

Recast of the Outputs of a Deterministic Model to Get a Better Estimate of Water Quality for Decision Makings

Ping Wang¹, Lewis Linker², Carl Cerco³, Gary Shenk², and Richard Batiuk²

¹University of Maryland Center for Environmental Science, Chesapeake Bay Program,
410 Severn Avenue, Annapolis, MD 21403, USA
pwang@chesapeakebay.net

²US Environmental Protection Agency/CBPO, 410 Severn Ave., Annapolis, MD 21403, USA
{linker.lewis, shenk.gary, batiuk.richard}@epa.gov

³US Army Corp of Engineers, 3909 Halls Ferry Road, Vicksburg, MS 39180, USA
cerco@erdc.usace.army.mil

Abstract. The outputs of a deterministic water quality model are retreated to recast the model simulation. A multi-variance regression method is used for daily model outputs versus observed data to assess the systematic errors in model simulation of individual model cell. The model outputs are re-adjusted to better represent the actual values, and to yield a better model calibration. Such a recast is important to model application for water quality analysis, especially for TMDL (Total Maximum Daily Load) which requires accurate simulation to exam criteria attainment. This paper addresses the recast method, its prerequisites, and how the results are extended for the scenarios with various load reductions.

1 Introduction

Deterministic simulation models have been used to generate spatially distributed time-series data and applied in many fields, including the environmental field [1,2]. Imperfect calibration, i.e., deviations of the model prediction from what are actually observed, may be due to differences in the observed and simulated values and/or having a phase- (or time-) shift. Here, a phase-shift means that the model simulates a similar value as the observed, but it appears one or several days earlier or later. The deviations of model estimates affect model applications, particularly when model application is aimed at achieving water quality conditions above or below a particular water quality criterion. This is the case in the current Chesapeake Bay application when the model is applied against criteria values of dissolved oxygen (DO), chlorophyll and water clarity.

The Chesapeake Bay is one of the largest and most productive estuaries in the world (Fig. 1). Monitoring efforts, computer models, and researches indicated that the degradation of the Bay's water quality is primarily due to excessive nutrients and sediment inputs from its 166,000 km² watershed [3]. The US government and the states of the Chesapeake region committed to protect the Bay, to improve its water quality, and to remove the Bay from the list of impaired waters under the Clean Water Act by year 2010 [4]. The deterministic Chesapeake Bay Estuarine Model (CBEM)

has been used to simulate the responses of water quality to nutrient and sediment loads. In overall, its calibration to the observed data is good and acceptable according to the Chesapeake Bay Program Modeling Subcommittee (CBPMS) [5], though with certain under- or over-estimates and phase-shift. Accurate simulation of key water quality indices in certain value ranges is important to policy decisions. For example, the draft water quality criterion for DO in the designate-use-area of “Open Water” is daily minimum of 3.5 mg/L. If the observed DO on some days are close to the criteria, an under-estimate or over-estimate of the model in DO may result in non-attainment (i.e., below the criteria) or attainment (i.e., at or above the criteria) of the criteria, respectively.

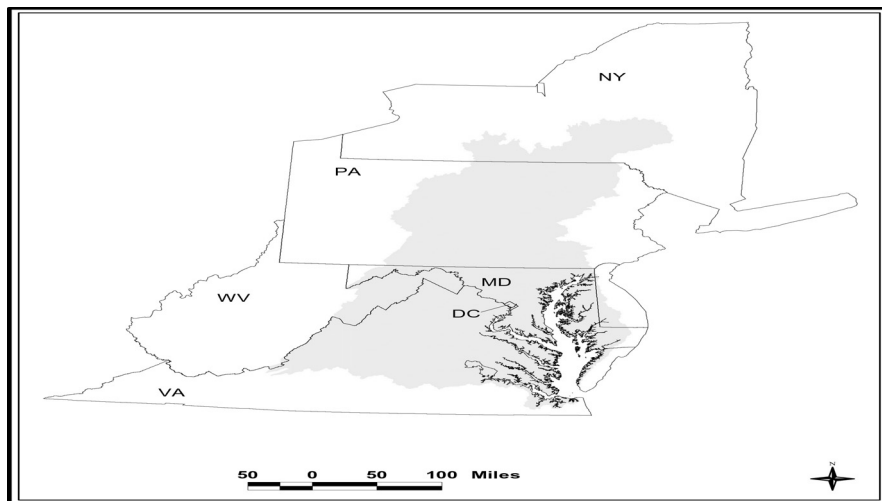


Fig. 1. Location of Chesapeake Bay and its watershed in the US mid-Atlantic states. Note: gray – the Chesapeake watershed; white within the gray – Chesapeake Bay

