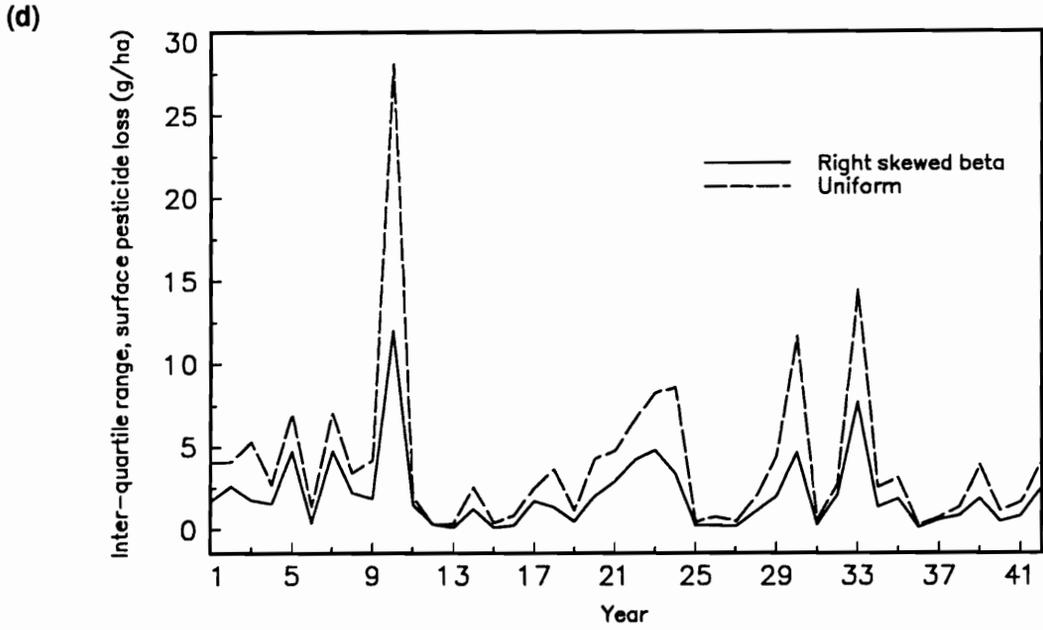


Figure 3.5 (continued) Impact of input distribution shape on output inter-quartile ranges over simulation period: (d) Surface pesticide loss (e) Pesticide percolation.

Table 3.18 Sign test p-values obtained for differences between scenarios 1 and 2 using right-skewed beta distributions for inputs.

Output variable	50% exceedence probability p-value	10% exceedence probability p-value
Runoff (cm)	< 0.001	< 0.001
Percolation (cm)	0.035	0.164
Sediment loss (Mg/ha)	< 0.001	< 0.001
Surface pesticide losses (g/ha)	< 0.001	< 0.001
Pesticide percolation (g/ha)	0.185	0.441

Summary and Conclusions

The simulation study showed that parameter uncertainty impacted both absolute and comparative predictions from GLEAMS for the two scenarios considered. When the effect of climatic variability on absolute predictions was compared at different EPs due to parameter uncertainty, it was found that the impact was greater at the 10% EP representing 'critical conditions' than at the 50% EP representing 'average conditions'. The magnitude of the increase in variability was found to depend on the extent to which output distributions changed over the simulation period. In comparative predictions, the impact of parameter uncertainty was significant for sediment loss, surface pesticide losses, and pesticide percolation, but not for runoff and percolation.

Correlations among inputs were found to produce changes in predicted medians and IQRs for all outputs, with somewhat larger changes evident for outputs at the 10% EP. The most significant impact of correlations for comparative predictions was on pesticide percolation. Predictions of surface pesticide losses and pesticide percolation showed the greatest sensitivity to input distribution shape. Comparative predictions of water and pesticide percolation were very strongly impacted by input distribution shape.

While it is difficult to draw broad conclusions from a simulation study considering only two scenarios, it is clear that parameter uncertainty can have important consequences not only for absolute predictions for a given scenario but also for comparative analysis between scenarios. For the two scenarios considered in the simulation study, the conclusion of increased water and pesticide percolation for scenario 1 compared to scenario 2 that would be made from a deterministic analysis is not justified when parameter uncertainty is considered. The MCS analysis indicates that when input variances and correlations are considered, there may be no significant differences in water and pesticide percolation between the two scenarios, particularly

under 'critical conditions'. While an analysis as to what was causing this response in terms of the model structure was not pursued, a possible intuitive explanation is the interaction between the evapotranspiration and percolation components of the model. With input parameter combinations that would result in wetter profiles for both scenarios, the model response for scenario 1 may result in increased evapotranspiration to such an extent that percolation differences between the scenarios are no longer significant.

Chapter IV

Probability distributions for hydrologic inputs

...the statistical method..., which in the present state of our knowledge is the only available method of studying the properties of real bodies, involves an abandonment of strict dynamical principles, and an adoption of the mathematical methods belonging to the theory of probability. It is probable that important results will be obtained by the application of this method which is as yet little known and is not familiar to our minds.

James Clerk Maxwell

INTRODUCTION

In the simulation study in the previous chapter we noted that GLEAMS is sensitive to the shape of input distributions for both absolute and comparative predictions. Hence it is important to be able to define the shape of input probability distributions. Characterizing probability distributions by direct fitting to measured values is very rarely possible because of cost and time constraints; indirect methods which achieve an acceptable degree of accuracy in defining distributions are more pragmatic. The sediment and chemical transport components of NPS pollution models are driven by the hydrologic component, and therefore, defining probability distributions for hydrologic inputs is necessarily the first step. In the current study, attention is focused on defining probability distributions for critical hydrologic inputs for GLEAMS, with distributions for other inputs assumed to be adequately described by heuristics described in the next chapter. The hydrologic inputs that are considered here are parameters which partition rainfall into runoff and infiltration, and soil hydraulic properties that define sub-surface unsaturated flow. The following description of the hydrologic component of GLEAMS is derived from Smith and Ferreira (1989).

Rainfall is partitioned into runoff and infiltration by a modification of the USDA Soil Conservation Service (SCS) runoff curve number method (Smith and Williams, 1980). The unsaturated soil

profile in the root zone is assumed to be composed of a set of linked linear storage elements, each having a designated storage capacity, and water content limit (field capacity or water held at a matric tension of 0.1 to 0.33 bar). Only water in excess of this limit is considered mobile. The estimated rainfall input water fills the storages starting at the top until the total storage depth is exhausted. The rate of movement of mobile water between storage elements depends on the saturated conductivity of the elements. No water moves when saturation is less than field capacity. In this case, water leaves a storage only by evapotranspiration. The water content of a storage element is not allowed to fall below a lower limit (wilting point or water held at a matric tension of 15 bar). The sub-surface profile can be characterized by up to five horizons, with each horizon described by its saturated hydraulic conductivity and water content at wilting point, field capacity, and full saturation (porosity).

In this chapter procedures for defining probability distributions for runoff curve number, saturated conductivity, field capacity, and wilting point under conditions of limited data availability are developed. Procedures for obtaining the complete runoff curve number distributions for all moisture conditions based on previous work by Hjelmfelt et al. (1982) and Mills et al. (1992) are outlined. The complete CN distribution is employed in a conceptual approach to determining the CN distribution for average moisture conditions, suitable for use in continuous simulation models. Procedures for obtaining probability distributions for saturated conductivity, field capacity, and wilting point, using multiple linear regression equations similar to those proposed by Rawls and Brakensiek (1985) are developed and evaluated. A new set of regression equations which are robust to prediction outside the construction sample data are proposed. The utility of using estimates of field-scale variability in soil texture, porosity, and organic matter content to approximate distributions of soil hydraulic properties employing the new regression equations is assessed.

RUNOFF CURVE NUMBER DISTRIBUTIONS

The SCS runoff equation provides predictions of runoff volume for a given rainfall amount, using a potential retention parameter (S) related to soil and cover conditions through the runoff curve number (CN) as follows

$$CN = \frac{1000K}{S + 10K} \quad (4.1)$$

where S is potential retention, mm (in), CN is a dimensionless curve number, and K is a unit conversion constant equal to 25.4(1). The CN may exhibit considerable variation from storm to storm due to several factors including storm duration, storm intensity, and antecedent moisture. Rawls et al. (1993) state that to account for this variation, CNs for a given soil-cover complex may be specified at different antecedent runoff conditions (ARCs). Runoff curve numbers are generally specified at three ARCs, referred to as ARC I, ARC II, and ARC III, representing increasing runoff potential. Values of S (or CN) corresponding to ARC I, ARC II, and ARC III will be denoted hereafter by S_I (CN_I), S_{II} (CN_{II}), and S_{III} (CN_{III}) respectively. Potential retention (or CN) at different ARCs may be approximated from the following relationship (Hawkins, 1985)

$$\frac{S_I}{S_{II}} \cong \frac{S_{II}}{S_{III}} \cong 2.3 \quad (4.2)$$

The use of the term ARC is relatively recent, and it was previously referred to as antecedent moisture condition (AMC). While apparently intended as a generic term (Hjelmfelt, 1994) reflecting CN variability from all sources, it has commonly been assumed (Ponce, 1988) that CNs specified at different AMCs reflected only the impact of soil moisture on runoff. Thus, CNs specified at different AMCs were taken to correspond to dry, average, and wet conditions.

Potential retention and runoff curve number distributions

Hjelmfelt et al. (1982) found that if the potential retention parameter (S) employed in the runoff equation is assumed to follow a lognormal distribution, values of S at 10%, 50% and 90% exceedence probabilities lead to CN values comparable to CN_I , CN_{II} , and CN_{III} respectively. The CN_{II} value was interpreted as a median corresponding approximately to the SCS tabulated value for the soil-cover complex under consideration. In their analysis, Hjelmfelt et al. (1982) used values of S determined from annual maximum events. The results of studies by Hauser and Jones (1991) and Yoo et al. (1993) indicate that the procedure may be employed to estimate median curve number values from all observed rainfall-runoff pairs.

With S assumed to follow a specified distribution, an expression for the probability density function (pdf) of CN may be derived using standard techniques for deriving probability laws of scalar functions of random variables (Haan, 1977). However, a closed form expression for the cumulative distribution function (cdf) of CN cannot be obtained when a lognormal distribution is assumed for S . Thus, to generate random variates from the CN probability distribution, it would be necessary to numerically integrate the CN pdf. An alternative approach is to employ direct Monte Carlo simulation of equation (4.1) to obtain the pdf and cdf of CN. To do this, given estimates of the scale ($\hat{\mu}$) and shape ($\hat{\sigma}$) lognormal parameters of S , random variates Y from a normal distribution $N(\hat{\mu}, \hat{\sigma})$ are generated and $\exp(Y)$ values, which will follow a lognormal distribution $LN(\hat{\mu}, \hat{\sigma})$ are substituted into equation (4.1) for S . The CN values obtained can be ranked to generate the cdf. As routines for generating normal variates are widely available, this procedure can be easily implemented.

To obtain lognormal parameter estimates for S from observed rainfall-runoff data, values of S are first obtained by solving the SCS runoff equation for S (Haan and Schulze, 1987)

$$S = 5R + 10Q - 10\sqrt{Q^2 + 1.25RQ} \quad (4.3)$$

where Q is the runoff volume, mm (in) from a rainfall volume R, mm (in). Mills et al. (1992) have noted that since values of S can only be computed from equation (4.3) when runoff occurs, distribution parameters estimated from these values will be biased towards lower S values. When estimating distribution parameters for S from short term rainfall-runoff records, it is important to insure that the parameters are unbiased. Mills et al. (1992) show that with a lognormal distribution for S and a Weibull distribution for R, with S and R assumed to be independent, theoretical considerations of the distribution of S conditioned on R and the joint distribution of S and R will result in the following expression

$$[1 - P(S_m \leq x | Q_{occ.})]^{(j/k)} = \int_x^\infty \left\{ [1 / s\sigma\sqrt{2\pi}] \exp\{-0.5[(\ln s - \mu) / \sigma]^2\} \exp[-(0.2s)^\alpha \beta^{-\alpha}] \right\} ds \quad (4.4)$$

where $P(S_m \leq x | Q_{occ.})$ is the empirical distribution function (EDF) of measurements of S conditioned on runoff, j is the number of runoff events, k is the number of rainfall events, μ and σ are unbiased scale and shape lognormal parameters of S, and α and β are the shape and scale Weibull parameters of R. The EDF of S conditioned on rainfall, j, and k may be obtained from observed rainfall-runoff data, and a Weibull distribution fitted to the rainfall data. Equation (4.4) can then be solved for μ and σ .

If no rainfall-runoff data are available, the CN_{II} value obtained from SCS tabulated values may be used to approximate $\hat{\mu}$ and $\hat{\sigma}$, with suitable exceedence probabilities, p% and (1-p)% assumed for the corresponding CN_I and CN_{III} values. The lognormal parameters are conveniently computed by working in standard normal space. We first note that if S follows a lognormal distribution, $(\ln S - \hat{\mu}) / \hat{\sigma}$ will follow the standard normal distribution $N(0,1)$. Points in standard normal space corresponding to S_I and S_{III} can thus be obtained from

$$C_i = \left(\frac{\ln S_i - \hat{\mu}}{\hat{\sigma}} \right) \quad (4.5)$$

where $i=I$ or III , C_I and C_{III} are the $(1-p)\%$ and $p\%$ exceedence values in standard normal space. Values of S_I and S_{III} are computed from the CN_I and CN_{III} values, substituted in equation (4.5), with C_I and C_{III} set equal to $\Phi^{-1}(p/100)$ and $\Phi^{-1}(1-p/100)$ respectively, where Φ^{-1} is the inverse of the standard normal cdf, and the resulting equations solved for $\hat{\mu}$ and $\hat{\sigma}$. As noted before, Hjelmfelt et al. (1982) suggested that using $p = 10\%$ may be a reasonable approximation.

Distributions resulting from factors other than antecedent moisture variation

It appears that the current consensus on S (or CN) variation is to lump all sources of variability (Hjelmfelt, 1994), and not attempt to explain how much of the variation can be attributed to individual factors such as antecedent moisture conditions. The S distributions obtained as described in the previous section would reflect variability from all sources. However, continuous hydrologic models that use the SCS runoff equation (e.g. GLEAMS, EPIC, Opus) generally adjust the value of CN based on current soil moisture conditions using modifications of a procedure originally employed by Williams and LaSeur (1976). Thus, these models directly account for variability due to antecedent moisture and the S distribution reflecting variability from all sources would not be appropriate. The S distribution for these models should ideally reflect variability only from sources other than antecedent moisture conditions.

The S distribution resulting from factors other than antecedent moisture may be derived from the S distribution obtained when all sources of variation are considered, under the assumptions that there is no interaction between antecedent moisture and other sources of variation and that the

median of both distributions are the same. To avoid ambiguity we will refer to the S distribution resulting from all sources of variation as the complete S distribution and denote the S distribution resulting from factors other than antecedent moisture by S^* .

We assume the pdf of the complete S distribution, $f(x)$ is defined, and denote the pdf of the S^* distribution by $f(y)$. Further, let m be the fraction of total variation in S explained by antecedent moisture conditions. This fraction defines an area under the $f(x)$ density less than one which includes the value $x = S_{||}$. If the points on the x-axis bounding the area less than and greater than $S_{||}$ are denoted by S_0 and S_1 respectively, and the fraction of m composed of S values greater than $S_{||}$ is n , then

$$S_0 = F^{-1}((1-m)(1-n)) \quad (4.6a)$$

$$S_1 = F^{-1}((1-m)n) \quad (4.6b)$$

where F^{-1} is the inverse cdf of the complete S distribution. It should be noted here that S_0 and S_1 do not necessarily equal $S_{||}$ and S_1 . The areas under the $f(x)$ density between $(0, S_0]$ and (S_1, ∞) can be regarded as representing the variation in S not explained by antecedent moisture conditions. The regions bounded under $f(x)$ between $(0, S_0]$ and (S_1, ∞) represent truncated pdfs which can be rescaled to have maximum and minimum values of $S_{||}$ by multiplying with appropriate factors. The combination of these pdfs is the pdf of S^* , which will then have the required median of $S_{||}$. From the definition of truncated pdfs and pdfs of linear functions of random variables, we can write an expression for the pdf of S^* as follows

$$f(y) = (1-q) \frac{S_0}{S_{||}} \frac{f(x)}{F(S_0)} I_{(0, S_0]}(x) + q \frac{S_1}{S_{||}} \frac{f(x)}{1-F(S_1)} I_{(S_1, \infty)}(x) \quad (4.7a)$$

where F is the cdf of the complete S distribution, q is the probability of obtaining S values greater than S_{II} on the S^* distribution and $I_A(x)$ denotes the indicator function of the set A defined by (Law and Kelton, 1991)

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases} \quad (4.7b)$$

Once $f(x)$ is defined, we need to know m , n and q in order to define $f(y)$. A natural assumption is that $q = n$, and if m is taken to be equal to the area bounded by $f(x)$, S_I and S_{III} , the symmetry imposed by equation (4.2) requires that $n = 0.5$. We note that in the limiting case, when none of the total variation in S can be explained by antecedent moisture conditions only, $m = 0.0$, and $S_0 = S_I = S_{III}$, in which case $f(y) = f(x)$ i.e. the S^* distribution will be the same as the complete S distribution. If observed rainfall-runoff data are available, S_I and S_{III} can be located on the complete S distribution to determine m . Here again, if no rainfall-runoff data are available, suitable exceedence probabilities have to be assumed for S_I and S_{III} .

Random variates from the S^* distribution can be generated employing the composition method (Law and Kelton, 1991). The procedure can be summarized as follows

1. Generate two standard uniform variates U_1 and U_2 .
2. If $U_1 \leq (1-q)$, use U_2 to obtain a random variate from the complete S distribution in the interval $(0, S_0]$, and rescale this variate by multiplying by S_{II}/S_0 .
3. If $U_1 > q$, use U_2 to obtain a random variate from the complete S distribution in the interval (S_1, ∞) , and rescale this variate by multiplying by S_{II}/S_1 .

Figure 4.1 illustrates the procedure where points 1 and 2 in the interval $(0, S_0]$ map onto 1^* and 2^* , while points 3 and 4 in the interval (S_1, ∞) map onto 3^* and 4^* . The procedure for obtaining the complete S distribution described previously, and the procedure for obtaining the S^*

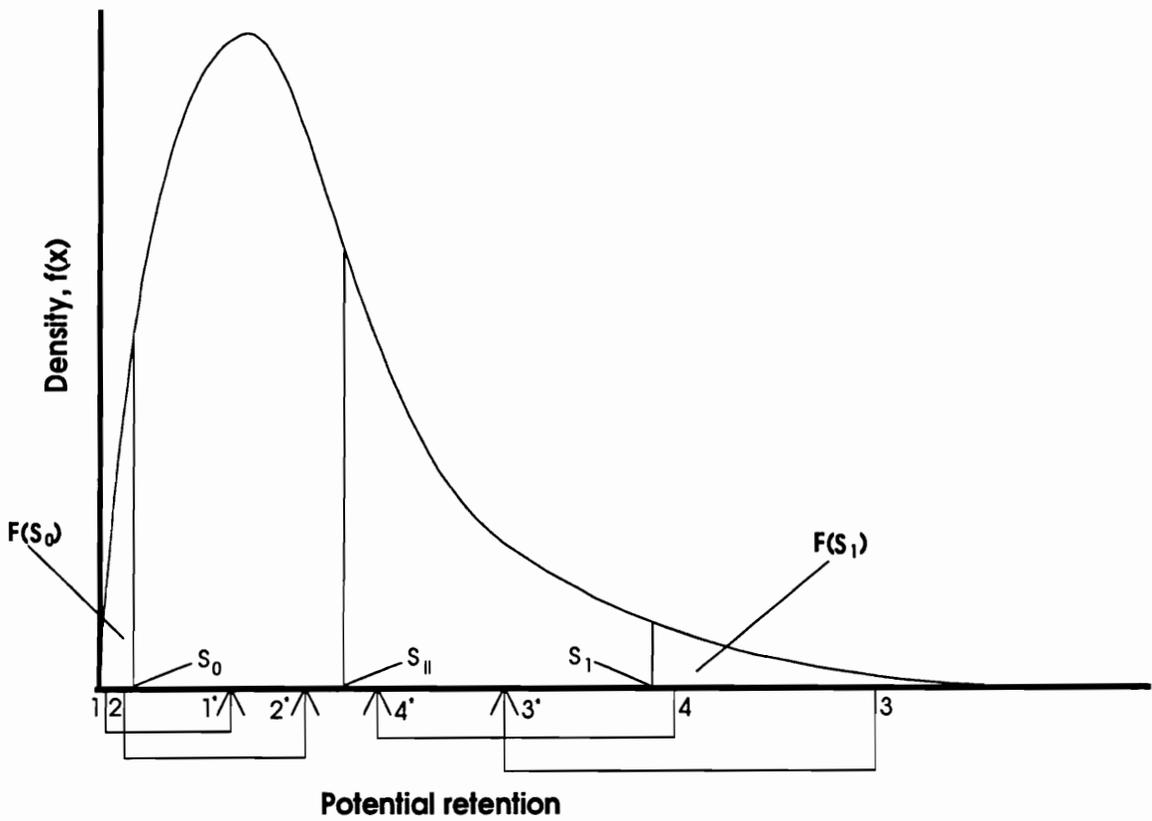


Figure 4.1 Conceptual procedure for obtaining the S^* distribution from the complete S distribution.

distribution from the complete S distribution described in this section were incorporated into a computer routine, listed in Appendix C. The routine allows generation of variates from the complete CN (analogous to complete S) and CN* (analogous to S*) distributions using inputs for either the lognormal parameter estimates for the complete S distribution or the exceedence probability for CN_{III}. Variates from the CN* distribution are generated assuming $q = n = 0.5$, and the input value for m.

Application to conventional and conservation tillage systems

Complete CN and CN* distributions were obtained under conservation and conventional tillage systems using rainfall-runoff data from small paired watershed plots for a single crop (corn) season (May-September, 1990). The soil type is a Suffolk sandy loam characterized as deep and well-drained belonging to Hydrologic Group B. A total of 43 rainfall events were observed over the season (storms producing total rainfall less than 0.5 mm were not considered), 18 of which produced runoff from the conventional tillage plot, and 10 of which produced runoff from the conservation tillage plot (Heatwole et al., 1992). EDFs for S conditioned on runoff were obtained from equation (4.3) and are shown in Figure 4.2.

A Weibull distribution was fitted to the rainfall data using VTFIT (Cooke et al., 1993), which estimates parameters by numerical solution of equations for the Weibull maximum likelihood estimators given by Law and Kelton (1991). Unbiased lognormal parameter estimates for S were obtained by solving equation (4.4) for μ and σ using IMSL (1987) routine DUNLSF which solves a nonlinear least squares problem using a modified Lavenberg-Marquardt algorithm and a finite-difference Jacobian. The IMSL routine DQDAGI, which allows integration over a semi-infinite interval, was used to provide function evaluations for routine DUNLSF. The Weibull parameters for rainfall and the unbiased lognormal parameter estimates of S for conventional and

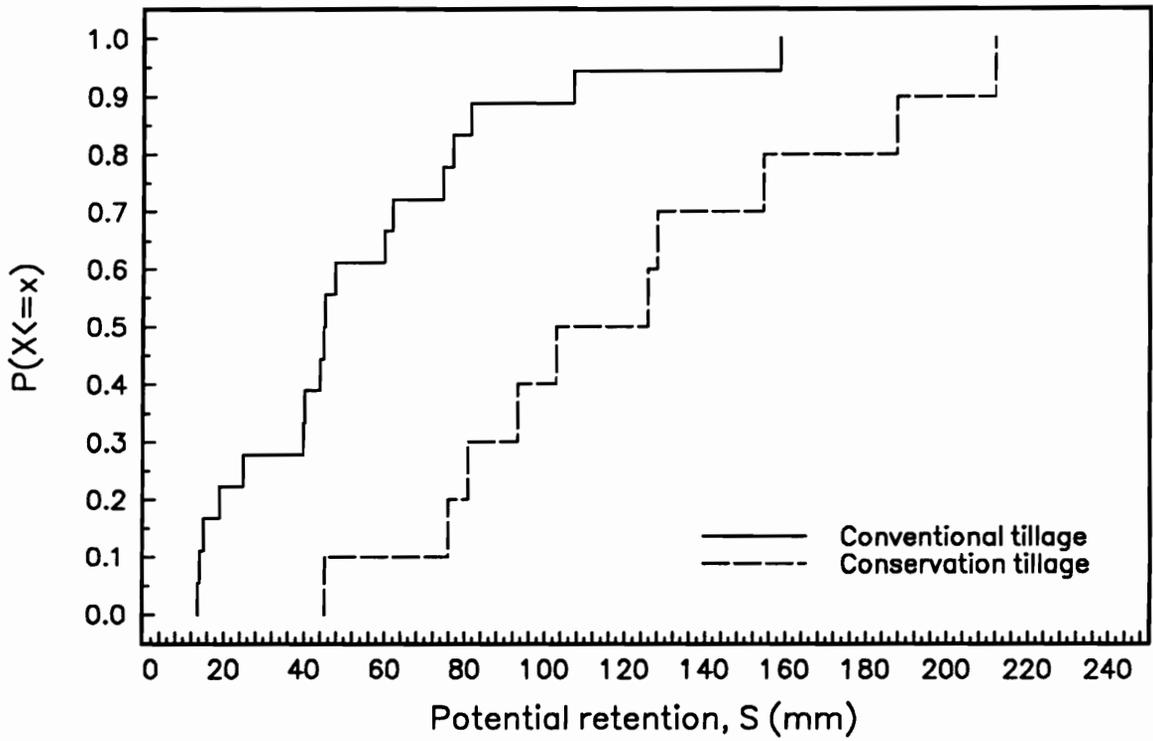


Figure 4.2 Empirical distribution functions for potential retention conditioned on runoff.

conservation tillage plots are shown in Table 4.1. The median of the complete S distribution ($S_{||}$), the corresponding $S_{||}$ values and exceedence probabilities are given in Table 4.2. The fraction of total variation in S explained by antecedent moisture conditions (m), calculated from $(1 - 2\Phi((\ln S_{||} - \mu)/\sigma))$, where Φ is the standard normal cdf, is also given in Table 4.2. The unbiased lognormal parameter estimates of S and computed values of m were used to generate 2000 variates from the complete CN and CN* distributions for both tillage systems. Histograms representing the complete CN and CN* pdfs are shown in Figures 4.3 through 4.6, and a plot of ranked values representing the complete CN and CN* cdfs are shown in Figures 4.7 and 4.8. A summary of central tendency and dispersion measures for the CN and CN* distributions for both tillage systems is given in Table 4.3.

Results and Discussion

The EDFs of S conditioned on runoff (Figure 4.2) suggest both location and dispersion differences between the two tillage systems, with an apparently higher variability in S indicated under conservation tillage. However, the unbiased lognormal parameter estimates for S (Table 4.1) show that the variance in S is actually smaller for the conservation tillage plot. The complete CN pdfs (Figures 4.3 and 4.4) are left-skewed for both tillage systems, with the skewness more marked for the conventional tillage system. The difference in complete CN distributions is reflected in the CN* pdfs, with the distribution for conservation tillage practically symmetrical, while the distribution for conventional tillage remains left-skewed (Figures 4.5 and 4.6).

For the rainfall-runoff data analyzed, as indicated by the unbiased parameter estimates for S, CNs under conventional tillage were more variable than CNs under conservation tillage (Table 4.3). The differences in distributions is also clear from the cdf plots (Figures 4.7 and 4.8) The complete CN distribution represents variability from all sources while the CN* distribution may be regarded as representing variability in CN values due to factors other than antecedent moisture

Table 4.1 Parameter estimates for rainfall and potential retention (S) distributions.

Variable	Distribution name	Scale parameter	Shape parameter
Rainfall (mm)	Weibull	14.799	0.804
Retention-CvT (mm)	Lognormal	4.216	0.937
Retention-CsT (mm)	Lognormal	4.892	0.524

CvT - Conventional tillage
CsT - Conservation tillage

Table 4.2 Median of complete S distributions (S_{II}), corresponding S_{III} values with exceedence probabilities (p_e), and fraction of S variation explained by antecedent moisture (m).

Tillage system	S_{II} (mm)	S_{III} (mm)	p_e (%)	m
Conventional	67.77	29.46	81.3	0.626
Conservation	133.22	57.92	94.4	0.888

Table 4.3 Summary of central tendency and dispersion measures for complete CN and CN* distributions.

