Optimum Sample Size for Development of Reaeration Coefficient Equation in Stream Water Quality Modeling

강물의 수질오염 modeling에 사용되는 再曝氣계수공식 개발을 위한 적정규모의 표본의 크기

Yoon, Chun Gyeong* · Charles S. Melching**
윤 춘 경

요 약

동일한 하천의 용존산소량(DO)을 예측하는 경우에도 사용하는 再曝氣계수(K_2)는 계산하는 공식에 따라 커다란 차이를 나타내며, 부적합한 공식의 사용에 의한 K_2의 계산은 하천의 수질관리 정책결정에 지장을 초래하므로 현장사정에 적합한 공식의 개발이 필요하다. 이러한 공식의 개발은 많은 현장측정 자료를 사용할수록 신뢰성이 높으나 현장측정은 소요되는 비용에 제약을 받기 때문에 신뢰성과 경제성을 동시에 고려한 표본의 크기의 적정규모를 산정하는 것이 필요하다.

본 연구에서는 Monte Carlo 방법에 의해 통계적으로 추출된 K_2를 사용해서, 주어진 자료에 의해 개발된 공식을 사용할 때 야기되는 오차가 K_2 개수의 증가에 따라 얼마나 감소하는지를 널리 사용되는 공식 중에 Owen공식과 Churchill공식을 New Jersey에 있는 Passaic River에 적용시켜 검토하였다. 표본의 크기가 10에서 20으로 증가할 때 오차가 크게 감소하였으며 20을 넘어 증가시켰을 때에는 오차의 감소폭이 미미하였다. 오차의 감소형태와 단위측정당 소용되는 비용을 고려할 때 약 20정도의 표본의 크기가 적정수준의 규모로 판단된다. 이러한 적용사례의 결과는 회귀모델의 이론적 계산결과에 의한 오차 감소와 흡사하여 본 연구결과는 여러 가지 K_2 공식과 광범위한 하천의 조건에 적용이 가능할 것이며, 본 연구에서 사용한 적정표본의 크기 산정방법은 회귀분석에 의해 실험식을 개발하는 다른 분야에도 적용이 가능하다.

---

* 건국대학교 농과대학
** Hydraulic Engineer, U. S. Geological Survey, Illinois District, 102 E. Main St., 4th Floor, Urbana, IL 61801 U. S. A.

키워드 : Reaeration coefficient, Water quality data, Sample size, Monte Carlo simulation, Regression analysis, Error analysis

I. Introduction

Reaeration is a physical absorption of oxygen from the atmosphere by water. It is the most important natural process by which streams with low dissolved-oxygen (DO) concentrations resulting from waste inputs may recover dissolved oxygen. DO concentration is a primary indicator of stream water quality and the reaeration coefficient (K_2) is a dominant input parameter affecting model-output reliability when estimating the DO concentration (Yoon and Melching 1992). Therefore, accurate estimation of K_2 is important for assessment of the assimilative capacity of streams. The methods for measurement of K_2 in streams are listed in Table-1.

Table-1. K_2 measurement methods and expected errors

Method	Expected Error	Reference
Dissolved-Oxygen Balance	65%	1
Disturbed-Equilibrium	115%	1
Radioactive Gas Tracer	15%	1
Modified Tracer Gas (Propane)	15%	2
Modified Tracer Gas (Ethylene)	25%	2

1. Bennett and Rathbun (1972).
2. Estimated error based on Grant and Skavroneck (1980) and Rathbun and Grant (1978).

Measurement of K_2 by these methods in streams may involve considerable error. Even if K_2 can be measured accurately for a particular period, it will vary with stream-flow conditions. Therefore, development of a reliable equation capable of estimating K_2 over a wide range of stream-flow conditions is necessary for prediction of DO concentrations. The development of site-specific K_2 estimation equation on the basis of reaeration measurements made in the stream is discussed in this paper with particular emphasis placed on the number of K_2 measurements required to obtain a reliable estimation equation.

Wilson and Macleod (1974) examined 16 published K_2 estimation equations with a data set of 400 K_2 measurements and found normalized mean errors varying from -35% to 701%. Grant and Skavroneck (1980) examined 20 published K_2 estimation equations with the results of 6 K_2 measurements from the radioactive tracer gas method, the most accurate method for measuring K_2 in a stream, and found normalized mean errors varying from -94% to 7,200%. St. John et al. (1984) illustrated the significance of K_2 on the oxygen balance and how the difference in estimates provided by various K_2 estimation equations can affect wastewater treatment plant design. The results indicated that the wastewater treatment level could vary from conventional secondary treatment to advanced treatment because of differences in K_2 estimation equations.

