PROBABILITY DISTRIBUTION FOR CRITICAL DO LOCATION IN STREAMS

Y.K. Tung
W.E. Hathhorn

1988

Journal Article WWRC-87-17

In

Ecological Modelling

Volume 42

1988

Yeou-Koung Tung
Wade E. Hathhorn
Wyoming Water Research Center
and
Deaprtment of Statistics
University of Wyoming
Laramie, Wyoming

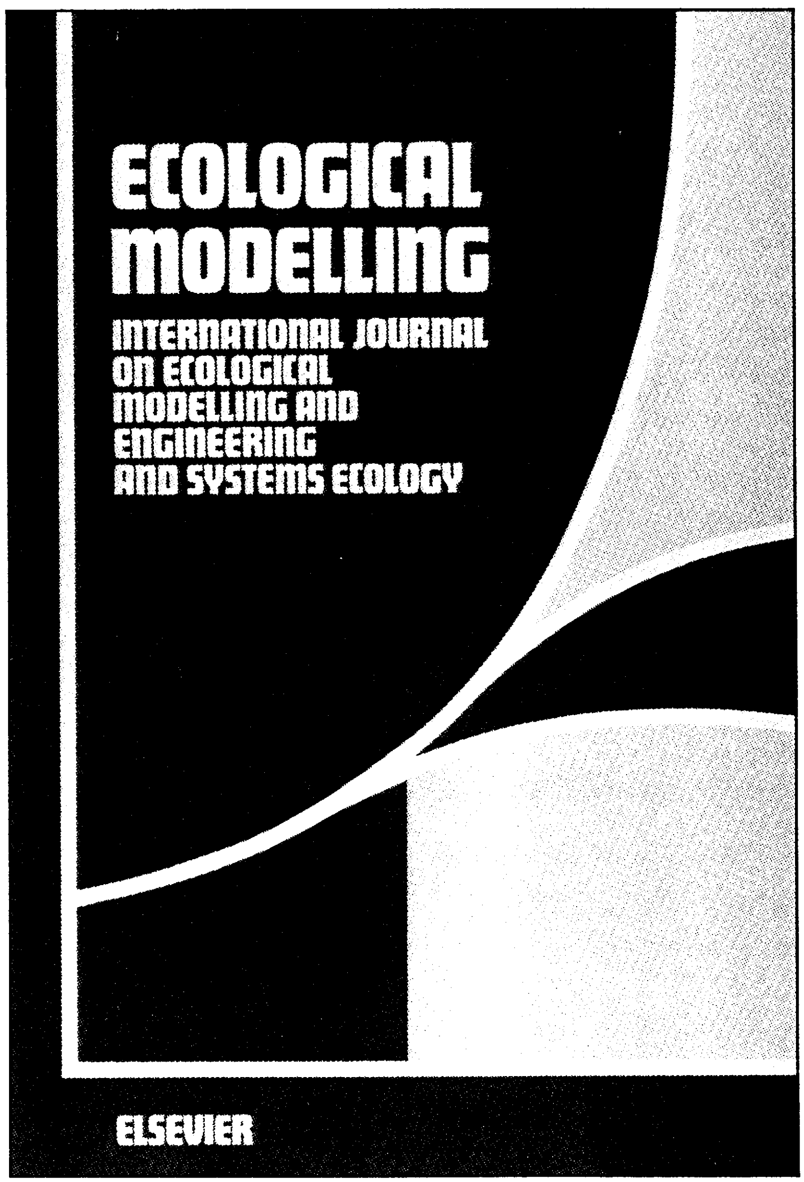
PROBABILITY DISTRIBUTION FOR CRITICAL DO LOCATION IN STREAMS

YEOU-KOUNG TUNG and WADE E. HATHHORN

Wyoming Water Research Center and Department of Statistics, University of Wyoming, Laramie, WY 82071 (U.S.A.)

(Accepted 28 October 1987)

REPRINTED FROM:



ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM

Ecological Modelling, 42 (1988) 45–60

Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

PROBABILITY DISTRIBUTION FOR CRITICAL DO LOCATION IN STREAMS

YEOU-KOUNG TUNG and WADE E. HATHHORN

Wyoming Water Research Center and Department of Statistics, University of Wyoming, Laramie, WY 82071 (U.S.A.)

(Accepted 28 October 1987)

ABSTRACT

Tung, Y.K. and Hathhorn, W.E., 1988. Probability distribution for critical DO location in streams. *Ecol. Modelling*, 42: 45–60.

Identifying the critical location in a stream environment system plays an important role in regulating and monitoring water quality. The critical location is defined as the point of maximum dissolved oxygen deficit within any reach of stream. It is at this location that the threat to the health of the aquatic biota is most severe. Unfortunately, due to the existence of random processes and parameter uncertainties within actual stream conditions, the critical location cannot always be determined with certainty. In recognizing the importance of identifying such a position, this paper attempts to assess the appropriateness of using some of the more common probability distributions to describe the random characteristics of the critical location in a stochastic stream environment. The results from such an assessment could enable one to estimate useful properties of the random critical location such as confidence interval information and the mode of its location. It is believed that this information would have important implications in managing and monitoring stream water quality.

INTRODUCTION

In water quality control and monitoring, focus is often placed on the critical location where the water quality condition is the most threatening. In an attempt to identify this location, water quality models are used to assist in estimating not only the critical location but also the condition of the water quality at that location.

