

ASSESSMENT OF PROBABILITY
DISTRIBUTION OF DISSOLVED OXYGEN
DEFICIT

Yeou-Koung Tung
Wade E. Hathhorn

1988

Journal Article
WWRC-88-24

In

Journal of Environmental Engineering

Volume 114, No. 6

December 1988

Yeou-Koung Tung
Wyoming Water Research Center
and Department of Statistics

Wade E. Hathhorn
Department of Civil Engineering

University of Wyoming
Laramie, Wyoming

V = vertical flow velocity through liner flaw;
 y = distance from middle plane;
 Γ = gamma function;
 η = absolute viscosity of liquid;
 θ = angular coordinate;
 ν = kinematic viscosity of liquid;
 ρ = density of liquid; and
 τ = viscous shear at plane at distance y from middle plane.

ASSESSMENT OF PROBABILITY DISTRIBUTION OF DISSOLVED OXYGEN DEFICIT

By Yeou-Koung Tung,¹ Associate Member, ASCE,
and Wade E. Hathhorn,² Student Member, ASCE

ABSTRACT: Water quality modeling and prediction in a stream environment is complicated by the occurrence of a number of random processes. Due to the complexity of water quality model, the analytical derivation of the exact probability distribution of the dissolved oxygen (DO) deficit is difficult. This paper proposes an approximate but practical method by using first-order analysis of uncertainty in estimating the statistical moments of the DO deficit. The statistical moments estimated were then used in an appropriate probability distribution for the DO deficit concentration. Practical issues involved are: (1) Which is the appropriate probability distribution for the DO deficit concentration? and (2) is this appropriate distribution function sensitive to the distribution of water quality parameters? This paper examines a number of commonly used probability distributions for their appropriateness in describing the random characteristics of the DO deficit concentration under various conditions. Of the distributions considered in the investigation, lognormal distribution proved to be the best model.

INTRODUCTION

A major portion of the complexity associated with water quality modeling and prediction is the inherent randomness exhibited throughout the stream environment. Not only are the physical and biological processes not clearly defined, but an imposing number of uncertainties are associated with the various processes occurring within the stream environment. Several researchers have already attempted to analyze these uncertainties. For example, Loucks and Lynn (1966) investigated the effect of inherent uncertainty due to the natural variations in streamflow and waste flow on the probability distribution of dissolved oxygen (DO); Padgett and Rao (1979) presented a joint probability distribution for biochemical oxygen demand (BOD) and DO; and Kothandaraman and Ewing (1969) and Chadderton et al. (1982) have investigated the stochastic nature of the model parameters in the Streeter-Phelps equation (Streeter and Phelps 1925).

Realizing the existence of such uncertainties in water quality modeling, the prediction of DO deficit or DO concentration within a given reach of stream is no longer a deterministic exercise. Rather, the DO deficit must be treated as a random variable. Fig. 1 is a schematic sketch illustrating the probability distribution for the DO deficit due to the presence of inherent randomness of water quality parameters. In probabilistic water quality analysis, it is typical to deal with the problem of assessing the probability of water quality violation. To perform such probability computations, knowledge about the statistical properties and the distribution of water quality in-

¹Assoc. Prof., Wyoming Water Res. Ctr. and Statistics Dept., Univ. of Wyoming, Laramie, WY 82071.

²Former Grad. Student, Dept. of Civ. Engrg., Univ. of Wyoming, Laramie, WY 82071.

Note. Discussion open until May 1, 1989. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on January 15, 1988. This paper is part of the *Journal of Environmental Engineering*, Vol. 114, No. 6, December, 1988. ©ASCE, ISSN 0733-9372/88/0006-1421/\$1.00 + \$.15 per page. Paper No. 22990.

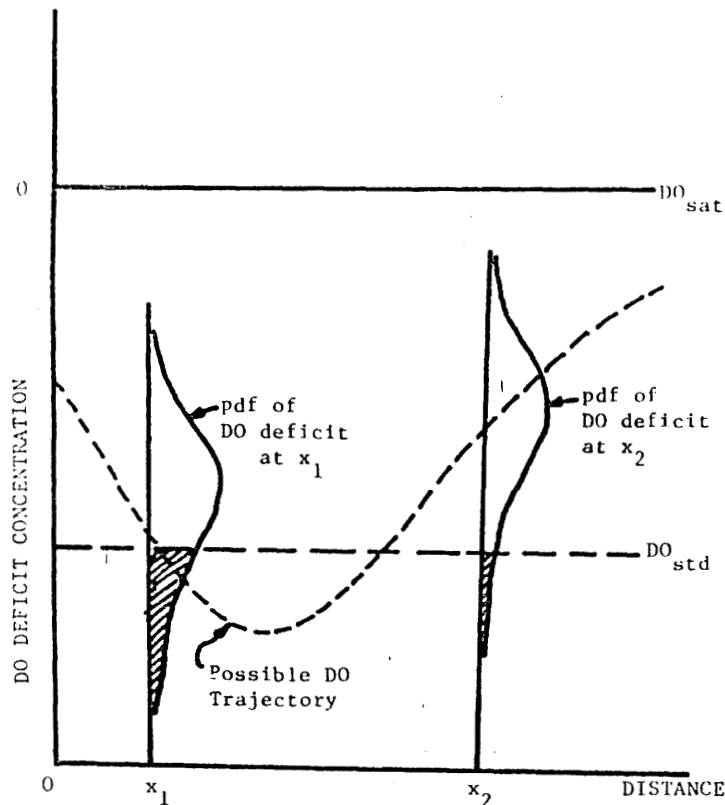


FIG. 1. Schematic Sketch of Probability Density Function (PDF) for DO Deficit Concentration

deficit must be known. However, in most cases, the exact distribution of the DO deficit is not known and is often assumed without any basis.

