EVALUATION OF GLEAMS CONSIDERING PARAMETER UNCERTAINTY

by

Randy W. Clouse

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Biological Systems Engineering

APPROVED:

Conrad D. Heatwole, Chairman

Frank E. Woeste

John V. Perumpal

Mary Leigh Wolfe

May 1996

Blacksburg, Virginia

Keywords: Parameter Uncertainty, Model Error, Model Validation

EVALUATION OF GLEAMS CONSIDERING PARAMETER UNCERTAINTY

by

Randy W. Clouse

Conrad D. Heatwole, Chairman

Biological Systems Engineering

(ABSTRACT)

A probabilistic procedure was applied to the evaluation of predictions from the GLEAMS nonpoint source pollution model. Assessment of both the procedure and model was made by comparing absolute and relative predictions made with both probabilistic and deterministic procedures. Field data used came from a study of pesticide fate and transport in both no-till and conventional tillage plots in a Coastal plain soil. Variables examined were: runoff, sediment yield, surface losses, mass in the root zone, and depth of center of mass for two pesticides and a tracer. Random inputs were characterized with probability distributions. Values for inputs were sampled from these distributions for 5000 model executions to create output distributions in the probabilistic procedure. Central tendency values from the probabilistic input distributions were used as inputs for the deterministic runs.

Model predictions generally followed expected trends and were within observed variability. Two exceptions were systematic under-predictions of runoff and pesticide losses and under-predictions of the depth of bromide in the root zone later in the observed period. These exceptions may indicate errors in the runoff and plant uptake components of the model. Neither procedure made relative predictions correctly all the time, however subjective assessment of the model results led to consistent decisions between the two procedures. The probabilistic procedure reduced parameter uncertainty by eliminating arbitrary parameter selection from available data by utilizing the complete range of data, however, it did not eliminate uncertainty in the data itself.

Acknowledgements

I would like to thank Dr. Conrad Heatwole for his patience and support over the course of my graduate study. Thanks also go out to Dr. John Perumpal, Dr. Frank Woeste, and Dr. Mary Leigh Wolfe for all their consultation during my course of study. I am most grateful for assistance from Dipmani Kumar and Sebastian Zacharias in performing the data analysis. Thanks to Joel Paz and Neil Zharadka for all their help with the FARMSCALE project during my course of study. The financial support from the Biological Systems Engineering Department and the Chesapeake Research Consortium is gratefully acknowledged. Finally, thanks go out to my parents and my brother for giving me the support to get through it all.

Table of Contents

List of Illustrations

List of Tables

INTRODUCTION

The water that we drink every day is a precious commodity that helps keep not only us as human beings, but the plants and animals that surround us in our environment, healthy and sustained. Whenever water is degraded from its original background condition it is "polluted." Pollution to water can be grouped into two distinct categories depending on the source of that pollution: point source or nonpoint source (NPS). Point sources are so named because there is a single "point" from which the pollution comes, such as the end of a pipe that discharges industrial wastes from a manufacturing plant. Nonpoint source pollution on the other hand is diffuse in nature, which means that it comes from a large area of land, is variable and unpredictable, and is not as concentrated as point source pollution.

Since point source pollution is readily identifiable as to its source, it is easier to target with regulations and physical controls than is NPS pollution. An example of the reduction of a point source discharge would be upgrading a wastewater treatment plant to reduce the amount of phosphorous discharged into surface waters. Nonpoint source pollution, however, is dependent more on human activities that cannot be as readily controlled or halted, because these activities by their nature are intended to alter the environment. Typical activities that generate NPS pollution include agriculture and construction. Another trait of NPS pollution which makes it unique is that it is highly dependent on meteorological events of which the magnitude and frequency are totally beyond human control.

Nonpoint source pollution can come in a variety of forms. In large enough quantities, eroded and transported soil, termed sediment, can cloud waterways and estuaries enough to cause the loss of aquatic plant and animal habitat. It is estimated that "annual in-stream damage by sediments is between $3.2 and $13 billion" (Novotony and Olem, 1994). Other potential pollutants include the nutrients nitrogen and phosphorous, which are often applied to agricultural crops. These nutrients, either separately or in

combination can spur the growth of algae in inland water bodies, thus clouding the water and destroying habitat for other plants and animals. Nitrogen also poses a hazard to drinking water in that in excessive quantities it can lead to methemoglobinemia, which is more commonly known as "blue baby disease." Methemoglobinemia primarily poses a threat of death to small infants who ingest water containing high concentrations of the nitrate form of nitrogen (Davis and Cornwell, 1991). Other common materials used by agriculture, which can find their way into drinking water and harm plants and animals in the environment, are agricultural pesticides. Residues of the herbicide atrazine have been found to reduce the yields of oats four years after their initial application (Burnside et al., 1969). Persistent pesticides such as atrazine can prevent the use of crop rotations on farms. Pesticides when found in surface waters can lead to fish kills and the deaths of other animals that may feed on fish (Rao and Hornsby, 1989).

In order to alleviate the problems from NPS pollution, practices known as Best Management Practices (BMPs) are implemented. Some examples of BMPs used in agriculture are: various types of reduced tillage, crop rotations, winter cover crops, contour planting, buffer strips of grass or forest land between crop lands and surface waters, and better management of the application of animal waste, chemical fertilizers, and pesticides to the land as compared with previous management (Novotony and Olem, 1994).

In order to determine which type of BMP to implement in a given situation to alleviate a NPS problem, there are two general approaches. The first of these is monitoring, which consists of implementing a BMP in a certain area and collecting data on how much a specific pollutant is reduced. While this method may collect a great deal of information about a specific BMP for a certain site and conditions, it is very costly and time consuming. The other method of determining BMP selection is modeling, which involves mathematical predictions of the fate and transport of pollutants in the environment. With new increases in computer speed, simulations of the effects of many different BMPs for a certain site can be made quickly and cheaply. For NPS pollution,

models have been promoted as tools to make relative comparisons between two or more BMPs (Leonard et al., 1987).

Model predictions fail to match observed data due to error in the observations and error in the modeling. Modeling error can be attributed to (1) errors in the equations and algorithms, and (2) errors in inputs due to inability to capture the spatial and temporal variations of the inputs. Many of the models developed thus far for the prediction of NPS pollution such as GLEAMS (Leonard et al., 1987) and PRZM (Carsel et al., 1985) ignore the uncertainty found in many soil inputs by using average point values. This method of quantifying inputs can be very misleading because some typical soil inputs, such as saturated hydraulic conductivity, can vary up to two or three orders of magnitude within several meters. Methodologies for applying these models have been developed that can account for and deal with these uncertainties (Haan et al., 1995; Kumar, 1995a).

Objectives

The overall objective of this study was to assess the probabilistic procedure developed by Kumar (1995a) by applying it in an evaluation of the prediction of pesticide fate and transport by the GLEAMS NPS model. Specific objectives are:

(1) Evaluate GLEAMS using the probabilistic validation procedure developed by Kumar (1995a) to compare absolute and relative model predictions with observed data. The data in this comparison, where available, will include runoff, sediment yield, pesticide surface losses, mass of solute in the root zone, and depth of center of solute mass, for the pesticides atrazine and metolachlor and the tracer bromide.

(2) Determine the effect of parameter uncertainty on model validation by comparing the conclusions reached by the probabilistic evaluation with those obtained from a deterministic evaluation of the model.

REVIEW OF LITERATURE

OVERVIEW OF NONPOINT SOURCE POLLUTION MODELS - CLASSIFICATION AND TERMINOLOGY

NPS models have proliferated with the advent of faster computers and due to modeling's inherent advantages over monitoring, in being able to compare many alternative management scenarios quickly and cheaply. NPS models vary widely in their complexity, predictive ability, and in the pollutants and transport processes that they attempt to model. In order to understand NPS models better, this section will try to explain what NPS models are, what they are intended to do, and how they are classified by modeling experts.

A model is an assemblage of concepts in the form of equations that represent a mathematical interpretation of natural phenomenon (American Society for Testing and Materials, 1992). Haan (1989) expanded this definition slightly when he defined a hydrologic model, as "a collection of physical laws and empirical observations written in mathematical terms and combined in such a way as to produce a set of results (outputs) based on a set of known and/or assumed conditions (inputs)." Generally, NPS models incorporate relationships for hydrology, sediment detachment, transport, and deposition, water movement in the unsaturated zone, along with overland and subsurface transport and processes for degradation of nutrients and pesticides.

Model classification can be based on several different characteristics of the model. Based on the complexity of the processes included in the model and the intended end use, NPS models can be grouped into three broad classes of research, management, and screening (Wagenet and Hutson, 1986), which are listed in decreasing order of complexity. These classifications are quite analogous to the three classifications, research, action-agency, and extension/farmer, that Loague and Green (1991) used in classifying models from a user's perspective. Research models are intended to further the scientific understanding of the processes and interactions occurring in a given system. A research

model also should be the type of model to most closely match observed data, since it tries to use physically-based equations as opposed to empirical equations. This type of model suffers disadvantages when compared to other types of models due to slow computational times, large data requirements, and difficulty in transferring the knowledge of how to use it from the model developer to the model user. An example of a research model would be LEACHMP (Wagenet and Hutson, 1986).

A second group of models are the management type of models. This type of model is used for making planning decisions based on relative predictions between scenarios (Leonard et al., 1987; Leonard and Knisel, 1990), with different model scenarios being characterized by changes or differences in management systems, soils, climate, chemical properties, and other factors (Kumar, 1995a). This type of model uses as many physically-based equations as possible, but where it is deemed necessary substitutes approximate mathematical solutions and empirical relationships in order to save computational time. Due to these mathematical approximations, the results from management models are not deemed to be as accurate as research models. Examples of management models include GLEAMS (Leonard et al., 1987) and PRZM (Carsel et al., 1985).

The simplest form of a model is a screening model. The purpose of a screening model is to rank chemicals by means of simple indices as to their potential to adversely affect an area of land (Loague et al., 1989). This type of model tries to represent the fate of pollutants by as few equations and relationships as possible, utilizing only those relationships that are most important in characterizing the system. The results from a screening model are generally not useful for making absolute predictions, but rather in classifying pollutants in broad groups according to their potential hazard. Examples of screening models are MOUSE (Nofziger and Hornsby, 1986) and BAM (Jury et al., 1983).

There are a variety of terms besides research, management, and screening that can be used to characterize models. Models can be either deterministic or stochastic

depending on how the inputs and outputs of the model are considered (Clarke, 1973; Addiscott and Wagenet, 1985; Novotony and Olem, 1994). Deterministic models "presume that a system or process operates such that the occurrence of a given set of events leads to a uniquely definable outcome" (Addiscott and Wagenet, 1985). Stochastic models on the other hand make no presuppositions about any one definite outcome, but assume that variables are random and are described by a probability distribution (Clarke, 1973; Loague and Green, 1991).

Further distinctions can be made between mechanistic and functional models (Addiscott and Wagenet, 1985). Mechanistic models try to represent a system using the "most fundamental mechanisms of the process" as understood by modern science. Functional models represent systems in simplified manners that require less computer time and fewer inputs. The definitions of mechanistic and functional models appear to be similar to those applied to physically-based or conceptual and quasi-physically based or empirical models by other authors (Clarke, 1973; Haan, 1989; Loague and Green, 1991). It should be noted that often times there is no clear distinction between a mechanistic and a functional model, but rather there are more or less mechanistic models. Also, models are generally classified by a combination of the above terms, such as in Clarke's (1973) breakdown of models into the following four categories: stochastic-conceptual, stochastic-empirical, deterministic-conceptual, and deterministic-empirical.

Some further characterizations include rate and capacity models (Addiscott and Wagenet, 1985) as these terms are applied to pesticide fate models. Rate models use equations to represent how fast processes occur, while capacity models use relationships dealing with the volume of substances in an area. These two classifications closely parallel the differences between mechanistic and functional models.

Models have been classified by the level of processes that they simulate. Loague and Green (1991) have separated solute transport models into the following four structural levels: single-process, multiple-process, comprehensive, and field-scale. As indicated by the structural level names, these categorizations vary from simulating a single

process to a comprehensive model that can accommodate field-scale heterogeneity and temporal variability. In a similar fashion, hydrologic models have been broken into four levels of complexity, individual processes, component models, integrated watershed models, and global watershed models by Woolhiser and Brakensiek (1982). Many of the management models used today are termed either field-scale or watershed-scale based on the size of the areas that they represent, with GLEAMS (Leonard et al., 1987) and ANSWERS (Beasley and Huggins, 1981), respectively, being representative of the two types.

Management models can also be classified into the categories of event-based and continuous models (Novotony and Olem, 1994). Event-based models simulate the response of an area of land to individual rainfall and snowfall events. Since these models only look at single events, they require little meteorological data and run quicker than other models. However, the specific storm event description and antecedent moisture conditions must be specified for these models to operate correctly. Continuous models simulate a range of processes such as precipitation, available surface storage, snow accumulation and melt, evapotranspiration, soil moisture, infiltration, runoff, and soil water movement, in the order that they occur over a period of time. This method can create a long time series of water and pollutant loadings, but can be computationally intensive. Generally, watershed models, such as ANSWERS, have been event-based, while field-scale models, such as GLEAMS, have been continuous, but this distinction is blurring with the advent of models such as ANSWERS-2000 (Bouraoui, 1994). Whether or not time is included as an independent variable for model equations has led to models being classified as static or dynamic (Woolhiser and Brakensiek, 1982).

Models can also be classified as distributed or lumped parameter (Novotony and Olem, 1994; Kumar, 1995a). Distributed parameter models attempt to divide the area being modeled into smaller, homogeneous units, which have uniform, physically-based inputs. Lumped parameter models fail to take into account the spatial variability of inputs, instead using a single effective value for the entire area modeled.

MODEL EVALUATION

Models are generally created to meet one of two purposes - either to further the scientific understanding of a given environment or pollutant fate process(es) or to allow planners and managers to make decisions based on model predictions. For both of these goals, there is a need to evaluate the model in order to assess how "valid" its representation of a given scenario is.

When creating or evaluating a model, the following seven general steps are recommended in a standard by the ASTM (1992):

(1) Model Conceptualization, involves determining the important processes that must be included in the model description.

(2) Program Examination, involves questioning whether anything fundamental was omitted in the initial model conceptualization.

(3) Algorithm Examination, evaluates whether appropriate numerical schemes have been adopted to represent the model in the form of computer code.

(4) Data Evaluation, is intended to ascertain the quality and the quantity of the input and output data for use with the model.

(5) Sensitivity Analyses, identify which parameters are most influential in determining model predictions.

(6) Validation, refers to simulations to determine the degree to which model predictions agree with field or laboratory measurements of those same quantities, without altering model coefficients following calibration.

(7) Code Comparison, is the process of comparing the strengths and weaknesses of various codes designed to perform similar tasks.

Other sources have broken evaluation into the areas of verification, calibration, and validation (Leonard and Knisel, 1990) with verification roughly corresponding to the third component of the ASTM process. Calibration and validation would match the sixth

ASTM component. In these definitions, model verification includes examination of the computer program to insure that the model builders' concepts of system behavior are represented accurately, program debugging, and testing through ranges of parameter values to insure computational stability (Leonard and Knisel, 1990). Model calibration is "a process involving iterative adjustments of selected parameter values in efforts to obtain a 'best fit' comparison between simulated results and actual data" (Leonard and Knisel, 1990). Calibrations should be performed with data that are different from that used for validation. Ideally, models could be used without calibration, but this would require that all the relationships that are incorporated into them are physically-based, which is not the case with current state of the art NPS models.

It should be noted that even though the above definitions are now in a standard, there are instances in the literature where the terms verification and validation are essentially switched in their meanings (Thomann, 1982). Other methods of defining model validation or similar terms can be found in Gilmour (1973), Wilmott et al. (1985), Parrish and Smith (1990), and Zacharias et al. (1993). The degree to which a model is considered to be valid is often dependent on how specific modeling goals are defined (Leonard and Knisel, 1990; ASTM, 1992). While the definitions given earlier seem rather straightforward, there is much debate in the hydrologic and solute modeling communities as to what criteria are used to determine a validated model. Questions that are commonly raised about model validity include, how accurate the model's predictions have to be or can a model be validated to any given degree of accuracy at all (Leonard and Knisel, 1990; Konikow and Bredehoft, 1992).

In order to evaluate the validity of a model, both statistical and graphical displays should be used (Leonard and Knisel, 1990; Loague and Green, 1991; ASCE, 1993; Zacharias et al., 1993). Statistical evaluations of models generally fall into one of three categories: (1) summary statistics such as the mean and standard deviation of the predicted and observed data (Zacharias and Heatwole, 1994), (2) use of a test statistic, for example hypothesis testing, to compare measured data against simulated results (Parrish

and Smith, 1990), and (3) analysis of residual errors (Loague and Green, 1991; Zacharias and Heatwole, 1994). An analysis of residual errors can include the following statistics, maximum error, root mean square error, coefficient of determination, modeling efficiency, and coefficient of residual mass. The equations to describe these statistics as they appear in Loague and Green (1991) are:

Maximum error (ME),

$$ME = \mathrm{Max}|P_i - O_i|_{i=1}^{n} \tag{1}$$

Root mean square error (RMSE) (Zacharias and Heatwole, 1994),

$$RMSE = \left[\sum_{i=1}^{n}(P_i - O_i)^2 / n\right]^{0.5} \tag{2}$$

Normalized objective function (NOF),

$$NOF = \left[\sum_{i=1}^{n}(P_i - O_i)^2 / n\right]^{0.5} * \frac{100}{\overline{O}} \tag{3}$$

Coefficient of determination (CD),

$$CD = \sum_{i=1}^{n}(O_i - \overline{O})^2 / \sum_{i=1}^{n}(P_i - \overline{O})^2 \tag{4}$$

Modeling efficiency (EF),

$$EF = \left(\sum_{i=1}^{n}(O_i - \overline{O})^2 - \sum_{i=1}^{n}(P_i - O_i)^2\right) / \sum_{i=1}^{n}(O_i - \overline{O})^2 \tag{5}$$

Coefficient of residual mass (CRM),

$$CRM = \left(\sum_{i=1}^{n}O_i - \sum_{i=1}^{n}P_i\right) \Big/ \sum_{i=1}^{n}O_i \tag{6}$$

where P_i are the predicted values; O_i are the observed values; n is the number of samples; and $\overline{O}$ is the mean of the observed data. The limits for these equations include zero as the lower limit for ME, NOF, and CD statistics and one as the maximum value for the EF statistic. Both EF and CRM can become negative. The NOF is the root mean square error (RMSE) normalized to the overall mean. The NOF is similar to the coefficient of

variation and is dimensionless. The RMSE is the sum of squares difference normalized to the number of observations and is a commonly used objective function for model evaluation (Zacharias and Heatwole, 1994). The CD is a measure of the proportion of the total variance of the observed data explained by the predicted data. Ideally, if the predicted and observed data matched exactly, the ME, RMSE, NOF, CD, EF, and CRM would equal 0.0, 0.0, 0.0, 1.0, 1.0, and 0.0, respectively (Loague and Green, 1991). If the EF is less than zero, the actual model-predicted values are worse than the observed mean (Loague and Green, 1991). Other lists of statistics for the determination of model error for hydrologic and solute transport modeling can be found in James and Burges (1982), Thomann (1982), Wilmott et al. (1985), Green and Stephenson (1986), and ASCE (1993). For solute transport models, it has been proposed that models are valid if they can make predictions within a factor of 2 of the true value (Parrish and Smith, 1990).

The types of graphical comparisons that might be used in evaluating solute transport models include: (1) comparison of observed and predicted concentration profiles; (2) comparison of ranges and medians of integrated values of predicted and observed data; (3) comparison of matched predicted and observed integrated values; and (4) comparison of cumulative distribution functions for integrated values (Loague and Green, 1991). Graphical technique (1) can be used for evaluating model performance at specific sites, while the last three techniques can evaluate model performances at several sites at once (Loague and Green, 1991). Systematic error can be evaluated by methods (2) and (3) while spatial variations are represented by (4).

VALIDATION METHODS ACCOUNTING FOR PARAMETER UNCERTAINTY

In the course of validating hydrologic and solute transport models, differences between the observed and predicted results, or errors, will be encountered. According to Haan (1989) these errors can be grouped into three categories:

(1) Inherent variability in natural processes.

(2) Model uncertainty.

(3) Parameter uncertainty.

Other classifications of sources of error have been put forth by Luis and McLaughlin (1992).

Inherent variability in natural processes refers to temporal and spatial variability such as changes in weather patterns from year-to-year and the great variability, up to several orders of magnitude over several meters, in soil properties such as saturated hydraulic conductivity. Also included in this category are the inaccuracies inherent in the measurement of these values. The second category addresses the limitations and simplifications involved in mathematically representing environmental phenomena. The third category refers to the huge number of possible numerical combinations that can be created among input values and the possibility of the same model results coming from entirely different input combinations (Beven and Binley, 1992).

The standard technique for accounting for variability has been to conduct a sensitivity analysis on the inputs and then perform a calibration of the most sensitive inputs in order to get outputs to match a set of measured values. A sensitivity analysis is conducted by establishing a base scenario of inputs and then varying them one at a time, usually by set increments (say, 5, 10, or 20%) to determine what percent change is obtained from outputs due to the given percent change in inputs. The percent change of the output gives an indication on the amount of variation that can be expected in an output due to input parameter uncertainty. The results of the sensitivity analysis can then be used to determine the accuracy and precision of a given modeling scenario (Kumar, 1995a). Calibration is varying inputs until a best fit between model outputs and measured values is obtained. Often times these adjustments are made manually using modeling expertise within the appropriate range of observed values for inputs and environmental characteristics (Zacharias, 1992). More objective methods for calibration generally involve the minimization of differences between observed and predicted values (Haan,

1989). These techniques include the method of moments, least squares, maximum likelihood, arbitrary objective functions, and Bayesian estimation (Haan, 1989).

Several techniques other than the standard validation techniques for deterministic models just described have been explored. In the case of variation in annual model outputs due to long term climatic variability, error has been accounted for by performing statistical analyses on the results from long term continuous simulations of models (Davis and Heatwole, 1990).

Attempts have been made at separating errors that are due to parameter uncertainty from those of model errors (Luis and McLaughlin, 1992; Haan et al., 1995; Kumar and Heatwole, 1995). The separation of parameter uncertainty from model uncertainty has focused on two methods, Monte Carlo Simulation (MCS) and First Order Uncertainty Analysis (FOUA).

With MCS, model inputs that are considered random are characterized by appropriate probability distributions. When the model is run in a Monte Carlo mode, inputs are randomly sampled a predetermined number of times from the probability distributions that define them. From the MCS, distributions of outputs are created that represent the entire range of possible results from the model. From this range of output solutions, characteristic statistics such as the mean and variance for outputs can be obtained. Examples of applications of MCS can be found in Warwick and Cale (1986), Carsel et al. (1988a,b), and Kumar and Heatwole (1995).

Warwick and Cale (1986) used MCS to investigate errors in model predictions due to input data uncertainty for predictions of in-stream pollutant concentrations. Carsel et al. (1988a,b) used MCS to assess regional leaching potentials for aldicarb applications to crops in Ohio and North Carolina. In these studies, the 90th, 95th, and 99th percentiles for pesticide movement past given depths were obtained. This study also found that the predicted outputs for the simulations began to converge after 500 simulations. Dean et al. (1989) included a MCS shell with their release of the RUSTIC model. The primary purpose of this shell was to estimate the uncertainty in model output given the uncertainty

in the input parameters for the fate and transport model. Beven and Binley (1992) used a Monte Carlo type procedure combined with general likelihood procedures in what they termed the Generalized Likelihood Uncertainty Estimation procedure. This procedure assigned a likelihood value to each set of input parameters which indicated how likely it was that the set was a simulator of the system being modeled. Luis and McLaughlin (1992) proposed a model validation procedure that used MCS and a series of hypothesis tests which are used to assess a model's ability to "predict the mean distribution of moisture content over time and space." This procedure assumed normal distributions for the variables of interest and used an indicator of the central tendency (mean) of model output distributions for comparisons rather than the entire empirical data set. Both of these traits of the Luis and McLaughlin (1992) procedure are often undesirable for the evaluation of NPS models (Kumar, 1995a). Assuming normal distributions for the variables of interest is undesirable because the assumption often does not hold. Using indicators of central tendency of data sets may not be warranted for NPS model evaluation because the distribution component of interest is usually the extreme tails.

Haan et al. (1995) used MCS to create probability density functions for a validation procedure that creates confidence intervals around model predictions. Kumar and Heatwole (1995) proposed a general model validation procedure for taking parameter uncertainty into account that also uses MCS. This procedure takes into account some possible problems in other validation procedures by accounting for multivariate correlations, eliminating the need for large numbers of hypothesis tests, allowing for comparisons between management practices, and taking model output variance into account.

Another technique for determining the error in the outputs of a model as caused by inputs errors is FOUA. First Order Uncertainty Analysis involves propagation of uncertainty in model inputs through a model with first order terms of a Taylor series expansion about the mean value of each input (Kumar et al., 1992). The theoretical basis of this procedure is defined by Benjamin and Cornell (1970). Applications of FOUA for

determining the uncertainty of outputs as caused by uncertainty in inputs can be found in Dettinger and Wilson (1981) and Loague et al. (1990). Loague et al. (1989, 1990) used FOUA to characterize uncertainties in leaching potentials of pesticides to groundwater in Hawaii.

There have been several studies to compare the results of uncertainties obtained by MCS and FOUA (Scavia et al., 1981; Smith and Charboneau, 1990). In these studies and the studies cited for each individual uncertainty estimation method, certain strengths and weaknesses of each method have become apparent. Monte Carlo simulations are computationally intensive, but MCS includes the entire relationships between the inputs and outputs. FOUA, on the other hand, requires only one computation for each input. Additionally, however, FOUA tends to lose accuracy due to its lack of higher order Taylor series terms when the models involved are nonlinear or if the coefficient of variation of any model input is > 10 - 20% (Zhang et al., 1993), which is characteristic of many soil inputs. An additional drawback to MCS is that it assumes complete representation of model parameter distributions, which may not always be the case (Zhang et al., 1993). An advantage for FOUA, where it can be applied, is that the amount of error in an output from each individual input can be determined.

CHARACTERIZATION OF INPUT PARAMETERS

As noted previously, one of the inherent problems in modeling is characterizing model input variables that vary at the field-scale or other scales. Some of the typical model inputs that can exhibit these variations include static soil properties (porosity and bulk density), water transport properties (saturated hydraulic conductivity and infiltration rate) (Jury, 1985), runoff curve number, and pesticide application and degradation rates.

Typically, in order to show input parameter variability, data for the given parameter is obtained either from a single field-scale study or from linear regression equations, which relate one input parameter to another (Rawls and Brakensiek, 1985).

Data from either source can then be fitted to a probability distribution. Typical distributions used for model parameters include normal, lognormal, gamma, triangular, and uniform distributions, (Carsel et al., 1988b; Kumar, 1995a) or other distributions as can be found in Law and Kelton (1991). Generally, the distribution that best fits the data is determined by an application of the Kolmogorov-Smirnov goodness-of-fit test or other types of goodness-of-fit tests (Carsel et al., 1988a; Kumar, 1995a).

If no field data are available for a given input, but a range of values can be located in the literature it has been assumed that the range equals four times the standard deviation of the data set (Kumar, 1995a). Additionally, common sense must be used when defining the distributions so as not to exceed the physical limits of the input parameter. Parameters that typically need constraints placed on the range of their values include total porosity, field capacity, and wilting point, which all cannot exceed 0.99 due to the physical impossibility of such a value (Kumar, 1995a). If all else fails, model users can use their best judgment and apply uniform or triangular distributions, based on data ranges, where necessary.

MODELING PESTICIDE FATE IN SOIL

Once a pesticide has been applied to an area of land, it can be degraded, either to a metabolite that is still toxic or to some nontoxic form, retained in soil, volatilized to a gaseous form, or transported away from the site. Even within the pesticide application process there are possible losses of chemicals. These losses can be affected by the nature of the pesticide formulation, atmospheric conditions, method of application, and sprayer droplet size (Zacharias, 1992). Degradation of the pesticide, while it is still on the crop canopy, crop residue, or the top of the soil surface, could come from the sun in a process known as photodegradation. Once the pesticide has reached the surface layer of the soil, one of its possible fates could be in runoff. Runoff occurs on a site when the infiltration capacity of a soil is exceeded. Four possible mechanisms for moving pesticides from the

surface layer into runoff are "(i) diffusion and turbulent transport of dissolved pesticide from soil pores to the runoff stream; (ii) desorption from soil particles into the moving liquid boundary; (iii) dissolution of stationary pesticide particulates; and (iv) scouring of pesticide particulates and their subsequent dissolution into the moving water" (Leonard, 1990 - originally Bailey et al., 1974).

One of the key relationships that is characterized mathematically for the overland flow transport of pesticides is the distribution of pesticide between the solution and adsorbed phases. One relationship that is typically used to describe this process is the Freundlich isotherm, which is defined as (Leonard, 1990):

$$S = KC^N \tag{7}$$

where S is adsorbed-phase concentration, C is the solution-phase concentration, and K and N are constants specific to the pesticide-soil combination. Other relationships which can be used for the representation of the adsorbed / solute phases include the Langmuir isotherm and the linear isotherm (Novotony and Olem, 1994). Both the Freundlich and Langmuir isotherms can reduce to the linear isotherm (Novotony and Olem, 1994).

One important idea to be considered when dealing with the transport of pollutants with sediments is the enrichment ratio. An enrichment ratio is defined as the concentration of the contaminant in the eroded material in runoff divided by the contaminant's concentration in the parent topsoil expressed on an oven dried basis (Novotony and Olem, 1994). In the cases of certain environments and pollutants, it has been found that the enrichment ratio is greater than 1, meaning that there is more pollutant on the sediment in the aquatic environment than there was attached to the parent soil (Dean, 1983; Novotony and Olem, 1994).

One of the most important factors in pesticide transport in runoff, whether it be in the liquid or adsorbed form, is the length of time between the application of the pesticide and the first runoff causing rainfall event. The sooner this event occurs, especially if it is within a day or two of pesticide application, the higher the concentration of pesticides in the runoff (Triplett et al., 1978; Baker and Johnson, 1979).

If the pesticide is not photodegraded or carried away with runoff, it could enter the soil with infiltration and be affected by chemical and biological degradation or the transport processes of leaching and diffusion. Along with these processes, the adsorption / desorption process as defined for runoff also plays a key role in holding pesticides in the soil matrix where degradation can occur. Often, the biological and chemical degradation of a pesticide are lumped together under the general title of degradation. Degradation is then represented by first order kinetics through the relation (Wagenet and Rao, 1990; Novotony and Olem, 1994):

$$\frac{dc_d}{dt} = -K_D c_d \tag{8}$$

where c_d = concentration of pesticide in the soil and K_D = a degradation rate constant. Some of the factors that can affect the rates of degradation in a soil are soil temperature and soil moisture (Walker and Zimdahl, 1981).

The movement of pesticides downward through the soil profile, which is known as leaching, is typically broken into two subprocesses, diffusion and mass flow (Zacharias, 1992). Diffusion occurs when a solute (pesticide) in water moves from an area of greater concentration to an area of less concentration (Wagenet and Rao, 1990; Fetter 1993). This process is usually represented by Fick's first law,

$$F = -D_d(dC/dx) \tag{9}$$

where F = mass flux of solute per unit area per unit time, D_d = diffusion coefficient (L^2/T), C = solute coefficient (M/L^3), and dC/dx = concentration gradient ($M/L^3/L$). Pesticides in dissolved form can be transported in what is known as mass flow. Other terms for this process are advective transport and convection (Fetter, 1993). It should be noted that even though in most laboratory studies pesticide transport is by mass flow and/or diffusion, in actual field situations transport can be speeded up by preferential flow through soil macropores (Patni et al., 1987).

Often when modeling pesticide fate, exact concentration profiles are not specified, but summary variables are. Summary variables which can be used to describe a typical

concentration profile include: total mass, center of mass, peak concentration, time for a critical concentration to leach to a depth of interest, and depth to the leaching front (Jones and Rao, 1988; Loague and Green, 1991).

The model used in this study was Groundwater Loading Effects of Agricultural Management Systems (GLEAMS) (Leonard et al., 1987). This model is a continuous, field-scale model which simulates hydrology, erosion, pesticide, and nutrient transport. It is based on the CREAMS model (Knisel, 1980), which was used to model surface losses of pesticides and nutrients, erosion/sediment yield, and runoff. GLEAMS differs from CREAMS in that a subsurface component was added to the model to simulate the movement of pollutants in the plant root zone. These two models have been well used in part due to the extensive documentation associated with the CREAMS manual.

The runoff component of the model uses a version of the SCS runoff equation (Williams and LaSeur, 1976), while the erosion component is a modification of the Universal Soil Loss Equation (Onstad and Foster, 1975). Evapotranspiration is calculated using daily air temperature and solar radiation. To describe the soil profile, the model allows input of descriptions for up to five soil layers. Soils inputs that can vary with the different layers include porosity, field capacity, wilting point, and organic matter content. When performing calculations, though, the model splits the soil profile into between three and twelve computational layers. Water movement is determined from a storage-routing technique using soil parameters, field capacity, wilting point, and saturated conductivity.

Pesticide transport in the subsurface is by advection since no dispersive component is included in the model. The model does not include volatilization. Adsorption of the pesticide to soil particles is describe using the pesticide-organic carbon partitioning coefficient. The degradation of pesticides are determined by using a first-order degradation equation. Other features of this model include its ability to simulate 10 pesticides and their metabolites simultaneously, to partition the application of the pesticide between plant foliage and the soil surface, to simulate four different pesticide application methods, and to provide outputs on a daily, monthly, or annual basis.

PROBABILISTIC EVALUATION

By using probabilistic evaluation procedures in the evaluation of NPS models, the focus of the model evaluation process can be moved from input parameter uncertainty to the model algorithms, equations, and assumptions used in the models. Kumar (1995a) proposed a methodology for the evaluation of NPS models which expands on the probabilistic modeling work proposed by Kumar et al. (1992) and Haan et al. (1995). Kumar's proposed methodology compares model calibration to the idea of a goodness-of-fit (g-o-f) procedure for determining how well the distribution for a given set of data matches the distribution for another set of data or a known probability distribution. This methodology also allows for both subjective (graphical) and objective evaluation of models as proposed by Loague and Green (1991).

The outline of the procedure as defined by Kumar (1995a) is as follows:

(1) Obtain observed data on the output 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.

It should be noted that the fourth step would not be valid if a calibration of the input parameters had been performed. Haan et al. (1995) have also presented a similar four step procedure for model evaluation.

METHODOLOGY

The material presented in this section follows the outline of Kumar's procedure, with the exception that the fourth step has been renamed to "Evaluation Criteria" to encompass all components of this procedure, not just the distributional comparisons.

Observed data

The field data used in this study was from a study conducted by Heatwole et al. (1992). The following description of the data comes from Zacharias and Heatwole (1993).

> 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 x 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 6 m 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 1.5 m 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.

The field data used for this study consisted of runoff, sediment yield, surface losses, solute mass in the root zone, and depth of center of solute mass, where available, for the pesticides atrazine and metolachlor and the tracer bromide. Surface losses are the combination of pesticides attached to sediment and in solution in runoff leaving a plot. A listing of these data appears in tables A-1 through A-13 in Appendix A. In this analysis, values of zero in the mass data set were replaced by the average of zero and the method detection limit. Solute mass was converted from the concentration data through unit conversion, while depth of center of solute mass was calculated using the formula,

$$z_c = \frac{\sum_{i=1}^{n} C_i d_i z_i}{\sum_{i=1}^{n} C_i d_i} \tag{10}$$

where z_c = depth of center of mass of chemical, C_i = chemical concentration in the ith layer, d_i = thickness of ith layer, and z_i = depth to center of ith layer from the soil surface (Zacharias and Heatwole, 1994).

Model Inputs

The inputs used for the Monte Carlo simulation runs of GLEAMS for the two management practices were obtained from Kumar (1995a). A complete description of these inputs and how they were obtained appears as Appendix B. The meteorologic inputs came from the previously mentioned field study (Heatwole et al., 1992) and a nearby class A weather station (Mostaghimi et al., 1989). The root zone was described using four horizons based on data from the county soil survey (SCS, 1982). Other inputs and input distributions for crop-related, soil, and chemical parameters came from a combination of site data, literature sources, and modeling expertise. Inputs considered as random variables appear in Table 1, with associated distribution parameters given in Table 2. Table 3 lists values for parameters that did not vary between the two management practices, while values for parameters that varied

Table 1. Inputs considered random variables and abbreviations. (Kumar, 1995a)

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 2. Distributions used for inputs in model evaluation. (Kumar, 1995a)

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

Table 3. Distributions and parameters common to conservation and conventional tillage simulations. (adapted from Kumar, 1995a)

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)	(-1.600,0.275)
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().

between management practices are in Tables 4 and 5. The input correlations used in generating the random variates appear in Table 6.

The random variate streams for the two management practices were generated by Kumar (1995b) using a general-purpose Monte Carlo pre-processor (Kumar et al., 1994). Separate random number data files were created for each of the three GLEAMS input files, hydrology, erosion, and pesticide, for both management scenarios, needed for this study.

Model Output Distributions

In order to generate the model output distributions, GLEAMS was executed in a batch mode as described by Kumar et al. (1994). This execution method used a DOS batch file that calls FORTRAN executable programs and executes DOS commands. The FORTRAN programs written for this procedure prompts the user for the number of trials, stores the total number of trials to be run and the current trial number in an ASCII data file, checks to see if the total number of trials has been reached, pulls random variates from the random number variate file and places them in the correct location in the GLEAMS input files, and extracts output variables of interest for observed dates and places them in a summary output file. Five thousand trials were conducted in this study using GLEAMS version 2.1.

Output distributions for mass in each of the four GLEAMS output layers, runoff, and sediment yield were generated previously by Kumar (1995b). Subsequent MCS runs were made to create output distributions for pesticide surface losses and depth of center of mass for both pesticides and bromide. All runs were made using the same input random number stream file.

Evaluation Criteria

The evaluation of results incorporate both graphical (qualitative) and quantitative assessment of the model as encouraged by Loague and Green (1991). Three types of

Table 4. Distributions and parameters used to simulate differences between conservation and conventional tillage systems - conservation tillage inputs. (adapted from Kumar, 1995a)

Input (units)*	Conservation tillage	
	Distribution	Parameter values
POR(1)	N(μ,σ)	(0.435,0.0534)
FC(1)	N(μ,σ)	(0.1962,0.0145)
FC(2)	N(μ,σ)	(0.2400,0.0293)
FC(3)	N(μ,σ)	(0.2589,0.0232)
FC(4)	N(μ,σ)	(0.2235,0.0628)
BR15(1)	N(μ,σ)	(0.0521,0.00459)
BR15(2)	LN(μ,σ)	(-2.582,0.235)
BR15(3)	N(μ,σ)	(0.1091,0.0213)
BR15(4)	N(μ,σ)	(0.0801,0.0322)
OM(1) (%)	N(μ,σ)	(0.948,0.185)
OM(2) (%)	LN(μ,σ)	(-0.726,0.432)
OM(3) (%)	N(μ,σ)	(0.3172,0.125)
OM(4) (%)	LN(μ,σ)	(-1.622,0.6025)
S (mm)	LN(μ,σ)	(4.892,0.524)
CFACT(1)	T(a,b,c)	(0.297,0.363,0.330)
CFACT(2)	T(a,b,c)	(0.297,0.363,0.330)
CFACT(3)	T(a,b,c)	(0.261,0.319,0.290)
CFACT(4)	T(a,b,c)	(0.225,0.275,0.250)
CFACT(5)	T(a,b,c)	(0.162,0.198,0.180)
CFACT(6)	T(a,b,c)	(0.126,0.154,0.140)
CFACT(7)	T(a,b,c)	(0.297,0.363,0.330)
NFACT	LN(μ,σ)	(-1.204,0.269)
FOLFRC	U(a,b)	(0.3,0.88)
APPL-Atrazine (g/ha)	B(α_1, α_2,a,b)	(1.231,1.515,0.52,1.49)
APPL-Metolachlor (g/ha)	B(α_1, α_2,a,b)	(1.701,3.968,0.46,2.35)

* No units indicate input is dimensionless

Table 5. Distributions and parameters used to simulate differences between conservation and conventional tillage systems - conventional tillage inputs. (adapted from Kumar, 1995a)

Input (units)*	Conventional tillage	
	Distribution	Parameter values
POR(1)	N(μ,σ)	(0.516,0.0534)
FC(1)	LN(μ,σ)	(-1.558,0.1203)
FC(2)	LN(μ,σ)	(-1.432,0.1221)
FC(3)	N(μ,σ)	(0.260,0.0231)
FC(4)	N(μ,σ)	(0.225,0.0625)
BR15(1)	N(μ,σ)	(0.051,0.00644)
BR15(2)	LN(μ,σ)	(-2.577,0.232)
BR15(3)	N(μ,σ)	(0.1096,0.0213)
BR15(4)	N(μ,σ)	(0.0807,0.0323)
OM(1) (%)	N(μ,σ)	(0.817,0.240)
OM(2) (%)	N(μ,σ)	(0.551,0.235)
OM(3) (%)	LN(μ,σ)	(-1.072,0.439)
OM(4) (%)	N(μ,σ)	(0.259,0.195)
S (mm)	LN(μ,σ)	(4.216,0.937)
CFACT(1)	T(a,b,c)	(0.297,0.363,0.330)
CFACT(2)	T(a,b,c)	(0.702,0.858,0.780)
CFACT(3)	T(a,b,c)	(0.585,0.715,0.650)
CFACT(4)	T(a,b,c)	(0.459,0.561,0.510)
CFACT(5)	T(a,b,c)	(0.270,0.330,0.300)
CFACT(6)	T(a,b,c)	(0.225,0.275,0.250)
CFACT(7)	T(a,b,c)	(0.333,0.407,0.370)
NFACT	LN(μ,σ)	(-1.386,0.333)
FOLFRC	Constant	0.0
APPL-Atrazine (g/ha)	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)	(2.825,5.780,0.67,1.34)

* No units indicate input is dimensionless

Table 6. Input correlations used in simulations.
(Kumar, 1995a)

Input variables	Correlation*
POR(i)/FC(i)	0.20
POR(i)/BR15(i)	0.20
POR(i)/OM(i)	0.25 (0.20)
FC(i)/BR15(i)	0.70
FC(i)/OM(i)	0.35 (0.30)
BR15(i)/OM(i)	0.45 (0.35)
FC(j)/FC(j+1)	0.10
BR15(j)/BR15(j+1)	0.10
OM(1)/OM(2)	0.50 (0.20)
OM(2)/OM(3)	0.30 (0.20)
OM(3)/OM(4)	0.60 (0.80)
CLAY(1)/FC(1)	0.70
SILT(1)/FC(1)	0.20
CLAY(1)/BR15(1)	0.70
SILT(1)/BR15(1)	0.20
CLAY(1)/OM(1)	0.30
SILT(1)/OM(1)	0.10
CFACT(k)/CFACT(k+1)	0.30

*If a different correlation value was used for conventional tillage, this is given in parentheses.
i,j = Horizon numbers (i = 1 to 4, j = 1 to 3)
k = Crop rotation period (1 to 7)

outputs were evaluated: variables with multiple observed data points; variables with single observed data points; and, comparison between different management scenarios.

Model outputs in the first category are mass of solute in the root zone and depth of center of solute mass each with 20 samples for a given sampling depth and date and plot. Visual comparisons were made between the model predicted distribution and the empirical distribution function (EDF), with assessment of how much distance separated the two distributions over the entire probability range, whether the empirical distribution was over- or under-predicted, and the similarity in the variances of the two distributions. Goodness-of-fit tests recommended for quantitative assessment by Kumar (1995a) are the Kolmogorov-Smirnov (KS) test, to assess the overall fit between the empirical and predicted distributions, and the Anderson-Darling (AD) test, which was used to assess the fit between the tails of two distributions. Test statistics defined by (Law and Kelton, 1991) are:

Kolmogorov-Smirnov (D) statistic,

$$D = \max\{D_n^+, D_n^-\} \tag{11a}$$

$$D_n^+ = \max_{1<i<n}\left\{\frac{i}{n} - \hat{F}(X_{(i)})\right\} \tag{11b}$$

$$D_n^- = \max_{1<i<n}\left\{\hat{F}(X_{(i)}) - \frac{i-1}{n}\right\} \tag{11c}$$

Anderson-Darling (A^2) statistic,

$$A^2 = (-\frac{1}{n}\sum_{i=1}^{n}\{(2i-1)\ln[\hat{F}(X_{(i)})\{1-\hat{F}(X_{(n-1+i)})\}]\}) - n \tag{12}$$

where $\hat{F}$ is the distribution function of a hypothesized distribution, n is the number of observed data points, and the X_i's are the observed data points sorted in ascending order.

In the case where the range of the empirical distribution exceeded the range of the predicted distribution, the extreme distribution value of the predicted distribution was rounded to 0.999 to eliminate division by zero in the A^2 calculation. Both test statistics were calculated and then categorized according to what p-value range (Law and Kelton, 1991) they fell into. Test statistics that fall into each p-value range can be evaluated based

on their relative significances, that is, a p-value of < 0.85 is more significant than a p-value in the 0.85-0.90 range, and thus indicates better fit between the empirical and predicted distributions. Depending on the specific model application, p-values could be used in conjunction with an alpha level to reject inadequate fits between the observed and predicted distributions. In order to assign an alpha level for determining the adequacy of fit between the distributions, the following considerations need to be made. From the standpoint of the model user, a high alpha value is desirable because it will ensure the best fit between distributions. Good fit indicates more accurate predictions. A lower alpha level would be desired by the model developer because predictions would not have to be as accurate, thus giving the model a better chance of being accepted. Exact alpha levels will also be affected by such things as the possible harmful effects of the chemical being modeled and the potential cost of alleviating these effects.

The second evaluation category was for the case when observed data consisted of a single value. Model outputs which fit this category included runoff, sediment yield, and pesticide surface losses. This category was evaluated with the probabilistic index of model performance, I_p, defined by Kumar (1995a) as:

$$I_p = \log_{10} \frac{p_e}{1-p_e} \tag{13}$$

where p_e is the probability of exceedence of the observed statistic in the predicted distribution. The probability of exceedence is obtained by subtracting the cumulative probability associated with an observed value from one. Perfect model prediction would be indicated by an I_p value of zero, while negative and positive values represent under- and over-prediction, respectively. I_p values with magnitudes between 0.95 and -0.95 would fall within 90 percent of a perfect prediction. Exceedence probabilities of zero or one were approximated by 0.001 or 0.999, respectively, in order to obtain finite I_p values of -3.00 and 3.00. Kumar (1995a) described the use of scatter graphs of computed indices for all spatio-temporal points as a means of comprehensive assessment of model performance. This method of model evaluation was implemented for the second category

of data by placing the indices obtained for each event for an output on a single scatter graph. In the visual evaluation of the I_p values, I_p values from -1 to 1 appear near the middle of the magnitudes of values on the probabilistic index scale, therefore are considered to be moderate under- and over-predictions, respectively, while I_p values on the extreme edges (-3.00 or 3.00) are considered to be extreme under- and over-predictions, respectively.

The third category for evaluation was comparisons between the scenarios for the two management practices. This category had two components in its evaluation, a scatter graph of I_p values and a table of probabilities that the observed sign of the differences between the management practices was correctly predicted. To show differences between observed data sets for two management practices, a value for the central tendency, the median, for the conventional tillage practice was subtracted from the median for the no-till practice. Samples consisting of a number of variates equal to those in the observed data sets were then drawn from the model output distributions. The median for the conventional tillage sample was then subtracted from the median for the no-till sample. The sampling process was performed 250 times for variables with multiple observations and 5000 times for variables with single observed data points in order to create sampling distributions. The probability of exceedence for the observed difference was then obtained from the sampling distribution and an I_p value computed.

The observed difference had two components, a magnitude and a sign, which were evaluated. Ideally, the model would match both the magnitude and the sign of the observed differences. The magnitude portion of this difference was assessed by how close the I_p values on the scatter graph were to zero. The sign of the difference was assessed by the portion of the sampling distribution, expressed as a probability, that fell on the same side of zero as the observed difference. While it is desirable to have the magnitude of the observed difference predicted correctly, management models have been promoted as tools for relative predictions between different management scenarios, which would at a minimum be shown by the correct prediction of the observed sign.

RESULTS AND DISCUSSION

Surface Outputs

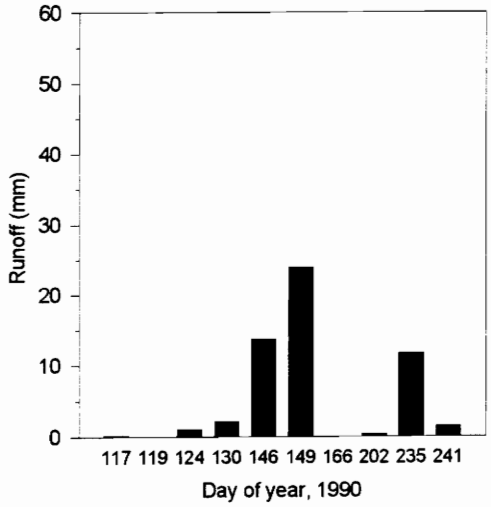
Runoff

Graphs of the magnitude of runoff events and scatter graphs comparing model predictions to the observed events for the no-till and conventional tillage plots appear in Figures 1 and 2. The magnitude graphs accompanied each scatter graph for all surface outputs, because the I_p index does not indicate the magnitude associated with each index value. Figure 3 is a scatter graph comparing runoff predictions for the two management practices, while the probability that the model correctly predicted the observed sign of the difference in runoff between the two management practices appears in Table 7. The graphs of the probability distributions for the observed values versus the predicted distributions appear in Appendix C, while the sampling distributions for relative comparisons appear in Appendix D.

