J. ENVIRONMENTAL SYSTEMS, Vol. 10(3), 1980-81

A POINT, NON-POINT SOURCE MODEL OF DISSOLVED OXYGEN FOR THE GREAT MIAMI RIVER*

STEVEN I. GORDON
Assistant Professor

RICHARD K. FROMUTH
Graduate Research Associate

Department of City and Regional Planning
The Ohio State University
Columbus Ohio

ABSTRACT

A procedure for forecasting bi-weekly dissolved oxygen (DO) levels has been validated for a portion of the Great Miami River in Southwestern Ohio. Point source pollution and land use (non-point) variables were related to mean annual DO levels at monitoring sites using multiple linear regression analysis. Bi-weekly DO concentrations were then generated assuming that they oscillate trigonometrically about annual mean DO as a function of time and historic variance. Test results indicate that standard errors of estimate were less than or equal to 1.5 mg/l for 50 per cent of all stations and less than or equal to 2.0 mg/l for 70 per cent of all stations sampled. Point source BOD, woodland, and commercial land were found to be significant variables in explaining the variance of mean annual DO concentrations.

INTRODUCTION

Several authors have noted the usefulness of statistical techniques in modeling water quality-land use relationships. Haith utilized a regression model to assess the contribution of various land uses to nitrogen, phosphorous, and suspended solids in New York State rivers [1]. Results of his study were used to estimate non-point pollution loads across twenty river basins. Stochastics Inc. has validated a probabilistic modeling procedure for the Ohio River and

---

*The effort was supported by a National Science Foundation Grant, No. NSF/ENV77-15020.

doi: 10.2190/E547-8WVN-N7WL-EY6F
http://baywood.com

the Potomac Estuary [2]. The model generates streamflow, BOD, and dissolved oxygen based on historic variations and predicts both temporal and spatial variation of DO under different flow and loading conditions. Carey et al. formulated a statistical model for the Raritan River in North Central New Jersey [3]. This model is an extension of work by Thomann [4]. The procedure assumes that annual dissolved oxygen variation is a trigonometric function of time. The Carey model simulated bi-weekly dissolved oxygen levels based on historic variation of bi-weekly concentrations about the mean. Mean Do is related to point BOD loads and land use/cover variables by means of a multiple linear regression equation developed from measured data.

These models contrast with more traditional approaches such as the Streeter-Phelps equation. Although they do not directly model the physical processes involved, they have the advantages of being relatively simple in structure, the capability of answering many important policy questions, and the ability to measure the relative importance and magnitude of non-point pollution problems. This ability is extremely important considering the ongoing section 208, areawide waste treatment management planning that is part of PL 92-500. The purpose of this paper is to report an adaptation of the Carey model to the Great Miami River Basin in Ohio and its application to water pollution policy questions.

METHODOLOGY

Calibration of a statistical model first requires the compilation of a consistent, geographic and time series data set. Data for the Great Miami was unfortunately scattered among many agencies in varied forms. However, we were fortunate to be able to obtain a consistent, long term data set on dissolved oxygen levels throughout the basin from the Miami Conservancy District. The study area is shown in Figure 1. Figure 2 shows the location of the twenty-four sampling stations for which data were compiled. Data sets of this nature are increasingly available for many river basins because of the monitoring activities of U.S. EPA, state environmental agencies, river basin commissions, water companies, and other public entities. Reformatting of these and other data resulted in the requisite information for deriving a statistical DO model:

- Bi-weekly dissolved oxygen data for monitoring stations along the main stem of the Great Miami River.
- Point BOD loads for all major sources on the main stem of the river. Major sources include industries and sewage treatment plants.
- BOD loads for all major tributaries.
- Slope of the stream bed between the dissolved oxygen monitoring stations
- The geologic provinces within the river basin.
- Rainfall data.

- The urban BOD component of each area analyzed (a reach area defined as the area, between oxygen stations, from which runoff drains into the main stem of the river). Reach areas were confined to the vicinity of the main stem since the effects of major tributaries on stream BOD were obtained from measured data.
- Variables representative of rural, non-point BOD loads for each reach.
- Cross section and streamflow data for each DO monitoring station.