The reliability of the K_2 estimation equation increases with sample size for development of a site-specific equation. However, increased sampling involves higher cost and financial constraints require engineers to find an appropriate number of samples for equation development that balances reliability and cost. Several commonly used K_2 estimation equations, the number of samples used to derive them, and equation accuracy for the given data are shown in Table-2. The accuracy is expressed in percent standard error, E_p

Table-2. Commonly used K estimation equations and accuracy for the given data

Eq. No	Reference	K_2^1 (1/day)	No. of K_2 Field	Values Used Lab.	Percent Standard Error
i	Churchill et at. (1962)	$5.02 \frac{V^{0.969}}{H^{1.673}}$	30	–	28
ii	Owens et al.(1964)	$6.94 \frac{V^{0.73}}{H^{1.75}}$	32	–	35
iii	Owens et al.(1964)	$5.34 \frac{V^{0.67}}{H^{1.85}}$	68	–	32
iv	Krenkel & Orlob(1962)	$48.85 \frac{E^{0.408}}{H^{0.66}}$	–	58	15
v	Thackson & Krenkel(1969)	$42.79 \frac{V^*}{H}$	–	58	24
vi	Isaacs & Gaudy(1968)	$4.17 \frac{V}{H^{1.50}}$	–	52	41
vii	Bennett & Rathbun(1972)	$5.57 \frac{V^{0.607}}{H^{1.689}}$	121	–	38
viii	O'Connor & Dobbins(1958)[2]	$821.34 \frac{D_L^{0.5} S^{0.25}}{H^{1.25}}$	15	–	31
ix	O'Connor & Dobbins(1958)[2]	$292.95 \frac{(D_L) V^{0.5}}{H^{1.673}}$	15	–	45

1. $K_2$from these equations is for use with DO prediction equations formulated in log base e at 20℃.

2. Equations derived theoretically and tested with published field data.

[V, average velocity in the stream, m/s; H, average depth of flow, m; E, energy dissipation rate per unit mass of fluid, erg/sec-gram; V*, shear velocity, m/s; D_L, molecular diffusion coefficient, m^2/day; S, channel slope, m/m]

(Bennett and Rathbun 1972),

$$E_p = 100(1-10^{-E_{SL}}) \quad \cdots\cdots\cdots\cdots (1)$$

where $E_{SL}=\left[\frac{\sum_{i=0}^{N}(Log\, K_{2c,\, i} - Log\, K_{2m,\, i})^2}{N}\right]^{0.5}$,

$K_{2c,\, i}$ is the calculated value of K_2 for case i, $K_{2m,\, i}$is the measured value of K_2 for case i, and N is the total number of cases. In this study, synthetic K_2 data generated by Monte Carlo simulation and multiple linear regression are applied to determine an appropriate sample size for development of a site-specific K_2 estimation equation.

Error Model for the Reaeration-Coefficient Estimation Equation

Because significant error is involved in the measurement of K_2, an estimation equation derived from erroneous K_2 measurements may result in erroneous estimates because of uncertainties in ① the form of the equation and ② the data used to develop the equation. It is difficult to determine how large the estimation error may be, therefore, it should be addressed in statistical terms. The errors are implicitly assumed to be random variables. The basic model is

$$K_{2i}^{*} = F_i(Z) + e_i \quad \cdots\cdots (2)$$

where Z is the total set of input data, $F(\cdot)$ is the estimation equation, K_2^{*} is the true K_2 value corresponding to the estimation of F (Z), e is the estimation error, and the subscript i refers to the i^{th} value of K_2.

The following assumptions are typically made about the errors, e_1, e_2, $\cdots$, e_n (Troutman 1985) :

(1) the errors are statistically independent of the predictions, $F_i(Z)$, and are identically distributed;

(2) the errors are statistically independent of each other;

(3) the errors have a mean and a variance of $E[e_i]=0$, $Var(e_i)=\sigma^2$; and

(4) the errors are normally distributed.

The advantage of imposing the above assumptions is that the value for σ^2 summarizes the accuracy of the estimation. In the development of a site-specific K_2 estimation equation by regression analysis or other methods, the veracity of these assumptions should be examined. Troutman (1985) provides details on ways these assumptions may be examined and the analyses modified to meet these assumptions.