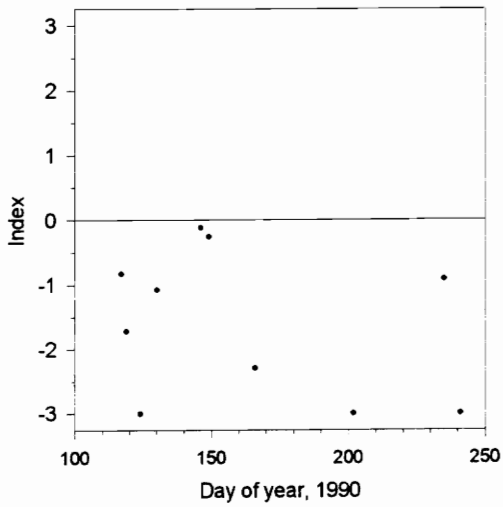
Overall, for both management practices all but two events are under-predicted (Figures 1 and 2). For the three largest events, though, on days 146, 149, and 235, the model performs fairly well with I_p values generally being within 0.5 of a perfect prediction of 0.0. The extreme under-predictions, which are indicated by I_p values of -3.0, tended to occur for observed runoff events that had a total runoff of 5 mm or less.

Under-prediction of runoff could be indicative of the model predicting too much soil moisture storage. Excess soil moisture storage could result from evapotranspiration that was too great or too small of a difference between field capacity and wilting point. Another reason that runoff could be under-predicted is due to too small a value of curve number being used as an input.

The scatter graph of differences between runoff predictions for the two management practices (Figure 3) shows that the model tends to over-predict differences between the predictions for the two management scenarios. As with the absolute value comparisons, the three runoff events with the largest magnitudes, the events on days 146,

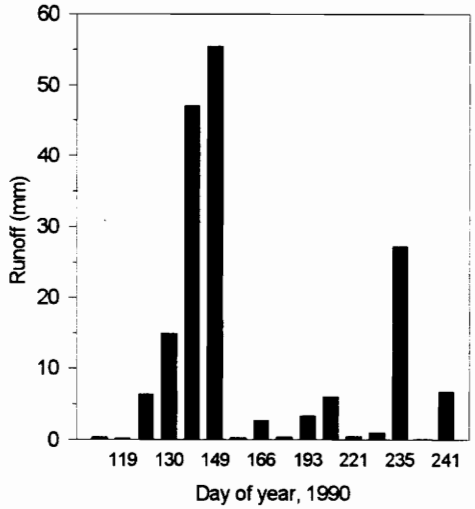


(a)

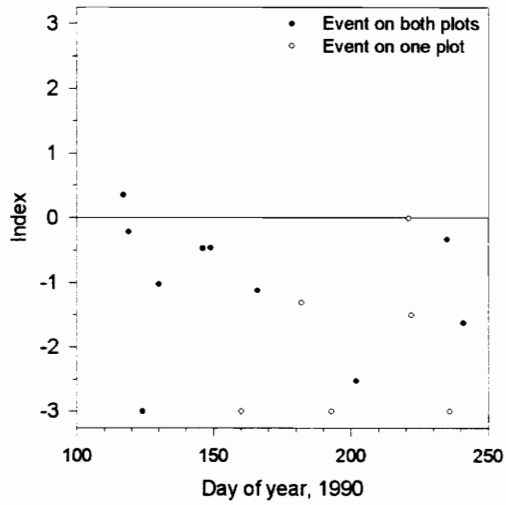


(b)

Figure 1. Runoff from no-till plot, (a) observed and (b) scatter graph of indices of model performance.



(a)



(b)

Figure 2. Runoff from conventional tillage plot, (a) observed and (b) scatter graph of indices of model performance.

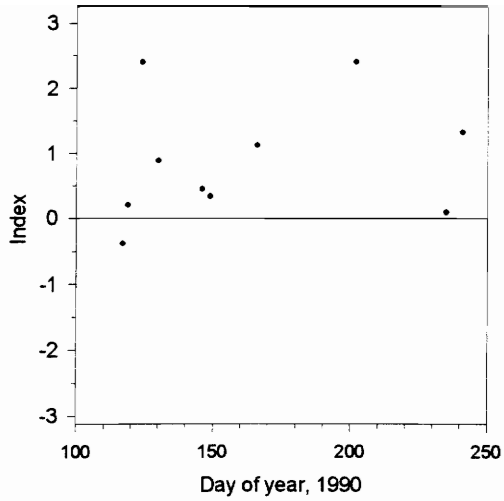


Figure 3. Scatter graph of I_p values for differences in runoff between the no-till and conventional tillage plots as predicted by GLEAMS.

Table 7. Probability of observed sign being correctly predicted by GLEAMS for differences in runoff between the no-till (NT) and conventional tillage (CT) plots.

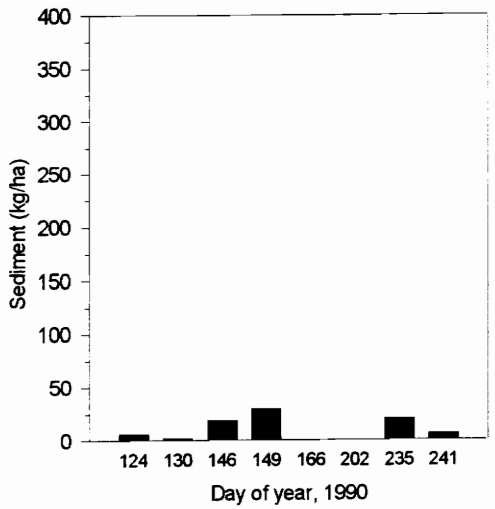
Day	Observed Sign (NT-CT)	Probability of Correctly Predicted Sign
117	-	0.767
119	-	0.516
124	-	0.319
130	-	0.810
146	-	0.834
149	-	0.815
166	-	0.438
202	-	0.272
235	-	0.819
241	-	0.575

149, and 235, produced I_p values that had magnitudes of less than 1.0.

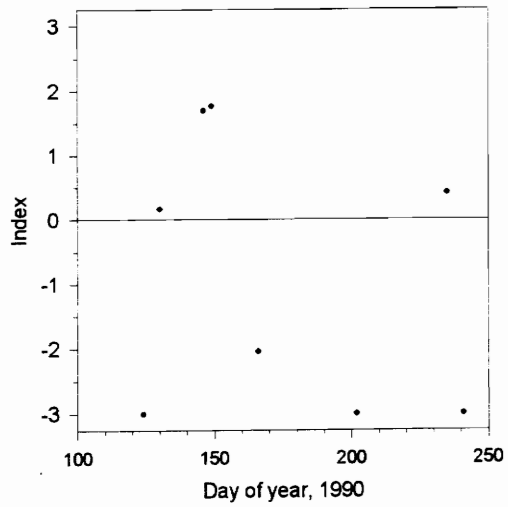
For runoff, the probability that the model predicted the correct sign for differences between management practices ranged from 0.27 to 0.83 (Table 7). Since NPS pollution problems usually come from large events, it is of interest to note that the probability of correct predictions for the days with the three largest events, days 146, 149, and 235, was always greater than 0.81. The probabilities in Table 7 which indicate that use of the model would be worse than random guessing (< 0.5) for relative prediction correspond to events with the smallest magnitudes. These low probabilities could be due to one of two causes. Either the model with these inputs is in error or there is an actual trend on this site for more runoff to occur from the conventional tillage plot than the no-tillage plot. Based on the negative observed differences in runoff between management practices for all dates for this site, the former reason, model error, is the most likely reason for the low probabilities of correct sign differences for runoff.

Sediment Yield

Observed sediment yields and I_p scatter graphs comparing predicted to observed sediment yield are shown in Figures 4 and 5 for the two plots. The predicted distributions used to obtain the I_p scatter graphs appear in Appendix C. These figures show that the model fluctuates between over- and under-prediction, although the under-predictions are more extreme (values of -3.00) than are any of the over-predictions. When comparing between the graphs for the two management practices, several trends in the values are evident in Figures 4 and 5. For instance, on Day 124 both management practices are extremely under-predicted, having an I_p value of -3.00. Next, on days 146 and 149 the model over-predicts with I_p values of approximately 1.5 resulting. Finally, both management practice prediction I_p values on day 202 are extremely low at -3.00, have relative highs at 0 on the next day, and then return to -3.00 at the end of the observed period. The similarity between the trends in the simulated results for the two management practices can be attributed to model algorithms rather than inputs, since (1) model input

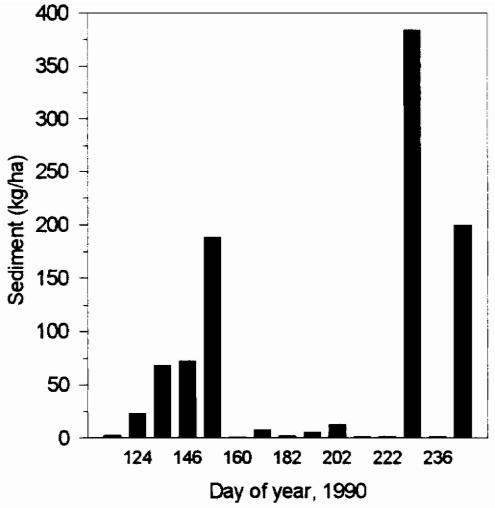


(a)

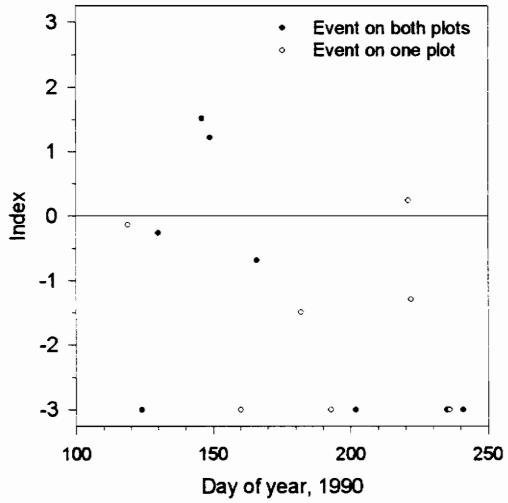


(b)

Figure 4. Sediment yield from no-till plot, (a) observed and (b) scatter graph of indices of model performance.



(a)



(b)

Figure 5. Sediment yield from conventional tillage plot, (a) observed and (b) scatter graph of indices of model performance.

sets were different to show management practice differences and (2) the observed data used for comparison was of two different magnitudes since it came from two different management practices.

The over-prediction of the three largest sediment yield events for the no-till plot is possibly due to the soil erodibility parameter being set too high in the model. On the conventional tillage graph, the initial over-prediction (days 146 and 149) followed by extreme under-prediction (days 235 and 241) does not represent a consistent prediction trend, therefore, these differences may be due to both parameter selection and algorithm errors in the model.

Figure 6 shows a scatter graph of the differences in sediment yield for the two management practices. This graph shows a tendency to over-predict the differences in sediment yield, with the majority of the I_p values being positive or only slightly negative. Figure 6 also shows an inverse relationship to the scatter graph results shown in Figures 4(b) and 5(b).

Table 8 lists the probability that the observed sign of the difference between sediment yield for the two management practices was predicted correctly by the model. These probabilities range from 0.272 to 0.936. For the larger observed sediment yields, which are of the most interest for nonpoint source pollution control, on days 146, 149, and 235, the probability of obtaining the correct sign for the differences between the management practices is above 0.90 for each day. The events with the lowest probability of correctly predicting sign differences were also the events which resulted in the smallest (<25 kg/ha) observed sediment yields (Figure 4 and 5). The reason that the probabilities for these events were so low is likely model error rather than the observed differences actually being positive since all observed events had negative differences.

Atrazine Surface Losses

Bar graphs showing the magnitudes of the surface losses (sediment bound and in solution) of atrazine for the no-till and conventional tillage plots appear in Figures 7(a)

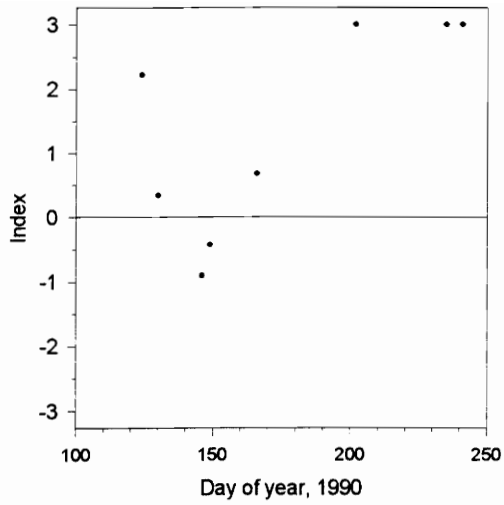
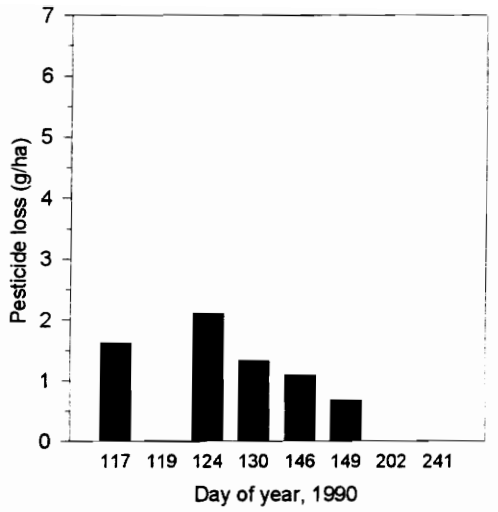


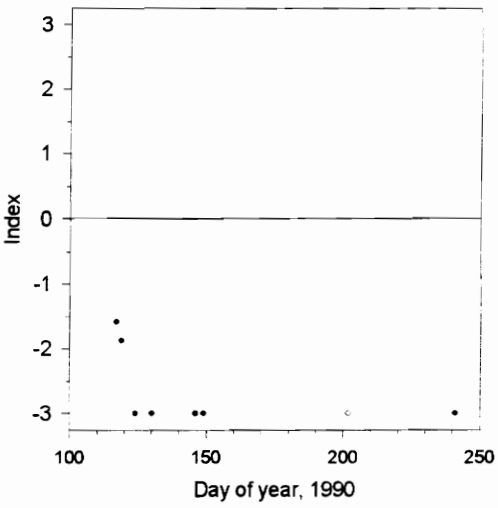
Figure 6. Scatter graph of I_p values for differences in sediment yield between the no-till and conventional tillage plots as predicted by GLEAMS.

Table 8. Probability of observed sign being correctly predicted by GLEAMS for differences in sediment yield between the no-till (NT) and conventional tillage (CT) plots.

Day	Observed Sign (NT-CT)	Probability of Correctly Predicted Sign
124	-	0.319
130	-	0.856
146	-	0.936
149	-	0.936
166	-	0.438
202	-	0.272
235	-	0.900
241	-	0.575

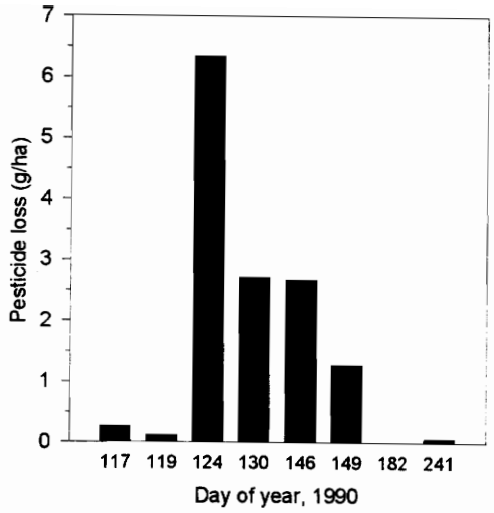


(a)

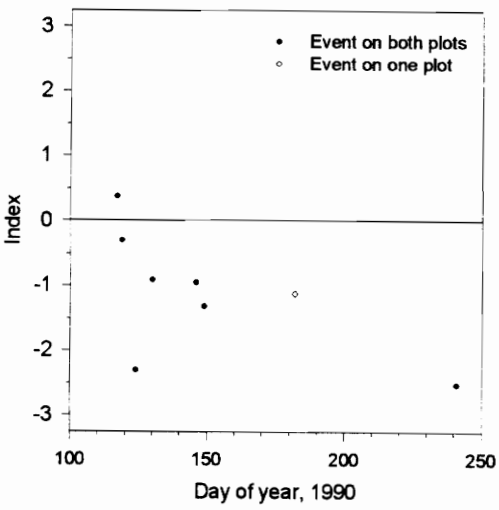


(b)

Figure 7. Surface losses of atrazine from no-till plot, (a) observed and (b) scatter graph of indices of model performance.



(a)



(b)

Figure 8. Surface losses of atrazine from conventional tillage plot, (a) observed and (b) scatter graph of indices of model performance.

and 8(a). Scatter graphs of I_p values comparing model predictions to observed values for atrazine surface losses appear in Figures 7(b) and 8(b). The probability density functions that these I_p values were drawn from appear in Appendix C. The I_p scatter graphs for atrazine surface losses show that the model, with the exception of one date for the conventional tillage plot, under-predicted these losses. In the case of the no-till predictions, the predictions are moderately under-predicted for the first two events and extremely under-predicted for the last six events. Atrazine surface loss predictions for the conventional tillage plot (Figure 8(b)) are over-predicted on earlier dates, then show a period of moderate under-prediction before showing extreme over-prediction on the later dates.

The under-predictions of atrazine surface losses (Figures 7(b) and 8(b)) could be due to over-prediction of pesticide transport or degradation. Rapid leaching of the pesticide could be due to the pesticide partitioning coefficient being too small. Since the prediction trends are similar, the differences in the magnitudes of under-prediction between the two management practices would most likely be due to differences in input parameter selection. These under-predictions could also be linked to under-predictions of the soluble portion of the chemical, which would be due to under-predictions in runoff by the model, which has already been shown to occur.

A scatter graph showing the differences in surface losses of atrazine between the no-till and conventional tillage plots appears in Figure 9. This scatter graph shows that on all but two days the model over-predicted the differences in surface losses between the two management practices. Table 9 lists the probability of a correct sign difference between the predictions for the no-till and conventional tillage plots. The probabilities in this table vary greatly, ranging from 0.081 to 0.812. Not unexpectedly, the lowest probability, 0.081, was on day 117, which exhibited one of the greatest under-predictions on the scatter graph in Figure 9 and which was the only day with positive observed differences. Unlike for the runoff and sediment yield predictions, the event with the largest observed loss did not have a better probability for predicting the correct sign

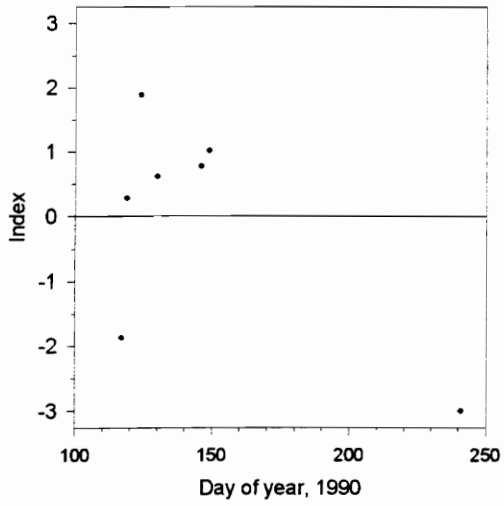


Figure 9. Scatter graph of I_p values for differences in atrazine surface losses between the no-till and conventional tillage plots as predicted by GLEAMS.

Table 9. Probability of observed sign being correctly predicted by GLEAMS for differences in atrazine surface losses between the no-till (NT) and conventional tillage (CT) plots.

Day	Observed Sign (NT-CT)	Probability of Correctly Predicted Sign
117	+	0.081
119	-	0.516
124	-	0.319
130	-	0.812
146	-	0.787
149	-	0.749
241	-	0.136

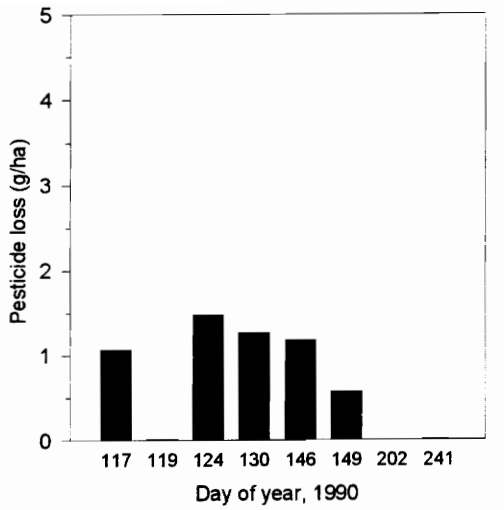
between the two management practices than any of the other atrazine surface loss events.

Probabilities below 0.5 in Table 9 (day 117, 124, and 241) could be caused by several factors. First, the observed sign of the differences could be incorrect. The positive difference for day 117 could be an example of this occurrence, since it is the only one of the seven days listed in this table that has a positive observed sign. Second, the observed sign and the model probability for correct sign prediction could both be correct, indicating that the observed data may just happen to be one of the few instances when this sign would be observed. Third would be a correct observed difference with the low probability caused by model error.

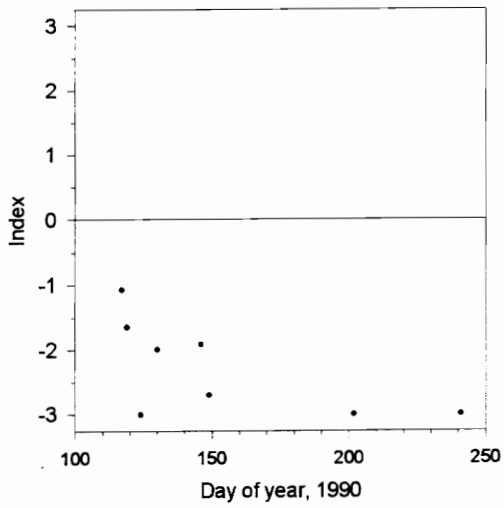
Metolachlor Surface Losses

Surface loss magnitudes for metolachlor are shown in Figures 10(a) and 11(a). I_p scatter graphs for comparisons between the predicted distributions and observed data for the metolachlor surface losses appear in Figures 10(b) and 11(b). In general, these scatter graphs show that GLEAMS under-predicted the surface losses of metolachlor with the exception of the conventional tillage prediction on the first day. Both tillage types exhibit the trend of having a high I_p value for the first event, a lower value for the second event, a pair of high values over the course of the third and fourth events, and lower values on the following days. The tendency for the no-till predictions to be more under-predicted than those for conventional tillage is most likely due to differences in parameter selection between the two tillage practices, while the temporal similarity in prediction trends likely comes from the model algorithms.

A scatter graph showing I_p values for the differences in the surface losses of metolachlor between the no-till and conventional tillage plots appears as Figure 12. This graph indicates that the model has a tendency to over-predict the differences between the two management practices. This graph also seems to follow an inverse pattern as compared to the I_p scatter graphs in Figures 10(b) and 11(b).

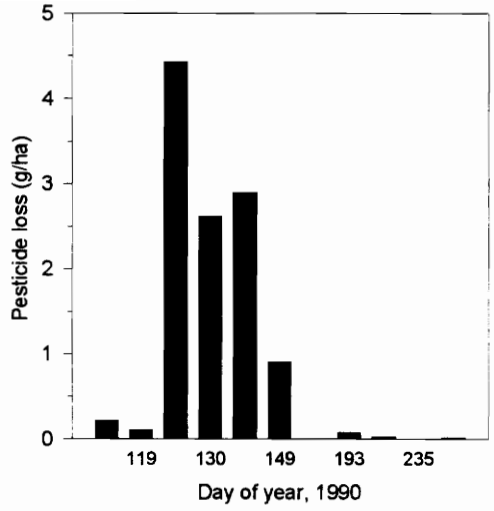


(a)

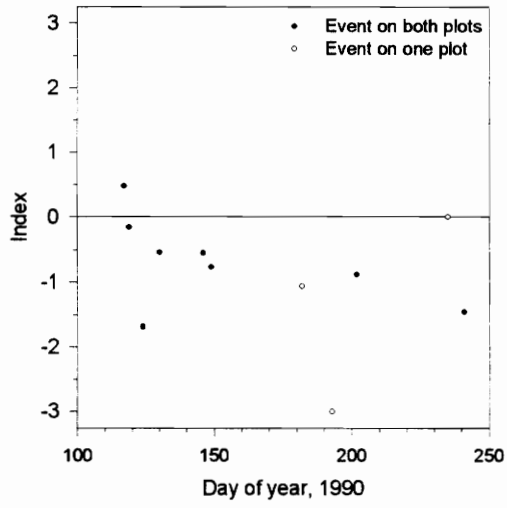


(b)

Figure 10. Surface losses of metolachlor from no-till plot, (a) observed and (b) scatter graph of indices of model performance.



(a)



(b)

Figure 11. Surface losses of metolachlor from conventional tillage plot, (a) observed and (b) scatter graph of indices of model performance.

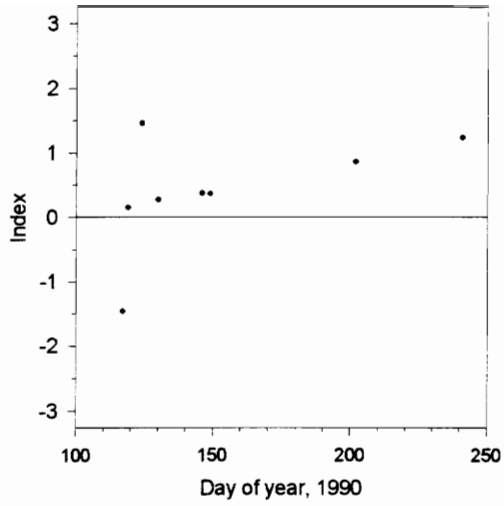


Figure 12. Scatter graph of I_p values for differences in metolachlor surface losses between the no-till and conventional tillage plots as predicted by GLEAMS.

The probability of the predicted differences being correct for differences between the no-till and conventional tillage plots for metolachlor surface losses is presented in Table 10. These probabilities range from a low of 0.081 to a high of 0.800. The probability for a correct prediction on the three days with the largest losses (days 124, 130, and 146) are 0.319, 0.800, and 0.771. Probabilities of correct predictions were less than 0.5 for four days, days 117, 124, 202, and 241. The possible reasons for these low probabilities are the same as for atrazine surface losses. These reasons include, observed error, both observed difference and predicted probability being correct, indicating that the observed value was a possible, but rare occurrence, and model error.

Since the model prediction trends for the metolachlor surface losses are similar to those for the atrazine surface losses, the causes of the under-prediction are also likely similar. The causes of under-prediction listed for atrazine surface losses were likely due to small losses of the soluble form of the pesticide due to the runoff under-predictions, rapid leaching of the pesticide due to the pesticide partitioning coefficient being too small, and over-prediction of pesticide degradation.

Summary

The results of the model evaluation for these outputs indicate that the observed values rarely fell in the middle of the predicted distribution. Runoff and pesticide surface losses were under-predicted by the model. Under-prediction of runoff by GLEAMS has been noted in other studies, and this study serves to further confirm under-prediction of runoff as a trait of GLEAMS. The majority of the observed values used for evaluating the four surface outputs fell within the range of the predicted distribution (graphs in Appendix C), indicating that the model was performing adequately for these outputs.

The relative differences between management practices were generally over-predicted for runoff, sediment yield, and the two pesticide surface losses. The probability that the sign of the model predicted difference between management practices matched the observed sign ranged from 0.081 to 0.94. The lower values in this probability range

Table 10. Probability of observed sign being correctly predicted by GLEAMS for differences in metolachlor surface losses between the no-till (NT) and conventional tillage (CT) plots.

Day	Observed Sign (NT-CT)	Probability of Correctly Predicted Sign
117	+	0.081
119	-	0.515
124	-	0.319
130	-	0.800
146	-	0.771
149	-	0.762
202	-	0.426
241	-	0.258

indicate poor relative predictive performance by the model. However for the largest events for these four outputs, which are of the most interest for NPS pollution control, the relative predictive performance of the model is better. For runoff and sediment yield, the correct relative predictions were made over 80 percent of the time for the three largest events. For the pesticide surface losses, the correct relative predictions for the three largest events were made between 32 and 81 percent of the time. These results indicate that the model makes incorrect relative model predictions some of the time. Incorrect results in the probabilistic procedure indicate that deterministic relative predictions could be in error. Management type models such as GLEAMS not being able to make relative predictions correctly 100 percent of the time has not been previously addressed.

The probabilistic index of model performance showed promise for future use in model evaluation studies by consistently showing similar model prediction trends between parameter selectiom for different model scenarios. An example of this index showing relative temporal trends between management practices can be seen on the scatter graphs for metolachlor surface losses (Figures 10(b) and 11(b)). Since the observed similarities in trends occurred between different data sets, they likely can be related to the model itself, rather than input parameter selection. This index does have its drawbacks though in that it fails to give an indication of the magnitude of the observed value that is behind the calculation of the index value and thus requires an additional graphs to view the magnitudes.

Solute Mass in the Root Zone

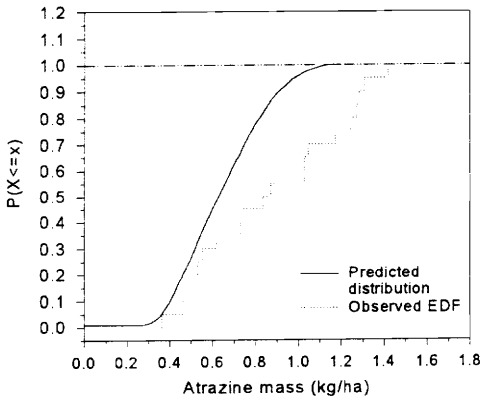
Atrazine

Empirical distribution functions and predicted distribution functions for atrazine mass in the root zone in the no-till and conventional tillage plots appear in Figures 13 and 14, respectively, with associated g-o-f test statistics in Table 11. For the no-till plot, KS p-values were less than 0.99 for two of five days. None of the p-values for the Anderson-Darling statistics for these days was below 0.99, however. The days that had p-values of greater than 0.99 had predicted distributions that were under-predicted and that had a smaller variance than the empirical distribution.

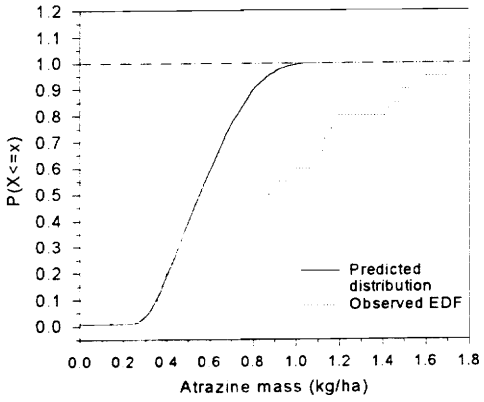
For the conventional tillage data, only one day, day 209, produced a p-value below 0.99 (Table 11). On three of the remaining days, days 118, 128, and 145, the predicted distribution did not have as large a variance as the empirical distribution and tended to slightly over-predict the empirical distribution. For the remaining day, day 272, the predicted distribution had a much smaller variance than the empirical distribution, but suffered most from a large under-prediction of the empirical distribution.

The under-prediction observed for all five days for the no-till distributions (Figure 13) and for the later dates for the conventional tillage distributions (Figure 14) possibly indicates that the model is simulating the degradation of atrazine too quickly, which means that the model inputs for atrazine half-life could be too short or the model algorithms are in error.

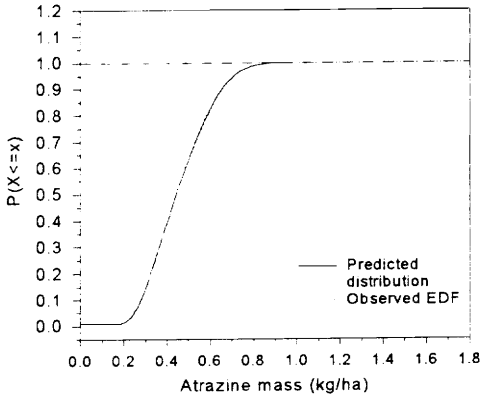
An I_p scatter graph for the differences in atrazine mass remaining in the root zone between the no-till and conventional tillage plots appears in Figure 15. The probability distributions used to obtain these I_p values appear in Appendix D. This plot shows a trend towards under-prediction of the differences between the two tillage types early in the observed period, which changes to extreme over-prediction of the differences in masses for the two tillage types towards the end of the observation period. Table 12 lists the probability that the observed sign of the differences in mass between the two tillage types is predicted correctly. The first two days of the data showed slight probability that the



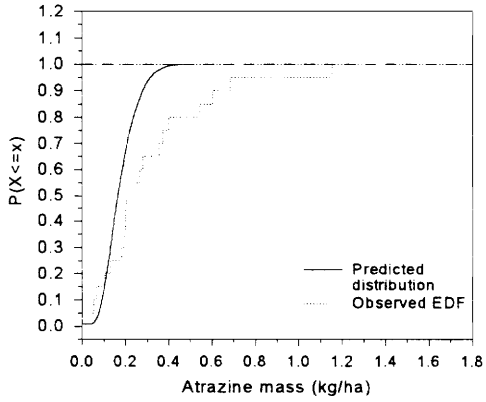
(a) Day 118



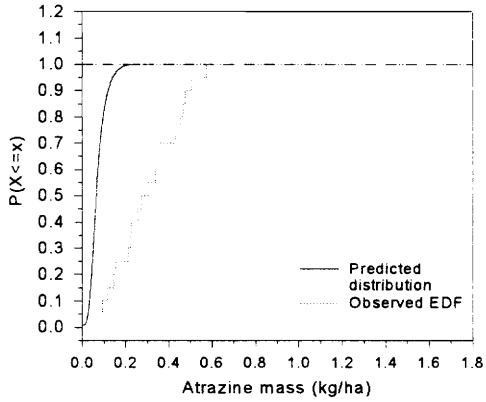
(b) Day 128



(c) Day 145

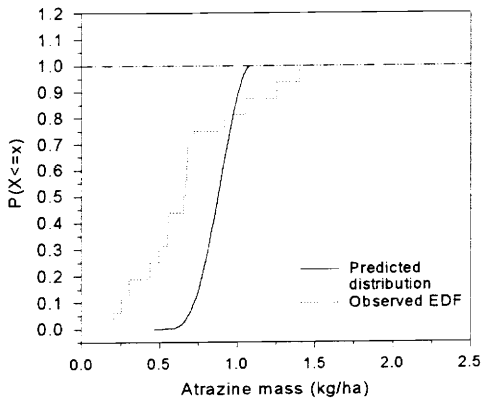


(d) Day 209

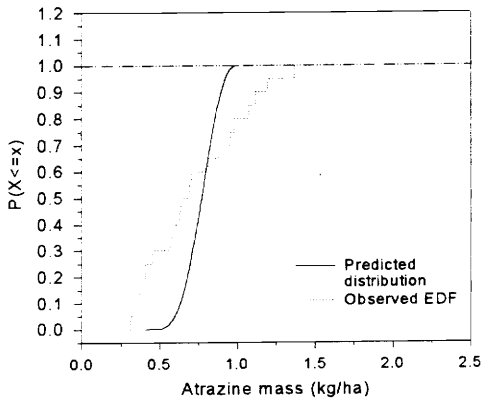


(e) Day 272

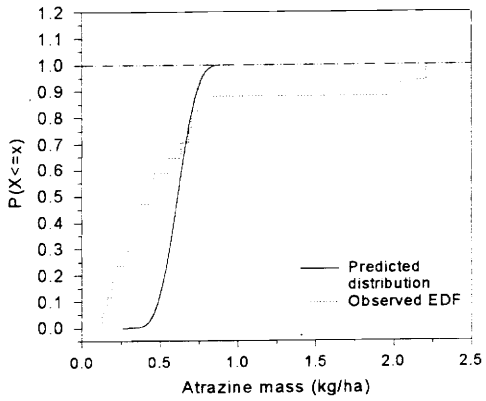
Figure 13. Predicted distribution and observed EDF for mass of atrazine in the root zone of the no-till plot on (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.



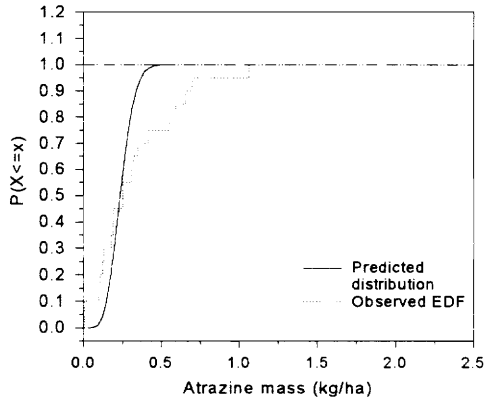
(a) Day 118



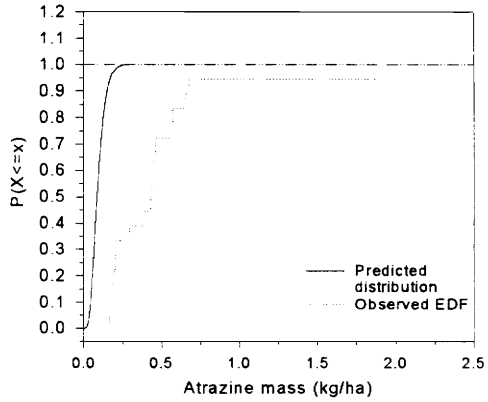
(b) Day 128



(c) Day 145



(d) Day 209



(e) Day 272

Figure 14. Predicted distribution and observed EDF for mass of atrazine in the root zone of the conventional tillage plot on (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.

Table 11. Goodness-of-fit test results for comparing between predicted distribution and observed EDF for atrazine mass in the root zone in (a) no-till plot and (b) conventional tillage plot.

(a) No-till

Day	Adjusted KS Test Statistic	KS P-value*	AD Test Statistic	AD P-value*
118	1.944	>0.99	17.381	>0.99
128	2.307	>0.99	30.000	>0.99
145	1.545	0.975 - 0.99	4.521	>0.99
209	1.583	0.975 - 0.99	12.161	>0.99
272	3.733	>0.99	89.371	>0.99

(b) Conventional tillage

Day	Adjusted KS Test Statistic	KS P-value*	AD Test Statistic	KS P-value*
118	2.704	>0.99	28.089	>0.99
128	1.808	>0.99	20.628	>0.99
145	2.287	>0.99	22.778	>0.99
209	1.311	0.90 - 0.95	13.011	>0.99
272	4.113	>0.99	88.361	>0.99

*Optimum fit between distributions indicated by a p-value of 0.0, while lack of fit is indicated by a p-value of >0.99.

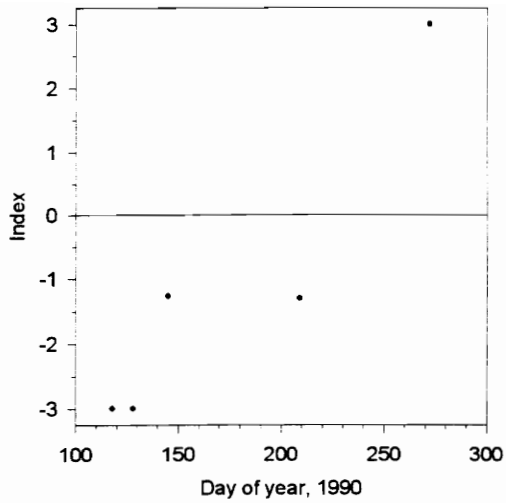


Figure 15. Scatter graph of I_p values for differences in atrazine mass in the root zone between the no-till and conventional tillage plots as predicted by GLEAMS.

Table 12. Probability of observed sign being correctly predicted by GLEAMS for differences in atrazine mass in the root zone between the no-till (NT) and conventional tillage (CT) plots.

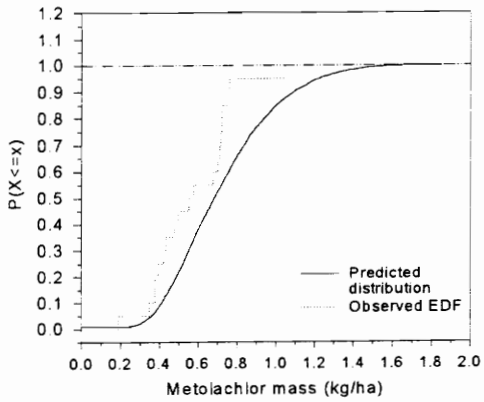
Day	Observed Sign (NT-CT)	Probability of Correctly Predicted Sign
118	+	0.004
128	+	0.004
145	-	0.996
209	-	0.988
272	-	0.940

sign of the differences would be predicted correctly with probability values of 0.004 - less than a 1 percent chance that the predicted sign would be correct. The remaining three days showed an extremely good probability that the sign will be predicted correctly with all three probabilities equaling or exceeding 0.94. The low probabilities of correct relative prediction when the sign of the observed differences between management practice was positive could be due to error in the observed data, correct observed data that actually does occur, but rarely, or errors in the model prediction.

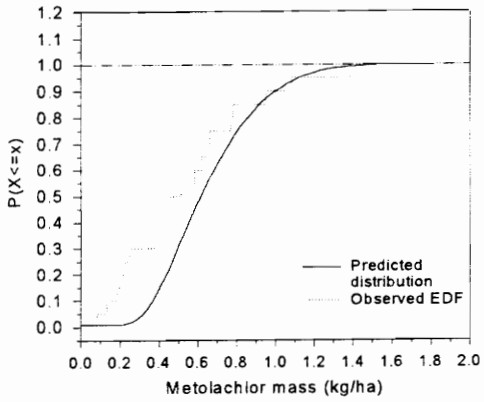
Metolachlor

Figures 16 and 17 show the empirical and predicted distributions for metolachlor mass in the root zone for the no-till and conventional tillage plots. Table 13 shows the KS and AD g-o-f test statistics for metolachlor mass for both plots. For the no-till data, the first two dates had p-values of less than 0.99 for the KS test, while the first date was the only date for any chemical for mass in the root zone which had a p-value of less than 0.99 for the AD statistic. The remaining three days tended to show a large over-prediction of the empirical data and thus the goodness-of fit tests for these dates showed minimum significance with p-values greater than 0.99. None of the days from the conventional tillage data had p-values of less than 0.99 for either the KS test or the AD test (Table 13). In general for all five dates, the observed data was over-predicted with the over-predictions becoming larger with time. Additionally, on the first two dates the variance of the observed data was larger than that of the predicted data, while on the last date the reverse was true.

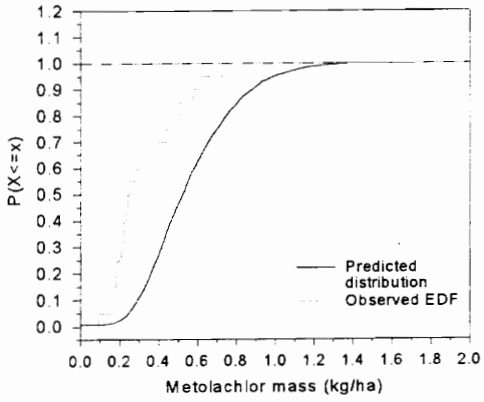
On the majority of the distributional graphs for mass of metolachlor in the root zone, the observed EDF seems to be over-predicted by the predicted distribution (Figures 16 and 17). This over-prediction could be caused by the model not simulating the degradation of metolachlor as fast as it actually occurred. Degradation that is predicted to occur too slowly, as is possible in this case, could possibly be fixed by using a shorter half-life in the inputs to the model.



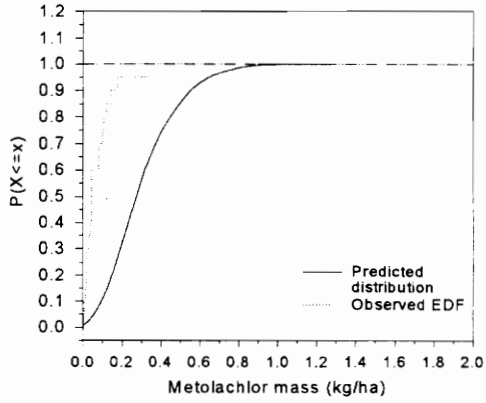
(a) Day 118



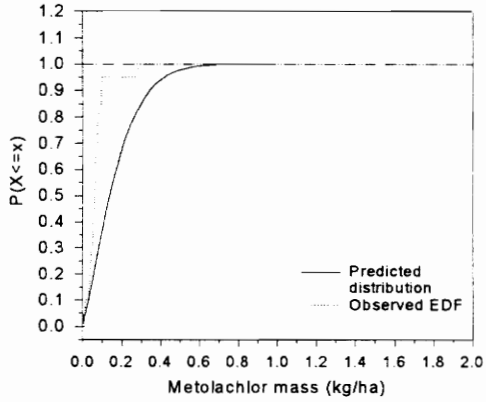
(b) Day 128



(c) Day 145

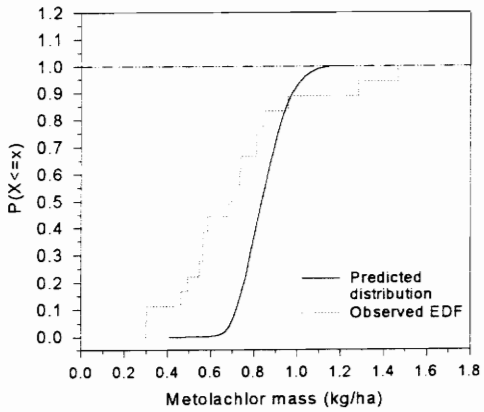


(d) Day 209

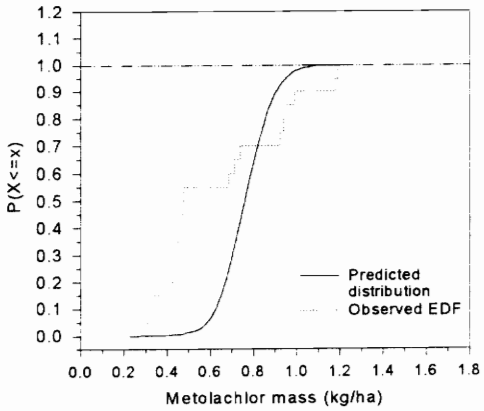


(e) Day 272

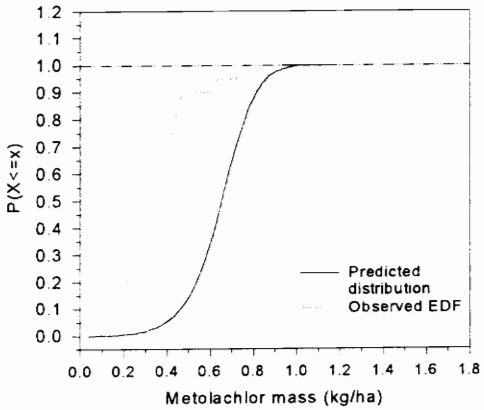
Figure 16. Predicted distribution and observed EDF for mass of metolachlor in the root zone of the no-till plot on (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.



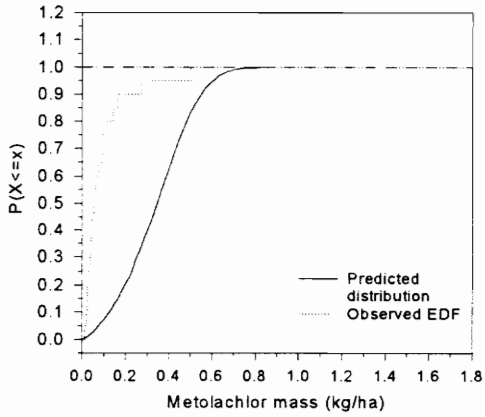
(a) Day 118



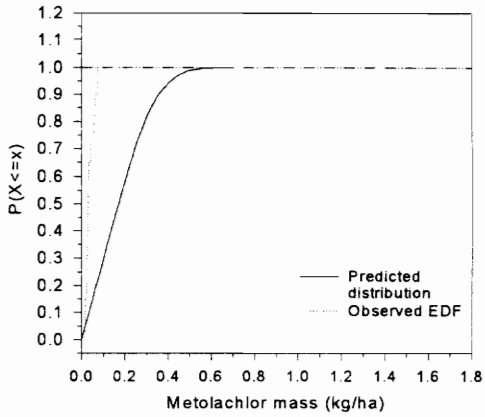
(b) Day 128



(c) Day 145



(d) Day 209



(e) Day 272

Figure 17. Predicted distribution and observed EDF for mass of metolachlor in the root zone of the conventional tillage plot on (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.

Table 13. Goodness-of-fit test results for comparing between predicted distribution and observed EDF for metolachlor mass in the root zone in (a) no-till plot and (b) conventional tillage plot.

(a) No-till

Day	Adjusted KS Test Statistic	KS P-value*	AD Test Statistic	AD P-value*
118	1.608	0.975 - 0.99	3.529	0.975 - 0.99
128	1.336	0.90 - 0.95	5.363	>0.99
145	2.421	>0.99	16.003	>0.99
209	3.293	>0.99	29.096	>0.99
272	2.745	>0.99	9.564	>0.99

(b) Conventional tillage

Day	Adjusted KS Test Statistic	KS P-value*	AD Test Statistic	AD P-value*
118	2.254	>0.99	24.237	>0.99
128	2.504	>0.99	23.616	>0.99
145	3.694	>0.99	34.073	>0.99
209	3.489	>0.99	32.147	>0.99
272	3.607	>0.99	23.246	>0.99

*Optimum fit between distributions indicated by a p-value of 0.0, while lack of fit is indicated by a p-value of >0.99.

