Quantifying and Communicating Model Uncertainty for Decision Making in the Everglades

Report of the

Comprehensive Everglades Restoration Plan's Model Uncertainty Workshop

U.S. Army Corps of Engineers

South Florida Water Management District

West Palm Beach, FL

January 15 – 17, 2002

Prepared by:

Upmanu Lall, Donald L. Phillips, Kenneth H. Reckhow
and Daniel P. Loucks (Chair)

April 2002

Quantifying and Communicating Model Uncertainty for Decision Making in the Everglades

Executive Summary

General aspects of model uncertainty

The Comprehensive Everglades Restoration Plan (CERP) is highly dependent on the results of dynamic regional hydrologic and ecologic simulation models. Even though these models, and those that may eventually replace them, are and will continue to be relatively complex and sophisticated, they, like all such models, only crudely approximate what actually takes place in the Everglades. Their predictive abilities, even in a statistical sense, are not, and will never be, perfect. Thus there is uncertainty in the predictions derived from these models. This uncertainty should be taken into account when evaluating model results.

The likelihood of capturing within simulation models all the processes occurring in a system as complex as the Everglades is not likely to happen, nor is it likely to be cost-effective to attempt such a task. Hence those involved in CERP and RECOVER will have to make decisions taking this inevitable uncertainty into account.

An uncertainty analysis is not the same as a sensitivity analysis. An uncertainty analysis attempts to describe the entire set of possible outcomes, together with their associated probabilities of occurrence. A sensitivity analysis attempts to determine the relative change in model output values given possible changes in model input values. A sensitivity analysis thus measures the change in the model output in a localized region of the space of inputs.

Performing sensitivity analyses is, or should be, standard procedure when modeling regions such as the Everglades. While one can often extend sensitivity analyses to a more comprehensive uncertainty analyses, but it may not be practical.

Collecting additional information on model parameters or using a longer period of data can reduce uncertainty. This reduction in uncertainty is manifest in “tighter” probability distributions of the variables of interest. Whether it is worth paying for such a reduction in uncertainty depends on the time and effort involved and in the subsequent potential impact on the decision. Quantitative procedures for exploring such questions are available.

Sometimes, increasing model complexity (higher resolution, more explanatory variables) is considered to reduce model uncertainty. Unfortunately, given a fixed amount of data one can reduce errors in fitting observations in this manner, but it may not increase the reliability or accuracy of prediction.

Model output uncertainty comes from input variability and measurement errors, parameter uncertainty, model structure uncertainty and algorithmic (numerical) uncertainty. These uncertainties are translated into uncertainty in the performance indicators and measures. In addition, there is uncertainty as to whether the specific performance indicators and measures used to characterize the overall system performance actually capture that overall performance. There is also uncertainty associated with performance measure targets that have been established.

The uncertainty analyses recommended for RECOVER involve identifying characteristics of various probability distributions of model inputs and then using those distributions to derive probability distributions of output variables. We recommend ways of doing this that reduce the computational burden compared to Monte Carlo simulations.

Models that have been designed to simulate hydrologic or ecologic processes typically have too many parameters for optimal estimation from observational data. As a consequence, the parameters are judgmentally chosen and model fitting becomes an art; more experienced modelers/artists will produce better fitting models. These models are designed to describe typical or average behavior; thus it is reasonable to expect that a good process model will yield a prediction trajectory that goes through the middle of the time series of observations, once the model is fully parameterized. Presumably, as more processes are adequately represented in the model, the model time trajectory will begin to capture the short-term fluctuations in the observations more accurately. However the model might be expected to underestimate the extremes, since its structure is more compatible with the central tendency. This point is important because it implies that parameter selection aimed at fitting the extremes (e.g., high phosphorus levels) is incompatible with the model structure that is designed to describe average system behavior.

Random error may be reduced by increasing the number of supporting data points included in the calculation of cell means. However, systematic error or bias, due for example to subsidence since the time when the elevations were measured, is not necessarily reduced as sample size within a cell increases.

Spatial-temporal models used by CERP and RECOVER contain variables whose values vary over space and time. The value of the information obtained from such models comes from their simultaneous depiction of temporal and spatial patterns. The influence of scale, i.e. the mismatches between the spatial and temporal resolutions of the input data compared to that of the output data, the mismatches between the times and places calibration data were obtained compared to those of the predicted data, data measurement and collection errors, methods such as kriging used for interpolation and extrapolation over space and time, and the quality of any analysis used to produce model input data, all contribute to the uncertainty of model output results. These uncertainties become important when one tries to develop and use spatial-temporal models for predicting hydrologic and ecologic performance over space and time.

Uncertainty in models that predict impacts over time and space resides largely within four components of model structure. These components are the inputs – the spatial and temporal interpolation of point data, the initial and boundary conditions, and in the calibration and verification procedures. Sources of uncertainty in spatial data are similar to the sources of uncertainty in temporal data – both are associated with data measurement and collection, data processing, model structure, natural variability, and human intervention.

Recommendations in the short run

It seems to us that in the short term it is important to have some way to derive estimates of the probability distributions of uncertain output variables in a very practical and transparent way, without requiring an enormous amount of work and time. More precise methods might produce more precise estimates, and that such methods could be pursued in the future should better estimates of uncertainty prove critical to decision making.

To summarize, the steps involve

- selecting the significant independent input variables (including model parameters) that contribute most significantly to the final model prediction,
- constructing probability density functions for each parameter to reflect the likelihood that the selected variable will take on various values within its possible range,
- propagating the uncertainties through the model to generate a probability density distributions of predicted output values and subsequently of the hydrologic attributes that impact important ecosystem indicators and features,
- deriving, if desired, confidence limits associated with the functions that convert hydrological attributes to indices indicating the relative suitability of conditions for the ecosystem indicators or features,
- using these confidence intervals plus the derived probability density distributions of the hydrological attributes to make quantitative statements about the probabilities (and the confidence in these probabilities) of meeting selected suitability index levels for each index.

RECOVER should strive for further reduction of the number of performance measures used to evaluate water management alternatives and find ways to collectively evaluate performance on a number of measures in order to have a clear-cut comparison between competing alternatives.

One way in which the number of hydrologic performance measures might be further reduced is to use the ecological models (e.g., ELM, ATLSS) or at least habitat suitability functions to determine which measures are of greatest importance to ecological processes and wildlife populations simulated in these models. While operationally CERP focuses on hydrologic restoration, ecological endpoints are of great concern and hydrologic restoration is viewed as the best way to effect ecological restoration as well. Hence, knowledge gained from the ecological models about which hydrologic performance measures are most important ecologically can be used to focus efforts on those measures.

Hydrologic performance measures generated by the South Florida Water Management Model (SFWMM) number in the hundreds. These have been reduced to about hundred key performance measures for examination and decision making about management alternatives. Further reduction is needed, perhaps by the use of indices.

If the plan evaluation process has difficulty handling all this it may indicate the need to focus the uncertainty analysis effort on just what is deemed important, doable, and beneficial. Then when the alternatives have been narrowed down to only a few plans that appear to be the better ones, a more complete uncertainty analysis can be performed. There is no need to nor benefit in performing sensitivity and uncertainty analyses on all possible management alternatives. Rather focus on those that look the most promising, and than only when the tradeoffs among important uncertain performance indicator values demands more scrutiny. Otherwise it is just too much work and it will likely not affect the decision anyway.

Acknowledgements

We have benefited from the reviews of an earlier draft of this report and from all who participated in the workshop. We especially want to thank Mike Choate, Don DeAngelis, Carl Fitz, Tom James, Tony King, Wasantha Lal, John Miller, Brenda Mills, Sherry Mitchell, Jayantha Obeysekera, John Ogden, Fred Sklar, Ken Tarboton, Randy van Zee, and many others who gave us a little of their wisdom. We can only hope we got it right – and that some of what we have written will be useful.

Quantifying and Communicating Model Uncertainty for Decision Making in the Everglades

Table of Contents:

1. Introduction

The Comprehensive Everglades Restoration Plan (CERP) involves numerous projects to be implemented over the next 30 years. This effort will also include monitoring of the response of the system to projects as they are implemented. It is the responsibility of the Restoration Coordination and Verification (RECOVER) team to evaluate the system-wide performance of each project and the overall performance of the Plan as it is carried out. The RECOVER team relies heavily on the various regional dynamic models (described in the Appendix) to predict project impacts expressed as performance indicators or performance measures. Dependent on the results of these performance evaluations, the Plan may be modified and refined.

By definition, models approximate reality. Spatially distributed hydrologic and ecologic models (such as SFWMM, NSM, and potentially ELM and ATLSS) used to study regional impacts of alternative water management practices or policies are approximate and have a relatively coarse resolution relative to the scale at which hydrology and ecology evolve in the Everglades. The coarse resolution reflects both computational and informational limitations. Data on hydrologic and ecologic conditions as well as on inputs to the models are available at a limited, often non-uniform resolution.

These models have been calibrated using measurements available at the gaging sites within the Everglades region. For example the SFWMM has been used successfully for over twenty years based on a calibration at 161 monitoring sites and 22 canal locations. The calibration process seeks to minimize errors of prediction at the locations where we have measured values. Given the approximation of reality, and the information and scale limitations, errors between predicted and observed quantities remain after calibration. Since the model extrapolates the observations to unsampled locations in space and time, the model "predictions" at those locations will also have errors. These errors are usually larger than those at sampled points. They can have well defined patterns in space or time.

Since a primary use of these models is predicting conditions at ungaged as well as at gaged locations and under conditions different from those observed, it is important to understand the nature of such errors. In addition to the errors of approximation, measurements of model input and output variable values often contain errors, and the conditions under which the model is calibrated may not be representative of conditions under which the model is to be used. For instance, the climate of the period used for calibration may be different from prior or future periods. Since models are used sequentially, e.g., a model is used to interpolate rainfall over space and time, and these data are inputs into a hydrologic model that computes water levels which in turn are used as inputs to an ecological model. The output from ecologic models can be input to models for calculating economic or other performance measures. Errors can propagate throughout this sequential modeling process. This leads to a varying degree of uncertainty in predicted outcomes at different locations and times. Loucks and Stedinger (1994) provide an example of this.

The RECOVER team that uses these models recognizes the need to consider the uncertainty in model output predictions when they make their decisions. The issue addressed in this report is just how uncertainty can be predicted and expressed in ways that will be of use to RECOVER. Any proposed uncertainty analysis method must be compatible with the time, data, computational and human resources available to the RECOVER team.

To identify just what types of uncertainty analyses might be most appropriate, a multi-agency workshop was held on model uncertainty analysis at the South Florida Water Management District in West Palm Beach during January 15 through 17, 2002. The primary goals of the uncertainty workshop were to:

1. Recommend a method to characterize and quantify the uncertainty in predictive system-wide models used for CERP,
2. Show how to characterize and quantify the uncertainty associated with various performance indicators and measures produced from the outputs of these models and,
3. Develop ways of identifying the tradeoffs among performance measure target values and their reliabilities for RECOVER evaluations.

In general the workshop goal was to provide guidelines on how to deal with uncertainty to those responsible for evaluating model alternatives.

This report supplements the discussions that took place in the Model Uncertainty Workshop. It outlines methods of quantifying uncertainty in model output including the performance measures used by RECOVER, and presents ways of incorporating risk assessment (or risk decision analysis) in the evaluation of various performance measures.

1.1 What is uncertain?

The Comprehensive Everglades Restoration Plan (CERP) is highly dependent on the results of dynamic regional simulation models. The ecological models (e.g., ELM and ATLSS) used to predict the responses of selected ecosystem variables to water management policies depend on the output of the hydrologic models (e.g., SFWMM and NSM). Even though these models, and the improved models that may eventually replace them, are relatively complex and sophisticated, they, like all such models, only crudely approximate what actually takes place in the Everglades. Their predictive abilities, even in a statistical sense, are not perfect.

A fixed period (e.g., 1965-1995) of historical data is used to calibrate each model such that the available measurements of output variables are reproduced with minimum error given the inputs and the calibrated parameters. The quality of both input and output data used often varies over the period of calibration. Thus, even though the model may be deterministic (i.e., the equations map the inputs into a fixed set of output values), there is uncertainty about the estimates from the model. In these situations a probabilistic approach is often considered to describe this uncertainty. Sometimes, increasing model complexity (higher resolution, more explanatory variables) is considered to reduce this uncertainty. Unfortunately, given a fixed amount of data one can reduce errors in fitting observations in this manner, but may decrease the reliability or accuracy of prediction during the use of the model as part of a decision process.

Further, for management applications, models are required to provide estimates of performance measures at locations for which we may have no measurements (either for calibration or validation). The estimation of probability distributions to characterize the joint (spatial and temporal) uncertainty of output variables from a model is rather difficult in this setting.

Traditionally, uncertainty in some boundary conditions (e.g., climate inputs, or altitude) is assessed through a scenario analysis. Scenario analysis methods generate sequences of representative and spatially and temporally consistent collections of values based on a set of assumptions, usually about the future. Such analyses allows one to explore alternate plausible conditions that may exist as inputs to the model. While, such analyses are relatively straightforward in concept, their practical application can require some assumptions that may or may not be verifiable.

Similar comments apply to uncertainty about model parameters that need to be calibrated grid cell by grid cell in spatially distributed models. As hydrological and ecological models are linked, input and output data may not be at the same locations or at the same resolution, necessitating interpolation or regridding procedures that potentially add errors. The practical assessment and characterization of such uncertainties, their propagation across models, across space and time, and their local and global variation and impact on the decision process through effective communication was addressed in the workshop. This assessment included going beyond the traditional framework of uncertainty analysis to consider the impact of factors such as climate change that may lead to much greater uncertainty in performance relative to model refinement.

Collecting additional information on model parameters or using a longer period of data can reduce uncertainty. This reduction in uncertainty is manifest in “tighter” probability distributions of the variables of interest. Whether it is worth paying for such a reduction in uncertainty depends on the time and effort involved and in the subsequent potential impact on the decision. Quantitative procedures for exploring such questions are available.

Key to refining and implementing projects for CERP is the evaluation of various performance indicators and measures derived from the output of simulation models. Model output uncertainty, originating from input variability and measurement errors, parameter uncertainty, model structure uncertainty and algorithmic (numerical) uncertainty, is translated into uncertainty in the performance indicators and measures. In addition, there is uncertainty as to whether the specific performance indicators and measures used to characterize the overall system performance actually capture that overall performance. There is also uncertainty associated with performance measure targets that have been established.

There is a need to quantify this uncertainty for the specific performance indicators and measures used in RECOVER evaluations. These uncertainties will vary depending on the locations at which they are produced. Once this uncertainty is quantified, those responsible for evaluating model alternatives can consider this uncertainty as they make decisions as to the relative merits of each alternative. They will be seeking alternatives that have the highest probability of achieving project goals while minimizing the risk of undesired outcomes. This will involve making tradeoffs among alternative target performance indicators, measures, and targets, together with their reliabilities (probabilities of not being met), and their costs. In some cases, reliable estimates of the quantities of interest or of the uncertainty associated with these estimates may not be possible. It is important to identify such a situation and to inform the decision makers of the need to acquire additional information and its potential value. The efficacy and utility of a probabilistic framework for uncertainty analysis for the CERP applications will also be discussed.

2. Uncertainty Analyses

This section begins with a brief description of some general concepts of uncertainty analysis that will serve as definitions of terms and methods referred to in the remainder of this report. This brief overview clearly does not serve to replace what can be found in textbooks on the subject, but it may help some better understand what is being presented in later sections without having to refer to textbooks. For those that already know these basic concepts, please go on to the next section.

2.1 General concepts

Describing uncertainty using probability distributions

Uncertainty involves the notion of randomness. If a value of a performance indicator or performance measure, or in fact any variable, like the phosphorus concentration at a particular location or the depth of water averaged over a 2x2 square mile area, varies and this variation over space and time cannot be predicted with certainty, it is called a random variable. One cannot say with certainty what the value of a random variable will be but only the likelihood or probability that it will be within some specified range of values. The probabilities of observing particular ranges of values of a random variable are described or defined by a probability distribution. There are many types of distributions and each can be expressed in several ways.

Suppose the random variable is X. If the observed values of this random variable can be only discrete values, the probability distribution of X can be expressed as a histogram, as shown in Figure 1a. If the random variable is a continuous variable that can assume any real value over a range of values, the probability distribution of X can be expressed as a continuous distribution as shown in Figure 1b. The shaded area under both the histogram and the continuous distribution is 1. Hence, the area between two values of the random variable, such as between u and v in Figure 1c, represents the probability that the observed value x of the random variable value X will be within that range of values. These probability distributions shown in Figure 1 are called probability density functions (pdf) denoted by $Pr_X(x)$ for the discrete case and $f_X(x)$ for the continuous case. The subscript X represents the random variable, and the variable x is some value of that random variable X.

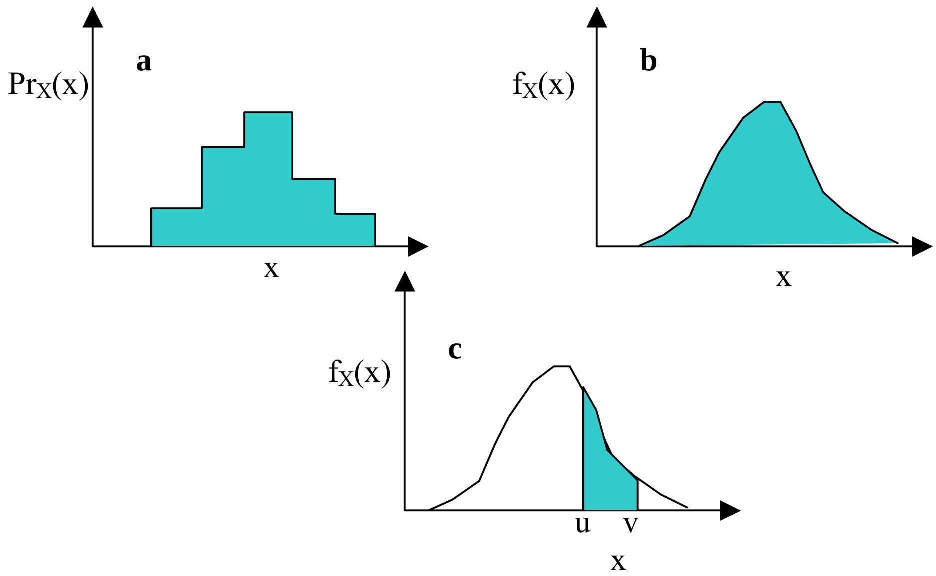


Figure 1. Probability density distributions for a discrete or continuous random variable X. The area under the distributions (shaded areas in a and b) is 1, and the shaded area in c is the probability that the observed value x of the random variable X will be between u and v.

A plot of the accumulated area under the continuous probability density distribution, $f_X(x)$, as x increases, is shown in Figure 2. This function, ranging from 0 to 1, is called the cumulative distribution function (cdf) and is denoted by $F_X(x)$.

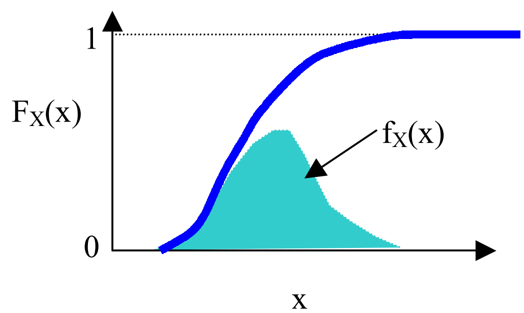


Figure 2. The cumulative distribution function of the random variable X. It represents the sum of the area under the probability density function, $f_x(x)$, as x increases.

$$F_X(x) = \int_0^X f_X(x)\,dx$$

The cumulative distribution function $F_X(x)$ specifies the probability that the observed value of the random variable X will be less than or equal to some selected value x.

$$F_X(x) = \text{Probability}\{X \le x\}$$

For the cumulative distribution shown in Figure 2, when x is 0, we know that probability of the random variable being less than or equal to this is 0. If x is very large, there is a much higher probability that the observed value of the random variable will be less than or equal to x.

Subtracting the cumulative distribution function from 1 defines the probability of exceedance function, as shown in Figure 3.

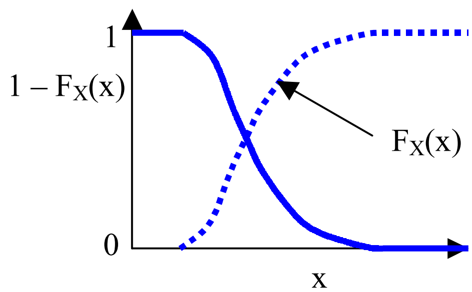


Figure 3. Probability of exceedance function, the solid line, is 1 minus the cumulative distribution function, the dashed line. The area under the probability of exceedance function is the mean or expected value of the random variable X.

This probability of exceedance function represents the probability that the observed value of the random variable X will be greater than x.

$$1 - F_X(x) = \text{Probability}\{X > x\}$$

Characterizing uncertainty using statistics

Various statistics are often used to characterize probability distributions. Common statistics include the mean, (or average or expected value) and a measure of the spread of a distribution about the mean. The mean, average, or expected value of the random variable X can be computed as

$$E[X] = \Sigma_x\, x\, Pr(x) \qquad \text{for the discrete case.}$$

$$E[X] = \int_0^{\infty} x\, f_X(x)\, dx \qquad \text{for the continuous case.}$$

A common measure of the spread of a probability distribution is called the variance. The variance of a discrete distribution is calculated by summing up the squares of the differences between all discrete values of x and the mean value, E[X], of the random variable X, multiplied by the probability of x.

$$\text{Var}[X] = \Sigma_x [(x - E(X))^2 Pr_x(x)]$$

The standard deviation SD[X], is the square root of the variance.

$$\text{SD}[X] = \{\text{Var}[X]\}^{0.5}$$

To illustrate what these measures or characteristics of a probability distribution are, consider a discrete uniform distribution of integer values from 1 to 10 each with a probability of 0.10. This distribution is shown in Figure 4.

The mean, E[X], variance, Var[X], and standard deviation SD[X], are:

$$E[X] = (0.10)(1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10) = 5.5$$

$$\text{Var}[X] = (0.10)\{(1\text{-}5.5)^2 + (2\text{-}5.5)^2 + (3\text{-}5.5)^2 + (4\text{-}5.5)^2 + (5\text{-}5.5)^2 + (6\text{-}5.5)^2 + (7\text{-}5.5)^2 + (8\text{-}5.5)^2 + (9\text{-}5.5)^2 + (10\text{-}5.5)^2\} = 8.25$$

$$\text{SD}[X] = (8.25)^{0.5} = 2.87$$

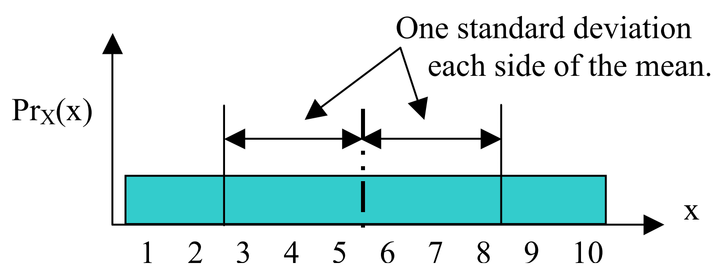


Figure 4. A uniform probability density distribution over the range of integer values from 1 to 10. The mean is shown at 5.5 and the arrows represent one standard deviation from the mean.

Note that one standard deviation each side of the mean value of the probability distribution includes about 2(2.87)/10 or 0.574 or over 57% of the total area of the distribution, i.e., from 5.5-2.87 to 5.5+2.87. Different distributions would have different areas associated with one standard deviation each side of the mean, but typically these areas are more than 50% of the area.

If the probability density distribution were a bell-shaped normal distribution, one standard deviation each side of the mean includes about 68% of the area under the probability density distribution – the bell curve. Two standard deviations each side of the mean will include about 95% of the area.

The uncertainty analyses in RECOVER involve identifying characteristics of various probability distributions of model input and output variables, and subsequently functions of those random output variables that have been called performance indicators or measures that are themselves uncertain or random.