The input data set Z includes two subsets, parameters and variables, denoted by b and X, respectively. It is customary to assume that X is observed correctly and all input errors are in the parameter vector b, where a true but unknown parameter vector β exists. By partitioning of the input data set, Eq. 2 becomes

$$K_{2i}^{*} = Fi(X, \beta) + e_i \quad \cdots\cdots (3)$$

In statistics, the population is characterized by the true but unknown parameter values, and inferences about the population are made based on sample data sets. Unless the sample size is infinite, it is impossible to estimate the population parameters exactly. In this study, the population parameters under consideration are denoted by β, and sample parameters are denoted by b. Practically, β is defined to be the parameter values that minimize errors in the objective function, $E[K_2^{*} - F(X, \beta)]^{2*} \approx \min E[K_2^{*} - F(X, b)]^2$. When the statistical model is run with any arbitrary set of parameters b, which may not equal the correct set β, the estimation error is

$$K_2^{*} - F(X, b) = K_2^{*} - F(X, \beta) + F(X, \beta) - F(X, b) = e + \gamma(X, b) \quad \cdots\cdots (4)$$

where $\gamma(X, b)$ is the difference between estimations made using the two sets of parameters β and b. The mean error with parameter set b is

$$E[K_2^* - F(X, b) \mid X)] = E[e \mid X] + E[\gamma(X, b) \mid X] = \gamma \quad \cdots (5)$$

Because e has a mean of zero for any input X, $\gamma = E[\gamma(X, b) \mid X]$ represents the bias or mean amount by which the estimated K_2 value with parameter vector b deviates from the best estimate of K_2 from the true model with parameter vector β. This bias results in a larger mean-squared error of estimation than with the parameter vector β (Troutman 1985), given by

$$E[(K_2^* - F(X, b))^2] = \sigma^2 + E[\gamma^2(X, b)] \quad \cdots (6)$$

Many K_2 estimation equations are of the form $K_2 = a_1 V^{a2} H^{a3}$ with two variables, V (velocity) and H(depth). This equation is obtained by multiple linear regression with experimental and field data in the form of

$$\ln K_2 = \ln a_1 + a_2 \ln V + a_3 \ln H + \varepsilon \quad \cdots (7)$$

where ε is an error term with mean=0 and standard deviation=σ_ε, and ε is normally distributed. For the untransformed data, Eq. 7 is

$$K_2 = e^{\varepsilon} a_1 V^{a2} H^{a3} = \varepsilon_t a_1 V^{a2} H^{a3} \quad \cdots (8)$$

where ε_t is the transformed error series which is lognormally distributed. For a lognormally distributed variable like ε_t,

$$\zeta^2 = \ln\left(1 + \frac{\sigma^2}{\mu^2}\right) \quad \cdots (9)$$

where ζ=standard deviation of the logarithmically transformed variable, σ= standard deviation of the non-transformed variable, and μ=mean of the non-transformed variable. If $\sigma/\mu \leq 0.30$, then

$$\zeta \approx \sigma/\mu \quad \cdots (10)$$

which is a coefficient of variation (COV) of the non-transformed variable (Ang and Tang, 1975, p. 105).

Monte Carlo Evaluation of Equation Bias

Monte Carlo simulation involves repeated simulation of a process using a set of random parameter values generated in accordance with the corresponding probability distributions.

Monte Carlo simulation was used to generate statistically based synthetic K_2 data for assessment of the relation between sample size and reliability. Multiple linear regression is applied to the synthetic K_2 data to develop K_2 estimation equations representative of various sample sizes. Additional synthetic K_2 data are used to evaluate the bias error in the K_2 estimation equations representative of various sample sizes. The procedure is applied to the Passaic River in New Jersey.

Application to the Passaic River

From Table-2, the typical error of a site-specific K_2 estimation equation developed with field measurements can be assumed to be about 35%, i.e., $\sigma_\varepsilon \approx 0.35$. The procedure for optimum sample-size determination for a site-specific K_2 estimation equation is applied to the Passaic River in New Jersey. To estimate the probability distributions of the velocity and the depth of the Passaic River, the hydraulic simulation part of the output from QUAL2E-Passaic (New Jersey Department of Environmental Protection 1987), which was calibrated with field data from August 3-5, 1983 was used. The stream was in a low-flow condition and representative values of the velocity and the depth were selected for each of the 32 reaches. The frequency distributions of the velocity and depth for these reaches are shown in Figure 1, and the distributions are assumed to follow a beta distribution. The basic statistics and assumptions made are :

