*Proceedings of the 1992 Winter Simulation Conference*
*ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson*

# USE OF SIMULATION METHOD FOR SURFACE WATER QUALITY DATA

Tiao J. Chang

Civil Engineering Department
Ohio University
Athens, Ohio 45701

Rabah Y. Odeh

Civil Engineering Department
Ohio University
Athens, Ohio 45701

## ABSTRACT

A simulation model, based on the first-order Markovian process, was introduced to generate synthetic monthly data for water quality, namely, pH, dissolved oxygen, water temperature, and specific conductance. Monthly means, standard deviations, and correlation coefficients of the synthetic and historic data were computed, respectively, for the purpose of comparison. It is found that the simulation model is able to preserve the statistics of means, standard deviations and correlations of the historic data. Furthermore, the estimated autocorrelations of the synthetic data were plotted to compared with their counterparts of the historic data. The results show that the autocorrelations of historic water quality data are preserved by the simulation model.

## 1 INTRODUCTION

Simulation is a useful tool that has been used in many applications for the long term studies. Particularly in the water resources engineering, planners use the simulation technique to reduce the design risk that could have been involved if the design is based on the available historic data, which are relatively short compared with the design life. A good example is the design of a reservoir to be used for water-supply storage or flood control. In such case, it is likely that the length of the historic flow data are shorter than the proposed economic life of the structure. Furthermore, it is highly unlikely for the exact pattern of the records to recur during the life time of the structure. This leads to the conclusion that the worst flood (or drought) based on the records may not be the worst possible flood (or drought) in the design life of the structure (Chang, 1989). Therefore, it is desirable to use the synthetic data for the analysis. In such situations, hydrologic simulation may be used for generating synthetic data that preserve certain statistics of the historical records.

Furthermore, water resources planners may be interested to know whether a certain water resource system will fail to meet quality standards. For instance, in studying the pattern of the dissolved oxygen level in a river, the simulated data can be used for an estimation of the expected frequency that the dissolved oxygen fails to meet the standard (Fiering and Jackson, 1981). It is important to emphasize the fact that simulation is a tool for planning when applied to the hydrologic data. The statistical methods used for generating synthetic data do not pretend to provide casual models for actual flows. Nevertheless, simulated data are realistic enough that their applications will improve the water resources planning process for the purpose of risk reduction.

One of the basic assumptions behind the simulation technique is that the synthetic data are considered to be the results of a random process that depends somehow on the elements of chance and probability. It does not assume that the actual values of the variable studied are to be predicted. However, what we do expect is that certain characteristics of the data are to be preserved. For example, a Markovian model for a simulation of flow series assumes that high flows tend to be followed by high flows and low flows tend to be followed by low flows, which is expressed by a deterministic term. Hence, a simulation model can be constructed by including a deterministic term and a random component to reflect the uncertainty of the process.

Since historic records of water quality data are relatively short, it is desirable to obtain their synthetic data. Based on the available monthly records of water quality, a simulation method was introduced in this study to generate synthetic data of monthly water quality, namely, pH, dissolved oxygen, water temperature, and specific conductance. The simulation model is designed to preserve certain statistical characteristics of historic data, namely, mean, standard deviation, and first-order correlation.

## 2 SIMULATION MODEL FOR WATER QUALITY

A first-order Markovian process is introduced in the following for the purpose of water quality simulation. Let the sequence $\{x_i\}$ start with the following form :

$$q_i = d_i + e_i, \qquad (1)$$

where $q_i$ is the $i^{th}$ value, $d_i$ is the $i^{th}$ deterministic component and $e_i$ is the $i^{th}$ random component in the process, which is assumed to be independently distributed with a mean of zero and a constant variance. Furthermore, the deterministic component, $d_i$, in Equ. 1 can take the linear autoregressive form as follows:

$$d_i = \beta_0 + \beta_1 q_{i-1} + \ldots + \beta_m q_{i-m}, \qquad (2)$$

where $ß_0$, $ß_1$, ......, $ß_m$ are constants, $q_{i-1}$, $q_{i-2}$, ....., $q_{i-m}$ are the proceeding values of $q_i$ in Equ. 1. Let m = 1 in Equ. 2, the process assumes that the current value depends on one immediate preceding value only. This is reduced to the first order Markovian model of the form:

$$q_i = \beta_0 + \beta_1 q_{i-1} + e_i, \qquad (3)$$

where $ß_0$ and $ß_1$ are constants.