A complete uncertainty analysis would involve a comprehensive identification of all sources of uncertainty that contribute to the probability distributions of each input or output variable. Assume this were done for two alternative project plans, A and B, and that the resulting probability density distributions for a specified performance measure were as shown in Figure 5. Figure 5 also identifies the costs of these two projects. The introduction of two performance criteria, cost and probability of exceeding a performance measure target, introduces a conflict where a tradeoff must be made.

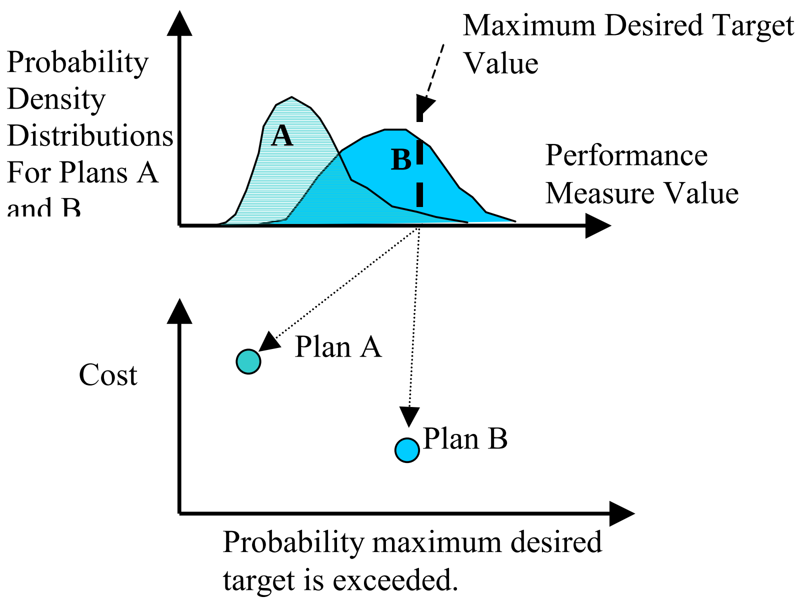


Figure 5. Tradeoffs involving cost and the probability that a maximum desired target value will be exceeded. In this illustration the performance measure is a maximum desired value (e.g., the area occupied by cattails in a region of the Everglades). Values in excess of these targets are considered undesirable or unsatisfactory.

2.2 Sensitivity and uncertainty

In a report on sensitivity and uncertainty analysis of the NSM in South Florida, Lal (1995) shows the how input parameter uncertainty and sensitivity can affect the uncertainty of an output variable or performance indicator. His schematic diagram is worth repeating here, in Figure 6.

Model error propagation, as illustrated in Figure 6, is discussed and referenced in Loucks and Stedinger (1994) and in Reckhow, Clements and Dodd (1990). Propagation error in water quality models has traditionally been estimated in two ways. For a comprehensive assessment, error propagation using Monte Carlo simulation can be applied to estimate the collective effect of individual error terms on the prediction error. However, this process is data intensive and may be computationally unwieldy for large models.

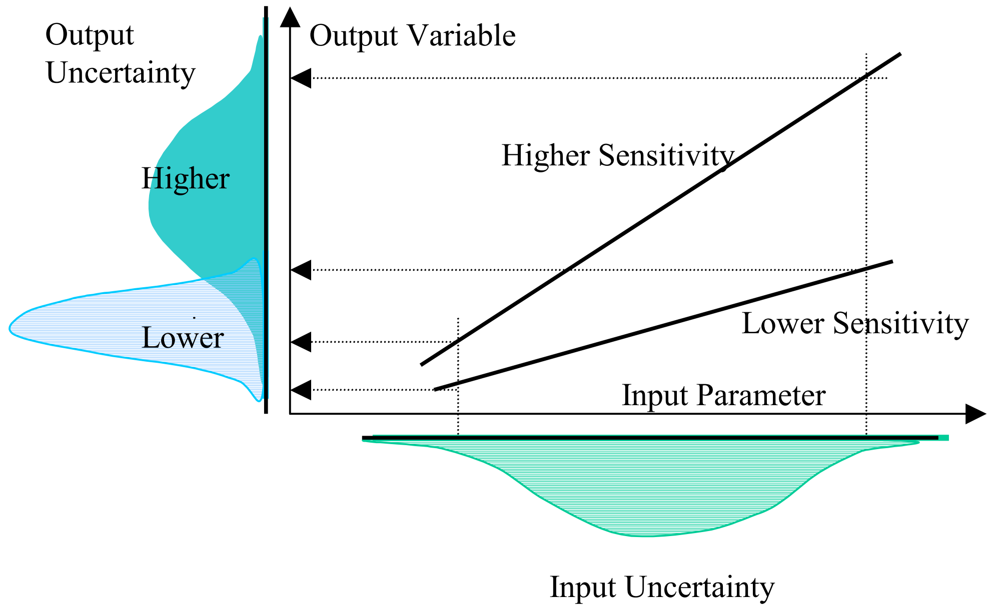


Figure 6. Schematic diagram showing relationship among input parameter uncertainty and sensitivity to output variable uncertainty (Lal, 1995).

A second and simpler approach is to compare predictions with observations, although the correct interpretation of this analysis is not as straightforward as it may seem. Prediction / observation differences will likely yield a smaller error estimate than that obtained from a thorough error propagation study, but it is unclear how to make these two estimates equivalent. Probably of greater importance, if a model is "over-fitted" to calibration data and the test or "verification" data are not substantially different from the calibration data, the prediction / observation differences will likely underestimate the prediction error in new scenarios. The best way to avoid this is to obtain independent verification data substantiated with a statistical comparison between calibration data and verification data.

In error propagation studies, we must be wary of the adequacy of the error estimate for individual model terms. For example, prediction error might be estimated using error propagation (e.g., Monte Carlo simulation) based on error terms for model parameters, inputs, initial and boundary conditions, and the model structure. However, if parameter covariance terms are ignored, if model equation error cannot be characterized, or if other deficiencies are evident but not justified, then it is unclear what the computed error represents. Thus, regardless of the rigor of the model fitting exercise and the statistical properties of the estimators, if the prediction error estimate is incomplete, biased, or in some way does not reflect the application scenario, then the error analysis can be misleading.

If a model is fitted to observational data using least squares, maximum likelihood, or Bayesian analysis, the error term has a clear meaning, with the prediction interval or Bayesian posterior distribution centered on the fitted observations. If the model is fitted using judgmental parameter selection and the error term is estimated from differences between predictions and observations, then the center of a prediction interval based on these differences may not be centered on the fitted observations. Much depends on the judgment of the modeler.

Models that have been designed with the primary objective of process description typically have too many parameters for optimal estimation from observational data. As a consequence, the parameters are judgmentally chosen and model fitting becomes an art; more experienced modelers/artists will produce better fitting models. These models are designed to describe typical or average behavior; thus it is reasonable to expect that a good process model will yield a prediction trajectory that goes through the middle of the time series of observations, once the model is fully parameterized. Presumably, as more processes are adequately represented in the model, the model time trajectory will begin to capture the short-term fluctuations in the observations more accurately, but otherwise the model might be expected to underestimate the extremes, since its structure is more compatible with the central tendency. This point is important because it implies that parameter selection aimed at fitting the extremes (e.g., high phosphorus levels) is incompatible with the model structure that is designed to describe average system behavior.

2.3 Types and sources of uncertainty

The unpredictable difference between the model output and the observed data is often labeled uncertainty. This uncertainty can result from either natural variability, say caused by unpredictable rainfall, evapotranspiration, water consumption, and the like, and/or by both known and unknown errors in the input data, the model parameters, or the model itself. The later is sometimes called knowledge uncertainty but it isn't always due to a lack of knowledge. Models are always simplifications of reality and 'errors' can result. Some errors are random, others are systematic. Sometimes errors occur because of a lack of knowledge, such as just how a particular species will react to various environmental and other habitat conditions.

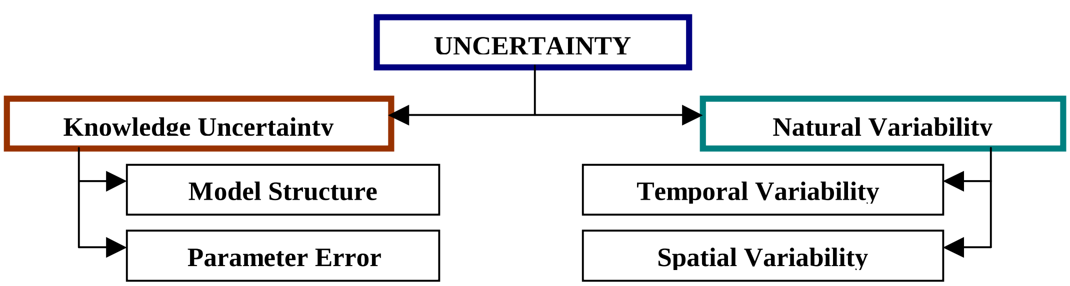


Figure 7. Types of uncertainty.

Imperfect representation of processes in a model constitutes model <u>structural error</u>, and imperfect knowledge of the values of parameters associated with these processes constitutes model parameter error. Natural variability includes both temporal variability and spatial variability, to which model input values may be subject. Figure 7 illustrates these different types of uncertainty. For example, the rainfall measured at a weather station within a particular model grid cell may be used as an input value for that cell, but the rainfall may actually vary at different points within that cell and may have a different mean. Knowledge uncertainty can be reduced through further measurement or improved models, whereas natural variability is a property of the natural system, and is usually not reducible at the scale being used.

Rather than pursue how to consider and handle 'knowledge' vs natural variability uncertainty it may be more helpful to classify uncertainty in another way, based on specific sources of uncertainty, and address ways of identifying and dealing with each source of uncertainty listed below.

Informational Uncertainties:

- errors in specifying the boundary and initial conditions at all prior times that impact the current output variable values
- errors in measuring observed output variable values

Structural Uncertainties:

- errors in model structure and parameter values
- variability of observed input and output values over a region smaller than the spatial scale of the model (e.g., topography within the 2x2 square mile area)
- variability of observed model input and output values within a time smaller than the temporal scale of the model. (e.g., rainfall and depths and flows within a day)
- errors in linking models of different spatial and temporal scales (e.g., the SFWMM and the ATLSS models)

Numerical Errors:

- errors in the model solution algorithm

Examples of the sources of uncertainty for the RECOVER models are presented in the Appendix.

First consider a situation, as shown in Figure 8, in which for a specific set of model inputs, the model outputs differ from the observed values, and the observed values, for those model inputs, are always the same. Here there is nothing random occurring and the model parameter values or model structure needs to be changed. This is typically done in a model calibration process.

Each of the models used by CERP is deterministic. Given specific inputs, the outputs are always going to be the same each time those inputs are simulated. If for specified inputs to any simulation model the predicted output does not agree with the observed value, as shown in Figure 8, this could result from measurement errors in both input and observed values. It could also result from errors in the model parameter values, the model structure, or the algorithm used to solve the model. None of these possible errors necessarily imply uncertainty exists. Errors do not necessarily translate to uncertainty, but the effect can be the same as if there were uncertainty.

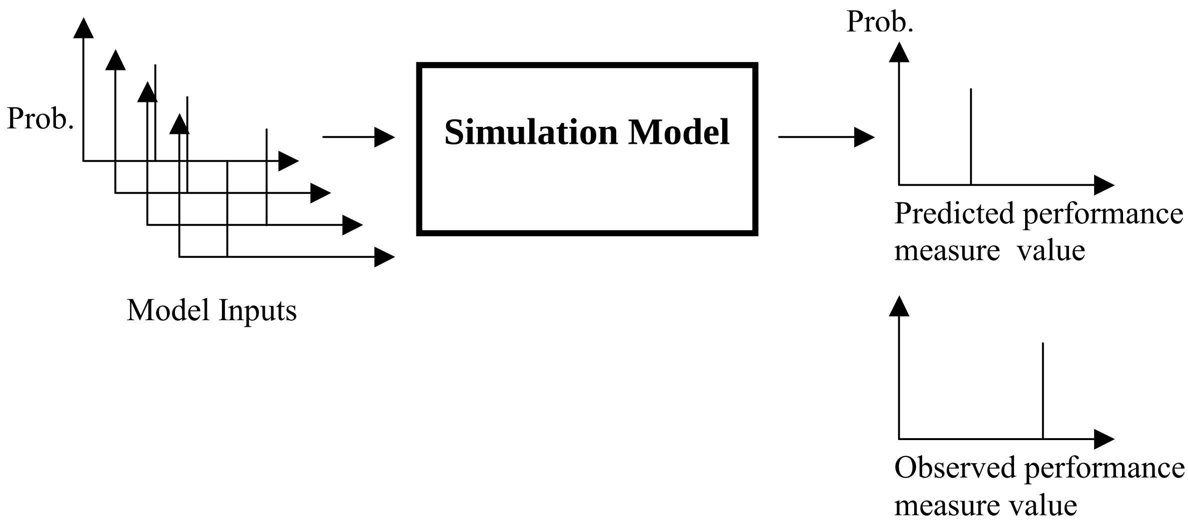


Figure 8. A deterministic system and a simulation model of that system needing calibration or modification in its structure. There is no uncertainty, only parameter value or model structure errors to be identified and corrected.

Next consider the same deterministic simulation model but now assume at least some of the inputs are random, i.e., not predictable, as may be case when random outputs of one model are used as inputs into the next model Random inputs will yield random outputs. Clearly the outputs of a deterministic model will not be predictable if what they are based on, i.e., the model inputs, are not predictable. While the model input and output values are not predictable, they can be described by probability distributions. If the uncertainty in the output is due only to the uncertainty in the input, the situation is similar to that shown in Figure 8. If the distribution of performance measure output values does not fit or is not identical to the distribution of observed performance measure values, then calibration of model parameter values or modification of model structure may be needed.

If a model calibration or 'identification' exercise finds the 'best' values of the parameters to be outside reasonable ranges of values based on scientific knowledge, then the model structure or algorithm might be in error. Assuming the algorithms used to solve each of the CERP models are correct and, as is the case, observed measurements of system performance vary for the same model inputs, as shown in Figure 9, it can be assumed that the model structure does not capture all the processes that are taking place that impact the value of the performance measures. This is often the case when relatively simple and low-resolution models are used to estimate the hydrological and ecological impacts of water and land management policies. However, it is worth noting that in the presence of informational uncertainties, there may be considerable uncertainty about the values of the "best" parameters during calibration.

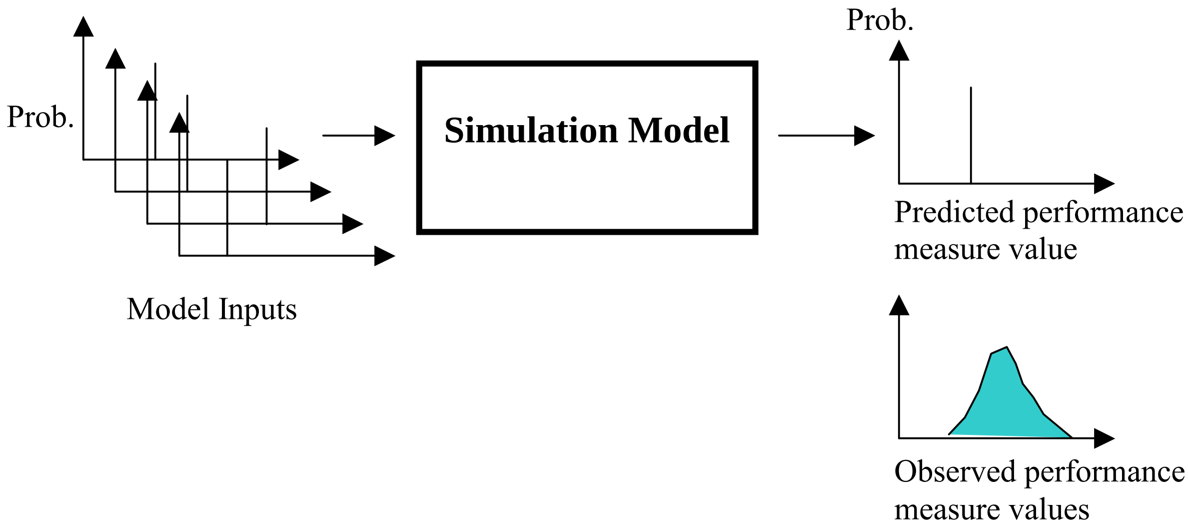


Figure 9. A deterministic simulation model of a 'random or stochastic' system. To account for the variability in observed results even given the same input values, the model's parameter values may need to vary over distributions of values and/or the model structure may need modification along with additional model inputs.

An example: Consider the prediction of a pollutant concentration at some site downstream of a pollutant discharge site. Given a streamflow Q (in units of 1000 m^3/day), the distance between the discharge site and the monitoring site, X (m), the pollutant decay rate constant k (day^{-1}), and the pollutant discharge W (Kg/day), we can use the following simplified model to predict the concentration of the pollutant C (g/m^3 = mg/l) at the downstream monitoring site:

$$C = (W/Q) \exp\{-k(X/U)\}$$

In the above equation assume the velocity U (m/day) is a known function of the streamflow Q.

In this case the actual value of the pollutant concentration C may vary unpredictably from the computed value of C even for the same inputs of W, Q, k, X, and U. This apparent uncertainty, as illustrated in Figure 7, can be simulated using the same model but by assuming a distribution of values for the decay rate constant k. Alternatively the model structure can be modified to include the impact of streamflow temperature T on the prediction of C.

$$C = (W/Q) \exp\{-k\theta^{T-20} (X/U)\}$$

Now there are two model parameters, the decay rate constant k and the dimensionless temperature correction factor θ, and an additional model input, the streamflow temperature, T. It could be that the variation in streamflow temperature was the sole cause of the first equation's 'uncertainty' and that the assumed parameter distribution of k was simply the result of the distribution of streamflow temperatures on the term $k\theta^{T-20}$.

If the output were still random given constant values of all the inputs, then another source of uncertainty exists. This uncertainty might be due to additional random loadings of the pollutant, possibly from non-point sources. Once again the model could be modified to include these additional loadings if they are knowable. Assuming these additional loadings are not known, a new random parameter could be added to the input variable W or to the right hand side of the equations above that would attempt to capture the impact on C of these additional loadings. A potential problem, however, might be the likely correlation between those additional loadings and the streamflow Q.

While adding model detail removed some of the uncertainty in the above example, increasing model complexity will not always eliminate or reduce model uncertainty. Adding complexity is generally not a good idea when the right equations are not known at the scale of application, and the amount of data for calibration is small compared to the number of parameters, and this seems to be the usual situation in RECOVER.

Even if more detailed models requiring more input data and more parameter values were to be developed, the likelihood of capturing all the processes occurring in a system as complex as the Everglades is not likely to happen, nor is it likely to be cost-effective to attempt such a task. Hence those involved in CERP and RECOVER will have to make decisions taking this inevitable uncertainty into account. The issues addressed in this workshop are how best to estimate this apparent uncertainty and then what to do with such information.

What is needed is a way to predict the variability evident in the system shown in Figure 9. Instead of a fixed output vector for each fixed input vector, a distribution of outputs are needed for each performance measure based on fixed inputs (Figure 9) or a distribution of inputs (Figure 10.). Furthermore the model output distribution for each performance measure must 'match' as much as possible the observed distribution of that performance measure.

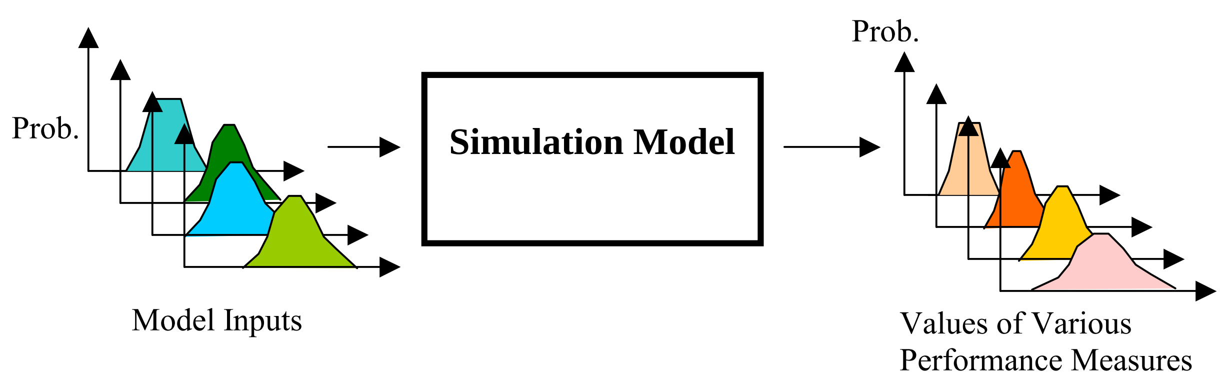


Figure 10. Simulating variable inputs to obtain probability distributions of predicted performance values that match the probability distributions of observed performance values.

Hydrologic Uncertainty

The use of a fixed 30+ year historic weather record for evaluating the effects of various water management alternatives in the models has both advantages and disadvantages. The primary advantage is that this historic record provides weather data that have all the appropriate spatial and temporal correlations built in. The data are realistic because they are real observed values. The disadvantage is that this historic period represents just one realization of what a 30+ year weather record might look like, and it is unlikely that the next 30+ year period will look just like it with respect to the weather. The overall temperature and precipitation means may be similar, for example, but the actual weather sequences and distributions may not. It is conceivable that the management alternatives may have somewhat different effects depending on the actual weather sequence over the next 30+ years. If it is desired to test the simulated effects of management alternatives on a range of possible weather for the next 30+ years, then the use of the same historic record cannot provide this.

There are two alternatives for how realistic statistical variations in 30+ year weather records might be provided for this type of evaluation. The first is to break the historic record into smaller pieces, say individual years or several year blocks, and then randomly recombine them in different sequences to provide a range of weather scenarios. This is somewhat analogous to the statistical technique of bootstrapping (Efron and Tibshirani, 1993), resampling from the existing samples to create a distribution of possible alternative samples. The advantage of this method is that all the temporal and spatial correlations within each block of data are retained. One disadvantage is that longer term (e.g., decadal, El Nino cycle, etc.) temporal patterns of weather may be broken up when the data are divided in this way. In addition, the range of weather is restricted to that observed in the historic 30+ year period, but there may be reasonable likelihood events which did not occur during that period (e.g., 50 year flood or drought), and it may be desired to examine how water management alternatives perform under such circumstances.

The second alternative is to use stochastic weather generators to provide future weather scenarios. A number of these weather generators are available, such as WGEN (Richardson and Wright, 1984), WXGEN (Nicks et al., 1990; Wallis and Griffiths, 1995), and CLIGEN (Nicks, 1995). Monthly statistical parameters are derived from long-term weather records for each weather station for input into the programs. These include statistical moments such as mean and variance of daily min and max temperatures, for example, as well as conditional probabilities such as the probability of rain given rain on the previous day, which describe the persistence of weather conditions. The advantages of this method are that any number of weather scenarios of any length can be produced, whose distribution mimics that of the long-term weather record from which the parameters were derived, and they can reflect a broader array of realizations than the use of a single historic record would allow. For example, scientists studying potential effects of climate change have used stochastic weather generators to produce weather scenarios with hypothesized shifts in temperature means and precipitation means, variances, intensities, etc. (Mearns et al., 1992; Wilks, 1992).

There are several disadvantages of stochastic weather generators, however. Some aspects of short-term temporal autocorrelation are incorporated in the algorithms. For example, the occurrence of precipitation is generally modeled as a first-order Markov process, where the probability of rain on a day is conditional on whether rain occurred the previous day. However, longer-term temporal autocorrelations (e.g., interannual) are generally not considered. In addition, the generators produce reasonable weather scenarios for a given site, but the spatial autocorrelation among sites across a landscape is not considered and unrealistic spatial patterns of weather may result.