The next step in the modeling process was the derivation and testing of three model equations. The first is a multiple linear regression equation for

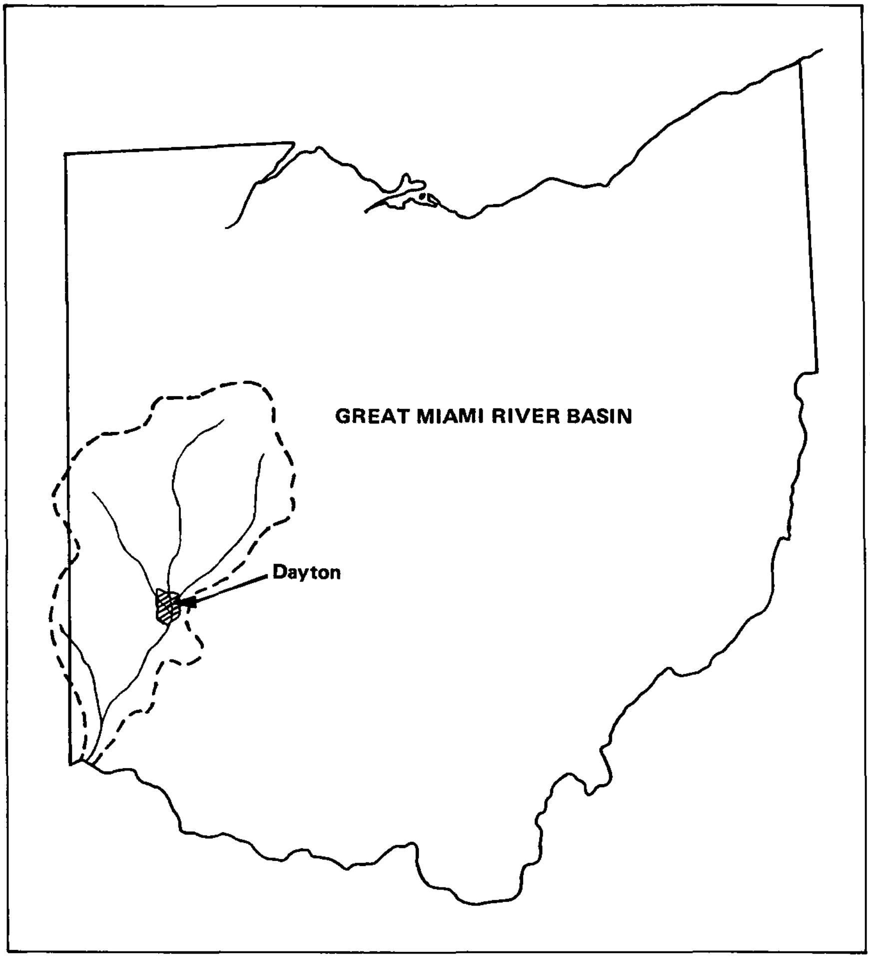


Figure 1. Location of study area in Ohio.