For a time series, $\{q_i\}$, with a mean of $\mu$, lag-one serial correlation coefficient of $\rho$ and a variance of $\sigma^2$, the process of the following form can be used:

$$q_i = \mu + \rho(q_{i-1} - \mu) + e_i, \qquad (4)$$

where $\{e_i\}$ are assumed to be normally distributed with a variance of $\sigma^2_e$, which is related to the variance of time series, $\sigma^2$, by

$$\sigma^2 = E[\rho(q_{i-1} - \mu) + e_i)]^2, \qquad (5)$$

where E denotes the expectation. This results in a relation between the variance of $\{q_i\}$ and that of $\{e_i\}$ by the following form:

$$\sigma_e^2 = \sigma^2(1 - \rho^2), \qquad (6)$$

where $\rho$ is the lag-one serial correlation coefficient of the time series.

Assuming that $\{t_i\}$ is a sequence of normally distributed and serially independent random variables with a mean of zero and a standard deviation of one, then $\{e_i\}$ can be replaced by $t_i$ by the following relation:

$$e_i = t_i \sigma \sqrt{(1 - \rho^2)} \qquad (7)$$

This process was used in this study for the simulation of monthly water quality by preserving the mean, variance, and first order correlation coefficient of the historic data.

## 3 APPLICATION AND DISCUSSION

Since water quality records are relatively short, it is desirable to obtain their synthetic data. Data of monthly water quality, namely, pH, water temperature, specific conductance, and dissolved oxygen, were obtained from a stream gaging station at Hibgy of the Scioto River, a tributary of the Ohio River. The gaging station has a drainage area of 13,283 square kilometers and the record length from 1972 to 1988 was used for the simulation study.

Though minimum, average, and maximum values of the water quality studied were available, it is of interest to investigate the worst scenarios. Therefore, the following were selected: (1) for specific conductance, maximum data were used because the conductivity usually increases with decreasing water quantity (Fair et al., 1968); (2) maximum water temperature data were used since a high water temperature usually leads to more chemical reactions, to less dissolved oxygen, and to a deterioration of water quality (Krenkel and Novotny, 1980); (3) minimum dissolved oxygen data were used since most regulations and standards from the environmental agencies require limitations on the minimum dissolved

**Table 1.** **Statistics of historic and simulated maximum pH in standard units for the Scioto River at Higby, Ohio.**

| Statistic | | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | Historical | 7.98 | 8.06 | 7.99 | 8.01 | 8.09 | 8.09 | 8.14 | 8.23 | 8.03 | 7.94 | 7.99 | 8.26 |
| | Simulated | 8.22 | 8.22 | 8.08 | 8.04 | 8.11 | 8.13 | 8.16 | 8.25 | 8.04 | 7.93 | 7.99 | 8.26 |
| St. Devi-ation | Historical | 0.22 | 0.16 | 0.20 | 0.20 | 0.29 | 0.37 | 0.32 | 0.31 | 0.27 | 0.18 | 0.16 | 0.21 |
| | Simulated | 0.16 | 0.16 | 0.19 | 0.20 | 0.26 | 0.32 | 0.32 | 0.31 | 0.27 | 0.19 | 0.16 | 0.19 |
| Corr. Coeff-icient | Historical | 0.67 | 0.58 | 0.26 | 0.64 | 0.73 | 0.68 | 0.41 | 0.61 | 0.25 | 0.45 | 0.64 | 0.26 |
| | Simulated | 0.71 | 0.49 | 0.24 | 0.52 | 0.60 | 0.69 | 0.45 | 0.64 | 0.36 | 0.46 | 0.47 | 0.02 |