A scatter graph of the I_p values for the predicted differences between mass remaining in the soil for the two tillage types appears as Figure 18. The predicted distributions used to obtain these I_p values appear in Appendix D. This figure shows that the model under-predicted the differences in masses in the root zone between the no-till and conventional tillage plots. For three of the days, the under-predictions are moderate with I_p values of approximately -1.0, while on two of the days, days 118 and 145, there was only slight under-prediction with I_p values close to 0.

Table 14 shows the probability that the sign predicted in the distributions used for Figure 18 matched the sign of the observed data. Three of the dates showed much better probabilities of predicting the correct sign than the rest, with probabilities on these dates all being above 0.878. The observed sign of the differences between management practices on these dates was negative, which meant that the mass of metolachlor remaining in the root zone was lower in the no-till plot than in the conventional tillage plot. On the dates when the observed mass of metolachlor in the no-till plot was greater than that in the conventional tillage plot, producing a positive difference, the predicted probabilities were quite low. The probabilities on these two days were 0.044 and 0.264. As mentioned with previous outputs, there are several reasons for low probabilities for correct relative predictions. These reasons include model error, error in measured data, and the chance that this is one of the rare instances that a positive difference occurs in the observed data.

Bromide

The empirical and predicted distributions for bromide mass in the root zone for both no-till and conventional tillage are shown in Figures 19 and 20. The statistics for the g-o-f tests for these distributions are in Table 15. For the no-till data, days 128 and 145 have p-values of less than 0.975. For the remaining KS tests and all the Anderson-Darling tests, the p-values were above 0.99. Even though the goodness-of-fit tests for days 128 and 145 showed some significance, the graphs of the data distributions (Figures 19 and 20) show that the model over-predicted the empirical distributions on these days. The

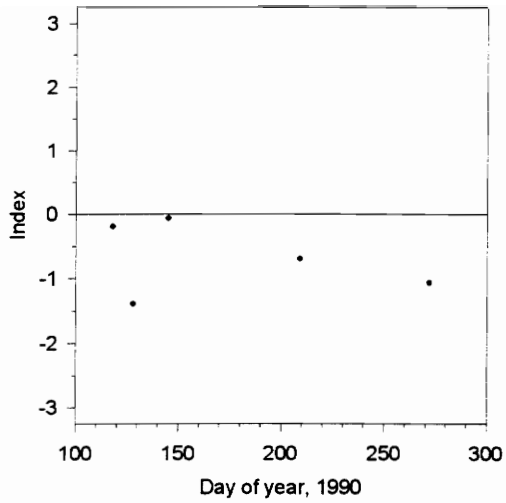
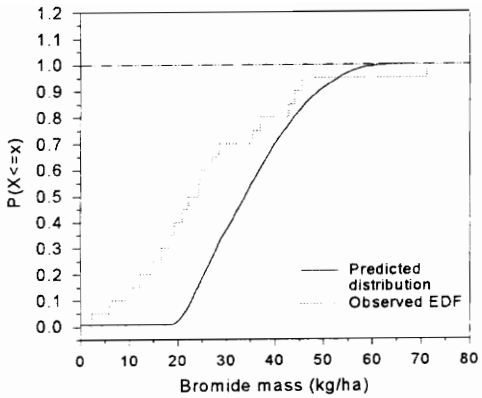


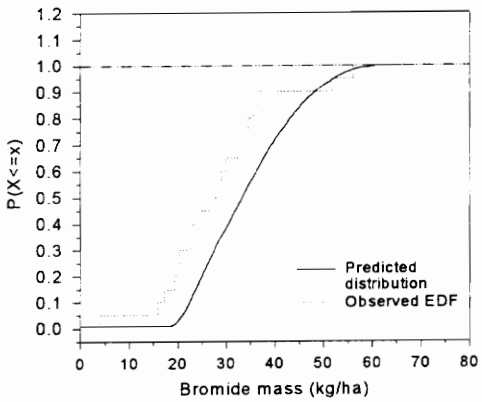
Figure 18. Scatter graph of I_p values for differences in metolachlor mass in the root zone between the no-till and conventional tillage plots as predicted by GLEAMS.

Table 14. Probability of observed sign being correctly predicted by GLEAMS for differences in metolachlor mass in the root zone between the no-till (NT) and conventional tillage (CT) plots.

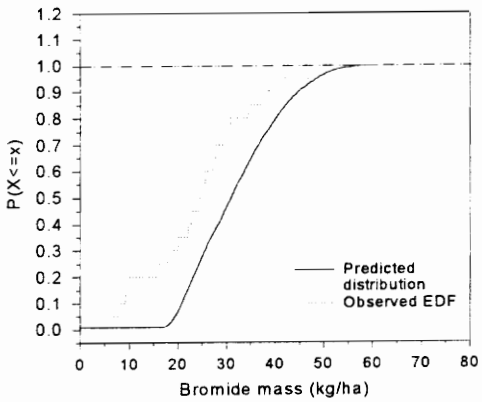
Day	Observed Sign (NT-CT)	Probability of Correctly Predicted Sign
118	-	0.944
128	+	0.044
145	-	0.946
209	-	0.878
272	+	0.264



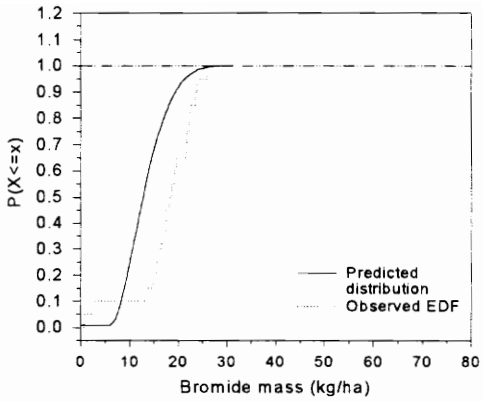
(a) Day 118



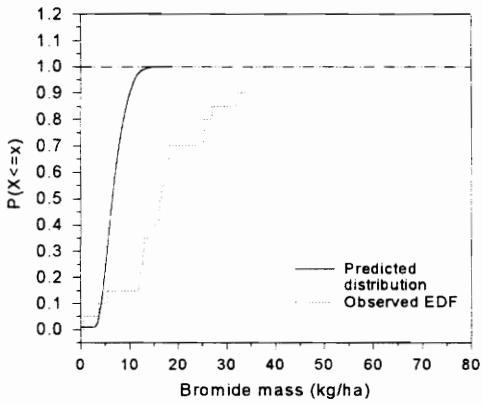
(b) Day 128



(c) Day 145

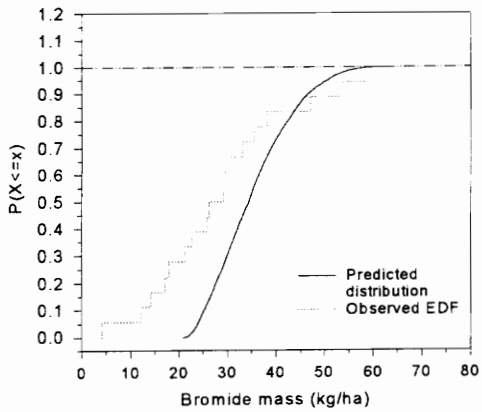


(d) Day 209

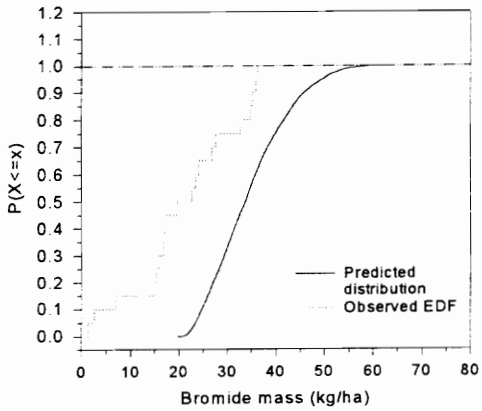


(e) Day 272

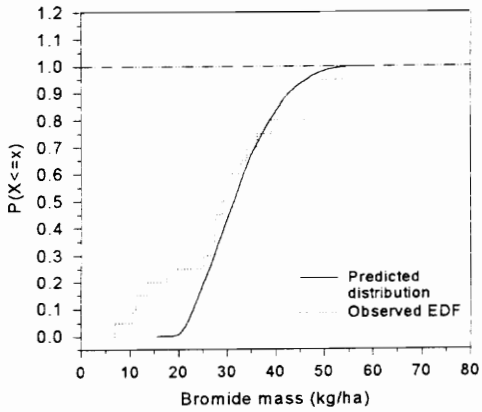
Figure 19. Predicted distribution and observed EDF for mass of bromide in the root zone of the no-till plot on (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.



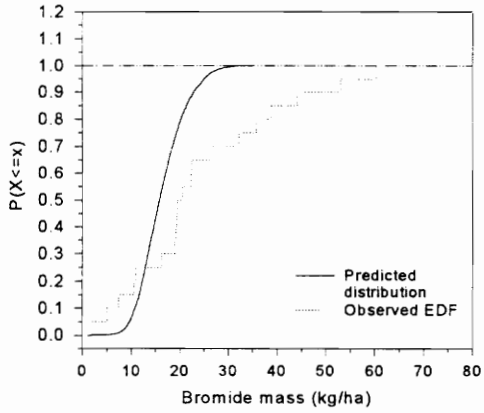
(a) Day 118



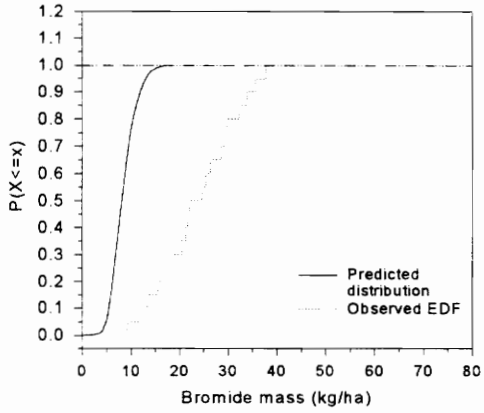
(b) Day 128



(c) Day 145



(d) Day 209



(e) Day 272

Figure 20. Predicted distribution and observed EDF for mass of bromide in the root zone of the conventional tillage plot on (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.

Table 15. Goodness-of-fit test results for comparing between predicted distribution and observed EDF for bromide mass in (a) no-till plot and (b) conventional tillage plot.

(a) No-till

Day	Adjusted KS Test Statistic	KS P-value*	AD Test Statistic	AD P-value*
118	1.979	>0.99	13.683	>0.99
128	1.391	0.95 - 0.975	5.614	>0.99
145	1.470	0.95 - 0.975	7.408	>0.99
209	2.405	>0.99	12.407	>0.99
272	3.816	>0.99	78.539	>0.99

(b) Conventional tillage

Day	Adjusted KS Test Statistic	KS P-value*	AD Test Statistic	AD P-value*
118	1.622	0.975 - 0.99	13.886	>0.99
128	2.637	>0.99	34.272	>0.99
145	1.148	0.85 - 0.90	6.246	>0.99
209	1.976	>0.99	16.750	>0.99
272	4.176	>0.99	111.278	>0.99

*Optimum fit between distributions indicated by a p-value of 0.0, while lack of fit is indicated by a p-value of >0.99.

trend observed on all the graphs is of over-prediction of mass in the root zone on the earlier dates, with large under-prediction on the later dates.

For the conventional tillage results, again two days show more than the minimum amount of significance for the KS test with p-values of less than 0.99 (Table 15). These results occur on days 118 and 145. A similar trend to that observed on the no-till graphs is observed on the conventional tillage graphs (Figure 20). The prediction of the mass in the root zone moves from over-prediction on the earlier dates, days 118 and 128, to under-prediction on days 209 and 272. Additionally, the predicted distributions on the later dates do not have as much variability as the observed data does.

Since bromide is a tracer, it should not be affected by degradation as the pesticides are, therefore any trends found on the distributional graphs (Figures 19 and 20) are likely due to the leaching component of the model. For both management practices, the predicted distributions tend to over-predict initially and under-predict at the end of the observed period. Under-prediction at the end of the observed period may be indicating movement of the tracer out of the root zone, which means that leaching is simulated too quickly. The initial over-predictions could be from slow infiltration of the bromide being simulated by the model.

Figure 21 shows a scatter graph of the I_p values for the differences in mass in the root zone between the no-till and conventional tillage plots. Graphs of the distributions used in obtaining these I_p values appear in Appendix D. There is no definite trend in the I_p values on this scatter graph, with three positive and two negative values being present in alternating order. Table 16 shows the probability that the sign of the predicted differences between the two management practices matches that of the observed data. The values in this table range from 0.46 to 0.96 with the values on the first three days around 0.50, and on the latter two dates greater than 0.96.

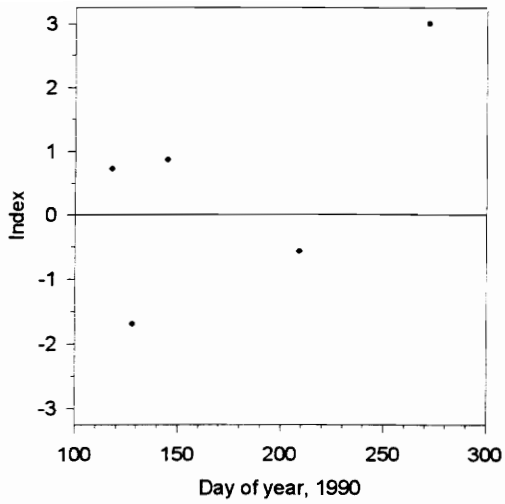


Figure 21. Scatter graph of I_p values for differences in bromide mass in the root zone between the no-till and conventional tillage plots as predicted by GLEAMS.

Table 16. Probability of observed sign being correctly predicted for differences in bromide mass in the root zone between the no-till (NT) and conventional tillage (CT) plots.

Day	Observed Sign (NT-CT)	Probability of Correctly Predicted Sign
118	-	0.552
128	+	0.460
145	-	0.524
209	-	0.964
272	-	0.968

Summary

The distributions for atrazine mass were generally under-predicted, while those for metolachlor mass were over-predicted. These trends are possibly due to errors in parameter selection for the half-life of these pesticides. The bromide distributions were over-predicted initially and under-predicted later. Trends in bromide prediction are more likely due to leaching algorithms and inputs than to degradation since bromide is not susceptible to degradation. There were few times when the goodness-of-fit tests produced p-values of less than 0.99, thus indicating that the model is not acceptable by these standards. However, these tests may be too exacting for this type of application. Despite the problems with the exactness of the goodness-of-fit tests, the probabilistic procedure delivered several additions to the model evaluation process. The primary addition was the ability to view both the observed and predicted variances. In the case of pesticide mass, the variance of the data generally reflected the spatial variability of the data over the area of the observed plots.

The differences between management practices for the pesticides were generally under-predicted, while the differences for bromide did not show any distinct trends. The probabilities for correctly predicting the sign of the differences between management practices varied over quite a large range with the range of probabilities extending from 0.004 to 0.996, which is essentially the entire range of possible probabilities. Higher probabilities occurred when the observed sign of the differences between management practices was negative, which indicated that the conventional tillage resulted in more mass in the root zone. The model's inability to predict larger mass in the root zone under no-till than under conventional tillage may signal a flaw in how the model simulates movement of pesticides through the root zone. Also, since none of these relative predictions matched the observed data all the time, there will be the possibility that the single predictions made in the deterministic procedure will fail to match observed data.

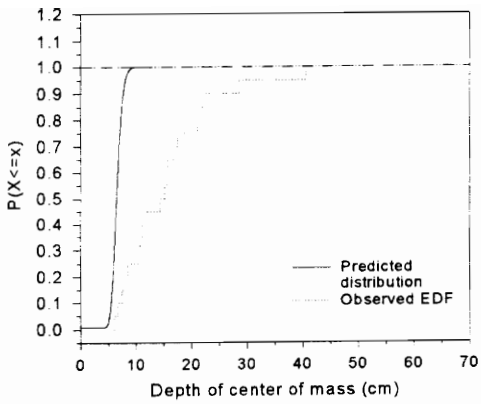
Depth of Solute Center of Mass

Atrazine

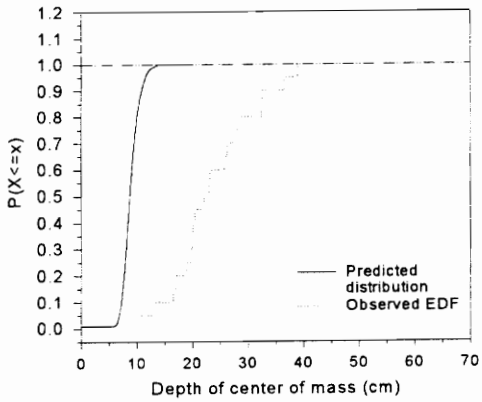
Figures 22 and 23 show the predicted and observed distributions for atrazine depth of center of mass for the no-till and conventional tillage plots, respectively. Table 17 shows the goodness-of-fit test statistics for these same distributions. For the no-till predictions, p-values of less than 0.99 occurred for the last three dates, indicating more than a minimal amount of significance. The p-values on these dates consistently decreased going from 0.975-0.99 to 0.95-0.975 and finally to <0.85. The decreasing p-values indicate increasing goodness-of-fit between the empirical and predicted distributions on these three dates. The date with the best KS test statistic, day 272, is the only day to show any significance above the minimum for the Anderson-Darling test from all three chemicals for depth of center of mass. The no-till distribution graphs (Figure 22) show that the model under-predicts depth of center of mass on the first two days, while the empirical and predicted distributions seem to be fairly centered on one another on the last three days. With the exception of the last day and especially on days 118, 128, and 209, the predicted distributions do not have as great a variance as the empirical distributions do.

The conventional tillage predictions of atrazine depth of center of mass are not as good as the no-till predictions since there is only one day with a p-value < 0.99 for the KS goodness-of-fit test for the conventional tillage predictions as compared to three days for the no-till predictions (Table 17). In Figure 23, the depth of center of mass is under-predicted on the first four days and over-predicted on the fifth date. Also, on the first three days, the variance of the empirical distribution is greater than the variance of the predicted distribution, while on the last two days the variances matched fairly well.

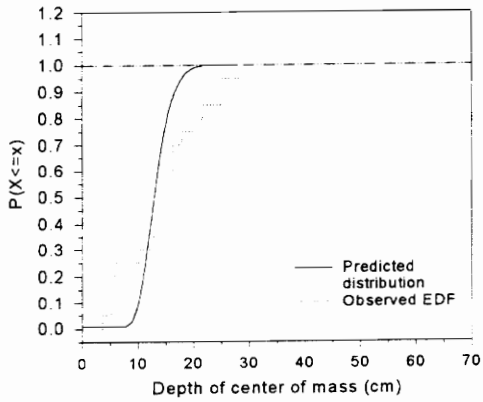
Figure 24 shows a graph of I_p values for differences between depth of center of mass predictions for the no-till and conventional tillage plots. This graph shows a possible trend towards under-prediction of the differences with four out of five dates having under-



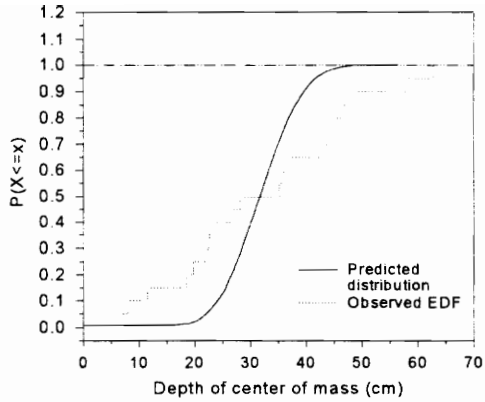
(a) Day 118



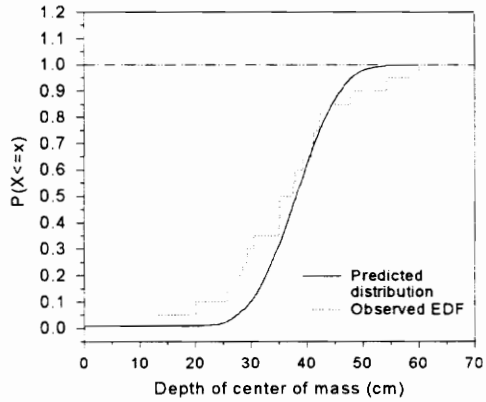
(b) Day 128



(c) Day 145

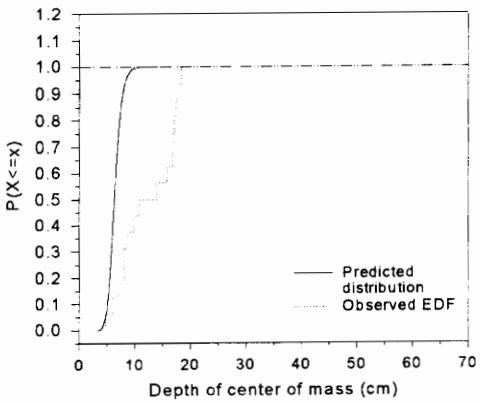


(d) Day 209

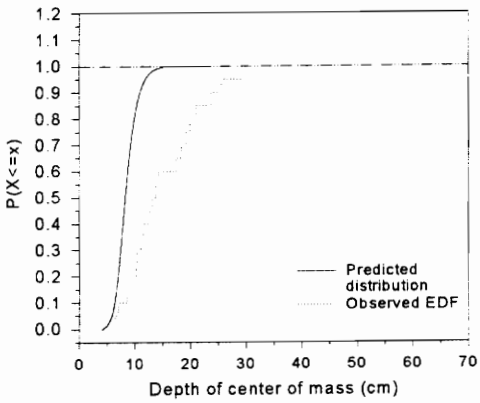


(e) Day 272

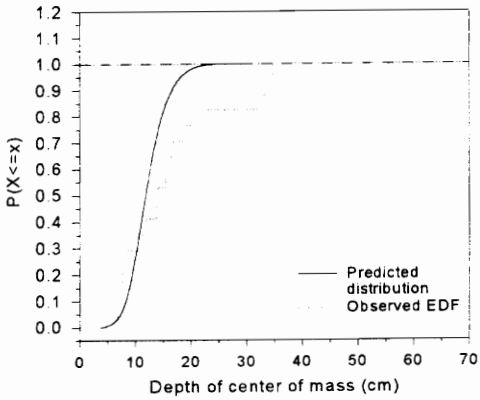
Figure 22. Predicted distribution and observed EDF for depth of center of mass for atrazine in the no-till plot on (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.



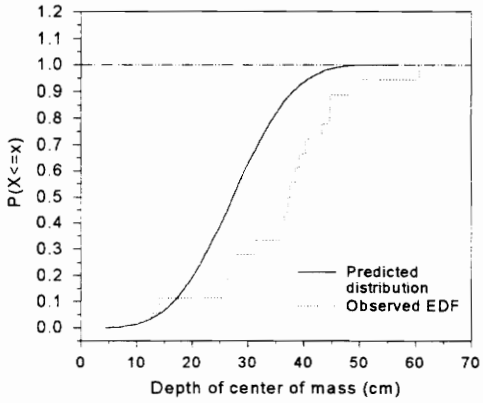
(a) Day 118



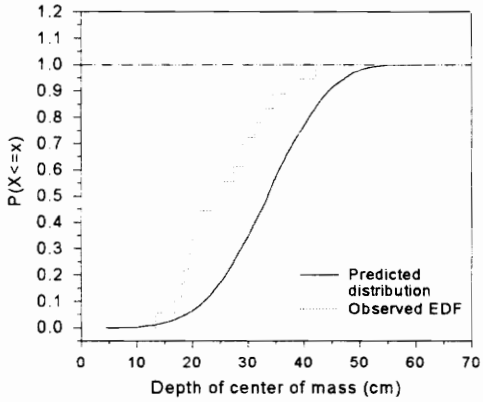
(b) Day 128



(c) Day 145



(d) Day 209



(e) Day 272

Figure 23. Predicted distribution and observed EDF for depth of center of mass for atrazine in the conventional tillage plot on (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.

Table 17. Goodness-of-fit test results for comparing between predicted distribution and observed EDF for atrazine depth of center of mass in (a) no-till plot and (b) conventional tillage plot.

(a) No-till

Day	Adjusted KS Test Statistic	KS P-value*	AD Test Statistic	AD P-value*
118	3.687	>0.99	99.718	>0.99
128	4.360	>0.99	141.873	>0.99
145	1.627	0.975 - 0.99	11.449	>0.99
209	1.467	0.95 - 0.975	10.303	>0.99
272	1.078	<0.85	2.370	0.90-0.95

(b) Conventional tillage

Day	Adjusted KS Test Statistic	KS P-value*	AD Test Statistic	AD P-value*
118	3.104	>0.99	49.881	>0.99
128	3.113	>0.99	40.986	>0.99
145	1.351	0.90 - 0.95	7.366	>0.99
209	2.249	>0.99	10.123	>0.99
272	1.788	>0.99	7.664	>0.99

* Optimum fit between distributions indicated by a p-value of 0.0, while lack of fit is indicated by a p-value of >0.99.

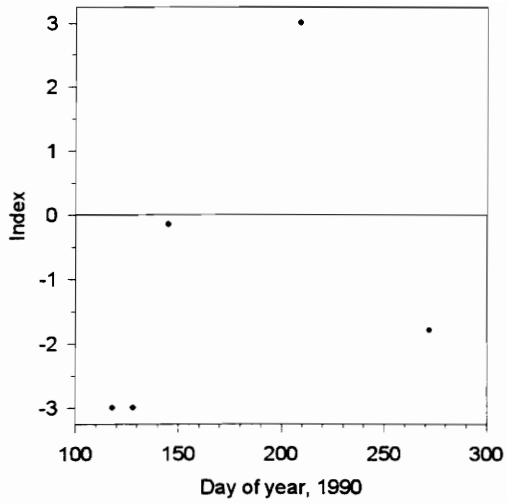


Figure 24. Scatter graph of I_p values for differences in atrazine depth of center of mass between the no-till and conventional tillage plots as predicted by GLEAMS.

predictions in them. The fourth date, however shows an extreme over-prediction of the differences in depth of center of mass between the two tillage types. The probabilities that the observed sign of the differences between the two tillage types is predicted correctly appear in Table 18. Four out of the five dates show probabilities of the correct sign being predicted of greater than 0.68. The remaining date, day 209, has a small probability (0.08) of being correct. This one small probability could be due to model error, an error in the observed value, or the chance that the observed difference happens to be one of the rare occurrences of this sign that are predicted by the model.

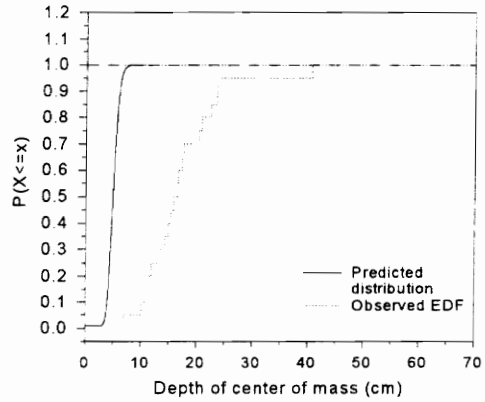
The trend of under-prediction of the depth of center of mass on the first two observed dates could be caused by a combination of preferential flow in the actual soil profile and errors in inputs that affect the initial movement of the pesticide into soil water, such as adsorption and washoff coefficients. Over-prediction of leaching by the model on the last two observed days could be due to the model predicting that leaching occurred too quickly. These two contradictory statements about the prediction of pesticide movement in the root zone could be explained as follows. The actual rate of pesticide leaching could be composed of two separate rates, one high rate controlled by preferential flow early in the observed period and the other controlled by mass flow later on. Thus, the constant leaching predicted by the model could be too small for the initial observed period and too large in the latter portion of the observed period.

Metolachlor

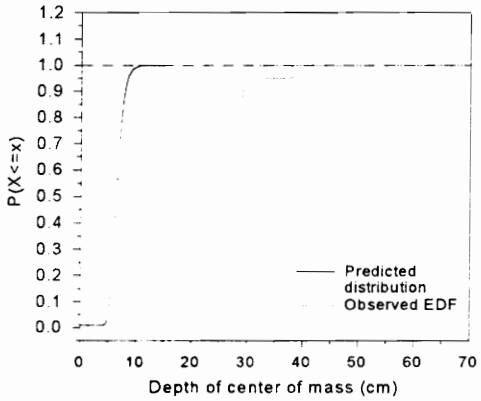
The predicted and empirical distributions for metolachlor depth of center of mass for the no-till and conventional tillage plots appear in Figures 25 and 26. Goodness-of-fit statistics for that same data appear in Table 19. For the no-till predictions, only day 272 had a goodness-of-fit test with a p-value < 0.99. On this day, the p-value for the KS test was between 0.975 and 0.99. The conventional tillage results also had only one day with a p-value below 0.99 for a goodness-of-fit test. This day was day 209 and the p-value for

Table 18. Probability of observed sign being correctly predicted by GLEAMS for differences in atrazine depth of center of mass between the no-till (NT) and conventional tillage (CT) plots.

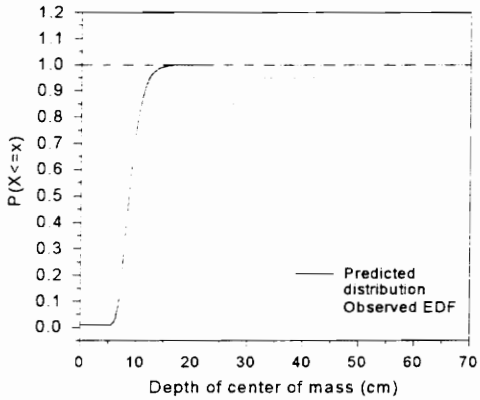
Day	Observed Sign (NT-CT)	Probability of Correctly Predicted Sign
118	+	0.684
128	+	0.788
145	+	0.848
209	-	0.080
272	+	0.924



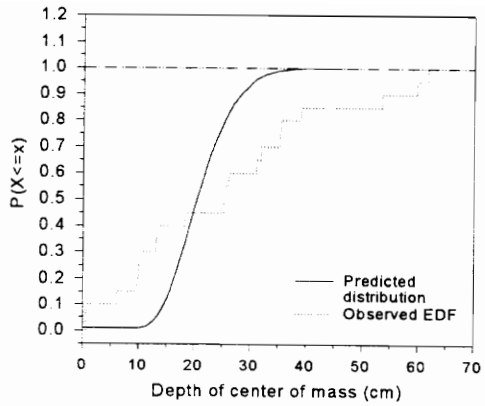
(a) Day 118



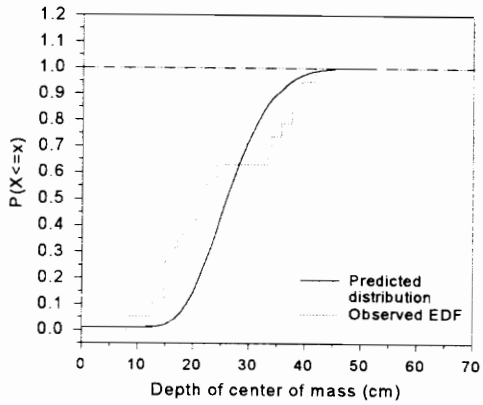
(b) Day 128



(c) Day 145

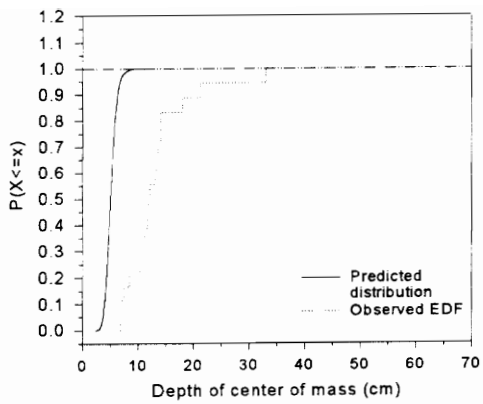


(d) Day 209

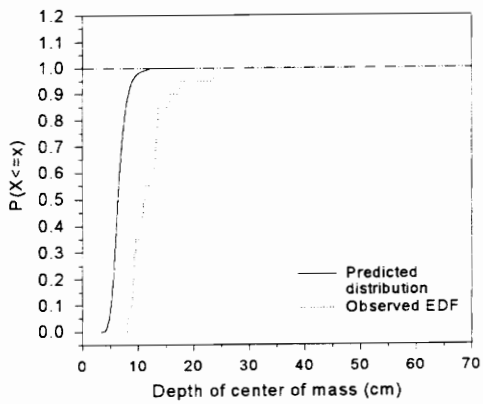


(e) Day 272

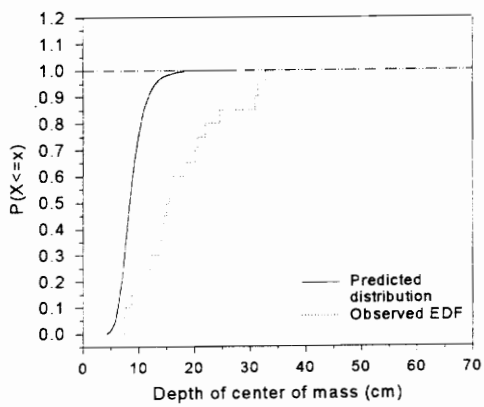
Figure 25. Predicted distribution and observed EDF for depth of center of mass for metolachlor in the no-till plot on (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.



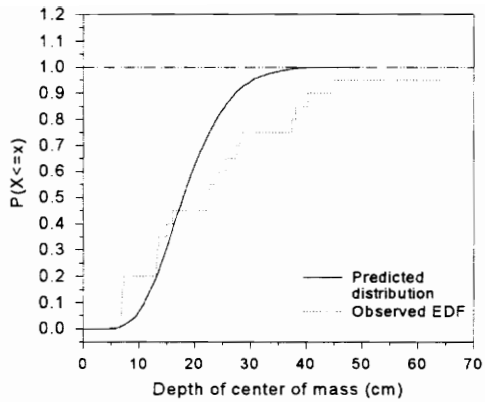
(a) Day 118



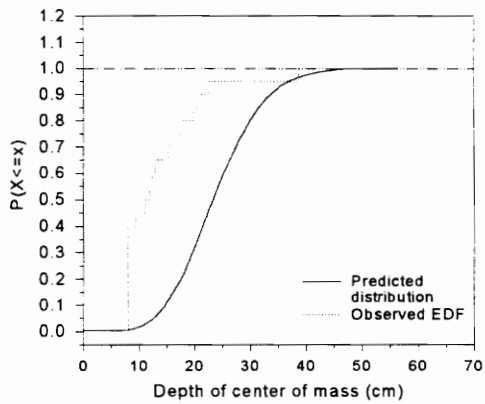
(b) Day 128



(c) Day 145



(d) Day 209



(e) Day 272

Figure 26. Predicted distribution and observed EDF for depth of center of mass for metolachlor in the conventional tillage plot on (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.

Table 19. Goodness-of-fit test results for comparing between predicted distribution and observed EDF for metolachlor depth of center of mass in (a) no-till plot and (b) conventional tillage plot.

(a) No-till

Day	Adjusted KS Test Statistic	KS P-value*	AD Test Statistic	AD P-value*
118	4.548	>0.99	154.462	>0.99
128	4.011	>0.99	118.026	>0.99
145	3.816	>0.99	76.258	>0.99
209	1.653	>0.99	15.219	>0.99
272	1.501	0.975 - 0.99	6.078	>0.99

(b) Conventional tillage

Day	Adjusted KS Test Statistic	KS P-value*	AD Test Statistic	KS P-value*
118	4.211	>0.99	102.112	>0.99
128	4.097	>0.99	60.667	>0.99
145	3.279	>0.99	41.612	>0.99
209	1.319	0.90 - 0.95	6.969	>0.99
272	2.860	>0.99	27.671	>0.99

* Optimum fit between distributions indicated by a p-value of 0.0, while lack of fit is indicated by a p-value of >0.99.

its KS test was between 0.90 and 0.95. The graphs for both tillage types (Figures 25 and 26) show that on the first three sampling dates the model under-predicted the depth of center of mass, while on the final two days, the model approximately predicted the correct magnitude of the empirical distribution. The only exception to the previous statement was the conventional tillage distribution for day 272, on which the empirical data was over-predicted.

A scatter graph of I_p values for metolachlor depth of center of mass differences between tillage types appears in Figure 27. The I_p values shown on this graph are quite variable with three extreme under-predictions, one moderate under-prediction, and one nearly perfect prediction. The general prediction trend observed on this graph is of model under-prediction of the differences between values of depth of center of mass for the two management practices. Table 20 lists the probabilities that the model will predict the same sign for the differences between management practices as the sign for the differences in the observed data. The probabilities in this table range from 0.368 to 0.844, with 0.368 being the probability for the first date, 0.844 the probability for the last date and there being a continuous increase for the probabilities in between.

As with the atrazine model results, it appears that the model under-predicts the metolachlor depth of center of mass in the first part of the observed period and over-predicts in the later portion. These results could also be due to the result of the lack of a preferential flow component in the model or errors due to improper parameter selection for the adsorption and washoff coefficients.

Bromide

Figures 28 and 29 show the empirical and predicted distributions for bromide depth of center of mass for both the no-till and conventional tillage plots. Table 21 lists the test statistics for the two goodness-of-fit tests performed on this data. For the no-till plot, there are no test results with a p-value less than 0.99 for any of the dates. The graphs of the distributions for this tillage type show a slight under-prediction on the first

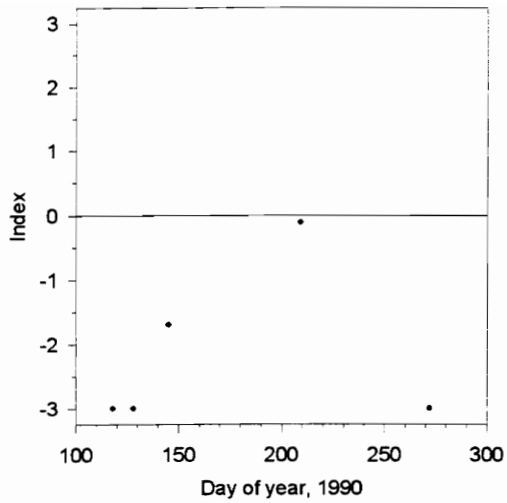
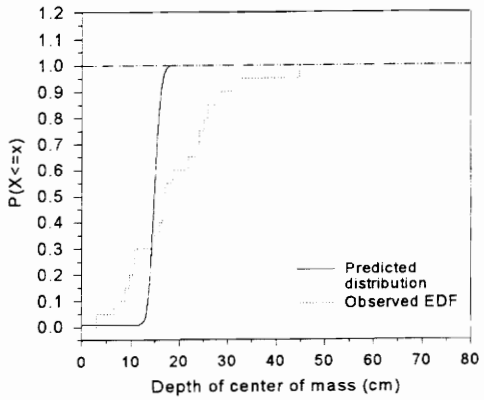


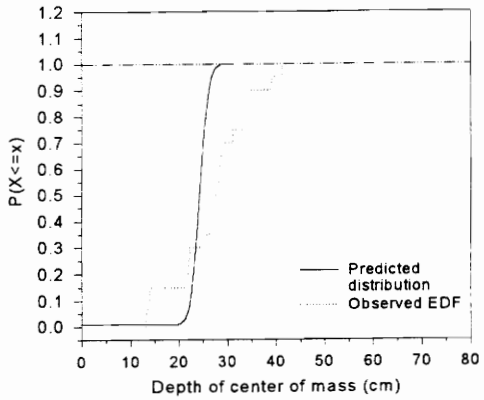
Figure 27. Scatter graph of I_p values for differences in metolachlor depth of center of mass between the no-till and conventional tillage plots as predicted by GLEAMS.

Table 20. Probability of observed sign being correctly predicted by GLEAMS for differences in metolachlor depth of center of mass between the no-till (NT) and conventional tillage (CT) plots.

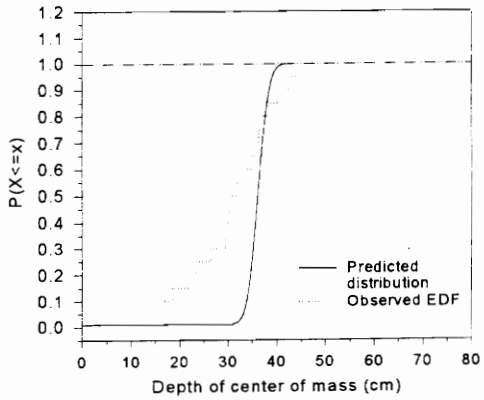
Day	Observed Sign (NT-CT)	Probability of Correctly Predicted Sign
118	+	0.368
128	+	0.564
145	+	0.640
209	+	0.832
272	+	0.844



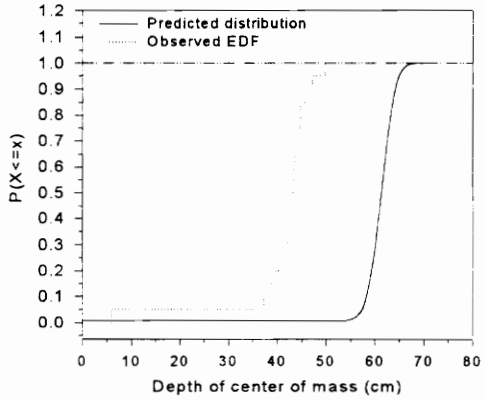
(a) Day 118



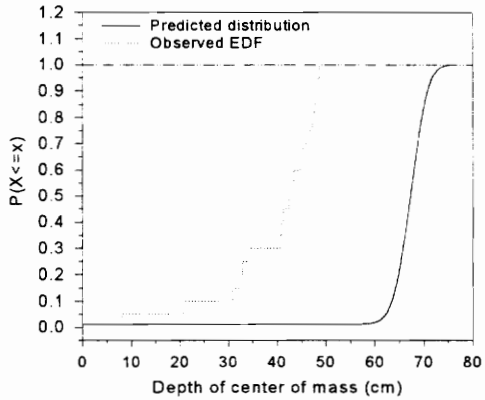
(b) Day 128



(c) Day 145

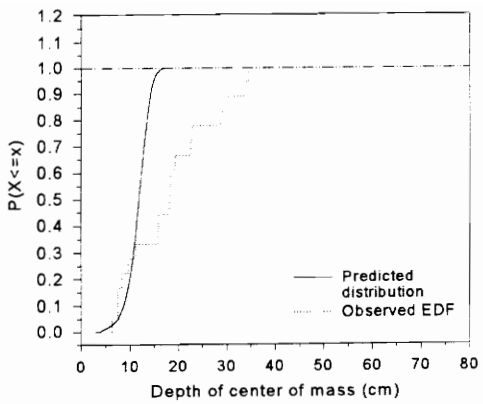


(d) Day 209

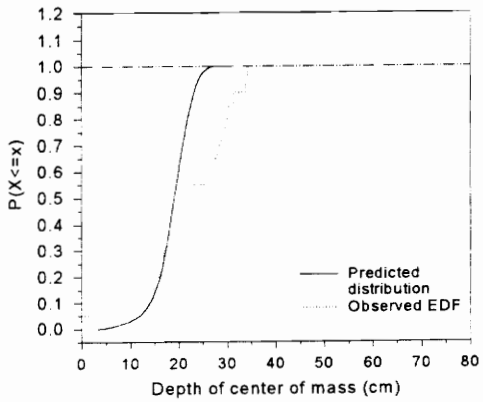


(e) Day 272

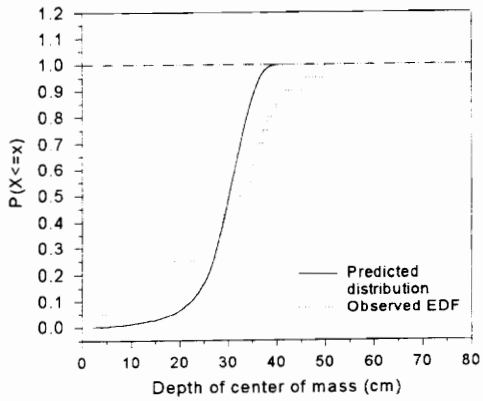
Figure 28. Predicted distribution and observed EDF for depth of center of mass for bromide in the no-till plot on (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.



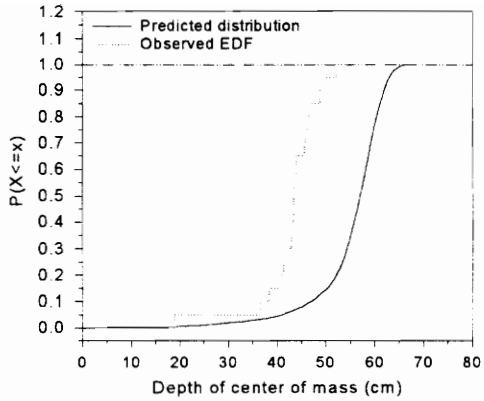
(a) Day 118



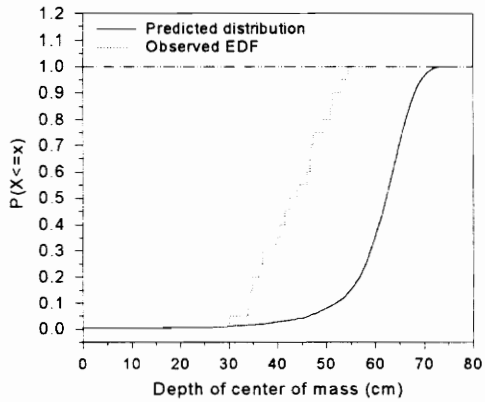
(b) Day 128



(c) Day 145



(d) Day 209



(e) Day 272

Figure 29. Predicted distribution and observed EDF for depth of center of mass for bromide in the conventional tillage plot on (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.

Table 21. Goodness-of-fit test results for comparing between predicted distribution and observed EDF for bromide depth of center of mass in (a) no-till plot and (b) conventional tillage plot.

(a) No-till

Day	Adjusted KS Test Statistic	KS P-value*	AD Test Statistic	AD P-value*
118	2.500	>0.99	39.330	>0.99
128	2.486	>0.99	31.089	>0.99
145	2.459	>0.99	20.920	>0.99
209	4.564	>0.99	71.654	>0.99
272	4.545	>0.99	65.166	>0.99

(b) Conventional tillage

Day	Adjusted KS Test Statistic	KS P-value*	AD Test Statistic	AD P-value*
118	2.857	>0.99	48.886	>0.99
128	2.029	>0.99	23.484	>0.99
145	1.499	0.975 - 0.99	8.308	>0.99
209	3.764	>0.99	29.545	>0.99
272	3.970	>0.99	37.824	>0.99

* Optimum fit between distributions indicated by a p-value of 0.0, while lack of fit is indicated by a p-value of >0.99.

two days, which then changes to a slight over-prediction by day 145 and large over-predictions on days 209 and 272. On the first three days, the predicted distributions have a much smaller variance than the empirical data does.

The conventional tillage graphs (Figure 29) show similar trends in over- and under- prediction with under-predictions on the first two days changing to over-predictions on the last two days. The difference in this set of data though is that on the third day the predicted and empirical distributions match fairly well and produce a p-value less than 0.99 for the KS test.

The bromide depth of center of mass graphs show slight under-predictions of the observed EDF for the first two days and large over-predictions on the last two observed days. This observation would indicate a small amount of under-prediction due to preferential flow in the earliest days and a general over-prediction of the speed of the leaching process.

The scatter graph of I_p values for differences in depth of center of bromide mass between the no-till and conventional tillage predictions appears in Figure 30. Four out of the five days on this graph show extreme over-predictions of the differences, while the fifth day shows a slight over-prediction. These results show that the model over-predicts the differences in the depth of center of mass between the no-till and conventional tillage plots. The probability that the model predicts the correct sign of the observed difference between depth of center of mass for the two tillage types appears as Table 22. The probabilities in this table are quite extreme with four days having a probability of a correct sign of 0.0 and the fifth day having a probability of 1.0. The four days with the 0.0 probability of a correct sign correspond to the four days with the extreme over-predictions in Figure 30, while the probability of one falls on day 128 which showed a slight over-prediction in I_p values in Figure 30. With the majority of the observed days with probabilities for relative predictions of zero, there may be an error in how the model simulates the leaching process, when degradation is not a factor, or there is error in the input parameters.

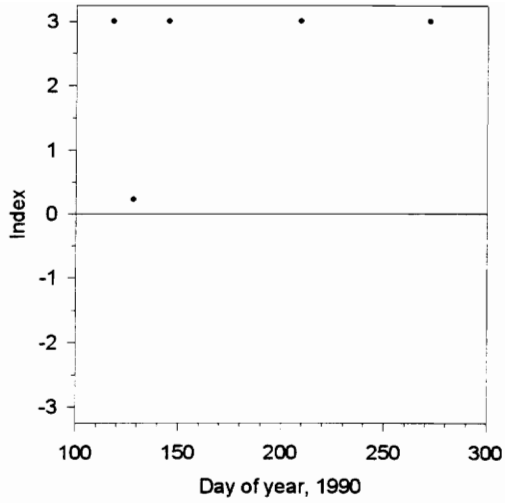


Figure 30. Scatter graph of I_p values for differences in bromide depth of center of mass between the no-till and conventional tillage plots as predicted by GLEAMS.

Table 22. Probability of observed sign being correctly predicted by GLEAMS for differences in bromide depth of center of mass between the no-till (NT) and conventional tillage (CT) plots.