Distribution	Mean	Median	Standard deviation	Inter-quartile range
Complete CN-CvT	75.65	78.80	15.33	20.78
CN*-CvT	76.86	79.02	11.42	13.98
Complete CN-CsT	64.82	65.49	11.17	15.76
CN*-CsT	65.20	65.64	6.63	7.84

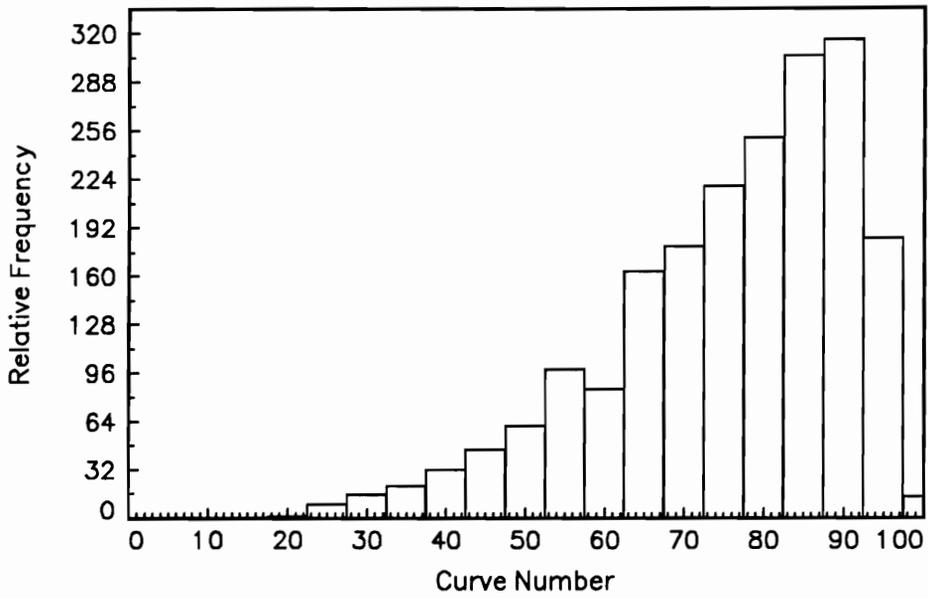


Figure 4.3 Histogram of generated values for complete CN distribution (conventional tillage).

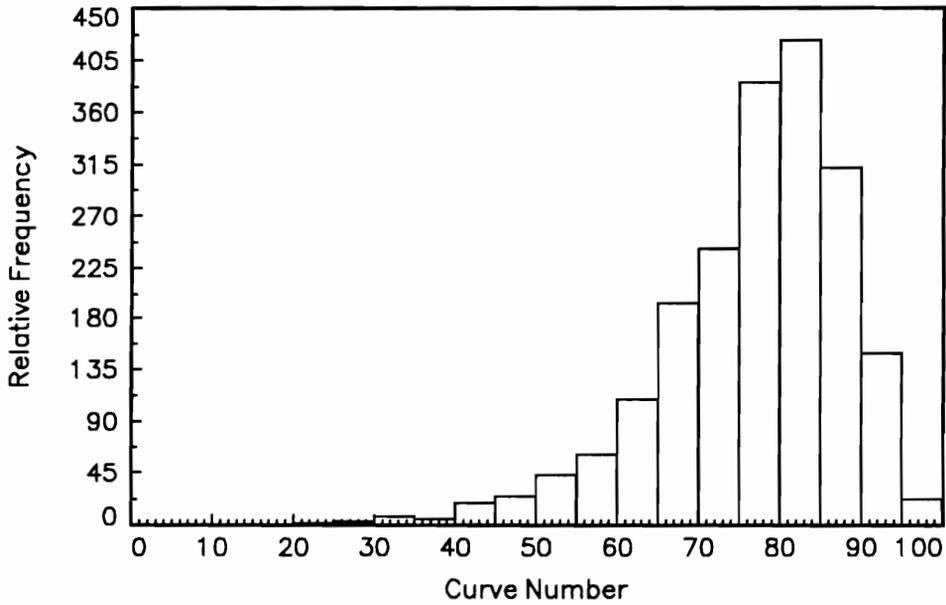


Figure 4.4 Histogram of generated values for CN* distribution (conventional tillage).

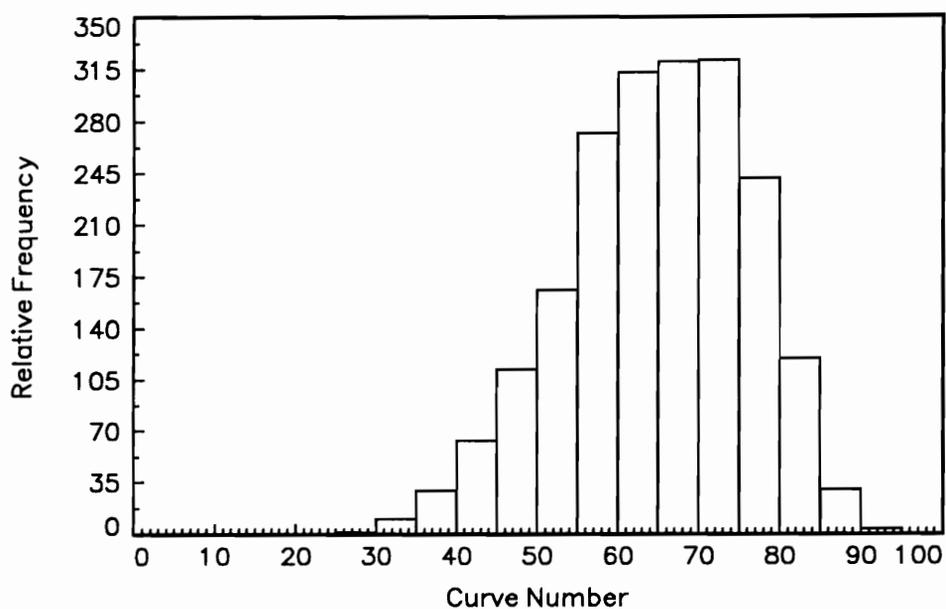


Figure 4.5 Histogram of generated values for complete CN distribution (conservation tillage).

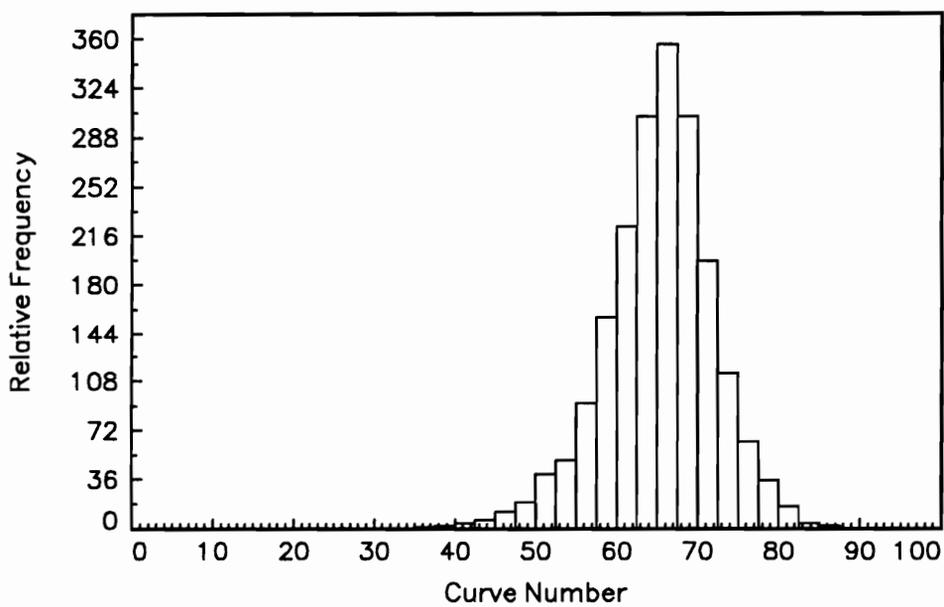


Figure 4.6 Histogram of generated values for CN* distribution (conservation tillage).

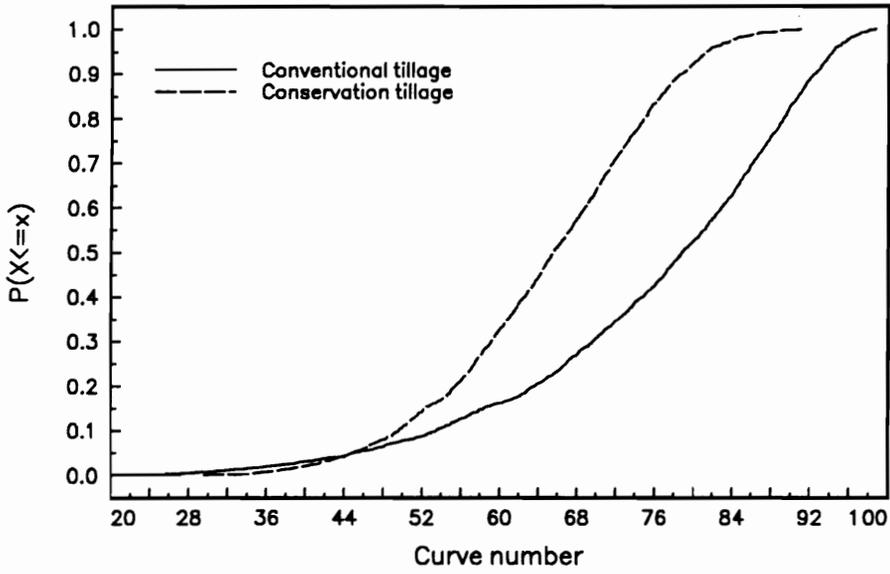


Figure 4.7 Plots of ranked values for complete CN.

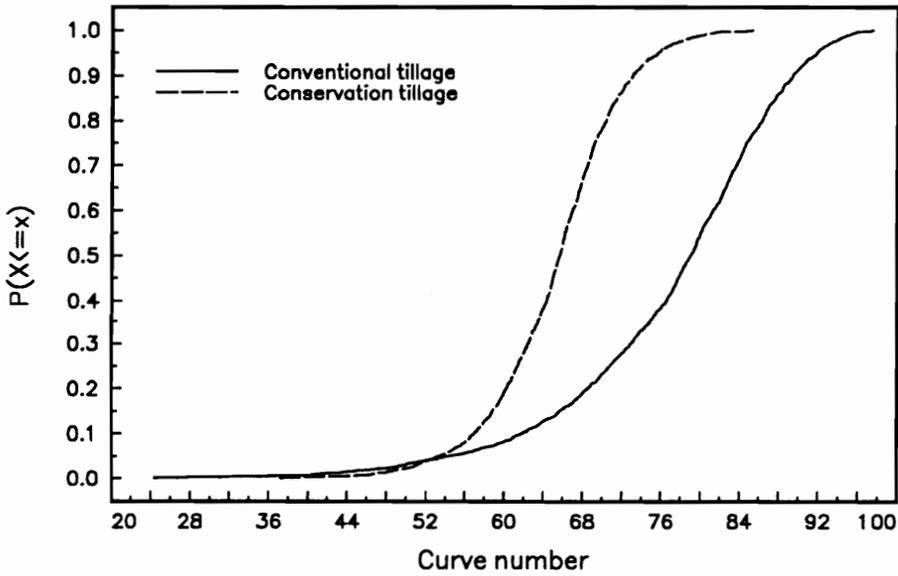


Figure 4.8 Plots of ranked values for CN^* .

conditions. This may be a simplistic interpretation as it assumes there is no interaction between other sources of variation e.g. rainfall intensity and antecedent moisture condition. However, there is very little information available on the nature of this interaction. The conceptual procedure developed here for obtaining the CN* distribution from the complete CN distribution provides a rational means of determining input distributions for CN in Monte Carlo studies using continuous hydrologic models that employ the SCS runoff equation with accounting of soil moisture.

DISTRIBUTIONS OF SOIL HYDRAULIC PROPERTIES

Rawls et al. (1982) and Rawls and Brakensiek (1985) have reported ordinary least squares multiple linear regression equations derived from an extensive soils database for predicting soil hydraulic properties from soil texture, porosity (or bulk density), and organic matter content. Although the equations are widely used to obtain point predictions, their utility in providing acceptable predictions of field-scale variation in soil hydraulic properties has not been rigorously examined. Loague (1992) reported a reduction in the performance of a quasi-physically based rainfall-runoff model when saturated hydraulic conductivity estimates were obtained from soil texture data rather than infiltration data. However, Loague (1992) did not use the actual textural data to estimate hydraulic conductivity but rather used the average values of saturated conductivity reported by Rawls et al. (1982) for the two soil textural classes at the catchment under study. Bonsu (1992) found that the equations of Rawls et al. (1982) provided reasonable estimates of saturated conductivity for coarse-textured soils but considerably underpredicted values for fine-textured soils. Ragab and Cooper (1993) compared estimates of saturated conductivity and parameters of the Brooks-Corey capillary-saturation model (Brooks and Corey, 1964) obtained from the Rawls and Brakensiek (1985) regression equations with *in situ*

measurements made using a Guelph permeameter. They suggest that predictions obtained from the regression equations are encouraging and worth further investigation.

A limitation of the Rawls and Brakensiek (1989) regression equations is that they are reported as point prediction functions and do not provide any indication of the variance associated with predictions. Beven (1989) and Beven and Binley (1992) have noted that there may be considerable uncertainty associated with estimates of soil hydraulic properties obtained from the regression equations. Myers (1990) has emphasized that quality fit (high coefficient of determination) and quality prediction do not necessarily coincide in a regression equation. To properly evaluate the usefulness of any regression equation and allow its judicious use for prediction, it is desirable to include the variance associated with any point prediction.

An extensive database of soil physical properties and measurements of moisture retention and saturated conductivity, compiled using information from various sources in the continental United States was obtained (Rawls, 1980). A least squares multiple linear regression analysis was carried out using saturated conductivity (K_s), and moisture contents at matric tensions of 0.1 bar ($\theta_{0.1}$), 0.33 bar ($\theta_{0.33}$), and 15 bar (θ_{15}) as response variables, and sand, clay, organic matter content, and porosity as potential predictors. The regression equations were evaluated to assess their utility in providing acceptable predictions of field-scale variability in soil hydraulic properties of interest.

Regression Analysis

In any regression analysis it is important to assess the validity of assumptions inherent in the analysis. The major assumptions in least-squares regression are that errors are normal and independent with mean zero and common variance. In addition, in multiple linear regression, it is important to assess and determine the impact of multicollinearity in the regressor variables.

developed. One set utilizes sand, clay, porosity, and organic matter as predictors. The second set may be used in cases where organic matter data is not available.

The following procedure was employed in the analysis:

1. The adjusted R^2 (coefficient of determination adjusted for degrees of freedom) of the full regression model (using all potential predictors) was obtained to provide a basis for determining the acceptability of models using predictor sub-sets with reduced multicollinearity.
2. Multicollinearity was assessed by computing the variance inflation factor (VIF) for regression coefficients. The VIF for the i^{th} regression coefficient can be written as (Myers, 1990)

$$VIF_i = \frac{1}{1 - R_i^2} \quad (4.10)$$

where R_i^2 is the coefficient of multiple determination obtained by regressing the i^{th} variable against the other regressor variables. Regressions were run using centered predictors and a sub-set of predictors selected based on the adjusted R^2 value and VIFs for regression coefficients. VIFs greater than 10 were taken to be indicative of multicollinearity (Myers, 1990).

3. Plots of standardized residuals against predicted response were used to assess possible violations of the assumption that the residuals have zero mean and common variance. These plots should exhibit no systematic trend and ideally would show a random scatter around zero. The standardized residual of the i^{th} observation, r_i can be written as (Myers, 1990)

$$r_i = \frac{y_i - \hat{y}_i}{\hat{\sigma} \sqrt{1 - \mathbf{x}_i' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_i}} \quad (4.11)$$

where y_i is i th observed response, \hat{y}_i is the i th predicted response, $\hat{\sigma}$ is the square root of the estimated root mean square error, and \mathbf{x}_i is the vector of predictors at the i th data point.

4. Normal probability (rankit) plots of standardized residuals against their expected value were used to examine possible departures from normality. These plots should ideally be a straight line through the origin with unit slope. The expected value of the i th order statistic from a standard normal sample size of n , $E(z_i)$ can be approximated from (Weisberg, 1980)

$$E(z_i) = \Phi^{-1} \left(\frac{i - 3/8}{n + 1/4} \right) \quad (4.12)$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution function.

5. The Durbin-Watson (D-W) statistic (Durbin and Watson, 1971) was used as an indicator for possible autocorrelation in the response variable. The (D-W) statistic, d , for n equally spaced data points may be written as (Sen and Srivastava, 1990)

$$d = \frac{\sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^n \varepsilon_i^2} \quad (4.13)$$

where ε_i is the i th ordinary residual, given by $(y_i - \hat{y}_i)$.

Myers (1990) notes that in many fields of application of regression analysis, failure of the homogenous variance assumption is very natural and expected. In such cases, variance stabilizing transformations on the response variable are often employed. Following Rawls and Brakensiek (1989) the natural logarithm transformation was applied to the saturated conductivity

data before performing any regressions. For all other response variables, regressions were initially carried out on the original data.

Table 4.4 shows the adjusted R^2 values obtained using the full models. From equation 4.9, the number of predictors for full third-order model using four and three first-order predictors is 34 and 19 respectively. The adjusted R^2 value for saturated conductivity with all predictors indicates using organic matter as a regressor does not improve prediction. A test for correlation between the natural logarithm of saturated conductivity and organic matter failed to reject the null hypothesis of no correlation ($p > 0.1$), and therefore, it was decided not to use organic matter as a predictor for saturated conductivity. Table 4.5 shows the number of predictors and adjusted R^2 for models selected using sub-sets of predictors in which all regression coefficients had VIFs less than 10.

The residual and rankit plots for K_s (Figure 4.9) show that there may be some departure from the assumption that the errors are normal and independent with common variance. However, these departures do not appear to be severe, and the fitted model was deemed to be acceptable. The residual plots for $\theta_{0.1}$, $\theta_{0.33}$ and θ_{15} using all predictors (Figures 4.10a through 4.12a) show a right-opening megaphone pattern which indicates that the error variance is a systematic function of the predicted value (Weisberg, 1980). The rankit plot for $\theta_{0.1}$ (Figure 4.10b) exhibits a flattening at both ends which indicates heavy-tailed errors suggesting too many extreme values for a normal sample. The rankit plots for $\theta_{0.33}$ and θ_{15} (Figures 4.11b and 4.12b) suggest light tailed errors which indicate fewer observations than would be expected for a normal sample.

Commonly used techniques to mitigate the impact of violating assumptions in least-squares regression include outlier analysis and the use of transformations on the response variable. The high value of standardized residuals of some of the data points suggests that an outlier analysis would be useful. However, an outlier analysis was not carried out due to the lack of information

Table 4.4 Number of observations and adjusted R² obtained with full models.

Response	Predictor set 1		Predictor set 2	
	Observations	Adjusted R ²	Observations	Adjusted R ²
$\theta_{0.1}$	1620	0.792	2274	0.770
$\theta_{0.33}$	1668	0.845	2591	0.805
θ_{15}	1986	0.862	3161	0.801
$\ln K_s$	1169	0.619	1379	0.639

Predictor set 1 refers to sand, clay, porosity and organic matter.
 Predictor set 2 does not include organic matter.

Table 4.5 Number of predictors and adjusted R² with selected model sub-sets.

Response	Predictor set 1		Predictor set 2	
	Predictors	Adjusted R ²	Predictors	Adjusted R ²
$\theta_{0.1}$	20	0.775	13	0.748
$\theta_{0.33}$	20	0.829	12	0.786
θ_{15}	19	0.855	14	0.796
$\ln K_s$	-	-	12	0.619

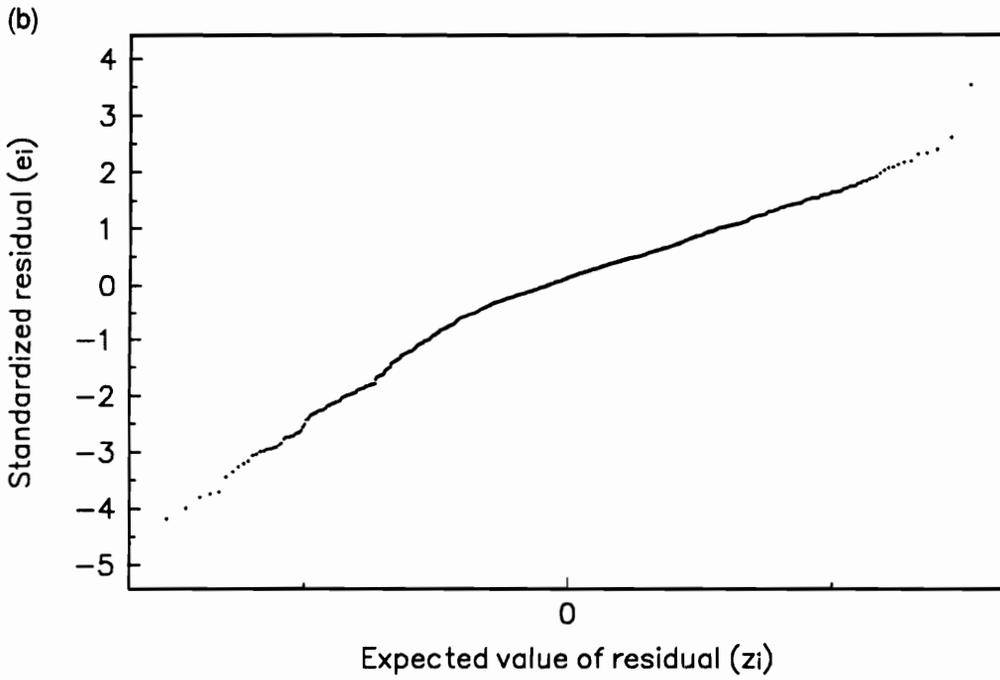
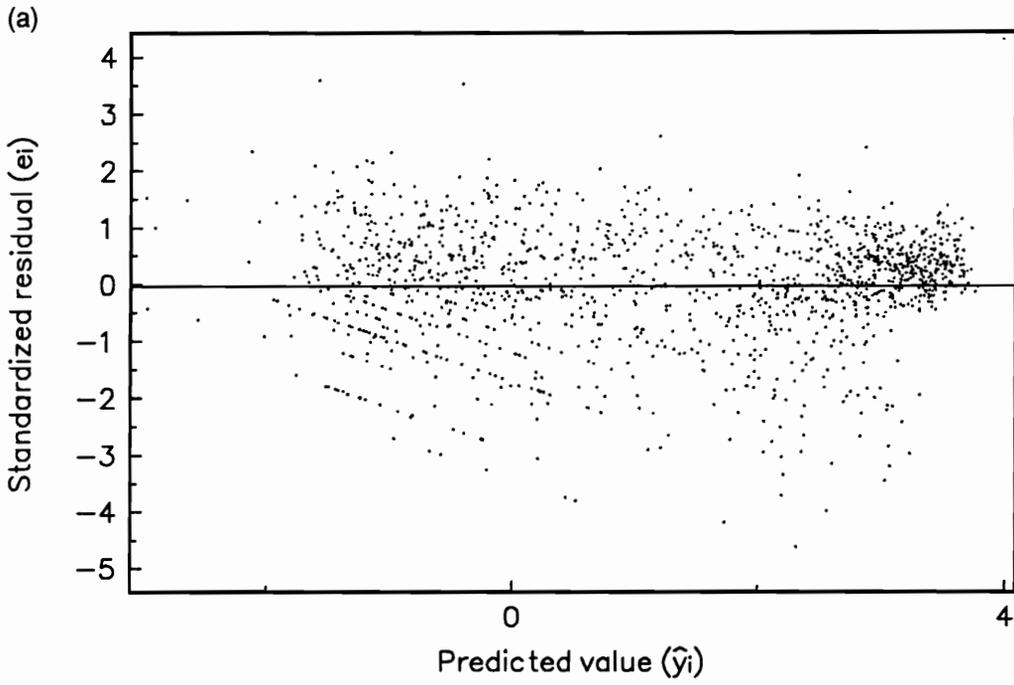
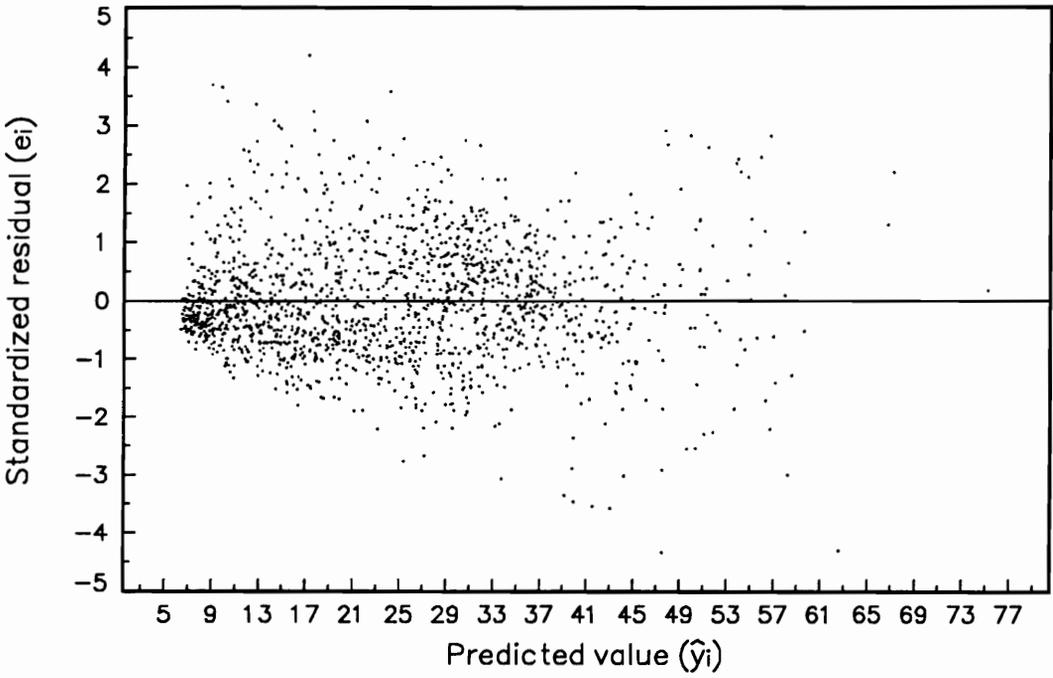


Figure 4.9 Diagnostic plots for $\ln K_s$ regression: (a) residual plot (b) rankit plot.

(a)



(b)

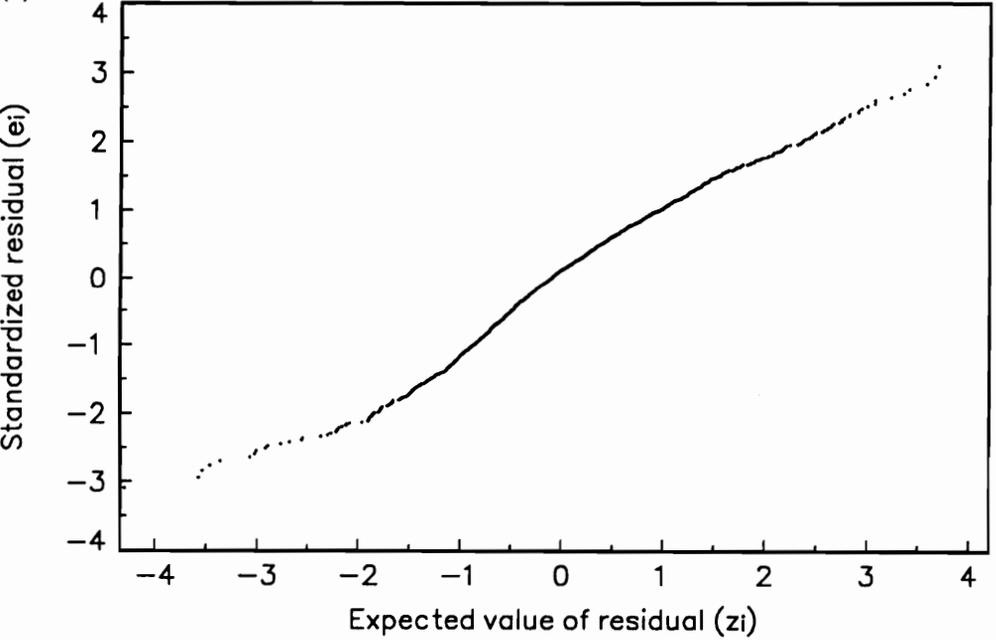


Figure 4.10 Diagnostic plots for $\theta_{0,1}$ regression: (a) residual plot (b) rankit plot.