Although significant research has been conducted in the uncertainty analysis of stream DO, most of these studies have been concerned with the variability of DO concentrations due to model parameter uncertainty (Kothandaraman and Ewing 1969; Hornberger 1980; Chadderton et al. 1982). However, there have been some attempts to derive analytical expressions for the exact probability distribution associated with the DO deficit. Thayer and Krutchkoff (1967) utilized a stochastic birth and death process to obtain an expression for the probability distribution of DO concentration without considering the uncertainties of the model parameters. Esen and Rathbun (1976) assumed the reaeration and deoxygenation rates to be normally distributed and investigated the probability distribution for DO and BOD using a random walk approach. Notably, Padgett et al. (1977) developed a joint probability density function for the BOD and DO concentrations by solving a stochastic differential equation, and Padgett and Rao (1979) later developed a non-parametric probability density function of the BOD and DO.

From a practical viewpoint, the main disadvantage of each of the aforementioned methods is that the resulting probability distributions derived for the DO deficit are very complicated. The required mathematical skills needed for such sophisticated approaches would make it difficult for most engineers to apply. Furthermore, all these analytically derived probability distribution functions for the DO can only be obtained by using very simple distributions for the model parameters such as uniform or normal. When distributions other than those simple ones are used to describe the randomness of water quality parameters (which could well be the case in reality), the analytical derivation of a probability distribution for the DO would be extremely difficult, if not impossible.

Another approach that is frequently applied by engineers is the Monte Carlo simulation. The method has recently been incorporated into the enhanced QUAL2 model, called QUAL2-UNCAS, by Brown and Barnwell (1987). This brutal force enumeration scheme requires a large number of repetitions, which could be computationally expensive. Of course, with the advent of computing power and efficiency of computers, the weight of such concern will be gradually diminishing. However, at the present time, the computation efficiency and cost remain an important concern in practical engineering problem solving.

In support of a more tractable methodology, this paper examines an approximate approach to probabilistic water quality analysis in that the statistical moments of the DO deficit are estimated by the first-order analysis. The statistical moments estimated are then incorporated with an appropriate probability distribution model for the DO deficit. However, practical issues that can be raised are: (1) Which is an appropriate probability model for the DO deficit? and (2) if there is one such probability model, how sensitive is it to the distribution of water quality parameters? Theoretically, the exact probability distribution of the DO deficit should change if the distribution of water quality parameters is changed. However, from a practical viewpoint, it is worthwhile to investigate the appropriateness of some commonly used probability models in describing the random characteristics of the DO deficit computed by the Streeter-Phelps equation. Because the distribution of the DO deficit may in theory be dependent on the statistical properties of the water quality parameters themselves, this paper also examines the sensitivity of these parameters on the selection of the appropriate probability model for the DO deficit. The candidate probability distributions included in the study were normal, two-parameter lognormal, two-parameter gamma, and Weibull distributions. To characterize these distributions completely, various statistical parameters must be known. To do this, the mean and variance of the DO deficit were estimated using first-order uncertainty analysis. This information was then used to compute the appropriate statistical parameters for each of the candidate probability models using the moment-parameter relationships that can be found in Hastings and Peacock (1974) and Patel et al. (1976).

In addition, rather than making such a strong assumption about the probability density function of any specific form for the DO deficit, an approach of using Fisher-Cornish asymptotic expansion of any continuous distribution (Fisher 1950; Fisher and Cornish 1960; Stuart and Ord 1987) was also applied as an alternative for comparison. This method relates the quantiles of standardized distribution of the DO deficit to the standard normal quantiles

and higher order moments such as skew coefficient and kurtosis of the DO deficit. As will be shown later, the results of numerical investigations indicated that the Fisher-Cornish asymptotic expansion does not yield any superior performance over its parametric probability distribution competitors considered herein. Furthermore, the approach requires the evaluation of higher statistical moments, which makes it computationally cumbersome in engineering analysis, especially when one deals with highly nonlinear equations involving uncertain parameters such as those encountered in water quality modeling. Nevertheless, it is important to note these results for the Fisher-Cornish asymptotic expansion so that others may become aware of the disadvantages of its use under similar circumstances.

FIRST-ORDER ANALYSIS OF UNCERTAINTY

The use of first-order uncertainty analysis is popular in all fields of engineering because of its relative ease in application to a wide array of problems. The detailed theory and mathematics of first-order uncertainty analysis can be found in Benjamin and Cornell (1970) and Cornell (1972). As an example of such use in the water quality field, Burges and Lettenmaier (1975) have utilized the method to investigate the uncertainty in predictions of BOD and DO within a stochastic stream environment.

Essentially, first-order analysis provides a methodology for obtaining an estimate for the moments of a random variable which is a function of one or several random variables. It estimates the uncertainty in a mathematical model involving parameters which are not known with certainty. By using first-order analysis, the combined effect of uncertainty in a model formulation, as well as the use of uncertain parameters, can be assessed.