Day	Observed Sign (NT-CT)	Probability of Correctly Predicted Sign
118	-	0.0
128	+	1.0
145	-	0.0
209	-	0.0
272	-	0.0

Summary

For all three solutes, the depth of center of mass was under-predicted initially and over-predicted on the later dates. The initial under-predictions could have been due to the lack of a preferential flow component for all three solutes and errors in adsorption and washoff coefficient parameter selection for the pesticides. The over-predictions of depth of center of mass on the later dates seem to indicate that the leaching process is simulated to occur too quickly by the model. As with the mass in the root zone evaluations, the goodness-of-fit tests rarely produced p-values of less than 0.99, which indicates a lack of fit between the predicted and empirical distributions.

The model generally under-predicted the difference in depth of center of mass between management practices for the pesticides and over-predicted the difference for bromide. The probabilities that the observed sign of the differences in depth of center of mass between management practices was predicted correctly were between 0.0 and 1.0. This extreme range in probabilities seems to run contradictory to the claim that management models can make relative predictions. The very low probabilities of correct sign prediction seem to indicate that the model is likely not accounting for some factors that are affecting solute transport.

Chapter Summary

The comparison between predicted and observed distributions included all values in the observed data sets which is important for studying the field-scale variation of model outputs. The probabilities in the relative prediction examination give model users a more realistic expectation of model capabilities than when using the model in a deterministic mode. The more realistic expectation of model capabilities comes from the probabilities that can be attached to the relative predictions through the use of the probabilistic procedure. By attaching probabilities to the relative predictions, model users will not have

a false sense of security about correctness of model results and will have an indication of how likely it is that the relative prediction is correct.

Overall, from the evaluation of the depth of center of mass output from the model with the probabilistic procedure, the following statements can be made about the model and the procedure itself. The variability shown in the graphs of the observed and predicted distributions reflects uncertainty in modeling the field-scale variability of depth of center of mass. Secondly, since only a few of the g-o-f tests show significance, these tests may be too strict for application to NPS pollution model evaluation or the model may be in error. The relative prediction results indicate that model users cannot be 100 percent certain about the relative results they are getting. Finally, the model results may be affected by the lack of a preferential flow component. Preferential flow was not actually studied at the field site, therefore there is no positive proof of its existence at this site, just evidence.

DETERMINISTIC EVALUATION

Deterministic evaluation is commonly used with NPS models. Thus, the results from the probabilistic procedure will be compared with those of a typical deterministic procedure on a qualitative basis. This comparison will be to determine if conclusions that are reached based on the predictions from the two procedures match and if the probabilistic predictions are more or less susceptible to parameter uncertainty. Initially though, the deterministic predictions will be compared to observed data.

METHODOLOGY

Model Inputs

Inputs for the deterministic simulations used the central tendency values from the MCS input distributions. For symmetric distributions, such as the uniform, normal, and triangular distributions, the mean of the data range was used. For distributions which were non-symmetric, like the lognormal and beta distributions, the median of the data distribution was obtained by generating 5000 random variates (MINITAB, 1994) using the MCS input distribution parameters and determining the median of each data set. The actual values used in the deterministic simulations appear in Tables 23 and 24, while the distributions used in obtaining these values appear in the previous chapter in Tables 3, 4, and 5. Copies of the input files for the deterministic simulations appear in Appendix E.

Output Analysis

The outputs generated in the deterministic simulations were analyzed using both subjective (graphical) and objective (statistical) assessments as encouraged by Loague and Green (1991). The graphical displays used for evaluating the model's accuracy in predicting the observed values were time based graphs with predicted and observed values plotted against the same x-axis, and scattergrams (James and Burges, 1982) with the x-axis representing the observed values and the y-axis the predicted values. In order to

Table 23. Central tendency values for deterministic inputs to GLEAMS for inputs common to no-till and conventional tillage simulations.**

Input (units)*	Input Value
CONA (mm day$^{-1/2}$)	3.650
POR(2)	0.435
POR(3)	0.374
POR(4)	0.434
KSOIL (Mg/ha)	0.200
KOC-Atrazine (ml/g))	100.409
KOC-Metolachlor (ml/g)	200.853
SOLLIF-Atrazine (days)	60.229
SOLLIF-Metolachlor (days)	91.010
HAFLIF-Atrazine (days)	5.019
HAFLIF-Metolachlor (days)	5.060
WSHFRC-Atrazine	0.605
WSHFRC-Metolachlor	0.705
COFUP	0.750
SILT(1)#(%)	23.863
SAND(1)# (%)	69.570
SAND(2)# (%)	60.111
SAND(3)# (%)	58.948
SAND(4)# (%)	75.375
CLAY(1)# (%)	6.844
CLAY(2)# (%)	11.805
CLAY(3)# (%)	16.602
CLAY(4)# (%)	13.674

* No units indicate input is dimensionless.

Used to obtain distributions for FC() and BR15.

** Median values obtained by generating 5000 random variates from the given distributions and ranges in Minitab Release 10 (MINITAB, 1994) and taking the 2500th value.

Table 24. Central tendency values for deterministic inputs to GLEAMS for inputs that vary between no-till and conventional tillage plots. **

Input (units)*	No-till Input Value	Conventional Tillage Input Value
POR(1)	0.435	0.516
FC(1)	0.196	0.211
FC(2)	0.240	0.239
FC(3)	0.259	0.260
FC(4)	0.224	0.225
BR15(1)	0.052	0.051
BR15(2)	0.076	0.076
BR15(3)	0.109	0.110
BR15(4)	0.080	0.081
OM(1) (%)	0.948	0.817
OM(2) (%)	0.487	0.551
OM(3) (%)	0.317	0.345
OM(4) (%)	0.199	0.259
S (mm)	134.352	68.800
CFACT(1)	0.330	0.330
CFACT(2)	0.330	0.780
CFACT(3)	0.290	0.650
CFACT(4)	0.250	0.510
CFACT(5)	0.180	0.300
CFACT(6)	0.140	0.250
CFACT(7)	0.330	0.370
NFACT	0.301	0.251
FOLFRC	0.590	0.990
APPL-Atrazine (g/ha)	0.944	0.930
APPL-Metolachlor (g/ha)	0.984	0.882

* No units indicate input is dimensionless.

** Median values obtained by generating 5000 random variates from the given distributions and ranges in Minitab Release 10 (MINITAB, 1994) and taking the 2500th value.

assess the model's ability to make comparative assessments between different management practices, scattergrams which showed comparisons between the observed and predicted ratios of values for the no-till and conventional tillage plots, were used.

Quantitative assessment of mass and depth of center of mass predictions was based on nonparametric versions of the NOF, CD, and EF. The normalized median absolute error, the robust coefficient of determination, and the robust modeling efficiency were used in place of the NOF, CD, and EF, respectively. The definitions for these statistics as they appear in Zacharias et al. (1996) are as follows:

Normalized median absolute error (MdAE),

$$\mathrm{MdAE} = \mathrm{median}\left\{\left|O_i - P_i\right| : i = 1,2,\ldots,n\right\} \times \left(\frac{100}{\tilde{O}}\right) \tag{14}$$

Robust coefficient of determination (CD*),

$$\mathrm{CD}^* = \frac{\mathrm{median}\left\{\left|O_i - \tilde{O}\right| : i = 1,2,\ldots,n\right\}}{\mathrm{median}\left\{\left|P_i - \tilde{O}\right| : i = 1,2,\ldots,n\right\}} \tag{15}$$

Robust modeling efficiency (EF*),

$$\mathrm{EF}^* = \frac{\mathrm{median}\left\{\left|O_i - \tilde{O}\right| i = 1,2,\ldots,n\right\} - \mathrm{median}\left\{\left|O_i - P_i\right| i = 1,2,\ldots,n\right\}}{\mathrm{median}\left\{\left|O_i - \tilde{O}\right| i = 1,2,\ldots,n\right\}} \tag{16}$$

where O_i and P_i represent the observed (median) and predicted values at a given location, n represents the number of observed and predicted values used in the comparison, and where:

$$\tilde{O} = \mathrm{median}\left\{O_i : i = 1,2,\ldots,n\right\} \tag{17}$$

Nonparametric statistics were chosen due to the small sample sizes in this study and the inability to determine with any degree of certainty the distributions of the data sets. These particular statistics were chosen to evaluate mass of solute in the root zone and depth of solute center of mass from the seven listed earlier because their parametric alternatives had been used in previous validation studies of solute transport models (Zacharias, 1992). The

model evaluation statistics used for evaluating the surface loss values were maximum error, root mean square error, and coefficient of residual mass. These statistics were chosen because their calculation did not depend on the type of distribution of the data and they were therefore essentially nonparametric to begin with.

Evaluation Criteria

A variety of graphical and quantitative methods were used to evaluate model performance. Actual observed and predicted values are listed in tables. Next are tables containing the model performance statistics. The model performance statistics for comparison between observed and predicted data are evaluated based on how close they are to their optimum values, either zero or one, as listed in previous sections. No guidelines for rating model performance based on these statistics have been established, therefore they are primarily useful in assessing which modeling scenarios are predicted better than other scenarios. The only definite quantitative statement that can be made based on these statistics would be in the case of a negative EF* value, in which case the median of the observed data set would be a better predictor of the data set than the predicted values.

The qualitative evaluation was based on how close the predicted and observed values appear to each other, whether the observed values are under- or over-predicted, and whether the closeness between the predicted and observed values and under- and over-predictions followed any temporal trends. The graph for comparing between ratios of values for different management practices can be evaluated in two different ways. Ideally, the predicted and observed ratios will match each other as indicated by a 45° line on the graph. Over- and under-predictions of this ratio can be evaluated based on whether predictions fall above or below the 45° line. Additionally, this graph can be used to determine whether the model has met the claim of being useful for relative predictions. If the model predicted the ratio on the same side of one as the ratio of observed values, the

model has made the correct relative prediction. Otherwise, the relative model prediction was in error.

RESULTS AND DISCUSSION

Surface Outputs

Runoff

The observed and predicted runoff values for both no-till and conventional tillage appear in Table 25. Model performance statistics comparing the observed values and deterministic predictions for runoff appear in Table 26. Figures 31 and 32 show graphical comparisons between the observed and predicted values using an event-based form (Figure 31) and a magnitude-based scattergram (Figure 32). Both graphs show a slight trend towards under-prediction, with systematic under-prediction appearing quite distinctly on the scattergram in Figure 32 (b). The model performance statistics (Table 26) show that the no-till predictions were closer to the observed values than the conventional tillage predictions since the no-till statistics were all closer to zero than the conventional tillage statistics. The conventional tillage performance statistics may be larger than the no-till statistics due to the larger magnitudes of the observed and predicted values involved in the calculation of these statistics. The under-predictions for runoff may be attributed to soil moisture storage being simulated as being too large due to errors in the selection of porosity, field capacity, and wilting point or errors in parameter selection for curve number.

Figure 33 shows observed and predicted ratios of no-till to conventional tillage runoff. The points on this graph fail to show any trends in prediction with the predicted ratios varying between over- and under-prediction. Of greater significance to the application of management models is the fact that the observed and predicted ratios all fall on the same side of one indicating that the model has correctly predicted which management practice produced the greatest amount of runoff.

Table 25. Observed values and deterministic predictions for runoff.

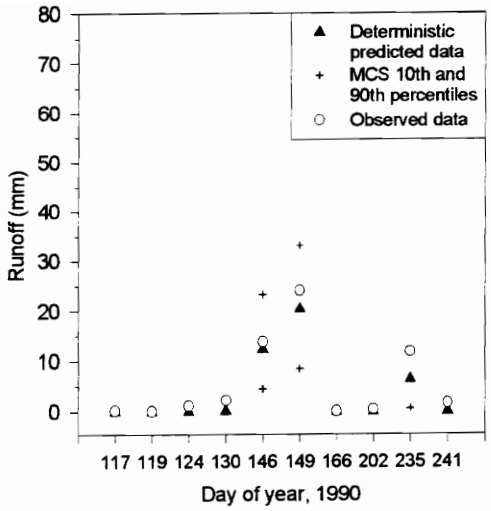
Dates	No-till Observed (mm)	No-till Predicted (mm)	Conventional Tillage Observed (mm)	Conventional Tillage Predicted (mm)
117	0.17	0.00	0.37	1.60
119	0.03	0.00	0.19	0.00
124	1.07	0.00	6.38	0.00
130	2.18	0.07	14.88	3.86
146	13.73	12.22	47.09	32.86
149	23.97	20.36	55.43	41.46
160			0.28	0.00
166	0.03	0.00	2.71	0.00
182			0.36	0.00
193			3.38	0.00
202	0.36	0.00	6.02	0.00
221			0.45	0.71
222			1.06	0.00
235	11.74	6.29	27.21	21.69
236			0.10	0.00
241	1.56	0.00	6.76	0.12

Table 26. Model performance statistics for comparison between deterministic predictions and observed values for runoff, sediment yield, atrazine surface losses, and metolachlor surface losses.

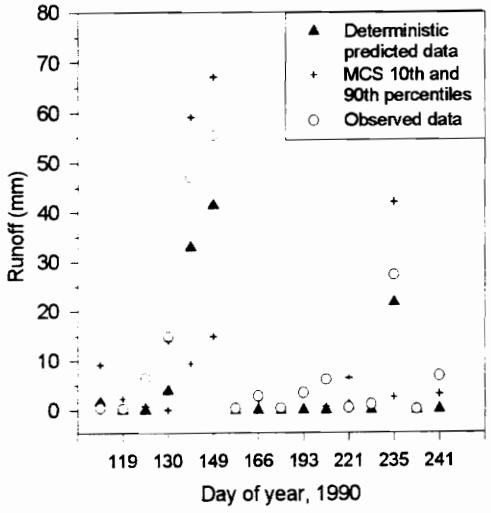
Chemical	Plot	ME*	RMSE (%)	CRM
Runoff	No-Till	5.45	2.31	0.29
	Conv. Till	14.24	6.58	0.41
Sediment yield	No-Till	139.53	62.95	-3.02
	Conv. Till	314.16	137.14	-0.04
Atrazine surface losses	No-Till	2.11	1.15	1.00
	Conv. Till	6.34	2.69	0.81
Metolachlor surface losses	No-Till	1.61	0.99	0.99
	Conv. Till	4.88	1.61	0.42

Note: If all predicted and observed values were the same, then the statistics would yield: Maximum error, ME = 0, Root mean square error, RMSE = 0, and Coefficient of residual mass, CRM = 0 (Loague and Green, 1991).

* The units for the maximum error column are as follows: runoff (mm), sediment yield (kg/ha), atrazine surface loss (g/ha), and metolachlor surface loss (g/ha).

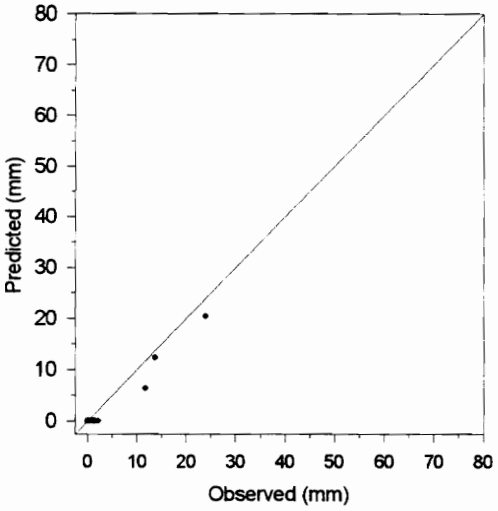


(a) No-till

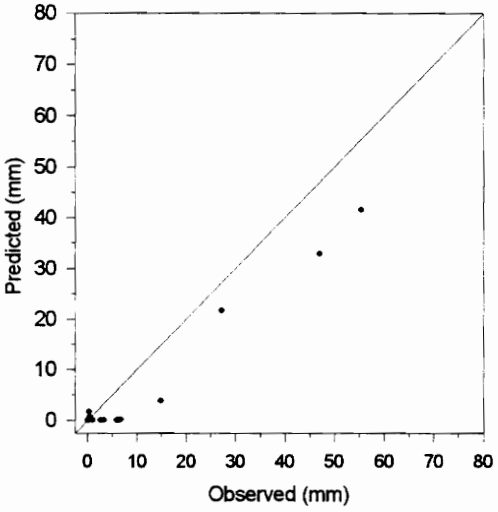


(b) Conventional tillage

Figure 31. Observed values, deterministic predictions, and MCS prediction percentiles for runoff from (a) no-till plot and (b) conventional tillage plot.



(a) No-till



(b) Conventional tillage

Figure 32. Observed values vs. deterministic predictions for runoff from (a) no-till plot and (b) conventional tillage plot.

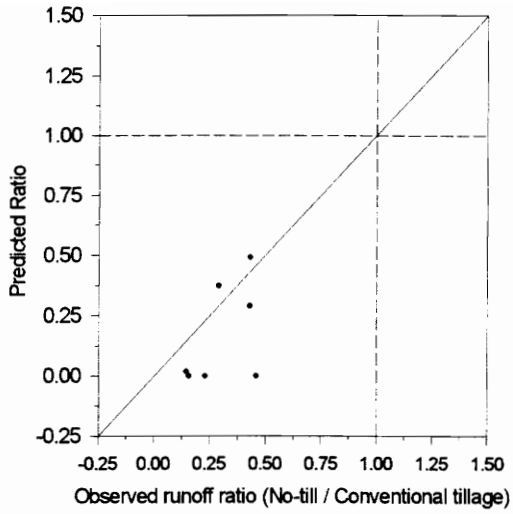


Figure 33. Observed vs. predicted ratios of no-till to conventional tillage runoff. (Three values not shown due to division by zero.)

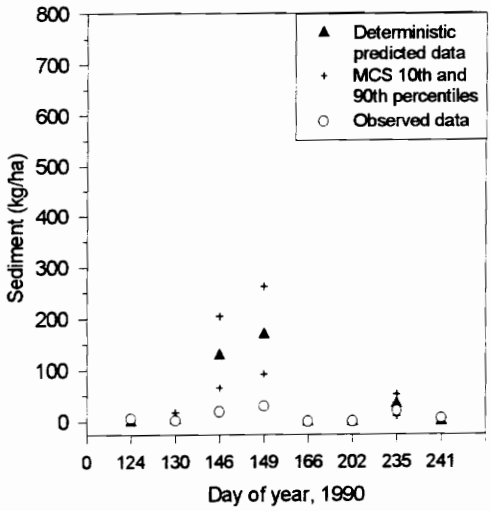
Sediment Yield

Observed and predicted values for sediment yield for both the no-till and conventional tillage plots appear in Table 27. Table 26 shows the model performance statistics for comparisons between the observed and predicted sediment yield values, while Figure 34 presents a graphical event-based view of this same comparison. A scattergram of the observed and predicted sediment yield values is in Figure 35. The graphs show that the model over-predicted sediment yield for all the largest sediment yield events (on days 146, 149, and 235) no matter what the date for the no-till plot (Figure 34(a)), while the model over-predicted large sediment yield events on the earliest dates (days 146 and 149) and under-predicted on the later dates (days 235 and 241) for the conventional tillage plot (Figure 34(b)). The model performance statistics for sediment yield (Table 26) indicate that the no-till sediment yield was better predicted than the conventional tillage. These statistics also indicate that the observed and predicted values are quite far apart with RMSE's of 63% and 137% for the no-till and conventional tillage predictions, respectively. The CRM value for the conventional tillage data (-0.04) is misleading since it is nearly zero despite the large differences between the observed and predicted values. This statistic is deceiving since the magnitude of the sum of the observed and predicted values matched, even though the individual differences between the observed and predicted values were quite large. The errors in the no-till predictions could be reduced by altering an erosion input such as soil erodibility. Since the conventional tillage differences seem to occur randomly, it is difficult to attribute them to a given input or component of the model.

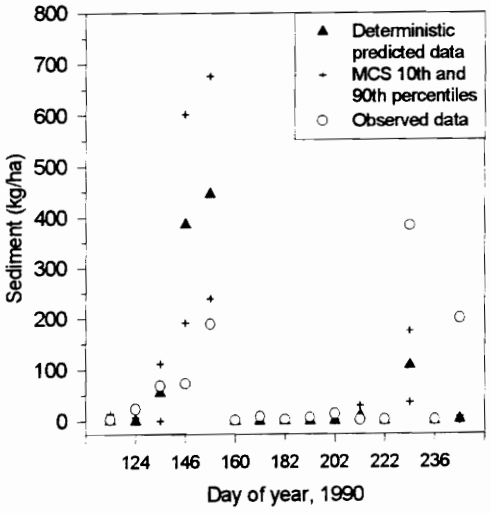
Figure 36 is a scattergram of the observed versus the predicted no-till to conventional tillage ratios. The ratios on this scattergram show that the model tends to over-predict the ratio of sediment yield between the management practices. This scattergram also shows that the model does, however, consistently predict which tillage type creates the most sediment yield, which is consistent with the intended usage of a management type model.

Table 27. Observed values and deterministic predictions for sediment yield.

Dates	No-till Observed (kg/ha)	No-till Predicted (kg/ha)	Conventional Tillage Observed (kg/ha)	Conventional Tillage Predicted (kg/ha)
119			2.53	0.24
124	6.15	0.00	22.96	0.00
130	1.74	3.13	68.45	55.13
146	18.81	128.31	72.99	387.14
149	29.96	169.49	188.45	446.4
160			0.64	0.00
166	0.09	0.00	8.03	0.00
182			2.12	0.00
193			5.64	0.00
202	0.05	0.00	12.58	0.00
221			1.20	9.65
222			1.66	0.00
235	20.19	33.29	384.19	107.63
236			1.53	0.00
241	6.06	0.00	199.85	1.44

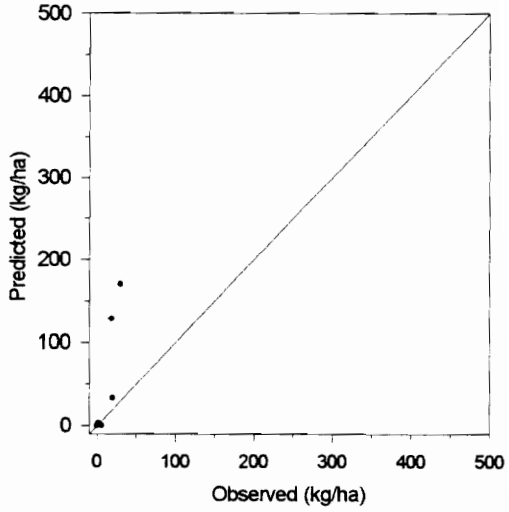


(a) No-till

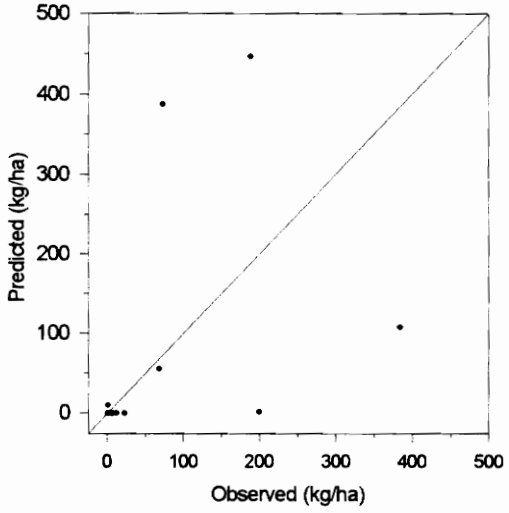


(b) Conventional tillage

Figure 34. Observed values, deterministic predictions, and MCS prediction percentiles for sediment yield from (a) no-till plot and (b) conventional tillage plot.



(a) No-till



(b) Conventional tillage

Figure 35. Observed values vs. deterministic predictions for sediment yield from (a) no-till plot and (b) conventional tillage plot.

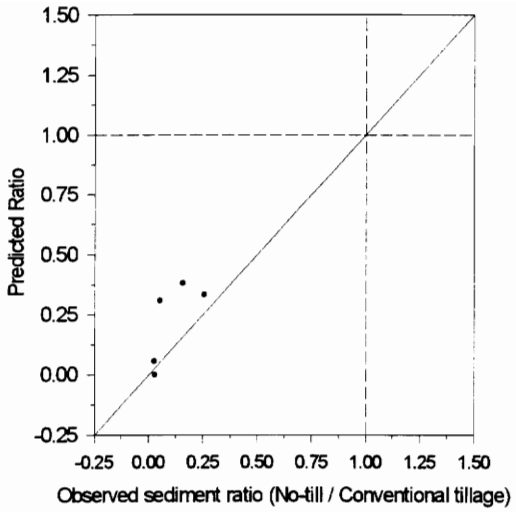


Figure 36. Observed vs. predicted ratios of no-till to conventional tillage sediment yield. (Three values not shown due to division by zero.)

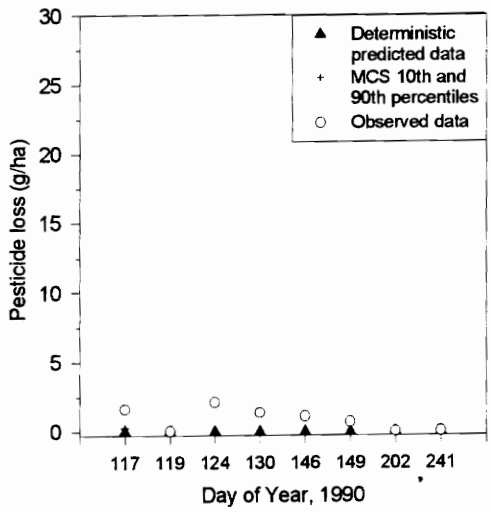
Atrazine Surface Losses

Table 28 shows the deterministic predictions and observed values for atrazine surface losses for both the no-till and conventional tillage plots. Table 26 lists the model performance statistics for the comparison between the observed atrazine surface losses and deterministic predictions for atrazine surface losses. Figures 37 and 38 graphically compare the median observed value and deterministic predictions for atrazine surface losses for the two plots using an event-based scale (Figure 37) and a magnitude-based scale (Figure 38). Both figures show that the model under-predicts the surface loss values for all dates and magnitudes of observed surface losses with the exception of the first date for the conventional tillage data. The low predictions for surface losses could be due to either the degradation of the pesticide, the movement of the pesticide into the soil being simulated too quickly, or could be associated with the low runoff predictions. The first condition could be affected by the selection of pesticide half-life, while the second could be affected by the selection of pesticide solubility and the distribution of the pesticide between the crop canopy and the soil surface upon application. Two of the three model performance statistics (ME and RMSE) in Table 26 indicate that the no-till predictions are closer to the observed values than the conventional tillage data while the third (CRM) indicates the opposite.

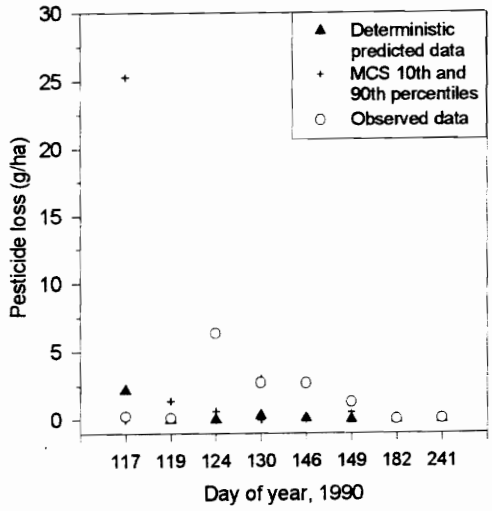
Figure 39 shows the observed and predicted ratios of surface losses between the no-till and conventional tillage plots. Since the model predictions were zero or close to zero, the predicted ratios showed up as zeros or could not be displayed due to division by zero. Of particular interest to the application of management type models was the observed ratio of greater than six, which was predicted as zero by the model. The event which produced this set of ratios was the first rainfall event after the application of the pesticides. The observed atrazine surface losses for the first rainfall event differ from the other observed events in that this is the only event in which the observed conventional tillage plot losses are less than those from the no-till plot.

Table 28. Observed values and deterministic predictions for atrazine surface losses.

	No-till		Conventional Tillage	
Dates	**Observed** (g/ha)	**Predicted** (g/ha)	**Observed** (g/ha)	**Predicted** (g/ha)
117	1.623	0.000	0.262	2.173
119	0.018	0.000	0.114	0.001
124	2.112	0.000	6.341	0.000
130	1.336	0.003	2.713	0.290
146	1.089	0.004	2.670	0.092
149	0.678	0.000	1.281	0.000
182			0.010	0.000
202	0.007	0.000		
241	0.012	0.000	0.066	0.000

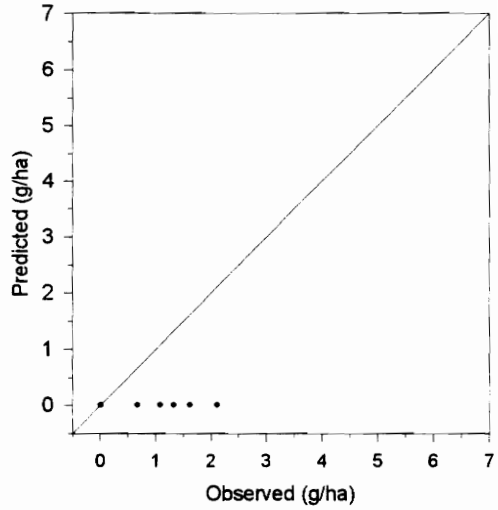


(a) No-till

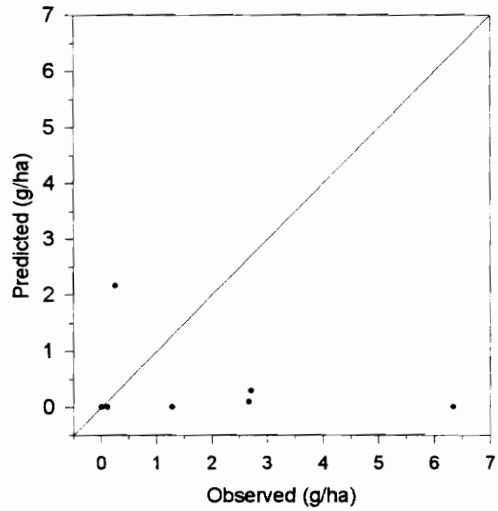


(b) Conventional tillage

Figure 37. Observed values, deterministic predictions, and MCS prediction percentiles for atrazine surface losses from (a) no-till plot and (b) conventional tillage plot.



(a) No-till



(b) Conventional tillage

Figure 38. Observed values vs. deterministic predictions for atrazine surface losses from (a) no-till plot and (b) conventional tillage plot.

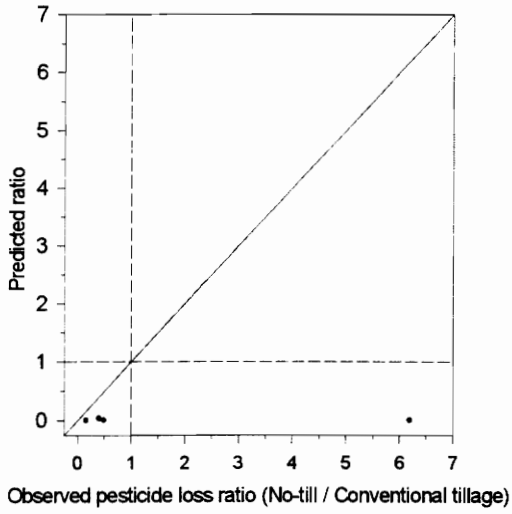


Figure 39. Observed vs. predicted ratios of no-till to conventional tillage atrazine surface losses. (Three values not shown due to division by zero.)

Metolachlor Surface Losses

Observed and predicted values for metolachlor surface losses appear in Table 29. The model performance statistics for comparing between the observed and predicted values for metolachlor surface losses appear in Table 26. The observed and deterministic model predictions for the metolachlor surface losses appear in Figures 40 and 41 with Figure 40 using an event-based scale and Figure 41 a magnitude-based scale. The two figures show that, just as for the atrazine surface loss predictions, the metolachlor surface losses are under-predicted on all but one date, the first date for the conventional tillage plot. On this one date, there is a large over-prediction in the metolachlor surface loss values. The magnitude-based graph in Figure 41 also shows mainly under-predictions. The model performance statistics (Table 26) again indicate that the no-till predictions are closer to the observed values than the conventional tillage predictions. The under-predictions observed for the majority of the events, as for the atrazine surface losses, could be attributed to incorrect parameter selection for the pesticide half-life, solubility, the initial surface distribution of the pesticide, or the low runoff predictions of the model.

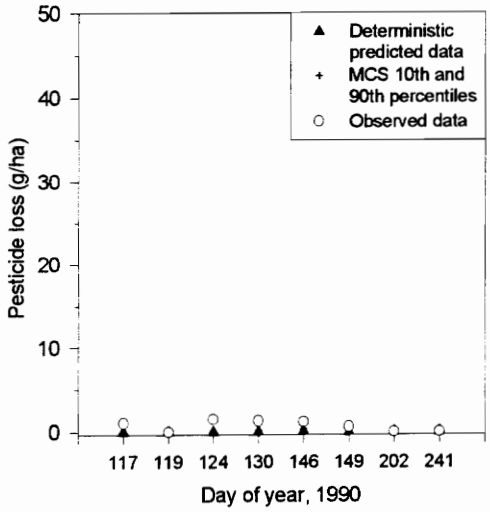
Figure 42 shows a scattergram with the observed and predicted ratios of metolachlor surface losses between the no-till and conventional tillage plots. This graph indicates that for all dates the predicted ratios remain quite close to zero, thus as the magnitude of the observed ratio increases, the model predicted ratio appears to worsen. The only day which has a predicted ratio on the opposite side of unity from the observed ratio is the first date, just as was the case for the atrazine surface losses.

Summary

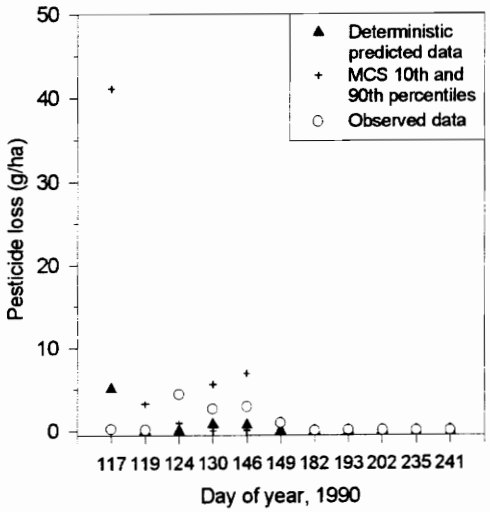
The deterministic runoff values were under-predicted for both management practices, while the sediment yield values were over-predicted with the exception of two events later in the observed period for the conventional tillage plot. All the pesticide surface losses were under-predicted with the exception of the first event for the

Table 29. Observed values and deterministic predictions for metolachlor surface losses.

Dates	No-till Observed (g/ha)	No-till Predicted (g/ha)	Conventional Tillage Observed (g/ha)	Conventional Tillage Predicted (g/ha)
117	1.068	0.000	0.220	5.102
119	0.015	0.000	0.107	0.004
124	1.478	0.000	4.424	0.000
130	1.264	0.011	2.612	0.793
146	1.185	0.075	2.894	0.692
149	0.570	0.001	0.906	0.002
182			0.012	0.000
193			0.078	0.000
202	0.005	0.000	0.030	0.000
235			0.001	0.001
241	0.007	0.000	0.014	0.000

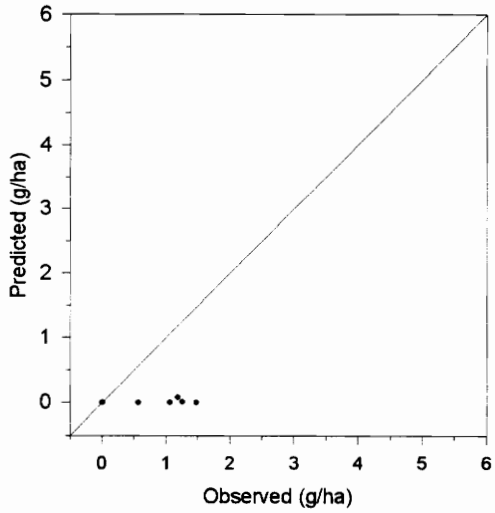


(a) No-till

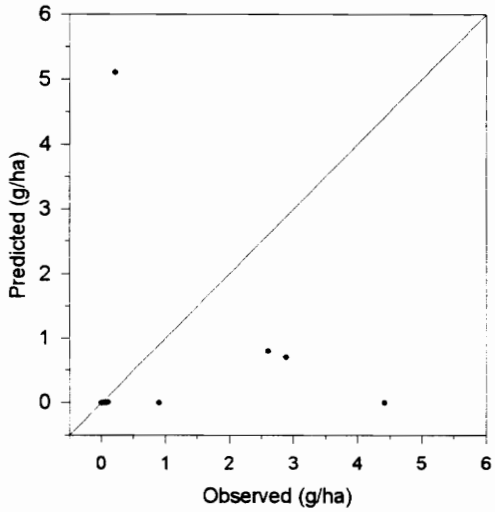


(b) Conventional tillage

Figure 40. Observed values, deterministic predictions, and MCS prediction percentiles for metolachlor surface losses from (a) no-till plot and (b) conventional tillage plot.



(a) No-till



(b) Conventional tillage

Figure 41. Observed values vs. deterministic predictions for metolachlor surface losses from (a) no-till plot and (b) conventional tillage plot.

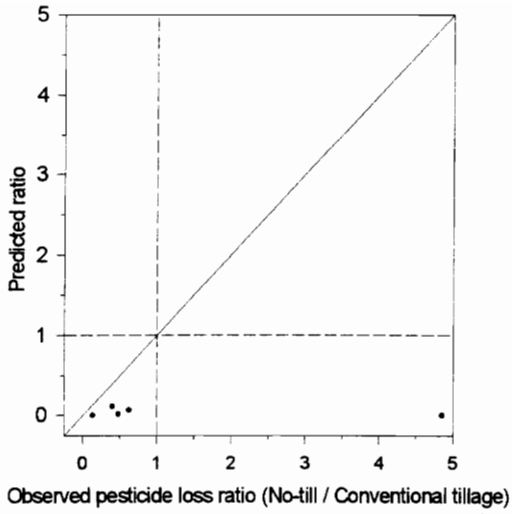


Figure 42. Observed vs. predicted ratios of no-till to conventional tillage metolachlor surface losses. (Three values not shown due to division by zero.)

conventional tillage plot for both pesticides. The results of this study serve to backup previous studies that have found that the model under-predicts runoff.

The ratio between management practice graphs indicated that correct relative predictions were made for all but one day for each of the pesticide predictions. The percentage of correct relative predictions seems to exceed many of the probabilities obtained for predicting the correct sign of the differences between management procedures with the probabilistic procedure. The two incorrect relative predictions indicate that even when applied as a management type model in a deterministic mode, GLEAMS is not perfect in its relative prediction capability.

Solute Mass in the Root Zone

Atrazine

The observed values and deterministic predictions for atrazine mass in the root zone appear in Table 30. Model performance statistics comparing the observed values and deterministic predictions appear in Table 31, with the graphical comparisons for these values displayed in Figures 43 and 44. The time-based graph of no-till mass in Figure 43 shows model under-prediction on the first two dates, slight over-prediction on the third date, and slight under-predictions on the last two days, with all predictions falling within the 10th and 90th percentiles of the observed data. The graph of atrazine mass in the root zone in the conventional tillage plot shows nearly the opposite trend from the no-till data with model over-prediction occurring on the first three days, a nearly perfect prediction on the fourth day and under-prediction on the last day. The predicted values all fell within the 10th and 90th percentiles of the observed data, except for the last date, which had a predicted value just below the 10th percentile of the observed data. In Figure 44, with the observed and predicted data graphed based on magnitudes, no real prediction trends are apparent, although the no-till data are under-predicted on four out of five days. Based on the results from previous studies, the tendency to under-predict could be due to how the model simulates leaching and pesticide degradation.

The model performance statistics comparing deterministic predictions and observed values for atrazine mass in the root zone show mixed degrees of agreement between the predictions and the observed values. The MdAE values fall in the middle of the three sets of mass statistics in Table 31 as far as being close to the optimum value of one. The EF* values are negative, which indicates that the observed median would better represent these data sets than the model predictions.

Figure 45 shows the observed and predicted ratios between atrazine masses in the root zones of the no-till and conventional tillage plots. This scattergram shows that the predicted ratio for atrazine mass stays nearly constant no matter what the observed ratio

Table 30. Observed values and deterministic predictions for (a) atrazine, (b) metolachlor, and (c) bromide mass in the root zone.

(a) Atrazine

Dates	No-till Median (kg/ha)	No-till Predicted (kg/ha)	Conventional Tillage Median (kg/ha)	Conventional Tillage Predicted (kg/ha)
118	0.85	0.62	0.66	0.89
128	0.87	0.55	0.66	0.79
145	0.36	0.45	0.44	0.64
209	0.23	0.18	0.25	0.26
272	0.29	0.07	0.43	0.10

(b) Metolachlor

Dates	No-till Median (kg/ha)	No-till Predicted (kg/ha)	Conventional Tillage Median (kg/ha)	Conventional Tillage Predicted (kg/ha)
118	0.56	0.70	0.70	0.85
128	0.48	0.65	0.47	0.79
145	0.25	0.56	0.38	0.68
209	0.04	0.31	0.05	0.37
272	0.06	0.16	0.03	0.18

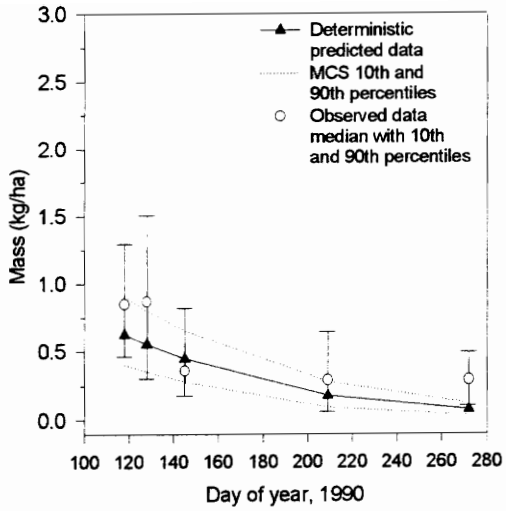
(c) Bromide

Dates	No-till Median (kg/ha)	No-till Predicted (kg/ha)	Conventional Tillage Median (kg/ha)	Conventional Tillage Predicted (kg/ha)
118	23.23	34.99	27.61	35.16
128	27.69	34.43	21.17	34.64
145	24.66	32.13	29.15	32.52
209	18.42	14.18	20.07	18.46
272	16.40	10.02	23.33	14.62

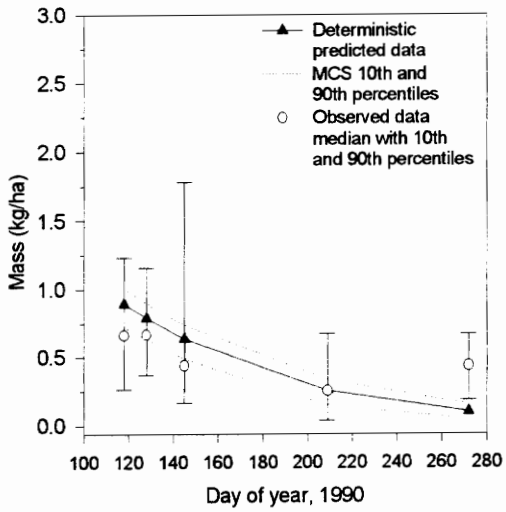
Table 31. Model performance statistics for comparison between deterministic predictions and observed medians of solute mass in the root zone.

Chemical	Plot	MdAE (%)	CD*	EF*
Atrazine	No-Till	61	0.67	-0.67
	Conv. Till	45	0.55	-0.06
Metolachlor	No-Till	68	0.65	0.18
	Conv. Till	82	1.05	0.05
Bromide	No-Till	29	0.40	-0.51
	Conv. Till	32	0.36	-1.31

Note: If all predicted and observed values were the same, then the statistics would yield: Normalized median absolute error MdAE = 0.0, Robust coefficient of determination, CD* = 1.0, and Robust coefficient of efficiency, EF* = 1.0 (Zacharias et al., 1996).

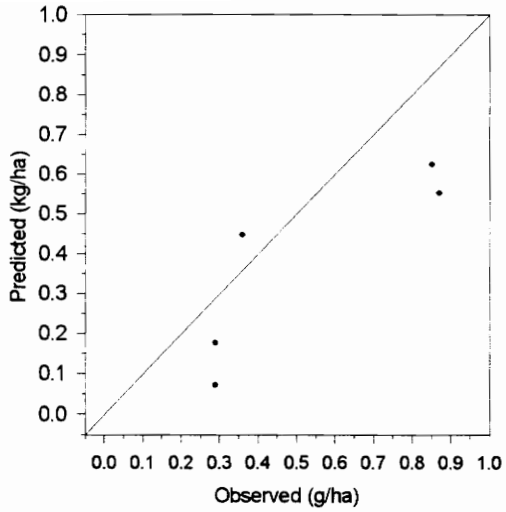


(a) No-till

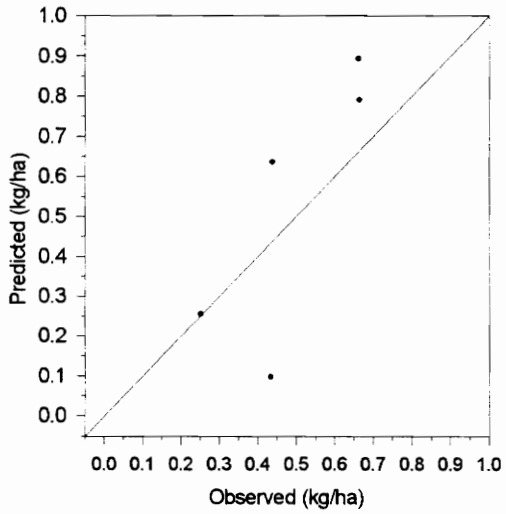


(b) Conventional tillage

Figure 43. Observed values, deterministic predictions, and MCS prediction percentiles for atrazine mass in the root zone for (a) no-till plot and (b) conventional tillage plot.



(a) No-till



(b) Conventional tillage

Figure 44. Observed values vs. deterministic predictions for atrazine mass in the root zone for (a) no-till plot and (b) conventional tillage plot.

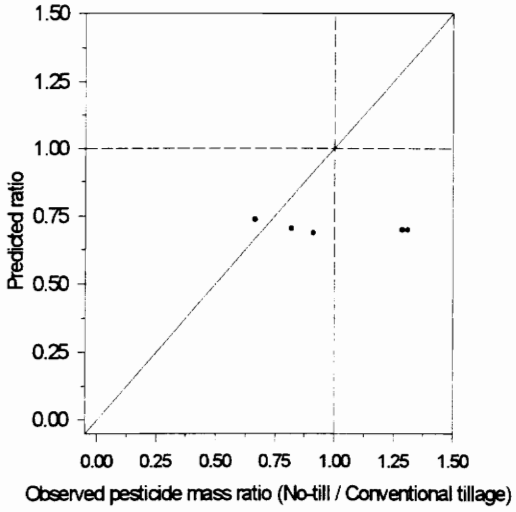


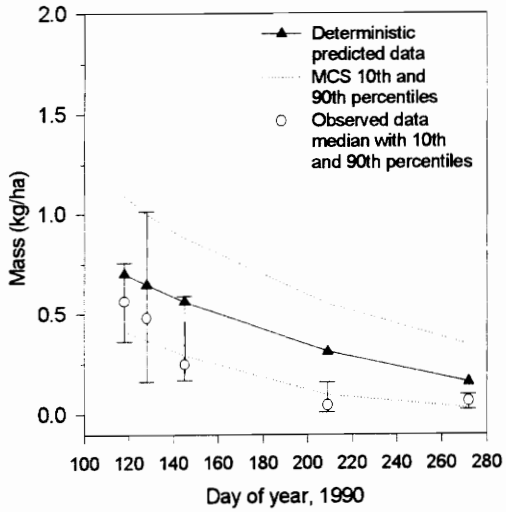
Figure 45. Observed vs. predicted ratios of no-till to conventional tillage atrazine mass in the root zone.

is. The constant predicted ratio between the values for the two management practices could be from a single sensitive parameter, such as curve number, or be due to the combined effects of several inputs, such as the soil matrix descriptors, which include porosity, wilting point, and field capacity. Figure 45 also shows that two of the predicted ratios fall on the opposite side of one as their corresponding observed ratios. This occurrence contradicts the purported relative predictive abilities of the model.

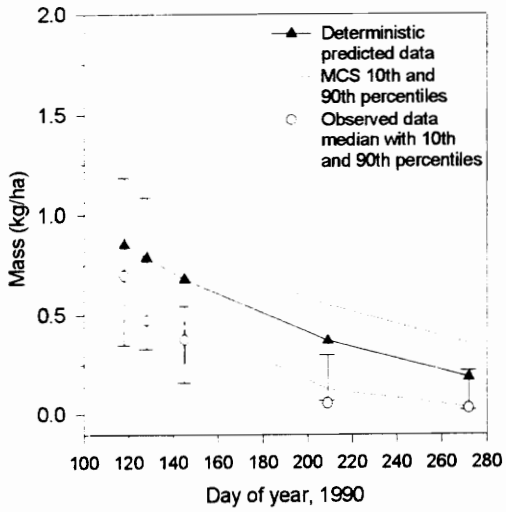
Metolachlor

Table 30 lists the observed and predicted values for metolachlor mass in the root zone. Table 31 lists the model performance statistics for comparing the deterministic model predictions to the observed values for metolachlor mass in the root zone. Figures 46 and 47 show graphical comparisons between the deterministic predictions and the observed values with Figure 46 using a time-based scale and Figure 47 a magnitude-based scale. The results for metolachlor mass in the root zone should be viewed in light of the fact that the half-life used as an input, which was obtained from literature sources, was nearly twice the half-life computed from field data for a previous study (Zacharias, 1992). Both graphical comparisons show that the model over-predicts the metolachlor mass in the root zone for both tillage types and on all dates. These results would tend to point to a systematic error in the model predictions as indicated by the metolachlor half-life being too large. This error could also be due to slower leaching being predicted by the model than is actually occurring.