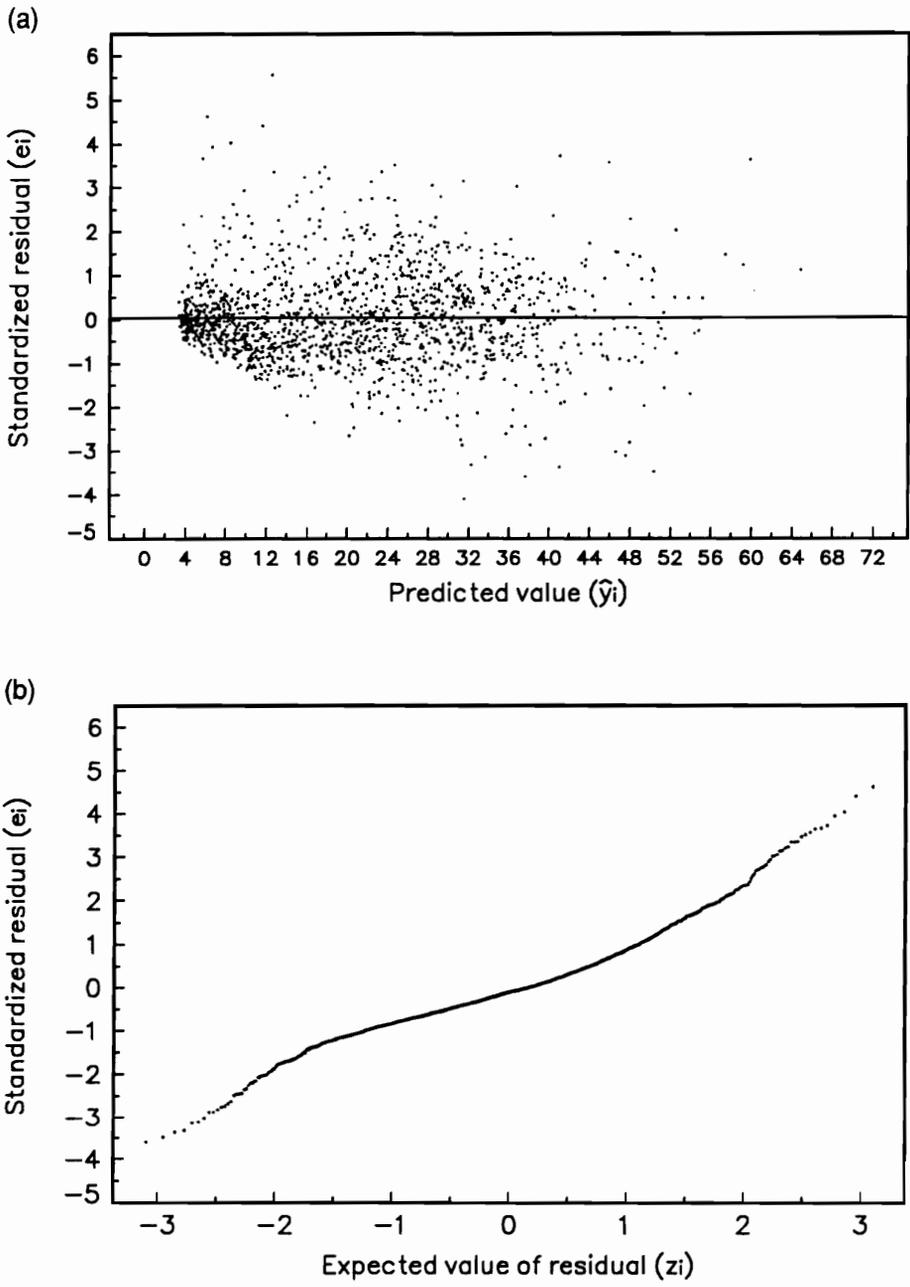


Figure 4.11 Diagnostic plots for $\theta_{0.33}$ regression: (a) residual plot (b) rankit plot.

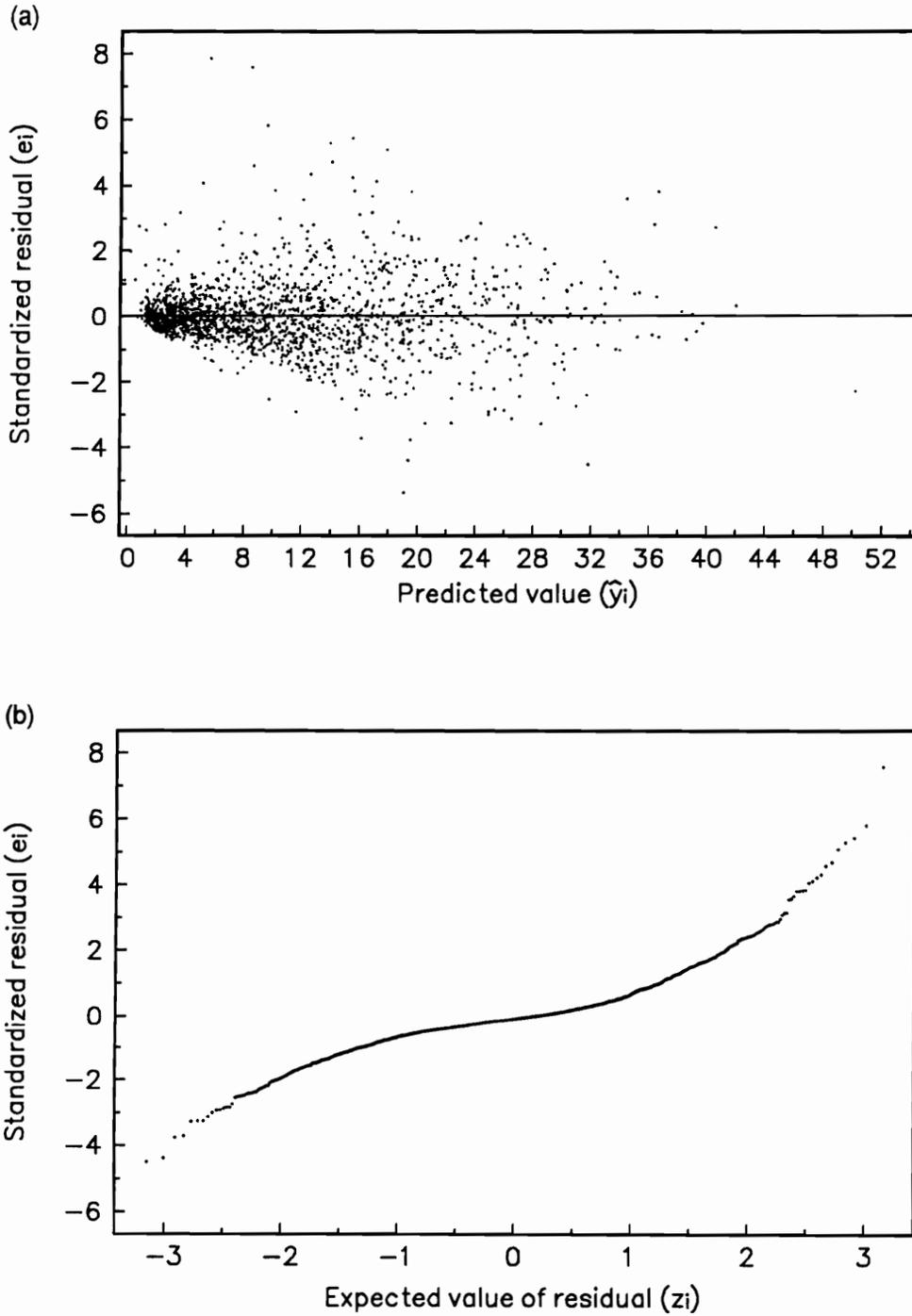


Figure 4.12 Diagnostic plots for θ_{15} regression: (a) residual plot (b) rankit plot.

on the substantive background of the data sets being analyzed. Knowledge of the data generating process is crucial in outlier analysis, which usually requires judgment on the elimination of data points that have been flagged as unusual through the use of various influence diagnostics.

Transformations of the response variable, as we have noted before, are often used as a variance-stabilizing measure. Weisberg (1980) suggests utilizing the square-root, logarithm, or reciprocal transformation for situations where a right-opening megaphone is observed in residual plots. These transformations were applied to the $\theta_{0.1}$, $\theta_{0.33}$ and θ_{15} response variables and regressions run using the sub-sets of predictors obtained previously. The square-root transform was selected as the most appropriate based on the adjusted R^2 values obtained and an examination of residual and rankit plots. Table 4.6 shows the adjusted R^2 of the transformed models. Residual and rankit plots for the transformed models are shown in Figures 4.13 through 4.15. As can be seen from the figures, while there is still some indication of heterogeneous variance and non-normality, significant improvement over the untransformed models has been realized.

Table 4.6 also shows the D-W statistic for the final regressions. The D-W statistic is commonly used to test whether the regression model errors follow a first-order autoregressive process. A formal test for autocorrelation based on the D-W statistic was not carried out, as this requires that observations be equally spaced. However, the D-W statistic was used to provide some indication of the presence of autocorrelation in the data. Values of the D-W statistic near 0 or 4 indicate the absolute value of the lag-1 correlation is close to unity. A D-W statistic near 2 indicates the lag-1 correlation is close to zero. The values of the D-W statistic obtained show that the assumption that observations are independent may be violated, particularly for the data sets utilizing only sand, clay, and porosity as predictors. Although techniques exist for removing autocorrelation in model errors (Myers, 1990), this was not pursued in the current study.

Table 4.6 Adjusted R² and Durbin-Watson (D-W) statistic for final models.

Response	Predictor set 1		Predictor set 2	
	Adjusted R ²	D-W statistic	Adjusted R ²	D-W statistic
$\sqrt{\theta_{0.1}}$	0.771	1.210	0.736	1.020
$\sqrt{\theta_{0.33}}$	0.830	1.290	0.789	1.050
$\sqrt{\theta_{15}}$	0.859	1.330	0.823	0.930
$\ln K_c$	-	-	0.619	0.940

Table 4.7 Input correlations estimated from the soil properties database (Rawls, 1980).

Sand	Clay	Porosity	Organic matter
1			
-0.694	1		
0.000	0.000	1	
-0.219	0.000	0.564	1

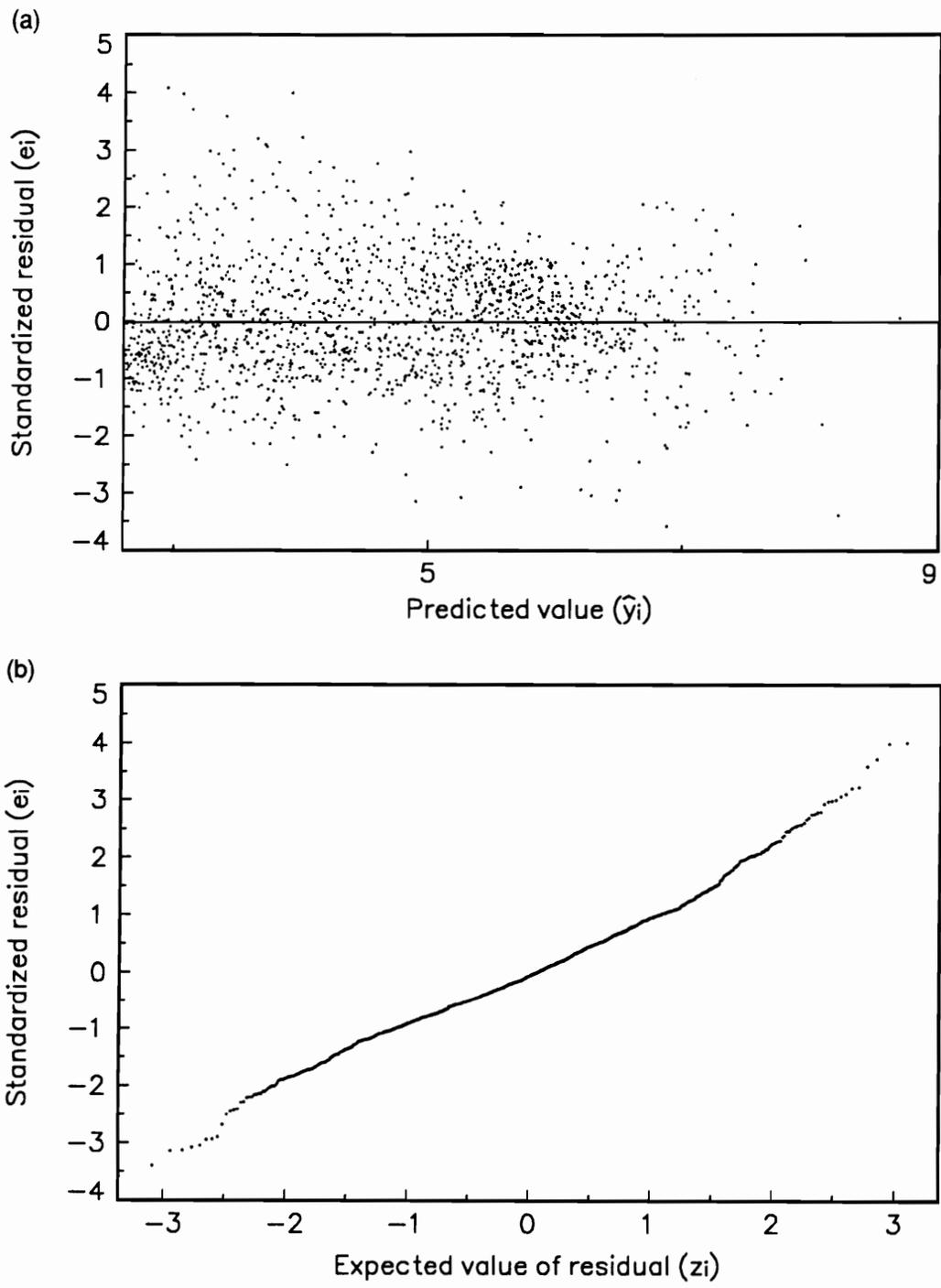


Figure 4.13 Diagnostic plots for $\sqrt{\theta_{0.1}}$ regression: (a) residual plot (b) rankit plot.

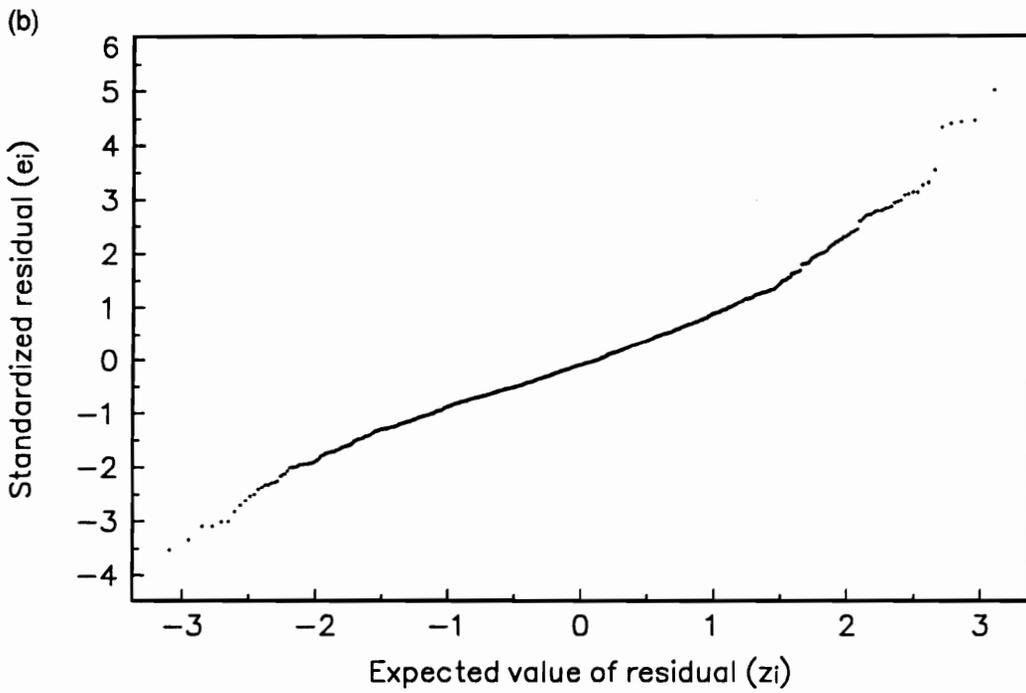
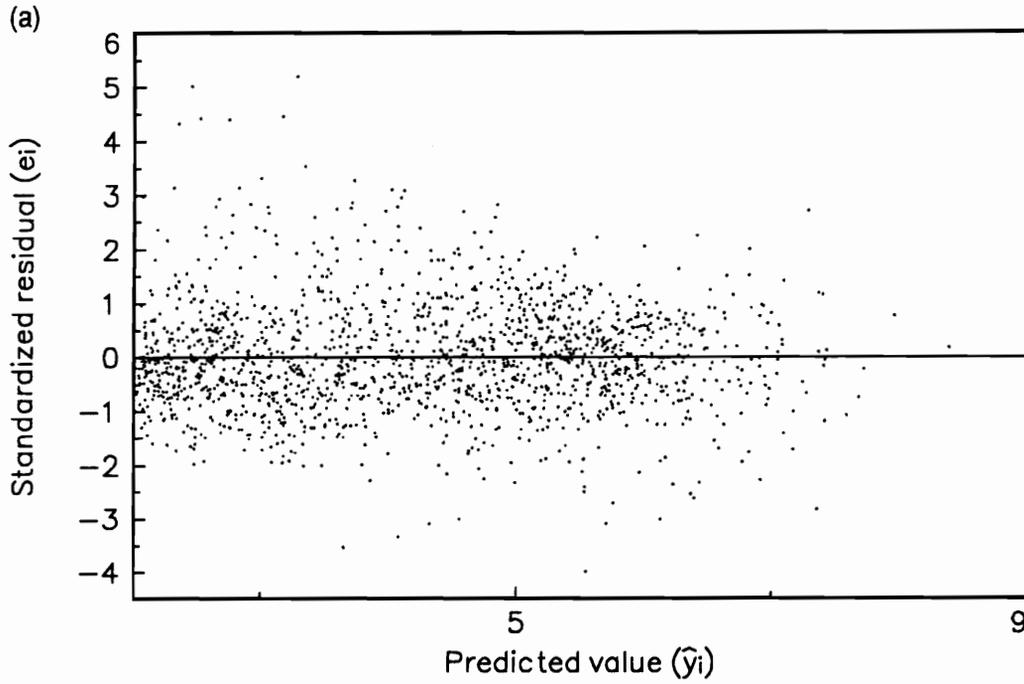


Figure 4.14 Diagnostic plots for $\sqrt{\theta_{0.33}}$ regression: (a) residual plot (b) rankit plot.

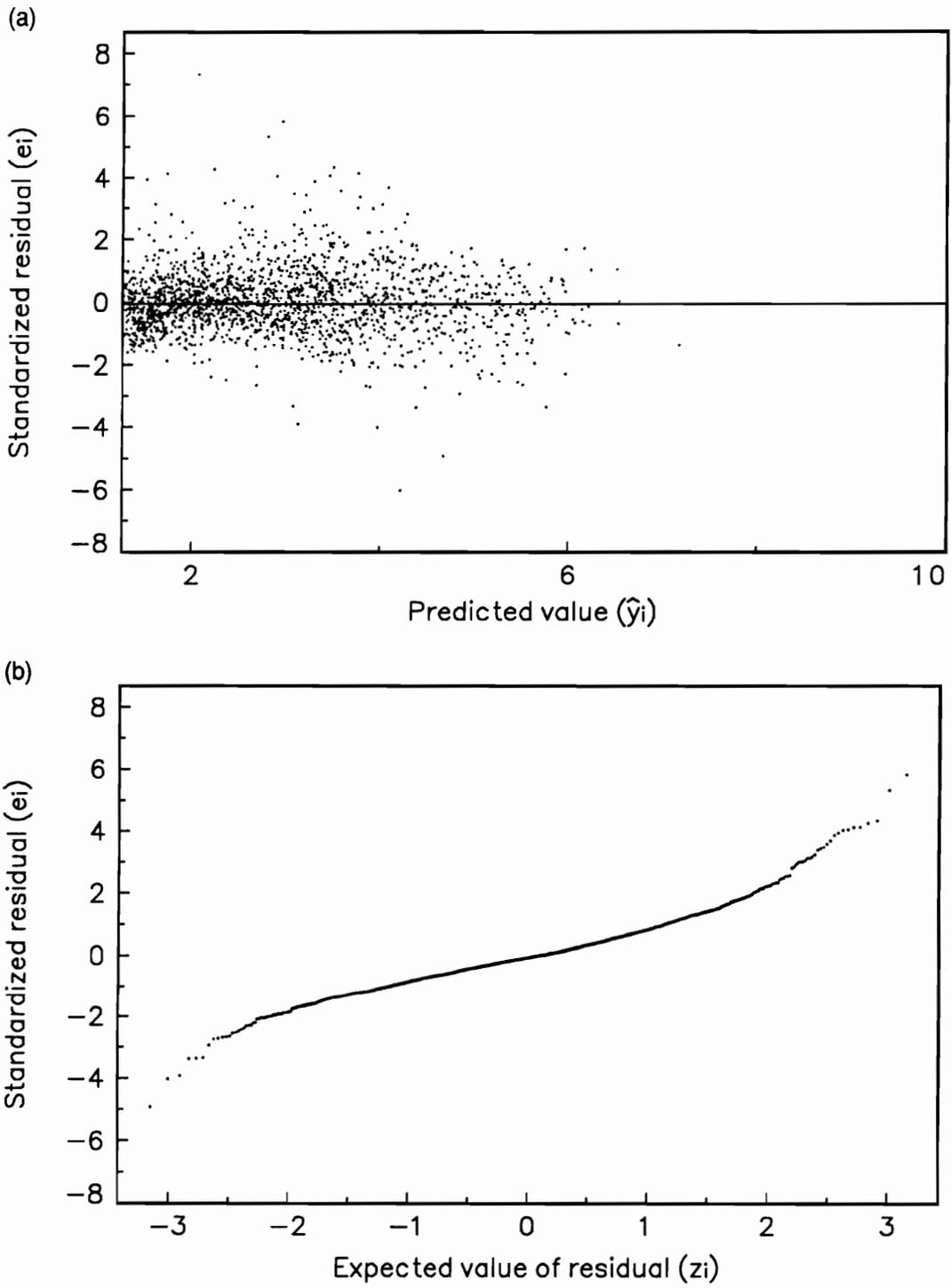


Figure 4.15 Diagnostic plots for $\sqrt{\theta_{15}}$ regression: (a) residual plot (b) rankit plot.

The equations obtained through the regression analysis were incorporated into a computer routine listed in Appendix C. The routine allows convenient generation of distributions for K_s , $\theta_{0.1}$, $\theta_{0.33}$ or θ_{15} , given information on field-scale variation in sand and clay content, porosity, and optionally, organic matter content. The routine may be used in a trace-driven simulation in which observed input vectors of predictor variables are used directly to generate the distributions. Recognizing that in several situations, data characterizing field-scale variability of predictor variables may not be available, the routine allows variation in predictor variables to be defined through expected ranges (maximum and minimum values) for the predictors. Information on ranges in predictor variables is generally more accessible (e.g. from soil survey reports) than observed data on field-scale variation in predictor variables.

If distributions are obtained using input for expected ranges in predictors, the routine assumes inputs follow a beta distribution, rescaled to the appropriate range, with both shape parameters equal to 1.5. This choice for parameters was calculated assuming a symmetrical beta distribution with range equal to four times the standard deviation. Correlations among inputs considered and the correlation matrix used, obtained from the soil properties database (Rawls, 1980) is given in Table 4.7.

In addition to providing point predictions, the routine computes the prediction variance, $\text{Var}(\hat{y} / \bar{\mathbf{x}})$, for an input vector $\bar{\mathbf{x}}$ outside the construction sample as

$$\text{Var}(y / \bar{\mathbf{x}}) = \hat{\sigma}^2 (1 + \bar{\mathbf{x}}' (\mathbf{X}'\mathbf{X})^{-1} \bar{\mathbf{x}}) \quad (4.14)$$

where $\hat{\sigma}^2$ is the mean square error. The square root of the right hand side of equation (4.14) is used to compute one standard deviation bounds for the point prediction in the transformed space of the response variable. However, the point prediction should strictly be regarded as the predicted median and the one standard deviation bounds as the 16th/84th percentile values as the point prediction and bounds are expressed in the original (untransformed) units.

The bounds provide important information on the quality of prediction. Regions in regressor space where prediction is poor will produce large bounds. In addition, the bounds are extremely useful when evaluating the predictive capability of the regression equations.

Evaluation

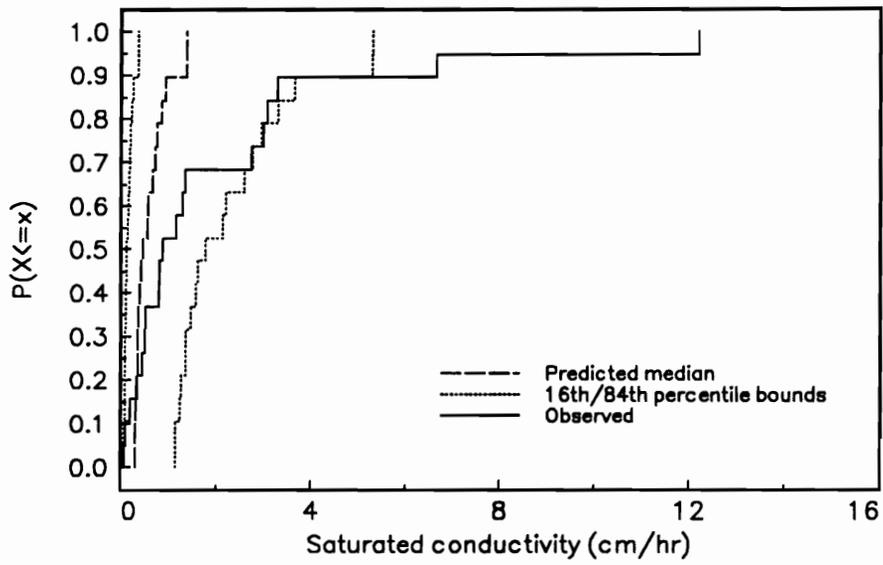
Kool et al. (1986) have reported data from a study to characterize soil physical properties at two sites of the Groseclose soil mapping unit. At each site, core samples were taken at various depths along a 150m transect. Data from the surface horizon at the first site was used to evaluate the utility of the regression equations developed in this section in providing acceptable predictions of field-scale variability in K_s , $\theta_{0.1}$, $\theta_{0.33}$ and θ_{15} .

The evaluation was carried out in two stages:

- i. The predicted median and 16th/84th percentile bounds were obtained for all response variables employing observed vectors of bulk density, and sand, clay, and organic matter content in a trace-driven simulation.
- ii. The predicted median and 16th/84th percentile bounds were obtained with 1000 generated values for all response variables using inputs for the ranges in predictor variables.

Observed EDFs and predicted EDFs from the trace-driven simulation are shown in Figures 4.16a through 4.19a. The observed dispersion in K_s is much greater than the dispersion in the predicted median EDF (Figure 4.16a). Although there are only two observed K_s values which lie outside the predicted 16th/84th percentile bounds, overall the prediction of field-scale variability in K_s cannot be regarded as acceptable. The predicted median EDFs for $\theta_{0.1}$ and $\theta_{0.33}$ correspond closely to the observed EDFs (Figures 4.17a and 4.18a) and indicate excellent prediction of field-scale variability for these response variables. The predicted median EDF for θ_{15} shows some deviation from the observed EDF (Figure 4.19a), particularly in the range $0.1 <$

(a)



(b)

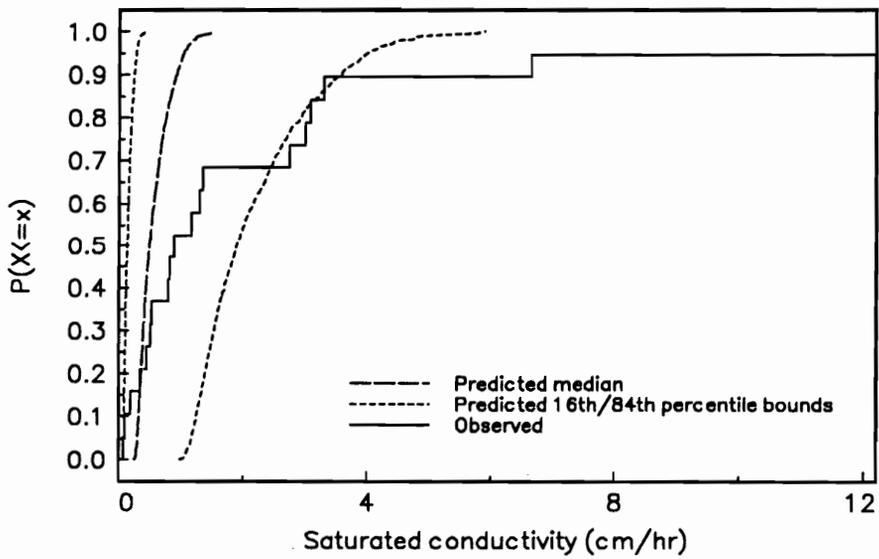


Figure 4.16 Observed and predicted cumulative frequency plots for K_s : (a) trace-driven (b) simulation using input ranges.