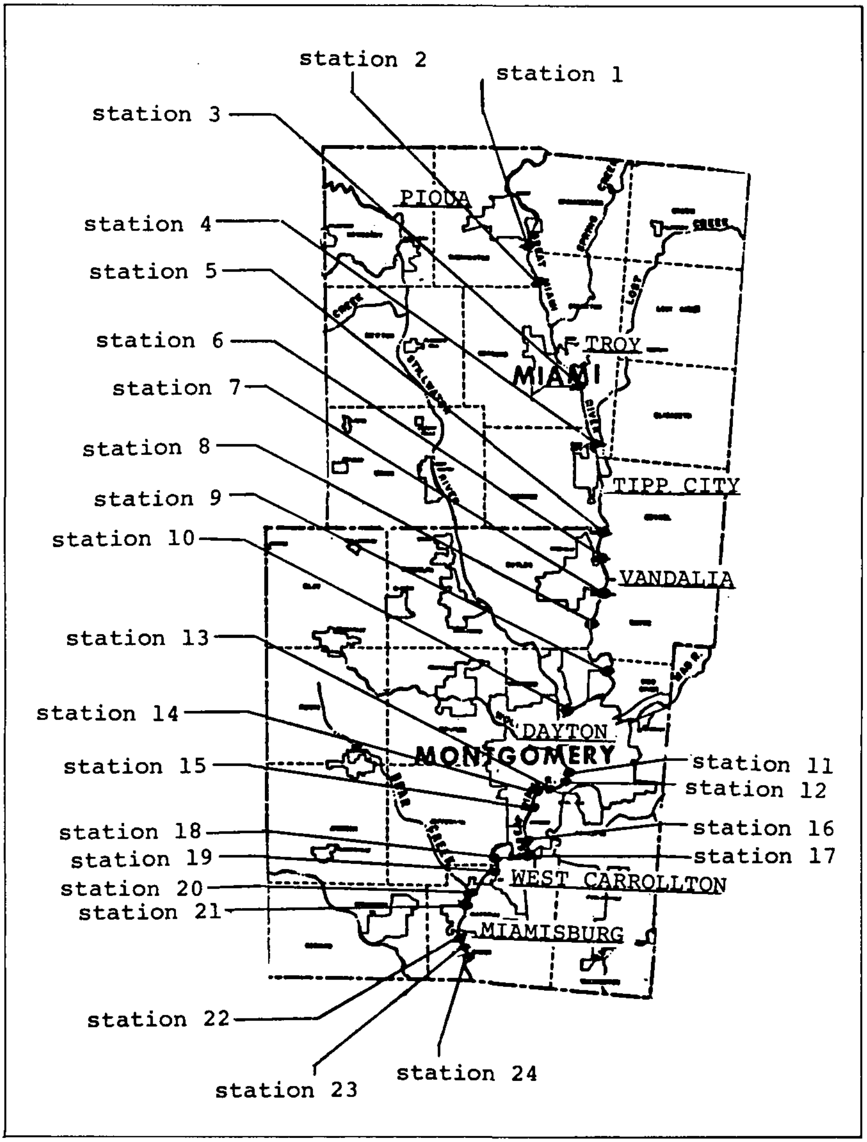


Figure 2. Map of study region showing D.O. stations and major municipalities.

predicting mean annual dissolved oxygen as a function of such variables as point source BOD, land use, land cover, mean annual streamflow, etc. Symbolically, the equation is given as:

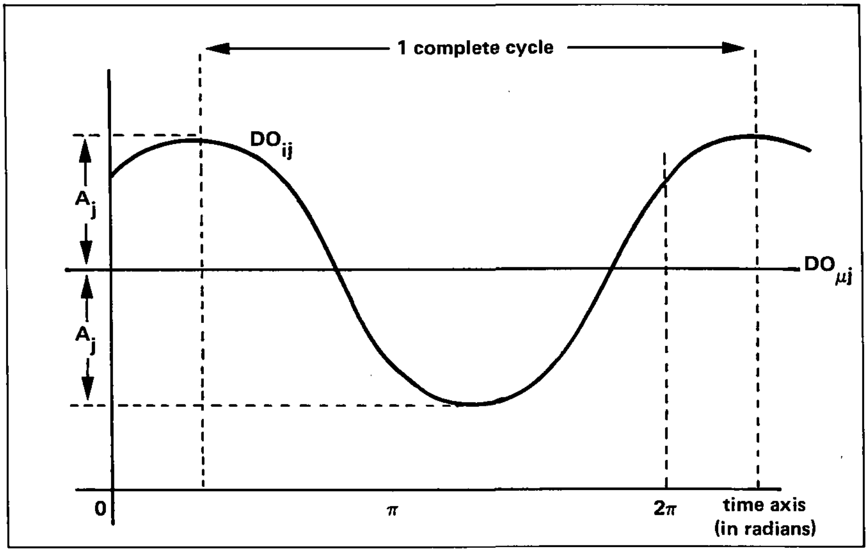
$$DO_{\mu j} = \sum_i a_{ij} P_{ij} + \sum_k b_{kj} F_{kj} + \epsilon \qquad (1)$$

where

$DO_{\mu j}$ = mean dissolved oxygen level for station j (mg/l)
P_{ij} = a set of i pollution variables related to mean DO at station j
F_{kj} = a set of k stream variables related to mean DO at station j
ϵ = an error term.