Water pollution is not a random process, it is deliberate. However, it should be pointed out that the modeling of water quality conditions in a stream environment is an extremely enigmatic task. Such complexities are

0304-3800/88/$03.50

the result of inherent randomness and uncertainty exhibited throughout the stream environment. Not only are the physical and biological processes not clearly understood, but an imposing number of uncertainties are also associated with the various processes occurring within the aquatic environment. Uncertainties in stream water quality modeling have been discussed by Hathhorn and Tung (1988). In this paper, the critical location in a stream environment is defined as the point, within a reach, where the dissolved oxygen (DO) concentration is at its minimum.

Recognizing the existence of such uncertainties, the prediction of the critical location within a given reach of stream is no longer deterministic. Rather, this critical location is a random variable associated with a probability distribution. However, the exact distribution of the critical location is not known, and the analytical derivation of such a distribution is generally impossible. Although there have been a number of investigations conducted to estimate the distributions of biochemical oxygen demand (BOD) and DO concentration (Loucks and Lynn, 1966; Thayer and Krutchkoff, 1967; Kothandaraman, 1970; Esen and Rathbun, 1976; Padget et al., 1977; Padget and Rao, 1979), the assessment of the probability distribution that describes the uncertain characteristics of the critical location has remained virtually unexplored. Although the analysis of the critical location in a stochastic stream environment remains relatively unaccounted for in the literature, this should not to be taken to mean that such information is of little significance. On the contrary, the identification of the critical location plays a major role in the regulatory process and monitoring of any stream system to which waste effluents are discharged. This location, from a monitoring viewpoint, has the greatest significance within any reach of the stream system. By knowing the distribution of the critical location and its statistical properties, such information as the confidence interval, the most likely value, and the like can be obtained. From this, regulatory agencies would be provided with information identifying a point or region likely to contain the most severe water quality conditions. Hence, in a stream system filled with uncertainty, such information could possibly narrow the length of monitoring networks by excluding locations which were unlikely to contain the actual critical point. This would obviously reduce the capital and operating costs of such a monitoring network.

Due to the complexity of the formula commonly used to evaluate the critical location, the analytical derivation of the exact distribution is, to say the least, a formidable task. As an alternative, this paper examines the appropriateness of using some of the more common probability distributions in describing the random characteristics of the critical location. The statistical properties of the critical location such as the mean, variance, and higher order moments are estimated using the first-order analysis.

BASIC WATER QUALITY MODEL

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 coupled equation set is given as:

$$\mathrm{d}L/\mathrm{d}t = -K_{\mathrm{d}}L$$

$$\mathrm{d}D/\mathrm{d}t = K_{\mathrm{d}}L - K_{\mathrm{a}}D \tag{1}$$

The solution to equation (1), replacing t by x/U, is:

$$D_x = \frac{K_{\mathrm{d}}L_0}{K_{\mathrm{a}} - K_{\mathrm{d}}}\left[\exp(-K_{\mathrm{d}}/U) - \exp(-K_{\mathrm{a}}x/U)\right] + D_0 \exp(-K_{\mathrm{a}}x/U) \tag{2}$$

where K_{d} is the deoxygenation coefficient (day^{-1}), K_{a} is the reaeration coefficient (day^{-1}), x is the distance downstream from the source of BOD (miles *), U is the average stream velocity (miles day^{-1}), D_x is the DO deficit concentration (mg L^{-1}) within a reach at a downstream distance x, D_0 is the initial DO deficit (at distance $x = 0$), and L_0 is the initial in-stream BOD concentration (both in mg L^{-1}).

The downstream location, X_{c} (miles), where the maximum DO deficit occurs can be found by differentiating equation (2) and solving for x:

$$X_{\mathrm{c}} = \frac{U}{K_{\mathrm{a}} - K_{\mathrm{d}}}\ln\left\{\frac{K_{\mathrm{a}}}{K_{\mathrm{d}}}\left[1 - \frac{(K_{\mathrm{a}} - K_{\mathrm{d}})D_0}{K_{\mathrm{d}}L_0}\right]\right\} \tag{3}$$

The point X_{c} will herein be referred to as the 'critical location'.

The original Streeter–Phelps equation is limited to only two instream processes: deoxygenation of the water due to bacterial decomposition of carbonaceous organic matter, and reaeration directly proportional to the DO deficit. It should also be noted that several assumptions were made in the development of the Streeter–Phelps equation: (a) steady, uniform flow; (b) DO deficits predicted by equation (2) are one-dimensional (functions only of the position downstream from a discharge point); and (c) rate of biodegradation and reaeration, expressed by K_{d} and K_{a}, are described by first-order kinetics. A typical DO profile for a single reach is shown in Fig. 1.

Although, equation (2) describes the response of DO in a single reach of stream as a result of the addition of a 'point-source' loading of waste at the upstream end of the reach, this equation can be used to determine the DO concentration in several reaches by applying the deficit at the downstream

---

* mile ≈ 1609 m.