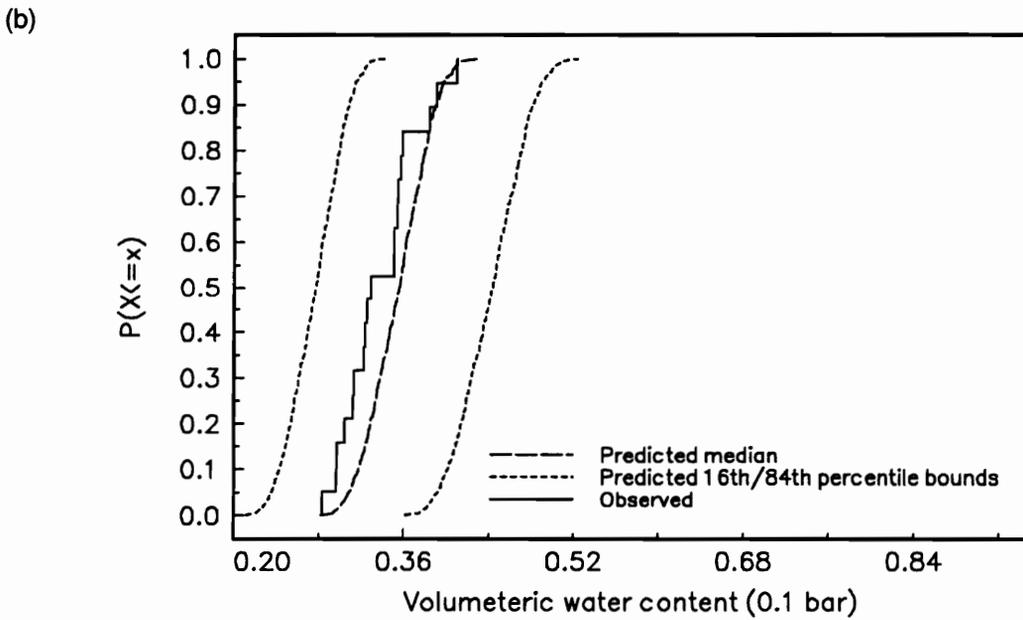
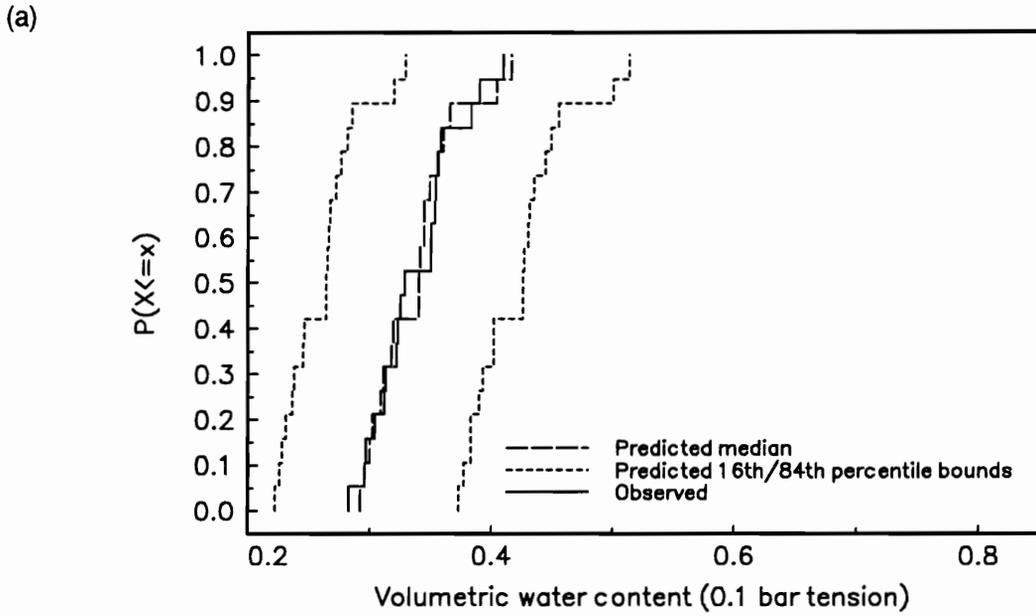


Figure 4.17 Observed and predicted cumulative frequency plots for $\theta_{0.1}$: (a) trace-driven (b) simulation using input ranges.

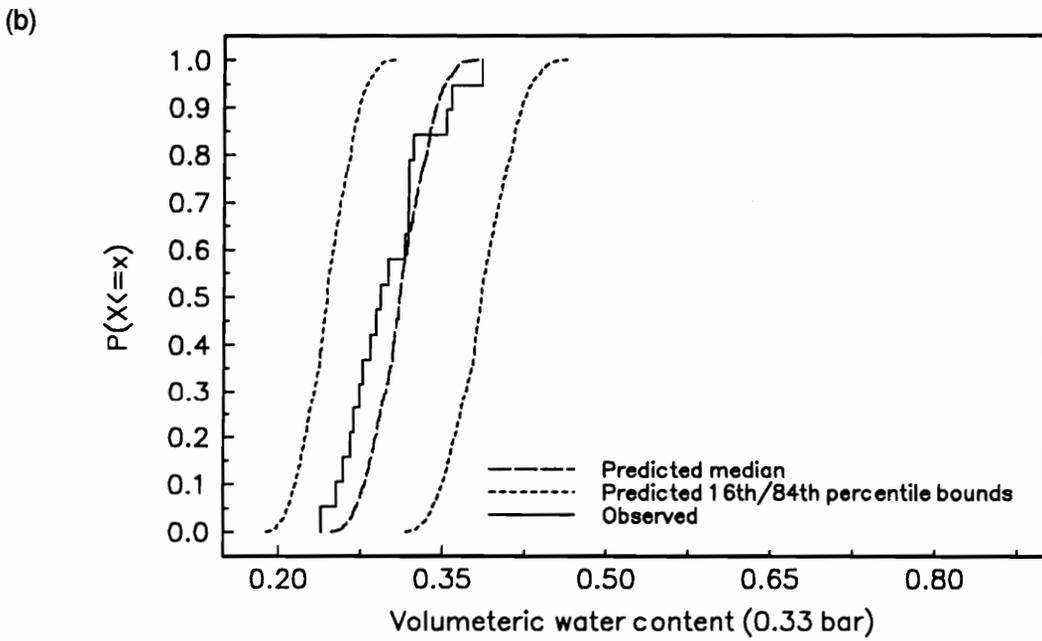
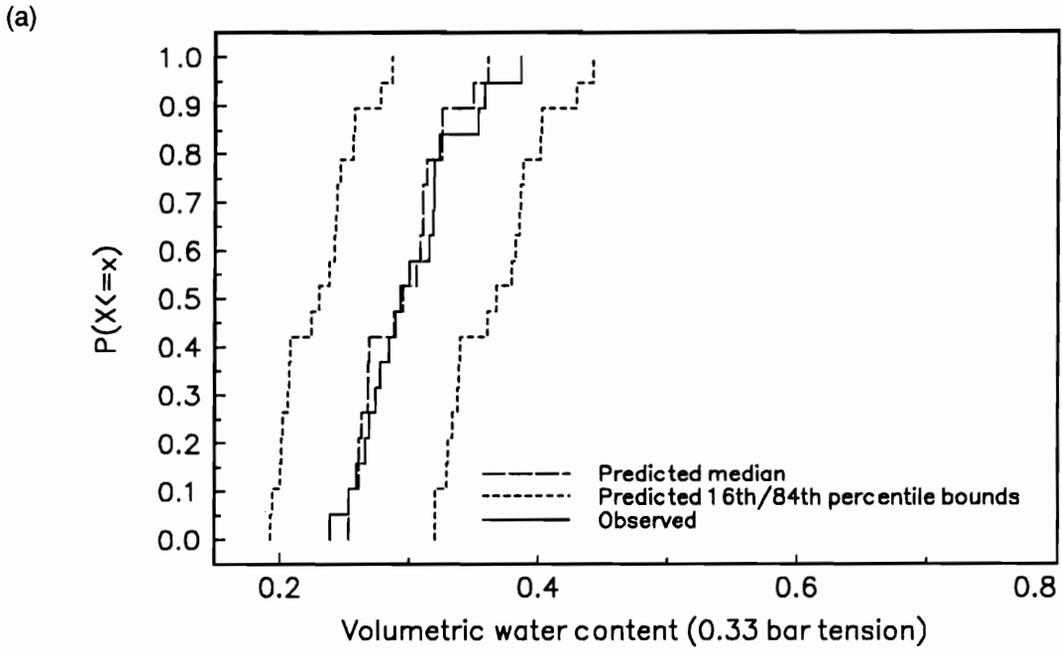
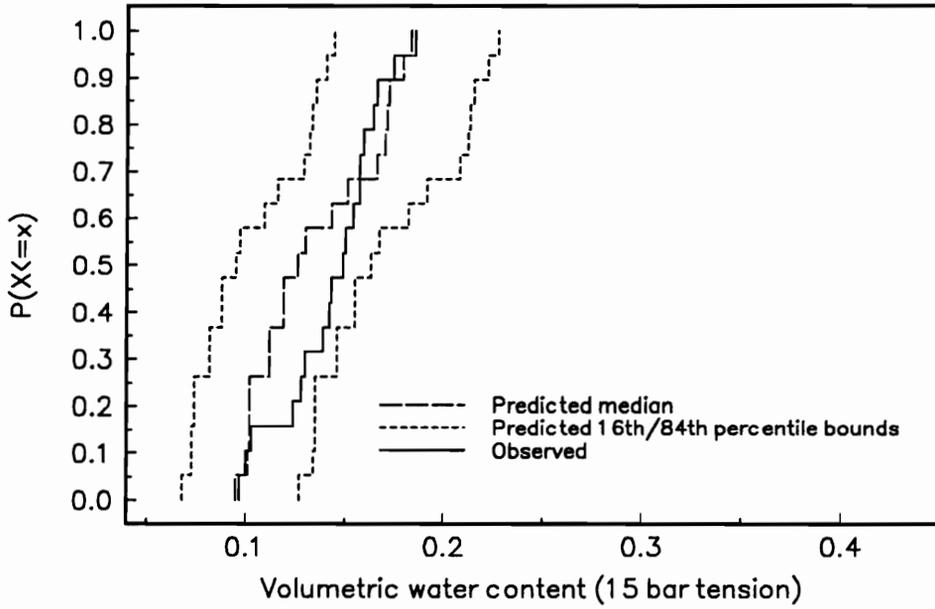


Figure 4.18 Observed and predicted cumulative frequency plots for $\theta_{0.33}$: (a) trace-driven (b) simulation using input ranges.

(a)



(b)

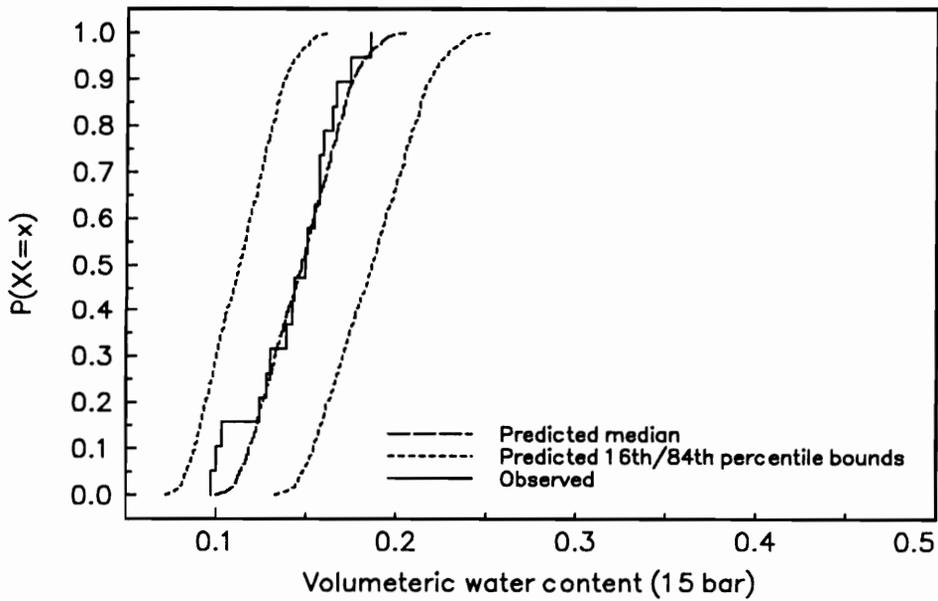


Figure 4.19 Observed and predicted cumulative frequency plots for θ_{15} : (a) trace-driven (b) simulation using input ranges.

$\theta_{15} < 0.15$. However, all observed values are contained well within the predicted 16th/84th percentile bounds and the overall prediction of field-scale variation was regarded as acceptable.

The results of using inputs for the ranges in predictor variables with the variables assumed to follow symmetrical beta distributions and a correlation structure as described previously is shown in Figures 4.16b through 4.19b. The impact of using input ranges compared to using observed input vectors is a slight positive location shift and decrease in dispersion for the predicted median. The width of the 16th/84th percentile bounds with input ranges show less variation than when observed input vectors are used.

Overall, there does not seem to be any large deterioration in prediction when input ranges are used to estimate field-scale variation in the response variables. The positive location shift and decrease in dispersion when input ranges are used actually results in better prediction for θ_{15} compared to using observed input vectors (Figure 4.19). However, in general some deterioration in prediction as observed for K_s , $\theta_{0.1}$, and $\theta_{0.33}$ would be expected when input ranges rather than observed input vectors are used as predictors.

A final assessment of the regression equations developed here was made by comparing point predictions with point predictions from the Rawls and Brakensiek (1989) equations. The observed predictor vectors used in the trace-driven simulation were used to obtain point predictions using the Rawls and Brakensiek (1989) equations. A measure of prediction error, E_p was computed using values of absolute deviations as

$$E_p = \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.15)$$

where y_i and \hat{y}_i are the observed and predicted i th value respectively, and n is the number of observations. A comparison of the prediction errors using the regression equations developed here and the Rawls and Brakensiek (1989) equations is given in Table 4.8. It is seen that for the

data set used here, there is an improvement in prediction for all variables. The improvement in prediction for $\theta_{0.1}$ and $\theta_{0.33}$ is particularly noteworthy.

The regression equations developed here appear to provide acceptable predictions of field-scale variability in field capacity and wilting point. The equations are robust to prediction outside the construction sample, and their predictive ability does not appear to deteriorate significantly when only information on input ranges rather than observed input vectors are used. However, predictions of field-scale variability in saturated conductivity are apparently poor.

Table 4.8 Comparison of point predictions from Rawls and Brakensiek (1985) regression equations and new regression equations.

Predicted variable	Sum of maximum absolute deviations		% improvement
	Rawls & Brakensiek	New	
$\theta_{0.1}$	2.856	0.484	83.1
$\theta_{0.33}$	1.484	0.502	66.2
θ_{15}	0.393	0.338	14.0
K_s	33.352	31.771	4.7

Chapter V

Development of probabilistic model evaluation procedures

...I felt that in a way one could invert a statement by Laplace. He asserts that the theory of probability is nothing but calculus applied to common sense. Monte Carlo is common sense applied to mathematical formulations of physical laws and processes.

Stansilaw M. Ulam

INTRODUCTION

An important issue in the development of a probabilistic framework for nonpoint source pollution modeling is the impact of uncertain inputs in procedures used to evaluate and validate models. It is necessary to ensure, as much as possible, that model evaluation studies focus on the model structure and algorithms and are not affected by uncertainty in inputs due to natural variability and information uncertainty. Haan et al. (1993) note that the evaluation of NPS pollution models by a simple comparison of observed and predicted results is fraught with ambiguities resulting from the large number of parameters and the inherent variability in natural systems. Kumar et al. (1992) have pointed out that the procedures used for uncertainty analysis of model predictions provide the basis for a more complete and thorough evaluation of models.

Probabilistic approaches to evaluating environmental models have been suggested by Luis and McLaughlin (1992) and Haan et al. (1993). However, these approaches cannot be directly applied for evaluating models for relative rather than absolute predictions. The procedures suggested by Luis and McLaughlin are essentially a series of hypotheses tests based on the assumption of normally distributed measurement residuals. The assumption of a normal distribution may not be valid for many variables of interest. Haan et al. (1993) have proposed placing confidence intervals on model results and suggest that decisions regarding model acceptability should be made based on the magnitude of measured data and performance criteria placed on the model. However, Haan et al. (1993) have not suggested any quantitative

measure of model performance which is important for interpreting results when observed data are available over a large spatio-temporal range.

In this chapter, new probabilistic procedures for model evaluation are developed. The procedures do not require the normality assumption and may be applied to assess model performance for both absolute and comparative predictions. A simple probabilistic index of model performance that can be easily interpreted is suggested. The procedures are illustrated by application to the GLEAMS model, using data from a short-term field study to determine pesticide fate and transport in a Coastal Plain soil under alternative tillage treatments.

A PROBABILISTIC APPROACH TO ASSESSING MODEL PERFORMANCE

In the development of a probabilistic approach to assessing model performance we draw an analogy to the goodness-of-fit (g-o-f) testing procedure. A g-o-f test is a statistical hypothesis that is used to formally assess whether a given set of independent observations can be regarded as coming from a population with a specified distribution function, $F(x)$. There are several g-o-f statistics that may be used in the testing procedure. Empirical distribution function (EDF) statistics, based on a comparison of $F(x)$ with the EDF of the sample, $F_n(x)$ are widely employed (Stephens, 1974; Cooke et al., 1993). Law and Kelton (1991) note that the original form of the EDF g-o-f tests are only valid if all the parameters of the hypothesized distribution are known, that is, the distribution parameters cannot be estimated from the sample data. Although Law and Kelton (1991) further note that in recent years tests based on EDF statistics have been extended to allow for estimation of parameters for some distributions, the original form of the EDF g-o-f tests is considered in the analogy drawn here.

We suggest that the model calibration process in which observed data on the response variable of interest are used to adjust the model parameters in some manner and a match attempted between the model predictions and observed data is analogous to fitting a distribution to sample data. In both cases, the goal is to estimate parameters. If the analogy is further extended, just as it is not possible to carry out a g-o-f hypothesis test (in the original form) if the distribution parameters are estimated from the sample data, model evaluation procedures based on calibration are intrinsically flawed. Haan et al. (1993) have also noted that confounding model evaluation with parameter estimation should be avoided.

It may be argued that if an independent set of observed data not employed in calibrating the model is available, this may be used for model evaluation. This would be acceptable if sufficient data were available for calibration and verification in all situations where the model was to be applied. However, with NPS pollution models, this is almost never the case.

With the preceding discussion in mind, a possible evaluation procedure for NPS models would be as follows:

1. Obtain observed data on the variable of interest.
2. Characterize variability or uncertainty in model inputs with appropriate probability distributions.
3. Obtain the model output distribution for the variable through direct Monte Carlo simulation.
4. Carry out a hypothesis test to determine if the observed data could come from the model output distribution, using suitable g-o-f statistics.

Clearly, step (4) is valid only if the observed data have not been used to adjust the parameters of the input distributions. The g-o-f approach is illustrated in Figure 5.1.

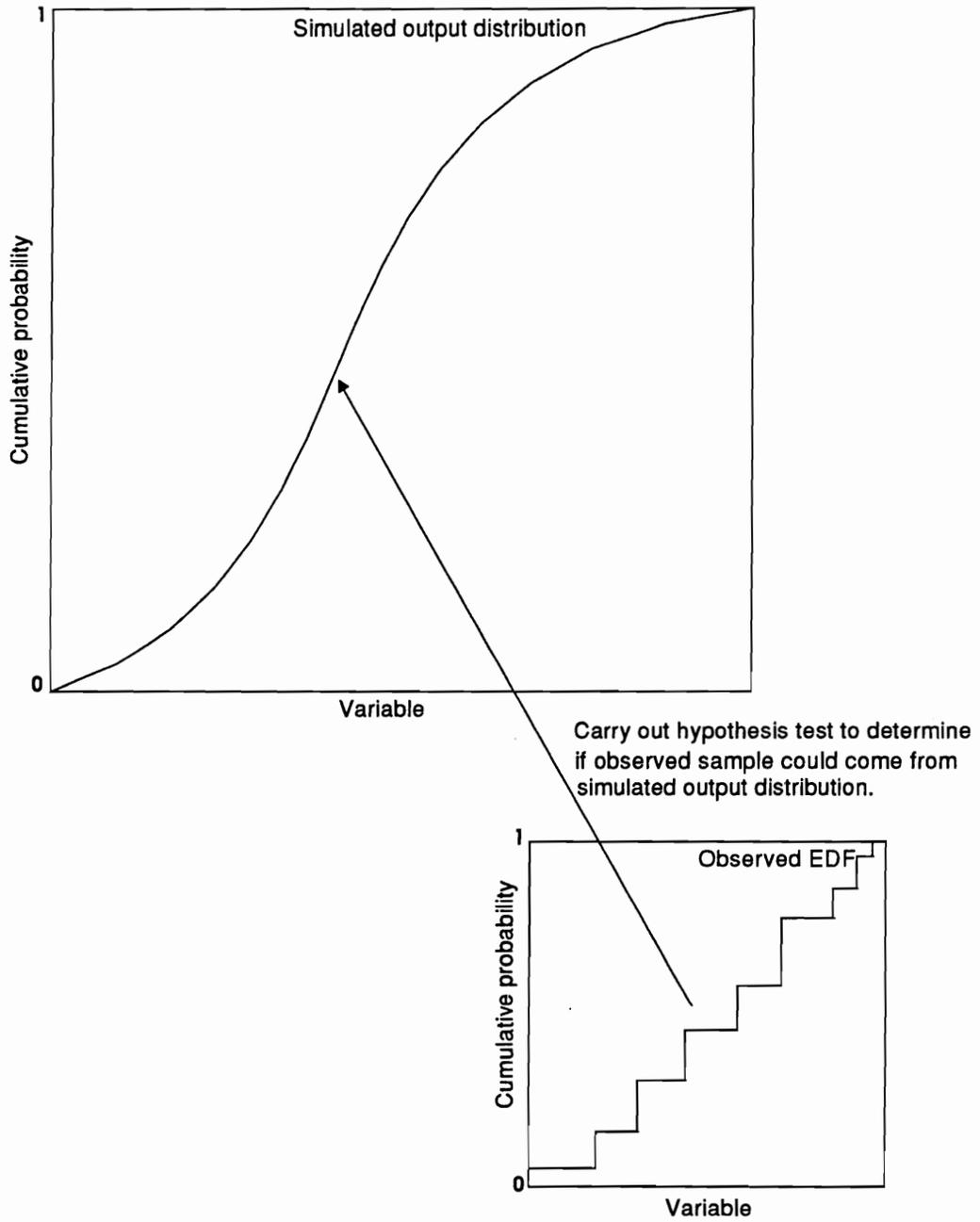


Figure 5.1 Goodness-of-fit approach to model evaluation.

Stephens (1974) described several EDF statistics that may be used in g-o-f hypothesis tests. Two are suggested here, the Kolmogorov-Smirnov (D) statistic and the Anderson-Darling (A^2) statistic. Denoting the model output distribution by $\hat{F}(x)$, the D statistic may be defined as (Law and Kelton, 1991)

$$D = \max \{D_n^+, D_n^-\} \quad (5.1a)$$

$$D_n^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - \hat{F}(X_{(i)}) \right\} \quad (5.1b)$$

$$D_n^- = \max_{1 \leq i \leq n} \left\{ \hat{F}(X_{(i)}) - \frac{i-1}{n} \right\} \quad (5.1c)$$

where n is the number of observed data points, and the X_i 's are the observed data points sorted in ascending order. The A^2 statistic is defined as (Law and Kelton, 1991)

$$A^2 = \left(-\frac{1}{n} \sum_{i=1}^n \left\{ (2i-1) \ln \left[\hat{F}(X_{(i)}) (1 - \hat{F}(X_{(n-i+1)})) \right] \right\} \right) - n \quad (5.2)$$

The D statistic is useful for obtaining an overall measure of goodness-of-fit while the A^2 statistic is designed to detect discrepancies in the distribution tails (Law and Kelton, 1991).

While the goodness-of-fit approach provides the most rigorous and general way to assess model performance, it may be difficult to interpret results if observed data are available over some spatio-temporal range and a large number of hypotheses tests have to be carried out. In several situations, primary interest may be focused on the model's ability to predict some statistic computed from the data - for example, the sample mean. In this case, the model sampling distribution of the statistic of interest can be easily obtained by resampling the output distribution using the required sample size. The observed statistic is then located in the sampling distribution and a probabilistic index of model performance, I_p , computed as

$$I_p = \log_{10} \frac{p_e}{1 - p_e} \quad (5.3)$$

where p_e is the probability of exceedence of the observed statistic in the sampling distribution. The index provides a convenient transformation of p_e which is easily interpreted. An index value of 0 is obtained when p_e is equal to 0.5, which would indicate excellent model prediction. Negative and positive values of I_p would indicate under- and over-prediction, respectively. It should be noted that if only a single observed value is available, I_p can be calculated directly from the model output distribution. Figure 5.2 illustrates how the value of p_e is obtained from the model output distribution or the output sampling distribution. A scatter plot of computed indices for all spatio-temporal points for which observed data are available provides a comprehensive assessment of model performance.

A significant advantage of the procedures proposed here is that they may be used to evaluate model performance for relative model predictions under alternative management scenarios. For example, if the difference in predicted medians of some output variable from alternative scenarios is of interest, the model output distributions for both scenarios would be obtained. The sampling distribution of the difference in predicted medians could then be determined from these distributions and I_p calculated as before. A problem that may arise in evaluating a model for comparative response is that with the same rainfall driving the response, the outputs tend to be correlated and can therefore only be analyzed as paired samples. However, the observed data is generally drawn from populations that are regarded as independent. In a Monte Carlo framework this problem can be circumvented by using different random number streams for simulating the two (or more) management options of interest. The use of common random numbers is recommended in simulation studies to reduce model output variance (Law and Kelton, 1991). However, when evaluating a model for comparative response with observed data from independent populations, separate random number streams are necessary to ensure independent model output.

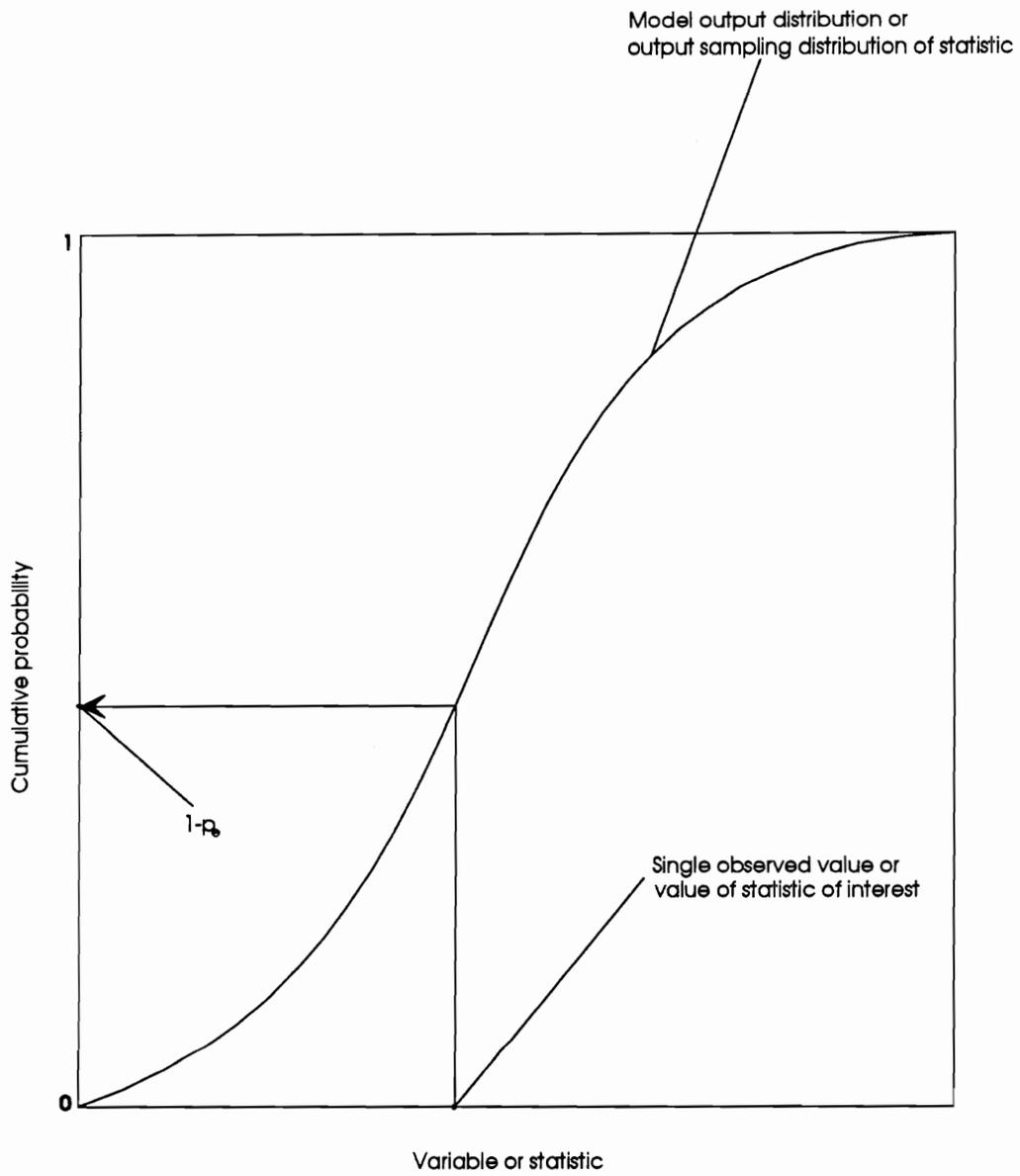


Figure 5.2 Obtaining the value of p_0 from the model output distribution or output sampling distribution.

APPLICATION

The model evaluation procedures were applied to the GLEAMS model. The intent of the application is not to carry out a rigorous evaluation of the model *per se*, but rather to illustrate the manner in which the procedures suggested here may be applied and the evaluation results interpreted.

Field Data

Data for the application were obtained from a field study conducted by Heatwole et al. (1992). The following description of the field data is derived from Zacharias and Heatwole (1993b).

Surface runoff monitoring and soil core sampling were used to characterize the fate and transport of atrazine, metolachlor, and bromide as a tracer from two 18 × 27 m field plots in the Coastal Plain region of Virginia (Westmoreland County). The soil is a Suffolk sandy loam (coarse-loamy, siliceous, thermic Typic Hapludult) characterized as deep and well drained. The plots were separated by a 6m buffer and located in a field that was in the second year of a two year conservation tillage wheat-soybean-corn rotation typical of the region. One plot was plowed and disked before corn was planted and chemicals applied, while the other plot remained under conservation tillage with a heavy soybean-wheat residue. Surface runoff quantity and quality measurements from both plots were made over a 22-week period. Soil core samples were collected up to 1.5m depth at 20 random locations in each plot on six dates over the 22-week period. Soil samples were placed on ice and stored at 4°C until analyzed for atrazine, metolachlor, and bromide concentration, as well as for organic matter content, pH, and gravimetric moisture content.