First-order uncertainty analysis is characterized by two major components: (1) Single moment (variance) treatment of the random variables; and (2) the use of first-order approximation of any functional relationship (e.g., the use of Taylor's series expansion). The first component implies that the random element of any variable is defined exclusively by its first nonzero moment or simply the variance of the random variable itself. Thus, information pertaining to the character of a random variable is provided solely by its mean and variance.

The second component states that only the first-order terms in a Taylor's expansion will be utilized in the analysis of a functional relationship containing random variables or processes. With exception of the evaluation of the mean (in which second-order terms may be included for the purpose of accounting for correlation among variables), any attempt to retain terms higher than first-order in the expansion requires more information about the random variables than those provided by their first and second moments (Cornell 1972).

To present the general methodology of first-order analysis, consider a random variable, Y , which is a function of N random variables (multivariate case). Mathematically, Y can be expressed as

$$Y = g(\mathbf{X}) \dots \dots \dots (1)$$

where $\mathbf{X} = (X_1, X_2, \dots, X_N)$, a vector containing N random variables X_i . Through the use of Taylor's expansion, the random variable Y can be approximated by

$$Y \hat{=} g(\bar{\mathbf{X}}) + \sum_{i=1}^N \left[\frac{\partial g}{\partial X_i} \right]_{\bar{\mathbf{X}}} (X_i - \bar{X}_i) + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left[\frac{\partial^2 g}{\partial X_i \partial X_j} \right]_{\bar{\mathbf{X}}} (X_i - \bar{X}_i)(X_j - \bar{X}_j) \dots \dots (2)$$

in which $\bar{\mathbf{X}} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N)$, a vector containing the mean of N random variables; and $\hat{=}$ represents equal in the sense of a second-order approximation.

To consider the correlation among random variables X_s , the second-order approximation of the mean (the expected value) of random variable Y is

$$\mu_Y = E[Y] \hat{=} g(\bar{\mathbf{X}}) + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left[\frac{\partial^2 g}{\partial X_i \partial X_j} \right]_{\bar{\mathbf{X}}} \text{cov} [X_i, X_j] \dots \dots \dots (3)$$

in which $\text{cov} [X_i, X_j]$ is the covariance between random variables X_i and X_j . It should be noted that the second term in Eq. 3 can be dropped if the random variables X_i are uncorrelated. In such a case, the resulting equation is the same as the first-order approximation.

It follows that the first-order approximation of the variable of Y is

$$\sigma_Y^2 = \text{var} [Y] \hat{=} \sum_{i=1}^N \sum_{j=1}^N \left[\frac{\partial g}{\partial X_i} \right]_{\bar{\mathbf{X}}} \left[\frac{\partial g}{\partial X_j} \right]_{\bar{\mathbf{X}}} \text{cov} [X_i, X_j] \dots \dots \dots (4)$$

If the X_s are uncorrelated, Eq. 4 reduces to

$$\sigma_Y^2 \hat{=} \sum_{i=1}^N \left[\frac{\partial g}{\partial X_i} \right]_{\bar{\mathbf{X}}}^2 \sigma_i^2 \dots \dots \dots (5)$$

where $\hat{=}$ means equal in a first-order sense; and $\sigma_i^2 =$ the variance corresponding to random variable X_i .

BASIC WATER QUALITY MODEL

To better understand the stream environment, several mathematical models have been developed to describe the interaction between the physical and biological processes occurring within the stream. The most well-known expression of this type is the Streeter-Phelps equation (Streeter and Phelps 1925). In differential form, the equation is given as:

$$\frac{dD}{dt} = K_d L - K_a D \dots \dots \dots (6)$$

The solution to Eq. 6, replacing t by x/U , is:

$$D_x = \frac{K_d L_0}{K_a - K_d} \left[\exp \left(-\frac{K_d x}{U} \right) - \exp \left(-\frac{K_a x}{U} \right) \right] + D_0 \exp \left(-\frac{K_a x}{U} \right) \dots \dots (7)$$

where $K_d =$ the deoxygenation coefficient (1/day); $K_a =$ the reaeration coefficient (1/day); $x =$ the distance downstream from the beginning of the reach (mi or km); $U =$ the average stream velocity (mi/day or km/day); $D_x =$ the DO deficit concentration (mg/L) within the reach at a downstream distance x ; $D_0 =$ the initial DO deficit (at distance $x = 0$); and $L_0 =$ the initial in-stream BOD concentration (both in mg/L).

Before the discussion proceeds, it is worthwhile to briefly describe the nature of the water quality model used here. It should be pointed out that

Eq. 6 is limited to only two in-stream processes: (1) Deoxygenation of the water due to bacterial decomposition of carbonaceous organic matter; and (2) atmospheric reoxygenation directly proportional to the DO deficit. Moreover, several assumptions have been made in the development of this model: (1) Steady, uniform flow with constant temperature; (2) DO deficits predicted by Eq. 7 are one-dimensional, i.e., a function only of the position downstream from a discharge point; and (3) the rate of biodegradation and reoxygenation, expressed by K_d and K_a , are described by first-order kinetics.

Since its conception, the original Streeter-Phelps equation has been modified to account for discrepancies between analytical estimations, computed from Eq. 7, and actual data collected in the field. These discrepancies have arisen as a result of the exclusion of a number of oxygen sources and sinks in the original equation. Dobbins (1964) pointed out eight other possible factors that could contribute to instream BOD and DO variations.