More problems may occur when running model scenarios, since such deviations will propagate. For instance, a DO in an “Open Water” is 2.8 mg/L, 0.7 mg/L lower than the criteria. A 30% nutrient reduction, which improves DO by 1.0 mg/L to be 3.8 mg/L, would comply with the criteria. If the model calibration is 0.5 mg/L lower, at 2.3 mg/L, the scenario of 30% nutrient reduction from the reference condition would probably yield a DO under 3.5 mg/L, resulting in a violation of the criteria. On the other hand, if the model calibration overestimates DO, it may give an impression that the water quality criteria have already been complied, and the model would not be able to provide meaningful information for management either. In short, although the CBEM simulates well the key water quality indices, the accuracy of the model around the critical values of these water quality indices may still not meet the requirements for decision making. It is desirable to have a better estimate in the interested water quality indices for the designate-use-area of the Bay. There are several data assimilation methods to get better model predictions [6,7], however, they cannot fulfil our requirements. This paper is to develop a method to retrieve the outputs of the deterministic CBEM to be better representative of the actual conditions.

2 Evaluation of CBEM Calibration

CBEM is a coupled three-dimensional finite-difference hydrodynamic model (based on CH3D [6]) and finite-volume water quality model (based on CE-QUAL-ICM [2]). The current model has 12960 cells with 2961 surface cells for the 11,000 km² surface area of the estuary. It simulates nutrient cycling and sedimentation, as well as interactions among key living resources and water quality parameters. Continuous daily values are output from model calibration and scenario runs. Discrete observations sampled on a specific cruise are used to represent the water conditions on that day.

There are many methods in model calibration and verification [9][10][11]. Visual analysis of scatter plots, frequency plots, and time-series plots are used for the CBEM calibration. Statistics is used to evaluate the confidence of the model by CBPMS [5].

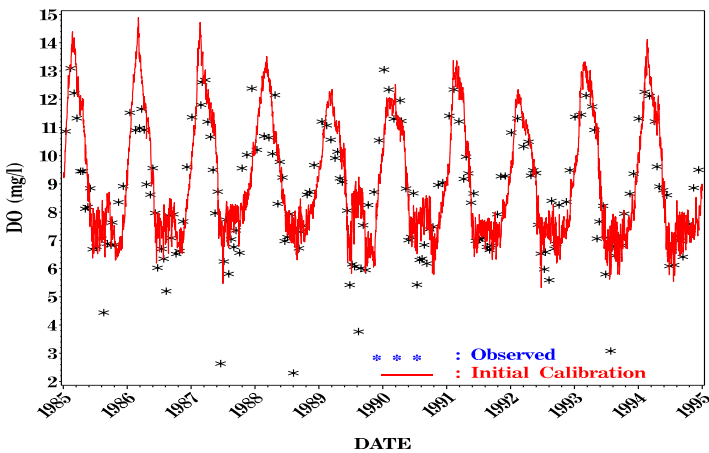


Fig. 2. Time series plot of initial calibration (-) and observed (*) DO in Cell 7563

The time-series plot of DO (Fig. 2), indicates that the model (the continuous line) is well calibrated for Cell 7563 (in “Open Water” of Segment 5), comparing to the observed (with stars). The days with low observed DO usually have low simulated DO, and the days with high observed DO usually have high simulated DO. The deviation of the model from the actual can be compared according to their values on the same date. The problem of phase-shift may also contribute the deviation of the model.

Figure 3 is a scatter plot of observed DO versus simulated DO for Cell 7563. The data are “paired”, which means that only the simulations on the day having observations are used in the plot. A perfect calibration will yield the points to orient a line with a slope of one and intercept of zero, called the ideal line. Since the plot is from paired observation and simulation, besides the deviation in absolute values, a phase-shift is another cause for the plot to deviate from the ideal line. Based on the paired values, we can apply statistics to assess the confidence of the model. The CBPMS set the following rules for a high confidence of DO simulation based on the statistics between the paired observed and modeled data: a) correlation coefficient > 0.5 (desir-

able), b) mean differences not greater than 1 mg/l (or roughly 10%), c) minimum concentration do not differ by more than 2 mg/l, and d) standard deviations do not differ by more than 0.5. Here the regression line is not used as a model to predict DO, but as the correlation of the paired data to evaluate the model. Since a phase shift and the complexity of the system, R^2 of 0.5 is considered to be an indication of good simulation. The statistics concludes the calibration of DO on Cell 7563 to be good.