**Table 2.** **Statistics of historic and simulated maximum pH in standard units for the Scioto River at Higby, Ohio.**

| Statistic | | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | Historical | 7.84 | 7.93 | 7.83 | 7.83 | 7.88 | 7.89 | 7.90 | 7.95 | 7.81 | 7.79 | 8.86 | 8.14 |
| | Simulated | 8.14 | 8.14 | 7.96 | 7.90 | 7.93 | 7.94 | 7.94 | 7.97 | 7.82 | 7.79 | 7.86 | 8.15 |
| St. Devi-ation | Historical | 0.22 | 0.16 | 0.24 | 0.21 | 0.24 | 0.30 | 0.28 | 0.29 | 0.24 | 0.18 | 0.18 | 0.20 |
| | Simulated | 0.16 | 0.16 | 0.21 | 0.21 | 0.23 | 0.26 | 0.27 | 0.29 | 0.24 | 0.20 | 0.18 | 0.19 |
| Corr. Coeff-icient | Historical | 0.73 | 0.49 | 0.46 | 0.64 | 0.68 | 0.73 | 0.48 | 0.56 | 0.68 | 0.54 | 0.44 | 0.01 |
| | Simulated | 0.70 | 0.64 | 0.45 | 0.67 | 0.77 | 0.75 | 0.48 | 0.53 | 0.57 | 0.53 | 0.54 | 0.25 |

**Table 3.** **Statistics of historic and simulated maximum specific conductance in microsiemens per centimeter for the Scioto River at Higby, Ohio.**

| Statistic | | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | Historical | 680 | 646 | 601 | 589 | 642 | 661 | 652 | 700 | 715 | 738 | 701 | 670 |
| | Simulated | 610 | 609 | 586 | 587 | 641 | 666 | 656 | 704 | 719 | 736 | 700 | 671 |
| St. Devi- ation | Historical | 109 | 106 | 83 | 68 | 77 | 82 | 93 | 100 | 115 | 103 | 75 | 96 |
| | Simulated | 108 | 107 | 86 | 70 | 75 | 80 | 85 | 95 | 102 | 102 | 84 | 90 |
| Corr. Coeff- icient | Historical | 0.53 | 0.34 | 0.36 | 0.56 | 0.52 | 0.81 | 0.66 | 0.63 | 0.57 | 0.68 | 0.65 | 0.48 |
| | Simulated | 0.59 | 0.41 | 0.45 | 0.56 | 0.49 | 0.77 | 0.62 | 0.52 | 0.57 | 0.78 | 0.55 | 0.01 |

**Table 4.** **Statistics of historic and simulated maximum water temperature in °C for the Scioto River at Higby, Ohio.**

| Statistic | | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | Historical | 3.17 | 4.06 | 8.92 | 14.1 | 19.9 | 23.7 | 26.0 | 26.1 | 23.1 | 16.6 | 11.1 | 5.91 |
| | Simulated | 6.01 | 6.00 | 10.2 | 14.6 | 20.1 | 23.9 | 26.1 | 26.2 | 23.1 | 16.6 | 11.1 | 5.94 |
| St. Devi- ation | Historical | 1.33 | 1.67 | 2.38 | 1.82 | 1.92 | 1.44 | 1.35 | 1.21 | 1.09 | 1.26 | 1.29 | 1.34 |
| | Simulated | 1.70 | 1.70 | 2.09 | 1.84 | 1.90 | 1.60 | 1.48 | 1.35 | 1.05 | 1.21 | 1.22 | 1.34 |
| Corr. Coeff- icient | Historical | 0.68 | 0.69 | 0.35 | 0.41 | 0.55 | 0.68 | 0.65 | 0.16 | 0.48 | 0.32 | 0.33 | 0.34 |
| | Simulated | 0.72 | 0.55 | 0.39 | 0.43 | 0.66 | 0.75 | 0.73 | 0.11 | 0.41 | 0.24 | 0.23 | 0.03 |