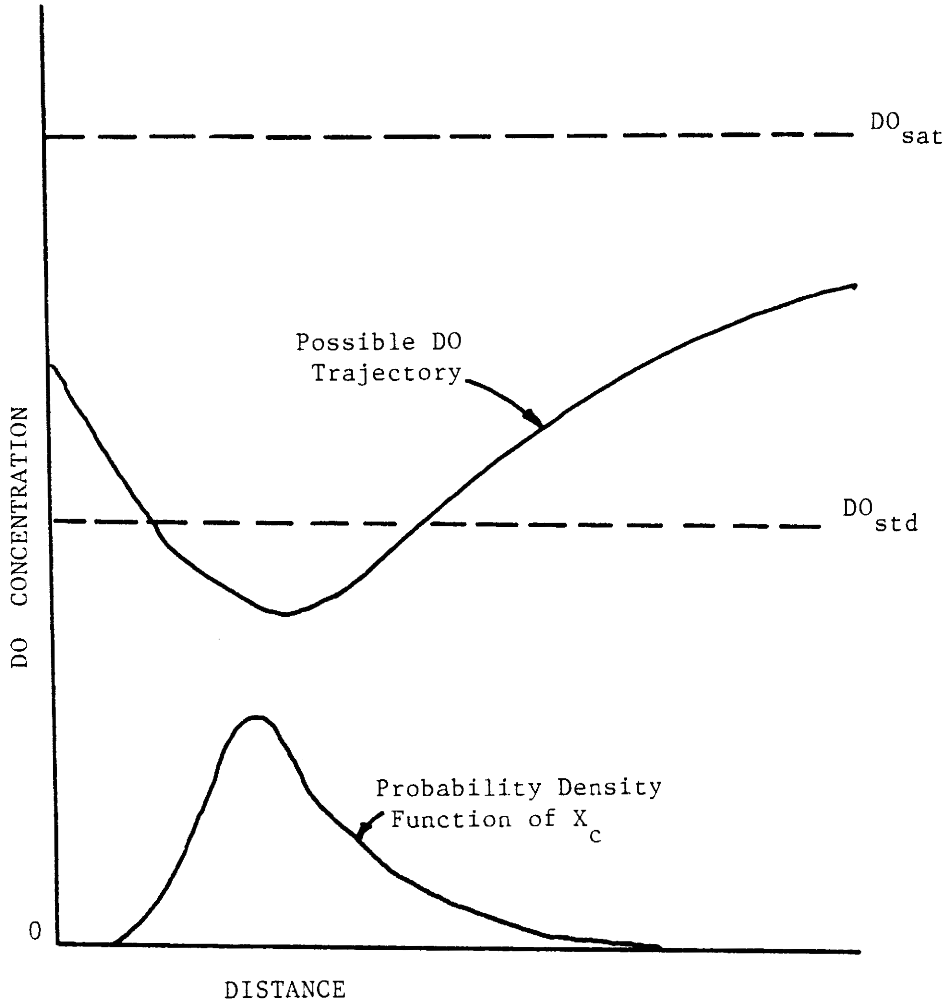


Fig. 1. Schematic diagram of the probability density function for the critical location.

end of one reach as the initial deficit of the succeeding reach. Thus, equation (2) can be applied recursively to determine the DO profile of an entire stream system in a multiple-discharge setting (Liebman and Lynn, 1966).

Since its conception, the Streeter–Phelps equation has been modified to account for discrepancies between analytical estimations, computed from equation (2), 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 processes which could contribute to instream BOD and DO variations. There have been several studies conducted in which one or more of the 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 modification can be made by simply adding terms to equation (2) to account for the various additional factors. The expression for the critical location can then be derived accordingly. However, to simplify the algebraic manipulations, the critical location derived from the original Streeter–Phelps equation will be utilized herein.

FIRST-ORDER ANALYSIS OF UNCERTAINTY

The use of first-order uncertainty analysis is quite popular in all fields of engineering and science. Its application can be found in a wide array of problems. Detailed analysis and development of first-order uncertainty methods are given by Benjamin and Cornell (1970) and Cornell (1972). Burges and Lettenmaier (1975) have utilized the method to investigate the uncertainty in predictions of BOD and DO within the stochastic stream environment.

Essential, first-order uncertainty analysis provides a methodology of obtaining an estimate for the moments of a single random variable or function of several random variables. The method estimates the uncertainty in a deterministic model formulation involving parameters which are not known with certainty. By using first-order analysis, the combined effect of uncertainty in a model formulation, resulting from the use of uncertain parameters, can be estimated.

First-order uncertainty analysis can be characterized by two major components: single moment (variance) treatment of the random variables, and the use of first-order approximation of any functional relationship (e.g., the use of Taylor's series expansion). The first major component implies that the random element of any variable is defined exclusively by its first two moments. Thus, information pertaining to the character of a random variable, Y, is provided solely by its mean ($\overline{Y}$) and variance (σ_y^2).

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

In order to illustrate the general methodology of first-order analysis, consider a random variable, Y, which is a function of n random variables X_i (multivariate case) will be considered. Mathematically, Y can be expressed as:

$$Y = g(\boldsymbol{X}) \tag{4}$$

where $\boldsymbol{X} = (X_1, X, \ldots, X_n)$. Through the use of Taylor's series expansion, the random variable Y can be approximated by:

$$Y \stackrel{2}{=} g(\overline{\boldsymbol{X}}) + \sum_{i=1}^{n} \left[\frac{\partial g}{\partial X_i} \right]_{\overline{\boldsymbol{X}}} (X_i - \overline{X}_i) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\frac{\partial^2 g}{\partial X_i \, \partial X_j} \right]_{\overline{\boldsymbol{X}}} (X_i - \overline{X}_i)(X_j - \overline{X}_j) \tag{5}$$

in which $\overline{X} = (\overline{X}_1, \overline{X}_2, ..., \overline{X}_n)$, is a vector containing the means of n random variables, $\overset{2}{=}$ represents equal in the sense of a second-order approximation.

Then, the second-order approximation of the expected value of Y is:

$$\overline{Y} \overset{2}{=} g(\overline{X}) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\frac{\partial^2 g}{\partial X_i \, \partial X_j} \right]_{\overline{X}} \mathrm{Cov}[X_i, X_j] \tag{6}$$

in which $\mathrm{Cov}[X_i, X_j]$ is the covariance between random variables X_i and X_j.

