UNCERTAINTY ANALYSIS OF HYDROLOGIC AND WATER QUALITY MODELS

Typically hydrologic and water quality models are complex collections of algorithms combined in such a way as to mathematically mimic some hydrologic system. We might write a generic model as:

$$\underline{O} = \underline{f}(\underline{I}, \underline{P}, t) + \underline{e} \quad (1)$$

where, $\underline{O}$ represents the outputs being modeled, $\underline{I}$ represents the inputs to the model such as rainfall, temperature, etc., $\underline{P}$ represents the parameters required by the model and $\underline{e}$ represents errors associated with the modelling process.

One axiom of stochastic processes is that any function of a random variable is itself an random variable. Thus if any of the variables in $\underline{I}$ and/ or $\underline{P}$ are uncertain and known only in a probabilistic sense, then $\underline{O}$ is also uncertain and can be known only in a probabilistic sense. What this means in the context of hydrologic and water quality modelling is that if we are uncertain about $\underline{I}$ or $\underline{P}$, then we are uncertain about $\underline{O}$ as well. Uncertainty is transferred from the inputs to the outputs.

As added complication with hydrologic and water quality models is that in general there are several uncertain parameters. We must be sure in any uncertainty analysis dealing with more thana one variable that we do not violate relationships that exist among the input parameters. If parameters tend to vary together in real life, then we must preserve this joint variability structure in our uncertainty analysis. We are talking about the correlation structure among the independent variables. Those of you who have worked with hydrologic and water quality models and attempted to find optimal parameters values known what I mean when I talk of a correlation structure among the parameters. Optimal values of one parameter are dependent to a certain degree on values assigned to the correlated parameters. Correlation and its computation are discussed in Haan (1977).

UNCERTAINTY ANALYSIS

Determining the uncertainty to assign to input parameters is one of the major hurdles that must be addressed in the overall evaluation of uncertainty associated with hydrologic and water quality modelling. If we are fortunate, the user manual will provide some guidance in estimating parameter values. Using this guidance and our own experience, we must come up with a single best estimate or expected value for each of the parameters. We need to investigate limits on parameters and suggested ranges of parameter values. If the parameters is a physically measurable parameter, we need to look into the literature and see what kind of variability is reported for the parameter. Our goal is to come up with the following quantities in order of priority:

1. Expected value
2. Variance
3. Distributional shape

Typically one might find a table of suggested values that give average values and ranges for the parameters under a variety of conditions. What can be done is to take the suggested value as the expected value or mean parameter estimate. The range might be taken as 2 or 3 standard deviations.

For example a table of Manning's n for a natupal stream that is winding with pools and riffles might show a minimum of 0.035 and a maximum of 0.050. From this several possibilities are available. We might assume a uniform distribution with = 0.035 and = 0.050. This assumes any value in the interval is as likely as any other and that values outside the interval are not possible. We might assume a triangular distribution with the minimum at 0.035, maximum at 0.50 and mode at 0.042. We might assume a normal distribution with mean = 0.042 and a standard deviation of 0.015/2 or 0.0075. This latter assumption would, of course, allow for values smaller than 0.035 and larger than 0.050. If we can only estimate x, we might take C_V = 0.2 or 0.3 and then s = C_V X.X and s are more important than the pdf that is used.

Assessing the correlation structure among parameters is much more difficult. For physical parameters such as bulk density, K %OM, % clay, etc., we might find some field data from which the correlation matrix might be calculated. In other cases we might have to rely on a rational analysis of the parameters. In the case of pseudo physical parameters or parameters designed to describe something physical but which in themselves are not physically measurable, one has to resort to experience and assumptions.

At this point it is apparent that considerable work may be involved in gathering the data required to characterize the uncertainty in each parameter and the parameters as a whole. One would not want to go to all this work unless in fact the parameter was important to the process being modeled. If a parameter has little impact on the output of a model, we don't want to spend a great deal of time estimating that parameter or worrying about uncertainty in that parameter.