There have been several studies conducted in which one or more of these additional processes have been included in the model formulation in an attempt to improve model predictability (Dobbins 1964; Hornberger 1980; Krenkel and Novotny 1980). In general, these modifications can be made by simply adding terms to Eq. 7 to account for the various additional factors. Although aware of the additional oxygen source and sink terms, the writers of this paper have selected, without loss of generality, the use of the original Streeter-Phelps equation in order to simplify the algebraic manipulation.

UNCERTAINTY ANALYSIS OF THE WATER QUALITY MODEL

A detailed discussion was given by Hathorn (1986) as to the type of uncertainties encountered in the modeling of the water quality process, including inherent, model, and parameter uncertainties. Knowing the existence of these uncertainties in the aquatic environment, such conditions should be incorporated into the modeling process in order to improve model accountability. In this study, first-order uncertainty analysis was utilized in which water quality parameters K_d , K_a , U , L_0 , and D_0 , as well as D_x in Eq. 7, are treated as random variables.

Based on the first-order analysis described in the previous section, the expected value of the DO deficit given by Eq. 7 at any location can be expressed as:

$$E[D_x] \approx D_x(\bar{K}_d, \bar{K}_a, \bar{U}, \bar{L}_0, \bar{D}_0) + \frac{1}{2} \text{var}[K_d]P''_{K_d, K_d} + \frac{1}{2} \text{var}[K_a]P''_{K_a, K_a} + \frac{1}{2} \text{var}[U]P''_{U, U} + \frac{1}{2} \text{var}[L_0]P''_{L_0, L_0} + \frac{1}{2} \text{var}[D_0]P''_{D_0, D_0} + \text{cov}[K_a, U]P''_{K_a, U} \dots (8)$$

in which $\text{var} [\]$ and $\text{cov} [\]$ represent the variance and covariance operators, respectively; $P'_x = \partial D_x / \partial X_j$ and $P''_{x,y} = \partial^2 D_x / \partial X_j \partial X_k$ evaluated at the mean of the model parameters (K_d , K_a , U , L_0 , and D_0). Detailed expression of P' and P'' of D_x were given by Hathorn (1986). It should also be noted that provisions for a positive correlation between parameters K_a and U were included in Eq. 8. Some investigators have considered a positive correlation between K_d and K_a (Esen and Rathbun 1976; Padgett 1978). Although statistical analysis of a given data set may reveal a correlation between these parameters,

it does not necessarily imply such a correlation has any meaningful physical representation of the system behavior. It has been known that K_d is a function of the physical characteristic of the stream, while K_a is characterized by the biological composition of the waste discharge and stream environment. It was the opinion of the writers that the correlation between K_d and K_a is spurious and, thus, was not considered in the study. Similarly, by Eq. 4 the variance of the DO deficit can be approximated as

$$\sigma_{D_x}^2 = \text{var}[D_x] \approx (P'_{K_d})^2 \text{var}[K_d] + (P'_{K_a})^2 \text{var}[K_a] + (P'_U)^2 \text{var}[U] + (P'_{L_0})^2 \text{var}[L_0] + (P'_{D_0})^2 \text{var}[D_0] + 2P'_{K_a} \cdot P'_U \text{cov}[K_a, U] \dots (9)$$

Expressions for the higher-order moments such as the skewness and kurtosis of the DO deficit can be derived with these same procedures under the assumption of independency between the various water quality parameters. The existence of correlation among water quality parameters, however, makes the evaluation of the higher-order moments difficult due to the lack of information about the cross-product moments of various orders in real-world problems. Fortunately, the skewness and kurtosis of the DO deficit are not required for the four parametric probability models considered in the study.

On the other hand, the Fisher-Cornish approach requires such information to define the shape of the DO deficit distribution. However, in the analysis that follows, the Fisher-Cornish method, with its moments estimated by first-order analysis, produced results that were no more superior than other parametric competitors. Therefore, the expressions for the skewness and kurtosis of the DO deficit (available from Hathorn 1986) are not presented here.

PERFORMANCE EVALUATION OF THE DISTRIBUTIONS FOR DO DEFICIT

The idea of applying first-order analysis for estimating the first two statistical moments of the DO deficit, along with a selection of an appropriate probability model for the DO deficit, is straightforward and practical. However, among the various probability models that are commonly used, a practical question to be raised is, "Which probability model (or models) best describes the random behavior of the DO deficit in a stream?" In the following investigation, four parametric (i.e., normal, lognormal, gamma, and Weibull) and one other (i.e., Fisher-Cornish asymptotic expansion) probability distributions have been selected as candidates such that a wide spectrum of shapes are represented.

To evaluate the relative performance of each of the five candidate probability distributions considered, three performance criteria are adopted herein: (1) Biasness (BIAS); (2) mean absolute error (MAE); and (3) root mean squared error (RMSE). These criteria are used simultaneously in an attempt to identify the best probability model. These criteria are mathematically defined as

1. Biasness:

$$\text{BIAS} = \int_0^1 (x_p - x_p) dp \dots (10)$$

2. Mean absolute error:

$$\text{MAE} = \int_0^1 |\hat{x}_{p,f} - x_p| dp \dots\dots\dots (11)$$

3. Root mean squared error,

$$\text{RMSE} = \left[\int_0^1 (\hat{x}_{p,f} - x_p)^2 dp \right]^{0.5} \dots\dots\dots (12)$$

in which x_p = the true value of the p th order quantile of the DO deficit; and $\hat{x}_{p,f}$ = the estimate of x_p determined from the assumed probability model, f , with its mean and variance estimated by the first-order analysis.