**Table 5. Statistics of historic and simulated minimum dissolved oxygen in milligrams per liter for the Scioto River at Higby, Ohio.**

| Statistic | | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | Historical | 10.8 | 11.0 | 10.0 | 8.82 | 7.59 | 6.74 | 6.43 | 6.29 | 6.29 | 7.28 | 8.80 | 11.0 |
| | Simulated | 12.1 | 12.1 | 10.6 | 9.07 | 7.75 | 6.94 | 6.59 | 6.44 | 6.42 | 7.30 | 8.82 | 11.0 |
| St. Devi- ation | Historical | 1.29 | 1.27 | 1.31 | 0.90 | 1.07 | 1.45 | 1.58 | 1.71 | 1.32 | 1.44 | 1.37 | 1.20 |
| | Simulated | 1.28 | 1.28 | 1.31 | 0.97 | 1.04 | 1.23 | 1.37 | 1.49 | 1.37 | 1.41 | 1.33 | 1.22 |
| Corr. Coeff- icient | Historical | 0.81 | 0.54 | 0.41 | 0.57 | 0.75 | 0.82 | 0.84 | 0.82 | 0.57 | 0.77 | 0.44 | 0.06 |
| | Simulated | 0.83 | 0.52 | 0.55 | 0.56 | 0.63 | 0.75 | 0.79 | 0.84 | 0.56 | 0.77 | 0.42 | 0.02 |