In the application of the model evaluation procedures, data on Metolachlor mass in the root-zone (up to 90cm) in the conservation tillage plot is selected to illustrate the goodness-of-fit approach to model evaluation. On this date, the maximum variation in Metolachlor mass in the root-zone of the no-till plot was observed. To illustrate the model evaluation procedures for comparative analysis, the observed value of the median difference in Metolachlor mass between the conservation tillage and conventional tillage plots on the last sampling date is used. Mann-Whitney tests (Gibbons, 1985) for differences in Metolachlor mass between the plots indicated the most significant difference on this date. Finally, data on surface losses of Atrazine from the conventional tillage plot over the monitoring period is used to illustrate the utility of scatterplots of computed I_p values in providing a comprehensive assessment of model performance.

Characterization of model inputs

GLEAMS inputs that were considered random variables based on physical considerations and the sensitivity analysis carried out previously (Chapter II) are shown in Table 5.1. The field-scale variability in many inputs have been reported to be adequately modeled by a normal or Lognormal distribution (Jury, 1991). Inputs for which distribution types could not be found in the literature were modeled by either the uniform, beta, or triangular distributions. The different distribution types used and their parameters are summarized in Table 5.2.

The soil profile was divided into 4 zones with bottom depths of 15cm, 30cm, 75cm, and 90cm based on information in horizon differentiation given for the Suffolk soil series in the soil survey of Westmoreland County (SCS, 1982). Available data for porosity (inferred from bulk density), sand, clay, and organic matter content for different horizons were used to obtain the probability distributions for these variables employing VTFIT (Cooke, 1993). Sufficient organic matter data was available to allow separate distributions to be fitted to the conventional and conservation tillage plots. However, distributions for sand and clay content, and porosity were the same for

Table 5.1 Inputs considered random variables and abbreviations.

Input description	Abbreviation
Soil evaporation parameter	CONA
Porosity by horizon	POR()
Field capacity by horizon	FC()
Immobile water content by horizon	BR15()
Organic matter content by horizon	OM()
Clay content by horizon	CLAY()
Silt content by horizon	SILT()
AMC II runoff curve number (potential retention)	CN2 (S)
Soil erodibility	KSOIL
Soil loss ratios for rotation	CFACT()
Manning's roughness factor	NFACT()
Pesticide application rate	APPL
Pesticide partitioning coefficient	KOC
Pesticide foliar half-life	HAFLIF
Pesticide soil half-life	SOLLIF
Pesticide wash-off fraction	WSHFRC
Coefficient for plant pesticide uptake	COFUP
Fraction of pesticide applied to crop residue	FOLFRC

() indicates multiple values may be input by horizon or crop rotation.

Table 5.2 Distributions used for inputs in model evaluation.

Distribution and parameters	Parameter interpretation
Uniform - $U(a,b)$	a =location, $b-a$ =scale
Normal - $N(\mu,\sigma)$	μ =location, σ =scale
Lognormal - $LN(\mu,\sigma)$	μ =scale, σ =shape
Beta - $B(\alpha_1,\alpha_2,a,b)$	α_1, α_2 =shape, $b-a$ =scale, a =location
Triangular - $T(a,b,c)$	$b-a$ =scale, c =shape

Table 5.3 Distributions and parameters common to conservation and conventional tillage simulations.

Input (units)*	Distribution	Parameter values
CONA (mm day ^{-1/2})	U(a,b)	(3.3,4.0)
POR(2)	N(μ,σ)	(0.435,0.0534)
POR(3)	LN(μ,σ)	(-0.9841,0.0805)
POR(4)	N(μ,σ)	(0.434,0.0377)
KSOIL (Mg/ha)	LN(μ,σ)	(-0.803,0.275)
KOC-Atrazine (ml/g)	LN(μ,σ)	(4.605,0.263)
KOC-Metolachlor (ml/g)	LN(μ,σ)	(5.298,0.283)
SOLLIF-Atrazine (days)	LN(μ,σ)	(4.094,0.257)
SOLLIF-Metolachlor (days)	LN(μ,σ)	(4.500,0.677)
HAFLIF-Atrazine (days)	LN(μ,σ)	(1.609,0.262)
HAFLIF-Metolachlor (days)	LN(μ,σ)	(1.609,0.703)
WSHFRC-Atrazine	U(a,b)	(0.450,0.760)
WSHFRC-Metolachlor	U(a,b)	(0.550,0.860)
COFUP	U(a,b)	(0.5,1.0)
SILT(1) (%)	B(α_1,α_2,a,b)	(1.385,1.235,17.4,29.5)
SAND(1)# (%)	B(α_1,α_2,a,b)	(1.889,2.524,63.0,78.7)
SAND(2)# (%)	B(α_1,α_2,a,b)	(1.785,2.986,52.2,74.3)
SAND(3)# (%)	B(α_1,α_2,a,b)	(0.807,0.946,44.8,77.0)
SAND(4)# (%)	B(α_1,α_2,a,b)	(1.376,0.896,43.2,93.2)
CLAY(1)# (%)	B(α_1,α_2,a,b)	(3.668,1.613,3.9,8.0)
CLAY(2)# (%)	B(α_1,α_2,a,b)	(1.796,2.462,5.7,20.6)
CLAY(3)# (%)	B(α_1,α_2,a,b)	(1.400,1.317,11.03,21.8)
CLAY(4)# (%)	B(α_1,α_2,a,b)	(1.133,1.003,3.2,22.7)

* No units indicate input is dimensionless

Used to obtain distributions for FC() and BR15().

Table 5.4 Distributions and parameters used to simulate differences between conservation and conventional tillage systems.

Input (units)*	Conservation tillage		Conventional tillage	
	Distribution	Parameter values	Distribution	Parameter values
POR(1)	N(μ,σ)	(0.435,0.0534)	N(μ,σ)	(0.516,0.0534)
FC(1)	N(μ,σ)	(0.1962,0.0145)	LN(μ,σ)	(-1.558,0.1203)
FC(2)	N(μ,σ)	(0.2400,0.0293)	LN(μ,σ)	(-1.432,0.1221)
FC(3)	N(μ,σ)	(0.2589,0.0232)	N(μ,σ)	(0.260,0.0231)
FC(4)	N(μ,σ)	(0.2235,0.0628)	N(μ,σ)	(0.225,0.0625)
BR15(1)	N(μ,σ)	(0.0521,0.00459)	N(μ,σ)	(0.051,0.00644)
BR15(2)	LN(μ,σ)	(-2.582,0.235)	LN(μ,σ)	(-2.577,0.232)
BR15(3)	N(μ,σ)	(0.1091,0.0213)	N(μ,σ)	(0.1096,0.0213)
BR15(4)	N(μ,σ)	(0.0801,0.0322)	N(μ,σ)	(0.0807,0.0323)
OM(1) (%)	N(μ,σ)	(0.948,0.185)	N(μ,σ)	(0.817,0.240)
OM(2) (%)	LN(μ,σ)	(-0.726,0.432)	N(μ,σ)	(0.551,0.235)
OM(3) (%)	N(μ,σ)	(0.3172,0.125)	LN(μ,σ)	(-1.072,0.439)
OM(4) (%)	LN(μ,σ)	(-1.622,0.6025)	N(μ,σ)	(0.259,0.195)
S (mm)	LN(μ,σ)	(4.892,0.524)	LN(μ,σ)	(4.216,0.937)
CFACT(1)	T(a,b,c)	(0.297,0.363,0.330)	T(a,b,c)	(0.297,0.363,0.330)
CFACT(2)	T(a,b,c)	(0.297,0.363,0.330)	T(a,b,c)	(0.702,0.858,0.780)
CFACT(3)	T(a,b,c)	(0.261,0.319,0.290)	T(a,b,c)	(0.585,0.715,0.650)
CFACT(4)	T(a,b,c)	(0.225,0.275,0.250)	T(a,b,c)	(0.459,0.561,0.510)
CFACT(5)	T(a,b,c)	(0.162,0.198,0.180)	T(a,b,c)	(0.270,0.330,0.300)
CFACT(6)	T(a,b,c)	(0.126,0.154,0.140)	T(a,b,c)	(0.225,0.275,0.250)
CFACT(7)	T(a,b,c)	(0.297,0.363,0.330)	T(a,b,c)	(0.333,0.407,0.370)
NFACT	LN(μ,σ)	(-1.204,0.269)	LN(μ,σ)	(-1.386,0.333)
FOLFR	U(a,b)	(0.3,0.88)	Constant	0.99
APPL-Atrazine (g/ha)	B(α_1,α_2,a,b)	(1.231,1.515,0.52,0.49)	B(α_1,α_2,a,b)	(2.863,1.979,0.61,1.14)
APPL-Metolachlor (g/ha)	B(α_1,α_2,a,b)	(1.701,3.968,0.46,2.35)	B(α_1,α_2,a,b)	(2.825,5.780,0.67,1.34)

* No units indicate input is dimensionless

predicted distribution is 0.289. The corresponding A^2 statistic is 5.379. From tables of critical values of D and A^2 statistics for various sample sizes provided by Stephens (1974), the D statistic will result in a failure to reject the null hypothesis that the observed data could be derived from the model output distribution at the 0.05 level. The computed statistic is very close to the critical value at the 0.05 level, implying that the goodness-of-fit may not be very good. The A^2 statistic results in a rejection of the null hypothesis ($p < 0.01$). On examining Figure 5.3, it can be seen that there are large discrepancies between the observed EDF and model output distribution in the left tail, which explains the rejection of the null hypothesis using the A^2 statistic. Although the model output distribution generally suggests over-prediction, agreement between observed and predicted values is reasonable. Thus, from a practical standpoint, and the fact that the Kolmogorov-Smirnov goodness-of-fit test fails to reject the null hypothesis of equivalent model and observed distributions at the 0.05 level, the model performance can be regarded as acceptable.

Figure 5.4 shows the model output sampling distribution (for a sample size of 20) of median difference in Metolachlor mass between the conservation and conventional tillage plots on the last sampling date. The observed value of the median difference is also shown in the figure. The sampling distribution was generated by computing the difference between the medians of twenty values obtained independently from the output distributions for the conventional and conservation tillage simulations. A test for association between the output distributions strongly failed to reject the null hypothesis of no association ($p > 0.24$), confirming that using separate random number streams for model inputs for different scenarios results in independent outputs. The computed value of I_p for the observed median difference is -1.07, indicating the model considerably under-predicts the median difference. The sampling distribution further shows that the probability of the median pesticide mass in the conventional tillage plot being higher than in the conservation tillage plot is about 27 %. Thus, even on an ordinal basis, the model does not seem to predict relative differences well on this particular sampling date.

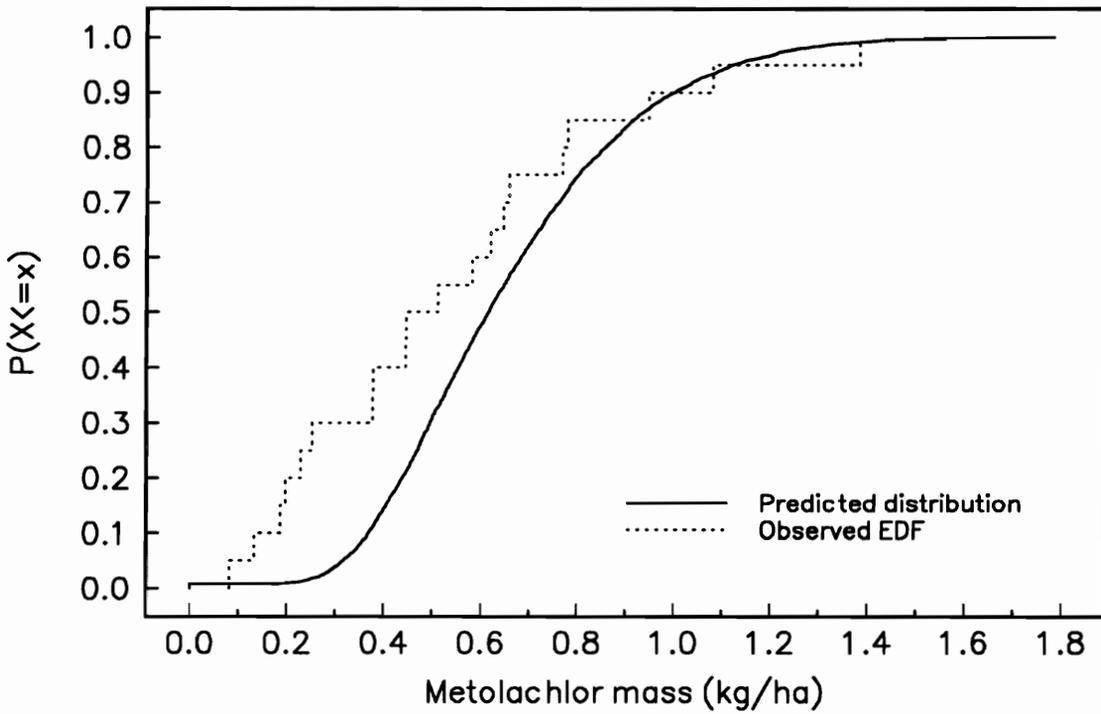


Figure 5.3 Predicted distribution and observed empirical distribution function (EDF) of Metolachlor mass in the soil profile on second sampling date.

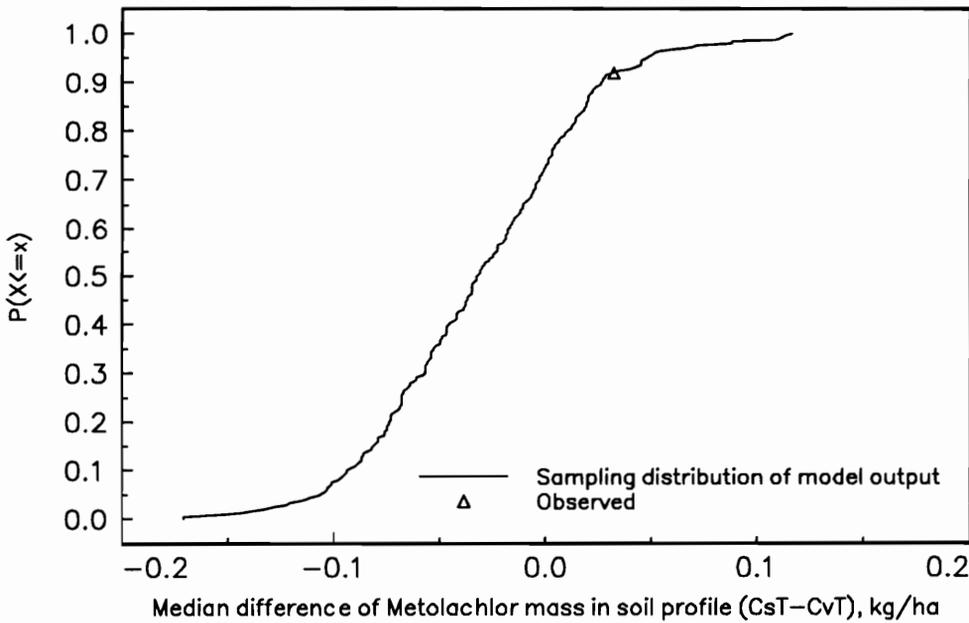


Figure 5.4 Model output sampling distribution ($n=20$) and observed value of median difference in Metolachlor mass between conservation tillage (CsT) and conventional tillage (CvT) on last sampling date.

Figure 5.5a shows the observed surface losses of Atrazine from the conventionally tilled plot over the monitored period. A scatter plot of I_p values, computed for the observed values from the model output distribution is shown in Figure 5.5b. Clearly, the model under-predicts Atrazine losses, and the degree of under-prediction increases with time as Atrazine concentrations on the soil surface decline. Although the I_p values for observed losses are between 1 and -1 on days 117, 119, 130 and 146, the general tendency to under-predict, which is apparently more pronounced when large losses occur (day 124), would lead to the conclusion that the model does not provide acceptable predictions of surface losses of Atrazine from the conventionally tilled plot.

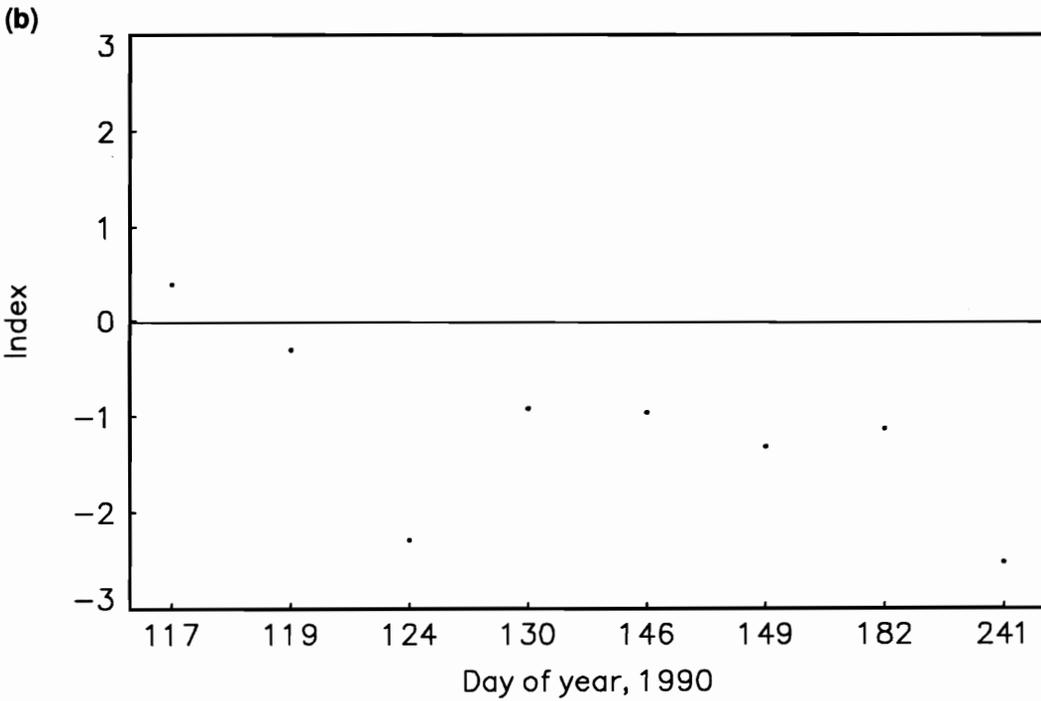
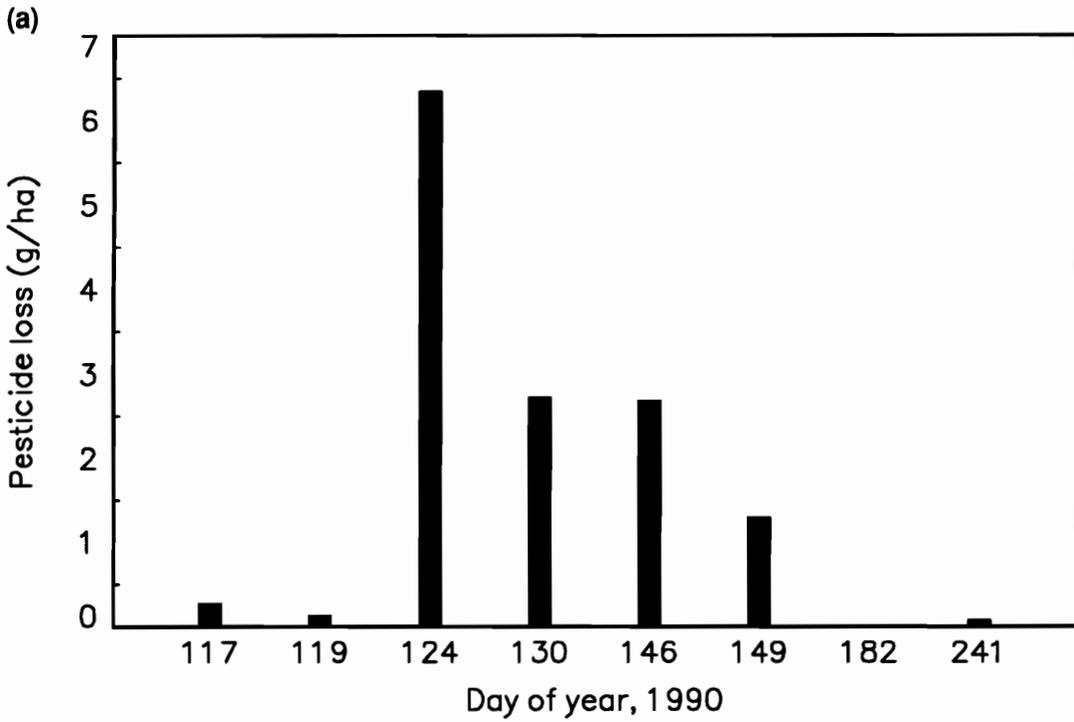


Figure 5.5 (a) Observed surface losses of Atrazine from conventionally tilled plot, and (b) scatter plot of indices of model performance over monitored period.

Chapter VI

Summary, Conclusions, and Recommendations

*The woods are lovely, dark, and deep,
But I have promises to keep,
And miles to go before I sleep,
And miles to go before I sleep.*

Robert Frost

SUMMARY AND CONCLUSIONS

Addiscott (1993) comments that today's computer modelers could very well find their activities contributing to decision making on environmental and related issues. Addiscott (1993) further notes that while this is to be welcomed, it emphasizes the "essentiality of getting models right - or as nearly right as possible - and of making sure that they are used properly, particularly in any form of decision-making process."

While current NPS pollution models are almost exclusively deterministic, the inherently stochastic nature of many model inputs, and the sometimes large degree of information uncertainty suggests that these models should be used in a probabilistic framework. Direct Monte Carlo simulation allows input uncertainty to be incorporated into deterministic models, and may be used to quantify the uncertainty associated with model predictions. However, the interpretation of output resulting from a dynamic Monte Carlo simulation of a continuous NPS pollution model is not straightforward. Indeed, previous Monte Carlo studies of continuous models have not attempted to examine the impact of input uncertainty in a dynamic framework.

In this dissertation, appropriate procedures for statistically analyzing output from continuous models for both absolute predictions and comparative analysis from dynamic Monte Carlo simulations are presented. The methodology is developed and demonstrated using the pesticide

component of a widely used management-oriented model, GLEAMS. Annual values of model output resulting from a single Monte Carlo trial were shown to be independent, which allows the explicit separation of the impact of parameter uncertainty and rainfall variability on model output by a simple re-ordering of the dynamic Monte Carlo output matrix. When analyzing model output from two or more alternative scenarios, it was shown that statistical procedures for paired or related samples are appropriate.

In a simulation study with two hypothetical scenarios, parameter uncertainty was found to have important consequences for both absolute predictions and comparative analysis between the scenarios. When the effect of climatic variability on absolute predictions was compared at different exceedence probabilities (EPs) due to parameter uncertainty, it was found that the impact was greater at the 10% EP representing 'critical conditions' than at the 50% EP representing 'average conditions'. The magnitude of the increase in variability was found to depend on the extent to which output distributions changed in shape over the simulation period. For the two scenarios considered in the simulation study, conclusions about comparative water and pesticide percolation that would be made from a deterministic analysis were found not to be justified when input variances and correlations were considered.

To facilitate the implementation of direct Monte Carlo simulation of deterministic computer models with correlated inputs, general-purpose routines were developed. A pre-processor module for generating correlated random variates from several commonly used probability distributions enables rapid batch implementation of direct Monte Carlo simulation. The module employs a modified multivariate-normal approach suggested by Taylor and Bender (1989), and allows truncation of distributions which are not bounded on both ends. To generate correlated random variates, it is necessary that the input correlation matrix be positive definite, and can therefore be factorized. Information on correlations among input variables in Monte Carlo studies is not readily available, and is commonly determined subjectively. A novel interactive approach

to generating a correlation matrix from subjective information was developed. The procedure, which guarantees a positive definite matrix, retains correlations between input variables with which a high degree of confidence is associated, and allows use of these correlations to examine possible correlations between other variables.

When seeking 'effective' input parameters for a field-scale model with no observed data on either the inputs or the response variables of interest (i.e. under information uncertainty), the use of uniform probability distributions over expected ranges may be a reasonable first assumption for Monte Carlo simulation. This would be analogous to the approach employed in Bayesian statistical procedures for defining prior probabilities when no information exists about the prior state of a variable (Clemmens, 1992). However, it is clearly desirable to reduce information uncertainty as much as possible, and consider only spatial and temporal variability in the physical system. Again, this is analogous to improving probability estimates in Bayesian statistics using information from prior processes. Recognizing that characterizing input probability distributions by direct fitting to measured values is rarely possible because of cost and time constraints, the use of indirect methods which achieve an acceptable degree of accuracy were developed and evaluated for critical hydrologic inputs in GLEAMS. The sediment and chemical transport components of NPS pollution models are driven by the hydrologic component, and therefore, defining probability distributions for hydrologic inputs is necessarily the first step.

Procedures were developed for defining probability distributions for runoff curve number, saturated conductivity, field capacity, and wilting point under conditions of limited data availability. A conceptual approach was used to derive the runoff curve number distribution for average moisture conditions, suitable for use in continuous models which employ the curve number method, from the complete curve number distribution for all moisture conditions. Procedures for estimating the complete curve number distribution from the SCS tabulated average curve number value and suitable assumptions about curve numbers specified at

different antecedent runoff conditions when no rainfall-runoff data is available were outlined. A new set of regression equations, similar to those proposed by Rawls and Brakensiek (1985) were developed for predicting saturated conductivity, field capacity, and wilting point. The equations, which employ sand and clay content, porosity, and if available, organic matter content as predictors are robust to prediction outside the construction sample data. Evaluation of the regression equations using an independent data set indicated the equations provided acceptable predictions of field-scale variability in field capacity and wilting point but not of saturated conductivity. There was no significant deterioration in prediction when input ranges rather than observed input vectors were used to estimate field-scale variation in the response variables. A comparison of point predictions from the new equations with those suggested by Rawls and Brakensiek (1985) showed improved prediction for all response variables.

The impact of uncertain inputs in procedures used to evaluate and validate models was considered an important issue in the development of a probabilistic framework for nonpoint source pollution modeling. Theoretically sound procedures for probabilistic evaluation of NPS pollution models were developed by analogy with the goodness-of-fit testing procedure. The procedures do not require the assumption of normality of prediction errors for application, and may be used to assess model performance for both absolute and comparative predictions. A simple probabilistic index of model performance that can be easily interpreted was suggested. The procedures were illustrated by application to the GLEAMS model using data from a short-term field-study to determine pesticide fate and transport in a Coastal Plain soil under alternative tillage treatments.

The use of management-oriented continuous models to assess NPS loading impacts in a deterministic framework has limitations for both absolute predictions and comparative analysis. In simulation studies where there is typically no data for calibration and several critical input parameters have wide ranges, there is little justification for selecting a value in the 'middle' of the

range as a 'best' estimate. Using best estimates will result in outputs which correspond approximately to the median of the model output distribution that would be obtained through direct Monte Carlo simulation. Assessing impacts and basing decisions on outputs which correspond to the distribution median (50% exceedence probability) may be meaningful in situations where 'steady-state' conditions are physically relevant, but in fact, with NPS pollutant loadings, outputs at a lower exceedence probability corresponding to 'critical-conditions' may be more significant.

The simulation study revealed that even with a management-oriented model such as GLEAMS, which would not be expected to exhibit a highly non-linear response, parameter uncertainty results in distributions for some outputs that may change dramatically over the simulation period, resulting in climatic variability having a greater impact on model results at lower EPs. Furthermore, in the simulation study, output distributions did not merely undergo location shifts when different management scenarios were considered, even though the input distribution shapes for the two scenarios were the same, which resulted in different estimators for output differences at different EPs. One result of the change in output distribution shapes for different scenarios is that a comparative analysis carried out at different EPs may result in different conclusions. It is therefore important to determine through experimental and expert opinion, the appropriate EP (or risk level) on which to base decisions. A deterministic analysis will generally result in a decision based on the 50% EP, which may not be appropriate for NPS pollutants.

Since many of the inputs in continuous NPS pollution models are highly variable spatially and temporally, procedures used for model evaluation should explicitly account for this variability. It is necessary to ensure, as much as possible, that model evaluation studies focus on the model structure and algorithms and are not overwhelmed by uncertainty in inputs due to natural variability and information uncertainty.

RECOMMENDATIONS FOR FUTURE WORK

There is clearly a need for further research before current NPS pollution models can be reliably used to assess impacts and evaluate alternative land management plans. Recommendations for future research work are given here. Some of the recommendations are extensions of work carried out in the current study, while others are related areas of research which would strengthen the basis on which protocols for probabilistic NPS pollution modeling could be developed.

Multivariate approaches to assessing model output

In the simulation study in Chapter III, model output variables were analyzed as univariate data. However, since all outputs are obtained for the same management scenario(s), there will generally be some dependence or correlation structure among the outputs, which indicates that the use of multivariate statistical techniques should be investigated. There have been few, if any, attempts to analyze model output using multivariate techniques. Principal component analysis could be used to obtain uncorrelated linear functions of the output variables, and physically meaningful components extracted. For example, a component corresponding to chemical leaching potential could be obtained, and another component obtained corresponding to surface chemical losses. This would allow a more holistic analysis of nonpoint loading impacts from alternative management scenarios.