SENSITIVITY ANALYSIS

The processes for identifying important parameters include sensitivity analysis. We desire to determine the sensitivity of model outputs to changes in values for model inputs. Two types of sensitivity coefficients are used. One is called an absolute sensitivity coefficient or simply the sensitivity coefficient, S, and the other a relative sensitivity coefficient, S. These are given by

$$S = \frac{\partial O}{\partial P} \qquad S_r = \frac{\partial O}{\partial P}\,\frac{P}{O} \tag{2}$$

where, O and P represent particular model outputs and parameters respectively. S gives the absolute change in O for a unit change in P while S gives the % change in O for a 1% change in P. Graphically the terms in these relationships are shown below.

Obviously for most hydrologic and water quality models, numerical procedures must be used since analytic partial derivatives can not be obtained. Thus one has to approximate the above derivatives by

$$S = \frac{O_2 - O_1}{P_2 - P_1} \qquad S_r = \frac{O_2 - O_1}{P_2 - P_1}\,\frac{P}{O} \tag{3}$$

where, P is given by $(P_1 + P_2)/2$ and O is given by $(O_1 + O_2)/2$. When evaluating these partial derivatives, all other parameters are set at their expected values. The derivatives are also taken about the expected value of P.

I prefer the relative sensitivity coefficients since they are dimensionless and can be compared across parameters while the absolute sensitivity coefficients have units of output over input and can not be directly compared across noncommensurate parameters.

Parameters can be ranked on the basis of their relative sensitivity coefficients and only the most sensitive ones retained for further analysis.

FIRST ORDER ANALYSIS

As has been previously indicated, FOA can be used to estimate the variance of a function or model in terms of the variance of the various parameters that go into the model. FOA equations are relatively simple to apply; however, the basic assumptions that are made must be kept in mind. Again these assumptions are that $f(X^0)$ is nearly linear and that $(x_1 - x_2^0)/2$ is small for I=1 to n. This latter assumption is often expressed in terms of a C_v being less than 0.2 or some other fixed value.

We can put the results of the above development of FOA in terms of hydrologic modelling using the notation previously adopted for a general hydrologic model

$$\underline{O} = \underline{f}(\underline{I}, \underline{P}, t) + \underline{e}$$

where, $\underline{O}$ represents model outputs and $\underline{P}$ represents the parameters of interest. The resulting estimates for the mean and variance are:

$$E(O) = f(I,P,t) \tag{4}$$

$$Var(O) = \sum_{i=1}^{p} \left[\left(\frac{\partial f(I, P, t)}{\partial P_i} \right)^2 \right]_{\bar{P}} Var(P_i) \tag{5}$$

$$+ 2 \sum_{i=1}^{p} \sum_{j=i+1}^{p} \left[\frac{\partial f(I, P, t)}{\partial f_i} \frac{\partial f(I, P, t)}{\partial P_j} \right]_{\bar{P}} Cov(P_i, P_j)$$

The expression for the variance appears to be quite complex and is unless the parameters are independent. The last term of the Variance expression contains Cov (P_i ,P_j). If P_i and P_j are independent then this term is zero. If P_i and P_j for all I and J are independent, then the Variance expression reduces to:

$$Var(\bar{O}) = \sum_{i=1}^{p} \left[\left(\frac{\partial f(I,P,t)}{\partial P_i} \right)^2 \right]_{\bar{P}} Var(P_i) \tag{6}$$

Recognizing the definition of the absolute sensitivity coefficients given by equation 2, this becomes

$$Var\ (O) = \sum_{i=1}^{P} S_i^2\ Var\ (P_i) \qquad (7)$$

where S_i is the absolute sensitivity with respect to P_i.

The fraction of the total variance due to the ith parameter is given by

$$F_i = \frac{S_i^2\ Var\ (P_i)}{\sum_{i=1}^{P} S_i^2\ Var\ (P_i)} \qquad (8)$$

F_i can be used to identify which parameters are the largest contributors to uncertainty in the output based on the FOA. Notice that F_i combines parameter uncertainty in terms of parameter variance and the sensitivity of the model to the parameter. Thus FOA has some very attractive features. Unfortunately, Taylor series expansion we use is truncated and thus is only an approximation for E(O) and Var(O). FOA is said to be valid for those situations where the model is nearly linear in the parameters of interest and the C_v on P_i is small. This can be seen from the following illustration.