A total of thirty-seven linear regression equations of this form were tested using various combinations of pollution and stream variables that should be related to DO levels. The final equation selected and reported below was chosen based on measures of statistical significance, accuracy, and theoretical meaningfulness of the equations.

The second model equation is based on the assumption that DO variation is a trigonometric function of time. This is represented in graphical form by Figure 3. Here, one can see that the variation around $DO_{\mu j}$ is a curve of amplitude A_j that is a function of time. According to Carey et al. this amplitude is related to dissolved oxygen variance for a given station [3]:



Source: Carey et al. [3, p. 151]

Figure 3. Graphical representation of annual DO cycle.

$$A_j = (2\sigma_j^2)^{1/2} \tag{2}$$

where

A_j = amplitude of the dissolved oxygen curve for station j (mg/l)
σ_j^2 = variance of bi-weekly dissolved oxygen readings for station j around $DO_{\mu j}$ (mg/l)

After computing variances for forty-one station years of DO data for 1973-75, amplitudes were computed and related to the variance around mean DO using a second linear regression:

$$A_j = b_1\, DO_{\mu j} + \epsilon \tag{3}$$

where ϵ = residual term.

The third model equation simulates the bi-weekly dissolved oxygen levels represented by the curve in Figure 3 and given as:

$$DO_{ij} = DO_{\mu j} + A_j \cos\left(\frac{\pi(T_{ij})}{13} - \theta_j\right) \tag{4}$$

where

DO_{ij} = the ith bi-weekly dissolved oxygen level for station j
$DO_{\mu j}$ = mean DO level from equation (1)
A_j = amplitude of the curve from equation (3)
T_{ij} = the bi-weekly period (fortnight) associated with the ith dissolved oxygen level at station j. Values range from 1 to 26.
θ_j = the phase constant for station j which permits DO_{ij} to reach a maximum value at the correct time. The value of θ was estimated empirically.

Data for the Great Miami River basin for 1973-75 were compiled in a format enabling the solution of equations (1) and (3) using multiple linear regression. The remainder of the work then consisted of:

1. **Verification** – in this test, the model was used to predict DO levels for 1973-75. A regression analysis of predicted versus observed DO levels was then performed. This test was used to check the agreement of the model with the data used to derive it.
2. **Validation** – the model was used to predict bi-weekly DO levels for 1976. Predicted and observed DO levels for 1976 were then compared using regression analysis. This procedure tested the accuracy of the model as a simulator of DO levels.
3. **Sensitivity Analysis** – changes in simulated DO levels caused by changes in the assumptions about BOD transfer downstream were evaluated using this test. The analysis was used to determine the accuracy with which

the BOD transfer must be measured when using the model. BOD transfer is simply a means of introducing the spatial character of deoxygenation into the model.

4. **Policy Analysis** – probable changes in the treatment and disposal of municipal sewage were modeled to determine their impacts on the DO levels in the Great Miami River. The changes include a regionalized sewage treatment plant and 50 per cent upgrading of a major sewage treatment plant.

Results

Mean annual dissolved oxygen was found to be a function of four variables:

$$DO_{\mu j} = 10.317 + 0.173\ BODU_j - 0.0617\ BODS_j - 0.399\ FOREST_j + 0.684\ COMM_j \qquad (5)$$

where

$DO_{\mu j}$ = mean annual DO level for station j

$BODU_j$ = mean daily point source BOD load transferred to station j from the upstream station j–1 (10^3 lb/day)

$BODS_j$ = the mean daily point BOD generated in the stream segment represented by station j (10^3 lb/day)

$FOREST_j$ = land in reach j that is forested plus forest land that is transferred from upstream (10^3 acres)

$COMM_j$ = land in reach j in commercial use plus commercial land from upstream (10^3 acres).