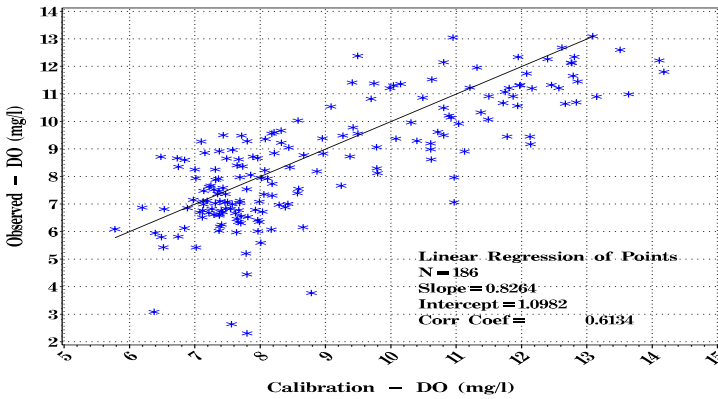


Fig. 3. Scatter plots for initial calibration and observed DO in Cell 7563

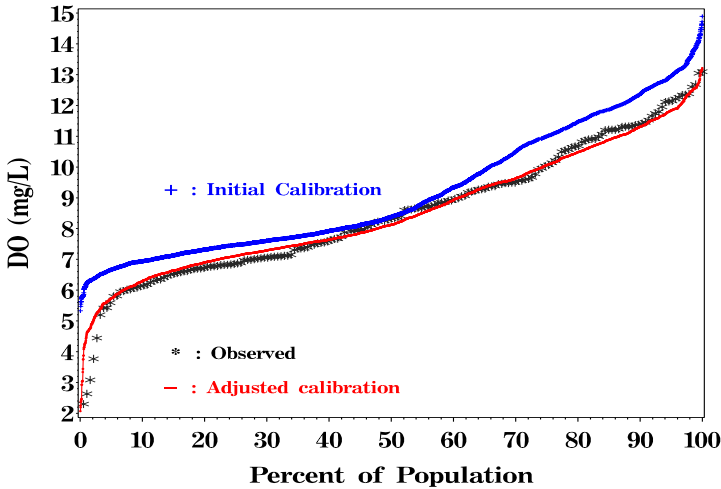


Fig. 4. Cumulative frequency of initial calibration (+), recast calibration (-), and observed (*) DO in Cell 7563

Figure 4 is a cumulative frequency plot of DO for the observed and initial calibration. A professional judgement based on the frequency plots by the CBPMS is also a key component to determine the confidence level of the model [5]. Note: the data for the cumulative frequency plot, both the modeled and the observed, are sorted, respec-

tively, from low to high values. If the two curves are close to each other, then the calibration is regarded as good. It means that the model simulates similar high and low values as the observed, without considering whether they appear on the same date. A separation of the two curves reveals deviations of the model from the observations. The deviation has three basic patterns. The curve of the simulations can be either 1) above (overestimate), or 2) below (underestimate), or 3) across the curve of the observations. One example of Pattern 3 deviation is that the model overestimates in the high value ranges, but underestimates in the low value ranges. If the model is used to assess the average DO (supposing with allowable ± 2 mg/l of deviation), then the calibration may be acceptable. However, for the policy-making as addressed in the introduction, we need higher accurate estimates in low DO (0–5 mg/l) values.

3 Method for Recast of the CBEM Calibration

The recast of the CBEM calibration is to use a stochastic method to adjust the output of the key water quality indices in the CBEM calibration run to be close to the observed. The adjustment is based on the relationship between the observed (e.g., y) and simulated (e.g., x) for individual model cell, which can be expressed as

$$y=f(x), \quad (1)$$

and established by regression. Once such an equation is established, for a modeled value in the calibration run (x), we get an adjusted model value, x' , ($x' + \text{residual} = f(x) = y$) that would be consistent with the observed within the desired accuracy. Theoretically, such a relationship can be established through a first-order linear regression of the paired observation and simulation that was used for Figure 3. Since the R^2 is too low, such a regressed equation is not suitable to recast the model calibration. The failure of such an approach is partly due to a phase-shift between the modeled and the observed.