ESTIMATION OF THE TRUE DISTRIBUTION OF D BY MONTE CARLO SIMULATION

It was assumed in this study that the DO deficit at any downstream location x can be computed using Eq. 7. Moreover, when examining Eq. 7 to a stream environment under uncertainty, the model parameters (K_d , K_a , U , L_0 , and D_0) were to be treated as random variables, each characterized by their own probability distribution. In turn, the DO deficit computed by Eq. 7 was itself a random variable characterized by its own distribution.

To determine the probability distribution of the DO deficit at a given location, Monte Carlo simulation technique (Rubinstein 1981) was employed, allowing each of the model parameters to be assigned one of four distributions: normal, lognormal, gamma, and Weibull. In conducting the Monte Carlo procedure, the statistical properties of the model parameter used throughout this study are listed in Table 1. This data set represents a stream environment described as a low velocity system according to the classification made by Chadderton et al. (1982). The difference of the data set used here and that of Chadderton et al. (1982) is that the coefficient of variation for the water quality parameters is increased from 0.2 to 0.3.

Simulation procedures were performed such that 10 groups of 999 DO deficits were generated using Eq. 7 with one of the four distributions previously mentioned for each of the model parameters. For example, during the first simulation (first row of Table 2), 10 groups of 999 DO deficits (using Eq. 7) were generated under the conditions of independence and normality for all the water quality parameters. Then, in successive simulations,

TABLE 1. Statistical Properties of Model Parameters Used to Investigate Distribution of Dissolved Oxygen Deficit and Critical Location

Model parameters (1)	Mean (2)	Standard deviation (3)	Units (4)
K_d	0.35	0.10	days ⁻¹
K_a	0.70	0.20	days ⁻¹
U	10.00	3.00	mi/day
L_0	18.00	5.00	mg/L
D_0	1.00	0.30	mg/L

Note: 1 mi = 1.6 km.

TABLE 2. List of Cases Using Various Distributions for Water Quality Parameters in Monte Carlo Simulation

Case number (1)	Distributions Assumed for Model Parameters					
	K_d (2)	K_a (3)	U (4)	L_0 (5)	D_0 (6)	$\rho(K_a, U)$ (7)
1	N	N	N	N	N	0.0
2	N	N	N	N	N	0.8
3	LN	LN	LN	LN	LN	0.0
4	LN	LN	LN	LN	LN	0.8
5	G	G	G	G	G	0.0
6	W	W	W	W	W	0.0
7	LN	N	G	W	LN	0.0
8	G	LN	W	N	G	0.0
9	LN	N	N	W	G	0.8
10	N	LN	LN	G	W	0.8

Note: N = normal; LN = lognormal; G = gamma; W = Weibull.

different distributions were assigned to each of model parameters. In the Monte Carlo simulation, a positive correlation ($\rho = 0.8$) between model parameters K_a and U was used in this study. When such a correlation was considered in the simulation, a bivariate normal or lognormal distribution was utilized for both parameters K_a and U .

During the simulations, each of the 10 groups of 999 DO deficits were ranked in ascending order. Specifically, the minimum value of the DO deficit generated was assigned to position 1 and the maximum value to position 999. Then, quantiles of the DO deficit were computed for several probability levels p by simply locating the value of the deficit in position $(999 + 1)p$. In order to reduce the sampling errors, each of the respective quantiles obtained for the 10 groups were then arithmetically averaged.

In the evaluation of the three performance criteria given in Eqs. 10, 11, and 12, numerical integration was performed at $p = 0.01, 0.025, 0.05, 0.1, 0.15, 0.2(0.1)0.7, 0.75(0.05)0.95, 0.975, \text{ and } 0.99$, where the estimated quantities, $\hat{x}_{p,f}$, were computed from the assumed distribution models.

RESULTS

The analysis of the goodness-of-fit performance criteria as described for the probability model of the DO deficit was conducted at downstream locations of 5 and 10 mi (8 and 16 km). These points represent locations on each side of the critical point found using the mean water quality parameters of Table 1.

The various cases considered in this study are shown in Table 2. In cases 1-6, the type of probability distributions assigned to the water quality parameters remained unchanged within each case. By contrast, these distributions were arbitrarily assigned in cases 7-10. The reason for considering this number of cases (although not exhaustively enumerated) was to examine the sensitivity of the selection of a probability distribution for the water quality parameters on the best-fit distribution obtained for the DO deficit. The

TABLE 3. Biasness for DO Deficit between Simulation Results and Various Assumed Distributions at Downstream Locations

Case number (1)	Biasness (in mg/L)				
	N (2)	LN (3)	G (4)	W (5)	FC (6)
(a) At 5 mi (8 km)					
1	0.416	0.407	0.413	0.415	0.416
2	0.389	0.379	0.385	0.387	0.389
3	0.421	0.412	0.417	0.420	0.417
4	0.421	0.412	0.418	0.420	0.419
5	0.413	0.405	0.410	0.412	0.411
6	0.411	0.403	0.408	0.410	0.417
7	0.392	0.383	0.389	0.391	0.394
8	0.415	0.406	0.411	0.414	0.412
9	0.383	0.374	0.380	0.382	0.385
10	0.418	0.409	0.415	0.417	0.418
(b) At 10 mi (16 km)					
1	0.593	0.580	0.587	0.591	0.593
2	0.561	0.547	0.557	0.558	0.561
3	0.560	0.547	0.554	0.557	0.554
4	0.568	0.554	0.564	0.565	0.563
5	0.560	0.547	0.554	0.558	0.556
6	0.571	0.558	0.565	0.568	0.580
7	0.537	0.524	0.531	0.535	0.541
8	0.583	0.570	0.577	0.580	0.581
9	0.543	0.529	0.539	0.540	0.547
10	0.567	0.553	0.563	0.564	0.565

Note: N = normal; LN = lognormal; G = gamma; W = Weibull; and FC = Fisher-Cornish.

results of a number of these experiments are given in Tables 3-5.