We also note that FOA produces only estimates for the E(O) and Var(O). If we want to look at probabilities of the O being in certain ranges, we must make pdf assumptions regarding the O. For example if we take O to be normally distributed with a mean of E(O) and a variance of Var(O), we can calculate probabilities of O in any range and we can put confidence intervals on O. Other pdf assumptions can be made as well.

FOA is computationally efficient requiring only 2p+1 model runs for a model in which p parameters are under consideration. For example, with 6 parameters, 13 model runs are required.

MONTE CARLO SIMULATION

MCS is a sampling procedure in which possible values for the input parameters are selected at random from appropriate pdfs and used in the model to produce estimates of the outputs. MCS is more complete than FOA and does not require assumptions of linearity and small parameter variances. It does require assumptions on the parameter pdfs and is computationally intensive generally requiring 1500 or more model runs. The approach illustrated in the following figure is actually quite simple in concept. After defining the probabilistic structure of the parameters, a parameter set is selected at random from the appropriate multivariate pdf. The model is run with these parameters and the output noted. This entire process is repeated many times (~1500) resulting in many (~1500) estimates for the outputs. These outputs are then analyzed probabilistically. The means and variances are calculated. Appropriate pdfs or probability plots are prepared and probabilistic statements made. Confidence intervals can be read directly from probability plots without making a distributional assumption about the outputs.

The importance of individual parameters in determining the uncertainty associated with a particular variable can be assessed by computing the correlation between the model outputs and the input parameters. Those parameters that are highly correlated with the model outputs are obviously important since changes in the values of these parameters will result in a corresponding change in the outputs from the models. On the other hand, if the correlations between outputs and parameters is low, the parameter is not very influential in determining the outputs.

Two checks that should be incorporated into any MCS are an examination of the correlation structure of the randomly generated parameter values and an examination of the pdfs of the randomly generated parameter values. The correlation structure of the parameters must statistically match the target correlation structure. If the intent was to generate independent rvs but in fact a high degree of correlation ends up in the generated parameters, the output variance will be incorrectly estimated. For example consider p1 and p2 as two parameters that are positively correlated with the model output but are uncorrelated with each other. If the generated "random" values of p1 and p2 are significantly positively correlated, the variance and the uncertainty in the output of the model will be exaggerated.

The fraction, F_i , of the total variance in model output attributable to the ith parameter based on a MCS can be estimated by computing

$$F_1 = \frac{r_{0,i}^2}{\sum_{i=1}^{p} r_{0,i}^2} \qquad (9)$$

where r_{oj} is the correlation between the output and the ith parameter and p is the number of uncertain parameters. This is a very rough approximation and may be used for guidance only. If the output variance is judged to be excessive, equations 8 and 9 can be used to determine which parameters are the biggest contributors to this uncertainty. Attempts can then be made to reduce the variance on these influential parameters by getting better estimates for then.

Equation 7 can be used to estimate how much variance a particular parameter can have and still achieve a target variance on O. For example if V_t is the target variance, the target variance on parameter p_j can be estimated from

$$Var\ (p_j) = \frac{1}{S_j^2}\left[V_t - \sum_{i=1_{\neq j}}^{p} S_t^2\ Var(p_t)\right] \qquad (10)$$

The variance reduction calculations provide guidance only. They are not exact (obviously) since neither FOA or MCS is exact. Furthermore, it is not always possible to reduce the variance of the input parameters. The results of a MCS can be subjected to a multiple regression analysis in a further effort to identify the important input parameters. Regression coefficients relating a model output to the input parameters that are not statistically significant are likely of little importance in determining the model outputs.

SAMPLE SIZE IN MONTE CARLO SIMULATION

Since Monte Carlo simulation (MCS) is a sampling procedure, one can expect that if several different simulations are made, they will each produce a different estimate for the mean of the process. Based on sampling theory and the assumption that the estimated mean from a MCS is approximately normally distributed, it is possible to estimate the sample size required to achieve a given level of accuracy in the estimate of the mean. If X is the mean and S_x^2 the variance of the output from a MCS, the variance of the mean is S_x^2 /n where n is the sample size. Letting w be the acceptable tolerance in the estimated mean at a (1-)% confidence level. The (1-)% confidence intervals on X are given by:

$$\left(\bar{X} - z\frac{S_x}{\sqrt{n}}, \quad \bar{X} + z\frac{S_x}{\sqrt{n}} \right) \tag{11}$$

where, z is a standard normal variate corresponding to a probability level of $\alpha/2$. These confidence intervals can be written in terms of the sample required to insure that the width of the confidence interval does not exceed a set level say w.