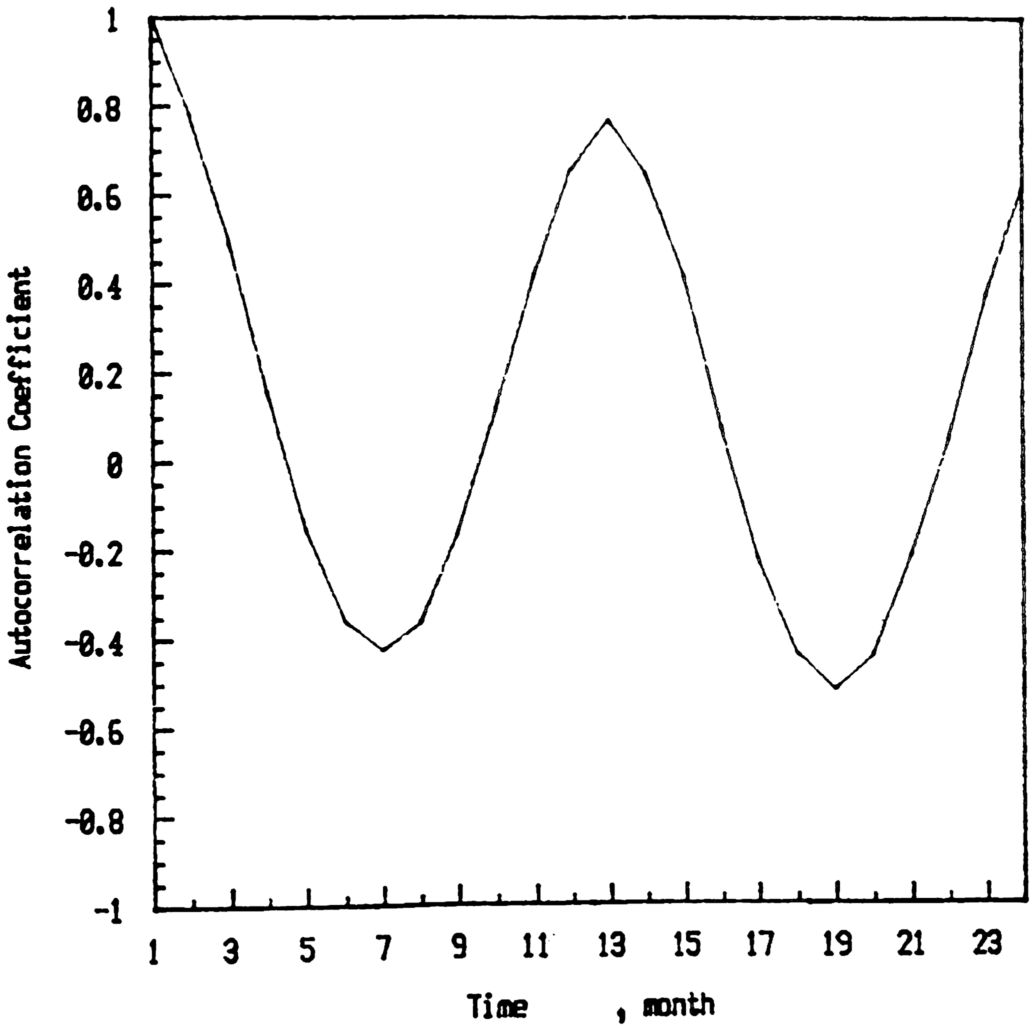


Figure 1. Autocorrelation plot of historic data for monthly dissolved oxygen at Higby of the Scioto River, Ohio

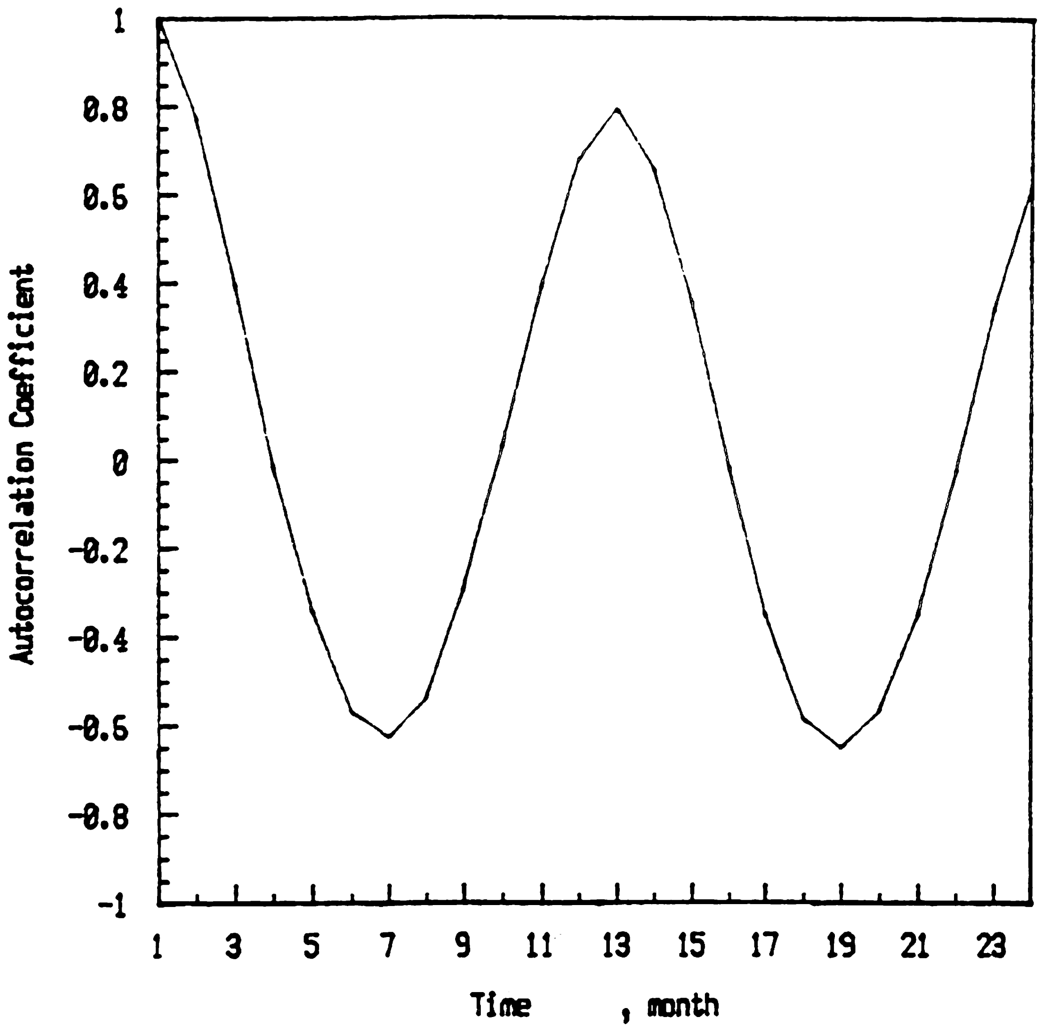


Figure 2. Autocorrelation plot of simulated data for monthly dissolved oxygen at Higby of the Scioto River, Ohio