When examining the biasness presented for each of the cases in Table 3, it was observed that the use of a two-parameter lognormal distribution for the DO deficit consistently yielded the lowest reported value. All probability distributions considered herein yielded positive biasness. Furthermore, upon review of Tables 4 and 5 in which the MAE and RMSE were tabulated, the two-parameter lognormal distribution was again consistently the lowest reported value for all combinations of distribution and correlation assumed for the model parameters. Combining the information provided by the three performance criteria, it was determined that among the five candidate distributions investigated the two parameter lognormal distribution presented the "best" fit to the simulated values for the DO deficit regardless of the distribution of water quality parameters. If one were to select a second-best distribution according to Tables 3-5, a selection of the two-parameter gamma distribution would be made.

Depending on the probability distribution assumed for the parameters in the water quality model, Eq. 7, the relative performance, according to the three performance criteria—the normal distribution, Weibull distribution, and

TABLE 4. Mean Absolute Error (MAE) for DO Deficit between Simulation Results and Various Assumed Distributions at Downstream Locations

Case number (1)	Mean Absolute Error (in mg/L)				
	N (2)	LN (3)	G (4)	W (5)	FC (6)
(a) At 5 mi (8 km)					
1	0.434	0.407	0.413	0.455	0.434
2	0.478	0.409	0.415	0.500	0.477
3	0.455	0.412	0.417	0.467	0.428
4	0.471	0.414	0.420	0.486	0.447
5	0.443	0.405	0.410	0.461	0.427
6	0.421	0.403	0.408	0.432	0.476
7	0.416	0.383	0.389	0.435	0.444
8	0.443	0.406	0.411	0.462	0.426
9	0.466	0.395	0.401	0.486	0.489
10	0.455	0.409	0.415	0.477	0.452
(b) At 10 mi (16 km)					
1	0.597	0.580	0.587	0.610	0.597
2	0.607	0.547	0.557	0.633	0.612
3	0.596	0.547	0.554	0.600	0.558
4	0.624	0.554	0.564	0.628	0.582
5	0.583	0.547	0.554	0.591	0.559
6	0.576	0.558	0.565	0.573	0.656
7	0.548	0.524	0.531	0.559	0.598
8	0.590	0.570	0.577	0.600	0.581
9	0.594	0.529	0.539	0.611	0.644
10	0.605	0.553	0.563	0.626	0.596

Note: N = normal; LN = lognormal; G = gamma; W = Weibull; and FC = Fisher-Cornish.

the Fisher-Cornish asymptotic expansion—varies. Of the majority of the cases considered, the Fisher-Cornish asymptotic expansion for the distribution of the DO deficit performs better or equally as well when compared with the results for the normal and Weibull distribution. However, after examining its marginal performance and knowing the relative complexity of this method, the use of the Fisher-Cornish method is seen as an impractical approach for obtaining the desired distribution for the DO deficit using the methods outlined in this paper.

Since the Fisher-Cornish asymptotic expansion is dependent on knowing of the higher-order moments to approximate the quantiles of standardized random DO deficit, the accuracy of approximating the distribution of the DO deficit using this method relies heavily on the accurate estimation of the skew coefficient and kurtosis, which are in turn estimated by the use of first-order analysis of uncertainty. Recognizing this fact, examinations were made to compare the values of the statistical moments of the DO deficit calculated by first-order analysis with those from the Monte Carlo simulation. From Table 6, significant discrepancies of the skewness and kurtosis between the first-order approximations and sample statistics from the Monte Carlo sim-

TABLE 5. Root Mean Squared Error (RMSE) for DO Deficit between Simulation Results and Various Assumed Distributions at Downstream Locations

Case number (1)	Root Mean Squared Error (in mg/L)				
	N (2)	LN (3)	G (4)	W (5)	FC (6)
(a) At 5 mi (8 km)					
1	0.457	0.415	0.418	0.478	0.457
2	0.514	0.427	0.431	0.548	0.521
3	0.485	0.418	0.429	0.501	0.452
4	0.506	0.423	0.434	0.523	0.477
5	0.473	0.412	0.420	0.493	0.451
6	0.446	0.405	0.411	0.459	0.531
7	0.443	0.388	0.394	0.462	0.481
8	0.469	0.413	0.419	0.490	0.446
9	0.503	0.408	0.417	0.532	0.528
10	0.486	0.420	0.426	0.509	0.485
(b) At 10 mi (16 km)					
1	0.623	0.590	0.592	0.635	0.623
2	0.644	0.562	0.567	0.669	0.654
3	0.638	0.551	0.566	0.644	0.584
4	0.669	0.558	0.577	0.676	0.623
5	0.622	0.551	0.561	0.633	0.589
6	0.601	0.574	0.579	0.600	0.740
7	0.585	0.530	0.538	0.593	0.666
8	0.623	0.578	0.583	0.630	0.609
9	0.634	0.535	0.548	0.650	0.693
10	0.643	0.562	0.570	0.662	0.632