$$w = \left(\bar{X} + z\frac{S_x}{\sqrt{n}} \right) - \left(\bar{X} - z\frac{S_x}{\sqrt{n}} \right) = 2z\frac{S_x}{\sqrt{n}}$$

where z is the standard normal deviate. This result can be solved for n resulting in

$$n = \left(\frac{2zS_x}{w} \right)^2 \tag{12}$$

In using this relationship, we can make a few MCS runs, calculate S_x, and then determine n. For example if 10 simulations result in an estimated standard deviation of 5 units and it is desired to be 95% confident that the mean is estimated to within 3 units, the required number of simulations can be calculated recognizing that $z_{1-\alpha\Omega} = z_{.975} = 1.96$ from

$$n = \left(\frac{2(1.96)(5)}{3} \right)^2 = 43$$

Therefore an additional 33 simulations are required. The following figure shows the expected change in the width of the 95% confidence intervals as a function of the sample size for this example with an assumed mean of 10 units.

We can also base the sample size requirement on the desired degree of accuracy for estimating a percentile from the distribution rather than the mean. We can rank the simulation outputs from smallest to largest so that $x_1 < x_2 < .. < x_m$. Here x_1 is an estimate for X_p where p=I/m. Morgan and Henrion (1990) state that the sample values x_i and x_k constitute the α confidence interval where

$$i = mp - z\sqrt{mp\,(1-p)} \qquad k = mp + z\sqrt{mp\,(1-p)} \tag{13}$$

and z is the deviation enclosing probability α of the standard normal distribution the value of x_i is rounded down and x_k is rounded up.

Consider that we want the α confidence interval on the pth fractile X_p given by (x_i, x_k) where x_i is an estimate of X_{p-} and x_k is an estimate of X_{p+p}. That is we want α confidence of X_p being between the sample values used as estimates of the (p-p)th and the (p+p)th fractiles. The sample size requirement can be determined by nothing that I = m(p-p) and k = m(p+p) or k-I = 2m p. From equations (13)

$$k - i = 2z\sqrt{mp(1-p)}$$

equating these two expressions for k-I, we get

$$m = p(1-p)\left(\frac{z}{\Delta P}\right)^2 \tag{14}$$

As an example, the 95% confidence that the 90th percentile, $X_{.90}$, is between estimates of the 85th and 95th percentiles can be determined. Here p = 0.05 and z = 1.96 so that

$$m = 0.90 \times (1-0.90) \times \left(\frac{2}{0.05}\right)^2 = 144$$

It should be noted that this result is independent of the shape of the distribution.

A case study the application of First order uncertainty analysis is described.

UNCERTAINTY ANALYSIS OF DISSOLVED OXYGEN MODEL USING FOA

First order uncertainty analysis of the basic Streeter Phelps Dissolved Oxygen Sag equation is presented in this section. The relationship among the parameters affecting the in-stream dissolved oxygen concentration is given by Equation 1:

$$D = \frac{K_d L_O}{K_a - K_d}\left(e^{-K_d t} - e^{K_a t}\right) + D_O e^{-K_a t} \tag{15}$$

in which D = dissolved oxygen deficit (C_s - C) in milligrams per liter; C_s = dissolved oxygen saturation limit, in milligrams per liter; C = dissolved oxygen concentration, in milligrams per liter K_a = reaeration rate coefficient (base e, per day); K_d = deoxygenation rate coefficient (base e, per day); L_o = initial instream total ultimate biochemical oxygen demand, in milligrams per liter, D_o = initial instream dissolved oxygen deficit, in milligrams per liter; and t = time of travel from D_o to D (days).

Basically the first order uncertainty analysis will provide a measure of the uncertainty of the dependent variable, D in terms only of the uncertainty in the dependent variables: K_a, K_d, L_o, D_o and t; i.e., percentage of the scatter of dissolved oxygen deficit predictions around the true deficit at any point along the sage curve can be assigned to reach of the independent variables.