Ecological uncertainty compared to hydrological uncertainty

Models used within the CERP include hydrological models (SFWMM and NSM), water quality models (LOWQM and to some extent ELM), and ecological models (ELM and ATLSS). All of these models are subject to the various types of uncertainty outlined earlier. However, some distinctions can be made between uncertainties in the hydrological and ecological models. The hydrological models are deterministic and each run with the same input values will give the same output. In contrast, some of the ATLSS models include demographic stochasticity. Probabilities of survival, reproduction, and movement of individuals at each time step may result in different outcomes (realizations) each time the model is run, even with identical input values. This type of stochasticity is an inherent property of populations needs to be included for the models to be realistic.

One other difference between the uncertainty in the ecological models and the hydrological models is that the latter provide input into the former. Since the ecological models use the output of hydrologic models as a starting point, any uncertainty in those boundary conditions will be propagated through the ecological models and combined with additional uncertainties within the ecological models. At present there is not a coupling of the models that would provide ecological feedback to the hydrological models (although this is done to some extent internally within ELM). If hydrologic changes in response to water management resulted in ecological changes, it is possible that these might have some impact on hydrology. For example, changes in vegetation such as expansion of *Typha* marshes could provide physical impediments affecting hydrologic flows, change important parameters such as Manning's roughness, or change rates of evapotranspiration. While these effects may be minor compared to effects of changing water control structures, they do contribute a degree of uncertainty that is difficult to evaluate at the present time.

Guidance on minimum criteria to put bounds on ecological model ELM

The ELM model includes a mixture of hydrological (e.g., stage, hydroperiod), water quality (e.g., P concentration), and ecological performance measures (e.g., areal extent of different vegetation types). By their nature, the hydrological and water quality performance measures have shorter term dynamics than the ecological performance measures, which are responding to these forcing functions on longer time scales. Changes in vegetation mosaics take place over multiple years, although the rate of some changes such as *Typha* invasion of sawgrass marshes can be quite rapid (Wu et al., 1997).

ELM is a spatially distributed model with a "unit model" or "general ecosystem model" (Fitz, et al. 1995, Fitz and Sklar 1999) for each grid cell across the landscape (typically 1 km^2 but the resolution may be varied). The unit model is run for each cell using inputs for that cell, and the cells are connected by flows to and from adjacent cells (Figure 11). The model has been calibrated for a variety of hydrological, water quality, and ecological performance measures in one basin (WCA-2A) and across the entire Everglades region. A sensitivity analysis has been performed on an early version of the model (Fitz et al., 1995).

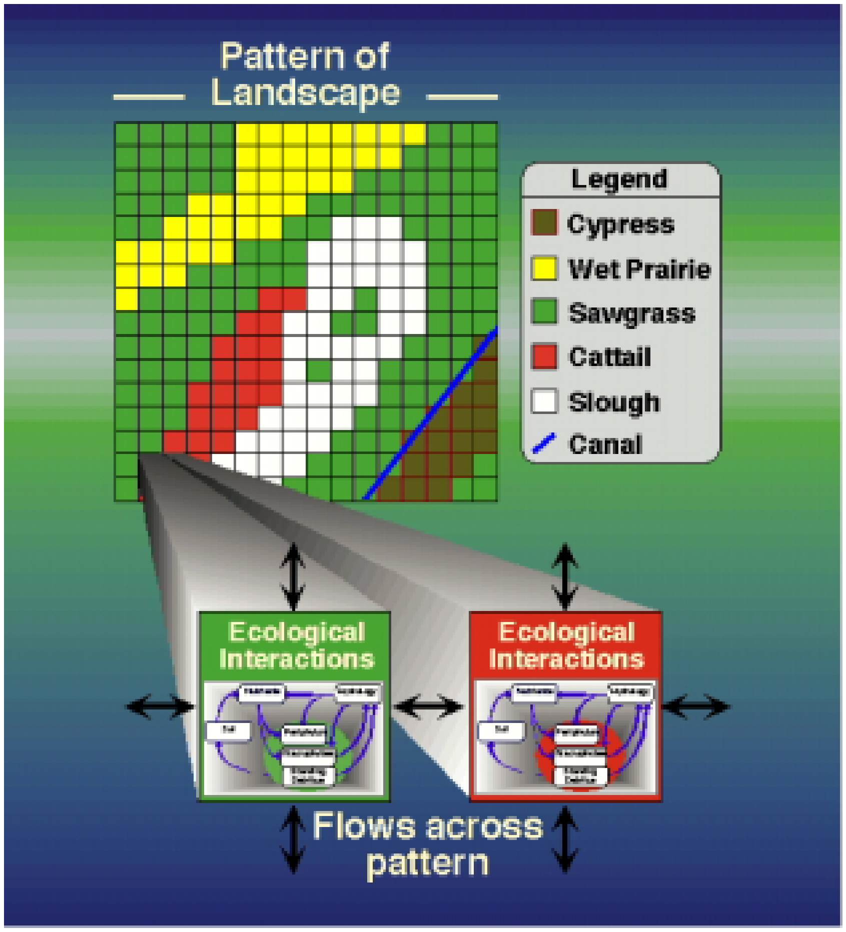


Figure 11. Spatially distributed ELM model, showing unit ecosystem models for each cell, connected by flows between cells.

Suggested next steps - If the model has changed substantially since the earlier version on which a sensitivity analysis was performed, this analysis might be repeated to identify the most sensitive parameters in the current version. This should include both habitat specific parameters for the unit model, and fluxes and variables that substantially affect fluxes between cells ("landscape drivers" – Fitz et al., 1995). While an uncertainty analysis incorporating the effect of parameter and input uncertainty on model output uncertainty for the entire region may be prohibitive due to the large number of interacting cells and associated parameter and input data sets, one approach would be to at least do this for the unit model. For a number of selected individual cells representing different habitat types, a Monte Carlo simulation could be performed using distributions of parameter and input data values that represent the degree of their uncertainty, at least for the ones identified as the most important in the sensitivity analyses. In the first step, this would not include uncertainty in specified flows across cell boundaries since this in itself depends on unit model uncertainty. In this way the variability of the model outputs in response to parameter and input variable uncertainty could be quantified. Part of these outputs includes flows across cell boundaries. This uncertainty could then be added and the Monte Carlo simulation repeated to incorporate the effect of these important landscape drivers on unit model uncertainty.

2.4 What uncertainty analysis can provide

An uncertainty analysis takes a set of randomly chosen input values (that can include parameter values), passes them through a model (or transfer function) to obtain the distributions (or statistical measures of the distributions) of the resulting outputs. As illustrated in Figure 12, the output distributions can be used to

- Describe the range of potential outputs of the system
- Estimate the probability that the output will exceed a specific threshold or performance measure target value.

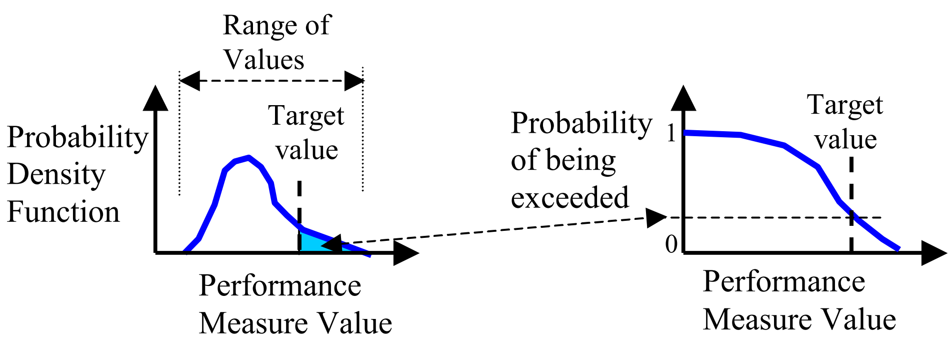


Figure 12. The distribution of performance measures defines range of potential values and the likelihood that any specified target value will be exceeded. The shaded area in the density distribution on the left represents the probability that the target value will be exceeded. This probability is shown in the probability of exceedance plot on the right.

The most common uses for uncertainty analyses are to make general inferences, such as the following:

- Estimating the mean value and the standard deviation of the outputs,
- Estimating the probability the performance measure will exceed a specific threshold.
- Putting a confidence interval on a function of the outputs, e.g., the range of function values that 90% likely to occur.
- Describing the range and likelihood of potential outputs of the system.

Implicit in any uncertainty analysis are the assumptions that statistical distributions for the input values are correct and that the model is a sufficiently realistic description of the processes taking place in the system. Neither of these assumptions is likely to be entirely correct.

The importance of first assumption is easy to check by using different distributions for the input parameters. If the outputs vary significantly, then the output is sensitive to the specification of the input distributions and hence they must be defined with care. A straight forward deterministic sensitivity analysis can be of value here. A sensitivity index similar to that proposed by McCuen (1973) can be used to measures the magnitude of change in an output variable Q per unit change in the magnitude of an input parameter value P from its base value Po. Letting the index i represent a decrease and j represent an increase in the parameter value from its base value Po, the sensitivity index SIpq for parameter P and output variable Q is

$$SI_{pq} = \max \{ | (Qo - Qi) / (Po - Pi) | , | (Qo - Qj) / (Po - Pj) | \}$$

Alternatively the index could measure the elasticity of parameter change, i.e., the maximum percent magnitude of change in an output variable per percent change in the magnitude of a parameter value from its base value.

$$SI_{pq} = \max \{ | [(Qo - Qi)/ Qo] / [(Po - Pi)/Po] | , | [(Qo - Qj)/ Qo] / [(Po - Pj)/Po] | \}$$

When the base or initial parameter value Po is zero a modified sensitivity index can be introduced (Nearing et al. 1990).

$$\text{mod } SI_{pq} = \{ | (Qi - Qj) | / (Pi - Pj) | \} \quad \text{or} \quad \{ | (Qi - Qj)/ Qij | / | (Pi - Pj) / Pij | \}$$

where Qij is the average of Qi and Qj, and Pij is the average of Pi and Pj.

The assumption of a correct model choice can be supported two ways. The first way is to obtain consensus among interested parties that the most appropriate or best model has been chosen. The second way is to develop and apply alternative models in a sensitivity analysis. Formal statistical approaches for handling multiple conceptual models are rarely used because they require the development of an entire suite of plausible conceptual models.

In the setting of the RECOVER models, a challenge in performing an uncertainty analysis of the sort described above is that information and parameter uncertainties are difficult to describe reliably through simple probability distributions considering the large number of parameters, the spatial correlation in inputs and inferred parameters. An uncertainty analysis is not the same as a sensitivity analysis. An uncertainty analysis attempts to describe the entire set of possible outcomes, together with their associated probabilities of occurrence. A sensitivity analysis attempts to determine the relative change in model output values given small changes in model input values. A sensitivity analysis thus measures the change in the model output in a localized region of the space of inputs. However, one can often use the same set of model runs for both uncertainty analyses and sensitivity analyses. It is possible to carry out a sensitivity analysis of the model around a current solution and then use it as part of a first order uncertainty analysis, but it may not be practical.

2.5 Performance Indicator or Performance Measure Values

Performance measures are functions of model outputs. Having quantified some of the uncertainty associated with model outputs, these distributions of output variable values can be used to obtain the distributions associated with any performance measure. The process or procedure for doing this may depend on the performance measure. This is illustrated in Figure 13.

To illustrate this consider an example involving the calculation of the probability distribution of a performance measure Z. Two independent random variables X and Y are required to define a performance indicator Z in each of a series of time periods t (t=1, 2, …, T). The different values of X and Y and their probabilities for each time period t are derived from an uncertainty analysis. For this illustrative example, the performance measure Z is the minimum of the sums X + Y over all T time periods. In other words it is a little more complicated than just the sum of X and Y in each period t.

Assume the values of each possible X and Y are discrete integers. From a simulation together with uncertainty analyses the values of each discrete random variable X and Y together with their probabilities are as listed in Table 1.

Table 1. Values of model outputs x_t and y_t together with their probabilities.

Time t	x_{1t} (PX_{1t})	x_{2t} (PX_{2t})	x_{3t} (PX_{3t})	y_{1t} (PY_{1t})	y_{2t} (PY_{2t})	y_{3t} (PY_{3t})
1	4 (0.2)	5 (0.5)	6 (0.3)	3 (0.1)	6 (0.7)	8 (0.2)
2	2 (0.3)	4 (0.6)	5 (0.1)	4 (0.2)	5 (0.5)	7 (0.3)
3	1 (0.1)	3 (0.5)	4 (0.4)	2 (0.1)	3 (0.4)	6 (0.5)

Now one can compute the probabilities of various possible sums of $x_t + y_t$ in each period t, as shown in Table 2.

Table 2. Possible values of sums of $x_t + y_t$ and their joint probabilities.

Time period t=1

$x_t + y_t$	7	8	9	10	11	12	13	14
$Pr(x_t + y_t)$	0.02	0.05	0.03	0.14	0.35	0.25	0.10	0.06

Time period t=2

$x_t + y_t$	6	7	8	9	10	11	12
$Pr(x_t + y_t)$	0.06	0.15	0.12	0.41	0.05	0.18	0.03

Time period t=3

$x_t + y_t$	3	4	5	6	7	9	10
$Pr(x_t + y_t)$	0.01	0.04	0.05	0.25	0.21	0.25	0.20

Finally computing the distribution of the minimum sum considering all the ways it could occur and their associated probabilities.

Min $\{x_t + y_t\}$	3	4	5	6	7	8	9	10
Pr ($\min_t \{x_t + y_t\}$)	0.01	0.04	0.05	0.28	0.26	0.10	0.22	0.04

This defines the probability distribution (a histogram in this discrete case) of the performance measure Z. This process for continuous distributions is illustrated in Figure 10.

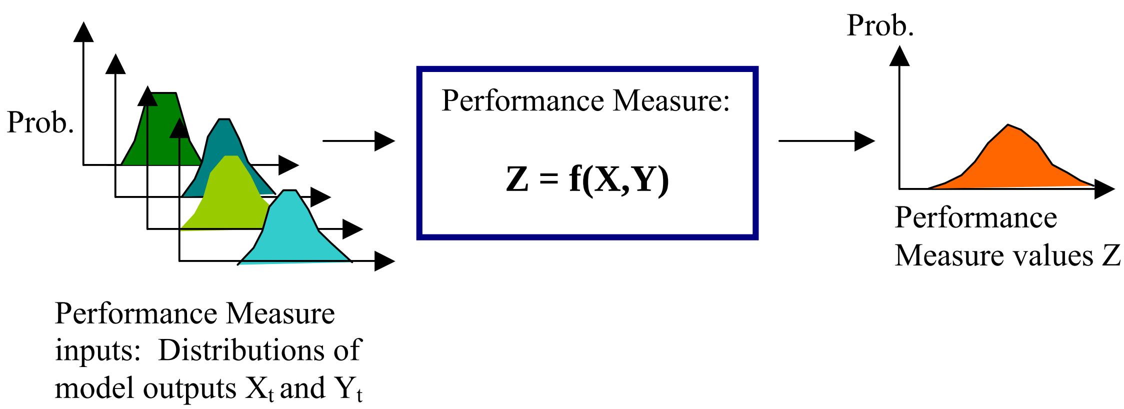


Figure 13. Calculating the distribution of a performance measure from the distributions of its input variables, in this case a time series of X_t and Y_t variable values. The transformation process of converting input distributions to performance measure distributions may differ for different performance measures.

2.6 Performance Measure Targets

Another possible source of uncertainty is the selection of a target value for a performance measure. Just what target value is correct? When this is not clear, various ways exist of expressing the uncertainty associated with any target value. One such method is the use of fuzzy sets. Use of 'grey' numbers or intervals instead of 'white' or fixed target values is another. When some uncertainty or disagreement exists over the selection of the best target value for a particular performance measure it seems to us the most direct and transparent way to do this is to subjectively assign confidence bands about a range of possible target values. Then this subjective distribution can be factored into the tradeoff analysis, as outlined in Figure 14.

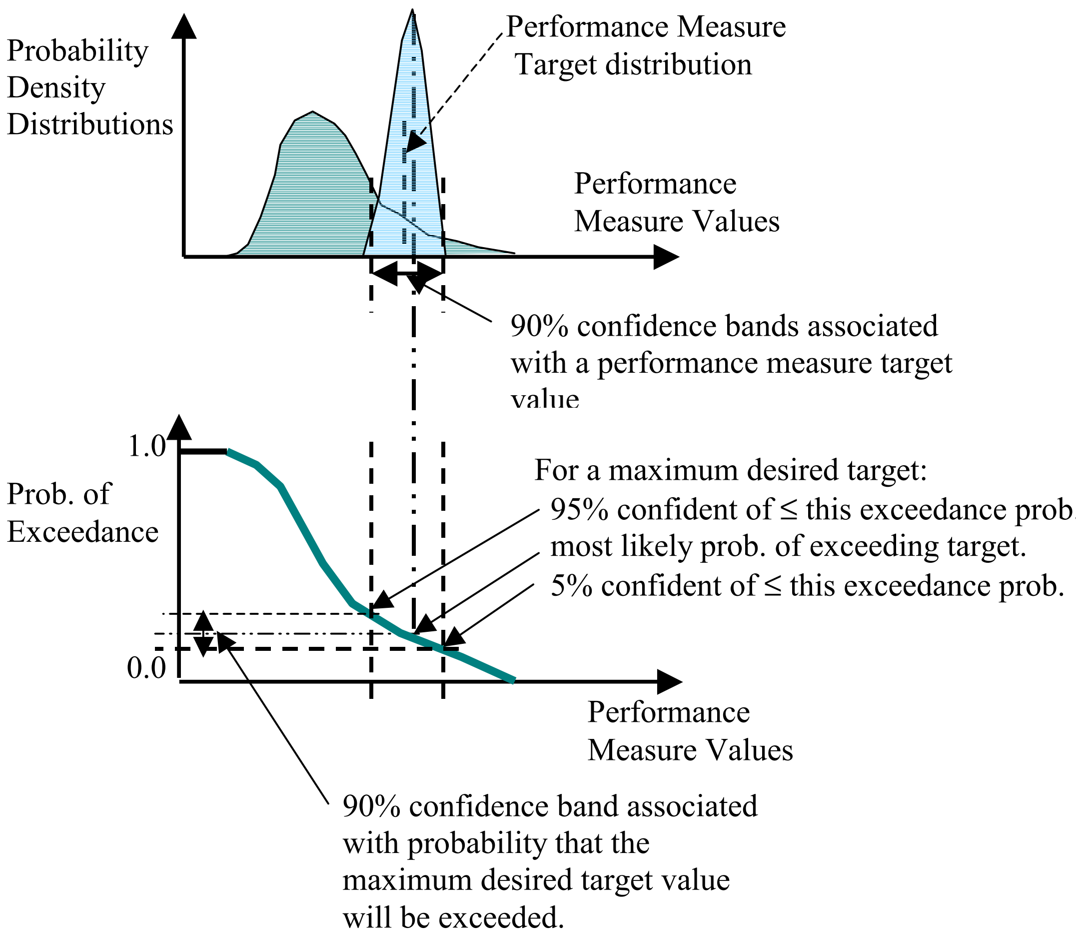


Figure 14. Combining the probability distribution of performance measure values with the probability distribution of performance measure target values to estimate the confidence one has in the probability of exceeding a maximum desired target value.

One of the challenges associated with defining and including in an uncertainty analysis the uncertainty associated with a target or threshold value for a performance measure is that of communicating just what the result of such an analysis means. Referring to Figure 14, suppose the target value represents some maximum limit of a pollutant, say phosphorus, concentration in the flow during a given period of time at a given site or region, and it is not certain just what that maximum limit should be. Subjectively defining the distribution of that maximum limit, and considering that uncertainty along with the uncertainty (probability of exceedance function) of pollutant concentrations – the performance measure – one can attach a confidence to any probability of exceeding the maximum desired concentration value.

The 95% confidence probability of exceedance shown on Figure 14, say $P_{0.95}$, should be interpreted as “we can be 95% confident that the probability of the maximum desired pollutant concentration being exceeded will be no greater than $P_{0.95}$.” We can only be 5% confident that the probability of exceeding the desired maximum concentration will be no greater than the lower $P_{0.05}$ value. Depending on whether the middle line through the subjective distribution of target values in Figure 14 represents the most likely or mean target value, the associated probability of exceedance is either the most likely, as indicated in Figure 14, or that for which we are only 50% confident.

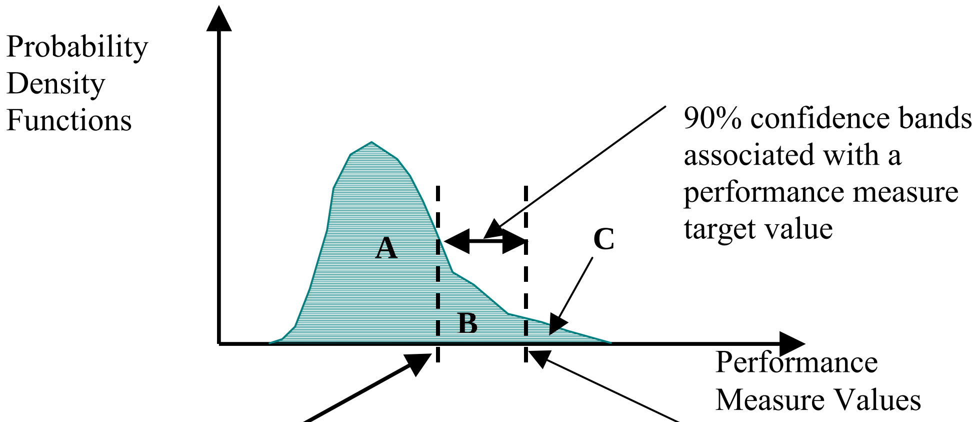
Figure 15 attempts to show how to interpret the confidence bounds when the uncertain performance targets are

- minimum acceptable levels that are to be maximized,
- maximum acceptable levels that are to be minimized or
- optimum levels.

An example of a minimum acceptable target level might be the population of wading birds in an area of the Everglades. An example of a maximum acceptable target level might be, again, the phosphorus concentration the flow in a specific region of the Everglades. An example of an optimum target level might be the depth of water most suitable for tree island development and maintenance during a period of the year.

For performance measure targets that are not expressed as minimum or maximum limits but that are the ‘best’ values, one can, referring to Figure 12, state that they are 90% confident that the probability of achieving the desired target is no more than B. The 90% confident level probability of not achieving the desired target at least A+C. The probability of the performance measure being too low is at least A and the probability of the performance measure being too high is at least C, again at the 90% confidence level. As the confidence level decreases the band width decreases, and the probability, at that lower confidence level, of not meeting the target, increases.

Now, clearly there is uncertainty associated with each of these uncertainty estimations, and this raises the question of just how valuable is the quantification of the uncertainty of each additional component of the plan evaluation process. Will plan evaluators and decision makers benefit from this additional information, and just how much additional uncertainty information is useful?



For a performance measure that is to be maximized:

5% confident that the probability of meeting or exceeding the desired target is at least B+C.

5% confident that the probability of not meeting or exceeding the desired target is less than A.

For a performance measure that is to be minimized:

95% confident that the probability of exceeding a desired target is no greater than B+C.

95% confident that the probability of not exceeding the desired target is at least A.

For a performance measure that is to be maximized:

95% confident that the probability of meeting or exceeding the desired target is at least C.

95% confident that the probability of not meeting or exceeding the desired target is less than A+B.

For a performance measure that is to be minimized:

5% confident that the probability of exceeding a desired target is no greater than C.

5% confident that the probability of not exceeding the desired target is at least A+B.

Figure 15. Interpreting the results of combining performance measure probabilities with performance measure target probabilities depends on the type of performance measure. The letters A, B, and C represent proportions of the probability density function of performance measure values. (A+B+C = 1.)

2.7 Spatially distributed data

Spatially distributed models such as those used in the RECOVER project require a variety of model inputs for each of the model cells. Examples include topography, rainfall, evapotranspiration, and others. Supporting measurements for some of these variables may be available at spatial resolutions higher or lower than that of the model cells, or from irregularly distributed measurement points. To be used as model inputs, these need to be transformed to the spatial resolution of the model grid.