oxygen (Tebbutt, 1983); (4) maximum and minimum pH values were used since both extreme situations would adversely affect the water quality (Hammer and Viessman, 1985).

Based on Equ. 7, the simulation model was used to generate synthetic monthly data of water quality, i.e., pH, dissolved oxygen, water temperature, and specific conductance. The simulation model is a first-order Markovian process, where a simulated value depends on its proceeding value by the first-order correlation. In order to preserve the monthly means of historic data, they were initially removed in the process as shown in Equ. 7. Furthermore, the monthly standard deviations and their serial correlations were separately used for monthly simulations to preserve the standard deviations and the serial correlations.

Monthly means, standard deviations, serial correlation coefficients of the historic and simulated data were calculated for the purpose of comparison. Tables 1 and 2 give the monthly means, standard deviations, and correlation coefficients for the historic and simulated data for maximum and minimum pH, respectively. The results show that the three statistics of the simulated data are agreeable with those of the historic data. Tables 3 to 5 provide the statistics of the historic and simulated monthly data for maximum specific conductance, maximum water temperature, and minimum dissolved oxygen, respectively. Except some variations for the months of January and February, the simulated data reasonably preserve the statistical properties of the historic data.

Furthermore the autocorrelation coefficients of the historic and simulated data for each water quality variable are computed and plotted against time lag, i.e., month, for comparison. Figures 1 and 2 are the examples of the autocorrelation plots for the historic and simulated dissolved oxygen data, respectively. They show that the simulation reasonably preserves the autocorrelation function of the historic data including the seasonal pattern.

## 4 CONCLUSIONS

To solve the problem of the short length for the analysis of water quality that is usually encountered in a long-term study, a simulation method was introduced to generate the synthetic data of water quality. Based on the comparison results, it is found that the simulation method can well preserve the mean, standard deviation, and correlation coefficient of the historic water quality data.

## ACKNOWLEDGEMENTS

This paper is based on the research supported in part by the U.S. Army Corps of Engineers under agreement ED-W-9129-035, and in part by Ohio University.

## REFERENCES

Chang, T.J., *Effects of Drought on Streamflow Characteristics*, Journal of Irrigation and Drainage Engineering, Vol. 116, No. 3, 332-341, 1989.

Fair, G.M., Geyer, J.C., and Okun, D.A., *Water and Wastewater Engineering*, John Wiley & Sons, Inc., New York, 1968.

Fiering, M.B., and Jackson, B.B, *Synthetic Streamflows*, 3rd ed., American Geophysical Union, Washington, D.C., 1981.

Hammer, M.J., and Viessman, W. Jr., *Water Supply and Pollution Control*, 4th ed., Harper & Row Publishers, NY, 1985.

Krenkel, P.A., and Novotny, V., *Water Quality Management*, pp. 46-59, Academic Press, Inc. 1980.

Tebbutt, T.H.Y., *Principles of Water Quality Control*, 3rd ed., Pergamon Press, NY, 1983.

## AUTHOR BIOGRAPHIES

**TIAO J. CHANG** is an Associate Professor in the Department of Civil Engineering at Ohio University in Athens, Ohio. He received his Ph.D. degree from the School of Civil Engineering at Purdue University in 1982. His research interest are focused on the statistical and computer applications in water resources engineering.

**RABAH Y. ODEH** was a former Graduate Associate in the Department of Civil Engineering at Ohio University in Athens, Ohio. He received his M.S. degree from the Ohio University in 1992.