Taking the partial derivative of D with respect to each of the independent variables.

$$\frac{\partial D}{\partial D_o} = e^{-k_a t} \tag{16}$$

$$\frac{\partial D}{\partial D_o} = \frac{K_d}{K_a - K_d} (e^{-k_a t} - e^{-k_a t}) \tag{17}$$

$$\frac{\partial D}{\partial K_a} = \frac{K_d L_o}{(K_a - K_d)^2} (e^{-k_a t} - e^{-k_a t}) + \frac{K_d L_o t e^{-k_a t}}{(K_a - K_d)^2} - D_o t e^{-k_a t} \tag{18}$$

$$\frac{\partial D}{\partial K_d} = \frac{K_d L_o}{(K_a - K_d)^2} (e^{-k_a t} - e^{-k_a t}) - \frac{K_d L_o t e^{-k_a t}}{K_a - K_d} \tag{19}$$

$$\frac{\partial D}{\partial t} = \frac{K_d L_o}{K_a - K_d} (K_a e^{-k_a t} - K_d e^{-k_a t}) - K_d D_o e^{-k_a t} \tag{20}$$

The first-order approximation to the total uncertainty in the dissolved oxygen deficit is given by Eq. 21.

$$S_D = (\sum_{i=1}^{5} |C_t|^2)^{1/2} \tag{21}$$

The terms of Eq. 21 are defined by Eq. 22.

$$C_1 = \frac{\partial D}{\partial D_o} S_{D_o}; C_2 = \frac{\partial D}{(\partial L_o)} S_{L_o}; C_3 = \frac{\partial D}{\partial K_d} S_{K_a}$$

$$C_4 = \frac{\partial D}{\partial K_d} S_{K_d}; \frac{\partial D}{\partial t} S_t \tag{22}$$

For Eqs. 21 and 22, the symbol S represents the standard deviation of particular variable. Thus, Eq. 21 shows that each of the independent variables contributes to the dispersion of D in a manner proportional to its own variance, S^2, and proportional to a factor which is related to the sensitivity of changes in D to changes in the independent variable.

Application of the method of first order uncertainty analysis to the BOD-DO system requires estimates of mean parameter values and standard deviations of D_o, L_o, K_d, K_a and t.

DATA SELECTION AND DEVELOPMENT

The independent variables of Eq. 15 are subject to wide variations from stream to stream or even from reach to reach of the same stream. Global or even regional mean values and standard deviations would be difficult to estimate and would probably not permit useful conclusions to be drawn. Therefore, because of the nature of the data requirements for the uncertainty analysis, calculations were restricted to particular, individual stream segments. However, a wide range of parameters values and combinations was selected to permit study of the relative importance of the variables, for a variety of stream conditions. The steps used in the development of the basic data are outlined below:

1. Four classes of streams, defined by the self-purification ratio, $f = K_a / K_d$, were selected.
2. A range of reaction rate coefficients was selected for each of four stream classes based on in-stream depth and velocity. Formulas considered for reaction were from the literature. For the data range examined, the O'Connor-Dobbins equation proved to be most relevant for stream classes, 1, 2 and 3 while that of Churchill, et al. was best suited for stream class 4.
3. Subsequently, deoxygenation rate constants were calculated for each pair of K_a. The range of data for the variables used in the analysis is shown in Table 1.

TABLE 1 - Ranges of Data by Stream class

Stream-Class description	f (dimensionless)	K_a (per day)	K_d (per day)	V (meter/sec.)	H (meters)
Sluggish	1.25-1.50	0.05-0.10	0.033-0.08	0.03-.015	3.05-6.10
Low-velocity	1.50-2.00	0.10-1.00	0.050-0.67	0.03-0.15	0.92-3.05
Moderate-velocity	2.00-3.00	1.00-5.00	0.500-2.50	0.15-0.61	0.61-1.52
Swift	3.00-5.00	1.00-10.0	0.200-3.33	0.61-1.83	0.61-3.05

4. The initial dissolved oxygen deficits were defined. A maximum allowable dissolved oxygen DO deficit of 4 mg/L was selected to permit a wide range of initial deficits. D_o values of 0, 1, 2 and 3 mg/L were assigned to each stream segment defined by a pair of K_a and K_d values derived previously.
5. With reaction coefficients defined, the maximum allowable instream ultimate oxygen demand can be calculated from Eq. 15. A listing of allowable value L_o in milligrams per liter for the variable range examined is given in Table-2.