Let's consider the following facts of the CBEM. The model overall simulates DO well. For example, it predicts a summer DO problem in areas where DO is observed to be problematic. Low DO values are simulated in some deep water in the summer, as is observed. Statistically, corresponding to an observed high (or low) value there is likely to have a high (or low) estimated model value near the day. The time-series and frequency plots support such an assessment. The problem is that the model estimates are systematically high or low in certain value ranges. What we want is to adjust such systematic errors. There are sufficient observed data for regression, about 1 or 2 each month for the entire 10 years of model simulation, which cover the periods of high and low DO in the Chesapeake Bay. Under such conditions, it is acceptable to order the data for the purpose of comparing the distribution of the two sets of data as we did for the cumulative frequency plots. The regression is basically on the two frequency curves, to make them to close to each other, with a focus on the low DO ranges. The days having low (or high) values before the adjustment would still have low (or high) values after the adjustment, and the recast values should be closer to the observed. The regression can be performed on the ordered paired observations and modeled data that were used for a paired frequency plot or scatter plot. However, we suggest the

following additional work to improve the recast, since there usually exists a phase shift in the model. Corresponding to an observation, its most likely corresponding modeled data may not be on the same date. Therefore, it is better not to only use the observed and simulated which are paired on same dates, but to use all simulated daily values and the available observations. The modeled has about 20 times more values than the observed. We firstly order the modeled daily values, then split the population equally into groups according to the number of the observations. The median value of each group is picked out to form a set of sample, which has the same number as the set of observed data. We apply regression on these two sets of data. In this method, all the highest and lowest simulated values are considered in the regression, which provides an additional way to overcome the phase-shift problem.

4 Results of the Recast of CBEM Calibration

The time-series plot (Fig. 5) shows that the adjusted values are very close to the observed, especially in low DO levels. For a criterion of 3.5 mg/l, the initial calibration (Fig. 2) predicts no violation of the DO criteria, whereas, the recast (Fig. 5) shows certain violations, consistent with the observation. For the management of water quality (such as DO, chlorophyll, and water clarity) in the Chesapeake Bay, a few days of time shift may not be critical, whereas, absolute value is more important in evaluation of the attainment of criteria. Figure 4 also plots the cumulative frequency of the recast DO (with joined dots). It is very close to the observation curve. Five percent of tolerance is set for the DO criteria. Therefore, a 5% error in DO estimate is acceptable. The model recast meets the required accuracy of the model in DO simulation. The recast method was also applied to all other cells that have monitoring data of DO, chlorophyll, and water clarity. No matter whether the initial calibration is predominated with underestimate or overestimate, satisfactory recasts were resulted. This indicates that the recast method is acceptable.

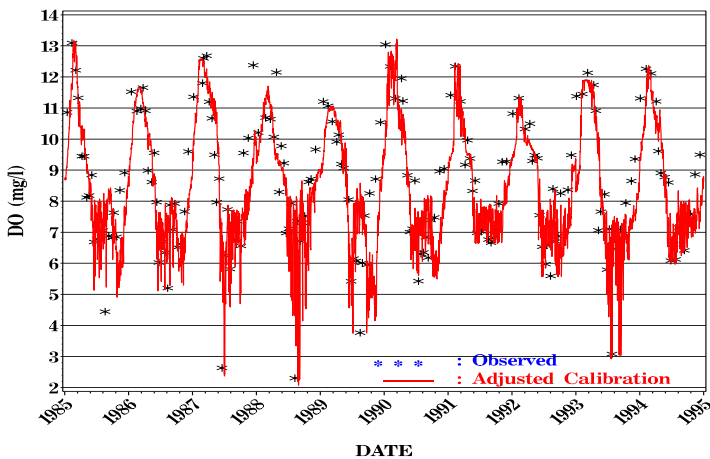


Fig 5. Time series plot of adjusted calibration (-) and observed (*) DO for Cell 7563

5 Recast for the Cells without Observed Data

The number of surface cells and monitoring stations varies among model segments, ranging from 4 to 100 and 1 to 10, respectively. The recast for the cells without observed data utilizes the recast results from the cells having observed data within the same model segment: use the relationships (i.e. the regressed coefficients) from three nearest calibration cells, and weigh them according to the inverse square distances to the cells (centroid).

We did such a check for a cell having a monitoring station. The recast by using its monitoring data directly is very close to that by using the weighed regression results from three nearby monitoring cells. We also did another type of check. For the cells having monitoring station, we aggregate all monitored and modeled DO in a certain depth range in a model segment to construct a regression, then applied the regression results uniformly to recast DO in each monitoring cell at the depth level. The result for each cell is close to the recast based on the individual cell monitoring info, although slightly worse. This indicates that the systematic errors in a segment are similar. We believe that the recast method for non-monitoring cells is acceptable.