Determining the global sensitivity of input parameters

A drawback of direct Monte Carlo simulation is that it cannot be used to obtain information on the relative contributions of input variances on model output variance. This information is extremely

important from the point of view of determining which inputs and correlations a model is most sensitive to, which may then be used to establish required accuracy and precision in defining model inputs. In previous studies it has been assumed that this information may be gleaned by partitioning the model output variance into relative contributions from input variances and correlations among inputs through a first-order error analysis (Song and Brown, 1990; Smith and Charbeneau, 1990). However, if the model response is highly non-linear or if input variances are very large, the results from a first-order error analysis may not be reliable. There is a need to develop a measure of global sensitivity of model outputs over the expected range of input parameters. This global measure rather than the local sensitivity of outputs at some base value of inputs would be a better indicator of which model inputs and correlations are the most important from the point of view of Monte Carlo simulation. One possible measure of global sensitivity could be obtained by a canonical correlation analysis using output from a Monte Carlo simulation (Kendall, 1975). Canonical correlation analysis is a multivariate statistical procedure concerned with determining the interdependence between two groups of variables.

Incorporation of spatial structure in field-scale models

The approach used in this dissertation to model field-scale variability of soil and chemical parameters is similar to the stream tube modeling concept for sub-surface solute transport described by Jury and Roth (1990). In this approach, the field is assumed to consist of stream tubes, which do not interact and each of which is described by a common set of transport laws that may have different parameter values. Implicit in this approach is the assumption that model parameters assumed to be random variables in the field have no spatial correlation, and can therefore be defined completely by a joint probability density function. However, many model parameters, particularly hydraulic properties may exhibit spatial correlation, and therefore require description as a regionalized variable employing a variogram or correlogram (Kitanidis, 1993). This could potentially result in a reduction in the variance of model outputs, and also provide

additional information as the model output may be mapped and possibly used to guide chemical application (Addiscott, 1993). There is a need to incorporate the spatial structure of input parameters in current NPS models and determine whether there is a significant reduction in output variance compared to using the stream tube approach. In practice, the main problem in incorporating spatial structure in any modeling study is the lack of data for estimating the variogram.

Assessment of curve number methodology in continuous models

In this dissertation, a conceptual approach to determining the potential retention or curve number (CN) probability distribution resulting from factors other than antecedent moisture condition was developed. Since continuous hydrologic models that use the SCS runoff equation adjust the value of CN based on current soil moisture conditions, CN distributions for continuous hydrologic models should ideally reflect variability only from sources other than antecedent moisture content. However, the current consensus on CN variation appears to be to lump all sources of variability and not attempt to explain how much of the variation can be attributed to individual factors such as antecedent moisture.

There is a need to further evaluate the implementation of curve number methodology in continuous models, and determine if there is a consistent improvement in runoff predictions compared to the original SCS runoff equation. To the author's knowledge, there has been no independent evaluation of the procedure since it was originally suggested by Williams and LaSeur (1976). The procedure developed in this dissertation for obtaining the CN probability distribution resulting from factors other than antecedent moisture may be used in a probabilistic evaluation procedure. A comparison could be made between runoff predictions using the original form of the SCS runoff equation using the CN distribution resulting from all sources of variation

and runoff predictions from a continuous model using the CN distribution from factors other than antecedent moisture.

Improvement of regression equations

The Durbin-Watson statistics for the regression equations developed in this dissertation for predicting field-scale variability in field capacity, wilting point, and saturated conductivity indicated that there may be autocorrelation present in the response variables. If methods for handling autocorrelation were incorporated in the regression analysis, an improvement in the quality of prediction from the regression equations could potentially be realized. Myers (1990) has suggested using lagged values of the response variable or transforming both response and regressor variables as possible methods for handling autocorrelation.

REFERENCES

- Abbot, M. B., J. C. Bathurst, J. A. Cunge, P. E. O'Connell, and J. Rasmussen. 1986. An introduction to the European Hydrological System - Systeme Hydrologique European, "SHE", 1. History and philosophy of a physically-based, distributed modelling system. *J. of Hydrology* 87:45-59.
- Addiscott, T. M., and R. J. Wagenet. 1985. Concepts of solute leaching in soils: a review of modelling approaches. *J. of Soil Science* 36:411-424.
- Addiscott, T. M. 1993. Simulation modelling and soil behaviour. *Geoderma* 60:15-40.
- Anderson, M. P., and W. W. Woessner. 1992. Applied groundwater modeling - simulation of flow and advective transport. Academic Press, Inc., San Diego, California.
- Benjamin, J. R., and Cornell, C. A. 1970. Probability, statistics, and decision for civil engineers. McGraw Hill, New York, N. Y.
- Beven, K. J. 1989. Changing ideas in hydrology. The case of physically based models. *J. of Hydrology* 105:157-172.
- Beven, K., and A. Binley. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* 6:279-298.
- Bonsu, M. 1992. A study of a texture-based equation for estimating the saturated hydraulic conductivity of an alfisol in the sudan savannah ecological zone, Ghana. *Hydrological Sciences* 37:599-606
- Bouzhara, A., P. G. Lakshminarayan, A. Carriquiry, P. W. Gassman, and J. F. Shogren. 1993. Metamodels and nonpoint policy in Agriculture. *Water Resource Research* 29(6):1579-1587.
- Brady, N. C. 1990. The nature and properties of Soils. Macmillan Publishing Company.
- Brooks, R. H., and A. T. Corey. 1964. Hydraulic properties of porous media. Hydrology Paper 3. Colorado State University: 27pp. Fort Collins, Colorado.
- Brown, B. W., and J. Lovato. 1991. RANLIB. Library of FORTRAN routines for random number generation. Dept. of Biomathematics. Univ. of Texas, M. D. Anderson Cancer Center, Houston, Texas.
- Carsel, R. F., L. A. Mulkey, M. N. Lorber, and L. B. Baskin. 1985. The Pesticide Root Zone Model (PRZM): a procedure for evaluating leaching threats to groundwater. *Ecological Modeling* 30: 49-69.
- Carsel, R. F., and R. S. Parrish. 1988. Developing joint probability distributions of soil water retention properties. *Water Resources Research* 24(5):755-769.
- Carsel, R. F., R. S. Parrish, R. L. Jones, J. L. Hansen, R. L. Lamb. 1988a. Characterizing the uncertainty of pesticide leaching in agricultural soils. *J. of Contaminant Hydrology* 2:11-124.

- Carsel, R. F., R. L. Jones, J. L. Hansen, R. L. Lamb, and M. P. Anderson. 1988b. A simulation procedure for groundwater quality assessments of pesticides. *J. of Contaminant Hydr* 2:125-138.
- Clemmens, A. J. 1992. Bayesian inference for feedback control. I: Theory. *J. of Irrigation and Drainage Engineering* 118(3):397-415.
- Cooke, R. A., S. Mostaghimi and F. E. Woeste. 1993. VTFIT: A microcomputer-based routine for fitting probability distribution functions to data. *Applied Engineering in Agriculture* 9(4):401-408.
- Davis, P. E., and C. D. Heatwole. 1990. Modeling impacts of alternative agricultural systems on groundwater in the Virginia Coastal Plain. ASAE Paper No. 90-2592. St. Joseph, MI: ASAE.
- Dean, J. D., P. S. Huyakorn, A. S. Donigian Jr., K. A. Voos, R. W. Schanz, Y. J. Meeks, and R. F. Carsel. 1989. Risk of Unsaturated/Saturated Transport and Transformation of Chemical Concentrations (RUSTIC): Volume I: Theory and code verification. EPA/600/3-89/048a. USEPA, Environmental Research Laboratory, Athens, Georgia.
- D'Elia, C. F., J. G. Sanders, and D. G. Capone. 1989. Analytical chemistry for environmental sciences: a question of confidence. *Environ. Sci. Technol.* 23(7):768-774.
- Destouni, G. 1993. Stochastic modelling of solute flux in the unsaturated zone at the field scale. *J. of Hydrology* 143:45-61.
- Dettinger, M. D., and J. L. Wilson. 1981. First-order analysis of uncertainty in numerical models of groundwater flow. Part 1: Mathematical development. *Water Resources Research* 17(1):149-161.
- Diebel, P. L. 1990. An economic analysis of low-input agriculture as a groundwater management strategy. Unpublished Ph.D. dissertation. Department of Agricultural Economics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.
- Duttweiler, D. W., and H. P. Nicolson. 1983. Environmental problems and issues of agricultural nonpoint source pollution. In *Agricultural management and water quality*, ed. F. W. Schaller and G. W. Bailey., 281-304. Ames: Iowa State University Press.
- Engman, E. T. 1986. Roughness coefficients for routing surface runoff. *J. of Irrigation and Drainage Engineering* 112(1):39-53.
- EPA. 1987. The Enhanced Stream Water Quality Models QUAL2E and QUAL2E-UNCAS: Documentation and User Model. EPA/600/3-87/007. USEPA, Environmental Research Laboratory, Athens, Georgia.
- Gassman, P. W., and C. Anderson. 1989. Long term leaching simulations using PRZM. ASAE Paper No. 89-2073. St. Joseph, MI: ASAE.
- Gibbons, J. D. 1985. Nonparametric methods for quantitative analysis. American Sciences Press, Columbus, Ohio.
- Gutjahr, A. 1988. Problem definition, variability, geostatistics and sampling in ground water contamination. Proceedings of a national workshop on ground water quality, Arlington, Virginia.

- Haan, C. T. 1989. Parametric uncertainty in hydrologic modeling. *Transactions of the ASAE* 32(1):137-146.
- Haan, C. T., J. B. Solie and B. N. Wilson. 1990. To tell the truth - hydrologic models in court. In *Transferring models to users*, ed. E. B. Janes and W. R. Hotchkiss, 337-348. Denver, CO: American Water Resources Association.
- Haan, C. T., B. Allred, D. E. Storm, G. Sabbagh, and S. Prabhu. 1993. Evaluation of hydrologic/water quality models - A statistical procedure. ASAE Paper No. 93-2505. St. Joseph, MI: ASAE.
- Han, M. B., D. A. Bender, and S. E. Taylor. 1991. Computer generation of highly skewed correlated random variables. *Transactions of the ASAE* 34(5):2279-2281.
- Hauser, V. L. and O. R. Jones. 1991. SCS curve numbers from short runoff records. ASAE Paper No. 91-2614. St. Joseph, MI: ASAE.
- Hawkins, R. H., A. T. Hjelmfelt Jr. and A. W. Zevenbergen. 1985. Runoff probability, storm depth and curve numbers. *Journal of Irrigation and Drainage Engineering (ASCE)* 111(4):330-340.
- Hern, S. C., S. M. Melancon, and J. E. Pollard. 1986. Generic steps in the field validation of vadose zone fate and transport models. In *Vadose zone modeling of organic pollutants*, ed. S. C. Hern and S. M. Melancon, 61-80. Chelsea, Michigan: Lewis Publishers.
- Heatwole, C. D., P. L. Diebel, J. M. Halstead. 1990. Management and policy effects on potential groundwater contamination from dairy waste. *Water Resources Bulletin* 26(1):25-34.
- Heatwole, C. D., T. A. Dillaha, and S. Mostaghimi. 1991. Agricultural BMPs applicable to Virginia. Bulletin 169, Blacksburg, VA: Virginia Water Resources Research Center
- Heatwole, C. D., S. Zacharias, T. A. Dillaha and S. Mostaghimi. 1992. Fate and transport of pesticides in a Virginia Coastal Plain soil. Bulletin 175, Blacksburg, VA: Virginia Water Resources Research Center.
- Hjelmfelt, A. T., Jr., L. A. Kramer and R. E. Burwell. 1982. Curve numbers as random variables. In *Proc. Rainfall-runoff Relationship*, ed. V. P. Singh. 365-370. Littleton, CO: Water Resources Publ.
- Hjelmfelt, A. T. 1994. Personal Communication. Research Hydraulic Engineer. ARS, USDA, Columbia, Missouri.
- Hockman, J. R., C. F. Neal, D. L. Racey, and D. F. Wagner. 1982. Soil Survey of Rockingham County, Virginia. United States Department of Agriculture, Soil Conservation Service and Forest Service.
- Hollander M., and D. A. Wolfe. 1973. Nonparametric statistical methods. John Wiley & Sons, New York.
- Hu, H., and I. Olkin. 1991. A numerical procedure for finding the positive definite matrix closest to a patterned matrix. *Statistics and probability letters* 12:511-515.
- IMSL. 1987. Stat/Math/Sfun Libraries. Fortran subroutines for statistical analysis, math, and special functions. *International Mathematical and Statistical Library*, Houston, Texas.

- Ince, P. J., and J. Buongiorno. 1991. Multivariate stochastic simulation with subjective multivariate normal distributions. In: Proceedings of the 1991 symposium on systems analysis in forest resources; March 3-6, Charleston, South Carolina: U. S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station, 1991:143-150.
- Johanson, R. C., J. C. Imhoff, H. H. Davis, J. L. Kittle, and A. S. Donigian. 1981. User's manual for hydrologic simulation program-Fortran (HSPF): Release 7.0. U. S. Environmental Protection Agency, Athens, Georgia.
- Jury, W. A., and K. Roth. 1990. Transfer functions and solute movement through soil. Theory and Applications. Birkhauser Verlag Basel.
- Jury, W. A., J. S. Dyson, and G. L. Butters. 1990. A transfer function model of field scale solute transport under transient water flow. *Soil Sci. Soc. Am J.* 54:327-332.
- Jury, W. A., W. R. Gardner, and W. H. Gardner. 1991. Soil Physics. 5th edition. John Wiley and Sons.
- Kendall, M. 1975. Multivariate analysis. Charles Griffin and Company.
- Khan, M. A., and T. Liang. 1989. Mapping pesticide contamination potential. *Environmental Management* 13(2):233-242.
- Kitanidis, P. K. 1993. Geostatistics. In *Handbook of Hydrology*, ed. D. R. Maidment, 20.1-20.39. McGraw Hill.
- Knisel, W. G., F. M. Davis, and R. A. Leonard. 1993. GLEAMS version 2.0. Part III - Users Manual.
- Konikow, L. F., and J. D. Bredehoeft. 1992. Ground-water models cannot be validated. *Advances in Water Resources* 15:75-83.
- Kool, J. B., K. A. Albrecht, J. C. Parker, and J. C. Baker. 1986. Physical and chemical characterization of the Groseclose soil mapping unit. Bulletin 86-4, Blacksburg, VA: Virginia Agricultural Experiment Station.
- Kumar, D, C. D. Heatwole, and F. E. Woeste. 1992. Towards a protocol for BMP assessment using field-scale NPS pollution models. ASAE Paper No. 92-2635. St. Joseph, MI: ASAE.
- Law, A. M., and W. D. Kelton. 1991. Simulation modeling and analysis. McGraw-Hill, Inc., New York.
- Leonard, R. A., W. G. Knisel, and D. A. Still. 1987. GLEAMS: Groundwater Loading Effects of Agricultural Management Systems. *Trans. of the ASAE* 30(5):1403-1418.
- Leonard, R. A., and W. G. Knisel. 1990. Can pesticide transport models be validated using field data: Now and in the future ? Departmental Publication Number 3, Agricultural Engineering Department, University of Georgia, Coastal Plain Experiment Station, Tifton, Georgia.
- Li, W. 1993. A subsurface water quality evaluation sytem for assessing NPS pollution potential by pesticides. Unpublished Ph.D. dissertation. Department of Agricultural Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

- Loague, K. M., R. S. Yost, R. E. Green, and T. C. Liang. 1989. Uncertainty in a pesticide leaching assessment for Hawaii. *J. of Contaminant Hydrology*. 4:139-161.
- Loague, K. M., R. E. Green, T. W. Giambelluca, T. C. Liang, and R. S. Yost,. 1990. Impact of uncertainty in soil, climatic, and chemical information in a pesticide leaching assessment. *J. of Contaminant Hydrology*. 5:171-194.
- Loague, K. M., and R. E. Green. 1991. Statistical and graphical methods for evaluating solute transport models: Overview and Application. *J. of Contaminant Hydrology*. 7:51-73.
- Loague, K. 1992. Using soil texture to estimate saturated hydraulic conductivity and the impact on rainfall-runoff simulations. *Water Resources Bulletin* 28(4):687-693.
- Luis, S. J., and D. McLaughlin. 1992. A stochastic approach to model validation. *Advances in Water Resources* 15:15-32.
- Marquardt, D. W. 1980. You should standardize the predictor variables in your regression models. *J. of the American Statistical Association* 75:87-91.
- Market Engineering Co. 1992. Crystal Ball: A forecasting and risk management program. Market Engineering Co., Denver.
- Menelik, G., R. B. Reneau, Jr., D. C. Martens, T. W. Simpson, and G. W. Hawkins. 1990. Effects of tillage and nitrogen fertilization on nitrogen losses from soils used for corn production. Bulletin 167. Virginia Water Resources Research Center, Blacksburg, Va. 167 pp.
- Microsoft. 1991. Microsoft FORTRAN Version 5.1. Users reference. Redmond, WA: Microsoft Corporation.
- Microsoft. 1992. Microsoft Excel Version 4.0 Users Guide. Redmond, WA: Microsoft Corporation.
- Mills, W. C., A. W. Thomas, A. L. Dillard and W. M. Snyder. 1992. Computing unbiased retention parameter probabilities from rainfall and runoff data. *Transactions of the ASAE* 35(3):891-897.
- MINITAB. 1993. Minitab Reference Manual, Release 9 for Windows. State College, PA: Minitab, Inc.
- Mostaghimi, S., U. S. Tim, P. W. McClellan, J. C. Carr, R. K. Byler, T. A. Dillaha, V. O. Shanholtz, and J. R. Pratt. 1989. Watershed/Water quality monitoring for evaluating BMP effectiveness - Nomini Creek watershed: Pre-BMP evaluation. Final Report No. N-P1-8096. Department of Agricultural Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Va.
- Myers, R. H. 1990. Classical and modern regression with applications. PWS-KENT Publishing Company.
- Notziger, D. L., and A. G. Hornsby. 1987. CMLS: Interactive simulation of chemical movement in layered soils. Inst. Food Agric. Sci., Univ. of Florida Circular 780, Gainesville, Florida.
- Novotny, V., and G. Chesters. 1981. Handbook of nonpoint source pollution: Sources and management. New York:Van Nostrand Reinhold Co.

- Oren, S. S. 1981. A conjugate directions method for subjective assessment of normal distributions. *Journal of Optimization Theory and Applications* 33(1):25-36.
- Parrish, R. S., and C. N. Smith. 1990. A method for testing whether model predictions fall within a prescribed factor of true values, with an application to pesticide leaching. *Ecological Modeling* 51:59-72.
- Pennell, K. D., A. G. Hornsby, R. E. Jessup, and P.S. C. Rao. 1990. Evaluation of five simulation models for predicting Aldicarb and Bromide behavior under field conditions. *Water Resources Research* 26(11):2679-2693.
- Phillips, D. L., P. D. Hardin, V. W. Benson, and J. V. Baglio. 1993. Nonpoint source pollution impacts of alternative agricultural management practices in Illinois: A simulation study. *J. of Soil and Water Conservation* 48(5):449-457.
- Ponce, V. M. 1989. Engineering hydrology - principles and practices. Prentice Hall.
- Ragab, R., and J. D. Cooper. 1993. Variability of unsaturated zone water transport parameters: implications for hydrological modelling. 2. Predicted vs. in situ measurements and evaluation of methods. *J. of Hydrology* 148:133-147.
- Rawls, W. J., L. R. Ahuja, D. L. Brakensiek, and A. Shirmohammadi. 1993. Infiltration and soil water movement. In *Handbook of Hydrology*, ed. D. R. Maidment. 5.1-5.51. McGraw Hill, Inc.
- Rawls, W. J., D. L. Brakensiek, and K. E. Saxton. 1982. Estimation of soil water properties. *Transactions of the ASAE* 25(5):1316-1220 and 1328.
- Rawls, W. J., and D. L. Brakensiek. 1985. Prediction of soil water properties for hydrologic modeling. In: *Proceedings of Symposium on Watershed Management*. American Society of Civil Engineers, New York, pp 293-299.
- Rawls, W.J. 1980. Soil moisture retention database documentation (unpublished). USDA, Beltsville, Md.
- SAS. 1988. SAS/STAT Users Guide, Release 6.03. SAS Institute, Cary, North Carolina.
- Sen, A. and M. Srivastava. 1990. Regression Analysis - Theory, Methods, and Applications. Springer-Verlag.
- Scheuer, E. M., and D. S. Stoller. 1962. On the generation of normal random vectors. *Technometrics* 4:278-281.
- SCS. 1982. *Soil Survey of Westmoreland County Virginia*. Richmond, Va: USDA-SCS.
- Smith, C. N., R. A. Leonard, G. W. Langdale, and G. W. Bailey. 1978. Transport of agricultural chemicals from small upland Piedmont watersheds. EPA/600/3-78-056. USEPA, Environmental Research Laboratory, Athens, Georgia.
- Smith, M. C., A. B. Bottcher, K. L. Campbell, and D. L. Thomas. 1991. Field testing and comparison of the PRZM and GLEAMS models. *Trans. of the ASAE* 34(3):838-847
- Smith, R. E., and V. A. Ferreira. 1989. Comparative evaluation of unsaturated flow methods in selected USDA simulation models. In H. J. Morel-Seytoux, ed., *Unsaturated flow in hydrologic*

- modeling and practice, pp 391-412. NATO, ASI Series, C, vol.275, Dordrecht, The Netherlands.
- Smith, R. E., and J. R. Williams. 1980. Simulation of the surface water hydrology, chapter 2. IN CREAMS, A field scale model for chemicals, runoff, and erosion from agricultural management systems, Volume I, pp. 13-35. USDA, Conservation Research Report No. 26.
- Smith, R. E. 1992. Opus, An integrated simulation model for transport of nonpoint source pollutants at the field scale: Volume I, Documentation, U. S. Department of Agriculture, ARS-98, 120 pp.
- Smith, V. J., and R. J. Charbeneau. 1990. Probabilistic soil contamination exposure assessment procedures. *J. of Environmental Engineering* 116(6):1143-1163.
- Song, Q., and L. C. Brown. 1990. DO model uncertainty with correlated inputs. *J. of Environmental Engineering*. 116(6):1164-1180.
- Spooner, J., R. P. Maas, S. A. Dressing, M. D. Smolen, and F. J. Humenik. 1985. Appropriate designs for documenting water quality improvements from agricultural NPS pollution control. In: *Perspectives on nonpoint source pollution: proceedings of a national conference*. Kansas City, Missouri.
- Spooner, J., S. L. Brichford, R. P. Maas, M. D. Smolen, D. D. Dickey, G. Ritter, and E. Flagg. 1988. Determining the statistical sensitivity of the water quality monitoring program in the Taylor Creek-Nubbin Slough, Florida project. *J. Lake and Reservoir Management* 4(2):113-124.
- Spooner, J., R. L. Huffman, D. E. Line, J. A. Gale, G. D. Jennings, S. W. Coffey, J. A. Arnold, L. Wyatt, D. L. Osmond, and S. K. Haseeb. 1992. Fate and effects of pollutants - Nonpoint sources. *Water Environment Research* 64(4):503-514.
- Stephens, M. A. 1974. EDF statistics for goodness of fit and some comparisons. *J. of the American Statistical Association*. 69(347):730-737.
- Taylor, S. E. and D. A. Bender. 1989. A method for simulating multiple correlated lumber properties. *Forest Products Journal* 39(7/8):71-74.
- Thomas, A. W., W. M. Snyder, and G. W. Langdale. 1988. Stochastic impacts on farming: V. Risk adjustment through conservation planning. *Trans. of the ASAE* 31(5):1368-1374.
- Thomas, D. L., M. C. Smith, R. A. Leonard, and F. J. K. daSilva. 1990. Simulated effects of rapeseed production alternatives on pollution potential in the Georgia Coastal Plain. *J. of Soil and Water Conservation* 45(1):148-154.
- VCE, 1992. Pest management guide for field crops. Virginia Cooperative Extension, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.
- Wagenet, R. J., and P. S. C. Rao. 1990. Modeling pesticide fate in soils. In: *Pesticides in the soil environment: Processes, Impacts, and Modeling* ed. H. H. Cheng. Soil Science Society of America, Madison, Wisconsin.

- Wauchope, R. D., T. M. Buttler, A. G. Hornsby, P. W. M. Augustijn-Beckers, and J. P. Burt. 1992. The SCS/ARS/CES pesticide properties database for environmental decision-making. *Environmental Contaminant Toxicology Reviews* 123:1-164.
- Weisberg, S. 1980. Applied linear regression. John Wiley and Sons.
- Weaver, R. D., and J. K. Harper. 1993. Analyzing water quality policy using microeconomic models of production practices and biophysical flow models of environmental processes. *Resource Management and Optimization* 9(2):95-105.
- Wischmeier, W. H., and D. D. Smith. 1978. Predicting rainfall erosion losses - a guide to conservation planning. United States Department of Agriculture Handbook Number 537.
- Williams, J. R. and W. V. LaSeur. 1976. Water yield model using SCS curve numbers. *J. of the Hydraulics Division ASCE* 102(HY9):1241-1253.
- Williams, J. R., C. A. Jones and P. T. Dyke. 1984. A modeling approach to determining the relationship between erosion and soil productivity, *Transactions of the ASAE* 27(1):129-144.
- Yoo, K. H., K. S. Yoon and J. M. Soileau. 1993. Runoff curve numbers determined by three methods under conventional and conservation tillages. *Transactions of the ASAE* 36(1):57-62.
- Young, R. A., C. A. Onstad, D. D. Bosch, and W. P. Anderson. 1989. AGNPS: a nonpoint source pollution model for evaluating agricultural watersheds. *J. Soil and Water Conserv.* 44(2):168-173.
- Zacharias, S. 1992. Personal communication. Graduate Research Assistant, Department of Agricultural Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.
- Zacharias, S., and C. D. Heatwole. 1993a. Comparison of quantitative techniques used for pesticide model validation. ASAE Paper No. 93-2506. St. Joseph, MI: ASAE.
- Zacharias, S., and C. D. Heatwole. 1993b. Predicting tillage treatment effects on pesticide transport: a validation study. ASAE Paper No. 93-2592. St. Joseph, MI: ASAE.
- Zhang, H., C. T. Haan, and D. L. Nofziger. 1993. An approach to estimating uncertainties in modeling transport of solutes through soils. *J. of Contaminant Hydrology* 12:35-50.
- Zhu, M. 1994. Personal Communication. Research Associate, Department of Agricultural and Applied Economics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Appendix A

General purpose Monte-Carlo pre-processor module

```

C-----
This program allows the generation of correlated random variates from the following
distributions: Uniform, Exponential, Normal, Lognormal, Weibull, Gumbel, Gamma,
Beta, and Triangular.
The Taylor-Bender(1989) method is used to generate correlated variates.
Routines for generating U(0,1) numbers, multivariate normal variates, and for
numerical evaluation of the normal, Gamma, and Beta distributions were obtained
from Brown and Lovato (1991).
Author: Dipmani Kumar
Last modified: 10/10/1994
C-----
parameter (nr=100)
implicit none
real*8 cov(nr,nr),meanv(nr),parm((nr*(nr+3)/2)+1),x(nr),work(nr),
&parmmat(nr,5)
integer nrv,nt,idismat(nr,2)
character*80 fileout
open (unit=1,file='mcpp.inp',status='old')
open (unit=2,file='mcpp.aux',status='unknown')
open (unit=3,file='mcpp.001',status='unknown')
read(1,*)
read(1, '(a)') fileout
open (unit=4,file=fileout,status='unknown')
C-----
C nrv=number of variables,nt=number of trials
C-----
read(1,*)
read(1,*) nrv,nt
call auxmain(nrv,cov,meanv,parm,x,work,nt,idismat,parmmat)
write(*,*) ' Normal program termination.'
write(*,1000) fileout
1000 format(' Output file: ',a20)
stop ' '
end
C-----
C Auxmain ensures arrays are correctly dimensioned when passed to subroutines
C-----
subroutine auxmain(nrv,cov,meanv,parm,x,work,nt,idismat,parmmat)
implicit none
integer i,j,nrv,nt,istream,idismat(nrv,2),icsstat,icorrel
real*8 cov(nrv,nrv),meanv(nr),parm((nrv*(nrv+3)/2)+1),x(nrv),
&work(nrv),parmmat(nrv,5),tm,tvar
external setgmn,genmn,cumnor,invnor,ranf
real*8 cumnor,invnor,ranf
C-----
C Random number stream:(1-32)
C-----
read(1,*)
read(1,*) istream,icorrel
if(icorrel.ne.0) then
read(1,*)
C-----
C Correlation matrix for N(0,1)
C-----
do 10,i=1,nrv
10 read(1,*)(cov(i,j),j=1,i)
do 20,i=1,nrv
20 cov(i,i)=1.0D0
call fillsym(cov,nrv)
write(2,*) '*** Inputs read in ***'
write(2,*)
write(2,*) 'Correlation matrix: '
write(2,*) '-----'
call wrmat(2,cov,nrv)
do 30,i=1,nrv
30 meanv(i)=0d0
call setcgn(istream)
call setgmn(meanv,cov,nrv,parm)
do 50,i=1,nt
write(*,1001) i,nt
1001 format('+ Generating correlated U(0,1) vectors: ',i5,' of ',i5)
call genmn(parm,x,work)
do 35,j=1,nrv
35 x(j)=cumnor(x(j))
50 write(3,*) x
write(*,*)
else
write(2,*) '*** Inputs read in ***'
write(2,*)
write(2,*) 'Variates specified to be uncorrelated. '