When the spatial resolution of the supporting data is higher than that of the model cells, the data must be resampled to the grid cell resolution. For example, topographic maps or digital elevation models (DEMs) may be used to provide elevations for numerous points within a model cell. Averaging the elevations for these points to compute a mean elevation for the cell is one simple way to accomplish this scale transformation. The degree of random error associated with these cell means may be calculated as the standard error of the measurements, which is equal to the standard deviation divided by the square root of the number of measurements. Thus random error may be reduced by increasing the number of supporting data points included in the calculation of cell means. However, systematic error or bias, due for example to subsidence since the time when the elevations were measured, is not necessarily reduced as sample size within a cell increases.

Often the supporting data are available only from an irregular network and require spatial interpolation to the model cells. Precipitation and potential evapotranspiration (PET) are examples of such variables. This interpolation involves uncertainty. Certain spatial interpolation procedures such as kriging provide standard deviations of the interpolated estimates as a measure of this uncertainty (Isaaks and Srivastava, 1989). When interpolation of values for grid points is the objective, ordinary kriging is the appropriate technique, while block kriging may be used when interpolating values to be applied to entire grid cells (Deutsch and Journel, 1992).

Evapotranspiration has been identified as one of the model parameters to which SFWMM and NSM water level performance measures are most sensitive (Lal, 1995; Trimble, 1995). The procedure has been to compute potential evapotranspiration for each weather station in the model domain, and then interpolate to the model grid by inverse distance interpolation (Ken Tarboton, personal communication). Thus, PET is a model input, but within the model PET is converted to AET (actual evapotranspiration) by the Penman-Monteith equation using evapotranspiration crop coefficients and Manning’s roughness coefficients appropriate for each grid cell’s vegetation (Ken Tarboton, personal communication). Therefore, the uncertainty of AET calculations depends both on model input uncertainty (PET) and model parameter uncertainty (crop and roughness coefficients). Specification of probability distributions of AET for model uncertainty analysis should include uncertainty from both sources. While the uncertainty of crop coefficients and roughness coefficients may be similar in all model grid cells, the uncertainty of interpolated PET values depends on the number and proximity of the supporting data points for which PET is known and varies spatially (Phillips and Marks, 1996). The following procedure is recommended to determine the spatially explicit uncertainty of PET model input values for model uncertainty analyses:

1. Calculate PET using the Penman-Monteith equation for each weather station in the model domain.
2. Interpolate PET to each model grid cell using kriging.
3. Repeat steps 1 and 2 for several time periods to determine average PET kriging standard deviations for each grid cell. (Since the proximity of the supporting data points does not change from one time period to another, it is anticipated that the magnitude and spatial pattern of kriging standard deviations may be reasonably constant over time.)
4. For each grid cell, compute the combined variance of the PET interpolation and the Penman-Monteith coefficients using the first order Taylor series expansion (including the covariances among the coefficients if known).
5. For each grid cell, use these variances and the calculated AET to characterize a normal probability distribution for AET for model uncertainty analyses.

2.8 Spatial-temporal scale issues in linking models

A spatial-temporal model contains variables whose values vary over space and time. The value of the information obtained from such models comes from their simultaneous depiction of temporal and spatial patterns. The influence of scale, i.e. the mismatches between the spatial and temporal resolutions of the input data compared to that of the output data, the mismatches between the times and places calibration data were obtained compared to those of the predicted data, data measurement and collection errors, methods such as kriging used for interpolation and extrapolation over space and time, and the quality of any analysis used to produce model input data, all contribute to the uncertainty of model output results. These uncertainties become important when one tries to develop and use spatial-temporal models for predicting hydrologic and ecologic performance over space and time. Lal (2000) discusses the effect of spatial and temporal resolution on hydrologic model output error. Sklar and Hunsaker (2001) provide an overview of the use and uncertainties of spatial-temporal ecological modeling issues applied to the Everglades. Here we merely highlight some of their insights.

Uncertainty in landscape models that predict ecological impacts over time and space resides largely within four components of model structure. These components are the inputs – the spatial and temporal interpolation of point data, the initial and boundary conditions, and in the calibration and verification procedures. Sources of uncertainty in spatial data are similar to the sources of uncertainty in temporal data – both are associated with data measurement and collection, data processing, model structure, natural variability, and human intervention.

Costanza and Sklar (1895) show the relation between model complexity and model uncertainty. Simpler models focused on relatively narrow questions can be made more accurate – less uncertain – than more complex models encompassing a broader range of outputs. They define the effectiveness of a model as a function of how much it attempts to explain (complexity) and how well it explains what was attempted (accuracy). They found there is a balance between complexity and accuracy – or uncertainty – that makes the model most effective. Spatial-temporal models tend to be very complex, and hence very uncertain and it is not always obvious where the sources of uncertainty are in specific situations. One way to assess the uncertainty of spatial-temporal models is to measure and collect more data. Nothing surprising is in this conclusion. With more data model calibration becomes more precise. Weaknesses in model structure might also become apparent. For such models, the calibration process should be an iterative one of model reconstruction to simplify temporal and spatial complexity.

The uncertainties of verification are the same as those for calibration except that verification errors increase with time. Predicted future events gets more and more uncertain as the difference in time between now and those future events increases. The validity of spatial-temporal models will diminish as prediction time increases because uncertainty is a cumulative component of models.

The Everglades landscape model (ELM) is a spatial-temporal model of nutrient cycling, plant growth, and hydrology (Fitz et al. 1996; Fitz and Sklar 1999). It is a complex process-based model for water management. It divides the Everglades into 10,000 1x1 km grid cells and computes the hydrology and the fate of plant and algae communities in response to nutrients, water, sunlight, and temperature at those grid cell corners. This spatial resolution is smaller than the outputs of water depths, flows, and hydroperiods from SFWMM that has a spatial resolution of 2x2 miles. Part of ELM's complexity is the feedback of the plant community on hydrology.

Efforts to measure and minimize landscape model uncertainties have been discussed by Colwell (1974, O'Neill and Gardner (1979), Turner et al. (1989), Ludweg et al. (1993), Lemons (1996), Anderson (1998) and Breckling and Dong (2000). They identify a number of ways to address model uncertainty. To reduce ignorance as a source of uncertainty, it is common to develop an experimental supplement for data collection and modeling. These usually involve sensor and model sensitivity tests both in the field and in the laboratory. To reduce model uncertainty associated with natural variability, it is common to develop a stochastic modeling approach for simulating likely variation in system states. For simple models, dealing with the uncertainty of chaotic events may be as knowable as stochastic events (Gleick 1987). Developing a rigorous quality control and assurance program best controls human error.

2.9 Computation Schemes for Uncertainty Analyses

A number of computation techniques have been used in uncertainty analyses for modeling projects. These include

- **Analytical Solution**: When there are few stochastic input parameters, and when the model is not too complicated, one can sometimes obtain an analytic form for the output statistical distribution. With a hydrologic system as complex as the Everglades where the transport of water and contaminants is over a large surface area constantly interacting with the groundwater aquifer, this approach becomes infeasible.

- **First-Order Analyses**: As outlined in Loucks and Stedinger (1994) and as implemented for some of the water management models used by SFWMD by Trimble (1995) and Lal (1995), this method is a way to obtain the variances of the parameter values in a model in order to obtain an estimate of the relative contribution of each parameter to the uncertainty in the model outputs. This approach is best suited to situations when the computed outputs respond linearly to parameter changes and the correlations among parameters is small.

- **Stochastic Numerical Models**: Developing and solving stochastic models using numerical methods is another approach to uncertainty analyses. For example stochastic differential equations defining a water quantity- quality or ecological model can include random parameter values as well as random variables reflecting limitations in model structure. Actual applications of these stochastic numerical models have been for relatively simple systems compared to the Everglades. It is unlikely the development of specialized numerical algorithms to handle the solution of stochastic differential equations instead of deterministic ones in the existing CERP models will be worth the considerable effort. The other disadvantage of this approach is that it is not very transparent to those unfamiliar with such methods.

- **Monte Carlo with Random Sampling**: This approach assumes that some of the model input variables and model parameters are random. Distributions of these input variables and parameters are defined using expert judgment and by calibration exercises. Once defined, the actual values of these random variables and parameters are drawn from these distributions for each simulation time step. Many simulations are performed, each using values drawn from these probability distributions, to generate a distribution of output values. After simulating a considerable number of time steps, one has a set of equally likely outputs that define a probability distribution for each selected performance measure. Although conceptually simple, a shortcoming is that many simulations may be required to obtain a satisfactory description of the output distribution.

  The Monte Carlo method is illustrated in Figure 16. Monte Carlo simulation involves multiple runs of the simulation model for randomly chosen input values. The result is a set of output values whose probability distribution can be defined and characterized, such as shown in Figure 9. To illustrate this method, consider a situation where only one input variable is random and only one output variable is to be determined. Let this random input variable be X. This random input variable X has a cumulative distribution function, $F_X(x)$, as shown in Figure 16. A value of this random variable X is needed for each simulation. For a particular value, x, of the random input variable, X, a particular value of the output variable, say y, will be generated by the simulation model. Many values of y define the probability distribution of the output variable Y, which is of interest. The resulting probability distribution of the random output variable Y based on all the simulated output values y, is valid only if the probability distribution of the set of input values $\{\mathbf{x}\}$ drawn from the probability distribution of the input variable X, is the same as the probability distribution of the random input variable X. So the challenge is to randomly select values of the random input variable X that will correspond to the probability distribution of X.

  There are some computer programs that can generate values x from any arbitrary probability distribution of a random variable X. However numerous computer programs exist that will generate random variable values that are equally likely over a range of values, say from 0 to 1. In other words, these values are drawn from a uniform probability distribution such as shown in Figure 4. Call these randomly generated values p, and since there are many of them, each will be designated as p_j.

These randomly generated p_j values can represent the value of the cumulative distribution function of the random input variable X. For each randomly generated p_j there is an x_j such that $F_X(x_j)$ equals p_j. The probability distribution of these values x_j will correspond to the probability distribution of the random variable X.

Each randomly generated input value x_j is simulated to obtain an output variable value y_j. Of interest of course is the distribution of the set of all y_j values. This is the distribution of the random output variable Y. This process is shown in Figure 16.

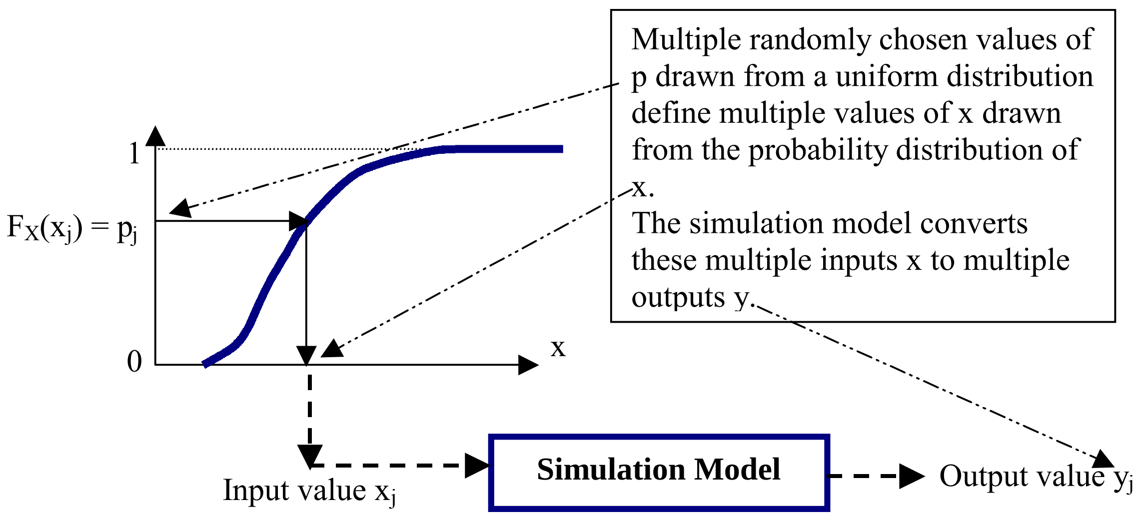


Figure 16. Monte Carlo sampling to obtain a set of input variable values x_j that are used to define the set of output variable values y_j.

Randomly selecting values of p_j from a uniform distribution and finding the x_j that equates the cumulative distribution $F_X(x_j)$ to p_j will insure that the set of x_j will be selected from the probability distribution of the input variable X. This is done for many equally likely random values of p_j to get sufficient number of output values y_j to define the probability distribution of the output random variable Y.

- **Bayesian Monte Carlo (BMC) analysis.** This approach estimates model uncertainty by combining prior information regarding the uncertainty of model inputs with the ability of different parameter input sets to describe available data on state variables. (The following paragraphs on BMC were from Dilks and James, 2002.)

Variants to Monte Carlo analysis have been developed that use Bayesian inference to generate improved estimates of parameter uncertainty by considering the ability of different parameter values to describe observed data (Spear and Hornberger, 1980; Fedra, 1983; Dilks et al, 1992). The approach starts out like traditional Monte Carlo analyses through the specification of statistical distributions for each uncertain model parameter.

These distributions are based upon prior knowledge of the variability in each parameter, literature reviews, professional judgment, etc. As with traditional Monte Carlo analyses, numerous iterations are performed with model inputs randomly selected from their pre-specified distributions. The unique aspect of Bayesian Monte Carlo analyses is that the results of each simulation are compared to field observations of model state variables, and "scored" with respect to the ability of a given parameter set to describe the observed data.

The original applications of this methodology (e.g. Spear and Hornberger, 1980; Fedra, 1983) used a simple yes/no scoring function; either a model simulation adequately described the observed data or it did not. Dilks et al (1992) took the methodology to its Bayesian roots, and used the statistical likelihood function to weight the acceptability of any parameter set. Performing sufficient simulations and tracking each result can generate an n-dimensional matrix generated describing the marginal uncertainty in each parameter as well as the entire error covariance structure. This covariance matrix can be used to define the uncertainty in model projections or investigate the contribution to overall uncertainty contributed by each parameter.

The first step in determining overall parameter uncertainty for the improved Monte Carlo analysis is to specify what is known about parameter distributions, prior to considering the ability of parameter values to describe observed data. The approach taken for specifying prior parameter distributions depends upon the type and quantity of data available. Relatively robust data sets allow for formalized statistical approaches and goodness of fit testing. Sparse data sets require more subjective judgment.

Computational issues

In spite of the simplicity of a Monte Carlo analyses, there can be are many potential problems in performing such analyses. One is the computer time required to obtain a sufficient number of output values to define their distributions. This is especially relevant here where the hydrologic simulation models (SFWMM and NSM) require hours for each simulation. Another is the problem of possibly obtaining infeasible combinations of input variable values, especially if the input variables are not independent. Because of these potential problems specialized sampling methods have been proposed.

Several specialized sampling techniques have been developed to reduce the number of simulations required in a Monte Carlo analysis to obtain a satisfactory description of the output distribution. One of the techniques, called Latin Hypercube Sampling (Iman and Conover 1982), has proven to be successful for hydrologic systems. The general approach is the same as for Monte Carlo modeling with random sampling, but the specific values of the input parameters are chosen differently. They are chosen from the same probability distributions, but the sampling scheme spreads the values in such a way as to reduce sampling variability. The probability distribution of each input variable is divided into segments of equal probability and one observation is randomly drawn from each range. This technique is outlined in Loucks and Stedinger (1994). It is likely some modification of this approach will be the most practical approach for carrying out uncertainty analyses for RECOVER, at least in the short run.

Design Constraints

There are a number of areas where design constraints must be carefully addressed. Three major design considerations are discussed briefly in this section: computation time, data storage, and use of stochastic inputs.

- **Computation Tim**e: Monte Carlo analyses producing a distribution of model outputs require many more simulations than do analyses where only one value for each output variable is needed in each simulation time step. A simulation model may be run 100 to 10000 times or more than it would otherwise, and hence computation time is a major concern. Design tradeoffs to comply with a time budget include choices on complexity of models, computer systems, and the use of statistical (black box) models or parallel computing techniques.

  Many are finding value in replacing relatively complex time-consuming process-based or conceptual simulation models with multivariate statistical models, including artificial neural network (ANN) models that once developed and calibrated (or trained) are much faster to run and can perform multiple simulations in only a fraction of the time required by the process-based model. ANNs are simply more complex regression equations, and, like regression equations, are best suited for interpolation within the range of data they have been calibrated for rather than for extrapolation beyond that range. For uncertainty analyses the development and use of statistical models such as ANNs is worthy of serious consideration. The development of stochastic ANNs involving distributions for some of the weights, might be needed, and certainly would be interesting research. These black-box (statistical) methods require data for training and this seems may be a problem over large parts of the Everglades.

- **Data Storage and Assimilation**: Results from the uncertainty analysis will be needed for a substantial number of hydrologic and ecologic performance measures, for a variety of geographic locations, for many time periods, and for a variety of water management policies. The large number of results can result in very large data storage requirements. If disk storage is limited design tradeoffs to comply with a disk storage budget can include limiting the number of locations and times steps at which impacts will be calculated and designing model repeatability. If results are repeatable, they may be recalculated if desired (rather than being computed once and stored). The biggest constraint will be the time required to evaluate all this information and to develop tradeoffs such as suggested in Figure 5.

- **Calibrating random parameter values:** The so-called inverse problem of finding the best parameter values that convert model inputs to realistic model outputs is difficult enough when just one value is assumed for each unknown model parameter. It is substantially more difficult when the parameter value for each simulation is to be selected from a distribution of possible values.

The calibration of parameter value distributions involves finding the distributions of the random parameter values that result in the best fit between the computed or predicted and observed distributions. Ways of doing this are discussed in the literature. For example, Warwick and Cale (1987) and Jaffe et al. (1988) present the calibration of the probability distributions of parameter values for water quality models with measurement errors. Warwick and Cale calibrated the probability distribution of the de-oxygenation rate constant of BOD for use in Monte Carlo simulations. Jaffe et al. discussed a heuristic calibration technique to obtain the probability distribution of the parameter values.

In a water quality modeling study using random parameter values carried out at Cornell University the genetic algorithms search method was used to find the means and standard deviations of three (assumed) normally distributed parameters (Lopez, 1999). Calibration was conducted sequentially for the mean and the variance of the parameters. First, the mean parameters were obtained from calibrating the deterministic model against the mean of the outputs. After calibrating the mean, the coefficients of variation were obtained. This two-step calibration procedure was chosen for two reasons:

> i) the mean of the stochastic models corresponds to the deterministic models' analytical solutions and these values are available. Using the analytical solutions, the search algorithm worked faster and was more accurate than numerical solution procedures.
>
> ii) numerical solution of the stochastic models was the best approach to estimate the variance of the outputs since it avoided the complex mathematics required for analytical solutions. The numerical solution was based on simultaneous multiple random paths of all the variables that must be simulated together.

2.10 Recommended Methods for Uncertainty Analyses

It seems reasonable to consider the application of uncertainty analysis to the District's water models in terms of short-term and long-term strategies. In the short-term, it must be recognized that the District's models are not conducive to complete uncertainty analysis given current computing technology. Thus strategies should be proposed for:

(1) conducting an informative, but incomplete uncertainty analysis, and
(2) using that incomplete uncertainty analysis to improve decision making.

In the long-term, recommendations are made to:

(1) structure the models so that a more complete uncertainty analysis is feasible, and
(2) employ Bayesian approaches that are compatible with adaptive assessment.

There are different ways to list, or identify, the uncertain variables or parameters in a simulation model. For example, we can say that errors, and therefore uncertainties, may exist in:

- Model parameters
- Model structure (equations)

- Model inputs (e.g., rainfall, temperature, pollutant loads)
- Initial conditions and boundary conditions

To further complicate matters, model parameters are often correlated (that is, there is frequently a relationship between selected pairs of parameters); this correlation should be included in the uncertainty analysis, as failure to include parameter correlation has been shown to significantly affect overall model prediction error.

At present, it does not seem likely that the District's models have the appropriate model structure and a sufficient observational database to support a complete uncertainty analysis (i.e., one that includes all of the sources of uncertainty (or errors) identified above). Of particular concern is the lack of observed data that would permit a complete parameter estimation using optimization (e.g., least squares or maximum likelihood), and thus judgmental parameter selection is necessary. At present, it is not clear that this judgment will be able to produce reasonable estimates of parameter correlation or model structure error.

Short-Term Strategy

If knowledge, data, or model structure prevents uncertainty analysis from being complete, is there any value in conducting uncertainty analysis? Stated another way, is it reasonable that decision making will be improved with even partial information on uncertainties, in comparison to current practice with no or very limited reporting of prediction uncertainties? Often, but not always, the answer is "yes," although the usefulness of incomplete uncertainty characterization, like the analysis itself, is limited. Some examples are illustrative.

From a scientific perspective, analysis of uncertainty yields insight on research needs, although uncertain terms that are ignored obviously cannot be assessed. Sensitivity analysis may be used to identify uncertain model terms that have a relatively large impact on predictions and are thought to be amenable to better quantification through research and/or monitoring. However, sensitivity analysis for one model parameter or input variable at a time, holding all others constant, can be misleading. For example, if two parameters are highly correlated such that each parameter individually is poorly known, but collectively their impact on predictions is well known, then single parameter sensitivity analysis can be misleading. This may be the case for correlated parameters such as the *maximum growth rate* and *half saturation constant* in growth kinetics included in water quality or ecologic models.

Using decision analysis as a prescriptive model, we know that uncertainty analysis improves decision making when prediction uncertainty is integrated with the utility (or loss, damage, net benefits) function to allow decision makers to maximize expected utility (or maximize net benefits). When uncertainty analysis is incomplete - and perhaps more likely, the utility function is poorly characterized if at all – the concepts of decision analysis still provide a useful guide.

Triangular distributions could be assessed for all uncertain model terms, assuming that correlation is negligible, and then sampling (e.g., using a modified Latin hypercube approach) could be used to simulate the prediction uncertainty. The result of this computation could be either over/under estimation of error, but it provides some indication of the magnitude of the uncertainty. However, this information alone, while perhaps helpful for research and monitoring, does not by itself aid decision-making. Estimates of prediction uncertainty need to be considered in conjunction with the attitudes toward risk on the performance measures.

For example, do the decision makers (or stakeholders, or other affected individuals/groups) want to particularly avoid loss of critical species? Are decision makers risk-averse with respect to ecological damage, such that they are willing to increase project costs in order to avoid species loss or to reduce the probability of species loss? This tradeoff, between species loss probability and cost, is not clear with deterministic (point) predictions of the performance measures, but it does become evident with a probability distribution (reflecting prediction uncertainty) on the performance measure of interest, as illustrated in Figure 5.

Steps in Short-term Uncertainty Analyses

Loucks and Stedinger (1994) outlined a series of procedures for sensitivity and uncertainty analyses for RECOVER models. The program outlined included several progressive steps for determining the model parameters to which model performance measures were most sensitive, and for propagating the uncertainties associated with model parameters. However, while complete uncertainty analyses, quantifying all sources of uncertainty and how they propagate within a model, is desirable to do (Reckhow, 2002), this is unlikely to be feasible in the near term for models used in RECOVER. This is due both to the large number of model input variables and parameters and to the amount of effort needed to make multiple model runs, which are very computationally intensive.

Some work as suggested by Loucks and Stedinger (1994) has been done. Trimble (1995) performed a sensitivity analysis for the SFWMM model and a first order uncertainty analysis for selected performance measures and parameters. Lal (1994, 1995) and Lal, Obeysekera and van Zee (1997) also performed a sensitivity analysis for the NSM model, and uncertainty analyses using first order, Rosenblueth, and Latin hypercube Monte Carlo simulation methods for selected performance measures and key parameters.

In the short term, the continuation and extension of SFWMM uncertainty analyses methods used by Trimble (1995), Lal (1994, 1995) and Lal, Obeysekera and van Zee (1997) may be in order. Several approaches are possible.

It seems to us that in the short term it is important to have some way to derive estimates of the probability distributions of uncertain output variables in a very practical and transparent way, without requiring an enormous amount of work and time. More precise methods might produce more precise estimates, and that such methods could be pursued in the future should better estimates of uncertainty prove critical to decision making.

Our recommended uncertainty analysis in the short run involves the following steps:

1. Identify those indicators or features of the Everglades ecosystem that will best characterize the condition of the ecosystem and the progress toward its restoration. These should include those aspects that the public is concerned about (e.g., alligators, fish and

birds) as well as other indicators or features that scientists believe are important (e.g., periphyton and phosphorus concentrations, and density and condition of tree islands).