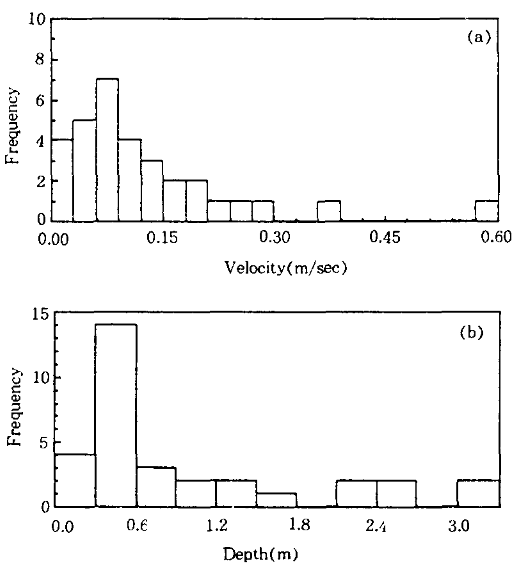

Fig. 1. Frequency distribution of velocity and the depth of the Passaic River from the low-flow survey of August 3-5, 1983

(i) the two selected "true" equations, $F(X, \beta)$, are

Owens et al. (1964) : $\ln K_2 = 1.94 + 0.73 \ln V - 1.75 \ln H + \varepsilon$ ·············· (11)

Churchill et al. (1962) : $\ln K_2 = 1.61 + 0.969 \ln V - 1.673 \ln H + \varepsilon$ ··········· (12)

(ii) e is normally distributed with mean and standard deviation

$\mu_\varepsilon = 0, \quad \sigma_\varepsilon = 0.35$

(iii) basic statistics of V and H for the Passaic River,

velocity (V, m/sec)

mean, $\mu_V = 0.129$

mode, $M_V = 0.076$

range, [r1, r2] = [0.016, 0.598]

depth (H, m) :

mean, $\mu_H = 0.950$

mode, $M_H = 0.457$

range, [r1, r2] = [0.160, 3.048]

(iv) beta distribution for both V and H.

Many random numbers were generated for V, H, and ε as per the appropriate distributions. Normal random numbers are generated for ε, and random numbers following a beta distribution are generated for V and H. Synthetic K_2 values are generated by Monte Carlo simulation with random numbers of V, H, and e and two assumed "true" equations, equations (11) and (12), and different data sets (5,000 K_2 values each) are prepared for multiple linear regression and error analysis. Two independent variables, V and H, and one dependent variable, K_2, are used in multiple linear regression to develop the K_2 estimation equation. An incorrect equation, F(X, b), is developed because of variation in ε, and the estimation of K_2 made with this equation differs from that made with F(X, β). The overall uncertainty in the estimates of K_2 from the erroneous equation is given by

$$VAR[\ln K_2]_j = \sigma_\varepsilon^2 + E[\gamma_j^2(X, b_j)], \text{ and} \quad \cdots (13)$$

$$SD[\ln K_2]_j = \sqrt{\sigma_\varepsilon^2 + E[\gamma_j^2(X, b_j)]} \quad \cdots (14)$$

where j denotes that this is the j^{th} realization of the value of SD[ln K_2] and E[γ_j^2 (X, b_j)] for the case of an estimation equation F (X, b_j) based on n samples and b_j are the model parameters for the j^{th} realization. For an accurate estimate of SD[ln K_2], many sets of n samples should be analyzed until the mean of the standard deviation of the natural logarithm of K_2

$$SD[\ln K_2] = \frac{1}{N_c}\sum_{j=1}^{N_c} SD[\ln K_2]_j \quad \cdots (15)$$

where N_c is the number of realizations necessary for SD[ln K_2] to converge to a consistent value.

Multiple linear regression was performed with the first 10 data sets and the equation obtained, F(X, b1), was

$$\ln K_2 = 2.24 + 0.875 \ln V - 1.77 \ln H \cdots (16)$$

This equation is different from the F(X, β) equation, therefore, a different estimate of K_2 results for the same velocity and depth. Equation 16 is applied to an independent data set to determine the variance due to bias. This independent data set is in descending order from data set number 5,000. Twenty data sets (data sets number 5,000 to 4, 981) are used to evaluate the expected value of γ_1^2(X, b_1), which is

$$E[\gamma_1(X, b_1)] = \frac{1}{m}\sum_{i=1}^{m}(F(X, \beta)_i - F(X, b_1)_i)^2 \cdots (17)$$

where m=20. Using Eq. 14, the SD[ln K_2]$_1$ can be obtained with E[γ_1^2(X, b1)] and $\sigma_\varepsilon^2 = 0.35^2$.