```

```

write(2,*)
call setcgn(istream)
do 70,i=1,nt
write(*,1002)i,nt
1002 format('+ Generating U(0,1) vectors: ',i5,' of ',i5)
do 60,j=1,nrv
60 x(j)=ranf()
70 write (3,*)x
write(*,*)
endif
100 format(30(e10.4,1x))
rewind(3)
call generate(nrv,nt,work,parmmat,idismat)
rewind(4)
read(1,*)
read(1,*)icsstat
if(icsstat.ne.0)then
call wsumstat(nrv,idismat,parmmat)
call chstat(x,meanv,work,cov,nrv,nt,icorrel)
write(2,*)
write(2,*)'*** Summary statistics for generated variates ***'
write(2,*)
write(2,1000)
write(2,1200)
write(2,1500)
do 200,i=1,nrv
read(5,*)tm,tvar
200 write(2,2000)i,tm,meanv(i),tvar,work(i)**2
write(2,2200)
write(2,2500)
write(2,3000)
write(2,3200)
c close(unit=5,status='delete')
if(icorrel.ne.0)then
write(2,*)
write(2,*)'Correlation matrix for generated variates: '
write(2,*)'-----'
call wrmat(2,cov,nrv)
endif
endif
1000 format(' Expected Actual Expected Actual')
1200 format(' Variable mean* mean variance* variance')
1500 format(' -----')
2000 format(i2,9x,e9.3,1x,e9.3,1x,e9.3,3x,e9.3)
2200 format(' -----')
2500 format(' *Expected mean and variance are without truncation.')
3000 format(' With truncation, the expected mean will be different')
3200 format(' and the expected variance will be smaller.')
return
end
subroutine wrmat(iod,mat,n)
implicit none
integer iod,i,j,n
real*8 mat(n,n)
do 10,i=1,n
10 write(iod,100)(mat(i,j),j=1,n)
100 format(30(f4.2,1x))
return
end
subroutine fillsym(mat,n)
implicit none
integer i,j,n
real*8 mat(n,n)
do 300,i=2,n
do 300,j=1,i-1
300 mat(j,i)=mat(i,j)
return
end
subroutine generate(nrv,nt,work,parmmat,idismat)
implicit none
integer nrv,nt,inparm(9),idismat(nrv,2),i,j,idis,itr,nvsd,iadd
integer i1,j1,k,idist,intrunc
real*8 work(nrv),parmmat(nrv,5),val,parminp(5)
data inparm/2,2,2,3,3,2,3,4,3/
read(1,*)
do 100, i=1,nrv
read (1,*)idis,itr
iadd=itr
if(itr.gt.1)iadd=itr-1
nvsd=inparm(idis)+iadd
backspace(1)

```

```

        read(1,*) (idismat(i,j),j=1,2), (parmmat(i,j),j=1,nvsvd)
c      write(*,*) idismat(i,1), idismat(i,2)
100    continue
        call wrdis(nrv, idismat, parmmat, inparm)
C-----
C      Start generating random variates
C-----
        do 200, i1=1, nt
write(*,1200) i1, nt
1200  format('+ Generating vector ', i5, ' of ', i5, ' vectors.')
        read(3,*) (work(j1), j1=1, nrv)
        do 150, i=1, nrv
            idist=idismat(i,1)
            intrunc=idismat(i,2)
            val=work(i)
            do 125, k=1, 5
125    parminp(k)=parmmat(i,k)
C-----
C      Call appropriate subroutine for evaluating cdf
C-----
            if (idist .eq. 1) then
                call genunf(parminp, val)
            elseif (idist .eq. 2) then
                call genexpo(parminp, val, intrunc)
            elseif (idist .eq. 3) then
                call gennor(parminp, val, intrunc)
            elseif (idist .eq. 4) then
                call genlnor(parminp, val, intrunc)
            elseif (idist .eq. 5) then
                call genwei(parminp, val, intrunc)
            elseif (idist .eq. 6) then
                call gengumb(parminp, val, intrunc)
            elseif (idist .eq. 7) then
                call gengam(parminp, val, intrunc)
            elseif (idist .eq. 8) then
                call genbet(parminp, val)
            elseif (idist .eq. 9) then
                call gentriag(parminp, val)
            endif
        work(i)=val
150    continue
        write(4,5000) (work(k), k=1, nrv)
c      write(*,*) work
5000  format(30(e14.8,1x))
200    continue
        return
        end
        subroutine wrdis(nrv, idismat, parmmat, inparm)
implicit none
integer nrv, idismat(nrv,2), i, j, inparm(9), nval, iadd
real*8 parmmat(nrv,5)
character*15 name(9)
character*3 trflag
data name(1)/'Uniform'/
data name(2)/'Exponential'/
data name(3)/'Normal'/
data name(4)/'Lognormal'/
data name(5)/'Weibull'/
data name(6)/'Gumbel'/
data name(7)/'Gamma'/
data name(8)/'Beta'/
data name(9)/'Triangular'/
write(2,*)
write(2,1000)
write(2,2000)
1000  format('Variable', ' Distribution', ' Truncation*', ' Parameters/(Tr
&uncation points) ')
2000  format('-----', ' -----', ' -----', ' -----'
&-----')
        do 100, i=1, nrv
            if(((idismat(i,1)).lt.1).or.((idismat(i,1)).gt.9)) then
                write(*,1200) i
                stop 'No random variates generated. Program terminated'
            endif
            iadd=idismat(i,2)
            if(iadd.eq.0) trflag=' N '
            if(iadd.eq.1) trflag=' L '
            if(iadd.eq.2) trflag=' R '
            if(iadd.eq.3) trflag='L,R'
            if(idismat(i,2).gt.1) iadd=idismat(i,2)-1
            nval=inparm(idismat(i,1))+iadd

```

```

write(2,3000)i,name(idismat(i,1)),trflag,(parmmat(i,j),j=1,nval)
100 continue
write(2,4000)
write(2,5000)
1200 format(' Distribution ID for variable ',i2,' is invalid.')
3000 format(i2,8x,a12,5x,a3,6x,5(f6.2,1x))
4000 format('-----')
&-----')
5000 format('*N=No truncation,L=Left truncated,R=Right truncated')
return
end
subroutine genunf(parminp,val)
implicit none
real*8 parminp(5),val,a,b,c
a=parminp(1)
b=parminp(2)
c=b-a
val=a+(val*c)
return
end
subroutine genexpo(parminp,val,itrunc)
implicit none
real*8 parminp(5),val,beta,gama,fa,fb,fd
integer itrunc
beta=parminp(1)
gama=parminp(2)
if(itrunc .ne. 0)then
if (itrunc .eq. 1)then
fa=(1d0-dexp(-((parminp(3)-gama)/beta)))
fb=1
endif
if (itrunc .eq. 2)then
fb=(1d0-dexp(-((parminp(3)-gama)/beta)))
fa=0
endif
if (itrunc .eq. 3)then
fa=(1d0-dexp(-((parminp(3)-gama)/beta)))
fb=(1d0-dexp(-((parminp(4)-gama)/beta)))
endif
fd=fb-fa
endif
if (itrunc .ne. 0)val=fa+(fd*val)
val=(-beta*dlog(1d0-val))+gama
return
end
subroutine gennor(parminp,val,itrunc)
implicit none
real*8 parminp(5),val,mu,sigma,fa,fb,fd,cumnor,invnor
external cumnor,invnor
integer itrunc
mu=parminp(1)
sigma=parminp(2)
if(itrunc .ne. 0)then
if (itrunc .eq. 1)then
fa=(parminp(3)-mu)/sigma
fa=(cumnor(fa))
fb=1
endif
if (itrunc .eq. 2)then
fb=(parminp(3)-mu)/sigma
fb=(cumnor(fb))
fa=0
endif
if (itrunc .eq. 3)then
fa=(parminp(3)-mu)/sigma
fb=(parminp(4)-mu)/sigma
fa=(cumnor(fa))
fb=(cumnor(fb))
endif
fd=fb-fa
endif
if (itrunc .ne. 0)val=fa+(fd*val)
val=invnor(val)
val=(val*sigma)+mu
return
end
subroutine genlnor(parminp,val,itrunc)
implicit none
real*8 parminp(5),val,mu,sigma,fa,fb,fd,cumnor,invnor,gama
integer itrunc
external cumnor,invnor

```

```

mu=parminp(1)
sigma=parminp(2)
gama=parminp(3)
if (itrunc .ne. 0) then
  if (itrunc .eq. 1) then
    parminp(4)=dlog(parminp(4)-gama)
    fa=(parminp(4)-mu)/sigma
    fa=(cumnor(fa))
    fb=1
  endif
  if (itrunc .eq. 2) then
    parminp(4)=dlog(parminp(4)-gama)
    fb=(parminp(4)-mu)/gama
    fb=(cumnor(fb))
    fa=0
  endif
  if (itrunc .eq. 3) then
    parminp(4)=dlog(parminp(4)-gama)
    parminp(5)=dlog(parminp(5)-gama)
    fa=(parminp(4)-mu)/sigma
    fb=(parminp(5)-mu)/sigma
    fa=(cumnor(fa))
    fb=(cumnor(fb))
  endif
  fd=fb-fa
endif
if (itrunc .ne. 0) val=fa+(fd*val)
val=invnor(val)
val=(val*sigma)+mu
val=dexp(val)+gama
return
end
subroutine genwei(parminp,val,itrunc)
implicit none
real*8 parminp(5),val,alpha,beta,gama,fa,fb,fd
integer itrunc
alpha=parminp(1)
beta=parminp(2)
gama=parminp(3)
if (itrunc .ne. 0) then
  if (itrunc .eq. 1) then
    fa=1d0-dexp(-(((parminp(4))-gama)/beta)**alpha))
    fb=1
  endif
  if (itrunc .eq. 2) then
    fa=0
    fb=1d0-dexp(-(((parminp(4))-gama)/beta)**alpha))
  endif
  if (itrunc .eq. 3) then
    fa=1d0-dexp(-(((parminp(4))-gama)/beta)**alpha))
    fb=1d0-dexp(-(((parminp(5))-gama)/beta)**alpha))
  endif
  fd=fb-fa
endif
if (itrunc .ne. 0) val=fa+(fd*val)
val=gama+(beta*(-(dlog(1-val))))**(1/alpha))
return
end
subroutine gengumb(parminp,val,itrunc)
implicit none
real*8 val,parminp(5),beta,alpha,fa,fb,fd
integer itrunc
beta=parminp(1)
alpha=parminp(2)
if (itrunc .ne. 0) then
  if (itrunc .eq. 1) then
    fa=dexp(-dexp((-parminp(3)-beta)/alpha))
    fb=1
  endif
  if (itrunc .eq. 2) then
    fa=0
    fb=dexp(-dexp((-parminp(3)-beta)/alpha))
  endif
  if (itrunc .eq. 3) then
    fa=dexp(-dexp((-parminp(3)-beta)/alpha))
    fb=dexp(-dexp((-parminp(4)-beta)/alpha))
  endif
  fd=fb-fa
endif
if (itrunc .ne. 0) val=fa+(fd*val)
val=(-dlog(-dlog(val))*alpha)+beta

```

```

return
end
subroutine gengam(parminp,val,itrunc)
implicit none
real*8 parminp(5),alpha,beta,gama,fa,fb,fd,invchi,cumgam,val
external invchi,cumgam
integer itrunc
alpha=parminp(1)
beta=parminp(2)
gama=parminp(3)
if(itrunc .ne. 0)then
  if (itrunc .eq. 1)then
    fa=cumgam(((parminp(4)-gama)/beta),alpha)
    fb=1
  endif
  if (itrunc .eq. 2)then
    fb=cumgam(((parminp(4)-gama)/beta),alpha)
    fa=0
  endif
  if (itrunc .eq. 3)then
    fa=cumgam(((parminp(4)-gama)/beta),alpha)
    fb=cumgam(((parminp(5)-gama)/beta),alpha)
  endif
  fd=fb-fa
endif
if (itrunc .ne. 0)val=fa+(fd*val)
val=gama+(((invchi(val,2*alpha)/2d0)*beta)
return
end
subroutine genbet(parminp,val)
implicit none
real*8 parminp(5),val,a,b,c,alp1,alp2,invbet
external invbet
alp1=parminp(1)
alp2=parminp(2)
a=parminp(3)
b=parminp(4)
c=b-a
val=invbet(val,alp1,alp2)
val=a+(val*c)
return
end
subroutine gentriag(parminp,val)
implicit none
real*8 parminp(5),val,a,b,c1,c2,c3
a=parminp(1)
b=parminp(2)
c1=b-a
c2=parminp(3)
c3=((c2-a)/c1)
  if((val .le. c3))then
    val=dsqrt(c3*val)
  else
    val=1d0-(dsqrt((1d0-c3)*(1d0-val)))
  endif
val=a+(c1*val)
return
end
subroutine chstat(vec,mv,sdv,corm,nrv,nt,icorrel)
real*8 vec(nrv),mv(nrv),sdv(nrv),corm(nrv,nrv),am,sd,am1,am2,
&sd1,sd2,cor
write(*,*)
do 10,i=1,nrv
write(*,299)i
299 format('+ Computing mean/variance for variable ',i2,')
rewind(4)
call avsd(i,am,sd,vec,nrv,nt)
mv(i)=am
10 sdv(i)=sd
write(*,*)
if(icorrel.ne.0)then
corm(1,1)=1d0
do 20,i=2,nrv
corm(i,i)=1d0
do 20,j=1,i-1
write(*,301)i,j
301 format('+ Computing correlation for variables ',i2,') and ',i2,')
am1=mv(i)
am2=mv(j)
sd1=sdv(i)
sd2=sdv(j)

```

```

    call correl(i,j,am1,am2,sd1,sd2,vec,nrv,nt,cor)
20   corm(i,j)=cor
    call fillsym(corm,nrv)
    endif
    return
    end
    subroutine avsd(c,am,sd,vec,nrv,nt)
    real*8 vec(nrv),sd,am,var
    integer c
    am=0d0
    do 10,k=1,nt
    read(4,*)(vec(i),i=1,nrv)
10   am=am+vec(c)
    am=am/dble(nt)
    rewind(4)
    var=0d0
    do 20,k=1,nt
    read(4,*)(vec(i),i=1,nrv)
20   var=var+((vec(c)-am)**2)
    sd=sqrt(var/(db1e(nt)-1d0))
    rewind(4)
    return
    end
    subroutine correl(c1,c2,am1,am2,sd1,sd2,vec,nrv,nt,cor)
    real*8 vec(nrv),am1,am2,sd1,sd2,cor
    integer c1,c2
    cor=0d0
    do 10,k=1,nt
    read(4,*)(vec(i),i=1,nrv)
10   cor=cor+((vec(c1)-am1)*(vec(c2)-am2))
    cor=cor/((db1e(nt)-1d0)*sd1*sd2)
    rewind(4)
    return
    end
    subroutine wsumstat(nrv,idismat,parmmat)
    implicit none
    integer nrv,idismat(nrv,2),i,j,idist
    real*8 parmmat(nrv,5),mean,var,p(4),mu,sigsq,gamma,var1,var2,
&alpha,beta,alp1,alp2,a,b,c
    external gamma
    open(unit=5,file='mcpp.002',status='unknown')
    do 100,i=1,nrv
    idist=idismat(i,1)
    do 10,j=1,4
10   p(j)=parmmat(i,j)
    if (idist .eq. 1)then
    mean=(p(1)+p(2))/2.d0
    var=((p(2)-p(1))**2)/12.d0
    write(5,1000)mean,var
    elseif (idist .eq. 2)then
    mean=p(1)+p(2)
    var=p(1)**2
    write(5,1000)mean,var
    elseif (idist .eq. 3)then
    mean=p(1)
    var=p(2)**2
    write(5,1000)mean,var
    elseif (idist .eq. 4)then
    mu=p(1)
    sigsq=p(2)**2
    mean=p(3)+(dexp(mu+(sigsq/2)))
    var=dexp((2*mu)+sigsq)*(dexp(sigsq)-1)
    write(5,1000)mean,var
    elseif (idist .eq. 5)then
    alpha=p(1)
    beta=p(2)
    mean=p(3)+(beta/alpha)*gamma(1d0/alpha)
    var1=2d0*gamma(2d0/alpha)
    var2=(1d0/alpha)*((gamma(1d0/alpha))**2)
    var=(beta**2/alpha)*(var1-var2)
    write(5,1000)mean,var
    elseif (idist .eq. 6)then
    mean=p(1)+(0.577d0*p(2))
    var=1.645d0*(p(2)**2)
    write(5,1000)mean,var
    elseif (idist .eq. 7)then
    mean=p(3)+(p(1)*p(2))
    var=p(1)*(p(2)**2)
    write(5,1000)mean,var
    elseif (idist .eq. 8)then
    alp1=p(1)

```

```

    alp2=p(2)
    a=p(3)
    b=p(4)
    mean=alp1/(alp2+alp1)
    mean=a+((b-a)*mean)
    var=((alp1*alp2)/(((alp1+alp2)**2)*(alp1+alp2+1d0)))*(b-a)**2
    write(5,1000)mean,var
elseif (idist .eq. 9)then
    a=p(1)
    b=p(2)
    c=p(3)
    mean=(a+b+c)/3d0
    var=(a**2+b**2+c**2-(a*b)-(a*c)-(b*c))/18d0
    write(5,1000)mean,var
endif
100 continue
1000 format(2(e9.3,1x))
rewind(5)
return
end

```

Example input file

***** Output file *****

test.out

***** Number of variables, Number of trials *****

9 2000

***** Random number stream(1-32), Flag for correlated variates(0,1)¹ *****

5 1

***** Correlation matrix (if flag for correlated variates=1) *****

```
1.000  .500  .300  .300  .400  .000  .300  .000  .000
.500  1.000  .400  -.400  .200  .300  .000  .000  .000
.300  .400  1.000  .500  .600  .600  .000  .300  .400
.300  -.400  .500  1.000  .500  .200  .300  .300  .500
.400  .200  .600  .500  1.000  .700  .440  .400  .500
.000  .300  .600  .200  .700  1.000  .300  .500  .600
.300  .000  .000  .300  .440  .300  1.000  .200  .600
.000  .000  .300  .300  .400  .500  .200  1.000  .500
.000  .000  .400  .500  .500  .600  .600  .500  1.000
```

***** Distribution id, Truncation flag(0,1,2,3)², Parameters³ *****

2 0 0.90 2.00

1 0 2.00 5.00

3 1 1.50 0.80 0.00

4 2 -0.80 0.20 0.00 20.00

7 0 3.00 1.00 0.00

6 0 7.70 3.80

5 3 1.50 1.00 0.00 0.5 3.50

9 0 0.30 1.00 0.50

8 0 1.50 5.00 0.00 1.00

**** Flag to compute summary statistics(0,1)⁴ *****

1

Notes:

¹ 0=Independent variates, 1=Correlated variates.

² 0=No truncation, 1=Left truncated, 2=Right truncated, 3=Left and right truncated.

³ If distribution is truncated (truncation flag is not 0), truncation points are also entered.

⁴ 0=Do not compute summary statistics, 1=Compute statistics.

Example output summary file

*** Inputs read in ***

Correlation matrix:

```

-----
1.00 .50 .30 .30 .40 .00 .30 .00 .00
.50 1.00 .40 -.40 .20 .30 .00 .00 .00
.30 .40 1.00 .50 .60 .60 .00 .30 .40
.30 -.40 .50 1.00 .50 .20 .30 .30 .50
.40 .20 .60 .50 1.00 .70 .44 .40 .50
.00 .30 .60 .20 .70 1.00 .30 .50 .60
.30 .00 .00 .30 .44 .30 1.00 .20 .60
.00 .00 .30 .30 .40 .50 .20 1.00 .50
.00 .00 .40 .50 .50 .60 .60 .50 1.00

```

Variable Distribution Truncation* Parameters/(Truncation points)

```

-----
1            Exponential            N            .90    2.00
2            Uniform                N            2.00   5.00
3            Normal                  L            1.50   .80    .00
4            Lognormal              R            -.80   .20    .00   20.00
5            Gamma                   N            3.00   1.00   .00
6            Gumbel                  N            7.70   3.80
7            Weibull                L,R          1.50   1.00   .00   .50   3.50
8            Triangular            N            .30   1.00   .50
9            Beta                    N            1.50   5.00   .00   1.00
-----

```

*N=No truncation,L=Left truncated,R=Right truncated

*** Summary statistics for generated variates ***

```

-----
Variable    Expected    Actual    Expected    Actual
            mean*        mean        variance*    variance
-----
1            .290E+01    .288E+01    .810E+00    .845E+00
2            .350E+01    .346E+01    .750E+00    .775E+00
3            .150E+01    .154E+01    .640E+00    .555E+00
4            .458E+00    .459E+00    .858E-02    .833E-02
5            .300E+01    .293E+01    .300E+01    .284E+01
6            .989E+01    .969E+01    .238E+02    .218E+02
7            .903E+00    .115E+01    .376E+00    .274E+00
8            .600E+00    .601E+00    .217E-01    .215E-01
9            .231E+00    .227E+00    .237E-01    .214E-01
-----

```

*Expected mean and variance are without truncation.
With truncation, the expected mean will be different
and the expected variance will be smaller.

Correlation matrix for generated variates:

```

-----
1.00 .46 .27 .25 .33 -.02 .26 -.02 -.02
.46 1.00 .40 -.38 .15 .26 .00 -.05 .00
.27 .40 1.00 .49 .57 .56 -.01 .25 .38
.25 -.38 .49 1.00 .50 .18 .26 .29 .47
.33 .15 .57 .50 1.00 .66 .40 .36 .46
-.02 .26 .56 .18 .66 1.00 .26 .48 .57
.26 .00 -.01 .26 .40 .26 1.00 .17 .55
-.02 -.05 .25 .29 .36 .48 .17 1.00 .47
-.02 .00 .38 .47 .46 .57 .55 .47 1.00

```

Appendix B

Listing of routine for interactive assessment of correlation structures

```

C-----
C This program allows the interactive generation of a correlation matrix that is
C positive-definite, based on a recursive method for factorization described by
C Scheur and Stoller (1962).
C To allow the user to jump back to a previous row at any time while generating the
C correlation matrix, liberal use was made of the GOTO statement.
C The GOTO's can be replaced with DOWHILE statements, if supported by the compiler.
C Author: Dipmani Kumar
C Last modify date: 8/27/1994
C-----

```

```

parameter (n=30)
integer ir,i,j,k,nv,ncc,icbn,icbn1,icbn2
real p(n,n),pr(n,n),c(n,n),termv(n-1),sumt,val,
&ubound,lbound,chval,smval
intrinsic TINY,EPSILON
character*25 infile
character*1 iyn,iyn1
write(*,*) ' INTERACTIVE CORRELATION MATRIX GENERATOR '
write(*,*) ' ***** '
write(*,*)
4 write(*, '(a\))') ' Enter number of variables (3 to 30): '
read(*,*,err=4)nv
if(nv.lt.3 .or. nv.gt.30)then
write(*,*) ' Number of variables must be between 3 and 30...'
goto 4
endif
5 format(i2)
do 10,i=1,nv
10 p(i,i)=1.0
8 write(*, '(a\))') ' Enter (2,1) correlation : '
read(*,*,err=8)p(2,1)
if(p(2,1).lt.-0.999 .or. p(2,1).gt.0.999)then
write(*,*) ' All correlations must lie in the interval [-0.999,0.9
&999]. '
goto 8
endif
irowst=3
9 continue
do 1000,ir=irowst,nv
99 write(*,100)ir
100 format(' Row ',i2,' computation')
write(*,*) ' -----'
write(*,105)ir-1
105 format(' Enter column number for which bounds are to be computed
&[1',',',',i2,','] : '\)
read(*,*,err=99)ncc
if(ncc.lt. 1 .or. ncc.gt.(ir-1))then
write(*,*) ' Invalid entry..'
goto 99
endif
111 continue
do 120,j=1,ir-1
if(j .ne. ncc)then
103 write(*,104)ir,j
104 format(' Enter ('i2,',',i2,') correlation : '\)
read(*,*,err=103)p(ir,j)
if(p(ir,j).lt.-0.999 .or. p(ir,j).gt.0.999)then
write(*,*) ' All correlations must lie in the interval [-0.999,
&0.999]. '
goto 103
endif
endif
120 continue
call reord1(((ir-1)-ncc),ir-1,p,pr,n)
call choles(c,pr,ir,ir-2,n)
do 250,k=1,(ir-2)
val=0.0
if(k .gt. 1)then
do 220,j=1,k-1
220 val=val+c(ir,j)*c(k,j)
endif
250 termv(k)=((pr(ir,k)-val)**2)/(c(k,k))**2
chval=0.0
do 300,k=1,ir-2
300 chval=chval+termv(k)
if (chval .gt. 1.0)then
write(*,*) ' Error - Values will violate conditions: '
write(*,3001)ir
3001 format (' Re-enter values for row ',i2,' ? (Y=Yes,N=No): '\)
read(*, '(a)',err=301)iyn
if(iyn.ne.'y'.and.iyn.ne.'Y'.and.iyn.ne.'n'.and.iyn.ne.'N')then

```



```

        goto 6003
    endif
    if (iyn1 .eq. 'Y' .or. iyn1 .eq. 'y')then
    write(*,5014)ir
5014  format(' Redoing row ',i2,'...')
        goto 99
    endif
6004  write(*,'(a\)\')' Go back to a previous row ? (Y=Yes,N=No): '
    read(*,'(a)',err=6004)iyn1
        if(iyn1.ne.'y'.and.iyn1.ne.'Y'.and.iyn1.ne.'n'.and.iyn1.ne.'N')
        & then
            write(*,*)' Invalid entry...'
            goto 6004
        endif
        if (iyn1 .eq. 'Y' .or. iyn1 .eq. 'y')then
6005  write(*,'(a\)\')' Enter row to go to: '
    read(*,*,err=6005)irobst
        if(irobst.gt.ir)then
            write(*,6013)ir
            goto 6005
6013  format(' Still doing row ',i2,' ....')
        endif
        if(irobst.ge.3)then
            write(*,502)irobst
502   format(' Going back to row ',i2,'....')
            goto 9
        else
            write(*,*)' Starting again with row 2...'
            goto 8
        endif
    endif
6015  write(*,'(a\)\')' Start again ? (Y=Yes,N=No): '
    read(*,'(a)',err=6015)iyn1
        if(iyn1.ne.'y'.and.iyn1.ne.'Y'.and.iyn1.ne.'n'.and.iyn1.ne.'N')
        & then
            write(*,*)' Invalid entry...'
            goto 6015
        endif
        if (iyn1 .eq. 'Y' .or. iyn1 .eq. 'y')then
            write(*,*)' Starting again with row 2...'
            goto 8
        endif
        stop 'Program terminated - correlation matrix not generated.'
    endif
7000  write(*,'(a\)\')' Enter value: '
    read(*,*,err=7000)p(ir,ncc)
    if(p(ir,ncc).lt.lbound.or.p(ir,ncc).gt.ubound)then
        write(*,7001)ir,ncc,lbound,ubound
        goto 7000
7001  format(' Value for (',i2,',',i2,') correlation is bounded between
        & ',f5.3,' and ',f5.3,')'
    endif
1000  continue
    write(*,'(a\)\')' Write to file ? (Y=Yes,N=No): '
6016  read(*,'(a)',err=6016)iyn1
        if(iyn1.ne.'y'.and.iyn1.ne.'Y'.and.iyn1.ne.'n'.and.iyn1.ne.'N')
        & then
            write(*,*)' Invalid entry...'
            goto 6016
        endif
        if(iyn1 .eq. 'Y' .or. iyn1 .eq. 'y')then
            write(*,'(a\)\')' Enter filename: '
            read(*,'(a)')infile
            open(unit=1,file=infile,status='unknown')
            do 600,i=2,nv
                do 600,j=1,i-1
600   p(j,i)=p(i,j)
                do 700,i=1,nv
700   write(1,800)(p(i,j),j=1,nv)
800   format(15(f6.3,1x))
            endif
            stop ' Normal termination.'
        end
C-----
C      Subroutine for computeing the elements of the Cholesky factor of p
C-----
        subroutine choles(c,p,i1,j1,n)
        integer i,j,i1,j1,k
        real c(n,n),p(n,n),valij,valjj
        do 100,i=1,i1
100  c(i,1)=p(i,1)

```

```

do 500,i=2,i1-1
do 500,j=2,i
if(j .eq. i)then
valjj=0.0
do 400,k=1,(i-1)
400 valjj=valjj+((c(i,k))**2)
c(i,j)=sqrt(1-valjj)
endif
if (j .ne. i)then
valij=0.0
do 450,k=1,(j-1)
450 valij=valij+((c(i,k)*c(j,k)))
c(i,j)=(p(i,j)-valij)/c(j,j)
endif
500 continue
if (j1 .gt. 1)then
i=i1
do 600,j=2,j1
valij=0.0
do 550,k=1,j-1
550 valij=valij+((c(i,k)*c(j,k)))
600 c(i,j)=(p(i,j)-valij)/c(j,j)
endif
return
end
C-----
C Subroutine for re-ordering matrix so that bounds are always computed for the
C (i,i-1) entry where i is the current row number
C-----
subroutine reord1(k,nrv,cov,cov1,nv)
real cov(nv,nv),cov1(nv,nv)
do 200,i=1,nrv
200 cov1(i,i)=1.0
do 100,i1=2,nrv
i=i1-k
if(i .le. 0)i=i+nrv
do 100, j1=1,i1-1
j=j1-k
if(j .le. 0)j=j+nrv
if((j .gt. i))then
cov1(i1,j1)=cov(j,i)
else
cov1(i1,j1)=cov(i,j)
endif
c write(*,*)i1,j1,i,j
100 continue
i1=nrv+1
do 300,j1=1,nrv
j=j1-k
if(j .gt. nrv)j=j-nrv
if(j .le. 0)j=j+nrv
300 cov1(i1,j1)=cov(i1,j)
c call prmat(cov,nrv+1,nv)
c call prmat(cov1,nrv+1,nv)
return
end
subroutine prmat(m,n,n1)
real m(n1,n1)
do 10,i=1,n
10 write(*,*)(m(i,j),j=1,i)
return
end