6 Application to the Recast of Model Scenarios

A model scenario of nutrient and sediment reduction usually yields lower chlorophyll and higher dissolved oxygen and water clarity than the reference conditions (such as the calibration run). Since the model calibration deviates from the actual, the errors will propagate in model scenario runs. Such modeled information may not be used to determine the level of load reduction to improve water quality.

For DO of the initial scenario (z) and initial calibration (x), their relationship, $z = g(x)$, can be established through regression. The CBEM has good proportional response of water quality indices to load reduction. A nearly linear relationship between them was observed for most cells. For the ideal or recast scenario (z') and recast calibration (x') the above relationship also holds approximately, i.e., $z' = g(x')$. Such an approximation requires the CBEM is fairly calibrated, and the errors in the estimate of the key water quality indices are mainly systematic.

An alternative, simpler way to recast daily DO of a scenario run is by referring to Equation 1, $y = f(x)$, i.e., the relationship between the observed, y , and the calibration (x). Since $x'=y$ for a perfect recast, therefore, $x'=f(x)$. Substituting scenario outputs, z , for x in Equation 1, yield $z' = f(z)$. Z' is the recast scenario. However, this method may not be as good as the previous one. The exploration of the errors related with these recast methods and the development of more sophisticated methods for scenario recast are beyond the scope of this paper.

7 Discussion

7.1 Assumption and Validity of the Method

The requisite of the proposed recast method is that the deterministic model has good simulations of the physical, chemical, and/or biological processes for the modeled ecosystem, and is reasonably calibrated both temporally and spatially. The CBEM is one of the most sophisticated surface water quality models in the world, and meets these requirements. It simulates well the target water quality indices of DO, chlorophyll, and water clarity. High chlorophyll was simulated in spring and summer in high nutrient loading areas, consistent with the historical data of high chlorophyll. Low DO, especially in bottom cells in the summer, was simulated in the historically DO problematic areas. Similarly is for water clarity. The responses of DO, chlorophyll, and water clarity to the scenarios of nutrient reduction are well simulated: i.e., a lower loads yields lower chlorophyll level, more water clarity, and higher DO. This can also be proved by the surface analysis [12], which is based on regression of the responses of water quality to various loading scenarios. Therefore, we have high confidence on the model. The modeled high (or low) values of a constituent are likely to have high (or low) observed values around the simulate dates. What is needed to adjust for the model output is to reduce the systematical over- or under-estimates. The recast is to bring the simulated value to be close to the observed level, as seen from the frequency plot (Figure 4). We applied the same method for all cells with monitoring data, and similar recasts were yielded. This method does not intend to improve the correlation for the paired data as in the scatter plot, since this method does not regress to match the pair-values directly. Nevertheless, the recast usually improves their correlation, leading the points to be close to the ideal line.

7.2 Problem with Ordered Data

For any two sets of unrelated data, if they are ordered by values, they are likely to show a certain correlation (e.g., near linear). Therefore, it is usually no meaningful for a regression on two sets of ordered values of time-series data. However, here we are trying to transpose the distributions of the modeled versus observed for the frequency plots, not the individual points, and to obtain the relationships between the distributions of these two sets of data. The CBEM is well calibrated. Statistically, the modeled higher values correspond to the observed higher values, and the modeled lower values correspond to the observed lower values. After the recast, the corresponding values get closer to each other, as Fig. 5 versus Fig. 2. The fact of good recast for all monitoring cells supports this recast method. People may have noticed that the data preparation for the recast method described in Sect. 3 is similar to that for quantile-quantile (Q-Q) plot. We can expect that such recast results would look good in Q-Q plot, as shown in Figure 6, which has great improvements over the initial calibration. Q-Q plots have been recommended by many researchers for model evaluation and exploratory data analysis [13], therefore, we believe the recast method is plausible.