2. Determine the hydrological attributes (that can and are to be managed) that impact these indicators and features of the ecosystem. This identifies just which hydrological attributes are most important at specific sites in the Everglades. There is little value in defining other hydrological attributes as performance indicators or measures if they do not influence any important indicator or feature in the ecosystem.

3. Using expert judgment, define the functional relationships between these hydrological attributes and the condition of these important indicators and features of the ecosystem. These functional relationships between hydrological attributes and the ecosystem indicators or features are useful as they identify the ranges of those attributes that are most influential and potentially worthy of further uncertainty analyses. The relative condition of the ecosystem indicators or features might range from 0 (the worst condition) to 1 (the best condition) as shown in Figure 17. These functional relationships may themselves be uncertain and this uncertainty can be included in the uncertainty analysis, as will be shown later.

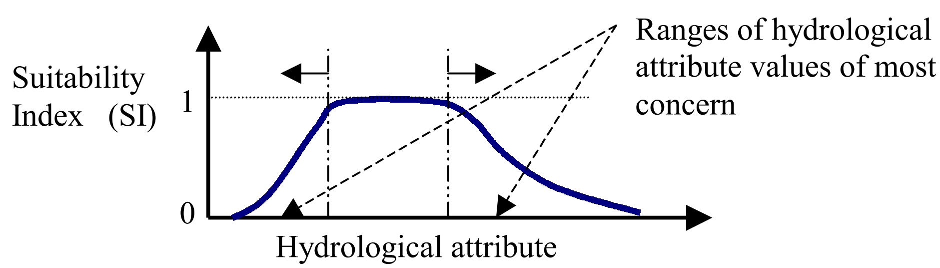
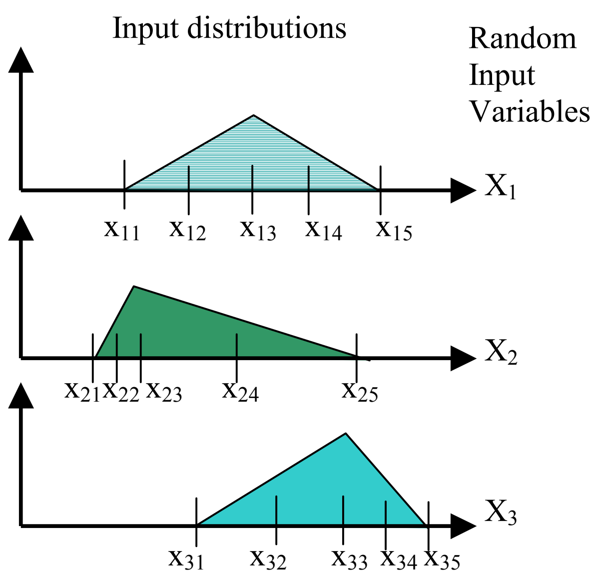


Figure 17. A function defining the relative suitability of a specified hydrological attribute for a specified ecosystem indicator or feature.

4. Perform a sensitivity analysis to determine just which input variables impact the hydrologic attributes the most over the ranges of those attributes of most interest, i.e., where they do not achieve optimal conditions for the appropriate ecosystem indicators or features. The values of these input variables should be independent of each other. Each selected independent input variable could serve to represent other input variables that are correlated with or dependent on them. It is feasible to consider only the most sensitive and independent input variables in this analysis.

   In highly parameterized models, performing sensitivity analyses to identify out of the hundreds of parameters a few of the most important ones can involve a considerable effort. To add to the difficulty, changes in some parameters may affect multiple model output variable or performance indicator values. For example, a parameter that controls evapotranspiration may also affect surface water flow and hydroperiod because both are dependent on the magnitude of surface water volume that is impacted by evapotranspiration. Clearly expert judgment will play an important role in this effort.

5. Identify the range of values these independent and significant input variables are likely to assume. For the input parameters that are characterized by a relatively limited number of values from the scientific literature, either uniform or triangular prior distributions can be used. Uniform distributions may be appropriate for those parameters where it is judged that the available information can only describe only a minimum and maximum value. Triangular distributions can be used for those parameters where it is possible to identify a "most likely" value as well a minimum and maximum value (Dilks and James, 2002). A triangular probability density distribution is shown in Figure 18.



Assuming low values of each input variable produce low values of the hydrological attribute value, and so on, the input variable values for the five simulations are:

1. x_{11}, x_{21}, x_{31}
2. x_{12}, x_{22}, x_{32}
3. x_{13}, x_{23}, x_{33}
4. x_{14}, x_{24}, x_{34}
5. x_{15}, x_{25}, x_{35}

Figure 18. Assumed triangular input distributions based on maximum and minimum values and the most likely value for three random input variables. (If better estimates of the input distributions are known, use them.) These three plus the midpoints between the limits and the most likely value are the five values from each distribution used in five simulations to generate five values for each output variable.

6. Perform simulations based on the selection of endpoints of each distribution, the most likely values, and points half way between the most likely and endpoints on either side of the most likely value, as shown in Figure 18. The combinations of input values from each independent input variable are selected so as to range from what would produce the lowest values of the hydrological attribute to the highest values of the hydrological attribute. This represents 5 separate simulations.

7. From the outputs of these 5 separate simulations, construct approximate cumulative probability distributions. For the five output variable values of say a random output variable Y, the smallest being 'y_1', the next smallest being 'y_2', the most likely being 'y_3' the next highest being 'y_4' and the highest being 'y_5', assume $F_Y(y_1) = 0$, $F_Y(y_2) = 0.15$, $F_Y(y_3) = 0.5$, $F_Y(y_4) = 0.85$, and $F_Y(y_5) = 1.0$. This is illustrated in Figure 19.

8. Compute the probability distribution of the hydrological attributes from these model output variable distributions if these output variables are not the hydrological attribute values themselves

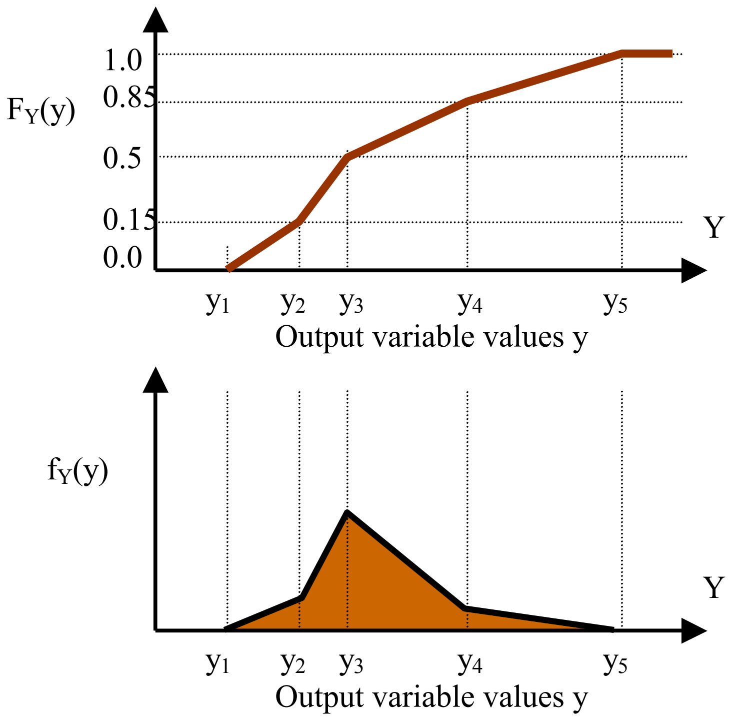


Figure 19. Derived estimated probability distributions of a single output variable.

9. Use these estimated and approximate probability distributions of the hydrological attributes together with the estimated confidence intervals associated with the functions identified in step 2 above for defining the ecosystem performance indices to make decisions and to compare different project alternatives. Figure 20 shows one approach for interpreting this uncertainty information.

This combines subjective probabilities (associated with the SI function) and approximate objective and/or subjective probabilities associated with the input variable values. The selection of a SI threshold such as 0.8 is arbitrary, but whatever value is selected it can be used as a basis to compare different project alternatives.

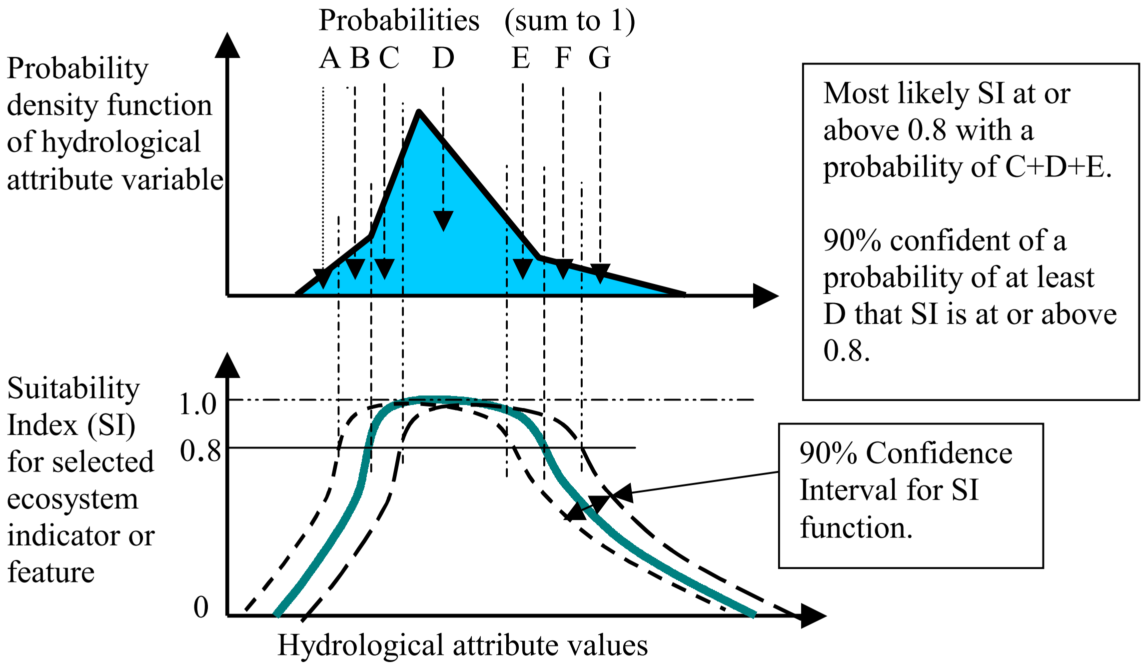


Figure 20. Conclusions drawn from this approximate uncertainty analysis for a SI value of 0.8. The confidence interval associated with the suitability index function may stem from different judgments of different experts, but the level of confidence remains a judgment.

To summarize, the steps involve

- selecting the significant independent input variables (including model parameters) that contribute most significantly to the final model prediction,
- constructing probability density functions for each parameter to reflect the likelihood that the selected variable will take on various values within its possible range,
- propagating the uncertainties through the model to generate a probability density distributions of predicted output values and subsequently of the hydrologic attributes that impact important ecosystem indicators and features,
- deriving, if desired, confidence limits associated with the functions that convert hydrological attributes to indices indicating the relative suitability of conditions for the ecosystem indicators or features,
- using these confidence intervals plus the derived probability density distributions of the hydrological attributes to make quantitative statements about the probabilities (and the confidence in these probabilities) of meeting selected suitability index levels for each index.

In theory this procedure must be carried out for each project alternative that is simulated. In practice, however, it would seem that as these analyses are being carried out it should become obvious just when a new uncertainty analysis needs to be done, and when the alternative being simulated will have approximately the same levels of uncertainty as some previously analyzed alternative.

Long-Term Strategy

In the longer term, some of the shortcomings of previous work and the short term procedures outlined above could be addressed and rectified. Two considerations in particular warrant more attention: more complete uncertainty analyses including the incorporation of the covariance structure of model inputs and parameters. In the very long term a whole set of newer hydrologic-ecologic models specifically designed to include uncertainty analysis may be worth considering.

Covariance – Some of the initial uncertainty analyses necessarily make simplifying assumptions about the independence of the model parameters and input variables used in the analysis, in the lack of knowledge to the contrary. While this does provide an initial look at error propagation through a model, falsely assuming independence may lead to either overestimates or underestimates of model uncertainty. Negative correlations among variables that are added or multiplied together, or positive correlations among variables that are subtracted or divided by one another can lead to a reduction in the uncertainty of the output compared to the results under the assumption of independence. This is true regardless of the uncertainty analysis method, e.g., first order uncertainty analysis or Monte Carlo simulation.

When Monte Carlo simulation is the method of choice, Latin hypercube sampling leads to much greater efficiency of multivariate space and a much smaller number of Monte Carlo simulation runs need be performed compared to simple random sampling (reference). Previous reports on sensitivity and uncertainty analysis of RECOVER models indicated that Latin hypercube sampling required the assumption of independence among the variables, but that is not the case. For example, the PRISM Monte Carlo simulation software uses the method of Iman & Conover (1982) to impose a specified correlation structure upon the Latin hypercube samples drawn from the multivariate parameter distribution (Gardner et al., 1983).

Correlations among parameters in an uncertainty analysis may be derived from expert opinion, inference from previous studies, or from observational data. Whenever the sources of data for pairwise correlations in a correlation (or covariance) matrix are different for different pairs of variables, this may result in a matrix in which the entries are not entirely consistent with each other. If the matrix is not positive definite, this is an indication of this type of problem. A solution to this problem is given by Iman and Davenport (1982), who describe a procedure to determine a correlation matrix that approximates the pairwise correlations.

If the parameters in an uncertainty analysis are assumed to be independent, but in actuality they are not, then many of the combinations of parameter values selected for the Monte Carlo simulations may not be feasible. For example, if two parameters are highly positively correlated, and a realization of values for a Monte Carlo simulation run has an extremely high value of one parameter and an extremely low value of the other parameter, this combination would have a very small probability of occurrence. As a result of this, many of the Monte Carlo simulation runs may result in model outputs that are unrealistic. If parameter independence is assumed, Spear and Hornberger (1980) suggest categorizing the simulation runs into reasonable and unreasonable outcomes. The unreasonable outcome runs are ignored, while the reasonable outcome runs are examined to infer the correlation structure of the parameters. This correlation structure may then be used in subsequent uncertainty analyses.

In the long run, the best strategy may be to restructure the models, emphasizing the development of models that are compatible with the need for error propagation and adaptive assessment / management. While they require considerable effort, Bayesian (probability) networks are particularly suitable for this task (Reckhow 1999, Borsuk, Stow and Reckhow 2000, Steinberg, Rekhow and Wolpert 1996, Jensen, 2001).

Bayesian Inference

Bayes Theorem lies at the heart of Bayesian inference (Gelman et al. 1995, Berry 1996); it is based on the use of probability to express knowledge and the combining of probabilities to characterize the advancement of knowledge. The simple, logical expression of Bayes Theorem stipulates that, when combining information, the resultant (or posterior) probability is proportional to the product of the probability reflecting *à priori* knowledge (the prior probability) and the probability representing newly acquired knowledge (the sample information, or likelihood). Expressed more formally, Bayes Theorem states that the probability for y conditional on experimental outcome x (written p(y | x)) is proportional to the probability of y before the experiment (written p(y)) times the probabilistic outcome of the experiment (written p(x | y)):

$$p(y \mid x) \propto p(x \mid y)\, p(y)$$

For example, suppose an ecologist is interested in the reduction in chlorophyll a in a lake associated with a 30% reduction in phosphorus concentration. The ecologist could use existing data from similar lakes to develop a simple chlorophyll – phosphorus regression model and predict the reduction in chlorophyll for the lake of interest. Alternatively, the ecologist could conduct dilution experiments on the lake, collecting new data to estimate the quantity of interest. Adopting a third option, a Bayesian ecologist would use "the best of both worlds" by combining the estimates using Bayes Theorem. In the language of Bayes Theorem, the regression model would yield the prior probability, since this estimator exists prior to the collection of new data, and the posterior probability would represent the revised estimate based on both prior knowledge and new experimental evidence (Reckhow, Clements, and Dodd. 1990).

At first glance, it seems hard to argue against this seemingly rational quantitative strategy for updating scientific knowledge. Indeed, one might ask why all ecologists aren't Bayesians? There are a number of reasons. Certainly the most important is that virtually all ecologists still learn probability and statistics from a classical, or frequentist, perspective, and Bayes Theorem is at best a minor topic within that curriculum. Beyond that, Bayesian inference has been widely regarded as subjective and thus not suitable for objective scientific analysis. The problem with this perspective is that most science is hardly the objective pursuit that many choose to believe.

Consider the judgments we make in a scientific analysis. Implicit (or explicit) in the ecologist's lake study on phosphorus and chlorophyll are judgments about the adequacy of the existing lakes data, the merits of dilution experiments, and the truth of the model relating phosphorus to chlorophyll. There are no purely scientific, absolutely correct, choices here; these represent "gray areas" about which reasonable scientists would disagree. Yet, these also represent judgments that must be made by the ecologist in order to carry out the lake study. Ordinary scientists are not unique in their reliance on judgment, however. Press and Tanur (2001) examine the scientific methods in the work of some of the most distinguished scientists (e.g., Galileo, Newton, Darwin, Einstein) in history, noting the substantial role of subjectivity in their work.

Further, consider how most scientists address the revision of scientific knowledge in light of their own new contributions. In some cases, the scientist simply states the conclusions of his work, making no attempt to quantitatively integrate new findings with existing knowledge. When integration is attempted in the concluding section of a research paper, it is typically a descriptive subjective assessment of the implication of the new knowledge. Bayesian inference has the potential to make combining evidence more analytically rigorous. It is ironic that the subjectivity of Bayesian analysis would be its undoing.

Bayesian analysis can be understood and applied in science in several different "ways." In perhaps the most fundamental sense, Bayesian inference provides a probability-based, normative approach for updating scientific knowledge on the basis of new information. Less comprehensive but more common, Bayesian analysis in mathematical models supports parameter estimation and yields predictive distributions for quantities of interest. For inferential statements, Bayesian statistical analysis, unlike classical or frequentist statistical analysis, is compatible with the statements that scientists are inclined to make following research, as they directly relate to the quantity/parameter of interest (e.g., a reaction rate), rather than to the value of a test statistic (e.g., a t-statistic).

2.11 Uncertain Uncertainty Considerations

If it is not possible to apply formal Monte Carlo analysis in a reasonable manner because not enough is known about the underlying relationships, it may still be possible to provide a quantitative or semi-quantitative rating of the uncertainty issues that may be useful in some contexts.

The following summary, adapted from guidance currently being used by the Intergovernmental Panel on Climate Change, shows one possible such method for rating the confidence in the underlying science.

1. For each of the major findings to be developed, **identify the most important factors and uncertainties that are likely to affect the conclusion**s. Also specify which important factors/variables are being treated exogenously or fixed, as it will almost always be the case that some important components will be treated in this way when addressing complex phenomena.

2. **Document ranges and distributions found in the literatur**e, including sources of information on the key causes of uncertainty. Note that it is important to consider the types

of evidence available to support a finding, such as distinguishing findings that are well established through observations and tested theory from those that are not so well established.

3. Given the nature of the uncertainties and the state of science, **make an initial determination of the appropriate level of precision—**is the state of science such that only qualitative estimates are possible, or is quantification possible, and if so, to how many significant digits? As the assessment proceeds, recalibrate the level of precision in response to your assessment of new information.

4. Quantitatively or qualitatively **characterize the distribution of values that a parameter, variable, or outcome may tak**e. First identify the end points of the range, and/or any high consequence, low probability outcomes or "outliers." Particular care needs to be taken to specify what portion of the range is included in the estimate (for example, this is a 90% confidence interval) and what the range is based on. Then provide an assessment of the general shape (for example, uniform, bell, bimodal, skewed, symmetric) of the distribution. Finally, provide an assessment of the central tendency of the distribution (if appropriate).

5. **Rate and describe the state of scientific information** upon which the conclusions and/or estimates (that is from Step 4) are based.

6. **Prepare a "traceable account"** of how the estimates were constructed that describes the reasons for adopting a particular probability distribution, including important lines of evidence used, standards of evidence applied, approaches to combining/reconciling multiple lines of evidence, and critical uncertainties.

7. **Use formal probabilistic frameworks for assessing expert judgment** (decision-analytic techniques), as appropriate.

In describing the state of scientific information, the following descriptors may be appropriate:

- **Well-Established**: models incorporate known processes; observations consistent with models, or multiple lines of evidence support the finding.

- **Established but Incomplete**: models incorporate most known processes, although some parameterizations may not be well tested; observations are somewhat consistent but incomplete, current empirical estimates are well founded, but the possibility of changes in governing processes over time is considerable, or only one or a few lines of evidence support the finding.

- **Competing Explanations**: different model representations account for different aspects of observations or evidence, or incorporate different aspects of key processes, leading to competing explanations.

- **Speculative**: conceptually plausible ideas that have not received much attention in the literature or that are laced with difficult to reduce uncertainties.

3. Selecting and Ranking Performance Measures

The Comprehensive Everglades Restoration Plan (CERP) seeks to restore hydrological and ecological functions in the South Florida region closer to their pre-drainage conditions. A large number of hydrologic performance measures have been identified by the RECOVER project as metrics of progress toward these restoration goals (Mills, 2002). These performance measures have targets which are set in various ways, including from runs of the Natural Systems Model (NSM), which simulates pre-drainage hydrology without any water control structures in place (RECOVER, 2001).

Hydrologic performance measures generated by the South Florida Water Management Model (SFWMM) number in the hundreds. These have been reduced to a hundred key performance measures (Mills, 2002) or fewer (RECOVER, 2001) for examination and decision making about management alternatives. However, operationally it is difficult to optimize a variety of measures at the same time. There will always be tradeoffs between alternatives that result in good performance in some measures and fair or poor performance in others, versus other alternatives that result in the opposite trends. It seems to us RECOVER should strive for:

1. further reduction of the number of performance measures used to evaluate water management alternatives, and
2. ways to collectively evaluate performance on a number of measures in order to have a clear-cut comparison between competing alternatives.

One way in which the number of hydrologic performance measures might be further reduced is to use the ecological models (e.g., ELM, ATLSS) or at least habitat suitability functions to determine which measures are of greatest importance to ecological processes and wildlife populations simulated in these models. While operationally CERP focuses on hydrologic restoration, ecological endpoints are of great concern and hydrologic restoration is viewed as the best way to effect ecological restoration as well (Appelbaum, 2002). Hence, knowledge gained from the ecological models about which hydrologic performance measures are most important ecologically can be used to focus efforts on those measures.

Even with a smaller number of hydrologic performance measures selected, the problem remains of how to choose between management alternatives with different results for the suite of performance measures. A quantitative objective function is needed to decide between competing management alternatives. One way to approach this would be to assign relative importance values for each of the performance measures and to compute an objective function which weights the results on these measures by their importance. The results for each performance measure could be normalized to its target value on a percentage basis. For example, for a performance measure in which higher values are better than lower values, with a target value representing a minimum desired level:

$$\text{relative performance} = \begin{cases} 100\% * (\text{performance} / \text{target value}) & \text{if performance} < \text{target} \\ 100\% & \text{if performance} \geq \text{target} \end{cases}$$

For a performance measure in which lower values are better, with a target value representing a maximum acceptable level:

relative performance =

100% * (target value / performance)   if performance > target
100%   if performance ≤ target

An objective function of overall performance for a management scenario could then be computed as a linear combination of the importance-weighted relative performances (scaled from 0 to 1):

overall performance =

Σ [relative importance of performance indicator * relative performance indicator value]

where: Σ relative importance = 1

In both examples above the relative performance indicator value is capped at 100%, so that each performance measure is scaled from 0 to 100%. (This restriction could be relaxed if it was desired to acknowledge or consider achievement of performance beyond the target value. In this case, relative performances of 50% and 150% on two performance measures would be viewed as favorably as relative performances of 100% on both, for example.)