It follows that the first-order approximation of the variance of Y is:

$$\sigma_y^2 = \mathrm{Var}[Y] \overset{1}{=} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\frac{\partial g}{\partial X_i} \right]_{\overline{X}} \left[\frac{\partial g}{\partial X_j} \right]_{\overline{X}} \mathrm{Cov}[X_i, X_j] \tag{7}$$

If the random variables are independent, equation (7) reduces to:

$$\sigma_y^2 \overset{2}{=} \sum_{i=1}^{n} \left[\frac{\partial g}{\partial X_i} \right]_{\overline{X}}^2 \sigma_i^2 \tag{8}$$

where $\overset{1}{=}$ means equal in a first-order sense (Benjamin and Cornell, 1970; Burges and Lettenmaier, 1975); σ_i^2 is the variance corresponding to random variable X_i.

STATISTICAL PROPERTIES OF THE CRITICAL LOCATION

As a result of the uncertainty involved in the stochastic stream environment, the critical location, computed by equation (3), is itself a random variable. Thus, to estimate the statistical properties of the critical location under such conditions, first-order analysis is employed. To illustrate the concept of the probability distribution associated with the critical location, a schematic diagram is provided in Fig. 1.

Using first-order analysis Taylor's series expansion of equation (3) leads to the following approximation:

$$\begin{aligned} X_c \overset{1}{=} & X_c(\overline{K}_d, \overline{K}_a, \overline{U}, \overline{L}_0, \overline{D}_0) + F'_{K_d}(K_d - \overline{K}_d) + F'_{K_a}(K_a - \overline{K}_a) \\ & + F'_U(U - \overline{U}) + F'_{L_0}(L_0 - \overline{L}_0) + F'_{D_0}(D_0 - \overline{D}_0) \end{aligned} \tag{9}$$

where $F'_X = \partial X_c / \partial X$ and $F''_{X,Y} = \partial^2 X_c / \partial X \, \partial Y$ are evaluated at the mean values of the model parameters. The analytical expressions for each partial derivative term was presented by Hathhorn (1986).

It follows that the first-order approximation of the expected critical location can be written as:

$$\mathrm{E}[X_c] \overset{1}{=} X_c(\overline{K}_d, \overline{K}_a, \overline{U}, \overline{L}_0, \overline{D}_0) \tag{10}$$

This is simply equation (3) evaluated at the means of the model parameters.

The use of first-order analysis can be continued to obtain estimates for the variance, skewness, and kurtosis of X_c as follows:

$$\begin{aligned}\mathrm{Var}[X_c] &= \sigma_{X_c}^2 = \mathrm{E}\left[\left(X_c - \overline{X}_c\right)^2\right] \\ &\overset{1}{=} F_{K_d}'^2\mathrm{Var}(K_d) + F_{K_a}'^2\mathrm{Var}(K_a) + F_U'^2\mathrm{Var}(U) \\ &\quad + F_{L_0}'^2\mathrm{Var}(L_0) + F_{D_0}'^2\mathrm{Var}(D_0)\end{aligned} \tag{11}$$

$$\begin{aligned}\gamma_{X_c} &= \mathrm{E}\left[\left(X_c - \overline{X}_c\right)^3\right]/\mathrm{Var}[X_c]^{1.5} \\ &\overset{1}{=} \Big\{F_{K_d}'^3\gamma_{K_d}[\mathrm{Var}(K_d)]^{1.5} + F_{K_a}'^3\gamma_{K_a}[\mathrm{Var}(K_a)]^{1.5} \\ &\quad + F_U'^3\gamma_U[\mathrm{Var}(U)]^{1.5} + F_{L_0}'^3\gamma_{L_0}[\mathrm{Var}(L_0)]^{1.5} \\ &\quad + F_{D_0}'^3\gamma_{D_0}[\mathrm{Var}(D_0)]^{1.5}\Big\}[\mathrm{Var}(X_c)]^{-1.5}\end{aligned} \tag{12}$$

$$\begin{aligned}\kappa_{X_c} &= \mathrm{E}\left[\left(X_c - \overline{X}_c\right)^4\right]\mathrm{Var}[X_c]^{-2} \\ &\overset{2}{=} \Big\{F_{K_d}'^4\kappa_{K_d}[\mathrm{Var}(K_d)]^2 + F_{K_a}'^4\kappa_{K_a}[\mathrm{Var}(K_a)]^2 \\ &\quad + F_U'^4\kappa_U[\mathrm{Var}(U)]^2 + F_{L_0}'^4\kappa_{L_0}[\mathrm{Var}(L_0)]^2 \\ &\quad + F_{D_0}'^4\kappa_{D_0}[\mathrm{Var}(D_0)]^2\Big\}[\mathrm{Var}(X_c)]^{-2}\end{aligned} \tag{13}$$

where γ_X and κ_X are the skew coefficient and kurtosis of the random variable X, respectively. It should be pointed out that the expressions in (11)–(13) assume that all water quality parameters are independent.

PROBABILITY DISTRIBUTION OF THE CRITICAL LOCATION

By considering the stream system to be an inherently random environment, the critical location X_c is itself a random variable. Thus, in order to evaluate the confidence interval or the location most likely to be critical, knowledge of the probability distribution associated with the X_c is required. Otherwise, the analysis of water quality conditions in a stream environment under uncertainty is, at best, simply conjecture.