The next 10 data sets (data sets number 11 to 20) are used for multiple linear regression and F(X, b_2) is obtained. The variance resulting from bias for this equation is determined for another 20 independent data sets (data sets number 4,980 to 4,961). This procedure is repeated until the mean of SD[ln K_2] converges to a consistent value, and this consistent value is the average error in K_2 estimation by an equation developed for the case of 10 samples. Once the mean of SD[ln K_2] for 10 samples was obtained, other sample sizes were examined to evaluate the effects of sample size on the estimation error. Sample sizes of 10, 20, 30, 40, 50, 60, 80, and 100 were selected and the K_2 estimation equations of Owens et al. (1964) and Churchill et al. (1962) were examined to determine the appropriate sample size.

Results and Discussion

Example with the Equation of Owens et al. (1964)

The mean of SD[ln K_2] and associated 95 % confidence limits are presented in Table-3 for different sample sizes and three cases with the equation of Owens et al. (1964)

Table-3. Mean of the standard deviation of ln (K_2) and 95-percent confidence limits for different sample sizes with the equation of Owens et al. (1964) assumed true

Sample Size	σ_ε=0.35, m=20 Mean of SD[ln K_2] (K_2 in 1/day)	σ_ε=0.35, m=30 Mean of SD[ln K_2] (K_2 in 1/day)	σ_ε=0.25, m=20 Mean of SD[ln K_2] (K_2 in 1/day)
10	0.4236±0.0187	0.4308±0.0220	0.2981±0.0083
20	0.3802±0.0078	0.3970±0.0081	0.2713±0.0037
30	0.3657±0.0026	0.3658±0.0024	0.2636±0.0023
40	0.3643±0.0032	0.3641±0.0027	0.2603±0.0018
50	0.3643±0.0028	0.3641±0.0026	0.2577±0.0013
60	0.3581±0.0016	0.3575±0.0016	0.2568±0.0015
80	0.3563±0.0016	0.3564±0.0017	0.2557±0.0012
100	0.3548±0.0011	0.3548±0.0010	0.2537±0.0009
∞ (expected)	0.3500±0.0000	0.3500±0.0000	0.2500±0.0000

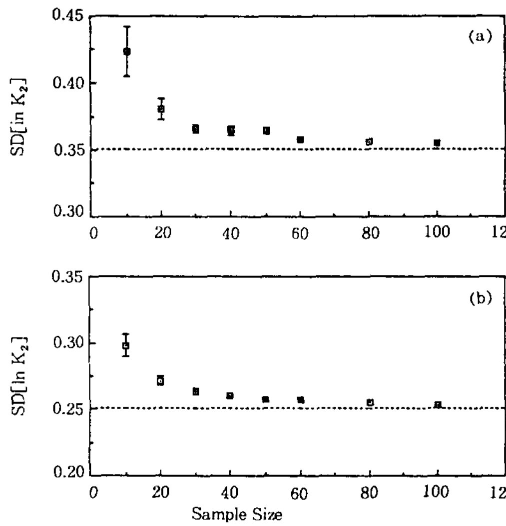

Fig. 2. Meamn of SD[ln K_2] and associated 95% confidence limits for K_2 estimation equation derived from different sample sizes generated from the equation of Owens et al. (1964) : (a) σ_ε=0.35, and (b) σ_ε=0.25

assumed to be the "true" equation. As sample size increases, average error decreases and the range of the 95% confidence limits also generally decreases. The error will be the same as σ_ε if sample size increases to infinity where F(X, b) becomes F(X, β) and no bias error is expected.

The mean of SD[ln K_2] with error bars of 95% confidence limits for different sample sizes with σ_ε=0.35 and the number of samples used to evaluate the variance resulting from bias (m) equal to 20 is illustrated in Fig. 2(a). Error bars for sample sizes of 20 or more are small and barely noticeable whereas the error bars are fairly large for a sample size of 10. This implies that, if a randomly taken set of 10 samples is used to develop a K_2 estimation equation, the error, SD [ln K_2], might deviate significantly from the mean of SD[ln K_2], whereas the potential deviations are not large for sample sizes of 20 or more. As sample size increases from 10 to 20, the mean of SD[ln K_2] decreases significantly. Further increases of sample size beyond 20 yield much smaller and more gradual reduction in the mean of SD[ln K_2]. Thus, the error resulting from equation bias can be reduced significantly by increasing the sample size from 10 to 20. However, collecting additional samples beyond 20 produces a small error reduction and is not justified. The result for the case of m=30 is presented in Table-3, and it is almost identical to that for m=20. Thus, m=20 is used hereafter.