This procedure presupposes assignment of quantitative importance values to each of the performance measures. Habitat suitability indices as functions of hydrological attributes provide this quantification. If ecological models (e.g., ELM, ATLSS) are used to determine which hydrologic performance measures are most important, it may be possible to devise a quantitative index of this importance. For example, if a number of different simulation runs were performed with different values of hydrologic performance measures as inputs, then the proportion of variance in an ecological performance measure (e.g., large wading bird populations, tree island area, etc.) that is accounted for by each hydrologic performance measure can be used to gauge its importance:

$$PM_i \; importance = \frac{\mathrm{var}\, accounted\;\; for\, by\, PM_i}{\sum_i \mathrm{var}\, accounted\;\; for\, by\, PM_i}$$

Alternatively, a Delphi approach may be used to assign the importance values to each of the hydrologic performance measures from expert opinion.

If the view is taken that it is not possible to assign importance values to performance measures, or that all performance measures are equally important, then some other way of judging overall performance for a management alternative is needed. In this instance, the objective function might be defined as:

overall performance = no. of performance measures meeting target values

Whichever alternative resulted in achieving the greatest number of target values would be preferred.

3.1 Statistical significance of differences among alternatives

When model uncertainty is taken into account it is reasonable to ask how one can distinguish whether model outcomes (performance measure values based on different management decisions) are significantly different from each other. While this question may be important to address, first there must be informed judgment about how large a difference is of any practical significance. If two outcomes are so close as to be of no practical difference, then it matters little whether one is statistically significantly different from the other. In evaluating two management alternatives in such a case, the alternative with better performance on other measures, or the one that is less costly would be preferred regardless of the statistical significance of a negligibly small difference.

The difference between two alternative management scenario model outcomes cannot be judged statistically when there is only a single result for each. Statistical comparison involves comparing multiple samples from the distributions of two different populations or samples. If model uncertainty analyses provided estimates of the width of 95% confidence bands around each model result, then a comparison could be made by determining if the confidence bands of each alternative overlapped the mean of the other alternative. This is analogous to a statistical t-test for differences between two populations.

However, such a test may be overly conservative in establishing statistical significance in the difference between the two management alternatives. The reason is that this type of test assumes that the samples from the two populations are independent (uncorrelated). This may or may not be the case. It is quite possible that values of the uncertain model parameters and inputs that lead to a performance measure value at the upper end of one alternative, would also lead to a value near the upper end of the other alternative, and similarly for values near the middle and lower ends of the confidence bands. So, it is possible that the confidence bands for the two alternatives might overlap extensively, suggesting no statistically significant difference, when for every case of model parameter and input values one alternative would consistently give higher results. When there is a logical pairing of samples (run results) from two populations (management alternative scenarios), the usual statistical procedure is to calculate the difference between each pair and test whether it is different from zero (as in a paired t-test).

The following procedures are recommended when a determination of statistically different model outcomes on a performance measure must be made between two management scenarios:

- Decide upon the smallest level of practical significance in model outcomes. If results from alternative management scenarios are less than this, ignore the question of statistical significance for this performance measure and choose between the alternatives based on other performance measures or cost.
- If generalized confidence bands are available for the model outcomes from uncertainty analyses, a conservative test for significant differences may be made by visual inspection

of whether the confidence bands of each alternative overlap the mean of the other alternative.

- If possible, perform a Monte Carlo uncertainty analysis for both management alternatives. Pair the results for both alternatives based on identical model parameter and input values. Test for significant differences by the Fisher sign test or Wilcoxon signed rank test (non-parametric analogs of the paired t-test; Hollander and Wolfe, 1973).

4. Communicating Uncertainty

I (we?) do not know of one best way to communicate concepts of uncertainty to the public. The best way may well depend on what the public, or members of the public, already know about risk and the various types of probability distributions (e.g., density, cumulative, exceedance) based on objective and subjective data, and the distinction between mean or average values and the most likely values. However, we believe graphical representations of these ways of describing uncertainty considerably facilitates communication.

The National Research Council report Science and Judgment in Risk Assessment (NRC 1994) addressed the extensive uncertainty and variability associated with estimating risk and concluded that risk characterizations should not be reduced to a single number or even to a range of numbers intended to portray uncertainty. Instead, the report recommended managers and the interested public should be given risk characterizations that are both qualitative and quantitative and both verbal and mathematical.

In some cases communicating qualitative information about uncertainty to stakeholders and the public in general may be more effective than quantitative information. There are, of course, situations in which quantitative uncertainty analyses are likely to provide information that is useful in a decision-making process. How else can tradeoffs such as illustrated in Figure 1 be identified? In addition quantitative uncertainty analysis often can be used to improve qualitative information about uncertainty, even if the quantitative information is not what is communicated to the public.

When presenting the results of any modeling applied to the Everglades, it should not be too difficult for the public to realize that, whether or not such models explicitly included uncertainty, model predictions will not necessarily correspond to what may be observed. There is no way to accurately predict any particular system performance measure. First of all, the models being used are relatively simple compared to reality. Secondly, many features of the Everglades are variable, like the weather, and this variability can be measured. Thirdly, we just do not know enough about the processes that take place in the Everglades to eliminate the possibility of surprises. To account for variability it is possible to identify and measure the range of possible outcomes and their probabilities or likelihood of occurring. Accounting for ignorance is much more difficult. One has to expect that events or outcomes will occur that were not foreseen, and even if they could be identified in advance, there is no way to estimate their likelihood of any of them happening in advance of their occurrence.

One should acknowledge to the public the widespread confusion regarding the differences between variability and uncertainty. Variability does not change through further measurement or study, although better sampling can improve our knowledge about variability. Uncertainty reflects gaps in information about scientifically observable phenomena. Uncertainty sometimes can be reduced through further measurement or study and then including this increased understanding within the simulation model.

Support for routine, formal quantitative analysis of uncertainty is based on the desire to move away from the use of average values or point estimates that do not identify the range of possible values or the confidence that can be associated with any particular performance measure value. Providing a numerical range of possible risks that reflects uncertainty and variability is thought to allow more-informed and more-transparent decisions than are possible when only a single point estimate is generated. The level of detail with respect to uncertainty and risk that needs to be effectively communicated to decision makers (or plan evaluators) will likely differ and be more quantitative than the information concerning uncertainty desired by other stakeholders.

Effectively communicating information about who or what is at risk or what might happen and just how severe and irreversible an adverse effect might be should a target value not be met, is just as important as the level of uncertainty and the confidence associated with such predictions. This may be qualitative information but it is often critical to informed decision-making. Risk and uncertainty communication is always complicated by the question of how much information is enough as well as how best to present it. Feedback and communication between those receiving such information and those giving it can help identify just what seems best for a particular audience.

Uncertainty characterizations must include information that is useful for all parties participating in a decision-making process involving tradeoffs between performance target values, probabilities of those targets not being met, and cost. Quantitative estimates of uncertainty are needed for this, but supplementary qualitative information on the nature of any possible adverse effects and how the uncertainty estimates were obtained is also likely to be useful.

Communicating a range or distribution of performance measure values reflecting uncertainty can be perplexing to nontechnical stakeholders who often want to know from technical staff whether a plan will work or not. Such information should not be perplexing to those involved in RECOVER. Yet the more simple and understandable the analyses, the more useful they may be in RECOVER. If there is a choice between academically interesting but rather complex ways of quantifying uncertainty and, on the other hand, conducting research (e.g., adaptive management experiments) focused on reducing important sources of uncertainty, spending money on reducing uncertainty would seem preferable to spending it on ways of calculating and describing it better.

In spite of some considerable efforts by those involved in risk assessment and management, we know very little about how to ensure effective risk communication that gains the confidence of stakeholders, incorporates their views and knowledge, and influences favorably the acceptability of risk assessments and risk-management decisions. This suggests that RECOVER will need to consider adopting comprehensive communication programs that helps stakeholders as well as decision makers better understand the implications of working or living with the uncertainties inherent in trying to restore an ecosystem as complex as the Everglades and having to make tradeoffs between targets, reliabilities, and costs.

Studies of the differences between technical and nontechnical perceptions of risk and uncertainty have identified many of the factors that contribute to misunderstandings and resentment when events do not turn out to be as expected. Our discussion here is not comprehensive; rather, it is intended to indicate the importance of effective communication and the potential for mistakes and misunderstandings.

Risk and uncertainty communication is a two-way street, however--it means both learning and teaching. Communicators dealing with uncertainty should learn about the concerns and values of their audience, their relevant knowledge, and their experience with uncertainty issues. Stakeholders' knowledge of sources and reasons for uncertainty analyses needs to be incorporated into assessment and management decisions. By listening, communicators can craft risk messages that better reflect the perspectives, technical knowledge, and concerns of the audience.

Effective communication must begin before important decisions have been made. It can be facilitated in communities by citizen advisory panels. Citizen advisory panels can give planners and decision makers a better understanding of the questions and concerns of the community and an opportunity to test its effectiveness in communicating concepts and specific issues regarding uncertainty.

One approach to make uncertainty more meaningful is to make risk comparisons. For example, a ten parts per billion, the target maximum phosphorus concentration in the Everglades, is equivalent to 10 seconds in over 31 years. If this is an average daily concentration target that is to be satisfied "99 percent," of the time, this is equivalent to an expected violation of less than one day every three months.

Even more difficult to communicate is the fact that a 1 in 100 probability of exceedance estimate is not an estimate of actual probability that an exceedance event will occur, but a statistical upper bound on the likelihood that it could happen. The actual risk may be much lower.

Many perceive the reduction of risk by an order of magnitude as though it were a linear reduction. A better way to illustrate orders of magnitude of risk reduction is shown in Figure 21, in which a bar graph depicts better than words that a reduction in risk from one in a 1,000 (10^{-3}) to one in 10,000 (10^{-4}) is a reduction of 90% and that a further reduction to one in 100,000 (10^{-5}) is a reduction 10-fold less than the first reduction of 90%. The percent of the risk that is reduced by whatever measures is a much easier concept to communicate than reductions expressed in terms of estimated absolute risk levels, such as 10^{-5}.

Risk comparisons can be helpful, but they should be used cautiously and tested if possible. There are dangers in comparing risks of diverse character, especially when the intent of the comparison is seen as minimizing a risk (NRC 1989). One difficulty in using risk comparisons is that it is not always easy to find risks that are sufficiently similar to make a comparison meaningful. How is someone able to compare two alternatives having two different costs and two different risk levels, for example, as is shown in Figure 5? One way is to perform an indifference analysis, but that can lead to different results depending who performs it. Another way is to develop utility functions using weights, where, for example reduced phosphorus load by half is equivalent to a 25 percent shorter hydroperiod in that area, but again each person's utility or tradeoff may differ.

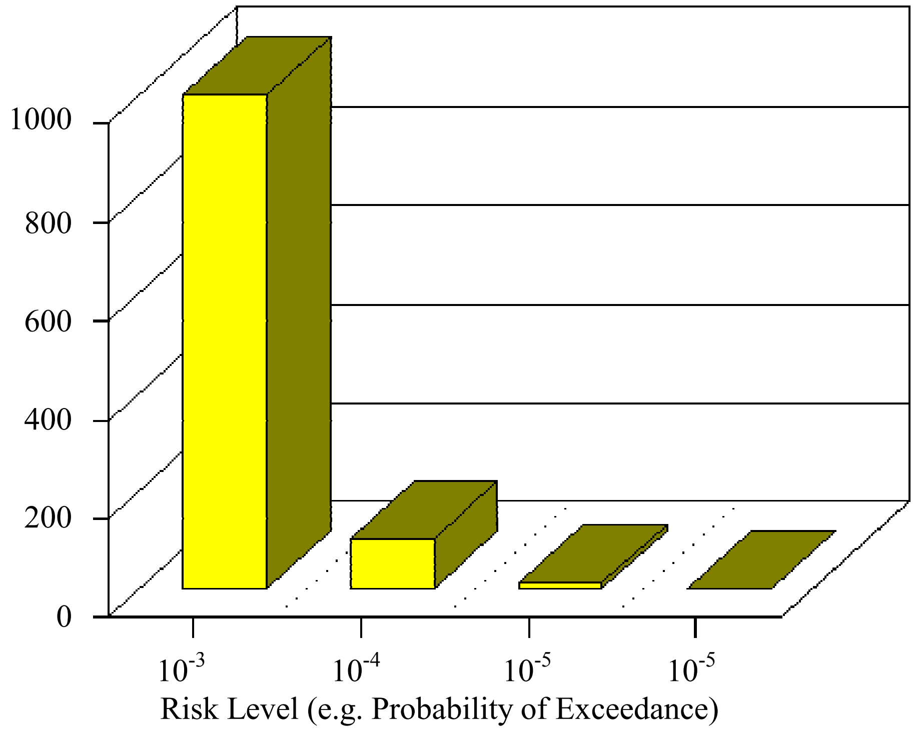


Figure 21. Reducing risk by orders of magnitude is not equivalent to linear reductions.

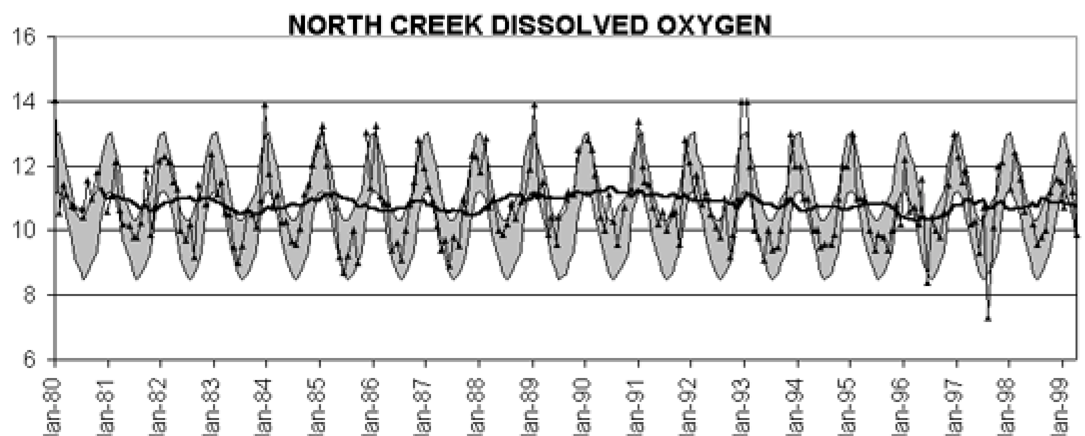


Figure 22. Graph showing the actual annual and monthly mean values of dissolved oxygen concentrations in a stream together with the ranges of the monthly average values occurring 90% of the time (North Creek data from King County, Washington).

Alternative ways of displaying uncertainty have been outlined in Lal (1994) and Loucks and Stedinger (1994). Figures 22 and 23 combine some of them to illustrate time series data and their trends and ranges within the middle 90% of their probability distributions. Tables 3 and 4 also include indicators of distributions associated with measured (or modeled) data. These plots and tables were taken from the web pages of King County in the state of Washington (http://dnr.metrokc.gov/wlr/waterres/streams/creekindex.htm).

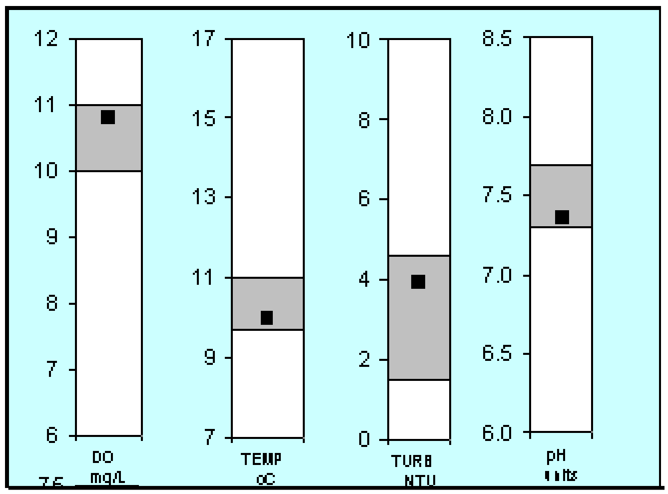


Figure 23. Values of selected water quality variables shown with gray bands identifying the mid 50 percent range of values for those variables (http://dnr.metrokc.gov/wlr/waterres/streams/creekindex.htm).

Table 3. Comparison of the Sediment Metal Concentrations (mg/kg dry weight) with State Freshwater Sediment Thresholds (http://dnr.metrokc.gov/wlr/waterres/streams/creekindex.htm).

Metal	Sediment Threshold	Mean	Min-Max
Arsenic	None	< 5.5	< Det-9.2
Silver	None	< Det	< Det
Mercury	2	< Det	< Det
Cadmium	10	< 0.23	< Det-0.39
Copper	110	7.0	5.7-8.7
Lead	250	3.1	3.1

Table 4. Summary of water quality characteristics of the stream under baseflow conditions and whether State or federal water quality criteria standards were met. Table headings are explained below.

Mouth of North Creek (0474)

	n=	Mean	Minimum	Maximun	Median +/- 1S.D.	# Non-standard	%Non-standard
FLOW (CFS)	175	38.39	1.3	300	10.8-58.0	N/A	N/A
D.O. (MG/L)	231	10.84	7.3	14.02	9.7-12.0	16	7.10
TEMPERATURE (°C)	245	9.98	0.1	20	4.7-16.4	17	6.94
TURBIDITY (NTU)	234	3.89	0.2	30	1.8-5.7	15	6.41
pH (UNITS)	231	7.35	6.09	8.4	7.2-7.8	1	0.43
CONDUCTIVITY (µMHO)	234	152.6	61	570	115-189	N/A	N/A
TSS (MG/L)	235	7.72	1.25	97.14	3.2-11.4	N/A	N/A
ORTHO-P (MG/L)	233	0.0506	0.007	0.23	.029-.071	N/A	N/A
TOTAL-P (MG/L)	233	0.0886	0.0394	0.373	.060-.115	N/A	N/A
AMMONIA (MG/L)	192	0.0375	0.001	0.112	.011-.069	N/A	N/A
NITRATE (MG/L)	232	0.8699	0.001	1.89	.487-1.374	N/A	N/A
TOTAL-N (MG/L)	76	1.2898	0.935	2.07	1.038-1.540	N/A	N/A
ENTEROCUCCUS (CFU/100ml)	127	100	10	2600	28-293	64	50.39
FECAL COLIFORM (CFU/100ml)	235	248	0	7500	71-737	172	73.19

Heading	Description
n =	Number of measurements between 1979 and 1999. Most sites have been sampled monthly.
Mean	Arithmetic average. (If the parameter has occasional extreme values, the mean can become misleading.)
Minimum	Lowest value measured at the site.
Maximum	Highest value measured
Median (+/- 1 s.d.)	Range into which the middle 70% of the values fell. This has the advantage of not being distorted by a few extreme values over the period of record. It is derived by first sorting the values in ascending or descending order and determining the middle value - the median. Second, the standard deviation (s.d.) is calculated to determine how variable the data are. Finally, the endpoints of the range are calculated - the median minus 1 standard deviation is the low end and the median plus 1standard deviation is the high end.
# Non - standard	Some parameters have water quality criteria assigned to them. The values in this column are the number of measurements that did not meet those criteria.
% Non - standard	The value described above expressed as a percentage.

5. Uncertainty and Decision Making

Consider the tradeoffs that need to be made as illustrated in Figure 5. That figure is repeated here as Figure 24, but now instead of considering a single target values as shown on Figure 5, assume there is a 90% confidence range associated with that single performance measure target value. Also assume that the target is a maximum desired upper limit (e.g., cattail area).

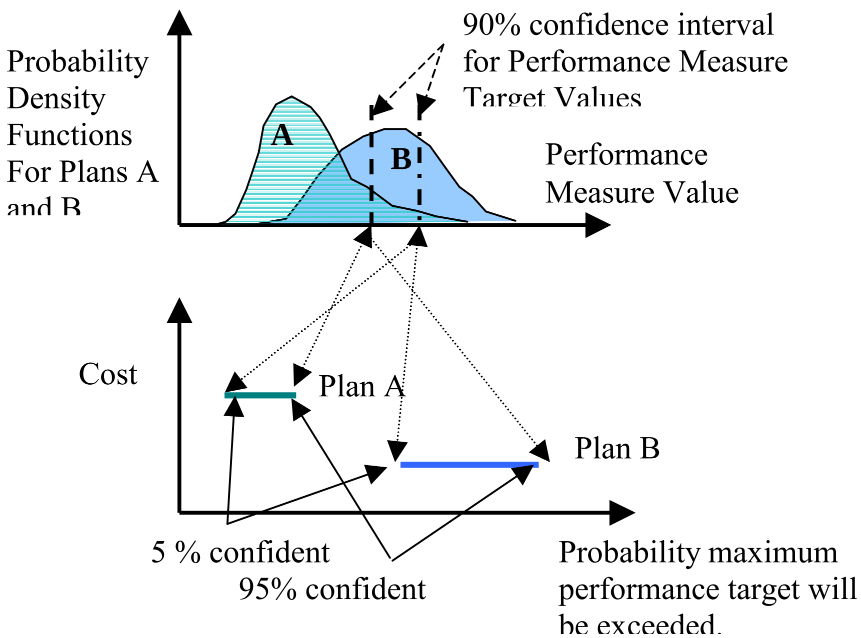


Figure 24. Two plans showing ranges of probabilities, depending on one's confidence, of the probability that an uncertain desired maximum (upper limit) performance target value will be exceeded. The 95% confidence levels are associated with the higher probabilities of exceeding the desired maximum target and the 5% confident levels are associated with the more desirable lower probabilities of exceeding the desired maximum target.

In the case shown in Figure 24, the tradeoff is clearly between cost and reliability, since no matter what confidence Plan A is preferred to Plan B with respect to reliability, but Plan A is more expensive than Plan B. The tradeoff is only between reliabilities and costs.

Consider however a third plan, as shown in Figure 25. This situation adds to the complexity of making appropriate tradeoffs, as now there are three criteria: cost, probability of exceedance (reliability) and the confidence in those reliabilities or probabilities. Add to this the fact that there will be multiple performance measure targets, each expressed in terms of their maximum probabilities of exceedance and the confidence in those probabilities.

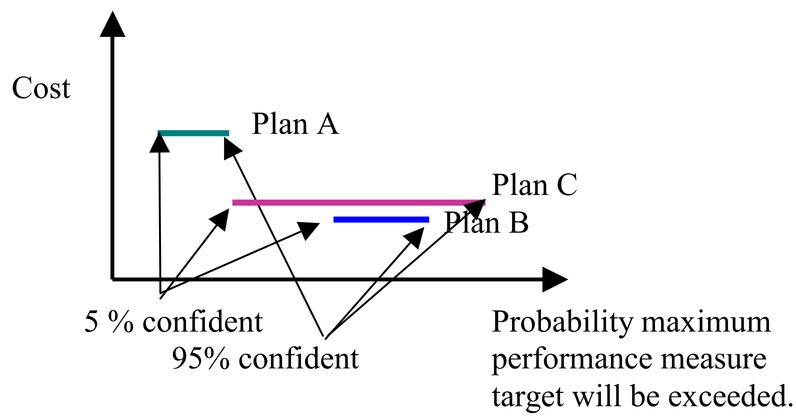


Figure 25. Tradeoffs among cost, reliabilities, and the confidence level of those reliabilities. The relative ranking of plans with respect to the probability of exceeding the desired (maximum limit) target may depend on the confidence given to that probability.

If the plan evaluation process has difficulty handling all this it may indicate the need to focus the uncertainty analysis effort on just what is deemed important, doable, and beneficial. Then when the alternatives have been narrowed down to only a few plans that appear to be the better ones, a more complete uncertainty analysis can be performed. There is no need nor benefit in performing sensitivity and uncertainty analyses on all possible management alternatives. Rather focus on those that look the most promising, and than only when the tradeoffs among important uncertain performance indicator values demands more scrutiny. Otherwise it is just too much work and it will likely not affect the decision anyway.