Although significant research has been conducted into the uncertainty analysis of stream dissolved oxygen, most of these studies have been concerned with variations in 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 probability distribution associated with the DO deficit as cited previously. From a practical viewpoint, the main disadvantage to the aforementioned methods is that the resulting probability distributions derived for the DO deficit are too complicated for most engineers to readily assess the probability of violating a given water quality standard. Moreover, for the case of assessing the probability distribution of the critical location, the analytical approaches developed for the DO and BOD studies cited above are no longer applicable.

Thus, the present study attempts to examine the utility of some commonly used parametric probability distributions to describe the random characteristics of the critical location computed by (3). Four parametric probability distribution are considered herein for their relative ease in use and versatile shape. These include the normal, lognormal, gamma, and Weibull distributions. When using these four probability models, only the mean and variance of X_c are needed and can be estimated by first-order analysis using (10) and (11). The parameters in each of the candidate probability distribution models can be obtained through moment-parameter relationships found in most statistics textbooks. Relations of the skew coefficient and kurtosis to the parameters in the candidate probability models can be found elsewhere Hastings and Peacock (1974) and Patel et al. (1976).

Instead of making such a strong assumption about the probability density function of a specific form for X_c, an approach of using Fisher–Cornish asymptotic expansion is also applied (Fisher, 1950; Fisher and Cornish, 1960; Kendall and Stuart, 1977). This method relates the quantile of any standardized distribution to the standard normal quantile and higher order moments. In this case, the quantile or order p for X_c can be approximated without making an assumption about its distribution as follows:

$$X_c(p) = \mathrm{E}[X_c] + \xi_p\sqrt{\mathrm{Var}(X_c)} \tag{14}$$

in which $X_c(p)$ is the pth-order quantile of standardized critical location. Because only the first four moments of X_c are approximated through first-order analysis in this study, i.e., (10)–(13), Fisher–Cornish asymptotic expansion for ξ_p can be expressed as:

$$\xi_p \approx z_p + \gamma_{X_c} H_2(z_p)/6 + \kappa_{X_c} H_3(z_p)/24 \\ -\gamma_{X_c}^2\left[(2H_3(z_p) + H_1(z_p)\right]/36 \tag{15}$$

in which z_p is pth-order quantile from standard normal distribution, $H_1(z_p)$, $H_2(z_p)$ and $H_3(z_p)$ are Hermit polynomials which can be computed by (Abramowitz and Stegun, 1970):

$$H_r(z_p) = z_p^r - \frac{r^2}{2\cdot 1!} z_p^{r-2} + \frac{r^4}{r^2\cdot 2!} z_p^{r-4} - \frac{r^6}{2^3\cdot 3!} z_p^{r-6} + \dots \tag{16}$$

PERFORMANCE EVALUATION CRITERIA OF THE DISTRIBUTIONS FOR THE CRITICAL LOCATION

The idea of applying first-order analysis for estimating the statistical moments of the critical location, along with an adoption of the probability density function for X_c, is straightforward and practical. However, among the various probability distribution models that are commonly used, a question to be raised is "Which probability distribution model (or models) best describe the random behavior of the critical location X_c in a stream?"

To evaluate the relative performance of each of the candidate probability distribution considered, three performance criteria are adopted herein: biasness (BIAS), mean absolute error (MAE), and root mean square error (RMSE). Each for the three criteria are used simultaneously in an attempt to identify the best probability model for describing the random characteristics of X_c at a given location.

These criteria are mathematically defined as:

– biasness

$$\text{BIAS} = \int_0^1 (\hat{x}_{p,f} - x_p)\, \mathrm{d}p \tag{17}$$

– mean absolute error

$$\text{MAE} = \int_0^1 |\hat{x}_{p,f} - x_p|\, \mathrm{d}p \tag{18}$$

– root mean square error

$$\text{RMSE} = \left[\int_0^1 (\hat{x}_{p,f} - x_p)^2\, \mathrm{d}p\right]^{0.5} \tag{19}$$

where x_p and $\hat{x}_{p,f}$ are, respectively, the true value and the estimate of the pth-order quantile determined from the assumed probability model, f. It should be noticed that the true value of the quantile for the X_c cannot be determined exactly due to the complexity of (3). As an alternative, Monte Carlo simulation (Rubinstein, 1981) is applied for obtaining the 'true' quantile for X_c. The Monte Carlo simulation for this task is described in the following section.

DERIVATION OF THE 'TRUE' DISTRIBUTION OF DO BY MONTE CARLO SIMULATION

To determine the probability distribution of X_c at a given location, Monte Carlo simulation techniques were employed, allowing each of the model parameters (K_d, K_a, U, L_0, and D_0) to be assigned one of four distributions: normal, lognormal, gamma, and Weibull. In addition, the

TABLE 1

Statistical properties of the model parameters used to investigate the distribution of the critical location for a hypothetical stream

Model parameters	Mean	Standard deviation	Units
K_d	0.35	0.10	day^{-1}
K_a	0.70	0.20	day^{-1}
U	10.00	3.00	miles day^{-1}
L_0	18.00	5.00	mg L^{-1}
D_0	1.00	0.30	mg L^{-1}

statistical properties of the model parameters for the hypothetical stream used in this paper are listed in Table 1. Simulation procedures were performed such that ten groups of 999 critical locations were generated using equation (3) and assigning one of the four distributions mentioned above to each of the model parameters. For example, during the first simulation run, ten groups of 999 critical locations were generated (using equation 3) under an independent and all normal assumption for all the water quality parameters. Then, in successive runs, different distributions were assigned to each of the model parameters, resulting in another set containing ten groups of 999 different critical locations.