The model performance statistics (Table 31) show various levels of goodness-of-fit. The MdAE values are the farthest from the optimum value of zero of the three masses in this table, while the EF* values are the only positive values for this statistic of all the masses. The positive EF* values indicate that the predicted values are a better representation of the observed data than the observed median. A scattergram of the observed and predicted ratios of the no-till to conventional tillage values appears as Figure 48. This scattergram as with the scattergram for the atrazine mass in the root zone

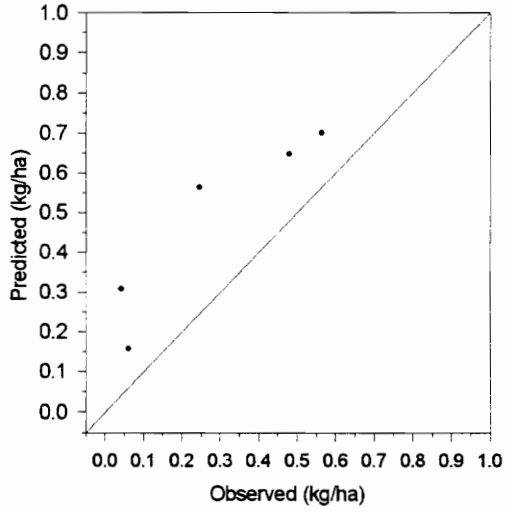


(a) No-till

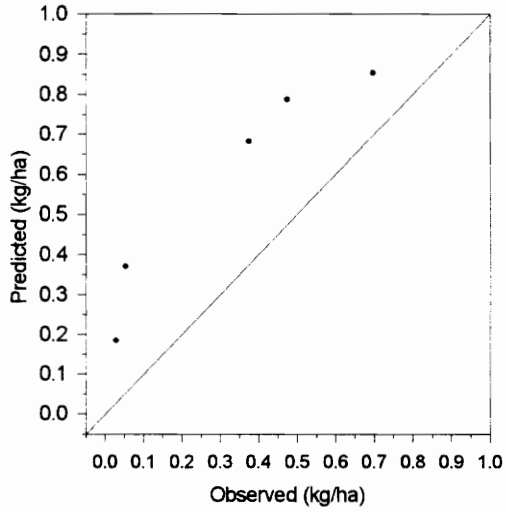


(b) Conventional tillage

Figure 46. Observed values, deterministic predictions, and MCS prediction percentiles for metolachlor mass in the root zone for (a) no-till plot and (b) conventional tillage plot.



(a) No-till



(b) Conventional tillage

Figure 47. Observed values vs. deterministic predictions for metolachlor mass in the root zone for (a) no-till plot and (b) conventional tillage plot.

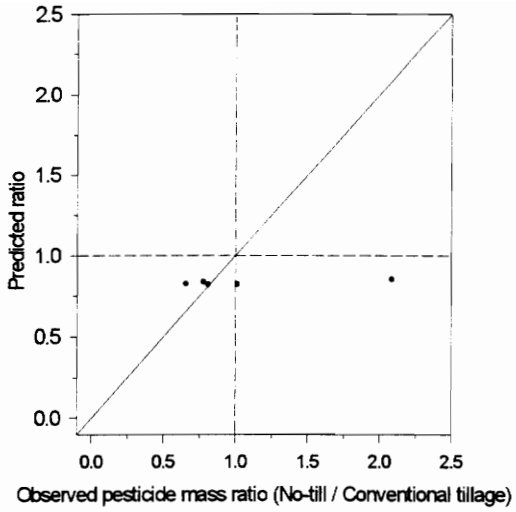


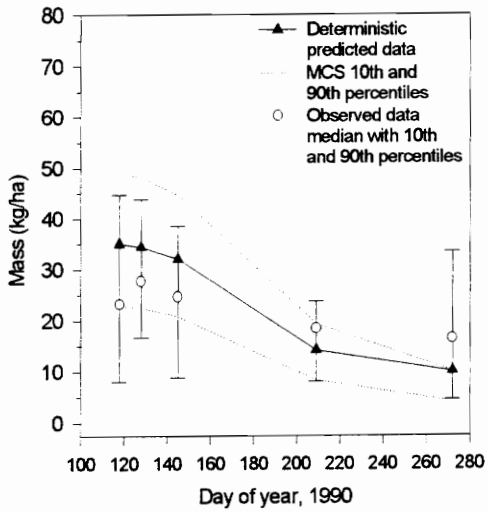
Figure 48. Observed vs. predicted ratios of no-till to conventional tillage metolachlor mass in the root zone.

indicates that the predicted ratio of no-till mass to the conventional tillage mass remains nearly constant no matter what the magnitude of the observed ratio. While no inputs have similar ratios between the two management practices, this predicted output ratio could be due to a single parameter or a combination of parameters. Figure 48 also has two predicted management practice ratios on the opposing side of one as their respective observed ratios, which contradicts the model's purported relative prediction ability.

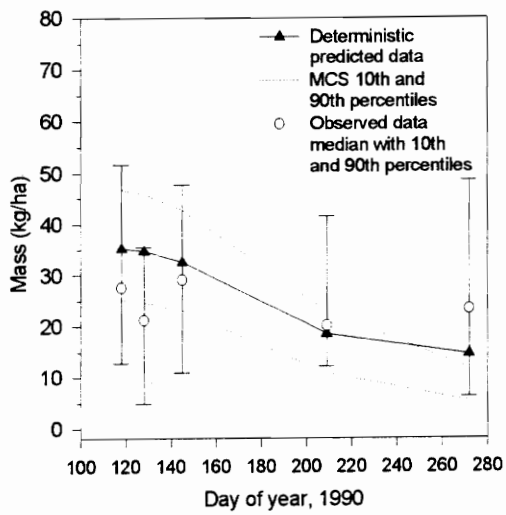
Bromide

Observed and predicted values of bromide mass in the root zone appear in Table 30. Model performance statistics for the comparison of the deterministic predictions of bromide in the root zone to observed values appear in Table 31. Graphical comparisons of the deterministic predictions of bromide mass in the root zone to the observed values appear in Figures 49 and 50, with Figure 49 being a time-based comparison and Figure 50 a magnitude-based comparison. The graphical comparison in Figure 49 shows that the model over-predicts the mass values on the first three days and then under-predicts on the last two days for both tillage types. All the deterministic predictions fall within the 10th and 90th percentiles of the observed data. The magnitude-based comparisons tend to show model under-prediction for lower magnitude observed masses and over-prediction for larger magnitude masses. Since bromide cannot be degraded, the early over-prediction of mass and later under-prediction can mainly be attributed to how the model is simulating leaching.

The model performance statistics also indicate various levels of goodness-of-fit for these predictions. The MdAE values of 29 and 32 for the no-till and conventional tillage plots, respectively, while not that close to the optimum MdAE value of 0.0, are nearly half of the MdAE values of any of the other solute masses in the root zone. On the other hand the CD* values for the bromide mass are 0.4 and 0.36, which are farther away from the optimum CD* value as compared to any of the CD* values for the other chemicals. The EF* values for both tillage types are negative, indicating that the median

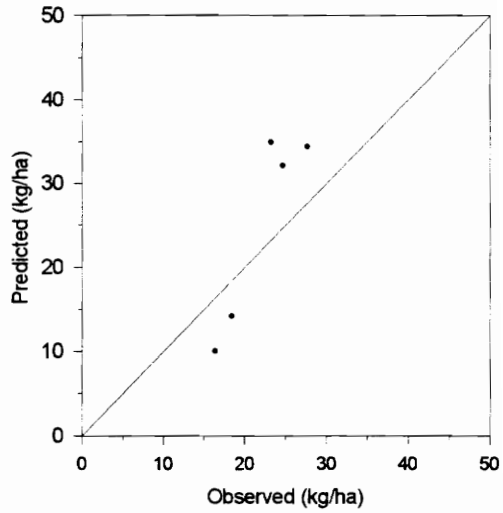


(a) No-till

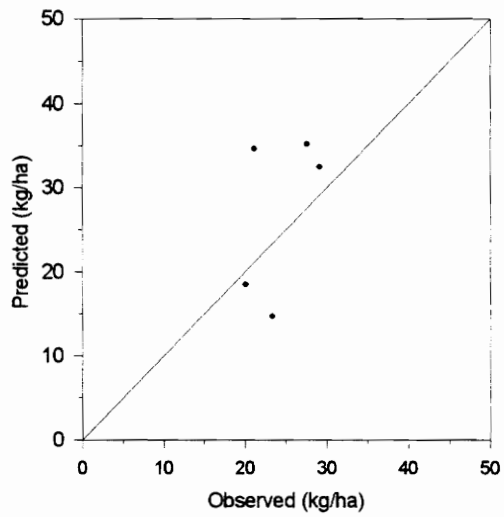


(b) Conventional tillage

Figure 49. Observed values, deterministic predictions, and MCS prediction percentiles for bromide mass in the root zone for (a) no-till plot and (b) conventional tillage plot.



(a) No-till



(b) Conventional tillage

Figure 50. Observed values vs. deterministic predictions for bromide mass in the root zone for (a) no-till plot and (b) conventional tillage plot.

values for the data sets for both tillage types would better represent the field data than the predicted values.

The observed and predicted ratios of bromide mass in the root zone between the no-till and conventional tillage plots are presented in Figure 51. The trend evidenced on this scattergram is slightly different than those for the atrazine and metolachlor masses in the root zone. Whereas on the scattergrams for the masses for the other chemicals, the predicted ratios remained constant, the bromide mass predicted ratio varies from about 0.6 to 1.0. The different ratio trends between the pesticide ratios and the bromide ratio indicate that degradation likely played a role in creating the constant management ratios for the pesticides (Figures 45 and 48).

Summary

No consistent prediction trends were evident in the atrazine mass predictions between the two management practices. The metolachlor mass graphs consistently showed over-prediction for both management practices. Bromide mass was over-predicted on the first three dates and under-predicted on the last two. Over the time period studied, the model shows the same trend of decreasing solute mass as was observed. Since no consistent prediction trends are evident between different solutes, no definitive statements can be made about general predictive traits of the model while using a deterministic procedure, except that there does not appear to be any systematic error in model predictions.

For the ratios between management practices, the two pesticides showed nearly constant ratios for all magnitudes of observed ratios. The evaluation of predicted ratios between management practices again showed problems with the relative predictive capability of GLEAMS since several ratios were not predicted on the same side of one as the observed ratios.

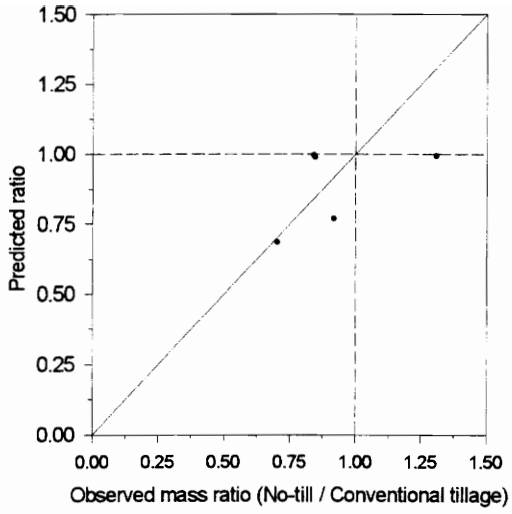


Figure 51. Observed vs. predicted ratios of no-till to conventional tillage bromide mass in the root zone.

Depth of Solute Center of Mass

Atrazine

Table 32 lists the observed and predicted values for atrazine depth of center of mass. Table 33 shows the model performance statistics for the comparisons between the observed values and the deterministic predictions of atrazine depth of center of mass. The graphical comparisons between the atrazine depth of center of mass observed values and deterministic predictions appear in Figures 52 and 53, with Figure 52 using a time-based scale and Figure 53 using a magnitude-based scale.

On the time-based scale graphs, the model under-predicted the depth of center of mass on the first three days and over-predicted on the fifth day. For no-till on the fourth day, the predicted and observed values matched each other nearly exactly. All but one of the model predictions, the second day of no-till predictions, fell between the 10th and 90th percentiles of the observed data. On the magnitude-based graph, the predictions for observed values of less than 25 cm are under-predicted. The initial under-predictions of atrazine depth of center of mass are likely due to the model under-predicting the rate at which the pesticide leaches through the soil. The later over-predictions are likely due to slow plant uptake of the pesticide being simulated.

The model performance statistics for the no-till predictions are the second closest to optimum values of the values listed in Table 33. The conventional tillage predictions were not as good though, having only the fourth closest to optimum values on this table for all the statistics. The EF* statistic for the conventional tillage plot indicates poor predictive performance, though, since it is negative meaning that the observed medians would better represent the observed data than the actual predictions.

Figure 54 shows the observed versus predicted ratios of no-till to conventional tillage atrazine depth of center of mass. This scattergram shows the predicted ratios vary little, going from 1.00 to 1.15, over the same range that the observed ratio goes from 0.75 to 1.75. The observed ratios are initially over-predicted, but become increasingly under-

Table 32. Observed values and deterministic predictions for depth of center of mass of (a) atrazine, (b) metolachlor, and (c) bromide.

(a) Atrazine

Dates	No-till Median (cm)	No-till Predicted (cm)	Conventional Tillage Median (cm)	Conventional Tillage Predicted (cm)
118	14.65	6.41	12.38	6.39
128	22.52	8.59	13.35	8.23
145	15.32	12.71	13.92	11.75
209	31.70	31.91	37.37	27.58
272	36.33	40.44	25.73	35.45

(b) Metolachlor

Dates	No-till Median (kg/ha)	No-till Predicted (kg/ha)	Conventional Tillage Median (kg/ha)	Conventional Tillage Predicted (kg/ha)
118	16.11	4.82	11.93	5.06
128	16.24	6.31	11.16	6.33
145	16.97	8.51	15.29	8.31
209	25.34	20.41	22.48	17.83
272	22.16	27.07	11.49	23.93

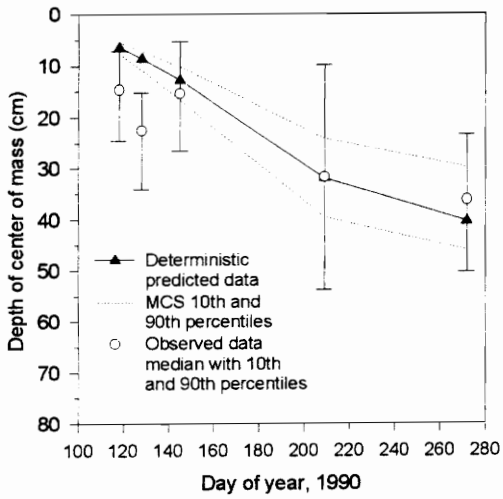
(c) Bromide

Dates	No-till Median (cm)	No-till Predicted (cm)	Conventional Tillage Median (cm)	Conventional Tillage Predicted (cm)
118	17.14	14.89	18.26	11.97
128	27.25	24.01	22.42	19.05
145	31.24	35.94	33.16	30.30
209	43.12	60.93	43.40	56.90
272	42.76	67.17	43.18	63.99

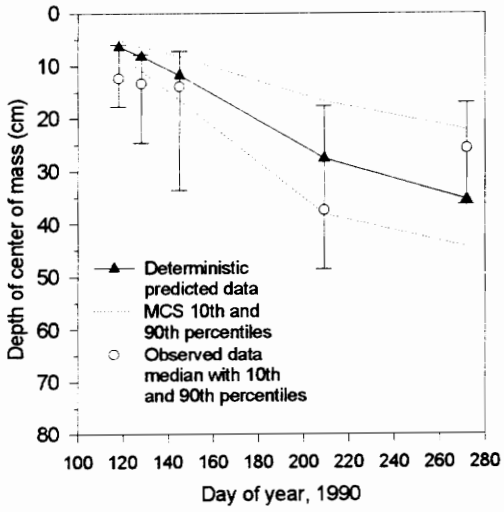
Table 33. Model performance statistics for comparison between deterministic predictions and observed medians for depth of center of mass in the soil.

Chemical	Plot	MdAE (%)	CD*	EF*
Atrazine	No-Till	18	0.56	0.48
	Conv. Till	43	0.20	-2.89
Metolachlor	No-Till	50	0.09	-8.84
	Conv. Till	58	0.13	-7.87
Bromide	No-Till	15	0.70	0.59
	Conv. Till	19	0.48	0.39

Note: If all predicted and observed values were the same, then the statistics would yield: Normalized median absolute error, MdAE = 0.0, Robust coefficient of determination, CD* = 1.0, and Robust coefficient of efficiency, EF* = 1.0 (Zacharias et al., 1996).

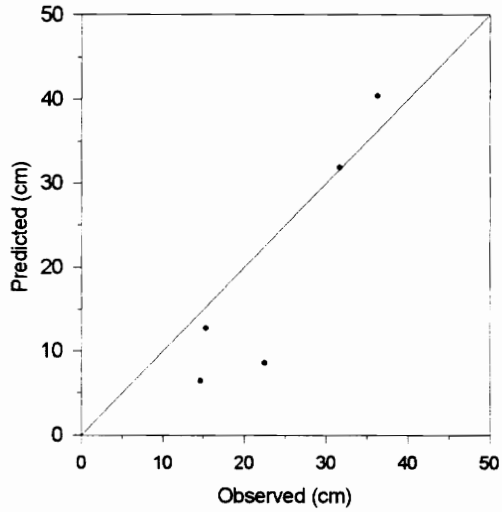


(a) No-till

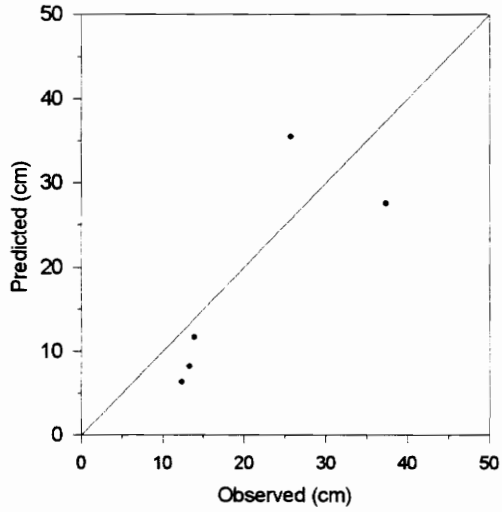


(b) Conventional tillage

Figure 52. Observed values, deterministic predictions, and MCS prediction percentiles for atrazine depth of center of mass for (a) no-till plot and (b) conventional tillage plot.



(a) No-till



(b) Conventional tillage

Figure 53. Observed values vs. deterministic predictions for atrazine depth of center of mass for (a) no-till plot and (b) conventional tillage plot.

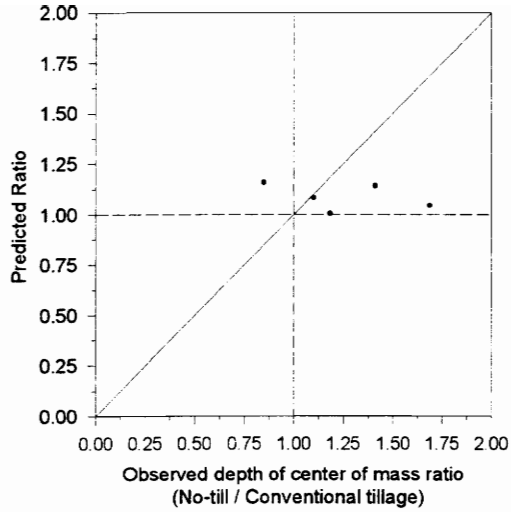


Figure 54. Observed vs. predicted ratios of no-till to conventional tillage atrazine depth of center of mass.

predicted. One of the predicted ratios on this graph falls on the opposite side of one as compared to its corresponding observed ratio, which indicates an incorrect relative prediction.

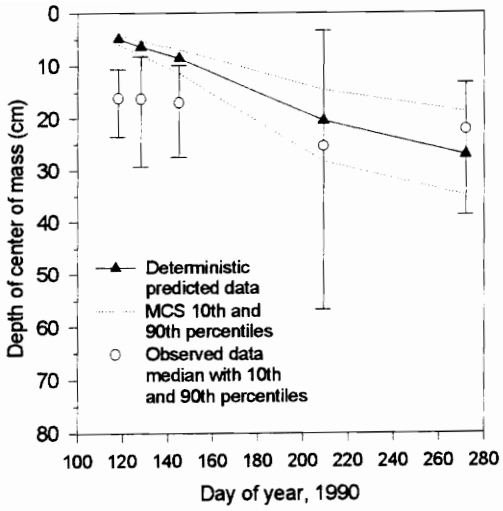
Metolachlor

The observed and predicted values for metolachlor depth of center of mass appear in Table 32. Model performance statistics for comparing between the observed values and deterministic predictions for the metolachlor depth of center of mass appear in Table 33. Graphical comparisons between observed values and deterministic predictions appear in Figures 55 and 56.

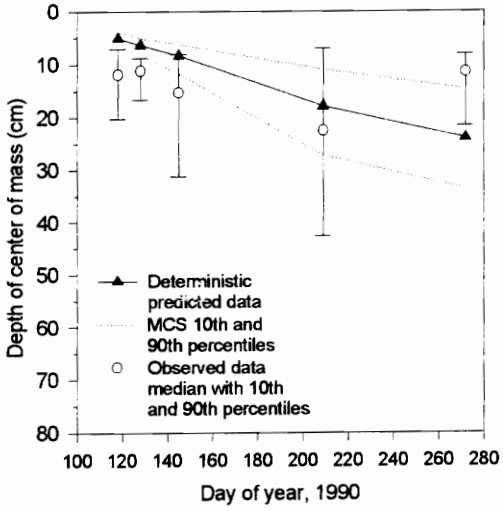
The time-based graphs in Figure 55 show that the model tends to under-predict the depth of center of mass value on the first four days and over-predict on the fifth day for both tillage types. The magnitude-based graphs in Figure 56 as would be expected show the same overall trend of model under-prediction with four of the five values being under-predicted. Since the prediction trends for metolachlor depth of center of mass are similar to those for atrazine, the causes are likely the same. The early under-prediction could be due to slow leaching being simulated by the model and the model over-prediction due to low amounts of plant uptake being simulated by the model.

The model performance statistics that are listed for metolachlor in Table 33 are collectively the fifth and sixth farthest away from the optimum values of the values in Table 33. The EF* statistics for both management practices indicate poor model performance since they are negative.

Observed versus predicted ratios of no-till to conventional tillage metolachlor depth of center of mass appear in Figure 57. As with the atrazine depth of center of mass ratios, this set of predicted ratios falls within a narrow range of values, 0.95 to 1.15, as compared to the range of its corresponding set of observed ratios, 1.00 to 2.00. This graph is significant from the standpoint of the usage of management type models, in that on the day where the predicted ratio was 0.95 the model failed to predict the management

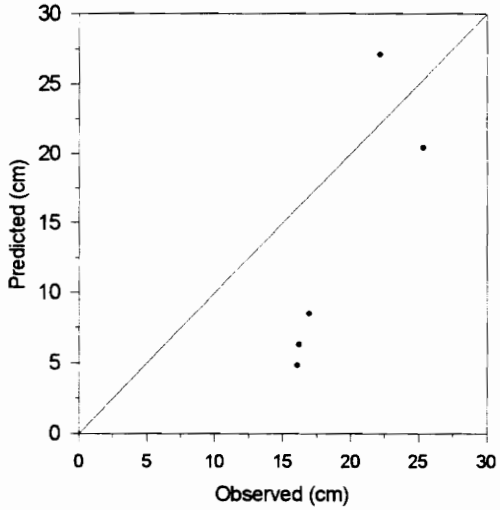


(a) No-till

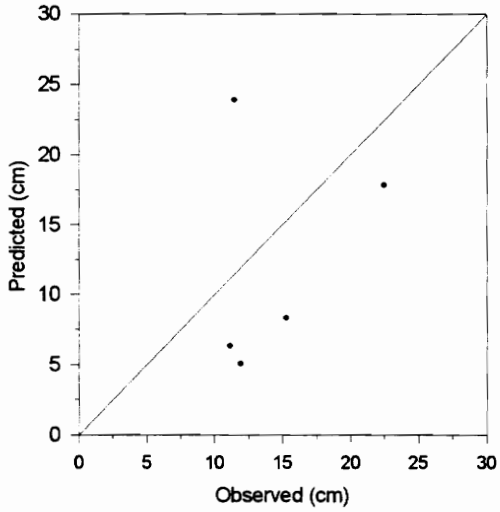


(b) Conventional tillage

Figure 55. Observed values, deterministic predictions, and MCS prediction percentiles for metolachlor depth of center of mass for (a) no-till plot and (b) conventional tillage plot.



(a) No-till



(b) Conventional tillage

Figure 56. Observed values vs. deterministic predictions for metolachlor depth of center of mass for (a) no-till plot and (b) conventional tillage plot.

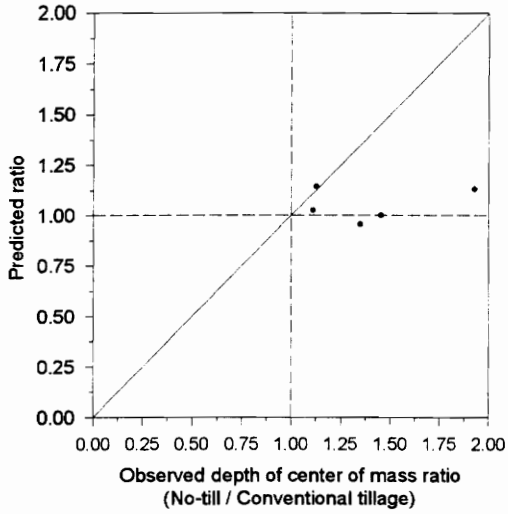


Figure 57. Observed vs. predicted ratios of no-till to conventional tillage metolachlor depth of center of mass.

type with the largest depth of center of mass as compared with the observed values.

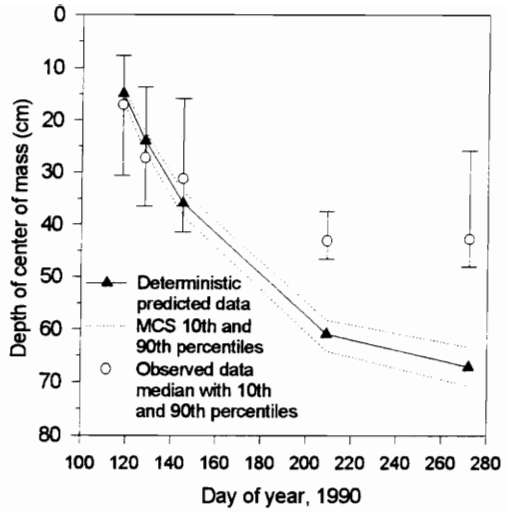
Bromide

Table 32 shows the observed and predicted values for the bromide depth of center of mass data. Table 33 shows the model performance statistics for comparison between the observed values and deterministic predictions for bromide depth of center of mass. Figures 58 and 59 show graphical comparisons between observed values and deterministic predictions for bromide depth of center of mass. Figure 58 uses a time-based scale, while Figure 59 uses a magnitude-based scale.

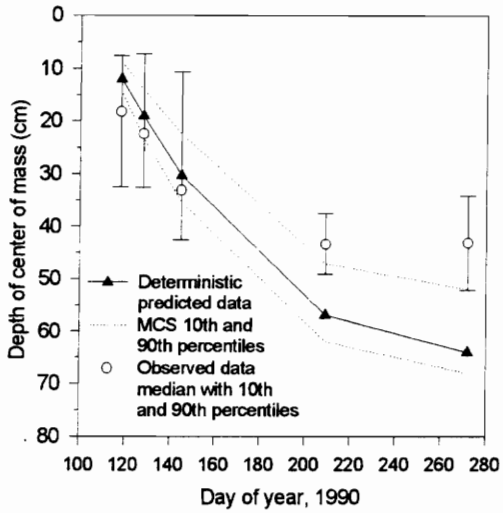
Figure 58 shows that the model predictions tend to be very close to the observed data medians on the first three days, while on the last two days the model produces large over-predictions. The no-till graph in Figure 59 shows a trend for the model to move from a slight under-prediction to over-predictions as the observed magnitudes increase. The conventional tillage graph shows a similar trend. The over-predictions of depth of center of mass on later dates could be attributable to leaching that is simulated to occur too fast or plant uptake that is simulated as being too small.

Surprisingly, considering the large differences between the observed values and deterministic predictions, the bromide model performance statistics are better than most of the other statistics in Table 33. The MdAE values are quite low at 15 and 19 and the EF* values, 0.59 and 0.39, are both positive. The CD* values of 0.70 and 0.48 for the no-till and conventional tillage predictions were the closest and third closest to optimum of the corresponding statistics in Table 33. These statistics may appear to be favorable because the errors in over- and under-prediction counteract each other.

Figure 60 shows the observed versus predicted ratios of no-till to conventional tillage bromide depth of center of mass values. This scattergram shows varying degrees of over-predictions of the ratios between the values for the two tillage types. Four of the predicted ratios fell on the opposite side of one as their corresponding observed ratios.

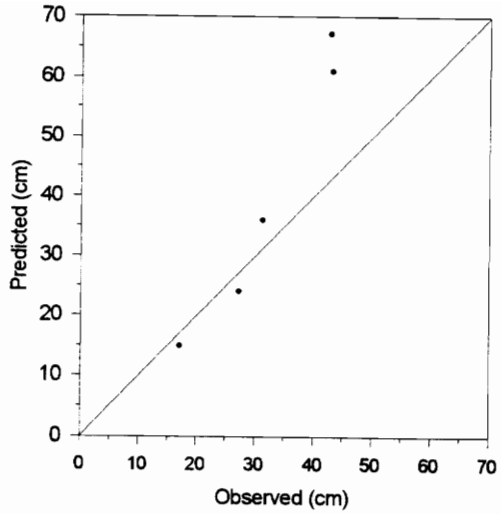


(a) No-till

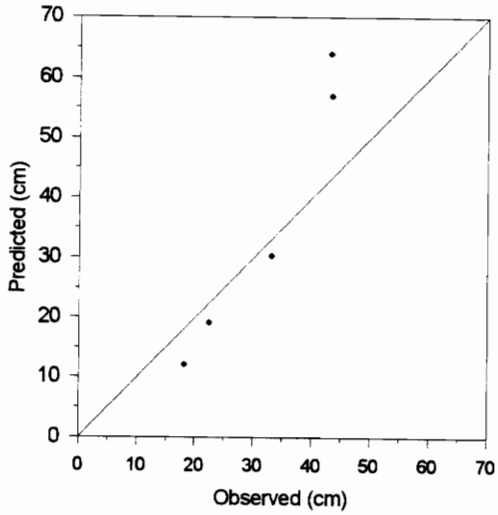


(b) Conventional tillage

Figure 58. Observed values, deterministic predictions, and MCS prediction percentiles for bromide depth of center of mass for (a) no-till plot and (b) conventional tillage plot.



(a) No-till



(b) Conventional tillage

Figure 59. Observed values vs. deterministic predictions for bromide depth of center of mass for (a) no-till plot and (b) conventional tillage plot.

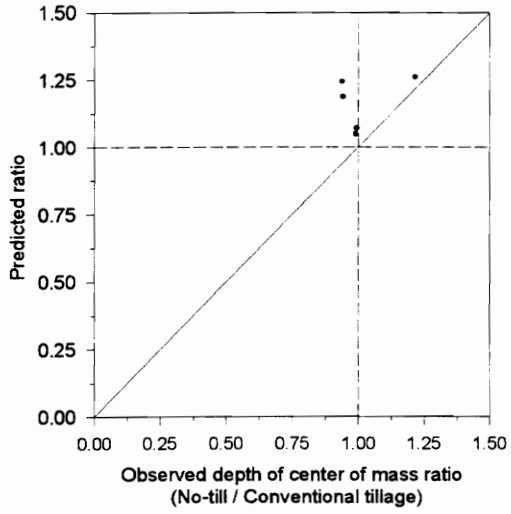


Figure 60. Observed vs. predicted ratios of no-till to conventional tillage bromide depth of center of mass.

Summary

For all three solutes the observed depth of center of mass was slightly under-predicted initially, and over-predicted on the later dates. The initial predictions all visually appeared close to the observed medians. The under-predictions for the later dates appeared to be of very large magnitudes indicating that the model algorithms may not describe the transport process very completely or accurately. Since the observed values indicate movement back to the surface, the errors in prediction may be due to the lack of a volatilization process or an inadequate description of plant uptake. For the depth of center of mass ratios, all the solutes exhibited at least one ratio that failed to match to corresponding observed ratio, thus indicating that the model will not make relative predictions correctly 100 percent of the time.

Chapter Summary

Consistent under-prediction was shown for runoff. Sediment yield was generally over-predicted and pesticide surface losses under-predicted. Mass in the root zone and depth of center of mass did not have consistent prediction trends for all five observed dates, although there were some common trends between management practices for individual dates. For relative predictions, the model failed to match the observed data for all dates, but generally was correct for more than half the dates.

COMPARISON BETWEEN PROBABILISTIC AND DETERMINISTIC EVALUATIONS

METHODOLOGY

Comparison Basis

A qualitative comparison between the deterministic results and the probabilistic results was made to assess how much parametric uncertainty affects the conclusions reached in a typical deterministic evaluation. This comparison was to determine whether the probabilistic evaluation provides model users and developers with any additional or better information in order to assess a model's predictive ability and if the probabilistic procedure can provide consistent evaluation results without uncertainty due to parameter selection. The basis for this evaluation was to determine whether both methods lead to the same conclusions concerning model over- and under-prediction and relative differences between predictions of values for different management practices, and whether or not the probabilistic evaluation provided any relevant information about model performance that the deterministic evaluation did not.

RESULTS AND DISCUSSION

A summary of the model prediction trends that were observed with each evaluation procedure appears as Table 34. The trends in the prediction of the differences between management practices for each of the model outputs appear as Table 35. The probabilities and percentages of relative predictions that matched observed relative relationships between management practices for both evaluation procedures are in Table 36.

Table 34. Prediction trends for model outputs for the probabilistic and deterministic evaluations.

Model Output	Probabilistic Evaluation	Deterministic Evaluation
Runoff	Under-prediction	Under-prediction
Sediment yield	Over- and under-prediction	Over- and under-prediction
Atrazine surface losses	Under-prediction	Under-prediction
Metolachlor surface losses	Under-prediction	Under-prediction
Atrazine mass	Over- and under-prediction	Over- and under-prediction
Metolachlor mass	Over-prediction on last 3 days	Over-prediction
Bromide mass	Over-prediction - first 2 days Under-prediction - last 2 days	Over-prediction - first 3 days Under-prediction - last 2 days
Atrazine depth	Under-prediction - first 2 days	Under-prediction - first 3 days Over-prediction - last day
Metolachlor depth	Under-prediction - first 3 days	Under-prediction - first 4 days Over-prediction - last day
Bromide depth	Over-prediction - last 2 days	Under-prediction - first 2 days Over-prediction - last 2 days

Table 35. Prediction trends in differences and ratios between management practices for model outputs as predicted in the probabilistic and deterministic evaluations.

Model Output	Probabilistic Evaluation (Difference)	Deterministic Evaluation (Ratio)
Runoff	Over-prediction	Over- and under-prediction
Sediment yield	Over-prediction	Over-prediction
Atrazine surface losses	Over-prediction	Under-prediction
Metolachlor surface losses	Over-prediction	Under-prediction
Atrazine mass	Over-prediction	Under-prediction
Metolachlor mass	Under-prediction	Over- and under-prediction
Bromide mass	Over- and under-prediction	Over- and under-prediction
Atrazine depth	Over-prediction	Under- prediction
Metolachlor depth	Under-prediction	Under-prediction
Bromide depth	Over-prediction	Over-prediction

Table 36. Probabilities and ratios of relative predictions which matched observed data for the probabilistic and deterministic evaluations. *

Model Output	Probabilistic Evaluation (Difference)	Deterministic Evaluation (Ratio)
Runoff	0.272-0.834	7/7
Sediment yield	0.272-0.936	5/5
Atrazine surface losses	0.000-0.812	3/4
Metolachlor surface losses	0.081-0.800	4/5
Atrazine mass	0.004-0.996	3/5
Metolachlor mass	0.044-0.946	3/5
Bromide mass	0.460-0.968	4/5
Atrazine depth	0.080-0.924	4/5
Metolachlor depth	0.368-0.844	3/5
Bromide depth	0.000-1.000	1/5

* Values in the second column are extreme high and low relative prediction probabilities for each output. Ratios in third column are the number of predicted relative ratios for each output which fell on the same side of one as the observed relative ratio for a given date versus the total number of dates.

Surface Outputs

Runoff

Both types of evaluations show varying degrees of under-prediction of observed runoff values (Figures 1, 2, 31, and 32). The under-predictions for runoff shown with both procedures reflect observations about the performance of GLEAMS for runoff from other studies. The deterministic graph (Figure 31) is more understandable over the entire set of observed events as compared to the I_p scatter graph (Figures 1 and 2) and its associated daily distribution graphs in the probabilistic evaluation. Part of the reason that the deterministic graph is more understandable is that the magnitudes of the runoff events are included right on the graph. With the probabilistic index of model performance graph, there are no indications of which events are the largest or what the magnitude of the difference between the observed and predicted values is. For the probabilistic index graph, including another graph of the event magnitudes next to it or looking at the distributional graph used in determining each index value can help with the problems mentioned above, but this process is more tedious than examining the deterministic graph. The MCS percentiles plotted on the deterministic graph show that the deterministic predictions fall near the middle of the percentiles, which indicates that the model's response for this output is approximately linear.

For the comparisons between management practices, the deterministic evaluation (Figure 33) shows a variety of over- and under-predictions of the observed ratios, while the probabilistic evaluation (Figure 3) shows a definite trend towards over-predicting the differences between the management practices. Of more significance to the application of models for management practice evaluation is that the deterministic evaluation (Figure 33) predicted the management practice with the highest runoff correctly for all dates, while the probabilistic evaluation (Table 7) predicted the observed sign correctly with a probability between 0.272 and 0.834. The probabilistic evaluation probabilities for correctly predicting the observed differences between management practices indicate that no

relative prediction is 100 percent certain, however, the correct relative deterministic predictions indicate that a deterministic simulation with carefully selected inputs can make correct relative predictions.

Sediment Yield

Over- and under-prediction of sediment yield values occur somewhat randomly for both evaluations (Figures 4, 5, 34, and 35). On the larger events, though, the trends seem to match each other between the two evaluation procedures, that is both procedures show over-predictions for the runoff events on days 146 and 149 and under-predictions for the events on day 241 for both tillage types. For the event on day 235, for both evaluation procedures under-predicted for the conventional tillage plot and over-predicted for the no-till plot. Since neither procedure showed a single prediction trend, the model does not seem to have any systematic error in its sediment yield component. As with the examination of runoff for the two procedures, the deterministic event-based graphs were easier to understand and provided more information by themselves than the corresponding probabilistic index graphs. The predictions made in the deterministic runs of the model for sediment yield fell near the center of the MCS percentiles as shown on the deterministic event-based graphs (Figure 34). The deterministic predictions being near the center of the MCS percentiles indicates that the model tends to have a linear response for sediment yield.

The deterministic comparison between management practices (Figure 36) generally showed over-prediction of ratios between the two management practices, while the probabilistic evaluation (Figure 6) showed a mix of over- and under-predictions of differences between the two management practices. The deterministic procedure always predicted which management practice produced the highest amount of sediment yield. For the probabilistic procedure though, the range of the probabilities for correct signs went from 0.272 to 0.936 (Table 8), indicating that there was a fair chance that the model would fail to predict the management practice with the highest value. From these results,

it would seem that the deterministic procedure would produce better results than the probabilistic procedure. Concluding that the deterministic procedure produced the best results, neglects the additional information gained by including observed and predicted variabilities in the deterministic procedure.

Atrazine Surface Losses

Both evaluation procedures (Figures 7, 8, 37, and 38) show that the model tended to under-predict atrazine surface losses with the exception of day 117 for the conventionally-tilled plot. Consistent under-prediction of atrazine surface losses indicates that there is systematic error in the model predictions and thus possible flaws in how the model simulates pesticide transport. This under-prediction may be linked to the under-prediction of runoff or some other component such as leaching or pesticide degradation. The deterministic event-based graphs were easier to comprehend than their probabilistic index counterparts.

The deterministic event-based graph also showed the MCS 10th and 90th percentiles for the atrazine surface losses. For the initial event on the conventional tillage plot, the deterministic prediction is much closer to the 10th percentile of the MCS than the 90th percentile. The closeness of the deterministic prediction to one of the extreme percentiles of the MCS indicates that the model has a non-linear response for this output variable. In this particular case, the smallness of the deterministic prediction in comparison to the 90th MCS percentile indicates that in a deterministic mode the model may drastically under-estimate the risk of a large surface loss of this pesticide.

The deterministic evaluation procedure (Figure 39) tended to under-predict the ratios between the predictions for the two management practices, except when both management practices resulted in 0's for their predictions, while the probabilistic procedure (Figure 9) over-predicted the differences in atrazine surface losses for the two practices. The probabilistic procedure (Table 9) showed extreme variability in predicting the correct sign of observed differences between management practices with probabilities

of a correct prediction ranging from 0.081 to 0.812, while all the non-zero deterministic predictions were on the correct side of one (Figure 39). This range of probabilities indicates that there may not be a single relative relationship between management practices for this output, but a chance that either management practice could result in the largest surface loss of atrazine.

Metolachlor Surface Losses

As with the atrazine surface losses, both evaluation procedures (Figures 10, 11, 40, and 41) tended to show under-predictions for all dates and management practices, with the exception of day 117 for the conventional tillage plot. As with the atrazine surface losses, these under-predictions indicate systematic error, and thus possibly a flaw in a component of the model.

The deterministic prediction for surface losses for the first event on the conventional tillage plot was much closer to the 10th MCS percentile than the 90th percentile, thus indicating non-linear response of the model for this output. The non-linear response of the model for this output indicates that the deterministic model simulation may under-estimate the threat of a large surface loss of this pesticide for the days immediately following pesticide application.

Of significance for the evaluation of model algorithms is the ability to see similar relative trends in the magnitude of the under-predictions between management practices for the probabilistic procedure when using a scatter graph of probabilistic indices for model performance (Figures 10 and 11). This trend can mainly be attributed to the model equations and algorithms since it occurs with two different sets of observed data and two different model parameter sets. The problem with a Monte Carlo-based procedure such as this one, as compared to First Order Uncertainty Analysis, is that errors cannot be related back to a specific parameter or equation, therefore, it will be difficult to determine what part of the model or input to change to correct the observed errors.

When comparing between management practices, the deterministic procedure (Figure 42) under-predicted the ratios between management practices, while the probabilistic procedure over-predicted the differences between the practices (Figure 12). The probability that the correct sign was predicted for differences between management practices ranged from 0.081 to 0.800 in the probabilistic procedure. The deterministic ratios between the management practices fell on the same side of one as the observed ratios on four of five days.

Solute Mass in the Root Zone

Atrazine

As would be expected, the deterministic evaluation graphs (Figure 43) reflect the differences between the predicted and observed values as they appear at the 50th percentile of the probabilistic procedure graphs (Figures 13 and 14). What the deterministic procedure does not show is where the model moves from over- to under-prediction of a certain day's values as occurs on day 145 for the no-till data (Figure 13). At the highest percentiles (0.65 and above) the model under-predicts, but below the 0.65 level the model tends to over-predict. This observation negates the use of model prediction trends at specific percentiles, because the model obviously can both under- and over-predict for the same day's data. Additionally, the probabilistic methodology shows that the MCS predicted distribution does not have as much variability in it as the empirical distribution (Figures 13 and 14). This observation cannot be made based on any of the deterministic graphs.

It can be seen from the deterministic graphs (Figures 43 and 44) that the deterministic predictions fall virtually in the middle of the range of the MCS distribution percentiles. This observation, which indicates that the model response for this output is linear, is slightly unexpected since the model as a whole would be considered a non-linear model. The deterministic graphs (Figures 43 and 44) also show that the model outputs are

within the range of the observed data for all but one day, which indicates that the model outputs seem reasonable.

In the two comparisons between management practices, the deterministic evaluation generally showed under-predictions of the ratios between no-till and conventional tillage mass in the root zone (Figure 45), while the probabilistic procedure showed under-predictions of the differences in atrazine mass between the two management practices (Figure 15). As for predicting which management practice resulted in more atrazine mass in the soil, the deterministic procedure correctly made this prediction for three out of the five days, while the probability for a correct prediction in the probabilistic procedure ranged from 0.080 to 0.924 (Figure 45 and Table 12).

Metolachlor

As with atrazine mass, the over-predictions shown on the deterministic graphs (Figure 46) are reflective of the differences between the empirical and predicted distributions at the 50th percentile in the probabilistic evaluation (Figures 16 and 17). This over-prediction trend is indicative of the trend for all percentiles for the no-till data, but not for the first two days of the conventional tillage probabilistic predictions. On the first two days of the conventional tillage probabilistic graphs (Figure 17), the empirical and predicted distributions cross each other, thus creating ranges of both model over- and under-prediction, which means that the predictive trend shown on the deterministic graphs does not hold for all of the probabilistic percentiles.

The deterministic graphs (Figure 46) show that the MCS percentiles are much wider than was observed for the atrazine mass percentiles and these percentiles do not fall as much within the range of the observed data. Both of these occurrences serve to point out that even after trying to account for parameter uncertainty with the probabilistic procedure, there is still possibly uncertainty due to input parameter selection. These graphs (Figure 46) emphasize that mass in the root zone as predicted by GLEAMS tends

to follow a linear response since the deterministic results again fall near the middle of the 10th and 90th MCS percentiles.

For the comparison between management practices with the two methods, the deterministic ratio graph (Figure 48) showed a combination of both over- and under-prediction of the ratios between the two management practices, while the probabilistic procedure (Figure 18) showed all under-predictions of the differences in metolachlor mass for the two management practices. The probabilistic procedure predicted the correct sign of differences between management practices with a probability between 0.044 and 0.946. The inability of the model to make relative predictions which match observed data all the time is indicated by the deterministic procedure only making the correct prediction on three of five dates for metolachlor mass.

Bromide

The results for the tracer, as with the previous two chemicals, show that the comparison between deterministic predictions and observed medians (Figure 49) matches the probabilistic evaluation results at the 50th percentile on the first three dates (Figures 19 and 20). One unique aspect for the tracer is that the deterministic results on the last two dates tended to fall close to the higher percentiles of the probabilistic evaluation results, rather than falling in the middle of them (Figure 49). This result indicates that leaching outputs from the model follow a nonlinear trend. In general, the differences between the predicted and observed data in the deterministic evaluation are indicative of the differences between the empirical and predicted distributions in the probabilistic evaluation except for where the two distributions cross at the extreme percentiles (Figures 19 and 20). The ability of the model to make realistic predictions is indicated by the majority of the MCS 10th and 90th percentiles being encompassed within the range of the observed percentiles (Figure 49). The only exceptions to this occur on the first three days of the no-till predictions when the MCS 90th percentiles are greater than the observed 90th percentiles.

As for the comparisons between management practices for the two procedures; generally, both show inconsistent trends in over- and under-predictions (Figures 21 and 51). The deterministic procedure failed to predict the ratio of bromide mass between the two management practices on the same side of one on one of the five days (Figure 51), while the probability of the probabilistic procedure predicting the correct sign for differences in management practices ranged between 0.460 and 0.968 (Table 16).

Depth of Solute Center of Mass

Atrazine

The deterministic procedure graphs (Figure 52) for atrazine depth of center of mass show the same magnitude of differences between observed and predicted data and under- and over-prediction trends as is exhibited on the probabilistic procedure graphs (Figures 22 and 23) at the 50th percentile levels. On about half the dates, the over- or under-prediction trend holds for nearly the entire range of probabilistic data, while on the other half the two distributions cross each other resulting in different prediction trends for different percentiles of the data (Figures 22 and 23). Observations of prediction trends for deterministic data on days where the probabilistic distributions cross each other could thus vary if based on different inputs.

The deterministic graphs (Figure 52) show that the model creates depth of center of mass values within an acceptable range since the MCS percentiles generally fall within the range of the observed data. Exceptions to this statement occur for the second day's values on the no-till plot and the fifth day on the conventional tillage plot. As with the mass in the root zone graphs, the deterministic predictions for atrazine depth of center of mass tend to fall in the middle of the MCS percentiles (Figure 52), thus indicating that over the range of values used as inputs for these simulations the model exhibits a linear response for depth of center of mass.

Both procedures tend to show under-prediction when comparing between management practices (Figures 24 and 54). Four of five days' ratios for comparison between no-till and conventional tillage in the deterministic evaluation have the correct magnitudes (Figure 54), while the probability of the correct sign being predicted for differences between the two management practices for the probabilistic procedure ranged between 0.080 and 0.924 (Table 18).

Metolachlor

The comparison between the deterministic and probabilistic evaluations for metolachlor depth of center of mass matched the same trends observed for the previous data sets, that is, the deterministic evaluations (Figure 55) show the same trends in over- and under-prediction as the probabilistic evaluations (Figure 25 and 26) at the 50th percentiles. These prediction trends are not always indicative of the over- and under-prediction trends over the rest of the probabilistic data range of percentiles. The metolachlor depth of center of mass predictions also tend to fall within the range of the observed data indicating that the model is performing adequately (Figure 55). The model shows a linear response in model output for metolachlor depth of center of mass as indicated by the deterministic simulations of the model falling nearly in the middle of the 10th and 90th MCS percentiles.