The case of σ_ε=0.35 is fairly realistic relative to the quality of several commonly applied K_2 estimation equations (Table-2). Under good stream-flow conditions (i.e., prismatic channels with small ranges of flow and biological activity), a more accurate basic equation might be derived based on tracer-method measurements of K_2, where the accuracy is approximately ±15% (Table-1). A best case scenario was examined by assuming that K_2 values obtained by a tracer measurement method would yield an equation with $\sigma_\varepsilon=0.25$, and the result is illustrated in Fig. 2 (b). Results shown in Figs. 2(a) and (b) are similar in shape and a sample size of 20 can be considered appropriate for both cases. Uncertainty reduction for a sample size increase of 10 to 20 is less significant for $\sigma_\varepsilon=0.25$ (Fig. 2(b)) than for $\sigma_\varepsilon=0.35$ (Fig. 2(a)).

The accuracy of the simulation results may be tested by comparison to the theoretical variance in predictions made with regression models. If for a linear regression model the independent variables are normally distributed, then the total prediction variance, σ_p^2 , is equal to (Troutman 1982)

$$\sigma_p^2 = \sigma_\varepsilon^2+\{[2(n-1)]/[n(n-3)]\}\sigma_\varepsilon^2 \quad (18)$$

The total prediction variance as a function of sample size computed from Eq. 18 is listed in Table-4. The agreement between the total prediction variance in Table-4 and the mean of SD[ln K_2] in Table-3 is very close. For a lower sample size (10 to 20) the difference is larger, reflecting the fact that the independent variables follow beta distributions, but the variance from Eq. 18 is still comparable. For higher sample sizes the differences between a normal distribution and a beta distribution are less significant, and the agreement between the results in Tables-3 and 4 is very close.

Table-4. Total predication variance as a function of sample size for a regression model

Sample Size	$\sigma_\varepsilon=0.35$	$\sigma_\varepsilon=0.25$
10	0.3924	0.2803
20	0.3690	0.2636
30	0.3623	0.2588
40	0.3591	0.2565
50	0.3572	0.2552
60	0.3550	0.2543
80	0.3545	0.2532
100	0.3536	0.2525

For practical applications, engineers and planners are more concerned with the bias in K_2 than in the bias in ln K_2. Numerous studies have shown that bias or smearing of the relations can result from the retransformation from logarithmic to standard relations. The expected value of γ^2(X, b) was computed for the simulation results for K_2 in the same manner as for ln K_2. For example, the bias in the K_2 estimates for the first realization of an equation developed from 10 samples presented above is

$$\gamma_i(X, b_1)=6.94V_i^{0.73}/H_i^{1.75} - 9.39V_i^{0.875}/H_i^{1.77} \quad \cdots\cdots (19)$$

The mean and 95% confidence limits of {E[γ^2(X, b)]}$^{1/2}$ (simillar to a standard deviation resulting from bias) for estimation of K_2 for the case of $\sigma_\varepsilon=0.35$ are shown in Fig. 3 for a range of sample sizes. As was the case for ln K_2, the variance resulting from the bias in K_2 estimation reduces significantly by increasing sample size from 10 to 20, but further increases in sample size produce only small reductions in the variance resulting from bias. Also, the magnitude of the confidence

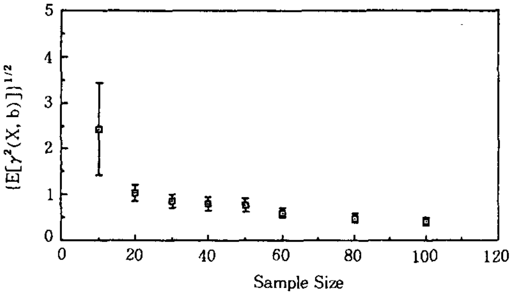

Fig. 3. Mean of $\{E[\gamma^2(X, b)]\}^{1/2}$ for K_2 and associated 95% confidence limits for K_2 estimation equation derived from different sample sizes generated from the equation of Owens et al.(1964)

limits decrease significantly as the sample size increases from 10 to 20, but further increases in sample size produce only small reductions in the magnitude of the confidence limits.