It should also be pointed out that provisions for considering a positive correlation ($\rho = 0.8$) between model parameters K_a and U, were also included in this simulation exercise. Extensive discussion of the correlation between water quality parameters is given by Hathhorn (1986). Noting that when such a correlation is considered, a bivariate normal distribution is utilized. Values of the statistical properties shown in Table 1 and the correlation coefficient of 0.8 were based on values adopted from various published articles.

During the simulation runs, each of the ten groups of 999 critical locations were ranked in ascending order. Specifically, the minimum value of the DO deficit generated is assigned to position 1 and the maximum value to position 999. Then, quantities of the critical location X_c are computed for several probability levels p by simply locating the value of X_c in position $(999 + 1)p$. In order to reduce sampling errors, each of the respective quantiles obtained for the ten groups were arithmetically averaged.

RESULTS AND DISCUSSIONS

Values of the performance criteria, i.e. BIAS, MAE, RMSE, under various conditions are given in Table 2–4. In examining Tables 3 and 4, the majority

TABLE 2

Biasness for the critical location between simulation results and various assumed distributions

Distribution assumed for model parameters						Biasness (in miles)				
K_d	K_a	U	L_0	D_0	$\rho(K_a, U)$	N	LN	G	W	FC
N	N	N	N	N	0.0	−0.782	−0.848	−0.808	−0.800	−0.782
N	N	N	N	N	0.8	0.0342	0.010	0.039	0.041	0.028
LN	LN	LN	LN	LN	0.0	−0.703	−0.768	−0.728	−0.720	−0.726
LN	LN	LN	LN	LN	0.8	0.011	−0.014	0.015	0.017	−0.059
G	G	G	G	G	0.0	−0.795	−0.861	−0.820	−0.813	−0.811
W	W	W	W	W	0.0	−0.703	−0.769	−0.728	−0.721	−0.670
G	LN	N	G	W	0.0	−0.645	−0.711	−0.670	−0.663	−0.638

N, normal; LN, lognormal; G, gamma; W. Weibull; FC, Fisher–Cornish.

of the results reveal that the two-parameter gamma distribution appears to best describe the randomness of X_c according to the minimum MAE and MSE criteria even though it has a slightly higher biasness than other distributions in some of the cases. In general, the overall performances of all five distributions are to be considered somewhat less than desirable when one examines the magnitude of the error criteria, especially for the Fisher–Cornish asymptotic expansion. These results place serious question on the estimation ability of first-order analysis in conditions where the functional relationship of interest, i.e., equation (3) is highly nonlinear. It appears that the ability of first-order analysis to accurately estimate higher order moments (such as skewness and kurtosis) of a functional relationship diminishes as the degree of nonlinearity of the function increases.

TABLE 3

Mean absolute error (MAE) for the critical location between simulation results and various assumed distributions

Distributions assumed for model parameters						Mean absolute error (in miles)				
K_d	K_a	U	L_0	D_0	$\rho(K_a, U)$	N	LN	G	W	FC
N	N	N	N	N	0.0	0.831	1.085	0.889	0.834	0.831
N	N	N	N	N	0.8	0.219	0.346	0.240	0.301	4.269
LN	LN	LN	LN	LN	0.0	0.985	0.768	0.728	0.934	0.735
LN	LN	LN	LN	LN	0.8	0.354	0.077	0.126	0.574	4.798
G	G	G	G	G	0.0	0.955	0.922	0.820	0.941	0.823
W	W	W	W	W	0.0	0.713	0.940	0.751	0.721	1.858
G	LN	N	G	W	0.0	0.645	0.954	0.751	0.664	0.699

N, Normal; LN, Lognormal; G, Gamman; W. Weibull; FC, Fisher–Cornish.

TABLE 4

Root mean square error (RMSE) for the critical location between simulation results and various assumed distributions

Distributions assumed for model parameters						Root mean square error (in miles)				
K_d	K_a	U	L_0	D_0	$\rho(K_a, U)$	N	LN	G	W	FC
N	N	N	N	N	0.0	1.697	1.306	1.124	1.890	1.697
N	N	N	N	N	0.8	0.364	0.571	0.444	0.462	5.099
LN	LN	LN	LN	LN	0.0	1.672	0.892	0.857	1.708	1.240
LN	LN	LN	LN	LN	0.8	0.464	0.113	0.167	0.739	5.787
G	G	G	G	G	0.0	1.653	1.129	1.004	1.765	1.402
W	W	W	W	W	0.0	1.374	1.012	0.821	1.497	2.506
G	LN	N	G	W	0.0	1.098	1.066	0.821	1.253	1.211

N, Normal; LN, Lognormal; G, Gamma; W, Weibull; FC, Fisher–Cornish.

There is an interesting result that can be observed from Tables 2–4. That is, when $\rho(K_a, U) = 0.8$, all candidate distributions considered except Fisher-Cornish asymptotical expansion result in slightly positive biasness while, in the case of $\rho(K_a, U) = 0$, a significantly negative biasness is observed. Furthermore, the values of RMSE and MAE are much smaller in the case of $\rho(K_a, U) = 0.8$ than that of $\rho(K_a, U) = 0$ for the four parametric probability distributions considered while it is totally opposite for Fisher–Cornish asymptotic expansion.

Before a final decision is made as to the type of distribution that can be considered most appropriate for the critical location amongst those considered here, the results given in Tables 5a and 5b should be examined. In these tables, the 90-percent confidence intervals of X_c for each of the assumed distributions are reported, along with the confidence intervals derived from Monte Carlo simulation and the Fisher–Cornish asymptotic expansion. It should be pointed out that the 90-percent confidence intervals reported for the assumed distributions in Table 5a are independent of the type of distribution assumed for the model parameters. This is due to the fact that each of the common distributions utilized here can be completely characterized by the mean and variance of X_c, which is in turn computed solely by the mean and variance of the model parameters. The mean and variance of the model parameters do not change as the distributions assumed for these parameters are varied. However, the value of variance is affected by the correlation between K_a and U. Separate values are reported for a zero and positive correlation between model parameters K_a and U.