From uncertainty analyses one can gain an appreciation of the magnitude of the uncertainty associated with model predictions and use that information when evaluating alternative CERP management plans or projects. In the specific area of water and ecosystem management, most science-based decisions have to deal with incomplete or inaccurate science. For example, every one involved in CERP knows models cannot predict with precision the effect of water management changes on water quantity and quality distribution over space and time, let alone their impacts on multiple ecological performance measures. Yet even with this uncertainty, planners and decision makers have to rely on those model predictions to guide decisions. Models are the best tools available, without actually implementing a management policy and waiting a long time to see how it works.

The fact that this workshop was planned and held indicates the belief that improved ways of quantifying the uncertainty in model predictions would be beneficial not only for plan evaluators but also to set monitoring and research priorities.

5.1 Meaningful performance measures for decision making and public participation

It is common practice for scientists to describe the systems they are trying to manage in terms of the variables they can simulate in their planning models. For example, surface water eutrophication is typically considered by scientists in terms of phosphorus and nitrogen concentrations, chlorophyll a levels, and dissolved oxygen depletion; these are "measurement endpoints." In contrast, the public (and decision makers) tend to describe eutrophication in terms of floating algal mats, fishing success, and water color and clarity; these are "assessment endpoints." If scientists are to provide useful analyses to guide eutrophication management, then these two groups, scientists and citizens, should be "talking the same language."

For improved communication between scientists and these stakeholders and decision makers, scientific predictions should be expressed in terms of the publicly-meaningful variables. Useful scientific analysis is more likely to result from an extension of the scientific assessment to explicitly link the measurement endpoints with the assessment endpoints. This can introduce additional uncertainty – just what is the link between chlorophyll levels and greenness? Here fuzzy sets might be appropriate, but keeping within the framework already outlined and illustrated in Figures 16 and 17, subjective distributions can be used to link measurement endpoints to assessment endpoints when such links are not deterministic.

There are reasons to try always to conduct uncertainty analyses, especially when uncertainty is large or poorly understood (to avoid the mistaken impression that assessments are precise and well-understood). These reasons include:

(1) Decision makers should know the expected uncertainty in the assessed response or outcome, so they may opt for another (more precisely-assessed) attribute, may consider the value of additional experimentation or monitoring, and/or may hedge decisions away from large losses.

(2) Those affected by the decision (but not making the decision) may want to know the extent of critical scientific uncertainties so that they can make their own judgments. In many predictive analyses, uncertainty is present in more than one component (e.g., parameters and models), so there is a need to estimate the combined effects of the uncertainties on the attribute of concern. This exercise of error propagation is usually undertaken with Monte Carlo simulation or first-order error analysis (Reckhow and Chapra 1983; Morgan and Henrion 1990; Loucks and Stedinger 1994; Lal 1995; Trimble 1995).

6. Appropriate model use – some general comments

The hydrological and ecological processes that take place in the Everglades are far more complex than what analysts have been able to model and simulate. The reason is not simply any computational limitations on the number of model variables, constraints, subroutines, or executable statements in those simulation programs. More importantly, it is because we do not understand sufficiently the multiple interdependent physical; biochemical; ecological; and social, legal and political (human) processes that influence and govern the behavior of such water resource systems. These natural and social processes are affected by uncertainties in things we can measure, even simple things like water supply and water demands. They are affected by the unpredictable actions of multiple individuals and institutions who are impacted by what they get or do not get from the management and operation of such systems, as well as by other events having nothing to do with water or ecology.

So, does the use of models in developing a systematic approach to planning and management really matter? Sure. Models can and do provide useful information and insight, but the message, we suggest, is that these models need to focus on the issues of concern to their clients, the planners, and managers and the interested public. Analysts need to be prepared to interact with the political or social structure of the institutions they are attempting to assist, as well as with the public and the press.

Analysts, i.e., model users, should also be prepared to have their work ignored. Even if the analysts are presenting 'facts" based on the current state of the sciences, sometimes these sciences are not considered relevant. Happily this is not always the case. The challenge of modelers or analysts interested in having an impact on the practice of water resource systems planning and management is to become a part of the largely political planning and management process and to contribute towards its improvement.

The development and application of models, or the practice of modeling, should be preceded by the recognition of what can and cannot be achieved from the use of such models. Models of real-world systems are always going to be simplified representations of those systems. What features of the actual system are incorporated into a model, and what features are not, will depend in part on what the modeler thinks is important with respect to the issues being discussed or the questions being asked. How well this is done will depend on the skill of the modeler, the time and money available, and, perhaps most importantly, the modeler's understanding of the real system and decision making process.

Developing models is an art, requiring a knowledge of the system being modeled, the client's objectives, goals and information needs, and some analytical and programming skills. Models are always based on numerous assumptions, and some of these may be at issue. Applying these approximations of reality in a way that contributes to everyone's improved understanding and eventually to more informed decisions clearly requires considerable modeling and communication skills as well as a little bit of experience.

Water resource planners and managers must accept the fact that decisions may not be influenced by their planning and management model results. To know, for example, that cloud seeding may, on average, reduce the strength of hurricanes over a large region like the Everglades does not mean that such cloud-seeding activities will or should be undertaken. Managers or operators may know that not everyone may benefit, and those who may lose will likely scream louder than those who may gain. Hence, decision-makers may feel safer in inaction than action (Shapiro 1990; Simon 1988). There is a strong feeling in many cultures and legal systems that failure to act (nonfeasance) is considered more acceptable than acts that fail (misfeasance or malfeasance). We all feel greater responsibility for what we do than for what we do not do. Yet our aversion to risk should not deter us from addressing such sensitive issues in our models. After all, our modeling efforts should be driven by the need for information and improved understanding. It is an improved understanding (not improved models per se) that may eventually lead to improved system design, management, and/or operation. Models used to aid water resource planners and managers are not intended to be, and rarely are (if ever), adequate to replace their judgment, only to enhance it. This we have learned, if nothing else, in our over 40 years of modeling experience.

6.1 Challenges of Applying Models in Practice

The clients of planners and managers are all who uses or are served by water. The clients of modelers or analysts are typically planners and managers who have problems to solve and who could benefit from a better understanding of what to do and why, what will happen given what they do, and who will care and how much. The aim of analysts is to provide planners and managers with meaningful (understandable), useful, accurate, and timely information. This information is to help them better understand their problems and how to solve them, and to help them better manage their resources - financial, human, and water.

Modeling involves the development of a mathematical or computational framework for describing a particular system and its operation in order to study, identify, and evaluate possible solutions to problems in that system. It is a process or procedure intended to focus and force clearer thinking and to promote better decision making. The approach involves problem recognition, system definition and bounding; identification of various goals or objectives; identification and evaluation of various alternatives; and, very importantly, effective communication of this information to those who need to know.

What is usually written about in reports such as this one, and talked about in professional conferences and workshops, are various system models and methods — i.e. the tools of modelers. Modelers talk about these tools because they are interested in them. What all of us should also be interested in, and discuss more than we do, is the use of these tools in the processes of planning and management. It we did, we could learn much from each other about what tools are needed and how they can be applied in practice. We could extend the thoughts of those who, in a more general way, addressed these issues over two decades ago (Majoni and Quade 1980; Tomlison 1980; Miser 1980; Stokey and Zeckhauser 1977).

There is always a gap between what the researchers in water resource systems modeling produce and publish, and what the practitioner finds useful and uses in addressing actual problems. Those involved in research are naturally interested in developing new and improved tools and methods for studying, identifying, and evaluating alternative water resource system designs and management and operation policies. If there were no gap between what is being developed or advocated by researchers and that which is actually used by practitioners ('those who work in the swampy lowlands of the real world' as one of our clan puts it), either the researcher or practitioner, or both would be responsible. Either the research community would be very ineffective in developing new technology or the practitioners would be incredibly skilled in reading, assimilating, evaluating, and adapting their research. Not all published research is ready or suited for implementation.

How can modelers help reduce the time it takes for new ideas and approaches to be used in practice? Clearly, practitioners are not likely to accept a new modeling approach or even modeling itself unless it is obvious that it will improve the performance of their work as well as help them reduce some problems they are trying to eliminate. Will some new model or computer program make it easier for practitioners to carry out their responsibilities? If it will, there is a good chance that the model or computer program might be successfully used, eventually. Successful use of the information derived from models or programs is, after all, the ultimate test of the value of any method produced by those involved in model development and research. Peer review and publication is only one, and perhaps not even necessary, step towards that ultimate test or measure of value of a particular model or modeling approach.

6.2 Using models in data poor situations

Models require data. They require input variable and parameter value data. They need sufficient data to enable their calibration and verification, ideally. But what if there are no, or very little, data? Some would argue that such situations would preclude model use. What is the alternative? Usually the alternative is just judgment. In other words some other model, in one's head, will be used. Such an analysis will probably not include any rigorous identification of assumptions and performance criteria. They will not include any sensitivity analysis to test just how critical are many of the assumptions that must be made to reach a conclusion. Analytical or mathematical models can allow one to do this. They, even if not calibrated or verified, can provide information on what assumptions warrant expenditures of time and money to get them right, and which don't really matter much. They can identify what data we should be collecting and monitoring, and what to ignore. Such information is useful to those who make the decisions, we would argue.

Model output values for sites where little data exist to calibrate and verify those outputs will probably be wrong, but at least the relative direction of change in system performance variables associated with changes in water management policies will probably be correct. And importantly, just the process of building a model of a region forces one to really think about what is going on even in the absence of detailed data. And this is a learning experience.

6.3 Evaluating Model Success

There are a number of ways one can judge success (or failure) in applying models in practice. Goeller (1988) suggested three measures as a basis for judging success:

- How the analysis was performed and presented (analysis success);
- How it was used or implemented in the planning and management processes (application success); and
- How the information derived from the model and its application affected the system design or operation and the lives of those who used the system (outcome success).

Each of the three measures applies to a different period of time. Outcome success, for example, may not be known until some time after a particular planning or management plan has been applied or implemented.

The extent to which the models and methods and style of presentation are appropriate for the problem being addressed, the resources and time available for the study, and the institutional environment of the client are often hard to judge. Publishing in peer-review journals is one basis for judging; review panels and reports are others. No model or method is without its limitations. Two other obvious measures are the feeling the analysts have about their own work and, very importantly, the feeling the clients have about the analysts' work. Client satisfaction may not be an appropriate indicator if, for example, they are unhappy only because they are learning something they don't want to accept. Producing results primarily to reinforce a client's prior position or opinions might result in client satisfaction but, most would agree, not in a professional modeling application.

Application or implementation success implies that the methods and/or results developed in the study were used in the planning and management process by those involved in that process. One should not, it seems to us, judge success or failure based on whether or not any of the model results were accepted and incorporated into the decision regarding a design, management plan, or operating policy. What one hopes for is that the information and understanding resulting from model application helps define and focus the debate about the problem and its possible solutions, a debate that should include stakeholders as well as decision makers. The extent to which this occurs is the extent to which one will have achieved application or implementation success. To us, this is the true test of success in model application in practice.

Outcome success is based on what happened to the problem situation once a decision (that was largely influenced by the results of modeling) was made and implemented. The extent to which the information and understanding resulting from modeling helped solve the problem or resolve the issue, if it can be determined, is a measure of the extent of outcome success.

It is clear that success based on any of the last two of the three criteria will be strongly dependent on the success of the preceding criteria. Modeling applications may be judged very successful based on the first two measures, but perhaps because of unpredicted events, the problem being addressed has become worse rather than improved, or while that particular problem was eliminated, its elimination created one or more even more severe problems. All of us can think of examples where this has happened in water resource development and operation. Who knows—a broader systems study might have helped planners, managers, and decision makers foresee such consequences, but one cannot count on that. Hindsight is always clearer than foresight. Much of what takes place in the world (that changes the way we do or think about things) is completely unpredictable. Given this, it is not clear whether we should hold modelers or analysts, or even planners or managers, completely responsible for any lack of "outcome success's if unforeseen events did indeed take place.

Problem situations and criteria for judging the extent of success will change over time, of course. By the time one can evaluate outcome success, the system itself may have changed enough for the outcome to be quite different than what was predicted in the analysis. Monitoring the performance of any decision, whether or not based on a successfully analyzed and implemented modeling effort, is often neglected. But monitoring is very important if changes in system design, management, and operation are to be made to adjust to changing and unforeseen conditions. The models themselves will need to be adjusted to address those same changing and unforeseen events

If the models, data, computer programs, documentation, and know-how are successfully maintained, updated, and used by the client institutions, there is a good chance that this methodology will be able to provide useful information relevant to the changes that are needed in system design, management, or operation. Until relatively recently, the successful transfer of models and their supporting technology has involved a considerable commitment of time and money for both the analysts as well as the potential users of the tools and techniques. The SFWMD modelers are well aware of this. Developments in interactive computer-based decision support systems that provide a more easily understood human-model-data-computer interface have substantially facilitated this technology transfer process. These interactive interface developments have had a major impact on the state of the practice in applying models in the processes of water resources planning and management.

Those who are involved in the development of water resource systems modeling methodology know that the use of these models cannot guarantee development of optimal plans for water resources development and management. Given the competing and changing objectives and priorities of different interest groups, the concept of an "optimal plan" is not very realistic. What modelers can do, however, is to help define and evaluate, in a rather detailed manner, numerous alternatives that represent various possible compromises among conflicting groups, values, and management objectives. A rigorous and objective analysis should help to identify the possible tradeoffs between quantifiable objectives so that further debate and analysis can be more informed. The art of modeling is to identify those issues and concerns that are important and significant and to structure the analysis to shed light on these issues.

A measure of the success of any systems study resides in the answer to the following questions: Did the study have a beneficial impact in the planning and decision-making process? Did the results of such studies lead to a more informed debate over the proper choice of alternatives? Did it introduce competitive alternatives which otherwise could not have been considered?

7. Some Conclusions

The recent National Research Council (NRC 2001) review of the TMDL process contained a primary finding that significantly more effort is required in determining the uncertainty of environmental models.

Uncertainty analysis in support of environmental management is motivated by the fact that environmental management is usually uncertain, and sometimes highly uncertain. The opportunity that is provided by uncertainty analysis should result in more informed decision making. But it comes at a cost, and this cost should be considered along with those benefits of having uncertainty information.

We have recommended some ways of including uncertainty analysis in the modeling work of CERP which we trust will be both practical and beneficial, at least in the short run. We have also suggested some extensions at least worth exploring in the long run. Undoubtedly if and when these recommendations are implemented, those who are doing the work will find better ways of meeting the needs for uncertainty information. We are certain these ways will be improvements and should be followed. (This report does have a half life, and it may not be very long!)

There are many interesting research opportunities and needs with respect to carrying out an uncertainty analysis on a system as complex as the Everglades and with the models being used to identify and evaluate alternative water management policies. Clearly there is uncertainty associated with any prediction of what might happen if any particular water management policy is implemented. Nevertheless, the focus of any scientific study in support of uncertainty analyses for improved decision making should be on the decisions being made (or objectives) associated with the resource. A predictive model (or more generally, a predictive scientific assessment) should be evaluated in terms of its use in addressing these decisions/objectives. This means that, ideally, predicted endpoints should be decision-based. In reality, good endpoints will reflect a compromise between what is desirable to aid decision-making and what is feasible for scientific assessment.

8. References

Ahn, H. and R. T. James, 2000. Variability, Uncertainty, and Sensitivity of Phosphorus Deposition Load Estimates in South Florida. Wat. Air and Soil Pollution, 00(1-15).

Anderson, J.L., 1998. Embracing uncertainty: The interface of Bayesian statistics and cognitive psychology. Conservation Ecology (online)**2**(1): 2. Available from the Internet. URL: http://www.consecol.org/vol2/iss1/art2

Appelbaum S. 2002. Overview of RECOVER. Presentation at Model Uncertainty Workshop, South Florida Water Management District, West Palm Beach, FL, January, 2002.

Benaman, J., 2002, A Systematic Approach to Uncertainty Analysis for a Distributed Watershed Model. Ph.D. Thesis in Progress. School of Civil and Environmental Engineering. Cornell University, Ithaca, NY

Berry, D.A. 1996, Statistics: A Bayesian Perspective. Belmont, CA: Duxbury.

BIOMOVS, 1993, Biospheric Model Validation Study, Phase II, Technical Report No. 1 Guidelines for Uncertainty Analysis, Swedish Radiation Protection Institute, Stockholm, Sweden.

Borsuk, M.E., Stow C.A., and Reckhow, K.H., 2002. Ecological prediction using causal Bayesian networks: A case study of eutrophication management in the Neuse River estuary. submitted to Ecosystems (downloadable from KHR website)

Breckling, B., and Dong, Q., 2000, Uncertainty in ecology and handling strategies in ecological modeling, in S.E. Joergensen and F. Muller, eds. Handbook of Ecosystem Theories and Management, CRC Press, Boca Raton, FL

Carlson, R. and P. Fox (1976). "A Northern Snowmelt-Flood Frequency Model." *Water Resources Research* **1**2(4): 786-794.

Chapra, S. C. and Reckhow, K. H., 1979, Expressing the Phosphorus Loading Concept in Probabilistic Terms. J. Fish. Res. Board Can., 36.

Colwell, R.K., 1974, Predictability, consistency, and contingency of periodic phenomena, Ecology Vol. 55, pp 1148-53.

Covello, V.T. and Merkhofer, M.W., 1993, Risk Assessment Methods. Plenum Press. London.

Deutsch C.V. and Journel A.G., 1992. GS-LIB: Geostatistical Software Library and User's Guide. Oxford University Press.

Dilks, D.W., Canale, R.P., and Meier, P.G., 1992, Development of Bayesian Monte Carlo Techniques for Water Quality Model Uncertainty, Ecological Modelling, Vol. 62, pp 149-162.

Dilks, D. W. and James, T. R., 2002, Application of Bayesian Monte Carlo Analysis to Determine the Uncertainty in the Lake Okeechobee Water Quality Model, Proceedings, Watershed 2002, Water Environment Federation

DOE, 1998b, Screening Assessment and Requirements for a Comprehensive Assessment, DOE/RL-96-16, Rev. 1, U.S. Department of Energy, Richland, Washington

Efron, B. and Tibshirani, R.J., 1993, An Introduction to the Bootstrap, London: Chapman & Hall.

Farris, W. T., B. A. Napier, P. W. Eslinger, T. A. Ikenberry, D. B. Shipler, and J. C. Simpson, 1994, Atmospheric Pathway Dosimetry Report*, 1944-199*2, PNWD-2228 HEDR, Battelle Pacific Northwest Laboratory, Richland, Washington.

Fedra, K. A Monte Carlo Approach to Estimation and Prediction. In M. B. Beck and G. Van Straten (Eds.), Uncertainty and Forecasting of Water Quality, pp 259-291.

Fitz, H.C., Voinov, A. and Costanza, R., 1995. The Everglades Landscape Model: multiscale sensitivity analysis. South Florida Water Management District, Everglades Systems Research Division, 88 pp.

Fitz, H.C., DeBellevue, E.B., Costanza, R., Boumans, R., Maxwell, T., Wainger, L., Sklar, F.H., 1996. Development of a general ecosystem model for a range of scales and ecosystems. Ecological Modelling 88: 263-295.

Fitz, H.C., and Sklar, F.H., 1999, Ecosystem analysis of phosphorus impacts and altered hydrology in the Everglades: A landscape modeling approach, pages 585-620 in K.R. Reddy et al., eds, Phosphorus Biogeochemistry in Subtropical Ecosystems, Lewis Publishers, Boca Raton, FL.

Fontaine, T. A. and Jacomino, V. M. F., 1997, "Sensitivity Analysis of Simulated Contaminated Sediment Transport." Journal of the American Water Resources Association **3**3(2): 313-326.

Francos, A., F. Elorza, J., Bouraoui, F., Bidoglio, G., and Galbiati, L., 2001, Sensitivity Analysis of Environmental Simulation Models: Hydrological Study at the Catchment Scale. Submitted for Publication.

Gardner RH, Rojder B, Bergstrom U. 1983. PRISM: A systematic method for determining the effect of parameter uncertainties on model predictions. Technical Report, Studsvik Energiteknik AB report NW-83-/555, Nykoping, Sweden.

Gelman, A., Carlin, J., Stern, H., and Rubin, D.B., 1995, *Bayesian Data Analysis*. Chapman and Hall, London:

Gilbert, and S. P. Clemence, eds., pp. 55-74, Geotechnical Special Publication Number 82, American Society of Civil Engineers, Reston, Virginia.

Gleick, J., 19987, Chaos: making a new science, Penguin, NY

Goeller, B. F. (1988). “A framework for evaluating success in systems analysis.” ***Report P-7454,*** The Rand Corp., Santa Monica, Calif.

Haimes, Y.Y., 1998, Risk Modelling, Assessment and Management, John Wiley & Sons, Inc. New York

Henrion, M., R. Sonnenblick, and K. S. Hoo, 1997, Lessons in the Collaborative Construction of an Environmental Integrated Assessment Model: The Tracking and Analysis Framework, Lumina Decision Systems, Los Gatos, California.

Hollander M. and Wolfe D.A., 1973, Nonparametric statistical methods. Wiley, New York.

Hornberger GM, Spear RC. 1980. Eutrophication in Peel Inlet – I. The problem-defining behavior and a mathematical model for the phosphorus scenario. Water Research 14: 29-42.

Iman R.L. and Davenport J.D., 1982, An iterative algorithm to produce a positive definite correlation matrix from an “approximate correlation matrix” (with a program user’s guide). Sandia Report SAND81-1376. Sandia National Laboratories, Albuquerque, New Mexico.

Iman, R. L. and W. J. Conover, 1982, “A Distribution-free Approach to Inducing Rank Correlations Among Input Variables,” in Communications in Statistics Vol. B11, No. 3, pp. 311-334.

Isaaks E.H. and Srivastava R.M., 1989, An introduction to applied geostatistics. Oxford University Press, New York.

Jaffe, P. R., Paniconi C., and Wood E. F., 1988, Model calibration based on random environmental fluctuations, J. Environ. Eng., 114(5), 1136-1145

James, R. T. and K. R. Jin. 2001. Lake Okeechobee Water Quality and Hydrodynamic Modeling:Background and Plans. South Florida Water Management District. West Palm Beach, Fl.

Jensen, F.V., 2001, Bayesian Networks and Decision Graphs, New York: Springer.

Kann, A. and J.P. Weyant, 1999, A Comparison of Approaches for Performing Uncertainty Analysis in Integrated Assessment Models, Journal of Environmental Management and Assessment, Vol. 5, No.1, pp 29-46.

Keeney, R.L., and H. Raiffa. 1976, Decisions with Multiple Objectives: Preferences and Value Tradeoffs, Wiley. New York.

Kelly, E. and Campbell, K., 2000, Separating variability and uncertainty-making choices, Human and Ecological Risk Assessment, An International Journal, In Press, February.

Kelly, E. and Roy-Harrison, W. 1998, A Mathematical Construct For Ecological Risk: A Useful Framework For Assessments, Human and Ecological Risk Assessment, An International Journal, Vol. 4, No. 2, pp 229-241.

Kelly, E.J., Campbell, K., and Henrion, M., 1997, To Separate or Not to Separate - That is the Question - A Discourse on Separating Variability and Uncertainty in Environmental Risk Assessments, Learned Discourses, Society of Environmental Toxicology and Chemistry (SETAC) News, November.

Lal W. 1994. Evaluation of the NSM Uncertainty Using SFWMM Model Results. Draft report. South Florida Water Management District, West Palm Beach, FL

Lal W. 1995. Sensitivity and Uncertainty Analysis of a Regional Model for the Natural System of South Florida. South Florida Water Management District, West Palm Beach, FL, Draft report, November

Lal, W., Obeysekera, J. and van Zee, R., 1997, Sensitivity and Uncertainty Analysis of a Regional Simulation Model for the Natural System in South Florida, in Managing Water: Coping with Scarcity and Abundance, Proceedings, 27th Congress of International Association for Hydraulic Research, San Francisco, CA. August.

Lemons, J., ed. 1996, Scientific uncertainty and environmental problem solving. Blackwell, Cambridge, MA

Lopez, A. and Loucks, D.P., 1999, Uncertainty representations in water quality prediction. Manuscript, Civil and Environmental Engineering, Cornell University, Ithaca, NY

Loucks D.P. and Stedinger J.R., 1994. Sensitivity and Uncertainty Analysis in Hydrologic Simulation Modeling of the South Florida Water Management District. Report of the Workshop on Reduction of Uncertainties in Regional Hydrologic Simulation Models (SFWMM and NSM), January, 1994.