Example with the Equation of Churchill et al. (1962)

An additional 5,000 K_2 values were computed with Eq. 12 (instead of Eq. 11) with the ln V, ln H, and ε remaining unchanged. Multiple linear regression was performed utilizing these K_2 values with the data in descending order from data set number 5,000, and determination of the variance resulting from bias was performed with the data in ascending order from the first data set. This is opposite to the case with the equation of Owens et al. (1964). The mean SD[ln K_2] and associated 95% confidence limits are presented in Table-5 for different sample sizes and two cases with the equation of Churchill et al. (1962) assumed to be the "true" equation. These results are also illustrated in Figs. 4 (a) and (b). The mean and

Table-5. Mean of the standard deviation of ln (K_2) and 95-percent confidence limits for different sample sizes with the equation of Owens et al. (1962) assumed true

Sample Size	σ_ε=0.35, m=20 Mean of SD[ln K_2] (K_2 in 1/day)	σ_ε=0.25, m=20 Mean of SD[ln K_2] (K_2 in 1/day)
10	0.4296±0.0224	0.2958±0.0077
20	0.3774±0.0060	0.2720±0.0042
30	0.3661±0.0027	0.2639±0.0024
40	0.3633±0.0024	0.2623±0.0032
50	0.3613±0.0023	0.2577±0.0014
60	0.3580±0.0015	0.2565±0.0013
80	0.3562±0.0015	00.2554±0.0012
100	0.3547±0.0011	0.2536±0.0009
∞ (expected)	0.3500±0.0000	0.2500±0.0000

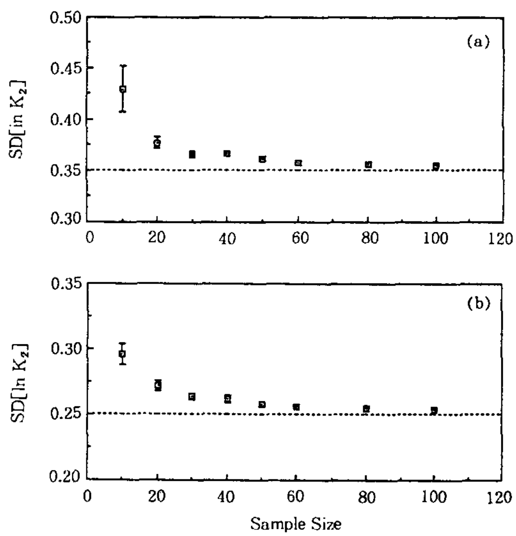

Fig. 4. Meamn of SD[ln K_2] and associated 95% confidence limits for K_2 estimation equation derived from different sample sizes generated from the equation of Owens et al. (1962) : (a) σ_ε=0.35, and (b) σ_ε=0.25

95% confidence limits of $\{E[\gamma^2(X, b)]\}^{1/2}$

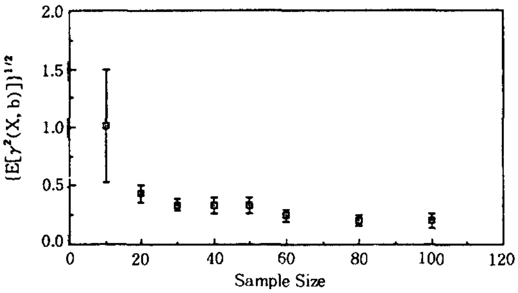

Fig. 5. Mean of $\{E[\gamma^2(X, b)]\}^{1/2}$ for K_2 and associated 95% confidence limits for K_2 estimation equation derived from different sample sizes generated from the equation of Owens et al. (1962)

for estimation of K_2 for the case of $\sigma_\varepsilon=0.35$ are shown in Fig. 5. The results from application of Eq. 12 are similar to those for Eq. 11, and the appropriate sample size is about 20.

Conclusions

The appropriate sample size for the development of site-specific K_2 estimation equations that balance cost and reliability was examined for the Passaic River in New Jersey. The K_2 estimation equations of Owens et al. (1964) and Churchill et al. (1962) were used in a Monte Carlo simulation analysis to generate synthetic K_2 data that were applied to evaluate reduction in prediction variance with sample size. All examples showed that about 20 samples can be considered an appropriate size for the Passaic River. This result is consistent with the suggestion of Ang and Tang (1975, p. 236) that when the sample size is greater than 20, the sample variance is a good estimator of the population variance. Further, the simulation results agree closely with theoretical results for regression models. Thus, the results of this study might be applicable for a wide range of reaeration-coefficient estimation equations and stream conditions. Testing of the multiple regression and equation-bias estimation procedures with field measurements of K_2 is needed to verify the bias reduction with sample size. Unfortunately, large data sets of field measured K_2 values are expensive to obtain and rare.