The deterministic evaluation (Figure 57) showed under-prediction of the ratio between the depth of center of mass for the no-till and the conventional tillage plot on four of the five days, while the probabilistic evaluation (Figure 27) also showed under-prediction of the difference in metolachlor depth of center of mass for the two management practices on four of the five days. The deterministic evaluation procedure predicted the magnitude of the ratio incorrectly for the observed data on one date (Figure 57), while the probability of the correct sign for the differences between the management practices ranged between 0.368 and 0.844 in the probabilistic procedure (Table 20).

Bromide

The comparison between the deterministic and probabilistic procedures for bromide depth of center of mass showed again that the prediction trend on the deterministic graphs (Figure 58) is generally reflective of the trends that occur at the 50th percentile level on the probabilistic evaluation graphs for the first three observed days (Figures 28 and 29). One difference between this evaluation and that of the evaluations of the depth of center of mass for the other chemicals, is that the deterministic predictions do not fall exactly in the middle of the probabilistic percentiles on the last two days for the conventional tillage data (Figure 58). The deterministic predictions on these two dates tend towards the 90th percentile edge of the data. This trend is similar to the trend observed in the bromide mass data for the same two days. The MCS percentiles shown on the deterministic graphs (Figure 58) for bromide depth of center of mass indicate that these depth of center of mass predictions may not be as adequate as the predictions of depth of center of mass for the pesticides. The MCS percentiles for bromide fall in the range of the observed data for the first three dates, but are much higher than the observed values on the last two dates, especially for the no-till plot.

The ratios between the two management practices are over-predicted in the deterministic evaluation for all days (Figure 60). In the probabilistic evaluation, the differences between the two management practices are all over-predicted, generally being extremely over-predicted (Figure 30). In the deterministic evaluation (Figure 60), two of the five days had the ratio between the management practices predicted incorrectly, while the probabilistic procedure (Table 22) had several probabilities of correctly predicting the differences between the two management practices of 0.000.

Chapter Summary

For the surface outputs, the deterministic graphs for viewing observed and predicted data were more understandable than the corresponding probabilistic index of model performance graphs. The index graphs tended to more consistently show similar

relative prediction trends between different model scenarios, though. Generally, the deterministic graphs showed over- or under-prediction of data on each day based on whether over- or under-prediction occurred at the 50th percentile on the probabilistic graphs. Over- or under-prediction at the 50th percentile of the probabilistic graphs was not always a reliable predictor of over- or under-prediction trends over the remaining range of data, therefore the deterministic prediction trends can not always be assumed to be constant over the entire observed time period. Also, the deterministic data tended to fall in the center of the probabilistic percentiles, with the exception of the last two days for the bromide data, which indicates that over this range of inputs the model has a linear response for its runoff, sediment yield, mass, and depth of center of mass outputs.

When predicting differences between management practices, the two evaluation procedures on several occasions failed to show the same over- and under-prediction trends. Additionally, both procedures were not guaranteed of predicting which management practice would produce a higher value than another practice as indicated by the deterministically predicted ratios appearing on the opposite side of one from the observed ratios and the probabilistic probability of a correct sign often being less than 0.5.

In general the probabilistic procedure did not eliminate the effects of parameter uncertainty in model evaluation. This procedure did, however, improve upon previous types of evaluation by including both model predicted and observed variability in the model evaluation process.

SUMMARY AND CONCLUSIONS

A comparison between results from a deterministic evaluation and a probabilistic evaluation of GLEAMS was conducted to assess the effects of parameter uncertainty on model evaluation. The data used for both types of evaluations came from a field study that was used to characterize the fate and transport of atrazine, metolachlor, and bromide from fields in the Coastal Plain region of Virginia (Zacharias, 1992). Data were taken from two 18x27 m plots in a field that was in the second year of a two-year no-till wheat-beans-corn rotation. One of the plots was conventionally tilled prior to the planting of the corn.

Inputs for the probabilistic evaluation of the model were created from a combination of field testing and best estimates of values from literature sources. The distributions used for the inputs were either normal, lognormal, beta, triangular, or uniform. The inputs for the deterministic evaluation came from central tendency values of the distributions that were used to describe the inputs for the probabilistic evaluation. For the probabilistic evaluation, the model was run 5000 times for each management practice in a Monte Carlo simulation mode to create output distributions. In the deterministic evaluation, the model was run once for each management practice to obtain a single set of output values. The results from both evaluations were judged using a combination of graphical and statistical methods. The ability of the model to predict differences in management practices was also evaluated.

Probabilistic Evaluation

(1) The model under-predicted runoff, atrazine surface losses, and metolachlor surface losses.

(2) The fit between model predicted distributions and empirical distribution functions for solute mass in the root zone and depth of center of mass was not very good based on visual and quantitative comparisons. Differences between the two types of

distributions came from differences in both the magnitude of the values in the distributions and in the variance of the distributions.

(3) The probability that relative predictions matched observed relations between management practices varied within a wide range of probabilities.

Deterministic Evaluation

(1) The model generally predicted the runoff and sediment yield observed values within an order of magnitude. Runoff was under-predicted.

(2) Pesticide surface losses were over-predicted for the first observed storm event and then under-predicted for subsequent storm events for both pesticides for the conventional tillage plots. Surface losses were under-predicted for all events for both pesticides for the no-till plots.

(3) The model generally over-predicted the mass of solute in the root zone on the first three days of observed data.

(4) The depth of center of mass for the three solutes was slightly under-predicted for the first three days of observed data.

(5) The model prediction of the management practice which produced the largest pollutant loading was the same as the observed relationship between pollutant loadings as a result of different management practices for the majority of the date and solute combinations.

Comparison Between Evaluations

(1) Both procedures showed under-prediction of runoff and pesticide surface losses.

(2) The two procedures did not always correctly predict the differences between management practices for the runoff, sediment yield, and surface loss values.

(3) The deterministic results for mass in the root zone and depth of center of mass reflected the differences between data distributions in the probabilistic evaluation at the 50th percentile level. The deterministic results did not consistently represent the

prediction trends observed over the entire range of percentiles from the probabilistic evaluation.

(4) For mass in the root zone and depth of center of mass, the probability of the correct observed sign of differences between management practices being predicted in the probabilistic procedure ranged from 0.0 to 1.00. This means that there is a possibility that a parameter set used in a deterministic evaluation may predict incorrect differences between the results for two management practices. Examples of incorrect relative predictions by the deterministic procedure were seen in several instances.

The results of this study must be viewed in light of the fact that they come from a relatively short time period and a data set that has not been replicated. Statements about model performance apply only to the limited set of soil and management conditions observed for one particular field and should not be arbitrarily extrapolated to other soil, management, and climatic conditions. The limited conditions that this evaluation applies to points out some of the inherent difficulties in validating nonpoint source pollution models when there is a limited number of data sets available to do so.

For many of the model outputs, adjustment of the parameters used to describe the input distributions would have improved model performance. Parameter uncertainty thus, will still exist in the probabilistic procedure, necessitating the need for careful parameter selection. Despite parameter uncertainty remaining in model evaluation, the probabilistic procedure is an improvement over deterministic procedures in that it includes both model output and observed variances in the evaluation. By including these variabilities, this procedure primarily accounted for the spatial variability of model variables throughout the observed plots. This procedure eliminates arbitrary selection of inputs from a data range since it uses the entire range. This characteristic should make model results obtained with this procedure more consistent between different users.

It appears that altering the central tendency values of the ranges of model inputs in a form of calibration could improve model performance for some outputs. Obtaining

additional site specific data instead of using values obtained from the literature could also improve model performance. An example of improved model performance by using site specific data could occur for model outputs involving metolachlor. The metalochlor half-life obtained from the literature was 90, while the half-life obtained from the field study at this site was 45. If the field study half-life of 45 were used in this study, the amount of metolachlor remaining in the soil would be reduced and in this case model performance improved. The existence of differences between the half-life obtained from field data and the literature reinforce the idea of uncertain input parameters and the need to obtain the best possible site-specific data for model simulations. The drawback to using calibration and site-specific data, such as this half-life, is that they are not always possible for every NPS modeling scenario and are therefore undesirable for a procedure for the general instance of evaluation and usage of NPS models. The inaccuracies that remained in the model predictions due to parameter selection and lack of field data, point out the need for the utmost care to be used in obtaining the best quality values possible for use as inputs to the model.

The model seems to make realistic predictions for most output variables, since the majority of the predicted percentiles fall within the ranges of the observed data. The model variances tended to be smaller than the observed variances, though. Model output variance being consistently smaller than the variance of the observed data is an indication that the model is a simplification of the real-world environment and has not accounted for all factors that possibly affect pesticide fate and transport. Several possible model flaws that were noted in the course of the evaluation included model under-prediction of runoff and pesticide losses and the inability of the model to simulate an increase in depth of center of mass for bromide on the last observed date. These statements about model errors are by no means definitive due to the possibility of parameter error also leading to these same conclusions.

In the case of the surface loss data for the pesticides, the usage of the probabilistic index of model performance made it possible to detect distinct trends in model predictions

for both management practices. Distinct trends in model prediction were not nearly as evident in the comparison between management practices in the deterministic procedure. The distribution comparison portion of the probabilistic evaluation procedure would allow model developers to evaluate model performance over the entire range of possible results rather than one specific percentile as occurs in the deterministic evaluation.

While the predicted data generally fell within the range of the observed data, the goodness-of-fit tests between distributions indicated that the predicted and empirical data did not fit very well. Since intuitively the two types of distributions seemed to match each other well, these types of goodness-of-fit tests may be too stringent to be applied to evaluating NPS models at this point. The goodness-of-fit tests being too stringent for this type of evaluation point out another difficulty in validating NPS models, which is the lack of criteria for acceptable validation of these models. With these models increasingly being utilized for regulatory purposes, determining a threshold for a model to be considered sufficiently accurate will become increasingly critical. Acceptance criteria are dependent on the intended application of a given model and the output from the model being examined.

Despite the imperfections in the relative predictions of both procedures, they can be used to make management decisions through qualitative analysis. Factoring the magnitude of the predicted events into the evaluation of the surface outputs would likely improve decisions made based on these outputs. As noted in the probabilistic results, the three largest events for surface losses of atrazine resulted in the largest pollutant loading coming from the conventional tillage practice in over 80 percent of the trials. Thus, if a manager were concerned about these types of loadings he/she would choose to implement no-till. Also, even though the probabilistic and deterministic procedures do not predict the largest pollutant loadings from the same practice every time, generally the majority of the predictions will indicate that the largest loadings come from a certain practice. For instance, for atrazine depth of center of mass in the probabilistic procedure, atrazine had leached the farthest for the no-till scenario in more than 68 percent of the trials for four

out of the five dates. With the deterministic procedure, atrazine was deeper in the no-till plot for four of five dates. Based on either of these two procedures, a decision maker would have chosen conventional tillage to reduce atrazine leaching, if that were the desired objective. Subjective evaluation of predictions from a management type model such as GLEAMS can thus be used for decision making for single criterion objectives.

RECOMMENDATIONS

Based on the results from the probabilistic and deterministic procedures and the comparison between the two procedures, the following recommendations are made for future work:

(1) Determine appropriate trial numbers for the Monte Carlo simulations used in creating model predicted distributions.

(2) Determine appropriate Kolmogorov-Smirnov and Anderson-Darling goodness-fit-test rejection levels for evaluating the fit between the model predicted distribution and observed EDF.

(3) Develop of a global sensitivity parameter as proposed by Kumar (1995a) in order to allow for the focus of resources in the probabilistic procedure on the most sensitive parameters.

(4) Compare relative model predictions in the probabilistic procedure using the mean and median as the central tendency indicators for the sampling distributions.

REFERENCES

Addiscott, T. M., and R. J. Wagenet. 1985. Concepts of solute leaching in soils: a review of modeling approaches. *J. of Soil Sci.* 36:411-424.

American Society for Testing and Materials. 1992. Standard practice for evaluating mathematical models for the environmental fate of chemicals. E 978-92, 704-711. Philadelphia, Pa.

American Society of Civil Engineers. 1993. Criteria for evaluation of watershed models. ASCE Task Committee of the Watershed Management Committee. *J. of Irrigation and Drainage Eng.* 119(3):429-443.

Bailey, G.W., A.R. Swank, Jr., and H.P. Nicholson. 1974. Predicting pesticide runoff from agricultural land: A conceptual model. *J. of Environmental Quality* 3:95-102.

Baker, J. L., and H. P. Johnson. 1979. The effect of tillage systems on pesticides in runoff from small watersheds. *Transactions of the ASAE* 22(3):554-559.

Beasley, D. B., and L. F. Huggins. 1981. *ANSWERS User's Manual.* EPA 905/9-82-001. Chicago: U.S. EPA.

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

Beven, K., and A. Binley. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* 6:279-298.

Bouraoui, F. 1994. Development of a Continuous, Physically-Based, Distributed Parameter, Nonpoint Source Model. Ph.D. diss. Dept. of Agricultural Engineering, Virginia Polytechnic Institute and State Univ., Blacksburg.

Brady, N. C. 1990. *The Nature and Properties of Soils.* New York: Macmillan Publishing Company.

Burnside, O. C., C. V. Fenster, G. A. Wicks, and J. V. Drew. 1969. Effects of soil and climate on herbicide dissipation. *Weed Sci.* 17:241-244.

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., R. S. Parrish, R. L. Jones, J. L. Hansen, and R. L. Lamb. 1988a. Characterizing the uncertainty of pesticide leaching in agricultural soils. *J. of Contaminant Hydrol.* 2:111-124.

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 Hydrol.* 2:125-138.

Clarke, R. T. 1973. A review of some mathematical models used in hydrology, with observations on their calibration and use. *J. of Hydrol.* 19:1-20.

Cooke, R. A, S. Mostaghimi, and F. E. Woeste. 1993. VTFIT: A microcomputer-based routine for fitting probability distribution functions to data. *Applied Eng. in Agriculture* 9(4):401-408.

Davis, M. L., and D. A. Cornwell. 1991. *Introduction to Environmental Engineering.* New York: McGraw-Hill, Inc.

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, Mich.: ASAE.

Dean, J. D. 1983. Potency factors and loading functions for predicting agricultural nonpoint source pollution. In *Agricultural Management and Water Quality*, eds. F. W. Schaller and G. W. Bailey, 153-177. Ames: Iowa State University Press.

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. Athens, Ga.: U.S. EPA.

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.

Engman, E. T. 1986. Roughness coefficients for routing surface runoff. *J. of Irrigation and Drainage Eng.* 112(1):39-53.

Fetter, C. W. 1993. *Contaminant Hydrogeology.* New York: Macmillan Publishing Company.

Gilmour, P. 1973. General validation procedure for computer simulation models. *The Australian Computer J.* 5:127-130.

Green, I. R. A. and D. Stephenson. 1986. Criteria for comparison of single event models. *Hydrological Science J.* 31(3):395-411.

Haan, C. T. 1989. Parametric uncertainty in hydrologic modeling. *Transactions of the ASAE* 32(1):137-146.

Haan, C. T., B. Allred, D. E. Storm, G. Sabbagh, and S. Prabhu. 1995. Statistical procedure for evaluating hydrologic / water quality models. *Transactions of the ASAE* 38(3):725-733.

Heatwole, C. D., S. Mostaghimi, T. A. Dillaha, S. Zacharias, and R. W. Young. 1992. Fate and transport of pesticides in a Virginia Coastal Plain soil. Bulletin 175. Blacksburg: Virginia Water Resources Research Center.

James, L. D., and S. J. Burges. 1982. Selection, calibration, and testing of hydrologic models. In *Hydrologic Modeling of Small Watersheds*, eds. C. T. Haan, H. P. Johnson, and D. L. Brakensiek, 237-472. St. Joseph, Mich.: ASAE.

Jones, R. L., and P. S. C. Rao. 1988. Reflections on validation and applications of unsaturated zone models. In *Proc. Int. Conference and Workshop on Validation of Flow and Transport Models for the Unsaturated Zone*, eds. P. J. Wieranga and D. Bachelet, 197-205. Las Cruces: New Mexico Univ.

Jury, W. A. 1985. Spatial variability of soil physical parameters in solute migration: A critical literature review. EA-4228. Palo Alto, Calif.: Electric Power Research Institute

Jury, W. A., W. R. Gardner, and W. H. Gardner. 1991. *Soil Physics.* 5th edition. John Wiley and Sons.

Jury, W. A., W. F. Spencer, and W. J. Farmer. 1983. Behavior assessment model for trace organics in soil: I. Model description. *J. of Environmental Quality* 12:558-564.

Knisel, W. G. 1980. CREAMS: A field scale model for chemicals, runoff, and erosion from agricultural management systems. Conservation Research Report No. 26. Washington, D.C.: USDA.

Konikow, L. F., and J. D. Bredehoeft. 1992. Ground-water models cannot be validated. *Advances in Water Resources* 15:75-83.

Kumar, D. 1995a. Parameter uncertainty in nonpoint source pollution modeling. Ph. D. diss. Dept. of Biological Systems Engineering, Virginia Polytechnic Institute and State Univ., Blacksburg.

Kumar, D. 1995b. Personal communication. Graduate Research Assistant, Dept. of Biological Systems Engineering, Virginia Polytechnic Institute and State Univ., Blacksburg.

Kumar, D and C. D. Heatwole. 1995. Probabilistic approaches to evaluating nonpoint source pollution models. In *Proc. Int. Symp. on Water Quality Modeling*, 164-176. St. Joseph, Mich.: ASAE.

Kumar, D., C. D. Heatwole, and S. J. Thomson. 1994. Direct Monte Carlo simulation of deterministic hydrologic models using batch processing. ASAE Paper No. 94 - 2504. St. Joseph, Mich.: ASAE.

Kumar, D., C. D. Heatwole, and F. E. Woeste. Towards a protocol for BMP assessment using field-scale NPS pollution models. ASAE Paper No. 92-2635. St. Joseph, Mich.: ASAE.

Law, A. M., and W. D. Kelton. 1991. *Simulation Modeling and Analysis.* New York: McGraw-Hill, Inc.

Leonard, R. A. 1990. Movement of pesticides into surface waters. In *Pesticides in the Soil Environment: Processes, Impacts, and Modeling*, ed. H. H. Cheng, 303-349. Madison, Wis.: Soil Science Society of America.

Leonard, R. A., and W. G. Knisel. 1990. Can pesticide transport models be validated using field data: Now and in the future? Departmental Publication No. 3. Tifton: Agricultural Engineering Dept., The Univ. of Georgia, Coastal Experiment Station.

Leonard, R. A., W. G. Knisel, and D. A. Still. 1987. GLEAMS: Groundwater loading effects of agricultural management systems. *Transactions of the ASAE* 30(5):1403-1418.

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.

Loague, K. M., and R. E. Green. 1991. Statistical and graphical methods for evaluating solute transport models: Overview and application. *J. of Contaminant Hydrol.* 7:51-73.

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 Hydrol.* 5:171-194.

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 Hydrol.* 4:139-161.

Luis, S. J. and D. McLaughlin. 1992. A stochastic approach to model validation. *Advances in Water Resources* 15:15-32.

Menelik, G., R. B. Reneau, Jr., D. C. Martens, T. W. Simpson, and G. W. Harkins. 1990. Effects of tillage and nitrogen fertilization on nitrogen losses from soils used for corn production. Bulletin 167. Blacksburg: Virginia Water Resources Research Center.

MINITAB. 1994. *Minitab Reference Manual. Release 10 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. Blacksburg: Dept. of Agricultural Engineering. Virginia Polytechnic Institute and State Univ.

Nofziger, D. L., and A. G. Hornsby. 1986. A microcomputer-based management tool for chemical movement in soil. *Applied Agricultural Research* 1:50-56.

Novotny, V., and H. Olem. 1994. *Water quality. Prevention, identification, and management of diffuse pollution.* New York:Van Nostrand Reinhold Co.

Onstad, C. A. and G. R. Foster. 1975. Erosion modeling on a watershed. *Transactions of the ASAE* 18(2):288-292.

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.

Patni, N. K., R. Frank, and S. Clegg. 1987. Pesticide persistence and movement under farm conditions. ASAE Paper No. 87-2627. St. Joseph, Mich.: ASAE.

Rawls, W. J., and D. L. Brakensiek. 1985. Prediction of soil water properties for hydrologic modeling. In *Proc. Symp. on Watershed Management*, 293-299. New York: ASCE.

Rao, P. S. C., and A. G. Hornsby. 1989. Pesticides and their behavior in soil and water. Soil Science Fact Sheet SL40 (Revised). Gainesville: Florida Co-operative Extension Service, Institute of Food and Agricultural Sciences, Univ. of Florida.

Scavia, D., W. F. Powers, R. P. Canale, and J. L. Moody. 1981. Comparison of First-Order Error Analysis and Monte Carlo Simulation in time-dependent lake eutrophication models. *Water Resources Research.* 17(4):1051-1059.

SCS. 1982. *Soil Survey of Westmoreland County, Virginia*. Richmond: USDA-SCS.

Smith, V. J., and R. J. Charbeneau. 1990. Probabilistic soil contamination exposure assessment procedures. *J. of Environmental Eng.* 116(6):1143-1163.

Thomann, R. V. 1982. Verification of water quality models. *J. of Environmental Eng. Division, Proceedings of the American Society of Civil Engineers* 108:923-939.

Thomas, A. W., W. M. Snyder, and G. W. Landale. 1988. Stochastic impacts on farming: V. Risk adjustment through conservation planning. *Transactions of the ASAE* 31(5):1368-1374.

Triplett, G. B., Jr., B. J. Conner, and W. M. Edwards. 1978. Transport of atrazine and simazine in runoff from conventional and no-tillage corn. *J. of Environmental Quality* 7:77-84.

Wagenet, R. J. and J. L. Hutson. 1986. Predicting the fate of nonvolatile pesticides in the unsaturated zone. *J. of Environmental Quality* 15(4):315-322.

Wagenet, R. J. and P. S. C. Rao. 1990. Modeling pesticide fate in soils. In: *Pesticides in the soil environment: Process, Impacts, and Modeling* ed. H. H. Cheng, 351-399. Madison, Wis.: Soil Science Society of America.

Walker, A. and R. L. Zimdahl. 1981. Simulation of the persistence of atrazine linuron and metolachlor in soil at different sites in the USA. *Weed Research* 21:255-265.

Warwick, J. J. and W. G. Cale. 1986. Effects of parameter uncertainty in stream modeling. *J. of Environmental Eng.* 112(3):479-489.

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. *Reviews of Environmental Contamination and Toxicology* 123:1-164.

Williams, J.R. and W. V. LaSeur. 1976. Water yield modeling using SCS curve numbers. *J. of Hydraulics Division., ASCE* 102:1241-1253.

Wilmott, C. J., S. G. Ackleson, R. E. Davis, J. J. Feddema, K. M. Klink, D. R. Legates, J. O'Connell, and C. M. Rowe. 1985. Statistics for the evaluation and comparison of models. *J. of Geophysical Research* 90(C5):8995-9005.

Wischmeier, W. H., and D. D. Smith. 1978. *Predicting Rainfall Erosion Losses - A Guide to Conservation Planning.* Agricultural Handbook Number 537. Washington, D.C.: USDA.

Woolhiser, D. A., and D. L. Brakensiek. 1982. Hydrologic modeling of small watersheds. In *Hydrologic modeling of small watersheds*, eds. C. T. Haan, H. P. Johnson, and D. L. Brakensiek, 3-16. St. Joseph, Mich.: ASAE.

Zacharias, S. 1992. Tillage Effects on Leaching and Persistence of Pesticides in a Coastal Plain Soil. M.S. thesis. Dept. of Agricultural Engineering. Virginia Polytechnic Institute and State Univ., Blacksburg.

Zacharias, S. 1995. Personal communication. Graduate Research Assistant, Dept. of Biological Systems Engineering, Virginia Polytechnic Institute and State Univ., Blacksburg.

Zacharias, S., and C. D. Heatwole. 1993. Predicting tillage treatment effects on pesticide transport: a validation study. ASAE Paper No. 93-2592. St. Joseph, Mich.: ASAE.

Zacharias, S., and C. D. Heatwole. 1994. Evaluation of GLEAMS and PRZM for predicting pesticide leaching under field conditions. *Transactions of the ASAE* 37(2):439-451.

Zacharias, S., C. D. Heatwole, and C. W. Coakley. 1993. Comparison of quantitative techniques used for pesticide model validation. ASAE Paper No. 93-2506. St. Joseph, Mich.: ASAE.

Zacharias, S., C. D. Heatwole, and C. W. Coakley. 1996. Robust quantitative techniques for validating pesticide transport model. *Transactions of the ASAE* (in press).

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 Hydrol.* 12:35-50.

APPENDICES

APPENDIX A

Table A-1. Surface runoff and transported sediment and chemicals from the no-till and conventional tillage plots. (Heatwole et al., 1992)

				ATRAZINE		METOLACHLOR	
Julian Day	**Plot**	**Runoff (mm)**	**Sediment (kg/ha)**	**Conc (ppb)**	**Mass (g/ha)**	**Conc (ppb)**	**Mass (g/ha)**
117	QNA	0.17		934.10	1.623	614.65	1.068
	QNB	0.37		71.07	0.262	59.74	0.220
119	QNA	0.03	*	66.96	0.018	53.46	0.015
	QNB	0.19	2.53	61.26	0.114	57.45	0.107
124	QNA	1.07	6.15	197.78	2.112	138.41	1.478
	QNB	6.38	22.96	99.44	6.341	69.37	4.424
130	QNA	2.18	1.74	61.25	1.336	57.99	1.264
	QNB	14.88	68.45	18.24	2.713	17.56	2.612
146	QNA	13.73	18.81	7.93	1.089	8.63	1.185
	QNB	47.09	72.99	5.67	2.670	6.15	2.894
149	QNA	23.97	29.96	2.83	0.678	2.38	0.570
	QNB	55.43	188.45	2.31	1.281	1.63	0.906
160	QNB	0.28	0.64	0.01	0.000	0.01	0.000
166	QNA	0.03	0.09	0.01	0.000	0.01	0.000
	QNB	2.71	8.03	0.01	0.000	0.01	0.000
182	QNB	0.36	2.12	2.84	0.010	3.28	0.012
193	QNB	3.38	5.64	0.00	0.000	2.32	0.078
202	QNA	0.36	0.05	1.92	0.007	1.52	0.005
	QNB	6.02	12.58	0.00	0.000	0.49	0.030
221	QNB	0.45	1.20	0.00	0.000	0.01	0.000
222	QNB	1.06	1.66	0.00	0.000	0.01	0.000
235	QNA	11.74	20.19	0.00	0.000	0.00	0.000
	QNB	27.21	384.19	0.00	0.000	0.00	0.001
236	QNA	0.00	0.00	0.00	0.000	0.23	0.000
	QNB	0.10	1.53	0.00	0.000	0.00	0.000
241	QNA	1.56	6.06	0.76	0.012	0.45	0.007
	QNB	6.76	199.85	0.97	0.066	0.21	0.014
Total	QNA	54.83	83.05		6.87		5.59
Total	QNB	172.67	972.83		13.46		11.30

*insufficient sample for all analyses

Table A-2. Observed atrazine mass in root zone (0-0.9m) from 20* cores on each date for the no-till plot (Zacharias, 1995).

Core Number	Day 118 (g/ha)	Day 128 (g/ha)	Day 145 (g/ha)	Day 209 (g/ha)	Day 272 (g/ha)
1	616.21	872.51	385.63	199.13	255.12
2	553.45	867.44	539.76	53.40	275.45
3	1284.32	535.90	287.34	46.52	222.41
4	735.88	560.93	164.51	544.86	156.55
5	1271.16	140.47	370.73	179.80	211.42
6	869.60	230.76	762.82	355.70	431.50
7	1418.99	1128.90	873.71	1153.51	146.62
8	468.90	1162.77	714.80	193.14	68.76
9	1026.31	1099.26	348.07	206.43	356.54
10	832.98	1535.08	1075.15	65.03	227.04
11	1305.71	1428.22	193.31	602.56	466.98
12	1171.47	741.49	570.24	279.74	575.07
13	1240.74	374.73	275.66	265.48	505.48
14	1025.93	549.59	481.96	252.38	355.81
15	460.90	398.06	126.75	131.49	454.11
16	528.02	1706.01	247.94	683.69	338.11
17	362.84	1119.50	263.74	373.19	303.20
18	1045.41	1480.32	524.61	198.70	473.54
19	731.93	970.04	284.99	399.87	90.57
20	521.69	617.25	274.81	104.85	115.42
Median	851.30	869.95	359.40	229.40	289.35
Average	873.62	875.96	438.33	314.47	301.48
Standard Deviation	337.03	453.13	254.20	266.27	149.03
Coefficient of Variation	39	52	58	85	49

* Number of actual cores varies from 16-20 depending on the data set.

Table A-3. Observed atrazine mass in the root zone (0-0.9m) from 20* cores on each date for the conventional tillage plot (Zacharias, 1995).

Core Number	Day 118 (g/ha)	Day 128 (g/ha)	Day 145 (g/ha)	Day 209 (g/ha)	Day 272 (g/ha)
1	556.99	628.53	368.64	593.65	451.03
2	499.08	691.01	275.41	319.29	184.11
3	680.87	1116.99	214.65	0.00	163.84
4	549.66	407.23	329.38	131.81	183.75
5	205.97	813.55	633.03	0.00	195.37
6	667.60	873.61	127.40	345.88	569.71
7	1254.98	582.75	458.70	106.81	291.91
8	258.41	374.15	438.64	120.96	376.01
9	437.76	1074.99	159.65	189.07	458.64
10	676.87	568.81	2045.77	653.64	652.17
11	1054.39	1195.05	551.19	96.60	440.75
12	304.48	369.11	2210.16	306.44	429.28
13	917.58	637.48	191.50	553.01	1883.09
14	1397.46	1366.91	743.45	412.72	561.73
15	655.39	410.83	686.32	1060.69	202.81
16	722.88	952.63	699.47	195.24	439.35
17		704.87	359.80	252.59	675.30
18		453.66		178.16	211.94
19		977.89		695.47	
20		311.48		251.16	
Median	661.50	664.25	438.60	251.90	434.35
Average	677.52	725.58	617.24	323.16	465.04
Standard. Deviation	337.78	308.82	601.44	269.17	391.48
Coefficient of Variation	50	43	97	83	84

* Number of actual cores varies from 16-20 depending on the data set.

Table A-4. Observed metolachlor mass in the root zone (0-0.9m) from 20* cores on each date for the no-till plot (Zacharias, 1995).

Core Number	Day 118 (g/ha)	Day 128 (g/ha)	Day 145 (g/ha)	Day 209 (g/ha)	Day 272 (g/ha)
1	575.53	620.66	607.44	100.23	285.66
2	708.47	659.26	162.57	119.48	55.82
3	721.34	377.15	236.17	104.50	67.10
4	753.26	1081.22	251.19	41.49	73.84
5	475.86	81.74	182.01	141.34	96.99
6	377.41	133.23	723.00	30.17	51.78
7	758.62	770.36	329.47	174.55	63.39
8	349.06	583.22	475.90	42.59	74.99
9	553.02	648.21	222.17	12.28	70.92
10	378.30	1384.55	435.90	40.27	95.69
11	1045.48	948.64	219.66	335.96	12.28
12	702.44	446.67	284.49	76.93	51.27
13	716.54	228.73	574.09	40.84	39.87
14	420.98	251.83	504.10	0.00	81.67
15	435.35	186.09	188.77	2.91	64.60
16	189.34	780.38	169.73	19.95	0.00
17	395.00	513.33	208.28	80.70	57.75
18	718.51	445.98	346.18	43.45	42.70
19	676.78	378.24	241.61	15.05	30.47
20	497.52	197.08	91.62	19.95	35.17
Median	564.25	480.00	246.40	42.05	60.55
Average	572.44	535.83	322.72	72.13	67.60
Standard. Deviation	200.35	339.29	173.04	78.72	57.00
Coefficient of Variation	35	63	54	109	84

* Number of actual cores varies from 16-20 depending on the data set.

Table A-5. Observed metolachlor mass in the root zone (0-0.9m) from 20* cores on each date for the conventional tillage plot (Zacharias, 1995).

Core Number	Day 118 (g/ha)	Day 128 (g/ha)	Day 145 (g/ha)	Day 209 (g/ha)	Day 272 (g/ha)
1	565.49	1177.97	721.42	23.66	28.42
2	493.12	737.42	461.83	62.87	29.25
3	711.44	923.25	244.02	162.33	17.55
4	561.73	452.45	401.36	47.99	28.82
5	298.40	992.23	405.29	75.95	23.40
6	585.84	711.99	281.46	271.11	42.03
7	847.53	475.21	209.26	20.02	25.42
8	304.37	434.99	401.66	37.23	74.90
9	458.59	684.07	140.96	23.56	74.69
10	732.38	455.02	441.36	110.96	31.27
11	683.51	1189.47	408.30	142.90	25.35
12	737.22	302.75	365.66	90.88	14.73
13	1283.81	433.88	438.30	52.79	23.40
14	545.65	391.95	379.94	55.41	61.97
15	958.00	466.07	629.52	90.70	55.33
16	1465.34	951.11	370.95	41.34	60.45
17	808.64	473.58	153.49	25.02	23.40
18	812.65	334.36	272.09	502.17	39.84
19		938.58	361.83	38.20	31.20
20		321.62	165.74	10.92	17.55
Median	697.45	474.40	375.45	54.10	29.05
Average	714.09	642.40	362.72	94.30	36.45
Standard. Deviation	300.16	290.84	147.44	114.27	18.83
Coefficient of Variation	42	45	41	121	52

* Number of actual cores varies from 16-20 depending on the data set.

Table A-6. Observed bromide mass in the root zone (0-0.9m) from 20 cores on each date for the no-till plot (Zacharias, 1995).

Core Number	Day 118 (g/ha)	Day 128 (g/ha)	Day 145 (g/ha)	Day 209 (g/ha)	Day 272 (g/ha)
1	20809.35	29923.61	27176.30	22289.64	17825.21
2	12108.03	19317.58	46877.41	23927.63	25246.71
3	16607.87	28979.38	34506.41	19267.37	16572.17
4	35378.37	15919.90	24202.96	17532.78	31930.15
5	10178.88	23151.46	6713.81	18161.80	14344.48
6	24617.27	27541.93	17950.03	25848.45	26932.80
7	18201.34	51646.27	20125.05	23529.62	16045.10
8	28409.64	33785.19	22212.96	18670.69	17246.04
9	14875.36	35564.50	8334.95	15468.70	34978.50
10	5872.82	34428.96	38175.78	17397.21	18151.84
11	71262.75	22582.51	25113.12	568.30	11981.64
12	44044.09	33847.46	30793.55	14839.53	35020.68
13	24339.19	36093.34	30412.00	16524.98	25316.74
14	36896.15	4007.93	27743.04	16338.04	16228.49
15	22118.25	20270.91	15944.72	19705.32	12243.39
16	2149.14	56105.89	38855.09	22484.35	3659.36
17	26901.43	27842.36	10045.96	21796.40	494.47
18	19253.14	17304.20	22673.21	2792.57	12946.17
19	42843.69	19922.32	9224.76	21636.79	5388.97
20	45541.93	24140.90	25572.88	13453.12	12761.27
Median	23228.75	27692.15	24658.05	18416.25	16400.35
Average	26120.43	28118.83	24132.70	17611.67	17765.71
Standard. Deviation	16279.08	11897.56	10879.34	6379.06	9660.64
Coefficient of Variation	62	42	45	36	54

* Number of actual cores varies from 16-20 depending on the data set.

Table A-7. Observed bromide mass in the root zone (0-0.9m) from 20* cores on each date for the conventional tillage plot (Zacharias, 1995).

Core Number	Day 118 (g/ha)	Day 128 (g/ha)	Day 145 (g/ha)	Day 209 (g/ha)	Day 272 (g/ha)
1	58583.30	32617.73	25037.20	19406.61	21673.51
2	26132.66	17230.66	17479.28	16249.54	26333.10
3	29088.51	27570.74	13713.07	5051.10	29080.11
4	4133.50	26833.14	28894.55	18933.31	35628.15
5	17170.12	35906.22	46983.28	11031.99	25357.49
6	46997.48	36286.78	29396.73	20627.62	37728.18
7	17845.98	34712.09	26568.24	22288.72	24610.53
8	21261.84	16883.91	39079.51	35747.65	22049.86
9	12190.15	35145.52	11247.81	26500.02	9265.45
10	22748.30	15127.36	10650.72	19256.07	21228.17
11	29348.07	15882.02	47743.78	10590.11	32321.30
12	33123.17	7103.06	30966.60	22418.86	18239.07
13	38132.70	24210.36	36194.95	19503.50	16410.50
14	25770.01	15408.84	32779.85	53035.76	33919.38
15	30222.08	17094.11	54541.89	1802.19	28498.29
16	53653.68	22747.40	6841.09	7402.84	15563.93
17	14225.83	2619.72	27615.79	60343.52	13145.84
18	35411.12	1302.39	47578.57	38814.17	20400.23
19		19593.09	33976.48	44143.41	29764.02
20		23552.11	27421.17	32123.06	13388.69
Median	27610.60	21170.25	29145.65	20065.55	23330.20
Average	28668.80	21391.36	29735.53	24263.50	23730.29
Standard. Deviation	14244.11	10593.60	13363.75	15626.60	8000.39
Coefficient of Variation	50	50	45	64	34

* Number of actual cores varies from 16-20 depending on the data set.

Table A-8. Observed atrazine depth of center of mass from 20* cores on each date for the no-till plot (Zacharias, 1995).

Core Number	Day 118 (g/ha)	Day 128 (g/ha)	Day 145 (g/ha)	Day 209 (g/ha)	Day 272 (g/ha)
1	8.45	25.69	5.65	26.72	35.05
2	7.98	23.13	16.07	19.69	35.15
3	21.05	19.95	17.94	6.92	27.67
4	6.64	22.04	5.75	62.77	42.26
5	28.43	13.33	21.67	11.51	54.32
6	15.06	26.26	25.11	45.42	47.76
7	21.18	10.67	14.62	35.25	39.28
8	10.63	28.06	16.28	37.50	37.50
9	17.37	20.12	14.79	22.72	41.29
10	21.96	38.90	16.41	8.00	60.08
11	11.06	16.97	5.21	43.88	20.08
12	40.60	18.96	28.74	36.09	42.50
13	16.82	22.99	13.22	46.57	40.87
14	15.52	36.51	15.85	58.29	25.68
15	5.96	20.46	3.59	28.15	35.08
16	10.61	32.43	5.19	18.41	28.52
17	7.43	19.61	21.51	49.27	29.32
18	14.24	27.28	12.54	22.27	37.91
19	11.97	32.55	10.14	42.38	13.21
20	15.65	16.45	24.81	23.48	30.48
Median	14.65	22.52	15.32	31.70	36.33
Average	15.43	23.62	14.75	32.26	36.20
Standard. Deviation	8.37	7.42	7.31	16.04	10.97
Coefficient of Variation	54	31	50	50	30

* Number of actual cores varies from 16-20 depending on the data set.

Table A-9. Observed atrazine depth of center of mass from 20* cores on each date for the conventional tillage plot (Zacharias, 1995).

Core Number	Day 118 (g/ha)	Day 128 (g/ha)	Day 145 (g/ha)	Day 209 (g/ha)	Day 272 (g/ha)
1	8.11	13.64	11.80	38.47	29.04
2	16.88	10.22	7.14	26.32	19.48
3	10.81	19.92	33.73	40.31	42.08
4	15.77	10.32	7.63	60.77	19.35
5	17.33	13.05	18.74	14.02	18.47
6	18.33	11.05	7.18	37.16	25.74
7	16.75	18.32	16.83	50.22	27.45
8	4.78	10.04	13.92	31.25	17.98
9	17.72	25.52	6.78	44.65	13.35
10	9.71	23.73	10.46	25.99	29.08
11	8.73	21.03	15.37	44.82	37.13
12	5.85	12.69	32.98	35.93	34.36
13	13.95	14.15	8.07	37.58	16.67
14	17.09	17.39	13.64	26.42	20.57
15	6.79	6.93	34.73	36.50	20.70
16	8.11	8.59	15.83	12.50	25.72
17		11.47	21.24	39.22	31.12
18		29.20		43.27	32.95
19		19.15			
20		6.34			
Median	12.38	13.35	13.92	37.37	25.73
Average	12.30	15.14	16.24	35.86	25.62
Standard. Deviation	4.86	6.38	9.42	11.89	7.89
Coefficient of Variation	40	42	58	33	31

* Number of actual cores varies from 16-20 depending on the data set.

Table A-10. Observed metolachlor depth of center of mass from 20* cores on each date for the no-till plot (Zacharias, 1995).

Core Number	Day 118 (g/ha)	Day 128 (g/ha)	Day 145 (g/ha)	Day 209 (g/ha)	Day 272 (g/ha)
1	15.76	21.92	14.77	9.92	21.51
2	23.44	25.86	20.27	31.07	17.19
3	6.79	11.04	21.03	31.89	39.30
4	17.29	29.56	27.54	26.07	24.16
5	11.39	19.55	16.14	39.08	37.99
6	11.20	7.15	26.81	10.08	33.93
7	20.82	13.63	15.33	35.19	18.31
8	16.45	16.19	13.23	25.21	35.69
9	16.55	15.23	15.89	18.24	22.25
10	17.52	37.88	15.29	13.24	16.21
11	11.87	18.43	18.95	25.46	37.50
12	40.57	16.49	27.26	35.47	22.16
13	13.07	12.58	41.60	9.62	14.82
14	14.86	15.78	25.22	0.50	41.59
15	15.26	8.03	11.72	0.50	12.14
16	10.06	28.91	3.70	6.06	33.29
17	20.28	13.34	17.80	61.79	8.00
18	14.30	16.28	15.41	53.44	14.69
19	22.38	27.46	21.69	59.84	14.77
20	23.58	8.27	8.12	13.06	
Median	16.11	16.24	16.97	25.34	22.16
Average	17.17	18.18	18.89	25.29	24.50
Standard. Deviation	7.15	8.17	8.21	18.42	10.65
Coefficient of Variation	42	45	43	73	43

* Number of actual cores varies from 16-20 depending on the data set.

Table A-11. Observed metolachlor depth of center of mass from 20* cores on each date for the conventional tillage plot (Zacharias, 1995).

Core Number	Day 118 (g/ha)	Day 128 (g/ha)	Day 145 (g/ha)	Day 209 (g/ha)	Day 272 (g/ha)
1	6.84	23.59	31.42	7.26	17.57
2	12.04	10.71	14.38	16.00	8.00
3	6.96	10.94	20.54	27.49	8.00
4	13.57	13.09	21.92	22.62	11.11
5	32.98	11.37	11.91	28.09	8.00
6	10.27	8.69	32.84	64.03	12.35
7	7.38	13.23	7.39	6.95	12.83
8	13.50	12.86	20.12	13.12	15.25
9	18.05	12.32	8.74	6.85	38.62
10	10.62	15.73	14.00	44.92	11.87
11	11.82	17.58	13.70	37.39	8.00
12	11.64	9.14	10.48	22.34	22.50
13	14.02	9.27	18.13	24.38	8.00
14	21.18	8.96	15.67	15.03	9.40
15	11.09	9.06	30.97	38.12	15.54
16	14.02	10.68	14.99	40.34	20.50
17	13.00	7.98	7.46	13.25	8.00
18	8.43	9.55	24.56	25.58	19.71
19		13.48	12.09	13.59	8.00
20		13.54	15.59	6.56	8.00
Median	11.93	11.16	15.29	22.48	11.49
Average	13.19	12.09	17.34	23.70	13.56
Standard. Deviation	6.14	3.69	7.73	15.18	7.58
Coefficient of Variation	47	31	45	64	56

* Number of actual cores varies from 16-20 depending on the data set.

Table A-12. Observed bromide depth of center of mass from 20* cores on each date for the no-till plot (Zacharias, 1995).

Core Number	Day 118 (g/ha)	Day 128 (g/ha)	Day 145 (g/ha)	Day 209 (g/ha)	Day 272 (g/ha)
1	16.62	33.31	29.46	40.24	21.51
2	8.84	28.37	29.70	49.79	17.19
3	9.75	13.54	42.86	47.12	39.30
4	25.90	33.35	23.50	42.65	24.16
5	6.59	13.90	13.35	43.44	37.99
6	10.91	26.11	31.89	42.65	33.93
7	17.17	26.10	30.59	44.19	18.31
8	23.66	21.93	36.60	43.61	35.69
9	15.90	27.76	34.79	44.49	22.25
10	10.89	22.17	43.82	44.72	16.21
11	25.07	26.74	30.08	5.94	37.50
12	44.73	39.04	38.12	44.75	22.16
13	24.18	31.10	18.55	42.91	14.82
14	14.35	21.48	40.09	39.75	41.59
15	18.70	13.09	13.13	42.11	12.14
16	3.00	41.14	21.79	37.69	33.29
17	21.92	33.95	36.30	41.14	8.00
18	17.10	28.18	26.43	37.22	14.69
19	32.80	25.13	35.68	43.32	14.77
20	28.46	29.23	33.38	46.09	
Median	17.14	27.25	31.24	43.12	42.76
Average	18.83	26.78	30.51	41.19	24.50
Standard. Deviation	9.89	7.71	8.85	8.81	10.65
Coefficient of Variation	53	29	29	21	43

* Number of actual cores varies from 16-20 depending on the data set.

Table A-13. Observed bromide depth of center of mass from 20* cores for each date for the conventional tillage plot (Zacharias, 1995).

Core Number	Day 118 (g/ha)	Day 128 (g/ha)	Day 145 (g/ha)	Day 209 (g/ha)	Day 272 (g/ha)
1	18.22	33.80	26.19	45.76	41.42
2	11.20	26.52	26.59	52.09	49.34
3	7.54	16.83	10.61	38.41	54.46
4	7.53	21.43	45.02	42.62	51.25
5	15.95	27.93	39.90	46.09	50.78
6	18.49	22.20	37.07	41.39	46.91
7	18.29	34.15	16.93	43.37	52.92
8	9.63	30.28	32.17	40.62	33.91
9	6.37	29.43	10.55	49.34	34.40
10	19.35	18.13	18.09	42.87	40.34
11	34.44	16.02	34.80	36.51	39.58
12	22.62	6.49	29.11	45.60	34.79
13	33.46	14.81	35.05	43.81	36.90
14	30.31	22.63	37.97	48.81	36.53
15	22.86	21.15	49.86	18.82	29.99
16	28.74	31.37	3.95	43.84	45.95
17	15.80	0.50	27.44	46.49	42.33
18	8.30	8.02	40.04	43.43	46.58
19		25.84	36.00	43.34	47.29
20		28.70	34.14	43.14	44.02
Median	18.26	22.42	33.16	43.40	36.33
Average	18.28	21.81	29.57	42.82	42.99
Standard. Deviation	9.06	9.30	12.17	6.69	7.07
Coefficient of Variation	50	43	41	16	16

* Number of actual cores varies from 16-20 depending on the data set.

APPENDIX B

Section B-1. Description of Inputs for the Probabilistic Procedure (Kumar, 1995a)

GLEAMS inputs that were considered random variables are shown in Table [1.] The field-scale variability in many inputs have been reported to be adequately modeled by a normal or lognormal distribution (Jury et al., 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 [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 both plots, with the exception of the porosity distribution for the first horizon which was assumed to have a higher mean for the conventional tillage plot based on a previous field-study carried out by Menelik et al. (1990) in the same area. Both normal and lognormal distributions were fitted to the organic matter and porosity data and the distribution with a lower D statistic selected for the Monte Carlo simulations. Beta distributions were used for the sand and clay contents as Loague (1992) reported them to not be adequately modeled by either normal or lognormal distributions.

Two hundred values of sand, clay, porosity, and organic matter were generated from the fitted distributions and used in regression equations developed

[by Kumar (1995a)] to obtain values of field capacity and immobile water content. Normal or lognormal distributions were then fitted to these values, and the distribution with the lower D statistic employed in the Monte Carlo simulations. Normal distributions for porosity, field capacity, and wilting point were truncated at 0.0 and 0.99 from physical considerations. Lognormal distributions for these inputs were also truncated at 0.99. Organic matter distributions were likewise truncated at 0.0 and 4.0, which represents the upper limit of organic matter content for Ultisols (Brady, 1990).

The antecedent moisture condition II (AMC II) curve number distributions were derived from the potential retention (S) distributions obtained from rainfall-runoff data for the plots using a conceptual procedure as described [by Kumar (1995a).] Lognormal parameters for soil erodibility, Manning's roughness factor, pesticide partitioning coefficients, and pesticide half-lives were estimated from best-estimate and ranges given in the literature (Wischmeier and Smith, 1978; Engman, 1986; Wauchope et al., 1992). Parameters were estimated assuming the natural log of the best estimate was equal to the mean and the difference of the natural logarithm of the maximum and minimum value was equal to four times the standard deviation in normal space.

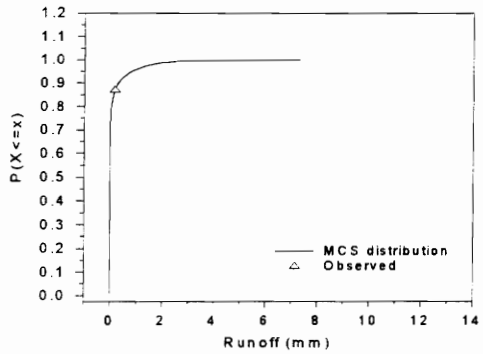
Symmetrical triangular distributions were used for soil loss ratios with a ±10% range around the best estimate (Thomas et al., 1988). The soil evaporation input (CONA), pesticide wash-off fraction, coefficient for plant pesticide uptake, and fraction of pesticide applied to crop residue were modeled with uniform distributions with ranges estimated from the GLEAMS manual. Application rates, which represent a source of extrinsic variability were modeled using beta distributions fitted to observed data.