Ludwig, D., Hilborn, R., and Walters, C., 1993, Uncertainty, resource exploitation, and conservation: lessons from history, Science vol. 260, pp 17-36

Ma, L., J. C. A. II, Ahuja, L. R., Shaffer, M. J., Hanson, J. D., and Rojas, K. W., 2000, Root Zone Water Quality Model Sensitivity Analysis using Monte Carlo Simulation. Transactions of the American Society of Agricultural Engineers 43(4): 883-895.

Majoni, H. and Quade, E. S. (1980). *Pitfalls of analysis.* John Wiley and Sons, New York

McCuen, R., 1973, The Role of Sensitivity Analysis in Hydrologic Modeling, Journal of Hydrology 18(1): 37-53.

McCuen, R. H. and Snyder, W. M., 1983, Hydrologic Modeling: Statistical Methods and Applications. Englewood Cliffs, NJ, Prentice-Hall.

Mearns, L.O., C. Rosenzweig, and R. Goldberg, 1992, Effect of changes in interannual climatic variability on CERES-Wheat yields: sensitivity and 2xCO_2 general circulation model studies. Agricultural and Forest Meteorology 62: 159-189.

Meixner, T., Gupta, H. V., Bastidas, L. A., and Bales, R. C., 1999, Sensitivity Analysis using Mass Flux and Concentration, Hydrological Processes 13: 2233-2244.

Mills B. 2002. Regional Evaluations and Performance Measures. Presentation at Model Uncertainty Workshop, South Florida Water Management District, West Palm Beach, FL, January, 2002.

Miser, H. J. (1980). "Operations research and systems analysis." Science., 209, 174 -182.

Morgan, M.G., and Henrion, M., 1990. Uncertainty, Cambridge University Press. Cambridge, UK.

Nearing, M. A., Deer-Ascough, L. and Laflen, J. F., 1990, Sensitivity Analysis of the WEPP Hillslope Profile Erosion Model, Transactions of the American Society of Agricultural Engineers 33(3): 839-849.

National Research Council (NRC), 1989, Improving Risk Communication, National Academy Press, Washington, DC.

National Research Council (NRC), 1994, Science and Judgment in Risk Assessment, National Academy Press, Washington, DC

National Research Council (NRC), 1996, Committee on Risk Characterization, Understanding Risk: Informing Decision in a Democratic Society," P.S. Stern and H.V. Fineberg, (eds.) National Academy Press, Washington, D.C.

National Research Council (NRC), 2001. Assessing the TMDL Approach to Water Quality Management. National Academy Press, Washington, D.C.

Nicks, A.D., Richardson, C.W., and Williams, J.R., 1990, Evaluation of the EPIC model weather generator. In A.N. Sharpley and J.R. Williams (eds.), Erosion/Productivity Impact Calculator, 1. Model Documentation. USDA-ARS Technical Bulletin 1768, Temple, Texas.

Nicks, A.D., 1995, WEPP Climate Generator (CLIGEN) and Parameter Data Base. USDA-ARS, Durant, Oklahoma.

O'Neill, R.V., and Gardner, R.H., 1979, Sources of uncertainty in ecological models. Pp 447 – 63 in Zeigler, B.P., et al., eds., Methodology in systems modeling and simulation. North-Holland Pub. Co., Amsterdam, NL

Phillips D.L., Marks D.G. 1996. Spatial uncertainty analysis: propagation of interpolation errors in spatially distributed models. Ecological Modelling 91: 213-229.

Press, S.J., and Tanur, J.M., 2001, The Subjectivity of Scientists and the Bayesian Approach. New York: Wiley.

Reckhow, K.H., 1994. Water quality simulation modeling and uncertainty analysis for risk assessment and decision making. Ecological Modelling. 72:1-20.

Reckhow, K.H. and Chapra, S.C., 1983. Engineering Approaches for Lake Management – Volume I: Data Analysis and Empirical Modeling. Butterworth. Boston, MA.

Reckhow, K.H., 1999, Water Quality Prediction and Probability Network Models. Canadian Journal of Fisheries and Aquatic Sciences. 56:1150-1158.

Reckhow, K.H., Clements, J.T., and Dodd, R.C., 1990, Statistical Evaluation of Mechanistic Water Quality Models, Journal of Environmental Engineering. 116:250-268.

Reckhow K., 2002, Applications of Water Models: Prediction Uncertainty and Decision Making. Presentation at Model Uncertainty Workshop, South Florida Water Management District, West Palm Beach, FL, January

RECOVER, 2002. RECOVER Regional Evaluation Team Performance Measures. May, 2001.

Reddy, K. R., 1991. Lake Okeechobee Phosphorus Dynamics Study. Volume II. Physico-chemical Properties in the Sediments. Final report submitted to the South Florida Water Management District, West Palm Beach, Florida, 269 pp.

*Resource Conservation and Recovery Act of 197*6, 42 U.S.C. 6901, et seq., as amended.

Richardson, C.W. and Wright, D.A., 1984, WGEN: A model for generating daily weather variables, USDA-ARS report ARS-8, Temple, Texas.

Shapiro, H. T., 1990, The willingness to risk failure, Science, 250(4981), 609.

Simon, I-I. A., 1988, Prediction and prescription in system modeling, 15th Anniversary of IIASA, Int, Inst. for Applied Systems Analysis, Laxenburg, Austria.

Sklar, F. H. and Hunsaker, C.T., 2001, The Use and Uncertainties of Spatial Data for Landscape Models: An Overview with Examples from the Florida Everglades, Chapter 2 in Spatial Uncertainty in Ecology, Implications for Remote Sensing and GIS Applications, Hunsaker et al. eds, Springer

Soutter, M. and Musy, A., 1999, Global Sensitivity Analyses of Three Pesticide Leaching Models Using a Monte-Carlo Approach, Journal of Environmental Quality 28: 1290-1297.

Spear R.C. and Hornberger G.M, 1980, Eutrophication in Peel Inlet – II. Identification of critical uncertainties via generalized sensitivity analysis. Water Research 14: 43-49.

Steinberg, L.J., Reckhow, K.H., and Wolpert, R.L., 1996. Bayesian Model for Fate and Transport of PCB in Upper Hudson River. Journal of Environmental Engineering, 122:341-349.

Stokey, E. and Zeckhauser, R., 1977, A primer for policy analysis,. W. W. Norton and Co., Inc., New York, N. Y.

Suter, G.W., II. 1993. Ecological Risk Assessment. Lewis Publishers. Boca Raton, FL

Tattari, S. and Barlund, I., 2001, The Concept of Sensitivity in Sediment Yield Modelling, Physics and Chemistry of the Earth 26(1): 27-31.

Tomlison, R., 1980, Doing something about the future, J. Operational Res. Soc., 31, 467-476.

Trimble, P.J., 1995, An evaluation of the certainty of system performance measures generated by the South Florida Water Management Model, Thesis, Florida Atlantic University, Boca Raton, FL, August

Turner, M.G., Costanza, R., and Sklar, F.H., 1989, Methods to evaluate the performance of spatial simulation models. Ecological Modelling, Vol. 48, pp 1-18.

USEPA (US Environmental Protection Agency). 1992. Framework for ecological risk assessment. EPA/630/R-92/001. Risk Assessment Forum, Washington, DC.

van Griensven, A., A. Francos and W. Bauwens (2001). "Sensitivity Analysis and Auto-calibration of an Integral Dynamic Model for River Water Quality." In Progress.

van Harn Adams, B.A. 1998. “Parameter Distributions for Uncertainty Propagation in Water Quality Modeling.” Doctoral Dissertation, Duke University.

von Winterfeldt, D., and Edwards, W., 1986, Decision Analysis and Behavioral Research, Cambridge University Press, Cambridge, UK.

Wallis, T.W.R. and J.F. Griffiths, 1995, An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator (EPIC). Agricultural and Forest Meteorology 73: 115-133.

Walker, W. E., 1987, Important hard problems in public policy analysis, Report P-7282, The Rand Corp., Santa Monica, Calif.

Warwick, J. J. and Cale, W. G., 1987, Determining likelihood of obtaining a reliable model, J. Env. Eng., 113(5), 1102-1119

Wilks, D.S., 1992, Adapting stochastic weather generation algorithms for climate change studies. Climatic Change 22: 67-84.

Wu, Y., Sklar, F.H., Rutchey, K., 1997, Analysis and simulations of fragmentation patterns in the Everglades, Ecological Applications 7: 268-276.

Appendix

Uncertainty topics in south Florida modeling, analysis, and decision systems.

This document is intended to provide some guidance to the panel of experts participating in the Comprehensive Everglades Restoration Plan's (CERP) Model Uncertainty Workshop. The goal of this workshop is to develop guidelines to quantify and communicate RECOVER model uncertainty for meaningful decision making. The models are briefly introduced in Section I. The uncertainty topics are outlined at increasing levels of detail in Sections II - IV. Of primary consideration are the Critical Topics, or the minimal set of topics for which we hope to get significant input from the panel of experts. The Short List section expands on these and related topics that are deemed to be the more significant problems faced by most CERP modelers and model users. The Long List section further details these uncertainty topics, including some subtopics that may prove to be of lesser overall importance in CERP decisions.

A.1. Model information

REGIONAL dynamic models for primary consideration:
Most are spatially distributed at various grid/mesh resolutions over >10,000 km^2
Period of analysis is approximately 3 decades
Hydrologic models (e.g., SFWMM, NSM) have been predominant tools
Ecological models (e.g., ELM, ATLSS) use output from SFWMM to varying extent

Refinement and development continues on these and other related hydrologic (SFRSM/HSE, SICS, MODBRANCH) and ecological models (ELVM, other ATLSS)

A.1.1 Current Models

South Florida Water Management Model (SFWMM)

The South Florida Water Management Model (SFWMM) is a regional-scale computer model that simulates the hydrology and the management of the water resources system from Lake Okeechobee to Florida Bay. It covers an area of 7600 square miles using a mesh of 2 mile x 2 mile cells. In addition, the model includes inflows from Kissimmee River, and runoff and in the Caloosahatchee River and St. Lucie canal basins. The model simulates the major components of the hydrologic cycle in south Florida including rainfall, evapotranspiration, infiltration, overland and groundwater flow, canal flow, canal-groundwater seepage levee seepage and groundwater pumping. It incorporates current or proposed water management control structures and current or proposed operational rules. The ability to simulate water shortage policies affecting urban, agricultural, and environmental water uses in South Florida is a major strength of this model. The SFWMM currently simulates hydrology on a daily basis using climatic data for the 1965-1995 period which includes many droughts and wet periods. In the near future it will use a 1965-2000 climatic record. The model has been calibrated and verified using water level and measurements at hundreds of locations distributed throughout the region within the model boundaries. Technical staffs of many /state/local agencies and public/private interest groups have accepted the SFWMM as the best available tool for analyzing regional-scale structural and/or operational changes to the complex water management system in south Florida. Documentation is available online at : http://www.sfwmd.gov/org/pld/hsm/models/sfwmm/index.html. For applications using the SFWMM see http://www.sfwmd.gov/org/pld/hsm/reg_app/index.html

Natural System Model (NSM)

The Natural System Model attempts to simulate the hydrologic response of the pre-drainage Everglades using the same climatic inputs, daily time step, calibrated model parameters and algorithms as the SFWMM. The NSM differs from the SFWMM in that it does not simulate the influences of any man-made features and uses estimates of pre-subsidence topography and historical vegetation cover. Use of the same climatic input allows for meaningful comparisons between the response of the managed system, simulated with the SFWMM, to that of the natural (pre-drained) system, simulated with the NSM. Some documentation of the NSM is available online at http://www.sfwmd.gov/org/pld/hsm/models/nsm/index.html

Everglades Landscape Model (ELM)

The ELM simulates the landscape response to different water management scenarios in the greater Everglades. It integrates some of the major ecosystem processes within a heterogeneous landscape, simulating dynamics of hydrology, soil & water nutrients, periphyton biomass & community type and macrophyte biomass & community type. Sensitivity analyses were used to evaluate model response to a large array of parameters across a spatial hierarchy. An early model version was calibrated to data on the full spectrum of ecological dynamics in a well-studied basin of the Everglades (WCA-2A). The current version is calibrated (1979-1995) for hydrology and surface water quality in the ~40 monitoring stations throughout the greater Everglades region. Rainfall and a variety of static physical data are shared with the SFWMM, which also provides daily managed flows through all water control structures (with the ELM simulating the subsequent spatial flow distributions). The full regional domain uses a 1 km^2 grid resolution, but the ELM is scalable. Results/documentation at: http://www.sfwmd.gov/org/erd/esr/elm.html

Across Trophic Level Systems Simulation (ATLSS)

ATLSS is an integrated system of simulation models representing the biotic community of the greater Everglades region and the abiotic factors that affect this community. The models are spatially explicit and have a resolution of 500 m x 500 m or finer. The abiotic processes simulated by the model are hydrology, fire and major storms. Presently, the ATLSS models are configured to use output from the SFWMM although it can be interchanged with any other hydrology model. The biotic modeling components integrate three approaches: (a) process models for lower trophic levels (including benthic insects, periphyton, and zooplankton), (b) structured population models for several important functional groups of fish and macroinvertebrates, and (c) individual-based models for endangered species or large consumers (Cape Sable seaside sparrow, wood storks, great blue herons, white ibis, American alligators, white-tailed deer, and Florida panther). The overall goal of the ATLSS models is to aid in understanding how the biotic communities of South Florida are affected by the hydrologic regime and other abiotic factors, and to provide a predictive tool for evaluating management alternatives. For more information on ATLSS see http://www.atlss.org

Lake Okeechobee Water Quality Model (LOWQM)

The Lake Okeechobee Water Quality Model (LOWQM) was developed to evaluate various management scenarios (including phosphorus load reduction and hydrology management) to determine the relative impacts on phosphorus and algal blooms within the lake. It uses the U.S. Environmental Protection Agency's Water Quality Analysis Simulation Program (WASP) (Ambrose et al., 1993) to simulate nutrient and algal dynamics in both the water column and underlying sediments. The original model was updated to include three algal groups: representing green algae, diatoms, and cyanobacteria; suspended solids, and processes related to sediment resuspension, the silica cycle, and nitrogen fixation. The model was further modified to include dissolved organic phosphorus and three forms of particulate organic phosphorus: labile, moderately labile and non-degradable. External forcings that drive the model include solar radiation, temperature, wind induced sediment resuspension, surface discharges into and out of the lake, rainfall, evaporation, and nutrient loads. The LOWQM treats the lake as a single homogenous entity. It is calibrated to monthly averaged observations of nutrients and chlorophyll *a* from the eight-station long-term network on the lake for 1982 to 2000. For more information go to, http://www.sfwmd.gov/org/wrp/wrp_okee/projects/lowqm.html

A.2. Critical topics

The primary goals of the uncertainty workshop are to:

1. Develop a method to characterize uncertainty in (a) predictive system-wide models, (b) performance measures produced from model output and (c) performance measure targets against which predictions are compared, for RECOVER evaluations.

2. Provide guidelines to those tasked with evaluating model alternatives on how to deal with uncertainty.

Critical topics for the uncertainty workshop panelists to address are,

- Methods to evaluate spatially distributed data (such as topography)
- Methods to calibrate and optimize parameters for models with varying complexity
- Minimum "standards" for quantifying uncertainty of models with varying complexity and objectives
- Spatio-temporal, other considerations when "soft" linking I/O among models
- Methods to estimate uncertainty in performance measures and their targets (hypotheses based on observed data)
- Visualization of model uncertainty relative to performance measure uncertainty
- Use of relative differences between model runs when models have limited calibration / verification
- Development of decision thresholds
- Effects of model structure on uncertainty

A.3. Short list of topics

A.3.1 Input data: uncertain data with high model sensitivity

Spatial data evaluation

Data: Land surface elevation
Data: Rainfall distribution and temporal frequency
Data: Soil attributes
Methods: required point distribution for different data types, statistically valid interpolation
Methods: understanding spatial and temporal aggregation biases
Methods: Spatio-temporal considerations when "soft" linking I/O among models

Parameter estimation & optimization

Data: Multiple coefficients (e.g., friction, shear stress) in surface flow for different habitats
Data: Plant & animal growth, uptake kinetics, mortality rates under varying hydrologic & nutrient conditions

Data: Decomposition rates of different soil types, varying hydrologic & nutrient conditions
Data: Soil accretion rates sediment resuspension and settling in different habitats under varying hydrologic conditions
Methods: Parameter optimization techniques for distributed models, many parameters
Methods: Aggregation from lab/field scale to model application

A.3.2 Model output/performance

Model structure

Methods: Selection of model grid/mesh resolution and extent in relation to objectives
Methods: Balance between statistical- vs. mechanistic/process- based models
Methods: Effects of discrete thresholds in model application

Calibration/validation/verification

Methods: Heuristic vs. predictive capabilities of models
Methods: Calibrate 5 yr, then verify 5 yr; or calibrate to entire 10 yr data set
Methods: Use of well-studied "reference" areas for model calibration, extrapolation to region
Methods: Need for, at minimum, sensitivity analysis of uncalibrated models

Calculating uncertainty

Methods: Minimum "standards" for sensitivity analysis
Methods: Feasibility of quantitative, multi-parameter uncertainty analysis of distributed models
Methods: Use of multi-scales, or spatial hierarchy, in sensitivity analysis of distributed models

A.3.3 Decision systems: analysis & communication

Model performance measures

Methods: Evaluate uncertainty in performance measures and performance targets
Methods: Too few, or too many, relative to their uncertainty
Methods: Appropriate and inappropriate spatial aggregation, temporal aggregation

Communicating uncertainty

Methods: Point time series graphs: confidence intervals etc.
Methods: Multi-cell indicator region time series graphs: confidence intervals etc.
Methods: Spatial maps: 3 dimensional surface; color=mean, height=uncertainty

Model comparisons

Methods: Necessary difference between uncertain target values and uncertain model values
Methods: Necessary model-model difference with "known" uncertainty bands

Methods: Necessary model-model difference with relatively unknown uncertainty
Methods: Necessary model-model difference with completely unknown uncertainty

Forecasting

Methods: Necessary extent of integrating scenarios of future climate, land use, etc

A.4. Long list of topics

A.4.1 Input data

Dynamic boundary conditions/forcing functions

Spatial considerations: methods to evaluate sparsely distributed spatial data

Extrapolation/interpolation of sparse points along continuous domain boundary
- Ex: tidal stage observations distributed along handful of observed points

Interpolated surface from sparsely distributed point observations
- Ex: Uneven, very sparse meteorological station distribution for ET (rare in Everglades)
- Ex: Uneven, sparse distribution of rainfall stations (few in Everglades)
- Ex: Atmospheric nutrient deposition stations rarely within Everglades

Use of coarse grid model for input to fine grid model
- Ex: SFWMM stage and/or flow boundary (border) conditions for different models
- Ex: SFWMM output stage maps post-processed to finer grid, based on vegetation

Temporal considerations: methods to accommodate missing data

Interpolation of missing/infrequent observations
- Ex: Monthly water quality observations, but require daily observation at inflow points
- Ex: Missing/uncertain data in daily rainfall observations, water control structure flows

Output from coarse time step model for input to fine time step model
- Ex: SFWMM daily flows vs. 15 minute required boundary conditions for some models

Static initial conditions and parameters

Spatial considerations

Aggregation/disaggregation
- Ex: Sampling bias/error in raw data relative to interpolation bias
- Ex: Land surface elevation: 1ha & 10 km^2 interpolations from 400 m point distribution

Ex: Soil attributes (e.g. bulk density, TP): 1ha & 1 km^2 grid from random, ~5 km spacing
Ex: Hydrogeological attributes of aquifer(s) from varying sources/distributions
Ex: Initial habitat maps at 1 ha, 1km^2, 10km^2 resolutions from ~10-30 m polygons

Temporal considerations

Maintaining static conditions for dynamic variables
Ex: Land elevation (changes ca. decadally)
Ex: Human land use and water demands (changes annually/decadally)
Ex: Water management infrastructure and operations (CERP/policy-dependent)
Ex: Habitat type, biomass, density (changes seasonally/annually)
Hindcasting initial conditions (calibration runs)
Ex: Land elevation
Ex: Soil nutrients
Ex: Plant community type

Parameters

Parameter estimation & optimization
Data: Multiple coefficients (e.g., friction, shear stress) in surface flow, different habitats
Data: Plant & animal growth, mortality under varying hydrologic & nutrient conditions
Data: Decomposition of different soil types, varying hydrologic & nutrient conditions
Data: Soil accretion rates in different habitats under varying hydrologic conditions
Feasibility of parameter optimization techniques for distributed models, many parameters
Aggregation from lab/field scale to model application

A.4.2 Model output/performance

Model structure

Selection of scale

Spatial aggregation & extent
Ex: Simulating fine scale features (vegetation, mgmt infrastructure) in coarse grid
Ex: Using "representative" fine-scale subregional models (flow in ridge/slough)
Ex: Spatial extent of uncertainty along boundaries of different models
Temporal scale
Ex: "Short term" (30 yr) models for longer term responses (soil accretion, tree island)

Complex system feedbacks

Process, or algorithm, aggregation
- Ex: Water quality simulation with single parameter that aggregates all processes
- Ex: Water quality simulation with interactions among processes, multiple parameters

Omission of feedbacks
- Ex: Importance of local/regional vegetation change on surface friction, local/regional ET

Error propagation

Interactions among uncertain variables
Accumulated errors in cumulative variables (e.g., 31 yr sum flows, 31 yr elevation change)
Impact of uncertainty in initial conditions
Spatial propagation of errors across grid (with flows)

Model output analysis

Complex system balance

Performance of multiple interacting variables as check on total system calibration
- Ex: "Known" rain, structure inflows, ET, recharge, & stages => flow estimation
- Ex: Ensure rates and stocks of material are realistic for all variables in complex system, not just the stock (stage, biomass etc.) of the "target" variable

Calibration "standards" for variables of different temporal response times

Fast response variables
- Stage – moderately dense daily observations in natural and urban/ag areas
- Flow – measured daily at structures; extremely rare observations elsewhere
- Water quality – measured ca. monthly at sparse network of marsh stations

Slow response variables
- Elevation – sparse and/or rare observation updates
- Vegetative habitat – infrequent but fine spatial observation updates

Adequacy
- Magnitude of model-observation deviations relative to natural & sampling variance
- Useful temporal and spatial extent of calibration vs. model application

Uncertainty analysis "standards"

Sensitivity and formal uncertainty analysis
- Ex: Sensitivity of all parameters, all combinations, and/or of selected parameters

Ex: Spatial hierarchical ("multi-scale") sensitivity in modular, distributed models
Ex: Quantitative uncertainty, considering variable interactions, error propagation, across time and space
Ex: Methods to communicate complex sensitivity or uncertainty analysis results

Aggregation

Temporal
Ex: Weekly, monthly, annual, and 30 yr aggregations of point data
Spatial
Ex: Aggregate entire basins (WCAs, Lake Okeechobee) vs. analyze subregions
Spatio-temporal
Ex: Weekly, annual, & 30 yr aggregations of spatially aggregated (Indicator Region) data

A.4.3 Decision systems

Relating model uncertainty to decision process

Model performance measures

Too few, or too many, relative to the target uncertainty and to model uncertainty
Appropriate and inappropriate spatial aggregation, temporal aggregation

Model comparisons

Necessary difference between uncertain target values and uncertain model values
Necessary model-model difference with "known" uncertainty bands
Necessary model-model difference with relatively unknown uncertainty
Adequacy of relative, model-model comparisons, when calibration & uncertainty are unknown

Visualization and communication

Visualization methods for comparing model runs

Explicit confidence bands in time series output for calibration and for scenarios
Error surfaces in spatial distributions of output for calibration and for scenarios

Forecasting

"Missing" data on future conditions

Ex: Future climate assumed to be represented by past ~30 dynamics
Ex: Sea level rise not accommodated in standard production model runs
Ex: Necessary extent of integrating ecological-economic benefits/risks