TABLE 2 - Allowable Ultimate BOD Loading Rates as Function of Initial DO Deficit

Self-purification ratio 'f'	Values for Initial Dissolved Oxygen Deficit, D_o, in milligram per liter			
	0	1	2	3
1.25	12.0	11.2	10.0	8.0
1.50	14.0	12.4	11.2	9.6
2.00	16.0	15.2	14.0	12.0
3.00	20.0	20.0	18.4	16.4
5.00	30.0	29.6	28.0	25.2

6. In all, a set consisting of 80 hypothetical stream segments defined by values for K_a, K_d, D_o and L_o was constructed. The time of travel to the point of maximum DO deficit, the critical time t_c was calculated for each of these.
7. To eliminate effects of possible interdependence among the basic data of the 80 original stream segments, 160 additional segments were developed. Eighty segments were obtained by increasing all of the L_o values by 50% while all other parameters were held constant. Another 80 segments were derived by increasing of t values by a factor of two of three while keeping all other parameters constant. Thus, a total of 240 hypothetical stream segments was included in the uncertainty analysis. An example of these data for some representative segments is presented in Table 3.
8. To complete the data requirements for the uncertainty analysis, knowledge about the distribution of parameter values for each stream segment was needed. This information must take the form of a standard deviation for each of the five independent parameters. In practice, this information is probably difficult, if not impossible, to develop without extensive monitoring. However, for the hypothetical case studied, the required standard deviations were derived from published sources. Standard deviations of + 1.0 mg/L for L_o and + 0.2 mg/L for D_o were assumed to be the same as for the laboratory analysis of these parameters. Churchill, et. al., (1962) indicated a possible range of 15% error for their equation. Therefore, to allow for a somewhat wider scatter in this analysis, an uncertainty of + 20% was assumed for both K_a and K_d. An estimate for the standard deviation of travel time is elusive. S_t was assumed to be + 10% of (Carter and Anderson 1963). Fortunately the analysis to follow will demonstrate that travel time estimates do not contribute a significant portion of the total scatter in D.

The respective estimates of parameter standard deviations are included with the basic data and in Table 3.

CALCULATION OF UNCERTAINTY MATRIX

For each line of data in Table-3, which represents a set of mean parameter values for each stream segment, a corresponding set of five partial derivatives was calculated by Eqs. 16-20. The values of the partial derivatives were multiplied by the appropriate standard deviation of the respective parameters, also from Table-3 as described by Eq. 22. Eq. 21 then gave the total uncertainty of dissolved oxygen deficit for each line of data from Table 3.

TABLE 3 : Representative Stream Segment Data for Uncertainty Analysis

Segment number	K_d per day	K_o per day	L_o mg/L	D_o mg/L	t_c days	t days	S_{kd} per day	S_{ka} per day
(a) Stream Class 1								
1	0.08	0.01	12.0	0.0	11.2	1.0	0.016	0.02
2	0.08	0.10	11.2	1.0	10.0	1.0	0.016	0.02
3	0.08	0.10	10.0	2.0	8.6	1.0	0.016	0.02
4	0.08	0.10	8.0	3.0	6.2	1.0	0.016	0.02
(b) Stream Class 2								
5	0.50	1.00	16.0	0.0	1.4	0.8	0.100	0.20
6	0.50	1.00	15.2	1.0	1.2	0.8	0.100	0.20
7	0.50	1.00	14.0	2.0	1.1	0.8	0.100	0.20
8	0.50	1.00	12.0	3.0	0.8	0.8	0.100	0.20
(c) Stream Class 3								
9	1.67	5.00	20.8	0.0	0.33	0.20	0.334	1.00
10	1.67	5.00	20.0	1.0	0.30	0.20	0.334	1.00
11	1.67	5.00	18.4	2.0	0.26	0.20	0.334	1.00
12	1.67	5.00	16.4	3.0	0.19	0.20	0.334	1.00
(d) Stream Class 4								
13	3.33	10.0	20.8	0.0	0.17	0.10	0.666	2.00
14	3.33	10.0	20.0	1.0	0.15	0.10	0.666	2.00
15	3.33	10.0	18.4	2.0	0.13	0.10	0.666	2.00
16	3.33	10.0	16.4	3.0	0.10	0.10	0.666	2.00

The contributions from the five independent parameters to total uncertainty of dissolved oxygen deficit were readily calculated for the 240 stream segments. Percentage contributions by variable to the total uncertainty are given in Table 4 for each stream segment shown in Table 3. The relative importance of the respective independent parameters to the prediction of dissolved oxygen deficits is easily seen by inspection.