```

Appendix C

Listing of routines for approximating probability distributions of soil hydraulic properties

Runoff curve number distributions

```

C-----
C   This routine allows generation of variates from the complete CN and CN*
C   distributions. Inputs are either lognormal parameter estimates for the
C   complete S distribution, or the ARC II CN and exceedence probability for
C   the ARC III CN. Input on the fraction of total CN variability explained by
C   antecedent moisture variation also has to be input to generate the CN*
C   distribution.
C   Author: Dipmani Kumar
C   Last modify date: 10/31/1994
C-----
      implicit double precision (a-h,o-z)
      character*15 cnoutf,cn2outf
      character*1 yninp
      double precision cumnor,invnor
      external cumnor,invnor
 8   write(*,'(a)\')' Enter complete CN distribution output file name:
      &
      read(*,'(a)',err=8)cnoutf
      open(unit=1,file=cnoutf,status='unknown')
 9   write(*,'(a)\')' Enter CN II distribution output file name: '
      read(*,'(a)',err=9)cn2outf
 7   write(*,'(a)\')' Enter number of variates to generate: '
      read(*,*,err=7)nvar
      open(unit=2,file=cn2outf,status='unknown')
11  write(*,'(a)\')' Generate CN/CN II distribution from available log
      &normal'
      write(*,'(a)\')' estimates for potential retention (Y/N)?: '
      read(*,'(a)',err=11)yninp
      if(yninp.eq. 'n' .or. yninp .eq. 'N')then
12  write(*,'(a)\')' Enter ARC II curve number: '
      read(*,*,err=12)cn2
14  write(*,'(a)\')' Enter exceedence probability (%) for ARC III cur
      &ve number: '
      read(*,*,err=14)cprob
      s2=(25400d0/cn2)-(254d0)
      s1=2.3d0*s2
      s3=s2/2.3d0
      c1=(invnor(cprob/100d0))
      c3=invnor(1d0-cprob/100d0)
      sig=(dlog(s3)-dlog(s1))/(c1-c3)
      amu1=dlog(s3)-(sig*c1)
      amu3=dlog(s1)-(sig*c3)
      amu=dlog(s2)
      write(*,*)amu,amu1,amu3,sig
      write(*,*)cprob,c1,c3
      call gencn(nvar,amu,sig,c1,c3)
      else
 2   write(*,'(a)\')' Enter lognormal scale parameter for potential ret
      &ention (mm): '
      read(*,*,err=2)amu
 3   write(*,'(a)\')' Enter lognormal shape parameter for potential ret
      &ention (mm): '
      read(*,*,err=3)sig
      coff=dexp(amu)/2.3d0
      cprob=(cumnor((dlog(coff)-amu)/sig))*100d0
      c1=invnor(cprob/100d0)
      c3=invnor(1d0-(cprob/100d0))
c   write(*,*)cprob,amu,sig,c1,c3
c   write(*,*)'Calling gencn...'
      call gencn(nvar,amu,sig,c1,c3)
      endif
      end
      subroutine gencn(nvar,amu,sig,c1,c3)
      implicit double precision (a-h,o-z)
      double precision invnor,ranf,val(2)
      external invnor,ranf
      write(*,'(a)\')' Enter percentage of curve number variation explai
      &ned by antecedent moisture variation: '
      read(*,*)feamc
      cprob=(50d0-(feamc/2d0))
      c1=invnor(cprob/100d0)
      c3=invnor(1d0-cprob/100d0)
      write(*,*)c1,c3,cprob,invnor(0.5d0)
      do 5,i=1,nvar
      do 2,k=1,2

```

```

2   val(k)=ranf()
   vals=invnorr(val(1))
   s=dexp((vals*sig)+amu)
   cn=25400d0/(s+254d0)
   write(1,1000)cn
   if(val(1).gt.0.5d0)then
   va2=(cprob/100d0)+((1d0-(cprob/100d0))*val(2))
   else
   va2=(cprob/100d0)*val(2)
   endif
   va2=invnorr(va2)
   if(va2.le.c1)then
   va2=va2-c1
   s2n=dexp((va2*sig)+amu)
   cn2n=25400d0/(s2n+254d0)
10  write(2,1000)cn2n
   elseif(va2.ge.c3)then
   va2=va2-c3
   s2n=dexp((va2*sig)+amu)
   cn2n=25400d0/(s2n+254d0)
20  write(2,1000)cn2n
   endif
c   write(*,*)i
   5   continue
1000 format(f6.2)
   return
   end

```

Distributions for saturated conductivity, and moisture contents at 0.1 bar, 0.33 bar and 15 bar tensions

```

C-----
C   This routine allows the generation of distributions for saturated conductivity, and
C   soil moisture contents at 0.1 bar, 0.33 bar, and 15 bar tensions. The routine may
C   be used in a trace-driven simulation in which observed input vectors of predictor
C   variables (sand, clay, porosity, and optionally organic matter content) are used to
C   directly generate the distributions.
C   Variation in predictor variables may also be defined by specifying the expected
C   ranges (maximum and minimum values) for the predictors, in which the variables are
C   assumed to follow a beta distribution, rescaled to the appropriate range, with both
C   shape parameters equal to 1.5.
C   The routine uses new regression equations developed by the author. Coefficients for
C   various regression equations are read from files. The elements of the inverse of
C   (X'X) matrix for various regression equations, used to compute
C   the prediction variance are also read from files.
C   Author: Dipmani Kumar
C   Last modify date: 7/29/1994
C-----
parameter(i1=13,i2=14,i3=21,i4=13,i5=21,i6=15,i7=20,max=35)
real*8 corm(4,4),meanv(4),work(4),x(4),parm(15)
real*8 barr(max),xpxim(max,max),vit(1,max),vi(max,1),prod(1,max)
character*1 iom,iid
character*15 infile,ofks,offc1,offc3,ofwp
write(*,*)'*** Distributions for soil hydraulic properties ** '
write(*,'(a)\')' Use organic matter as a predictor for field capac
&ity/wilting point ? (Y/N): '
read(*,'(a)\')iom
if(iom.eq.'n' .or. iom.eq.'N')then
nivar=3
write(*,'(a)\')' Use available observations for sand/clay/porosity
&? (Y/N): '
else
nivar=4
write(*,'(a)\')' Use available observations for sand/clay/porosity
&/organic matter ? (Y/N): '
endif
read(*,'(a)\')iid
if(iid.eq.'Y'.or. iid.eq.'y')then
write(*,'(a)\')' Enter input file name: '
read(*,'(a)\')infile
open(unit=10,file=infile,status='old')
else
call geninp0(nivar,corm,meanv,work,x,parm)
endif
write(*,'(a)\')' Enter saturated conductivity output file name: '

```

```

read(*,'(a)')ofks
open(unit=1,file=ofks,status='unknown')
write(*,'(a\)'') Enter 0.1 bar water content output file name: '
read(*,'(a)')offc1
open(unit=2,file=offc1,status='unknown')
write(*,'(a\)'') Enter 0.33 bar water content output file name: '
read(*,'(a)')offc3
open(unit=3,file=offc3,status='unknown')
write(*,'(a\)'') Enter 15 bar water content output file name: '
read(*,'(a)')ofwp
open(unit=4,file=ofwp,status='unknown')
call genks0(barr,xpxim,vi,vit,prod,il)
if(iom.eq.'Y' .or. iom.eq.'y')then
call genfc1a0(barr,xpxim,vit,vi,prod,i3)
call genfc3a0(barr,xpxim,vit,vi,prod,i5)
call genwpa0(barr,xpxim,vit,vi,prod,i7)
else
call genfc10(barr,xpxim,vit,vi,prod,i2)
call genfc30(barr,xpxim,vit,vi,prod,i4)
call genwp0(barr,xpxim,vit,vi,prod,i6)
endif
stop ' '
end
subroutine geninp(nv,corm,meanv,work,x,parm)
implicit double precision (a-h,o-z)
real*8 corm(nv,nv),work(nv),x(nv),meanv(nv),parm((nv*(nv+3)/2)+1)
character*1 stext
external setgmn,genmn,cumnor,ranf
open(unit=10,file='shpdis.001',status='unknown')
open(unit=11,file='shpdis.002',status='unknown')
write(*,'(a\)'') Enter percentage range for sand (minimum,maximum)
&:
read(*,*)amins,amaxs
write(*,'(a\)'') Enter percentage range for clay (minimum,maximum)
&:
read(*,*)aminc,amaxc
write(*,'(a\)'') Enter fraction range for porosity (minimum,maximu
&m):
read(*,*)aminp,amaxp
if(nv.eq.4)then
write(*,'(a\)'') Enter percentage range for organic matter (minim
&um,maximum):
read(*,*)aminom,amaxom
endif
write(*,'(a\)'') Enter number of variates to generate: '
read(*,*)nranv
write(*,'(a\)'') Enter random number stream (1-32): '
read(*,*)istream
if(amins.gt.65d0 .and. amaxc.le.35d0)then
corm(2,1)=-0.774d0
elseif(amaxs.le.65d0 .and. amaxc.le.35d0)then
corm(2,1)=-0.237d0
elseif(amaxs.le.65d0 .and. aminc.gt.35d0)then
corm(2,1)=-0.353d0
else
write(*,'(a\)'') Use sand/clay correlation for clayey(C), loamy
&(L), or sandy(S) soils?: '
read(*,'(a)')stext
if(stext.eq.'C' .or. stext.eq.'c')corm(2,1)=-0.353d0
if(stext.eq.'L' .or. stext.eq.'l')corm(2,1)=-0.237d0
if(stext.eq.'S' .or. stext.eq.'s')corm(2,1)=-0.774d0
endif
corm(3,1)=0.0d0
corm(3,2)=0.0d0
if(nv.eq.4)then
corm(4,1)=0.0d0
corm(4,2)=0.0d0
corm(4,3)=0.564d0
endif
do 10,i=1,nv
corm(i,i)=1.0d0
10 meanv(i)=0.0d0
call fillsym(corm,nv)
call setcgn(istream)
call setgmn(meanv,corm,nv,parm)
do 50,i=1,nranv*2
call genmn(parm,x,work)
do 35,j=1,nv
35 x(j)=cumnor(x(j))
50 write(11,*)x
rewind(11)

```

```

i1=0
dowhile(i1.lt.nranv)
  read(11,*)(work(j1),j1=1,nv)
  do 150,i=1,nv
    val=work(i)
    if(i.eq.1)then
      a=amins
      b=amaxs
    elseif(i.eq.2)then
      a=aminc
      b=amaxc
    elseif(i.eq.3)then
      a=aminp
      b=amaxp
    elseif(i.eq.4)then
      a=aminom
      b=amaxom
    else
      write(*,*)' ERROR !'
      stop ' '
    endif
    call genbet(a,b,val)
    work(i)=val
    continue
150   if((work(1)+work(2)).lt.99.9)then
      write(10,1000)(work(k),k=1,nv)
      i1=i1+1
    endif
1000  enddo
200   format(4(f6.2,1x))
      continue
      rewind(10)
      close(11)
      return
      end
      subroutine genbet(a,b,val)
      implicit none
      real*8 val,a,b,c,alp1,alp2,invbet
      external invbet
      parameter(alp1=3.5d0,alp2=3.5d0)
      c=b-a
      val=invbet(val,alp1,alp2)
      val=a+(val*c)
      return
      end
      subroutine fillsym(mat,n)
      implicit none
      integer i,j,n
      real*8 mat(n,n)
      do 300,i=2,n
        do 300,j=1,i-1
300   mat(j,i)=mat(i,j)
      return
      end
      subroutine genks(barr,xpxim,vi,vit,prod,i1)
      implicit double precision (a-h,o-z)
      real*8 barr(i1),xpxim(i1,i1),vit(1,i1),vi(i1,1),prod(1,i1)
      real*8 ksm,ksl,ksu
      data srms/1.348d0/
      iend=0
      open(unit=8,file='ksb.dat',status='old')
      open(unit=9,file='ksmat.dat',status='old')
      do 10,i=1,i1
10   read(8,*)barr(i)
1001 format(20(e10.4,1x))
      do 20,i=1,i1
20   read(9,*)(xpxim(i,j),j=1,i1)
      close(8)
      close(9)
      dowhile(iend.eq.0)
      read(10,*,end=50)s,c,p
      s=s-73.604d0
      c=c-11.179d0
      p=p-0.4295d0
      vit(1,1)=1d0
      vit(1,2)=s
      vit(1,3)=c
      vit(1,4)=p
      vit(1,5)=s**2
      vit(1,6)=p**2
      vit(1,7)=s*p

```

```

    vit(1,8)=c*p
    vit(1,9)=c**3
    vit(1,10)=(s**2)*p
    vit(1,11)=(c**2)*p
    vit(1,12)=(p**2)*s
    vit(1,13)=(p**2)*c
25   do 25,i=1,i1
      vi(i,1)=vit(1,i)
      ksm=0d0
30   do 30,i=1,i1
      ksm=ksm+(barr(i)*vit(1,i))
      call amatmul(vit,xpxim,prod,1,i1,i1,i1)
      vx=0d0
      do 35,i=1,i1
35   vx=vx+(prod(1,i)*vi(i,1))
      sep=srms*dsqrt(1d0+vx)
      ksl=ksm-sep
      ksu=ksm+sep
      ksm=dexp(ksm)
      ksl=dexp(ksl)
      ksu=dexp(ksu)
      write(1,1000)ksl,ksm,ksu
    enddo
50   continue
1000 format(3(f10.3,1x))
    rewind(10)
    return
end
subroutine genfcla(barr,xpxim,vi,vit,prod,i3)
implicit double precision (a-h,o-z)
real*8 barr(i3),xpxim(i3,i3),vit(1,i3),vi(i3,1),prod(1,i3)
data srms/0.68790d0/
iend=0
open(unit=8,file='fclab.dat',status='old')
open(unit=9,file='fclamat.dat',status='old')
do 10,i=1,i3
10  read(8,*)barr(i)
1001 format(20(e10.4,1x))
do 20,i=1,i3
20  read(9,*)(xpxim(i,j),j=1,i3)
close(8)
close(9)
dowhile(iend.eq.0)
read(10,*,end=50)s,c,p,o
s=s-66.263d0
c=c-11.066d0
p=p-0.4634d0
o=o-1.0335d0
vit(1,1)=1d0
vit(1,2)=s
vit(1,3)=c
vit(1,4)=p
vit(1,5)=o
vit(1,6)=c**2
vit(1,7)=p**2
vit(1,8)=s*c
vit(1,9)=s*p
vit(1,10)=s*o
vit(1,11)=c*p
vit(1,12)=(s**2)*o
vit(1,13)=(c**2)*p
vit(1,14)=(c**2)*o
vit(1,15)=(p**2)*o
vit(1,16)=(o**2)*c
vit(1,17)=(p**2)*s
vit(1,18)=(p**2)*c
vit(1,19)=s*c*p
vit(1,20)=c*p*o
vit(1,21)=p*o*s
25  do 25,i=1,i3
      vi(i,1)=vit(1,i)
      fclam=0d0
30  do 30,i=1,i3
      fclam=fclam+(barr(i)*vit(1,i))
      call amatmul(vit,xpxim,prod,1,i3,i3,i3)
      vx=0d0
35  do 35,i=1,i3
      vx=vx+(prod(1,i)*vi(i,1))
      sep=srms*dsqrt(1d0+vx)
      fclal=fclam-sep
      fclau=fclam+sep

```

```

        fc1am=(fc1am**2)/100d0
        fc1al=(fc1al**2)/100d0
        fc1au=(fc1au**2)/100d0
        write(2,1000) fc1al, fc1am, fc1au
    enddo
50  continue
1000 format(3(f6.3,1x))
    rewind(10)
    return
end
subroutine genfc3a(barr,xpxim,vi,vit,prod,i5)
implicit double precision (a-h,o-z)
real*8 barr(i5),xpxim(i5,i5),vit(1,i5),vi(i5,1),prod(1,i5)
data srms/0.62570d0/
iend=0
open(unit=8,file='fc3ab.dat',status='old')
open(unit=9,file='fc3amat.dat',status='old')
do 10,i=1,i5
10  read(8,*)barr(i)
1001 format(20(e10.4,1x))
do 20,i=1,i5
20  read(9,*)(xpxim(i,j),j=1,i5)
    close(8)
    close(9)
    dowhile(iend.eq.0)
    read(10,*,end=50)s,c,p,o
        s=s-65.857d0
        c=c-11.993d0
        p=p-0.46258d0
        o=o-1.0296d0
        vit(1,1)=1d0
        vit(1,2)=s
        vit(1,3)=c
        vit(1,4)=p
        vit(1,5)=o
        vit(1,6)=c**2
        vit(1,7)=p**2
        vit(1,8)=s*c
        vit(1,9)=s*p
        vit(1,10)=s*o
        vit(1,11)=c*p
        vit(1,12)=(s**2)*o
        vit(1,13)=(c**2)*p
        vit(1,14)=(c**2)*o
        vit(1,15)=(p**2)*o
        vit(1,16)=(o**2)*c
        vit(1,17)=(p**2)*s
        vit(1,18)=(p**2)*c
        vit(1,19)=s*c*p
        vit(1,20)=c*p*o
        vit(1,21)=p*o*s
        do 25,i=1,i5
25  vi(i,1)=vit(1,i)
        fc3am=0d0
        do 30,i=1,i5
30  fc3am=fc3am+(barr(i)*vit(1,i))
        call amatmul(vit,xpxim,prod,1,i5,i5,i5)
        vx=0d0
        do 35,i=1,i5
35  vx=vx+(prod(1,i)*vi(i,1))
        sep=srms*dsqrt(1d0+vx)
        fc3al=fc3am-sep
        fc3au=fc3am+sep
        fc3am=(fc3am**2)/100d0
        fc3al=(fc3al**2)/100d0
        fc3au=(fc3au**2)/100d0
        write(3,1000) fc3al, fc3am, fc3au
    enddo
50  continue
1000 format(3(f6.3,1x))
    rewind(10)
    return
end
subroutine genwpa(barr,xpxim,vi,vit,prod,i7)
implicit double precision (a-h,o-z)
real*8 barr(i7),xpxim(i7,i7),vit(1,i7),vi(i7,1),prod(1,i7)
data srms/0.4792d0/
iend=0
open(unit=8,file='wpab.dat',status='old')
open(unit=9,file='wpamat.dat',status='old')
do 10,i=1,i7

```

```

10  read(8,*)barr(i)
1001 format(20(e10.4,1x))
do 20,i=1,i7
20  read(9,*)(xpxim(i,j),j=1,i7)
close(8)
close(9)
dowhile(iend.eq.0)
read(10,*,end=50)s,c,p,o
s=s-62.345d0
c=c-13.448d0
p=p-0.45687d0
o=o-0.98097d0
vit(1,1)=1d0
vit(1,2)=s
vit(1,3)=c
vit(1,4)=p
vit(1,5)=o
vit(1,6)=s**2
vit(1,7)=p**2
vit(1,8)=s*p
vit(1,9)=s*o
vit(1,10)=c*p
vit(1,11)=c**3
vit(1,12)=(s**2)*p
vit(1,13)=(c**2)*p
vit(1,14)=(c**2)*o
vit(1,15)=(o**2)*c
vit(1,16)=(o**2)*s
vit(1,17)=(p**2)*s
vit(1,18)=(p**2)*c
vit(1,19)=s*c*p
vit(1,20)=c*p*o
do 25,i=1,i7
25  vi(i,1)=vit(1,i)
wpam=0d0
do 30,i=1,i7
30  wpam=wpam+(barr(i)*vit(1,i))
call amatmul(vit,xpxim,prod,1,i7,i7,i7)
vx=0d0
do 35,i=1,i7
35  vx=vx+(prod(1,i)*vi(i,1))
sep=srms*dsqrt(1d0+vx)
wpal=wpam-sep
wpau=wpam+sep
wpam=(wpam**2)/100d0
wpal=(wpal**2)/100d0
wpau=(wpau**2)/100d0
write(4,1000)wpal,wpam,wpau
enddo
50  continue
1000 format(3(f6.3,1x))
rewind(10)
return
end
subroutine genfcl(barr,xpxim,vi,vit,prod,i2)
implicit double precision (a-h,o-z)
real*8 barr(i2),xpxim(i2,i2),vit(1,i2),vi(i2,1),prod(1,i2)
data srms/0.7576d0/
iend=0
open(unit=8,file='fclb.dat',status='old')
open(unit=9,file='fclmat.dat',status='old')
do 10,i=1,i2
10  read(8,*)barr(i)
1001 format(20(e10.4,1x))
do 20,i=1,i2
20  read(9,*)(xpxim(i,j),j=1,i2)
close(8)
close(9)
dowhile(iend.eq.0)
read(10,*,end=50)s,c,p
s=s-61.211d0
c=c-14.896d0
p=p-0.46469d0
vit(1,1)=1d0
vit(1,2)=s
vit(1,3)=c
vit(1,4)=p
vit(1,5)=s**2
vit(1,6)=p**2
vit(1,7)=s*p
vit(1,8)=c*p

```

```

        vit(1,9)=c**3
        vit(1,10)=p**3
        vit(1,11)=(s**2)*p
        vit(1,12)=(c**2)*p
        vit(1,13)=(p**2)*s
        vit(1,14)=(p**2)*c
25      do 25,i=1,i2
        vi(i,1)=vit(1,i)
        fclm=0d0
30      do 30,i=1,i2
        fclm=fclm+(barr(i)*vit(1,i))
        call amatmul(vit,xpxim,prod,1,i2,i2,i2)
        vx=0d0
        do 35,i=1,i2
35      vx=vx+(prod(1,i)*vi(i,1))
        sep=srms*dsqrt(1d0+vx)
        fcll=fclm-sep
        fclu=fclm+sep
        fclm=(fclm**2)/100d0
        fcll=(fcll**2)/100d0
        fclu=(fclu**2)/100d0
        write(2,1000) fcll, fclm, fclu
    enddo
50      continue
1000     format(3(f6.3,1x))
        rewind(10)
        return
    end
    subroutine genfc3(barr,xpxim,vi,vit,prod,i4)
    implicit double precision (a-h,o-z)
    real*8 barr(i4),xpxim(i4,i4),vit(1,i4),vi(i4,1),prod(1,i4)
    data srms/0.6929d0/
    iend=0
    open(unit=8, file='fc3b.dat', status='old')
    open(unit=9, file='fc3mat.dat', status='old')
    do 10,i=1,i4
10      read(8,*)barr(i)
1001     format(20(e10.4,1x))
        do 20,i=1,i4
20      read(9,*)(xpxim(i,j),j=1,i4)
        close(8)
        close(9)
    dowhile(iend.eq.0)
    read(10,*,end=50) s,c,p
        s=s-57.893d0
        c=c-16.814d0
        p=p-0.46602d0
        vit(1,1)=1d0
        vit(1,2)=s
        vit(1,3)=c
        vit(1,4)=p
        vit(1,5)=s**2
        vit(1,6)=p**2
        vit(1,7)=s*p
        vit(1,8)=c*p
        vit(1,9)=p**3
        vit(1,10)=(s**2)*p
        vit(1,11)=(c**2)*p
        vit(1,12)=(p**2)*s
        vit(1,13)=(p**2)*c
25      do 25,i=1,i4
        vi(i,1)=vit(1,i)
        fc3m=0d0
30      do 30,i=1,i4
        fc3m=fc3m+(barr(i)*vit(1,i))
        call amatmul(vit,xpxim,prod,1,i4,i4,i4)
        vx=0d0
        do 35,i=1,i4
35      vx=vx+(prod(1,i)*vi(i,1))
        sep=srms*dsqrt(1d0+vx)
        fc3l=fc3m-sep
        fc3u=fc3m+sep
        fc3m=(fc3m**2)/100d0
        fc3l=(fc3l**2)/100d0
        fc3u=(fc3u**2)/100d0
        write(3,1000) fc3l, fc3m, fc3u
    enddo
50      continue
1000     format(3(f6.3,1x))
        rewind(10)
        return

```

```

end
subroutine genwp(barr,xpxim,vi,vit,prod,i6)
implicit double precision (a-h,o-z)
real*8 barr(i6),xpxim(i6,i6),vit(1,i6),vi(i6,1),prod(1,i6)
data srms/0.5618d0/
iend=0
open(unit=8,file='wpb.dat',status='old')
open(unit=9,file='wpmat.dat',status='old')
do 10,i=1,i6
10 read(8,*)barr(i)
1001 format(20(e10.4,1x))
do 20,i=1,i6
20 read(9,*)(xpxim(i,j),j=1,i6)
close(8)
close(9)
dowhile(iend.eq.0)
read(10,*,end=50)s,c,p
s=s-55.054d0
c=c-18.615d0
p=p-0.45739d0
vit(1,1)=1d0
vit(1,2)=s
vit(1,3)=c
vit(1,4)=p
vit(1,5)=s**2
vit(1,6)=p**2
vit(1,7)=s*c
vit(1,8)=s*p
vit(1,9)=c*p
vit(1,10)=c**3
vit(1,11)=p**3
vit(1,12)=(s**2)*p
vit(1,13)=(c**2)*p
vit(1,14)=(p**2)*s
vit(1,15)=(p**2)*c
do 25,i=1,i6
25 vi(i,1)=vit(1,i)
wpm=0d0
do 30,i=1,i6
30 wpm=wpm+(barr(i)*vit(1,i))
call amatmul(vit,xpxim,prod,1,i6,i6,i6)
vx=0d0
do 35,i=1,i6
35 vx=vx+(prod(1,i)*vi(i,1))
sep=srms*dsqrt(1d0+vx)
wpl=wpm-sep
wpu=wpm+sep
wpm=(wpm**2)/100d0
wpl=(wpl**2)/100d0
wpu=(wpu**2)/100d0
write(4,1000)wpl,wpm,wpu
enddo
50 continue
1000 format(3(f6.3,1x))
rewind(10)
return
end
subroutine amatmul(a1,a2,p,n1,n2,n3,i1)
real*8 a1(1,i1),a2(i1,i1),p(1,i1),sval
do 10,i=1,n1
do 10,j=1,n3
sval=0.0d0
do 5,k=1,n2
5 sval=sval+(a1(i,k)*a2(k,j))
10 p(i,j)=sval
return
end
subroutine geninp0(nv,corm,meanv,work,x,parm)
implicit double precision (a-h,o-z)
real*8 corm(nv,nv),work(nv),x(nv),meanv(nv),parm((nv*(nv+3)/2)+1)
call geninp(nv,corm,meanv,work,x,parm)
return
end
subroutine genks0(barr,xpxim,vi,vit,prod,i1)
implicit double precision (a-h,o-z)
real*8 barr(i1),xpxim(i1,i1),vit(1,i1),vi(i1,1),prod(1,i1)
call genks(barr,xpxim,vi,vit,prod,i1)
return
end
subroutine genfcl1a0(barr,xpxim,vi,vit,prod,i3)

```

```

implicit double precision (a-h,o-z)
real*8 barr(i3),xpxim(i3,i3),vit(1,i3),vi(i3,1),prod(1,i3)
call genfcl1a(barr,xpxim,vi,vit,prod,i3)
return
end
subroutine genfc3a0(barr,xpxim,vi,vit,prod,i5)
implicit double precision (a-h,o-z)
real*8 barr(i5),xpxim(i5,i5),vit(1,i5),vi(i5,1),prod(1,i5)
call genfc3a(barr,xpxim,vi,vit,prod,i5)
return
end
subroutine genwpa0(barr,xpxim,vi,vit,prod,i7)
implicit double precision (a-h,o-z)
real*8 barr(i7),xpxim(i7,i7),vit(1,i7),vi(i7,1),prod(1,i7)
call genwpa(barr,xpxim,vi,vit,prod,i7)
return
end
subroutine genfc10(barr,xpxim,vi,vit,prod,i2)
implicit double precision (a-h,o-z)
real*8 barr(i2),xpxim(i2,i2),vit(1,i2),vi(i2,1),prod(1,i2)
call genfc1(barr,xpxim,vi,vit,prod,i2)
return
end
subroutine genfc30(barr,xpxim,vi,vit,prod,i4)
implicit double precision (a-h,o-z)
real*8 barr(i4),xpxim(i4,i4),vit(1,i4),vi(i4,1),prod(1,i4)
call genfc3(barr,xpxim,vi,vit,prod,i4)
return
end
subroutine genwp0(barr,xpxim,vi,vit,prod,i6)
implicit double precision (a-h,o-z)
real*8 barr(i6),xpxim(i6,i6),vit(1,i6),vi(i6,1),prod(1,i6)
call genwp(barr,xpxim,vi,vit,prod,i6)
return
end

```

Vita

Dipmani Kumar was born on January 14, 1963 in Ranchi, India. He obtained a B.Tech in Agricultural Engineering from PAU, India in 1986. He was a lecturer at Gwebi College, Harare, Zimbabwe for two years before starting graduate studies at Virginia Tech, where he completed an M.S. in 1990. He is currently working as an Environmental Engineer with Tetra Tech, Inc.

Dipmani Kumar