Note: N = normal; LN = lognormal; G = gamma; W = Weibull; and FC = Fisher-Cornish.

ulation could be observed. The results of this investigation revealed that such discrepancies become more pronounced as the order of moment is increased. This observation indicated that the skew coefficient and kurtosis of the DO deficit estimated by first-order analysis was not satisfactory. Its inaccuracy was likely attributed to the nonlinearity involved in computing the DO deficit by Eq. 7, making the use of first-order analysis less desirable for estimating the high order moments. This phenomenon was also observed by other researchers (Gardner et al. 1981; Hornberger 1980). Therefore, from both the practical and computational viewpoints, the Fisher-Cornish asymptotic expansion, in conjunction with first-order analysis of uncertainty, is not a desirable approach for describing the distribution of the DO deficit. In addition, by comparing the mean values of the DO deficit in Table 6, it was observed that the mean DO deficit estimated by first-order analysis was somewhat higher than that based on Monte Carlo simulation for all the cases investigated. This is the main reason for the positive biasness observed in Table 3. In other words, first-order analysis yields an overestimation of the mean DO deficit. The difference in standard deviation between first-order analysis and simulation for the DO deficit is less significant.

TABLE 6. Comparison of Statistical Moments of DO Deficit between First-Order Analysis and Simulation at $x = 10$ mi (16 km)

Case number (1)	Mean		Standard Deviation		Skewness		Kurtosis	
	FOA ^a (2)	SIMU ^b (3)	FOA (4)	SIMU (5)	FOA (6)	SIMU (7)	FOA (8)	SIMU (9)
1	4.774	4.186	1.505	1.514	0	0.644	3.000	3.811
2	4.902	4.351	1.607	1.731	0	1.003	2.305	5.038
3	4.774	4.223	1.505	1.486	0.408	1.102	3.486	5.196
4	4.902	4.345	1.607	1.629	0.335	1.166	2.678	5.362
5	4.774	4.219	1.505	1.507	0.265	0.873	3.169	4.108
6	4.774	4.207	1.505	1.432	-0.739	0.594	6.312	3.661
7	4.774	4.242	1.505	1.482	-0.377	0.801	5.403	4.216
8	4.774	4.198	1.505	1.490	0.124	0.733	3.127	3.981
9	4.902	4.367	1.607	1.672	-0.298	1.123	4.148	5.679
10	4.902	4.342	1.607	1.683	0.111	0.910	2.404	4.374

^aFOA = first-order analysis.

^bSIMU = Monte Carlo simulation.

SUMMARY AND CONCLUSIONS

It has been recognized that processes in natural stream environments are inherently random. The ability to quantify the probabilistic status of stream environments is of vital importance in water quality management decision making. This paper presented a methodology to assess the probability distribution of dissolved oxygen deficit based on the simple Streeter-Phelps equation. The methodology involves the use of first-order analysis of uncertainty of the DO deficit equation. Mathematical operation in the exercise includes only taking the first- and second-derivatives that all engineers are capable of doing. Furthermore, the statistical analysis involves merely elementary operations of finding expectation and variance. The methodological framework is general enough to be applied to more sophisticated versions of the Streeter-Phelps equation which incorporate various sources and sinks affecting the DO levels in stream.

Various commonly used statistical probability distribution models were considered as the candidates for describing random characteristics of the DO deficit. Results of numerical experiments indicated that the use of a two-parameter lognormal distribution, along with the statistical mean and variance estimated by the first-order analysis, appears to be the best model with respect to the three goodness-of-fit performance criteria used regardless of the distributional properties of random water quality parameters involved in calculating the DO deficit. The practical implication of this observation is that, in probabilistic water quality analysis, one can relax the accuracy requirement on the probability distribution for the water quality parameters because the best-fit distribution for the DO deficit would be insensitive to the form of the parameter's distribution. The second-best parametric distribution model for describing the DO deficit is the two-parameter gamma distribution. Although the normal distribution has been used frequently in the past (mainly for its ease in use), it was found less suitable, in this study, in

describing the random behavior of the DO deficit. The nonparametric distribution model using Fisher-Cornish asymptotic expansion was also included in the study for the purpose of comparison. The model did not yield any better results than its parametric competitors and was substantially poorer than the lognormal distribution for most of the cases investigated. This is probably because the estimations of higher-order moments for random DO deficit by first-order analysis are not accurate enough due to the highly nonlinear nature of the DO deficit equation. This same reason may be applied to explain the observation in the study that first-order analysis produces a positively biased estimation of DO deficit when compared with the simulated values.

ACKNOWLEDGMENTS

The study was funded by the Wyoming Water Research Center at the University of Wyoming. The writers wish to thank the three anonymous reviewers for their constructive criticisms. Furthermore, we like to thank Dr. Steve McCutcheon from U.S. EPA Environmental Research Laboratory, Athens, Georgia, for his review of the paper and for bringing the recent work on Enhanced QUAL2-UNCAS to our attention.