When actually comparing the numerical values presented in Tables 5a and 5b, it is obvious that the range of values presented are quite extended. For example, Table 5b reports the 90-percent confidence interval for X_c,

TABLE 5

90-percent confidence intervals (miles) for the critical location

(a) under various distribution assumptions

$\rho(K_a, U)$	Normal	Lognormal	Gamma	Weibull
0.0	(7.24, 29.1)	(9.53, 30.6)	(8.80, 30.6)	(8.10, 28.4)
0.8	(11.5, 24.8)	(12.3, 25.5)	(12.1, 25.5)	(10.9, 24.6)

(b) using monte carlo simulation and the Fisher–Cornish asymptotic expansion

$\rho(K_a, U)$	Distribution assumed for model parameters					90-percent confidence interval	
	K_1	K_2	U	L_0	D_0	Simulation	Fisher–Cornish
0.0	N	N	N	N	N	(8.57, 32.9)	(7.24, 29.1)
	LN	LN	LN	Ln	Ln	(9.78, 32.3)	(8.10, 29.7)
	G	G	G	G	G	(9.24, 32.8)	(7.76, 29.5)
	W	W	W	W	W	(10.6, 31.7)	(6.73, 27.4)
	G	LN	N	G	W	(8.55, 31.6)	(7.04, 28.9)
0.8	N	N	N	N	N	(11.1, 25.5)	(13.0, 23.3)
	LN	LN	LN	LN	LN	(12.3, 25.2)	(15.7, 24.6)

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

using the simulation procedures, lies between 8.57 and 32.9 miles (13.8 and 53 km) under all normal and uncorrelated assumptions for the model parameters. It is also interesting to observe that, when the correlation coefficient of 0.8 between K_a and U exists, the resulting 90-percent confidence interval length for X_c is reduced by nearly half. This could be related to the observation of smaller errors in Tables 2–4 associated with $\rho(K_a, U) = 0.8$.

Finally, the percentage of overlap between the confidence intervals computed under each of the assumed distributions (normal, lognormal, gamma, Weibull, and Fisher–Cornish) and that obtained through simulation procedures are reported in Table 6. Again, the assumption of a gamma distribution for the critical location results in the closest characterization of the 'true' confidence intervals obtained through simulation. This provides an additional piece of evidence supporting the use of a gamma distribution to model the unknown behavior of the critical location with respect to the data set considered here.

Unfortunately, from a practical viewpoint, the results obtained for the confidence intervals, in Tables 5a and 5b, provide little, if any, significant information in identifying an exact or narrow range containing the critical location in a stochastic stream system. The results from this approach are

TABLE 6

Percentage of overlapping for 90% confidence intervals with that of simulation under various distributional assumptions

Distributions assumed for model parameters						Percentage of overlapping for 90% C.I.				
$\rho(K_a, U)$	K_1	K_2	U	L_0	D_0	N	LN	G	W	FC
0.0	N	N	N	N	N	84.4	86.6	89.6	81.5	84.4
	LN	LN	LN	LN	LN	85.8	92.4	92.4	82.7	88.4
	G	G	G	G	G	94.3	89.4	90.7	81.3	86.0
	W	W	W	W	W	87.7	94.8	94.8	94.4	79.6
	G	LN	N	G	W	89.2	91.4	94.6	86.1	88.3
0.8	N	N	N	N	N	96.9	100.0	100.0	95.3	69.0

simply too widespread to be of any use in improving the monitoring or sampling process. The wide range of values reported can again be explained by the high nonlinearity associated with equation (3).

On the other hand, the results of this study have revealed that the gamma distribution best describes the random character of the critical location in a stochastic stream setting. Knowing this information, one is able to determine the most likely point to be the critical location by computing the mode of the resulting gamma distribution. This information may in fact lead those responsible for monitoring water quality to a location which is at or near the actual critical location in a stochastic stream environment. Consequently, a more effective water quality monitoring program could be established. Recall that the dissolved oxygen sag curve in a specific reach is unimodel. Hence, the true critical location is unique within each reach. By assertaining the approximate position of the critical through the use of determining the mode of the distribution given for the critical location, monitoring networks could be centralized in a more narrow range of stream length around this position, resulting in reduced expenditures for such monitoring.

SUMMARY AND CONCLUSIONS

Because the critical location defines a point where the threat to the health of aquatic biota is the most vulnerable, it often becomes the focus of attention in regulating and monitoring the quality of a stream environment. Moreover, due to the existence of uncertainties in the various related physical and biological processes, the critical location, in general, cannot be predicted with certainty. To estimate the whereabouts of the critical location in a stochastic stream environment, it is necessary to have knowledge about the probability distribution of the critical location itself. Information provided by such a statistical analysis could direct water quality regulatory agencies in establishing a monitoring network centered in a narrow region which likely contained the most severe water quality conditions in any reach. As noted earlier, this could ultimately lead to reduced costs, both in capital set-up and operation, of such a monitoring network.