Downstream transfer of both point BOD ($BODU_j$) and non-point BOD ($FOREST_j$ and $COMM_j$) was calculated using a step function related to stream slope as derived by Carey et al. This is shown in Table 1. The greater the slope, the lower the overall efficiency of water transfer. This means there is more time for the BOD to be assimilated in a stream segment and therefore less BOD load remains to be transferred downstream. The transfer function was tested by Carey et al. and by the authors. Our results are given in our sensitivity analysis section below.

The amplitude of the bi-weekly DO curve was found to be

$$A_j = 6.234 - 0.348\ DO_{\mu j} \qquad (6)$$

Given these equations, we could then solve equation (4) for the bi-weekly DO levels for each station j. The bi-weekly interval was chosen as most compatible with the sampling frequencies used in the Great Miami River Basin.

Table 1. Step Function of BOD Passed Downstream Under Different Slope Conditions

Status of Stream Flow	*Slope*	*Percent BOD Accumulated at Downstream Station*	*Remarks*
Fast	$S < 2.5$	100%	BOD transported: no assimilation
Medium	$2.5 < S < 2.5$	50%	Some BOD assimilated
Slow	$S > 5$	0%	All BOD assimilated

Source: Carey et al., [3, p. 156].

Model Verification

Using 1973-1975 data, bi-weekly dissolved oxygen levels were predicted using the model and then compared to the actual measurements made during those years. Table 2 shows that the equation is an excellent predictor of bi-weekly DO levels. Regressions of observed versus expected (model generated) values show most stations with over 95 per cent of the variance explained. Only four stations exhibit a greater level of error. Another way to look at the error is in terms of DO level directly. This is shown in Table 2 in terms of the standard error of estimate and mean error. This table shows that the DO levels at the majority of stations was below 2.0 mg/l. This is not as good as the results of Carey et al. but still represents an acceptable level of model error for some policy applications [3].

Model Validation

Data for 1976, not used in model derivation, were employed to validate the model's predictive power. The 1976 dissolved oxygen levels were simulated for nineteen stations in the study region and then compared with observed data to check the model's validity as a predictive tool. Standard errors of estimate, mean error, and R^2 values for each station are given in Table 3. The mean of the standard errors for all of the stations was 2.0 mg/l. Sixty-eight per cent of the stations had standard errors of 2.0 mg/l or less and nine of the stations (approximately 50 per cent of the total) had standard errors between 1.0 and 1.5 mg/l. Dissolved oxygen concentrations at stations 1, 9, and 10 (all grab sampling sites) were predicted inaccurately by the model. Large discrepancies in predicted and measured results relate to sampling instrument error, the influence of algal blooms and possibly to such factors as illegal dumping of organic material. Still, R^2 values remain high for all stations showing the model performs quite well as a predictor of DO.

Table 2. Standard Error of Estimate, Mean Error of Estimate, and R^2 Without Intercept for 1973-75 for Predicted Bi-Weekly DO Levels

Station I.D.	*River Mile*	*Type of Station*	*No. of OBS.*[c]	*Standard Error (MG/L)*	*Mean Error (MG/L)*	*R^2 Without Intercept*	*Mean Observed DO (MG/L)*
1	111.90	G[a]	31	1.5	1.2	.99	9.1
2	108.04	G	33	2.2	1.5	.95	9.6
4	99.00	M[b]	66	1.4	1.0	.98	10.2
5	93.98	G	31	2.5	1.6	.93	9.3
6	92.45	G	26	2.2	1.7	.97	11.0
7	90.87	G	34	1.5	1.2	.97	9.0
8	89.45	G	31	2.2	1.7	.95	9.9
10	82.68	G	31	2.5	1.9	.94	10.0
11	77.96	M	73	1.3	1.0	.99	10.8
14	76.36	G	44	1.7	1.3	.97	9.7
16	72.91	G	33	1.3	1.1	.98	9.4
17	72.72	G	31	1.3	1.2	.99	10.2
18	69.00	G	48	1.9	1.6	.96	9.0
20	66.43	G	25	2.0	1.5	.94	8.3
21	65.75	M	69	1.6	1.2	.97	8.6
22	64.34	M	63	2.0	1.2	.94	8.1
24	63.82	M	58	1.4	1.0	.97	8.0

[a] Grab Station: sampling frequency less than once per week.