APPENDIX I. REFERENCES

- Benjamin, J. R., and Cornell, C. A. (1970). *Probability, statistics, and decisions for civil engineers*. McGraw-Hill, Inc., New York, N.Y.
- Brown, L. C., and Barnwell, T. O. (1987). "The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: documentation and user manual." Report EPA/600/3-87/007, U.S. Environmental Protection Agency, Washington, D.C.
- Burges, S. J., and Lettenmaier, D. P. (1975). "Probabilistic methods in stream quality management." *Water Resour. Bull.*, 11(1), 115-130.
- Chadderton, R. A., Miller, A. C., and McDonnell, A. J. (1982). "Uncertainty analysis of dissolved oxygen model." *J. Envir. Engrs. Div.*, ASCE, 108(5), 1003-1012.
- Cornell, C. A. (1972). "First-order analysis of model and parameter uncertainty." *Proc. Int. Symp. on Uncertainties in Hydrologic and Water Resour. Systems*, Univ. of Arizona, vol. III, 1245-1276.
- Dobbins, W. E. (1964). "BOD and oxygen relationships in streams." *J. Sanit. Engrg. Div.*, ASCE, 90(4), 53-78.
- Esen, I. I., and Rathbun, R. E. (1976). "A stochastic model for predicting the probability distribution of the dissolved oxygen deficit in streams." *Professional Paper 913*, U.S. Geological Survey, Washington, D.C.
- Fisher, R. A. (1950). *Contributions to mathematical statistics*. John Wiley and Sons, Inc., New York, N.Y.
- Fisher, R. A., and Cornish, E. A. (1960). "The percentile points of distributions having known cumulants." *Technometrics*, 2(2), 209-225.
- Gardner, G. W., et al. (1981). "A comparison of sensitivity analysis and error analysis based on a stream ecosystem model." *Ecological Modelling*, 12(1), 173-190.
- Hastings, N. A. J., and Peacock, J. B. (1974). *Statistical distributions*. John Wiley and Sons, Inc., New York, N.Y.
- Hathorn, W. E. (1986). "Stochastic optimal waste load allocation in a stream environment under uncertainty," thesis presented to the Univ. of Wyoming, Laramie, Wyo., in partial fulfillment of the requirements for the degree of Master of Science.
- Hornberger, G. M. (1980). "Uncertainty in dissolved oxygen prediction due to variability in algal photosynthesis." *Water Res.*, 14(4), 355-361.

- Kothandaraman, V., and Ewing, B. B. (1969). "A probabilistic analysis of dissolved oxygen-biochemical oxygen demand relationship in stream." *J. Water Poll. Cont. Fed.*, 41(2), R73-R90.
- Krenkel, P. A., and Novotony, V. (1980). *Water quality management*. Academic Press, New York, N.Y.
- Loucks, D. P., and Lynn, W. R. (1966). "Probabilistic models for predicting stream quality." *Water Resour. Res.*, 2(3), 593-605.
- Padgett, W. J., Schultz, C., and Tsokos, C. P. (1977). "A random differential equation approach to the probability distribution of BOD and DO in stream." *J. Appl. Math.*, 32(2), 467-483.
- Padgett, W. J. (1978). "A stream pollution model with random deoxygenation and reaeration coefficients." *Math. Biosci.*, 42(1/2), 137-148.
- Padgett, W. J., and Rao, A. N. V. (1979). "Estimation of BOD and DO probability distribution." *J. Envir. Engrg. Div.*, ASCE, 105(3), 525-533.
- Patel, J. K., Kapadia, C. H., and Owen, D. B. (1976). *Handbook of statistical distributions*. John Wiley and Sons, Inc., New York, N.Y.
- Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo method*. John Wiley and Sons, Inc., New York, N.Y.
- Streeter, H. W., and Phelps, E. B. (1925). "A study of the pollution and natural purification of the Ohio River." *Public Health Bulletin*, U.S. Public Health Service, Washington, D.C.
- Thayer, R. P., and Krutchkoff, R. G. (1967). "Stochastic model for BOD and DO in streams." *J. Sanit. Engrg. Div.*, ASCE, 93(3), 59-72.
- Stuart, A., and Ord, J. K. (1987). *Kendall's advanced theory of statistics: distribution theory*. 5th Ed., vol. 1, Oxford University Press, New York, N.Y.

APPENDIX II. NOTATION

The following symbols are used in this paper:

- BIAS = biasness defined in Eq. 10;
- cov [] = covariance operator;
- D_0 = initial DO deficit;
- D_x = DO deficit concentration at location x mi (or $1.6 \times$ km) from discharge point;
- $E[]$ = expectation operator;
- $g()$ = general representation of function;
- L_0 = initial BOD concentration;
- MSE = mean absolute error defined in Eq. 11;
- RMSE = root mean squared error defined in Eq. 12;
- K_a = reaeration coefficient;
- K_d = deoxygenation coefficient;
- P'_x = first-order partial derivative of D_x with respect to parameter X ;
- $P''_{x,y}$ = second-order partial derivative of D_x with respect to X and Y ;
- U = average flow velocity;
- var [] = variance operator;
- \mathbf{X} = vector of random parameter in first-order analysis;
- $\bar{\mathbf{X}}$ = vector of mean value of random parameter X_i ;
- X_i = i th random parameter;
- x_p = true p th quantile of DO deficit concentration;
- $\hat{x}_{p,f}$ = p th quantile of DO deficit estimated by using probability model, f ;
- Y = dependent random variable in first-order analysis;
- μ = mean value; and
- σ = standard deviation.