ANALYSIS OF THE UNCERTAINTY MATRIX

Examination of Table 4 reveals some general patterns as characterised by percentage contributions to uncertainty. The parameters of initial deficit, D_o, and deoxygenation coefficient, K_d, are shown to contribute most to uncertainty in the dissolved oxygen deficit for Type 1 streams. For stream Types 2 through 4, K_d is of primary importance, although the reaeration coefficient, K_a, is seen to exhibit a more significant role, especially at higher levels of the instream initial deficit. The significance of the initial deficit suggests the importance of post-aeration of waste waters prior to release to the stream.

Manipulations on insteam ultimate biochemical oxygen demand (UBOD) values, L_o, demonstrated only minor effects on the uncertainty analysis by slightly increasing the importance of the parameter

K_d. Variations of the travel time, t, imposed a shift in parameter importance as a function of whether or not t exceeded the time of flow, t_c, to the point of maximum deficit. The major shift from K_d to K_a as the most influential parameter in the uncertainty of DO deficit predictions was caused by increasing travel times beyond the critical travel time. Thus, the most important parameter of the BOD-DO equation is seen to correspond to the predominant aspect of the oxygen balancing mechanism for the portion of the DO sag curve where deoxygenation exceeds reaeration, the uncertainty analysed shows that K_d is the most influential parameter, generally. However, for the part of the sag curve where reaeration exceeds deoxygenation, K_a is generally the most influential parameter from the uncertainty analysis.

Similar responses were observed when analyses were made for situations using $S_{kd}=41\%$; $S_{ka}=35\%$; and $S_{kd}=41\%$; and $S_{ka}=10\%$, respectively. The latter cases were examined to address the situation for which empirical formulas are used to estimate the deoxygenation and reaeration coefficients for a particular stream. Standard deviations for such prediction have been given for the deoxygenation coefficient and for the reaeration coefficient.

TABLE 4 - Percentage Contributions to Uncertainty in DO Deficit (S^2_D)

Segment number	Percent of S_D^2 Contributed By:				
	D_o	L_o	K_a	K_d	t
(a) Stream Class 1					
1	44	7	0	40	9
2	47	8	1	38	6
3	52	8	3	33	4
4	59	10	6	24	1
(b) Stream Class 2					
5	2	10	15	69	4
6	2	10	22	64	2
7	2	11	31	56	1
8	2	12	41	45	0
(c) Stream Class 3					
9	1	6	20	70	3
10	1	6	27	64	2
11	1	6	36	56	0
12	1	6	46	47	0
(d) Stream Class 4					
13	1	6	19	70	3
14	1	6	27	64	2
15	1	6	36	57	0
16	1	6	46	47	0

SUMMARY OF STEPS IN AN UNCERTAINTY ANALYSIS

1. Estimate the best parameter values for the situation being modeled.
2. Define the correlation structure among the input parameters.
3. Estimate the variances of the input parameters.
4. Conduct a sensitivity analysis and select parameters to be used in the uncertainty analysis.

FOA

5. Conduct a FOA to estimate the variances for the model inputs of interest.
6. Calculate the fraction of the total variance attributable to each parameter.
7. Assume a pdf for the model output and calculate confidence intervals.

MCS

8. Determine the pdfs appropriate for the input parameters.
9. Perform a MCS.
10. Calculate the correlation matrix and a multiple regression on the MCS results.
11. Plot the MCS generated outputs as probability plots.
12. Determine appropriate pdfs for the outputs.
13. Determine confidence intervals on the outputs.

REFERENCES

1. Haan, C.T., 1977. *Statistical Methods in Hydrology.* Iowa State University Press, Ames.

2. Helsel, D.R., and R.M. Hirsch, 1992. *Statistical Methods in Water Resources.* Elseveier, New York.

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

::: ::::: :::