[b] Monitor Station: DO level recorded at 2 hour intervals.

[c] An observation is the average DO level for a bi-weekly period. In the case of monitor data, an observation represents the mean of two weeks of 2-hourly DO readings. In the case of grab sampling stations, a single observation may be the only DO measurement for the bi-weekly period.

Model Sensitivity Analysis

The slope transfer assumptions given in Table 1 were used, in the development and application of the dissolved oxygen model, as a means of transferring point and non-point source BOD from upstream to downstream reaches. Because of possible errors in these assumptions as well as possible errors in the measurement of stream bed slope, the additional slope transfer assumptions shown in Table 4 were tested for their effects on the predictive accuracy of the model. The results of this analysis are summarized in Table 5. Both tests were performed assuming 1976 conditions and results of each were compared to those obtained using the original assumptions. The original slope transfer assumptions, overall, give the best results. However, even in the case of test 2, which nearly eliminated BOD transfer, only minor changes in the standard error values resulted. This confirms the model's insensitivity to changes in assumptions concerning the BOD transfer process.

Table 3. Standard Error of Estimate, Mean Error of Estimate, and R^2 Without Intercept for 1976 for Predicted Bi-Weekly DO Levels

Station I.D.	*River Mile*	*Type of Station*	*No. of OBS.*[c]	*Standard Error (MG/L)*	*Mean Error (MG/L)*	*R^2 Without Intercept*	*Mean Observed DO (MG/L)*
1	111.90	G[a]	26	3.1	2.4	.91	9.9
2	108.04	G	10	0.9	0.8	.99	10.3
3	103.55	G	24	1.1	0.9	.99	9.9
4	99.00	M[b]	25	1.2	0.9	.99	10.4
5	93.98	G	10	1.4	1.2	.99	10.6
6	92.45	G	26	1.5	1.1	.99	10.7
7	90.87	G	19	1.8	1.4	.97	10.0
8	89.45	G	10	2.2	1.6	.97	11.3
9	86.60	G	25	4.7	4.3	.97	14.2
10	82.68	G	10	5.2	3.5	.88	13.0
11	77.96	M	26	1.7	1.3	.98	11.4
14	76.36	G	15	0.7	0.6	.99	10.6
16	72.91	G	10	1.2	1.0	.99	10.1
17	72.72	G	26	1.5	1.2	.98	9.8
18	69.00	G	10	1.8	1.4	.96	8.9
20	66.43	G	10	1.6	1.2	.98	8.6
21	65.75	M	25	1.8	1.4	.97	7.7
22	64.34	M	25	1.9	1.6	.97	7.1
24	63.82	M	24	2.7	2.0	.93	7.0

[a] Grab Station: sampling frequency less than once per week.

[b] Monitor Station: DO level recorded at 2 hour intervals.

[c] An observation is the average DO level for a bi-weekly period. In the case of monitor data, an observation represents the mean of two weeks of 2-hourly DO oberservations. In the case of grab sampling stations, a single observation may be the only DO measurement for the bi-weekly period.

Discussion

The final mean dissolved oxygen equation (5) gives a number of insights into point and non-point water pollution in the Great Miami River Basin. First, it is important to note that cropland, although tested as an explicit land use variable, did not prove to be a significant explanatory variable. This could be an accurate representation of conditions but might also reflect a bias in the basic data. Since the land use data were derived from LANDSAT interpretations that might be prone to error [5], data accuracy was verified using a sample from aerial photography.