The distributions and parameter values for inputs assumed to be the same for both conservation and conventional tillage simulations are summarized in Table [3.] Distributions and parameter values used to simulate differences between

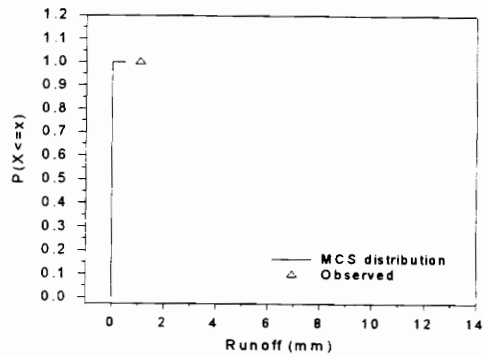
conservation and conventional tillage simulations are summarized in Tables [4 and 5.] Correlations among inputs considered in the simulations are summarized in Table [6.] The only observed correlation values available were the organic matter lag-1 serial correlations. Other correlations were assigned from subjective considerations and values reported in the literature. Correlation matrices were generated using an interactive procedure developed by Kumar (1995a) for assessing correlation structures with subjective information.

Meteorologic inputs (daily rainfall, daily maximum and minimum temperature, and monthly radiation) for the model were obtained from Heatwole et al. (1992), and data from a class A weather station located approximately 500m from the study site (Mostaghimi et al., 1989).

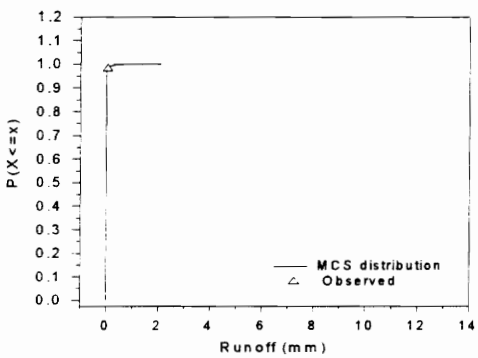
APPENDIX C



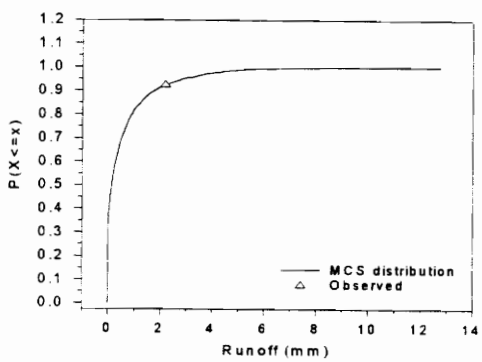
(a) Day 117



(d) Day 124

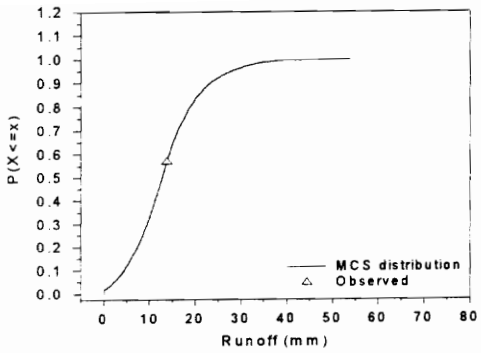


(b) Day 119

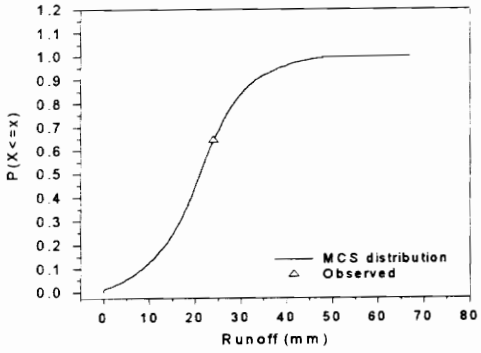


(d) Day 130

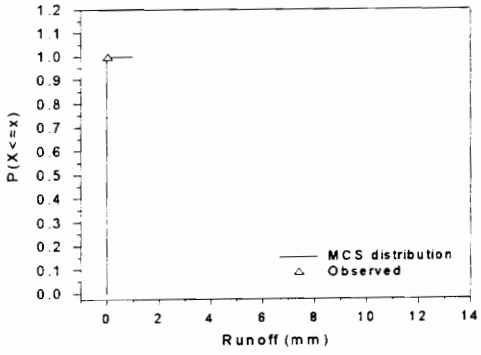
Figure C-1. MCS predicted distribution and observed value for runoff from the no-till plot on (a) day 117, (b) day 119, (c) day 124, and (d) day 130.



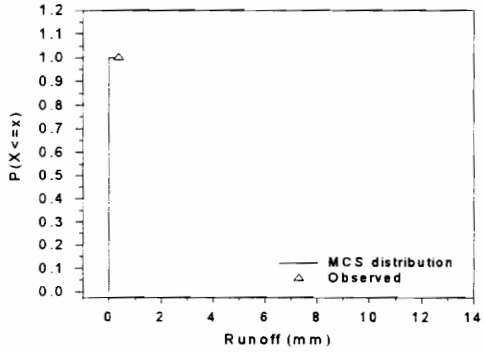
(a) Day 146



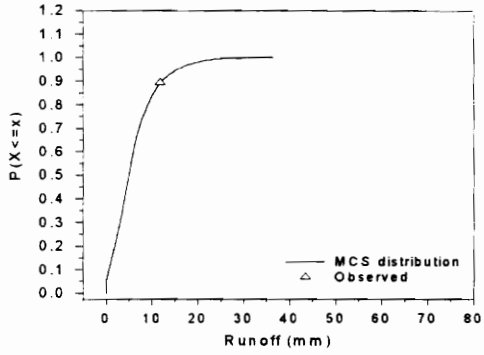
(b) Day 149



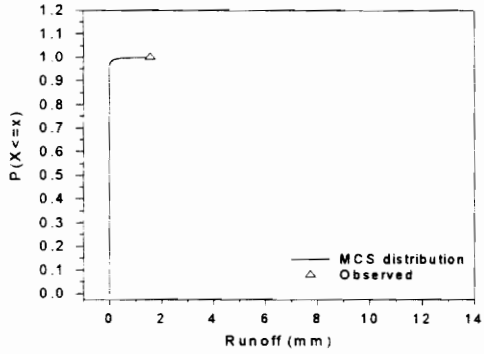
(c) Day 166



(d) Day 202

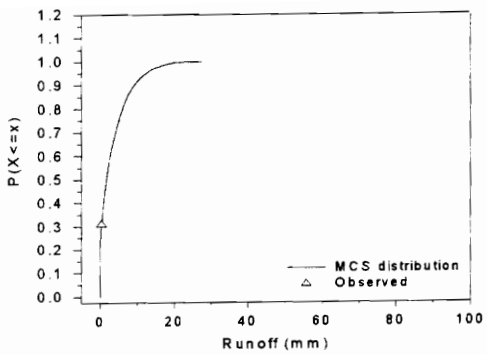


(e) Day 235

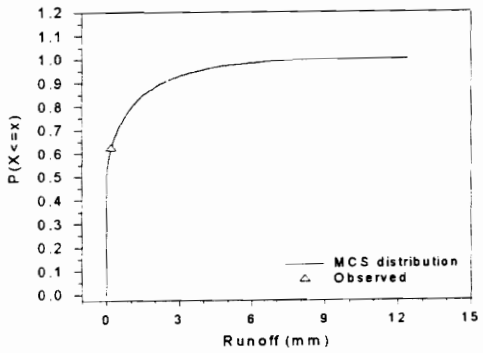


(f) Day 241

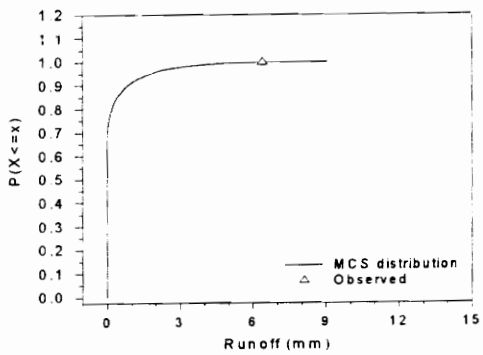
Figure C-2. MCS predicted distribution and observed value for runoff from the no-till plot on (a) day 146, (b) day 149, (c) day 166, (d) day 202, (e) day 235, and (f) day 241.



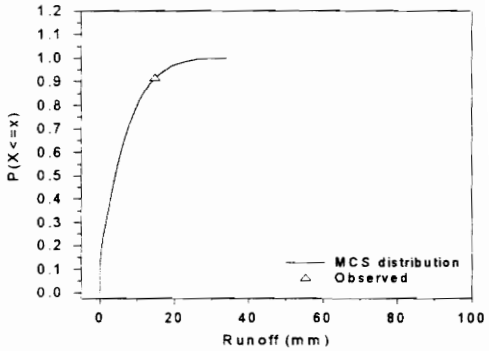
(a) Day 117



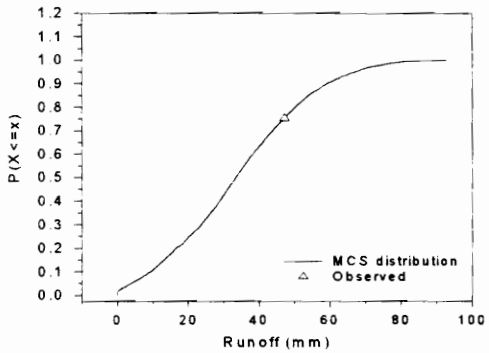
(b) Day 119



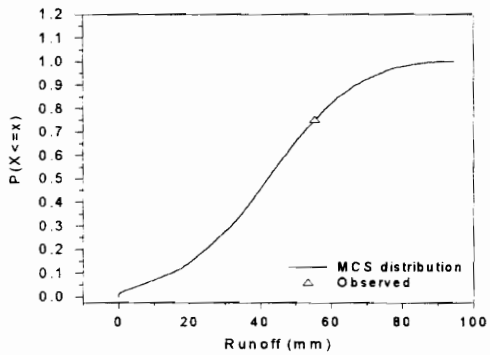
(c) Day 124



(d) Day 130

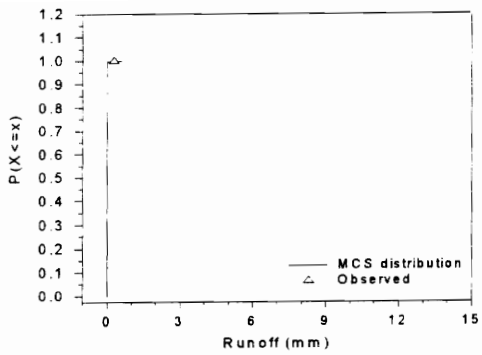


(e) Day 146

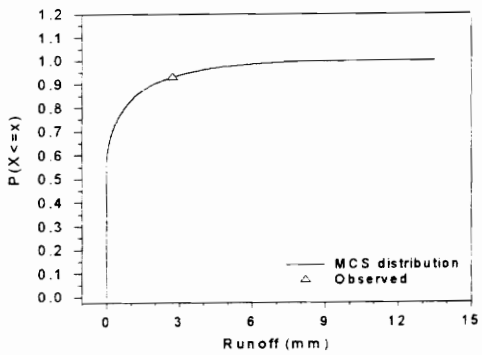


(f) Day 149

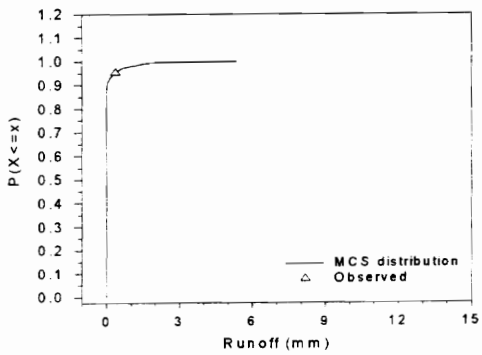
Figure C-3. MCS predicted distribution and observed value for runoff from the conventional tillage plot on (a) day 117, (b) day 119, (c) day 124, (d) day 130, (e) day 146, and (f) day 149.



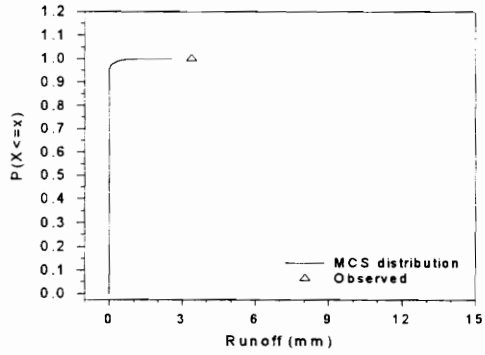
(a) Day 160



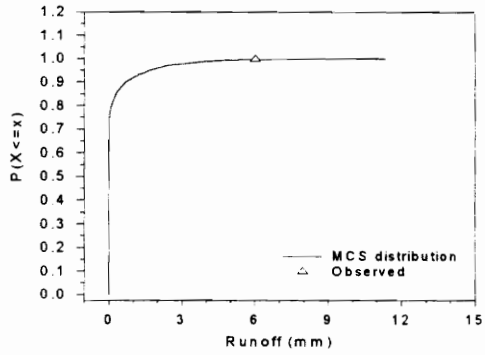
(b) Day 166



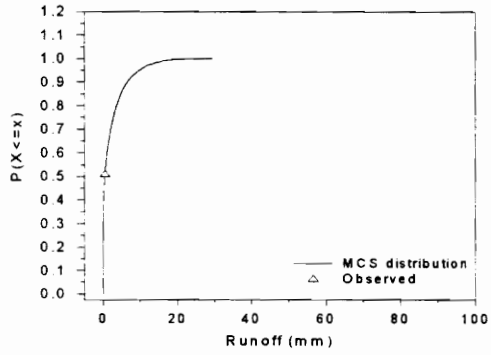
(c) Day 182



(d) Day 193

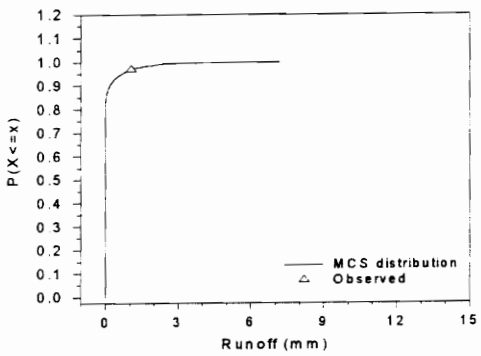


(e) Day 202

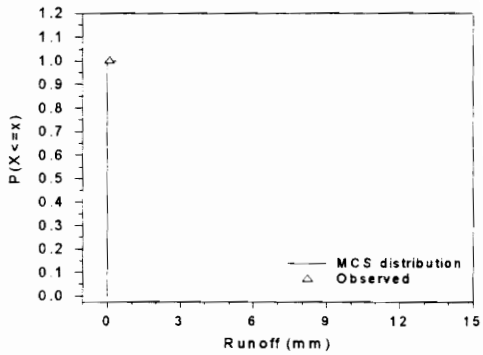


(f) Day 221

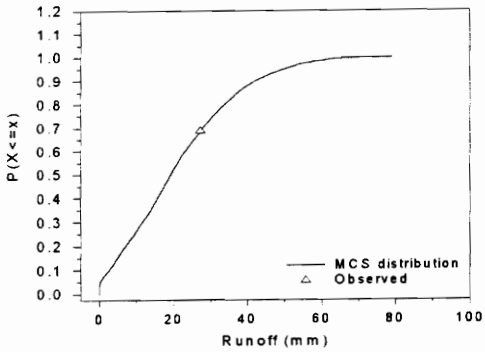
Figure C-4. MCS predicted distribution and observed value for runoff from the conventional tillage plot on (a) day 160, (b) day 166, (c) day 182, (d) day 193, (e) day 202, and (f) day 221.



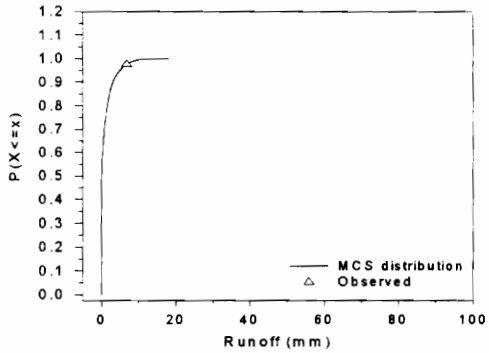
(a) Day 222



(c) Day 236

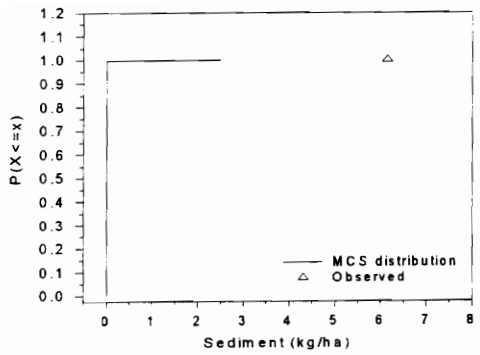


(b) Day 235

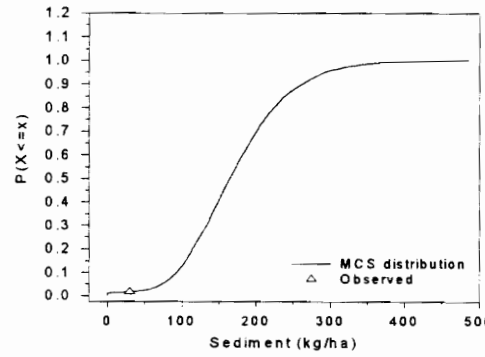


(d) Day 241

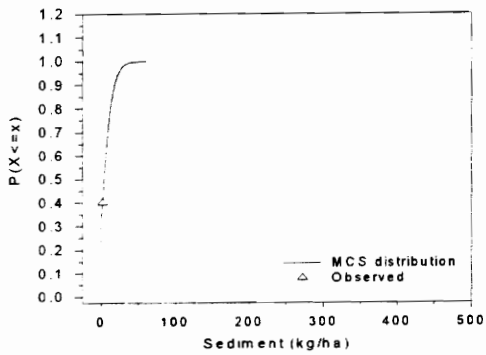
Figure C-5. MCS predicted distribution and observed value for runoff from the conventional tillage plot on (a) day 222, (b) day 235, (c) day 236, and (d) day 241.



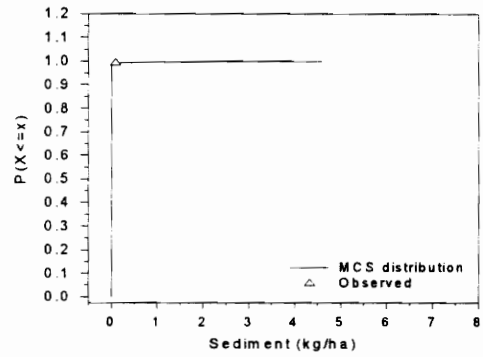
(a) Day 124



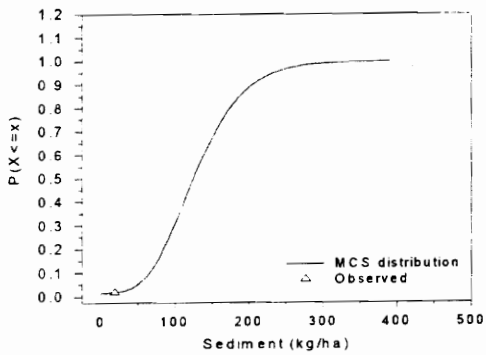
(d) Day 149



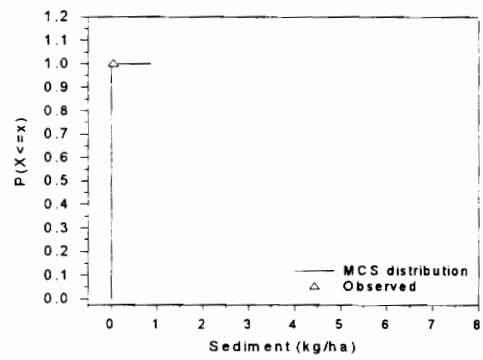
(b) Day 130



(e) Day 166

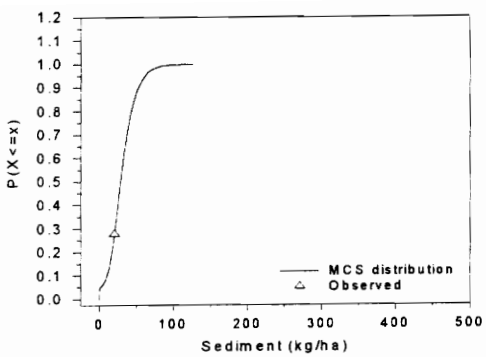


(c) Day 146

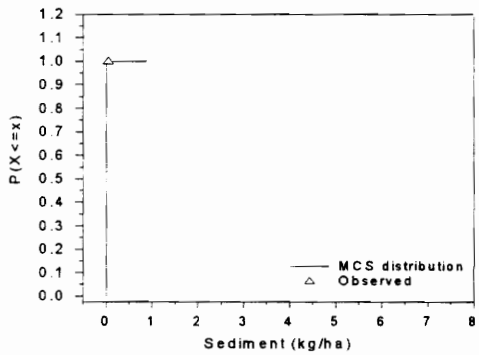


(f) Day 202

Figure C-6. MCS predicted distribution and observed value for sediment yield from the no-till plot for (a) day 124, (b) day 130, (c) day 146, (d) day 149, (e) day 166, and (f) day 202.

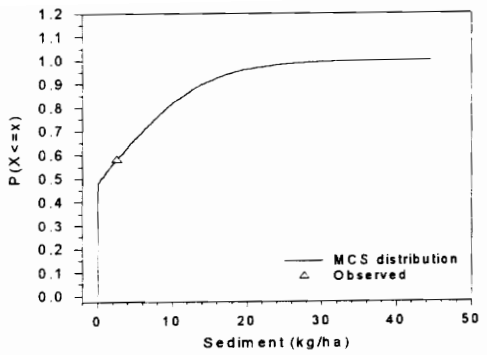


(a) Day 235

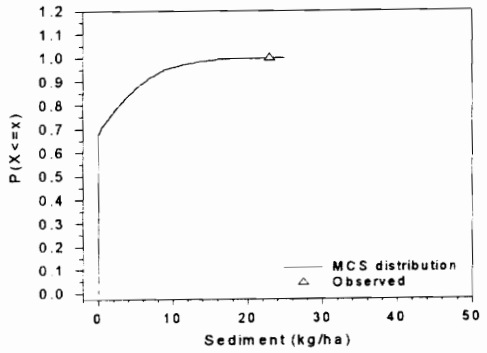


(b) Day 241

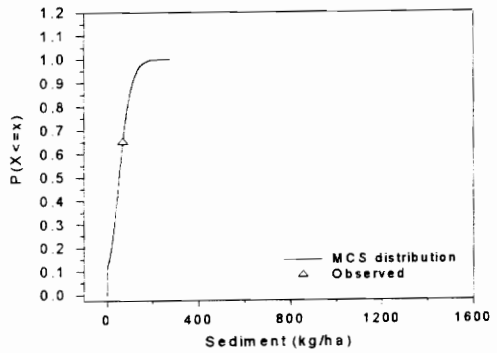
Figure C-7. MCS predicted distribution and observed value for sediment yield from the no-till plot for (a) day 235 and (b) day 241.



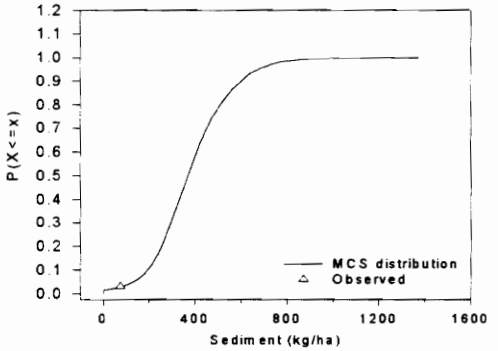
(a) Day 119



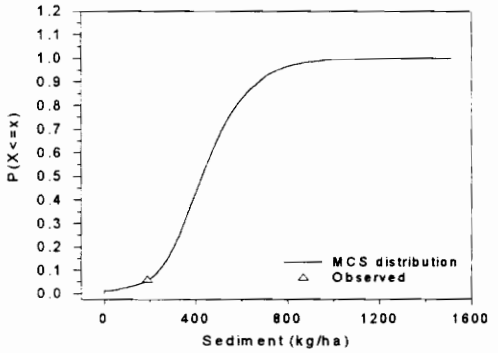
(b) Day 124



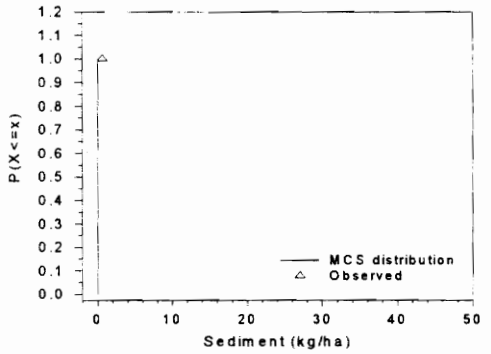
(c) Day 130



(d) Day 146

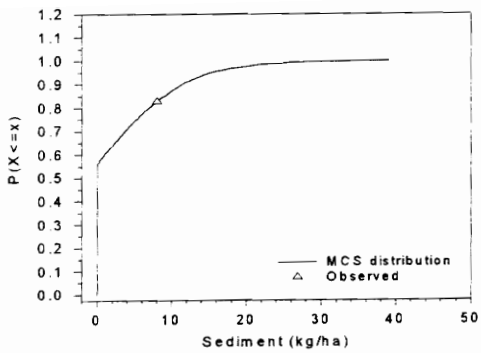


(e) Day 149

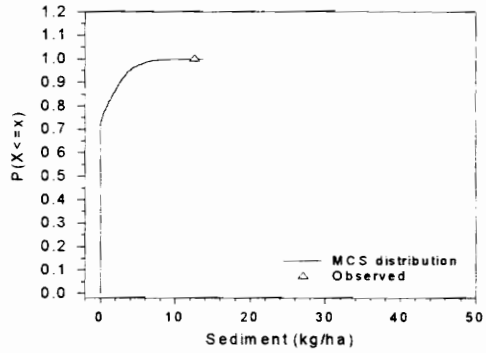


(f) Day 160

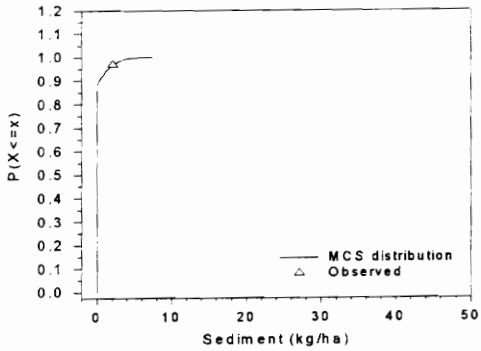
Figure C-8. MCS predicted distribution and observed value for sediment yield from the conventional tillage plot for (a) day 119, (b) day 124, (c) day 130, (d) day 146, (e) day 149, and (f) day 160.



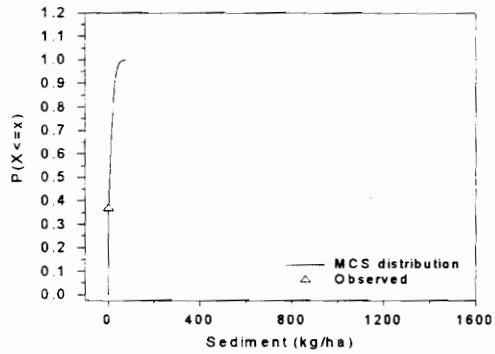
(a) Day 166



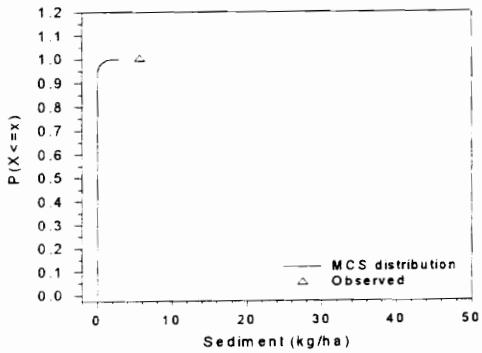
(d) Day 202



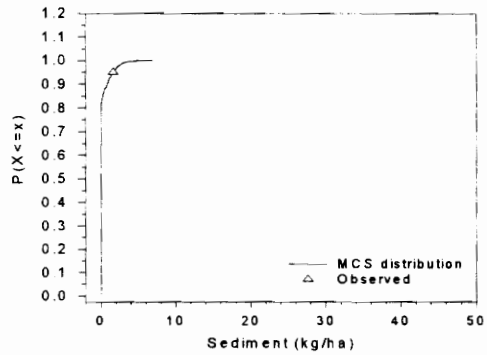
(b) Day 182



(e) Day 221

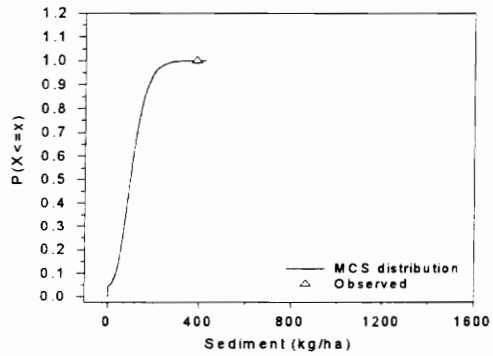


(c) Day 193

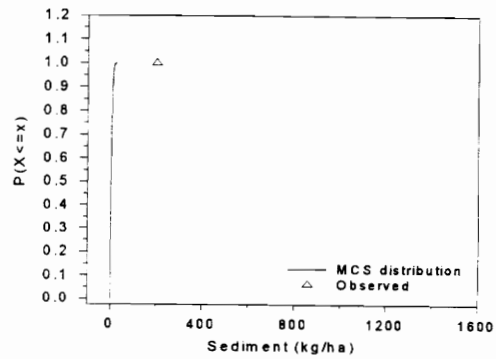


(f) Day 222

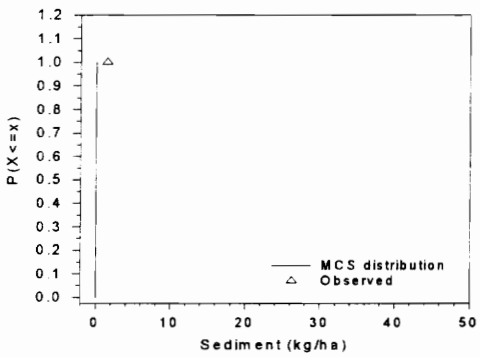
Figure C-9. MCS predicted distribution and observed value for sediment yield from the conventional tillage plot for (a) day 166, (b) day 182, (c) day 193, (d) day 202, (e) day 221, and (f) day 222.



(a) Day 235

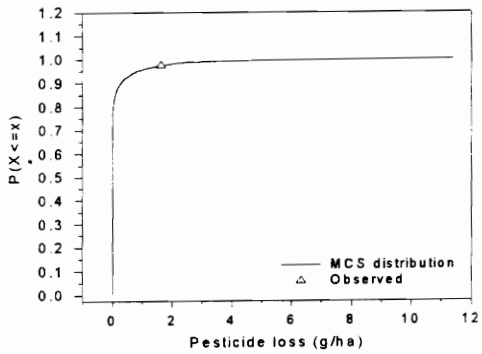


(c) Day 241

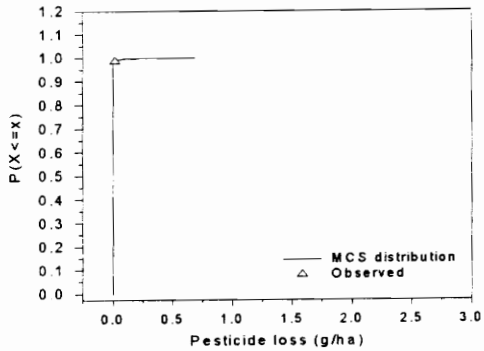


(b) Day 236

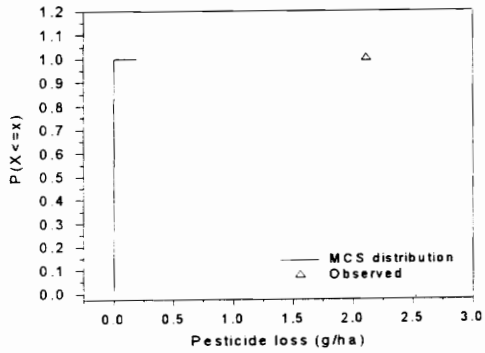
Figure C-10. MCS predicted distribution and observed value for sediment yield from the conventional tillage plot for (a) day 235, (b) day 236, and (c) day 241.



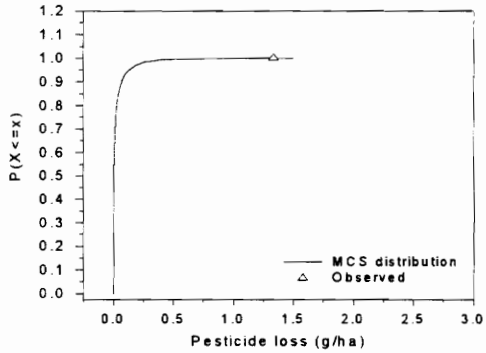
(a) Day 117



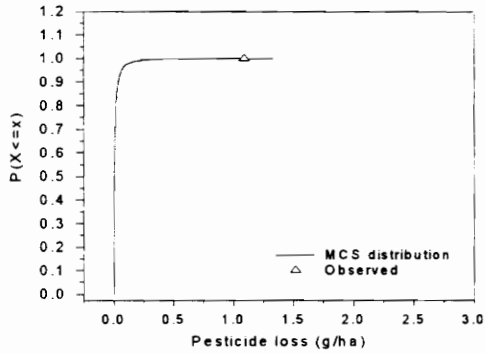
(b) Day 119



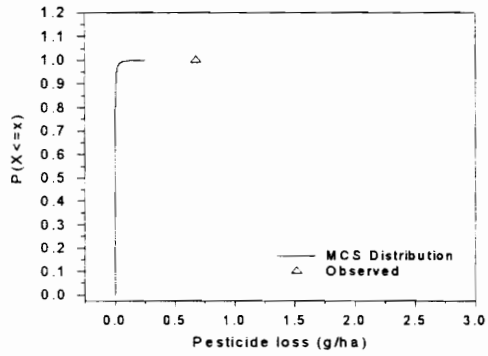
(c) Day 124



(d) Day 130

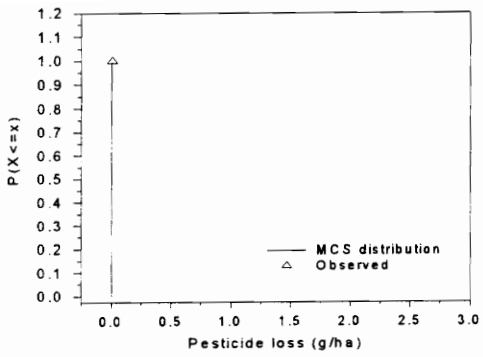


(e) Day 146

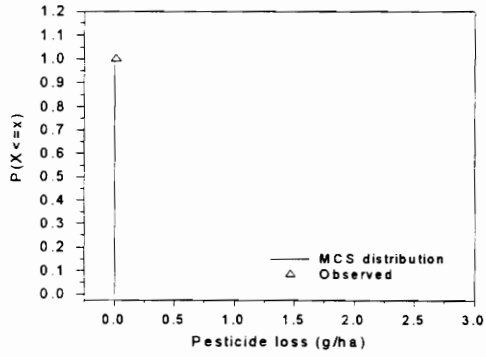


(f) Day 149

Figure C-11. MCS predicted distribution and observed value for atrazine surface losses from the no-till plot for (a) day 117, (b) day 119, (c) day 124, (d) day 130, (e) day 146, and (f) day 149.

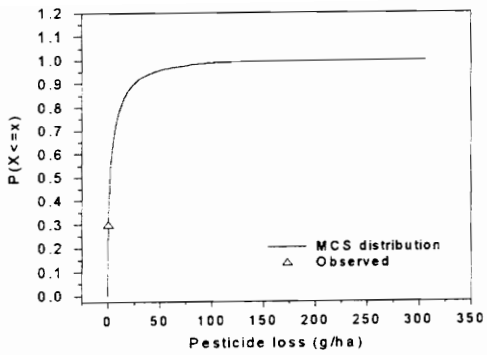


(a) Day 202

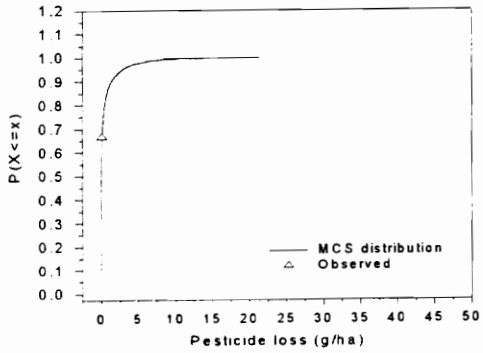


(b) Day 241

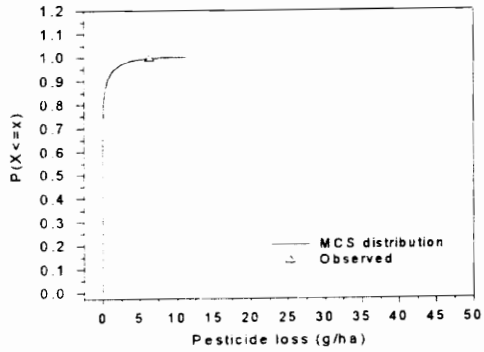
Figure C-12. MCS predicted distribution and observed value for atrazine surface losses from the no-till plot for (a) day 202 and (b) day 241.



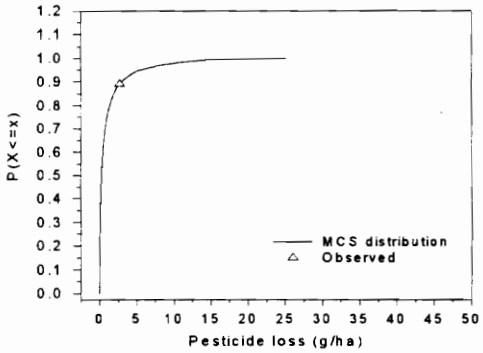
(a) Day 117



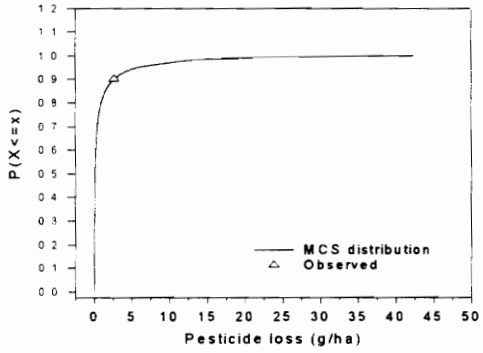
(b) Day 119



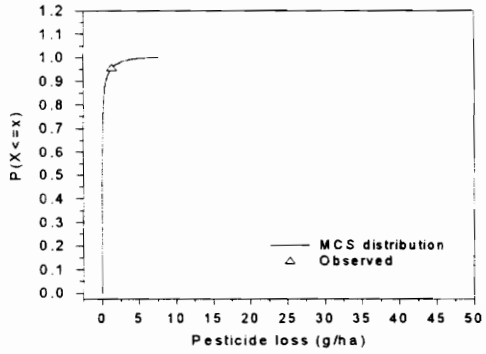
(c) Day 124



(d) Day 130

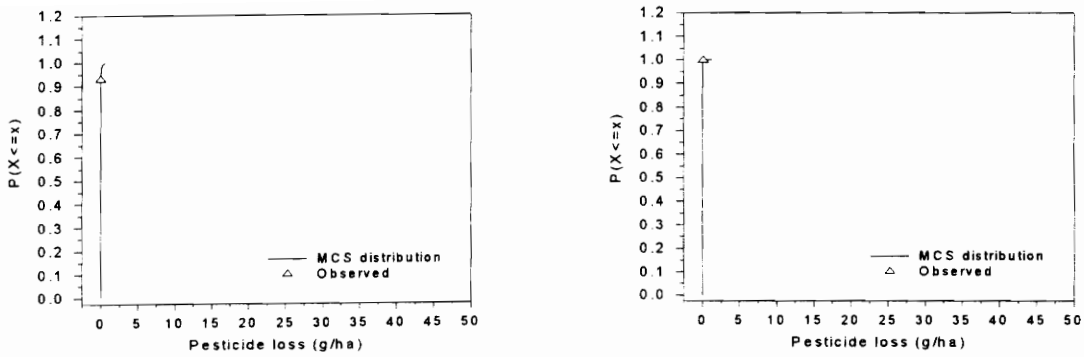


(e) Day 146



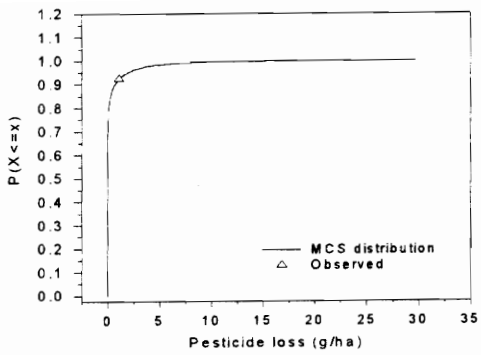
(f) Day 149

Figure C-13. MCS predicted distribution and observed value for atrazine surface losses from the conventional tillage plot for (a) day 117, (b) day 119, (c) day 124, (d) day 130, (e) day 146, and (f) day 149.

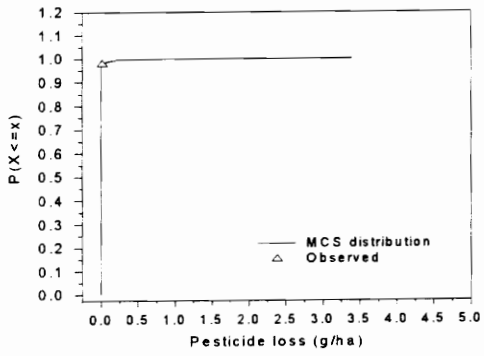


(a) Day 182 **(b) Day 241**

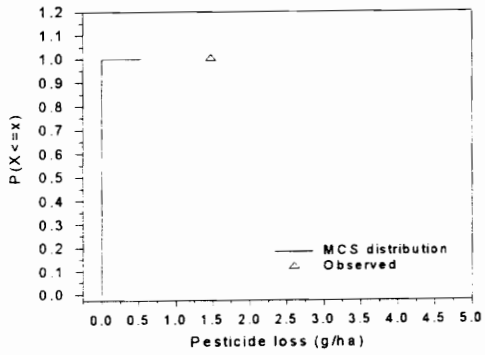
Figure C-14. MCS predicted distribution and observed value for atrazine surface losses from the conventional tillage plot for (a) day 182 and (b) day 241.



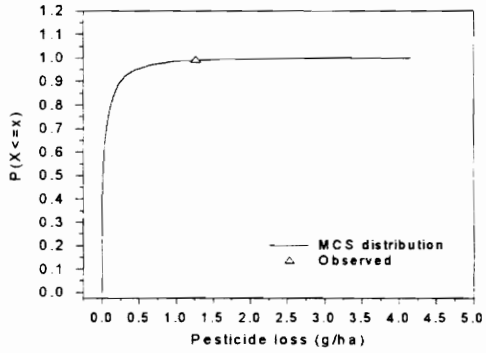
(a) Day 117



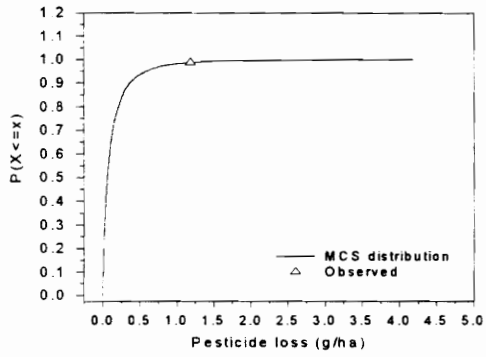
(b) Day 119



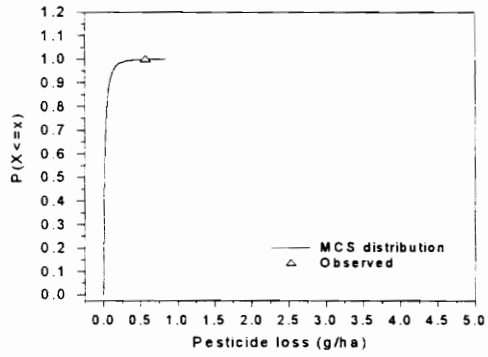
(c) Day 124



(d) Day 130

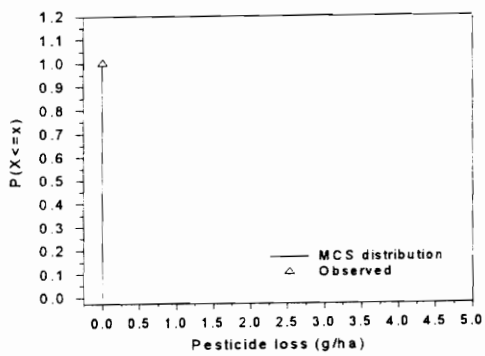


(e) Day 146

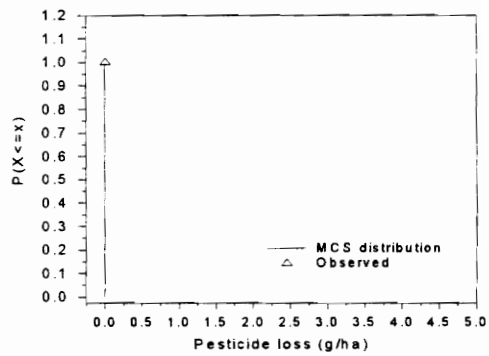


(f) Day 149

Figure C-15. MCS predicted distribution and observed value for metolachlor surface losses from the no-till plot for (a) day 117, (b) day 119, (c) day 124, (d) day 130, (e) day 146, and (f) day 149.

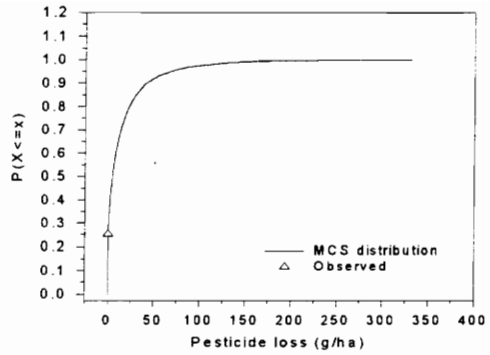


(a) Day 202

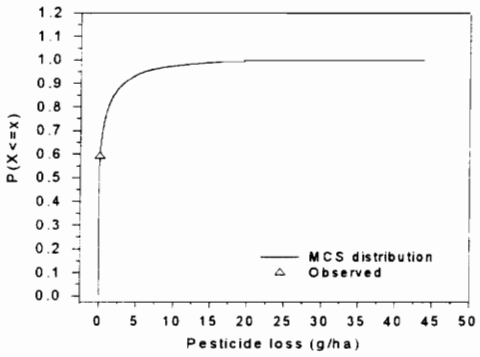


(b) Day 241

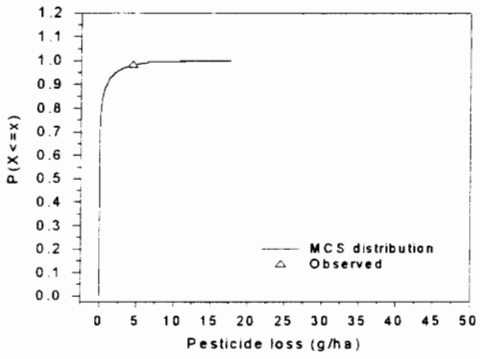
Figure C-16. MCS predicted distribution and observed value for metolachlor surface losses from the no-till plot for (a) day 202 and (b) day 241.