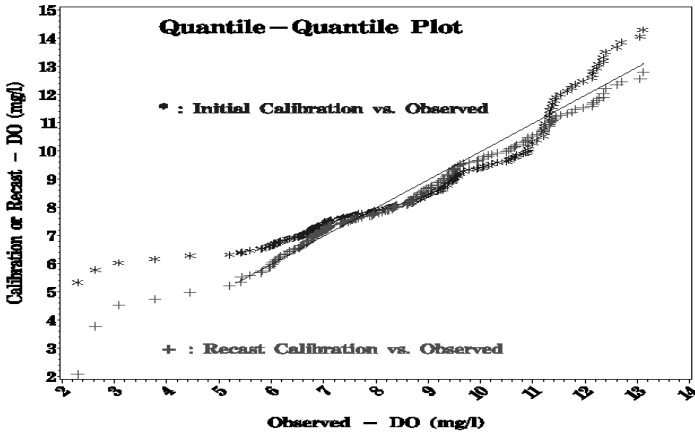


Fig. 6. Q-Q plot of initial calibration and recast calibration DO vs. observed DO

7.3 Regression on Paired or Non-paired Data

Our work reveals that a recast using paired observation and model data for regression provides a better estimate than the initial calibration. The regression using all daily model data (i.e., non-paired data) can remedy some problems associated with the phase-shift by a model. The latter has an even better recast than the method using paired values. The results of recast presented in this paper are by the latter approach.

7.4 Subdivision of Time Period for Regression

DO, chlorophyll, and water clarity during the period of May through September have contrast levels of water quality from those in the other period. In many cases, splitting data into these two periods improves regression. However, sometimes, for example, there are similar low DO values in the summers of two different years, but the model simulates somehow different levels of water quality in the two summers. In such a case, it is better to split data yearly. In fact, the model shows a strong correspondence of low-to-low or high-to-high values between simulated and observed within a year. Therefore, splitting data yearly for regression generally yields good results. Finer splitting data is not recommended, which may reduce sample population in the regression. It is useful to investigate the optimal subdivision of data. However, this is beyond the scope of this paper.

8 Conclusion

An accurate calibration of CBEM is required for the key water quality indices in the designate use area of the Chesapeake Bay. The method of this paper recasts the model calibration of the CBEM to yield more accurate daily estimates than the initial cali-

bration. Care should be taken that the requirements of this method are satisfied -- the deterministic model should first be well calibrated with reasonable responses to seasonal change and load changes.

References

1. Thomann, R.V., Mueller, J.A.: Principles of surface water quality modeling and control, Harper Collins Pub., New York, (1987) 644pp
2. Cerco, C.F., Cole, T.M.: Three-dimensional eutrophication model of Chesapeake Bay. *J. Envir. Engrg.*, (1993) **119**(6): 1006–1025
3. Officer, C.B., Biggs, R.B., Taft, J.L., Cronin, L.B., Tyler, M.A., Boynton, W.R.: Chesapeake Bay anoxia: origin, development and significance. *Science* (1984) **233**: 22–27
4. Chesapeake Executive Council: *Chesapeake 2000*. Annapolis, MD, USA (2002)
5. CBPMS: A comparison of Chesapeake Bay Estuary Model calibration with 1985–1994 observed data and adjustment of model estimates of dissolved oxygen, clarity, and chlorophyll to better represent observed data. A report of the Modeling Subcommittee, Chesapeake Bay Program Office, Annapolis, MD, USA, October (2002) pp300
6. Malanotte-Rizzoli, P.: Modern approaches to data assimilation in ocean modeling. Elsevier, (1996) pp455
7. Emery, W.J., Thomson, R.E.: Data analysis methods in physical oceanography, Pergamon, (1998) pp634
8. Johnson, B., Kim, K., Heath, R., Hsieh, B., Butler, L.: Validation of a three-dimensional hydrodynamic model of Chesapeake Bay. *J. Hydr. Engrg.*, ASCE (1993), **199**(1): 2–20
9. Thomann, R.V.: Verification of water quality models. *J. Envir. Eng.* (1982) 108(5) 923–940
10. Rechhow, K.H., Clements, T.J., Dodd, R.C.: Statistical evaluation of mechanistic water-quality models. *J. Envir. Eng.* (1992) 116(2) 250–268
11. Wang, P., Storrick, J., Linker, L.: Confirmation of the Chesapeake Bay Watershed Model. Proceedings of the 16th International conference of the Coastal Society, Williamsburg, VA, USA (1998) 381–390
12. Wang, P., Linker, L.C., Batiuk, R.A.: Surface analysis of water quality response to load. in: Spaulding, M.L. (ed.) Proc. 7th Intern. Conf. in Estuarine and Coastal Modeling, Nov. 2001, St. Petersburg, FL, USA (2002) 566–584
13. Nosteller, F., Tukey, J.: Data analysis and regression, Addison-Wesley (1977)