Use of this information in conjunction with statistical results provides an interesting insight into the non-point pollution problems in this basin. The

Table 4. BOD Transfer Assumptions Tested for Dissolved Oxygen Model Sensitivity Analysis

Test Number	*Slope Category (ft./mile)*	*Percent of BOD Transferred Downstream (%)*
1	Greater than 5.0	0
	Between 2.5 and 5.0	30
	Less than 2.5	80
2	Greater than 5.0	0
	Between 2.5 and 5.0	1
	Less than 2.5	10
Original Slope Assumption	Greater than 5.0	0
	Between 2.5 and 5.0	50
	Less than 2.5	100

coefficients of COMM and FOREST in equation (5) are opposite in sign of what one might expect. Our analysis indicates several possible explanations:

1. The regions within the Great Miami Basin which are classified as forested by LANDSAT imagery, are often lightly wooded, more steeply sloped than surrounding cropland and located adjacent to streams. Animal wastes, decaying leaves or other vegetation, and wastes from malfunctioning septic tanks may be contributing to lower dissolved oxygen levels in reaches of the Great Miami River adjacent to or near wooded areas.
2. The long term effect of urban development in the Great Miami River Basin, as measured by the amount of commercial land, is to reduce non-point BOD. Although there is a BOD component in urban non-point runoff, the concentration is sufficiently low so as to actually improve the water quality of the stream in terms of mean dissolved oxygen levels. Consequently, it is the rural, not the urban, component of non-point runoff that is contributing to the low dissolved oxygen levels in the Great Miami Basin. This finding is in agreement with that of Jalal [6].
3. Cropland is not a major contributor to rural, non-point BOD along the Great Miami River. Variables associated with cropland did not correlate with mean dissolved oxygen. It appears, however, that land used for pasture, which is often near streams, is a prime source of rural non-point BOD.

Table 5. Results of Slope Sensitivity Analysis

Standard Errors of Estimate for 1976 Simulation (mg/l)

Station	*Original Slope Assumptions*	*Test 1*	*Test 2*
1	3.1	3.1	3.2
2	0.9	1.0	1.3
3	1.1	1.1	1.3
4	1.2	1.2	1.4
5	1.4	1.3	1.6
6	1.5	1.4	1.7
7	1.8	1.8	1.9
8	2.2	2.2	2.4
9	4.7	4.7	4.9
10	5.2	5.1	5.0
11	1.7	1.6	1.6
14	0.7	0.8	1.1
16	1.2	1.1	1.1
17	1.5	1.6	1.6
18	1.8	1.8	1.8
20	1.6	1.6	1.7
21	1.8	1.8	1.6
22	1.9	2.0	2.3
24	2.7	2.9	3.2
	$\bar{x} = 2.0$	$\bar{x} = 2.0$	$\bar{x} = 2.1$

Policy Analysis

There are a number of water quality proposals in the study region which have important planning implications. In order to test the usefulness of the dissolved oxygen model as a planning tool, the following proposals were analyzed:

1. construction of a regional sewage treatment plant near the city of Vandalia, Ohio;
2. upgrading of the Dayton sewage treatment facility.

The modeling analysis indicated that both of these proposals, if implemented, would increase dissolved oxygen concentrations in the river. However, the improvement caused by the Dayton plant upgrading was small, indicating the high contribution of other pollution sources to the low dissolved oxygen levels in the vicinity.

CONCLUSION

Although the exact form of the basic model equations differs between the two studies, the results of the research indicate that the linear modeling procedure used by Carey, et al. in the Raritan River study is transferrable to the Great Miami River. In general, the Great Miami version of the model is less accurate than that obtained for the Raritan. The chief cause of this is most likely the absence of an algae parameter in the predicting equations.

This model cannot replace more deterministic approaches to modeling the short term behavior of dissolved oxygen. However, it is inexpensive to operate and should be quite useful in the preliminary analysis of land use or water quality management proposals.

Its advantages over other approaches include the following:

1. provides an explicit estimation of the impact of non-point sources on water quality and identifies the major sources and their location.
2. takes advantage of increasingly available water quality monitoring data.
3. is easily and explicitly tested relative to its reliability and degree of error.
4. can be utilized to test the effects of long term point and non-point pollution control policies on the DO levels.

Model application is limited by the following:

1. The model is one dimensional. That is, it predicts dissolved oxygen levels for only those points where DO has been measured (i.e., monitoring stations). Dissolved oxygen concentrations may vary over the length, width and depth of a stream. Some two dimensional models are available for studying this variation.
2. Maximum and minimum dissolved oxygen levels at stations for a given bi-weekly period are not predicted by the model. Only estimates of bi-weekly means are forecast.
3. The model is sensitive to land use and land cover variables (commercial land and forested land). It is important that inventories of such data are kept up to date if accurate forecasts of dissolved oxygen are to be obtained.