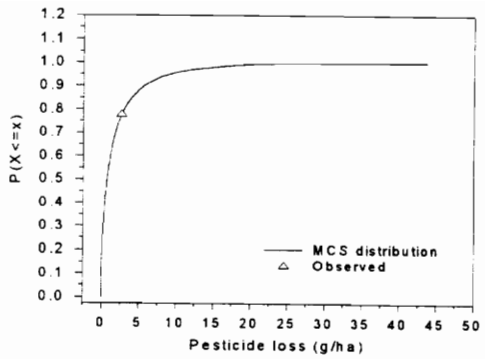
(a) Day 117



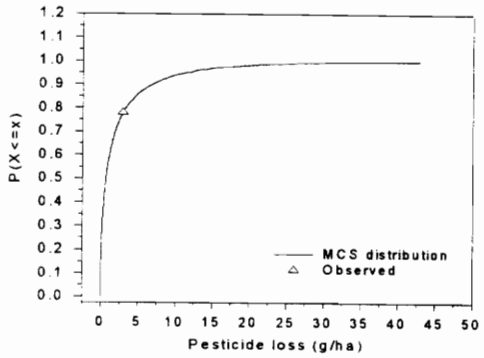
(b) Day 119



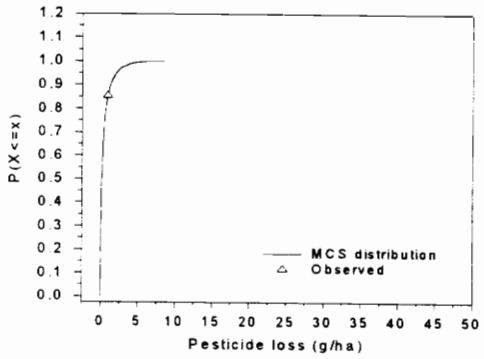
(c) Day 124



(d) Day 130

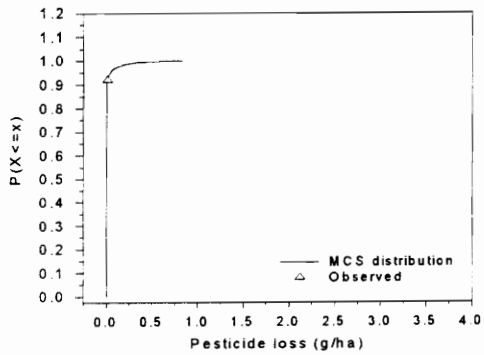


(e) Day 146

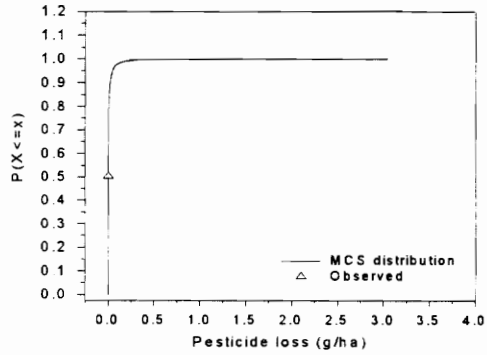


(f) Day 149

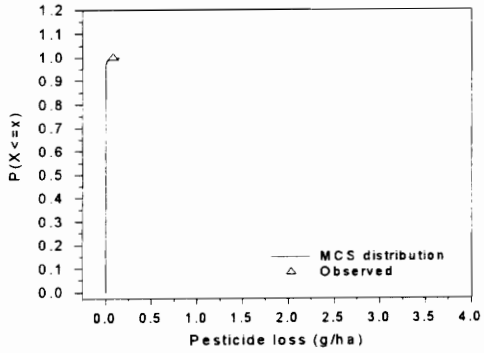
Figure C-17. MCS predicted distribution and observed value for metolachlor surface losses from the conventional tillage plot for (a) day 117, (b) day 119, (c) day 124, (d) day 130, (e) day 146, and (f) day 149.



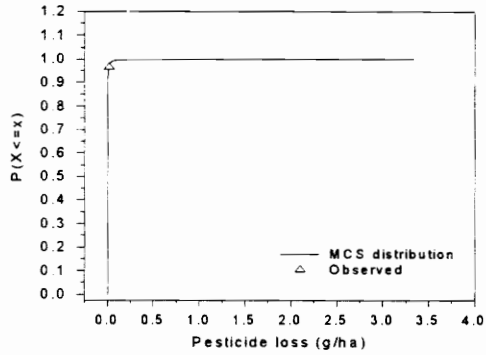
(a) Day 182



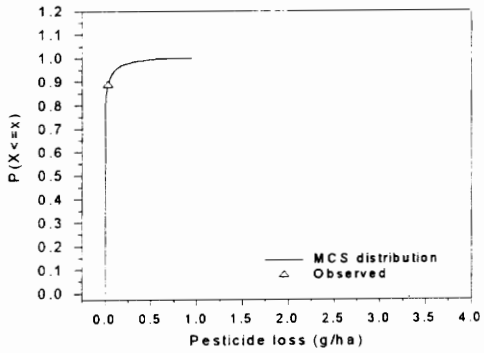
(d) Day 235



(b) Day 193



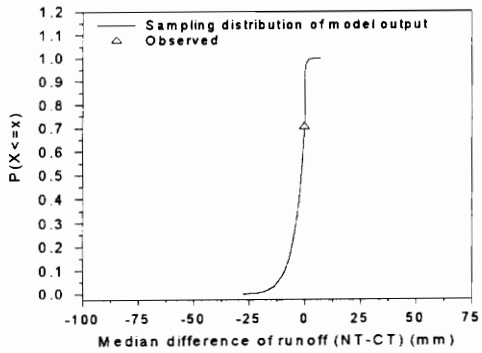
(e) Day 241



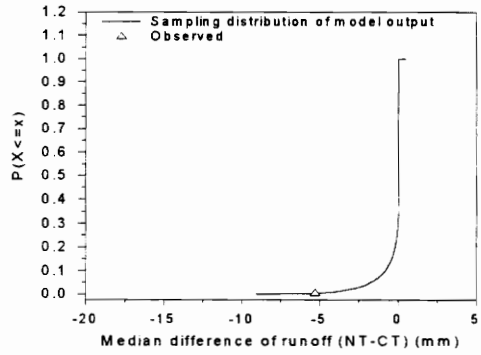
(c) Day 202

Figure C-18. MCS predicted distribution and observed value for metolachlor surface losses from the conventional tillage plot for (a) day 182, (b) day 193, (c) day 202, (d) day 235, (e) day 241.

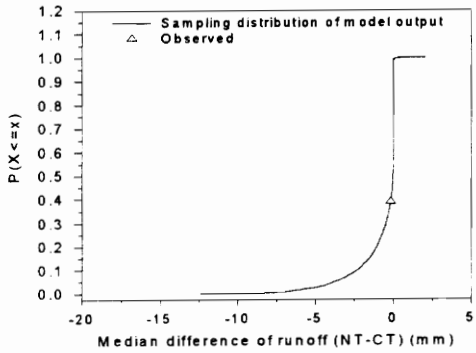
APPENDIX D



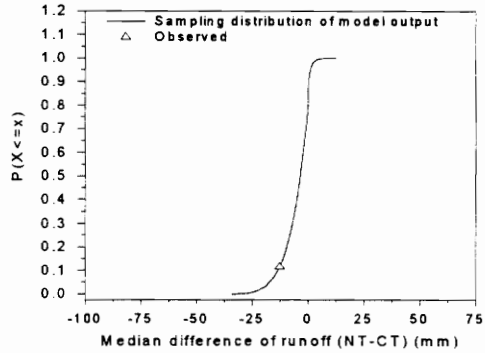
(a) Day 117



(c) Day 124

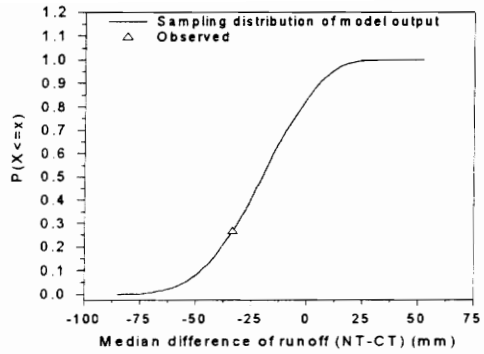


(b) Day 119

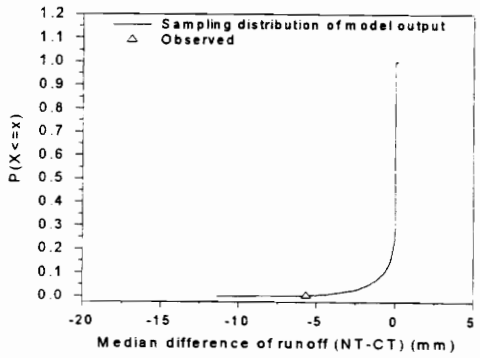


(d) Day 130

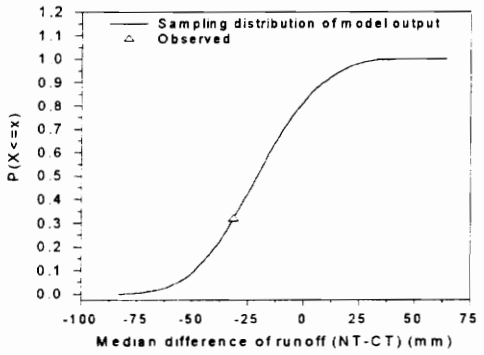
Figure D-1. Sampling distribution and observed value for differences in runoff between the no-till (NT) and conventional tillage (CT) plots for (a) day 117, (b) day 119, (c) day 124, and (d) day 130.



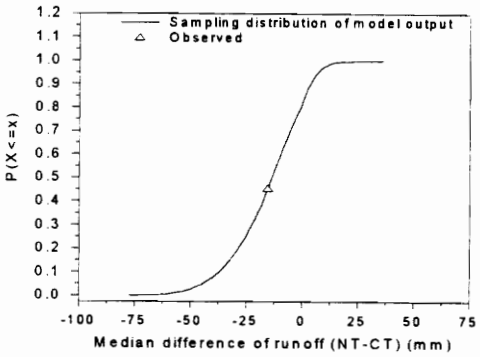
(a) Day 146



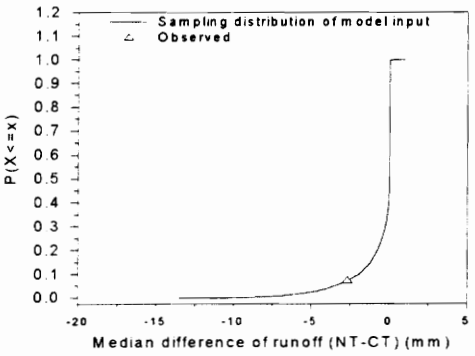
(d) Day 202



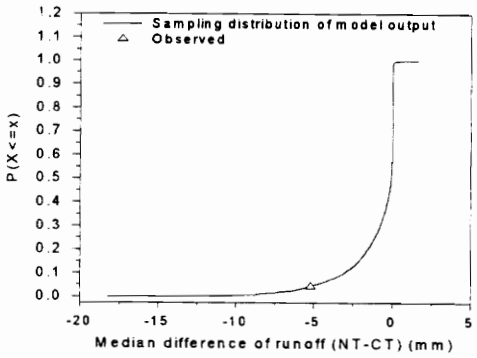
(b) Day 149



(e) Day 235

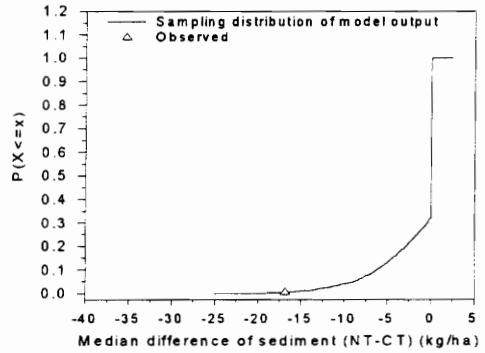


(c) Day 166

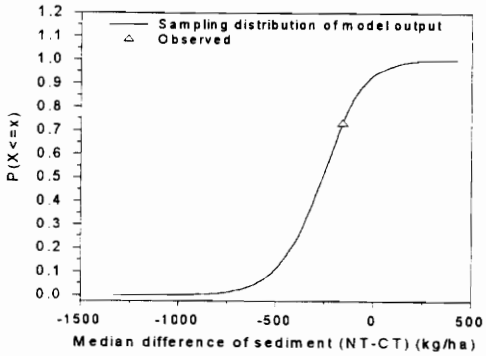


(f) Day 241

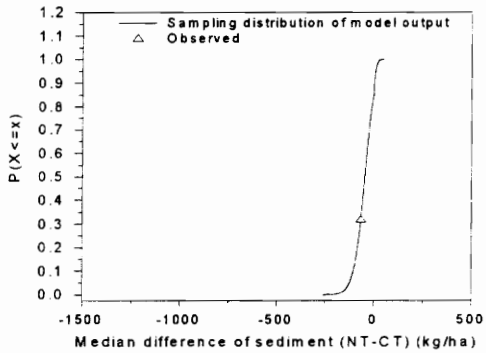
Figure D-2. Sampling distribution and observed value for differences in runoff between the no-till (NT) and conventional tillage (CT) plots for (a) day 146, (b) day 149, (c) day 166, (d) day 202, (e) day 235, and (f) day 241.



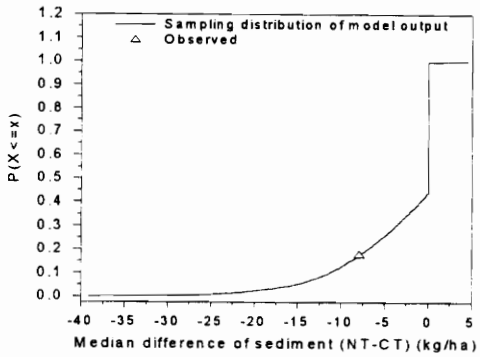
(a) Day 124



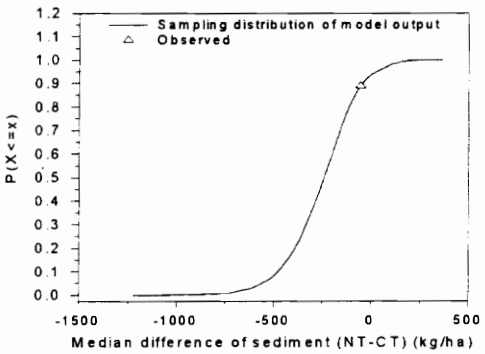
(d) Day 149



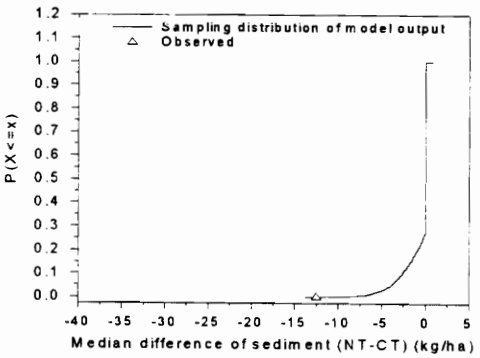
(b) Day 130



(e) Day 166

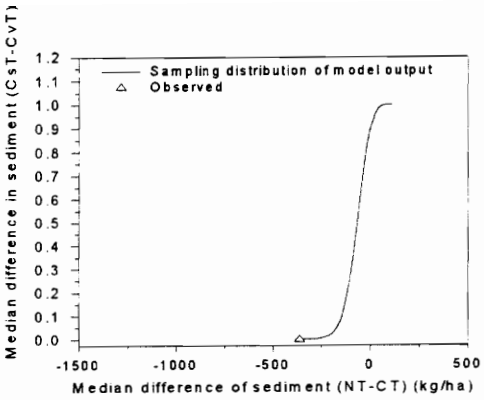


(c) Day 146

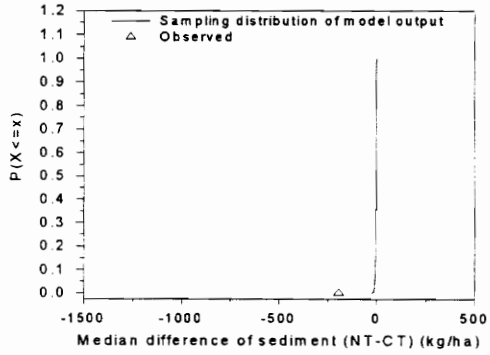


(f) Day 202

Figure D-3. Sampling distribution and observed value for differences in sediment yield between the no-till (NT) and conventional tillage (CT) plots for (a) day 124, (b) day 130, (c) day 146, (d) day 149, (e) day 166, and (f) day 202.

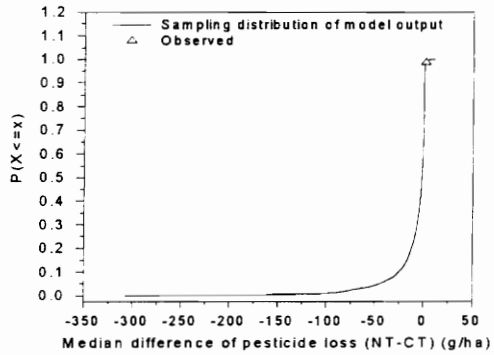


(a) Day 235

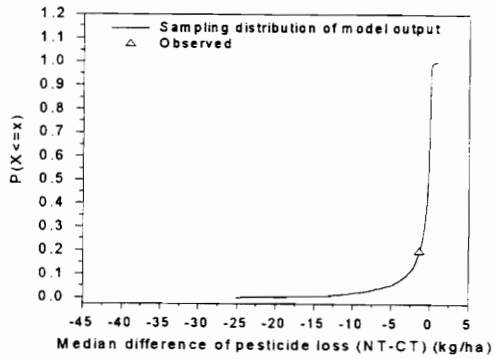


(b) Day 241

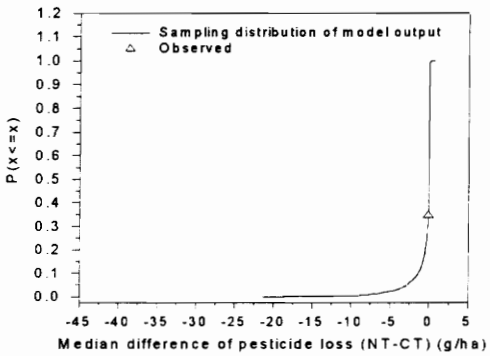
Figure D-4. Sampling distribution and observed value for differences in sediment yield between the no-till (NT) and conventional tillage (CT) plots for (a) day 235 and (b) day 241.



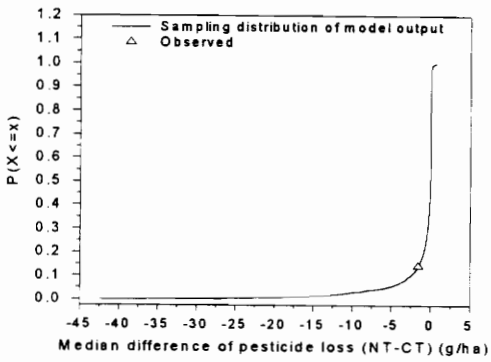
(a) Day 117



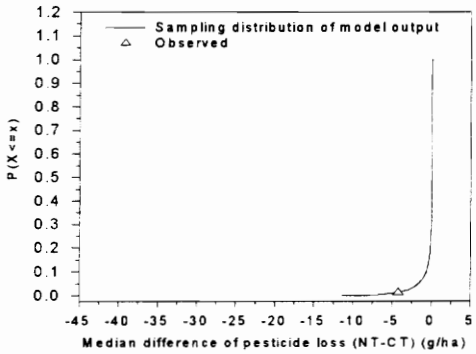
(d) Day 130



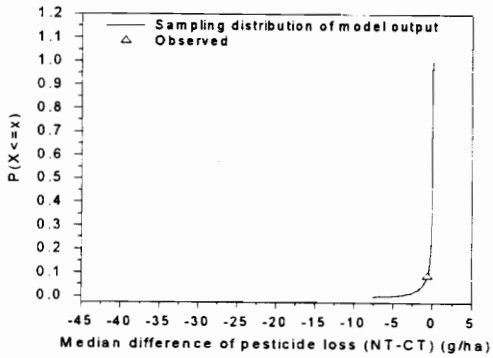
(b) Day 119



(e) Day 146

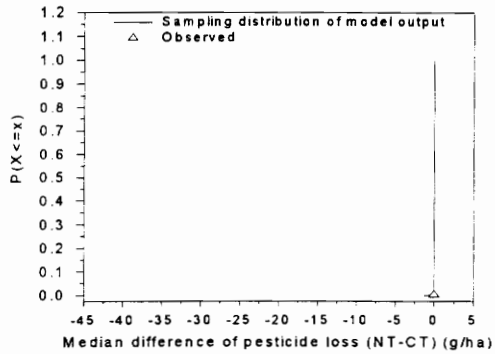


(c) Day 124



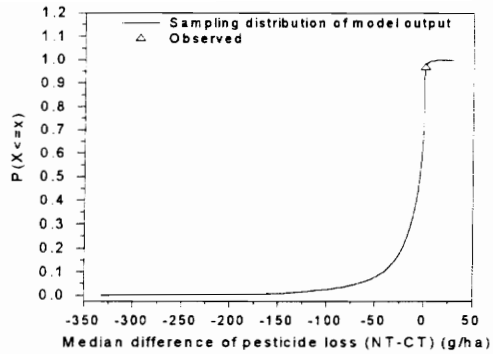
(f) Day 149

Figure D-5. Sampling distribution and observed value for differences in atrazine surface losses between the no-till (NT) and conventional tillage (CT) plots for (a) day 117, (b) day 119, (c) day 124, (d) day 130, (e) day 146, and (f) day 149.

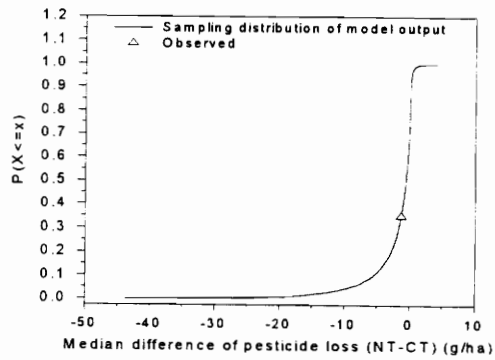


(a) Day 241

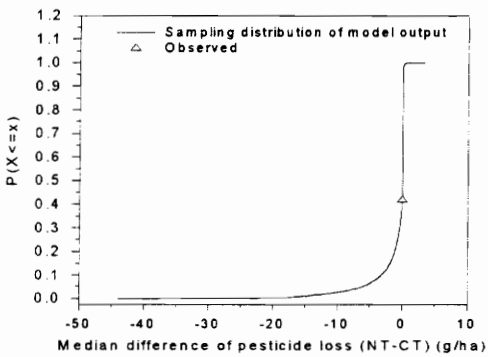
Figure D-6. Sampling distribution and observed value for differences in atrazine surface losses between the no-till (NT) and conventional tillage (CT) plots for (a) day 241.



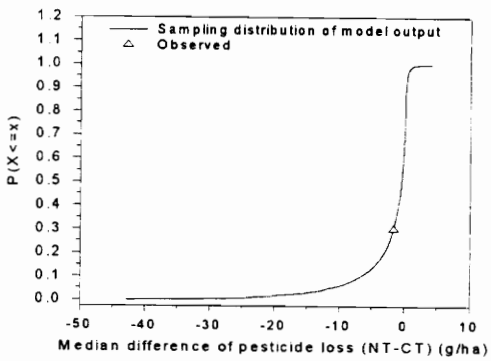
(a) Day 117



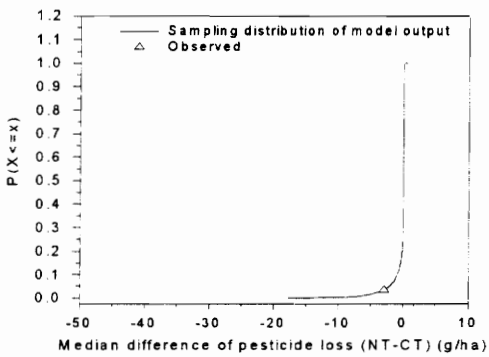
(d) Day 130



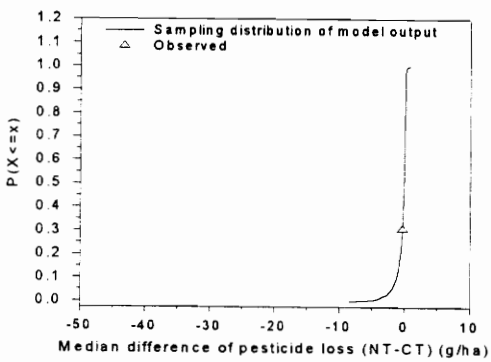
b) Day 119



(e) Day 146

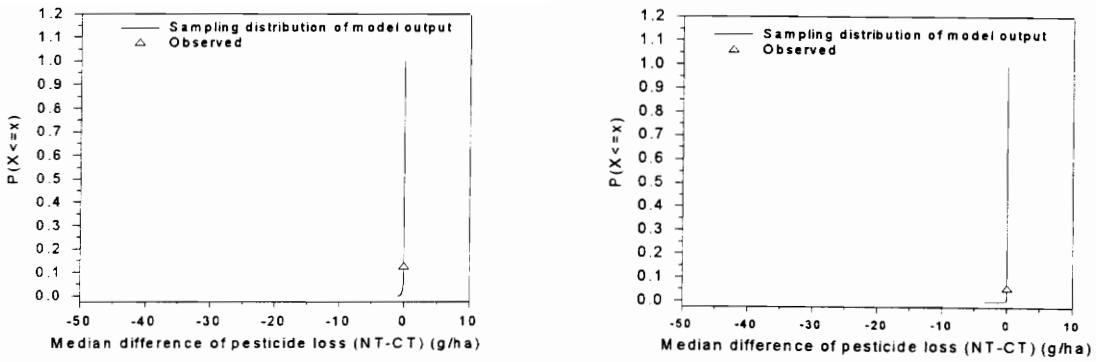


(c) Day 124



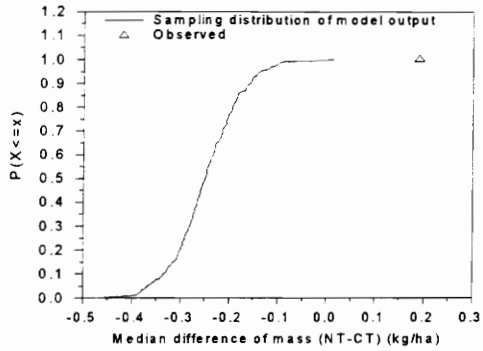
(f) Day 149

Figure D-7. Sampling distribution and observed value for differences in metolachlor surface losses between the no-till (NT) and conventional tillage (CT) plots for (a) day 117, (b) day 119, (c) day 124, (d) day 130, (e) day 146, and (f) day 149.

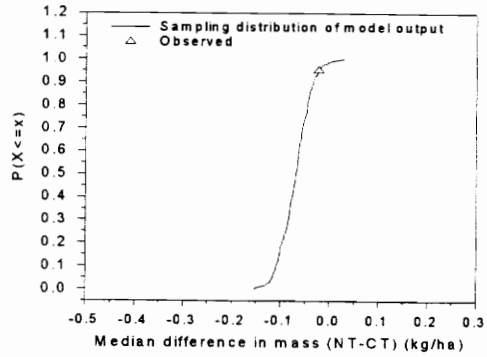


(a) Day 202 **(b) Day 241**

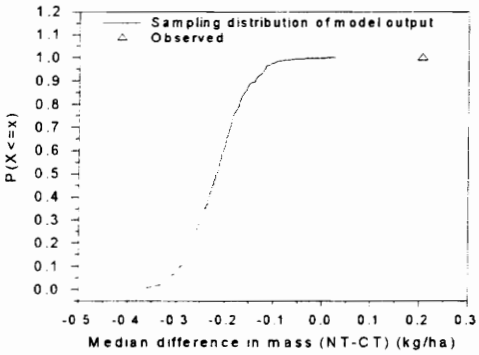
Figure D-8. Sampling distribution and observed value for differences in metolachlor surface losses between the no-till (NT) and conventional tillage (CT) plots for (a) day 202 and (b) day 241.



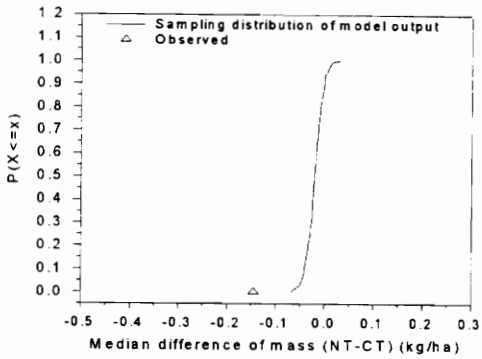
(a) Day 118



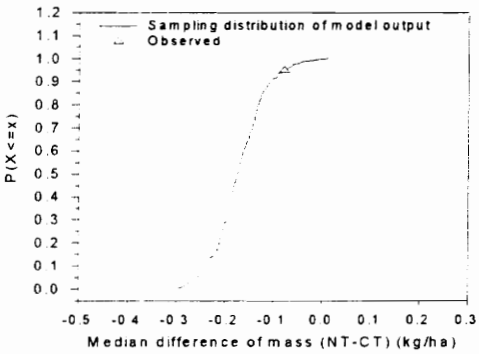
(d) Day 209



(b) Day 128

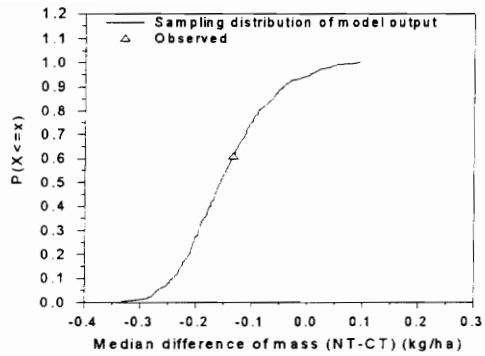


(e) Day 272

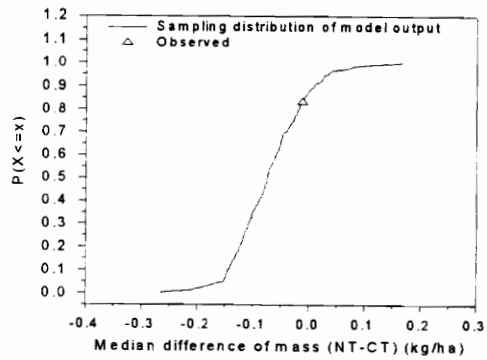


(c) Day 145

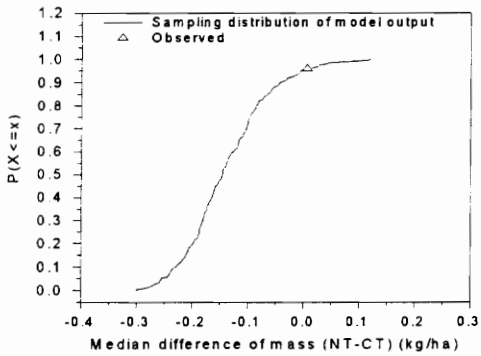
Figure D-9. Sampling distribution and observed value for differences in atrazine mass in the root zone between the no-till (NT) and conventional tillage (CT) plots for (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.



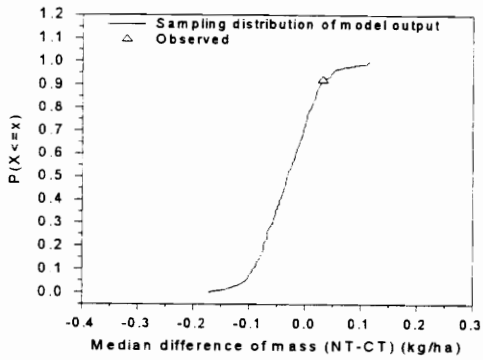
(a) Day 118



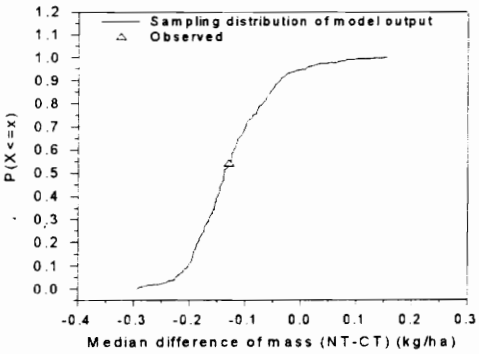
(d) Day 209



(b) Day 128

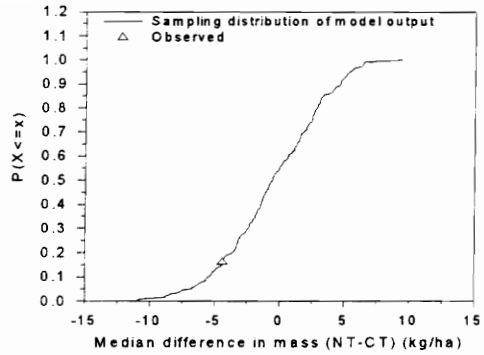


(e) Day 272

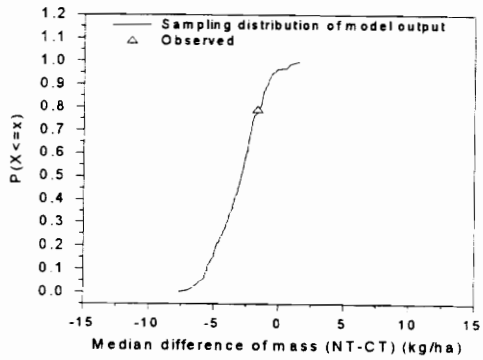


(c) Day 145

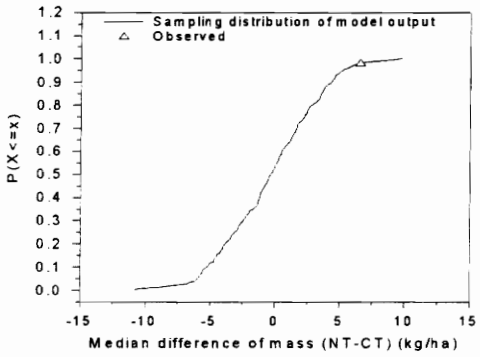
Figure D-10. Sampling distribution and observed value for differences in metolachlor mass in the root zone between the no-till (NT) and conventional tillage (CT) plots for (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.



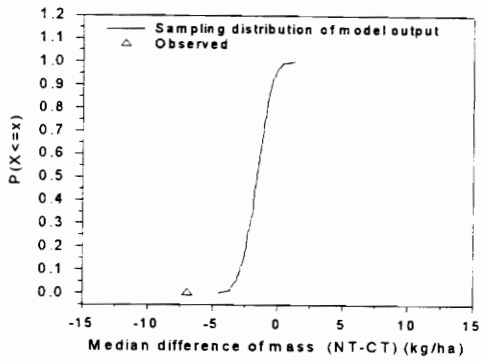
(a) Day 118



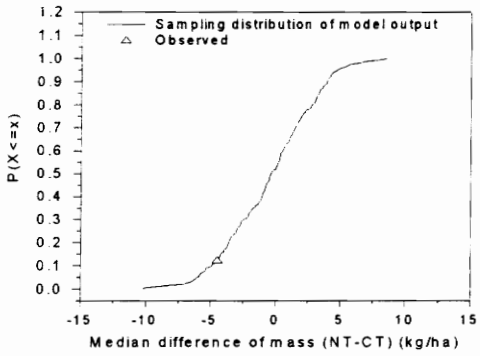
(d) Day 209



(b) Day 128

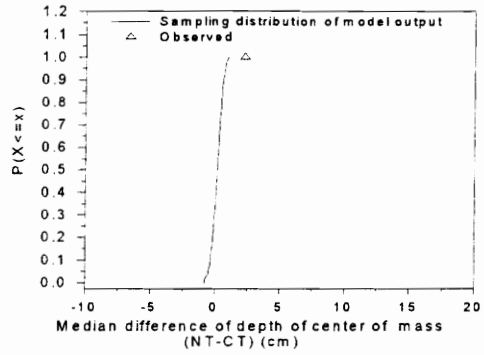


(e) Day 272

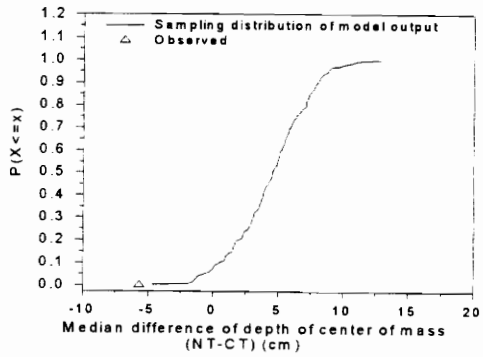


(c) Day 145

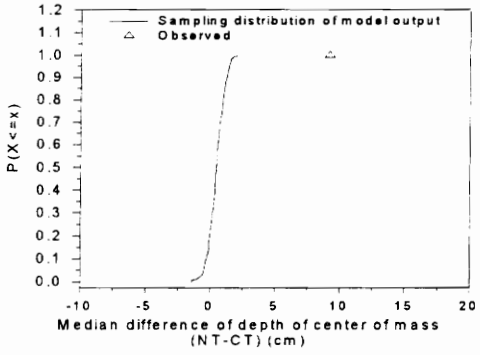
Figure D-11. Sampling distribution and observed value for differences in bromide mass in the root zone between the no-till (NT) and conventional tillage (CT) plots for (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.



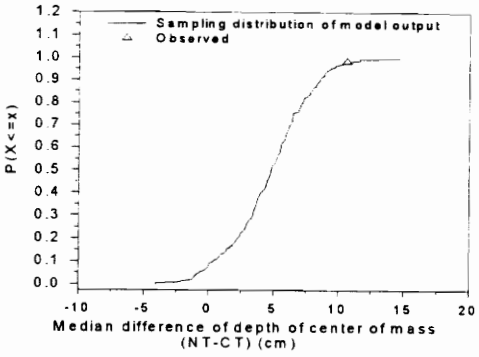
(a) Day 118



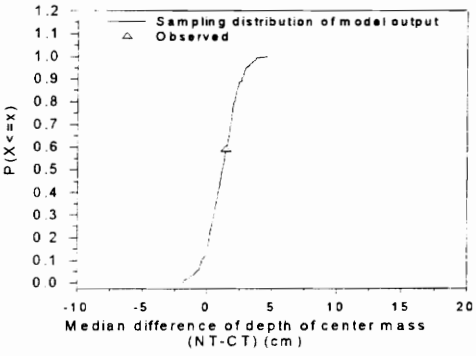
(d) Day 209



(b) Day 128

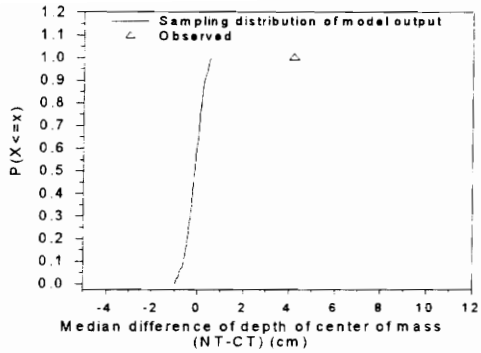


(e) Day 272

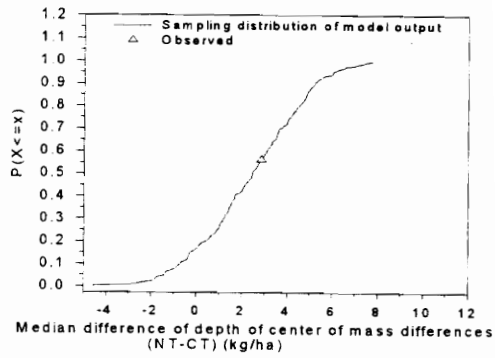


(c) Day 145

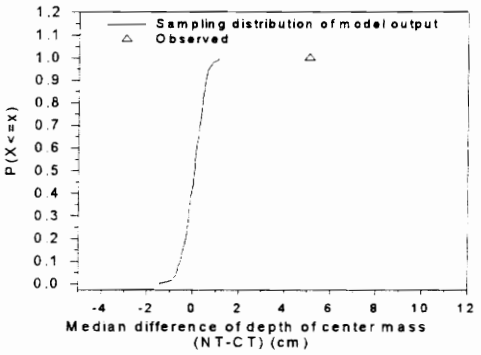
Figure D-12. Sampling distribution and observed value for differences in atrazine depth of center of mass between the no-till (NT) and conventional tillage (CT) plots for (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.



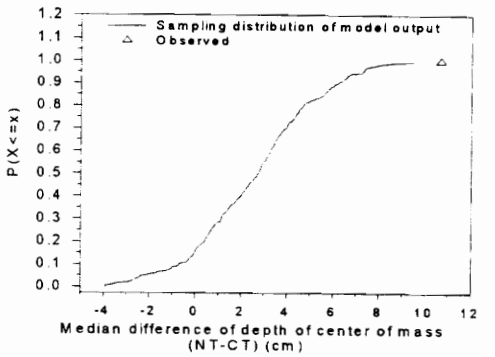
(a) Day 118



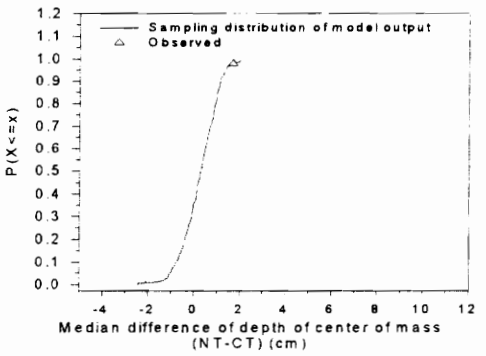
(d) Day 209



(b) Day 128

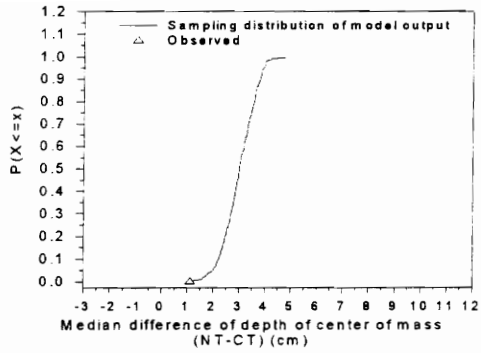


(e) Day 272

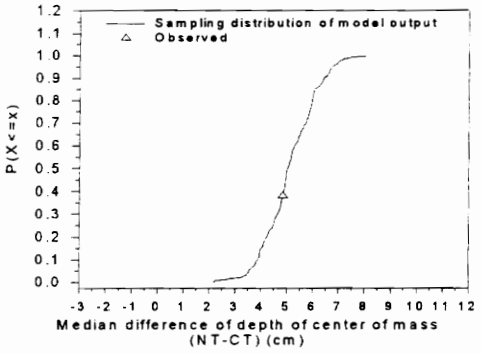


(c) Day 145

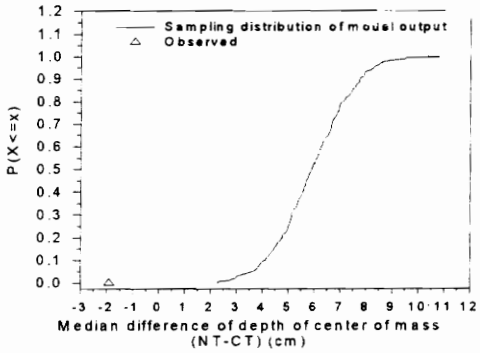
Figure D-13. Sampling distribution and observed value for differences in metolachlor depth of center of mass between the no-till (NT) and conventional tillage (CT) plots for (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.



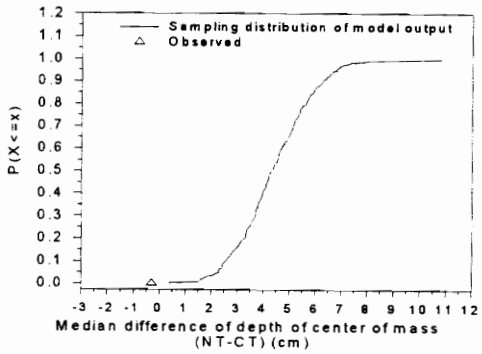
(a) Day 118



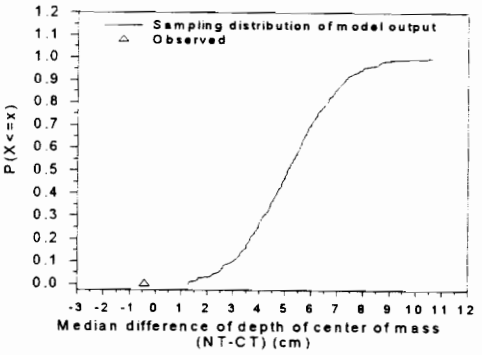
(b) Day 128



(c) Day 145



(d) Day 209



(e) Day 272

Figure D-14. Sampling distribution and observed value for differences in bromide depth of center of mass between the no-till (NT) and conventional tillage (CT) plots for (a) day 118, (b) day 128, (c) day 145, (d) day 209, and (e) day 272.

APPENDIX E

Table E-1. GLEAMS hydrology parameter set for the no-till plot.

```
Base parameter file for hydrology. Monte-Carlo based deterministic
evaluation of GLEAMS 2.10. Suffolk sandy loam (Typic Hapludults), Westmoreland
Co., VA. No-till corn (1 yr simulation)
 90000     0     0     0     1     0     1     1     1     0
     2   301   601   602   603   651   652   653   801   802
   803   811   812   813
0.0486  8.800   0.5   3.7  65.4  0.03  1.50  90.0
     0     4  15.0  30.0  75.0  90.0
 0.435 0.435 0.374 0.434
 0.196 0.240 0.259 0.224
 0.052 0.076 0.109 0.080
 1.990 3.870 10.26 8.800
 0.948 0.487 0.317 0.199
  6.84 11.81 16.60 13.67
 23.86  28.1  24.5  10.9
 11.49 12.35 14.28 18.72 21.97 28.87 30.61 28.09 24.98 22.44
 16.48 11.26
 -0.29  0.02  2.96  6.74 11.10 16.88 20.40 18.38 13.09  8.95
  2.69  0.65
 260.0 289.0 350.0 468.0 534.0 578.0 531.0 468.0 388.0 300.0
 228.0 189.0



    90    90     1
    58  0167  0289        45.0

    20  1113  1288        90.0

    74  1289  2166        22.5

     0
    -1     0
```

Table E-2. GLEAMS erosion parameter set for the no-till plot.

```
Best estimate erosion inputs for Monte-Carlo based deterministic
evaluation of GLEAMS 2.10. Suffolk sandy loam (Typic Hapludults), Westmoreland
Co., VA. No-till corn (1 yr simulation).
   90    90     0     1     1
  20.0
    1  0.0486
  27.0   0.03
    1    1.0  0.200
    1
   001    085    113    135    195    233    273
    1    1.0
   .33    .33    .29    .25    .18    .14    .33
   1.0    1.0    1.0    1.0    1.0    1.0    1.0
   .30    .30    .30    .30    .30    .30    .30
```

Table E-3. GLEAMS pesticide parameter set for the no-till plot.

```
Best estimate pesticide inputs for Monte-Carlo based deterministic
evaluation of GLEAMS 2.10. Suffolk sandy loam (Typic Hapludults), Westmoreland
Co., VA. No-till corn (1 yr simulation).
  90000  90365     3     0     0
    1       Bromide     0
    2      Atrazine     0
    3    Metolachlor     0
    1 1000.0099999.00    .00    .00   1.00    .00   1.00
99999.00
    2  33.00   5.02  100.41    .00    .61    .00    .75
  60.23
    3  530.00   5.06  200.85    .00    .71    .00    .75
  91.01
   1115     3
    1  35.43   1.00    .55  0.449     0    .00
    2    .94   1.00    .59  0.409     0    .00
    3   0.98   1.00    .59  0.409     0    .00
    0
```

Table E-4. GLEAMS hydrology parameter set for the conventional tillage plot.

```
Base parameter file for hydrology. Monte-Carlo based deterministic
evaluation of GLEAMS 2.10. Suffolk sandy loam (Typic Hapludults), Westmoreland
Co., VA. Conventionally tilled corn (1 yr simulation)
 90000     0     0     0     1     0     1     1     1     0
     2   301   601   602   603   651   652   653   801   802
   803   811   812   813
0.0486  8.800   0.5   3.7  78.7  0.03  1.50  90.0
     0     4  15.0  30.0  75.0  90.0
 0.516 0.435 0.374 0.434
 0.211 0.239 0.260 0.225
 0.051 0.076 0.110 0.081
 1.990 3.870 10.26 8.800
 0.817 0.551 0.345 0.259
  6.84  11.8 16.60 13.67
 23.86  28.1  24.5  10.9
 11.49 12.35 14.28 18.72 21.97 28.87 30.61 28.09 24.98 22.44
 16.48 11.26
 -0.29  0.02  2.96  6.74 11.10 16.88 20.40 18.38 13.09  8.95
  2.69  0.65
 260.0 289.0 350.0 468.0 534.0 578.0 531.0 468.0 388.0 300.0
 228.0 189.0



   90    90     1
   58  0167  0289        45.0

   20  1113  1288        90.0

   74  1289  2166        22.5

    0
   -1     0
```

Table E-5. GLEAMS erosion parameter set for the conventional tillage plot.

```
Best estimate erosion inputs for Monte-Carlo based deterministic
evaluation of GLEAMS 2.10. Suffolk sandy loam (Typic Hapludults), Westmoreland
Co., VA. Conventionally tilled field planted with corn (1 yr simulation).
   90    90    0    1    1
  20.0
    1  0.0486
  27.0   0.03
    1    1.0  0.200
    1
   001   085   113   135   195   233   273
    1   1.0
   .33   .78   .65   .51   .30   .25   .37
   1.0   1.0   1.0   1.0   1.0   1.0   1.0
   .25   .25   .25   .25   .25   .25   .25
```

Table E-6. GLEAMS pesticide parameter set for the conventional tillage plot.

```
Best estimate pesticide inputs for Monte-Carlo based deterministic
evaluation of GLEAMS 2.10. Suffolk sandy loam (Typic Hapludults), Westmoreland
Co., VA. Conventionally tilled field planted with corn (1 yr simulation).
  90000  90365    3    0    0
    1      Bromide    0
    2     Atrazine    0
    3   Metolachlor    0
    1 1000.0099999.00    .00    .00   1.00    .00   1.00
99999.00
    2  33.00   5.02  100.41    .00    .61    .00    .75
  60.23
    3  530.00   5.06  200.85    .00    .71    .00    .75
  91.01
   1115    3
    1  35.43   1.00    .00  0.999    0    .00
    2    .93   1.00    .00  0.999    0    .00
    3   0.88   1.00    .00  0.999    0    .00
    0
```

Vita

Randy W. Clouse was born on September 26, 1971 in Frederick County, Virginia. He received a Bachelor of Science degree in Agricultural Engineering from The Pennsylvania State University in May of 1993. From June 1993 to December 1993, he worked as a Biological Engineering intern at The Land pavilion at EPCOT Center in Lake Buena Vista, Florida. He began his Master's degree in Biological Systems Engineering at Virginia Polytechnic Institute and State University in January 1994.

Randy W Clouse