In typical water-quality modeling, standard K_2 estimation equations, which are functions built into the model, are used to calculate the K_2 values. Generally, the equations built into water-quality models are well-known equations. However, these were developed under conditions prevailing in a given river or for a specific set of laboratory conditions, and the estimation reliability for general field conditions is poorly known. For the stream system of interest, it is recommended that a site-specific K_2 estimation equation be developed rather than applying built-in equations because DO is one of the most important indicators of the stream water quality, and K_2 is often the dominant input parameter affecting the reliability of DO prediction. A site- specific K_2 estimation equation should be incorporated into the model. The cost for field measurement of the K_2 values by the propane gas method is approximately \$3,000～\$4,000 per measurement, depending on field conditions. Considering the high cost to develop a water-quality model, the use of the model in making multi-million dollar decisions regarding upgrading or expanding wastewater treatment plants, and the significance of the reaeration coefficient on the decision, spending about $60,000 for 20 K2 measurements to develop a more reliable model may be justified. The simulation procedure demonstrated in this study can be applied to other rivers to help make decisions regarding the appropriate sample size for development of a site-specific K_2 estimation equation before actual sampling starts.

Acknowledgements

The authors thank Dr. Brent Troutman and Dr. Ronald Rathbun of the Central Region Branch of Regional Research, U.S. Geological Survey, Lakewood, Colorado, for their detailed reviews and helpful comments that greatly improved this paper.

References

1. Ang, A. H-S. and W. H. Tang, 1975, Probability Concepts in Engineering Planning and Design, Vol. I, Basic Principles, John Wiley & Sons, New York.
2. Bennett, J. P., and R. E. Rathbun, 1972, "Reaeration in Open Channel Flow", U. S. Geological Survey Professional Paper 737.
3. Churchill, M. A., H. L. Elmore, and R. A. Buckingham, 1962, "The Prediction of Stream Reaeration Rates", Journal of Sanitary Engineering, ASCE, 88(SA4), 1-46.
4. Grant, R. S., and S. Skavroneck, 1980, "Comparison of Tracer Methods and Predictive Equations for Determination of Stream-Reaeration Coefficients on Three Small Streams in Wisconsin", U. S. Geological Survey, Water-Resources Investigations Report 80-19.
5. Isaacs, W. P., and A. F. Gaudy, 1968, "Atmospheric Oxygenation in Simulated Streams", Journal of Sanitary Engineering, ASCE, 94(SA2), 319-344.
6. Krenkel, P. A., and G. T. Orlob, 1962, "Turbulent Diffusion and the Reaeration Coefficient", Journal of Sanitary Engineering, ASCE, 88(SA2), 53-83.
7. New Jersey Department of Environmental Protection, 1987, "Passaic River Water Quality Management Study", Trenton, New Jersey.
8. O'Connor, D. J., and W. E. Dobbins, 1958, "Mechanism of Reaeration in Natural Streams", Transactions, ASCE, 123, 641-684.
9. Owens, M., R. W. Edwards, and J. W. Gibbs, 1964, "Some Reaeration Studies in Streams", International Journal of Air and Water Pollution, 8, 469-486.
10. St. John, J. P., T. W. Gallagher, and P. R. Paquin, 1984, "The Sensitivity of the Dissolved Oxygen Balance to Predictive Reaeration Equations", in Gas Transfer at Water Surfaces, W. Brutsaert and G. H. Jirka (Eds)., D. Reidel Publishing Co., Boston, 577-589.
11. Thackston, E. L., and P. A. Krenkel, 1969, "Reaeration Prediction in Natural Streams", Journal of Sanitary Engineering, ASCE, 95(SA1), 65-94.
12. Troutman, B. M., 1982, "An Analysis of Input Errors in Precipitation-Runoff Models Using Regression with Errors in

the Independent Variables", Water Resources Research, 18(4), 947-964.

13. Troutman, B. M., 1985, "Errors and Parameter Estimation in Precipitation-Runoff Modeling : Part-1. Theory", Water Resources Research, 21(8), 1195-1213.
14. Wilson, G. T., and N. Macleod, 1974, "A Critical Appraisal of Empirical Equations and Models for the Prediction of the Coefficient of Reaeration of Deoxygenated Water", Water Research, 8, 341-366.
15. Yoon, C. G., and C. S. Melching, 1992, "Sources and Reduction of Uncertainty in Stream Water Quality Modeling", Final Report to the New Jersey Water Resources Research Institute,New Brunswick, New Jersey.

(접수일자 : 1996년 6월 8일)