This paper presents the results of study which assesses the appropriateness of using several common parametric probability distributions, along with a nonparametric Fisher–Cornish asymptotic expansion to describe the probability distribution of the critical location based on the statistical moments of X_c estimated using first-order analysis. Preliminarily, based on the three curve fitting criteria, i.e., BIAS, MAS, and RMSE, it was found that the two-parameter gamma distribution best describes the unknown character of the critical location in the majority of the cases examined here. Readers, however, should be cautioned that the gamma distribution found in this study for best describing the distribution of X_c may not be entirely valid if different water quality models and statistical properties of water quality parameters are considered. The design and execution of experiments from which the results can be made in general remains a challenging task for future research.

Although the 90-percent confidence interval for X_c was found to be too wide to have any practical usefulness, knowing the distribution of X_c would enable one to find the location that is most likely to be the critical point which in turn could be used as the basis for setting up an effective water quality monitoring network. Another interesting observation from the study is that the inclusion of a positive correlation between K_a and U, which exists in reality, shortens the length of confidence interval for the critical location. Unfortunately, the length of the confidence interval obtained under these assumptions remains too extended to provide any useful information for identifying the range of location containing the actual critical location.

ACKNOWLEDGEMENTS

The support given by the Wyoming Water Research Center to this study is greatly appreciated. The authors are indebted to the two reviewers for the paper. Gratitude is also extended to Mrs. Ruth Daniels for her preparation of the manuscript.

REFERENCES

Abramowitz, M. and Stegun, I.A. (Editors), 1972. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (9th Edition). Dover, New York, 1046 pp.

Benjamin, J.R. and Cornell, C.A., 1970. Probability, Statistics, and Decisions for Civil Engineers, McGraw-Hill, New York, 684 pp.

Burges, S.J. and Lettenmaier, D.P., 1975. Probabilistic methods in stream quality management. Water Res. Bull., 11: 115–130.

Chadderton, R.A., Miller, A.C. and McDonnell, A.J., 1982. Uncertainty analysis of dissolved oxygen model. J. Environ. Eng. ASCE, 108: 1003–1012.

Cornell, C.A., 1972. First-order analysis of model and parameter uncertainty. In: C.C. Kisiel and L. Duckstein (Editors), Proc. Int. Symp. Uncertainties in Hydrologic and Water Resource Systems, 11–14 December, III. University of Arizona, Tucson, AZ, pp. 1245–1276.

Dobbins, W.E., 1964. BOD and oxygen relationships in streams. J. San. Eng. Div. ASCE, 90: 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. USGS Prof. Pap. 913.

Fisher, R.A., 1950. Contributions to Mathematical Statistics. Wiley, New York.

Fisher, R.A. and Cornish, E.A., 1960. The percentile points of distributions having known cumulants. Technometrics, 2: 209–225.

Gardner, G.W., O'Neill, R.V., Mankin, J.B. and Carney, J.H., 1981. A comparison of sensitivity analysis and error analysis based on a stream ecosystem model. Ecol. Modelling, 12: 173–190.

Hastings, N.A.J. and Peacock, J.B., 1974. Statistical Distributions. Wiley, New York, 130 pp.

Hathhorn, W.E., 1986. Stochastic optimal waste load allocation in a stream environment under uncertainty. M.S. thesis, Department of Civil Engineering, University of Wyoming, Laramie, WY, 392 pp.

Hathhorn, W.E. and Tung, Y.K., 1988. Assessing the risk of violating stream water quality standards. J. Environ. Manage. (in press).

Hornberger, G.M., 1980. Uncertainty in dissolved oxygen prediction due to variability in algal photosynthesis. Water Resour. Res., 14: 355–361.

Kendall, M. and Stuart, A., 1977. The Advanced Theory of Statistics, Vol. 1 (4th Edition). Griffin, 604 pp.

Kothandaraman, V., 1970. Probabilistic variations in ultimate first stage BOD. J. Sanit. Eng. Div. ASCE, 96: 27–34.

Kothandaraman, V. and Ewing, B.B., 1969. A probabilistic analysis of dissolved oxygen – biochemical oxygen demand relationship in streams. J. Water Pollut. Control Fed., 41: R72–R90.

Krenkel, P.A. and Novotny, V., 1980. Water Quality Management. Academic Press, New York, 671 pp.

Liebman, J.C. and Lynn, W.R., 1966. The optimal allocation of stream dissolved oxygen. Water Resour Res., 2: 581–591.

Loucks, D.P. and Lynn, W.R., 1966. Probabilistic models for predicting stream quality. Water Resour. Res., 2: 593–605.

Padgett, W.J. and Rao, A.N.V., 1979. Estimation of BOD and DO probability distribution. J. Environ. Eng. ASCE, 105: 525–533.

Padgett, W.J., Schultz, C. and Tsokos, C.P., 1977. A random differential equation approach to the probability distribution of BOD and DO in streams, J. Appl. Math. SIAM, 32: 467–483.

Patel, J.K., Kapadia, C.H. and Owen, D.B., 1976. Handbook of Statistical Distribution. Wiley, New York, 302 pp.

Rubinstein, R.Y., 1981. Simulation and the Monte Carlo method. Wiley, New York, 278 pp.

Streeter, H.W. and Phelps, E.B., 1925. A study of the pollution and natural purification of the Ohio river. Public Health Bull. 146, U.S. Public Health Service, Washington, DC, pp. 127–147.

Thayer, R.P. and Krutchkoff, R.G., 1967. Stochastic model for BOD and DO in streams. J. Sanit. Eng. Div. ASCE, 92: 59–72.

Tung, Y.K. and Mays, L.W., 1980. Optimal risk-based design of hydraulic structures. Tech. Rep. CRWR-171, Center for Research in Water Resources, University of Texas, Austin, TX, 396 pp.