Overall, the model provides a reasonable estimate of long-term dissolved oxygen trends and a measure of the relative impacts of non-point pollution problems on DO. It cannot be expected to forecast the impacts of such short term trends as periodic low flow, floods, and sewage treatment plant overflows. Statistical models such as this should complement design level models used to determine site specific waste load allocations and should take advantage of increasingly available water quality monitoring and land use data.

APPENDIX I

Notation

The following symbols are used in this paper:

A_j = amplitude of the dissolved oxygen curve for station j (mg/lb)

b_i = coefficient for mean annual dissolved oxygen term in equation relating dissolved oxygen curve amplitude to mean annual dissolved oxygen

b_{kl} = coefficient for k^{th} stream variable related to mean annual dissolved oxygen at station j

$BODU_j$ = mean daily point source BOD load transferred to station j from the upstream station j − 1 (10^3 lb/day)

$BODS_j$ = mean daily point BOD generated in the stream segment represented by station j (10^3 lb/day)

$COMM_j$ = land in reach j in commerical use plus commercial land transferred from upstream (10^3 acres)

DO_{ij} = the i^{th} bi-weekly dissolved oxygen level for station j

$DO_{\mu j}$ = the mean annual dissolved oxygen level for station j (mg/l)

F_{kj} = the k^{th} pollution variable related to mean annual DO at station j

$FOREST_j$ = land in reach j that is forested plus forest land in upstream reaches that is transferred (10^3 acres)

P_{ij} = i^{th} pollution variable related to mean annual dissolved oxygen at station j

S = stream bed slope (pt/mile)

T_{ij} = the i^{th} bi-weekly time period for which the dissolved oxygen level is being predicted at station j

α_{ij} = coefficient for the i^{th} pollution variable related to mean annual dissolved oxygen at station j

ϵ = residual term associated with the regression equation relating dissolved oxygen curve amplitude to mean annual dissolved oxygen

π = 3.14176

σ_j^2 = the variance of all bi-weekly dissolved oxygen levels observed at station j

θ_j = the phase constant for station j used to control the time during the year at which dissolved oxygen reaches a maximum. θ is calibrated empirically.

REFERENCES

1. D. A. Haith, Land Use and Water Quality in New York, River, Journal of the Environmental Engineering Division, *Proceedings of the American Society of Civil Engineers, 102*:EE1, February, pp. 1-28, 1976.
2. U.S. Army Corps of Engineers, *Models and Methods Applicable to Corps of Engineers Urban Studies,* AD 786 516, National Technical Information Service, Springfield, Virginia, August, 1974.
3. G.W. Carey, L. Zobler, M.R. Greenberg and R.M. Hordon, *Urbanization, Water Pollution and Public Policy,* Center for Urban Policy Research, Rutgers University, New Brunswick, New Jersey, pp. 150-158, 1972.
4. R.V. Thomann, Time Series Analysis of Water Quality Data, Journal of the Sanitary Engineering Division, *Proceedings of the American Society of Civil Engineers, 93*, pp. 1-23, February, 1967.
5. S. I. Gordon, Utilizing LANDSAT Imagery to Monitor Land Use Change, a Case Study in Ohio, *Remote Sensing of Environment, 9,* pp. 189-196, 1980.
6. K. F. Jalal, Water Quality Impacts of Urbanization—A Methodology, *Journal of the Environmental Engineering Division, Proceedings of the American Society of Civil Engineers, 103*:EE1, pp. 49-57, February, 1977.

ACKNOWLEDGEMENT

A number of state and regional agencies in Ohio were helpful in providing data for this research project. They include the Miami Conservancy District, The Water Resources Division of the U.S. Geological Survey, The Ohio Environmental Protection Agency, and The Ohio Department of Natural Resources.

Direct reprint requests to:

Steven I. Gordon
Dept. of City and Regional Planning
The Ohio State University
289 Brown Hall
190 West 17th Avenue
Columbus OH 42210