Engineering Water Quality Models and TMDLs

Steven C. Chapra, M.ASCE

Professor and Berger Chair, Civil and Environmental Engineering Dept., Tufts Univ., Medford, MA 02155. E-mail: steven.chapra@tufts.edu

Introduction

Over the past 75 years, engineers have developed water quality models to simulate a wide variety of pollutants in a broad range of receiving waters. In recent years, these receiving water models are being coupled with watersheds, groundwater, bottom sediments, and airsheds to provide comprehensive frameworks predicting the impact of human activities on water quality. As Thomann (1998) terms it, "Golden Age" of water quality modeling is upon us.

At face value, these developments should bode well for water quality assessments such as TMDLs (total maximum daily loadings). However, if not conceived and implemented properly, models could also detract from such efforts. The present paper describes some of these pitfalls as well as related opportunities.

The discussion will be organized around the model development process. As in Fig. 1, water quality modeling is embedded within the larger context of the TMDL decision process. The primary function of modeling is to provide a decision support model that can be used in TMDL prescriptions; in particular, the model provides a means to predict water quality as a function of loads and system modifications.

The process [Fig. 1(b)] starts with model selection and development. The latter relates to situations where existing models are inadequate. After selecting or developing the model, existing data are used to construct a preliminary model application. This exercise should include thorough data mining to ensure that all possible historical data are considered. For cases where adequate historical data do not exist, additional data are collected. The model is then calibrated by adjusting uncertain parameters so that the model performs adequately. This is followed by a series of confirmation tests. These typically involve applying the calibrated model to cases that differ significantly from those with which the model was calibrated. This is followed by an analysis phase to assess model sensitivity and uncertainty. Finally, the model can be integrated into a decision support system (DSS) to facilitate the model's use in decision making.

It should be stressed that although Fig. 1(b) defines a sequence of events, the use of two-way feedback arrows on the left denotes that the process is adaptive. For example, after calibration, it might be necessary to collect additional data for confirmation. Similarly, an uncertainty analysis might be performed on the preliminary model in order to help focus the data collection and

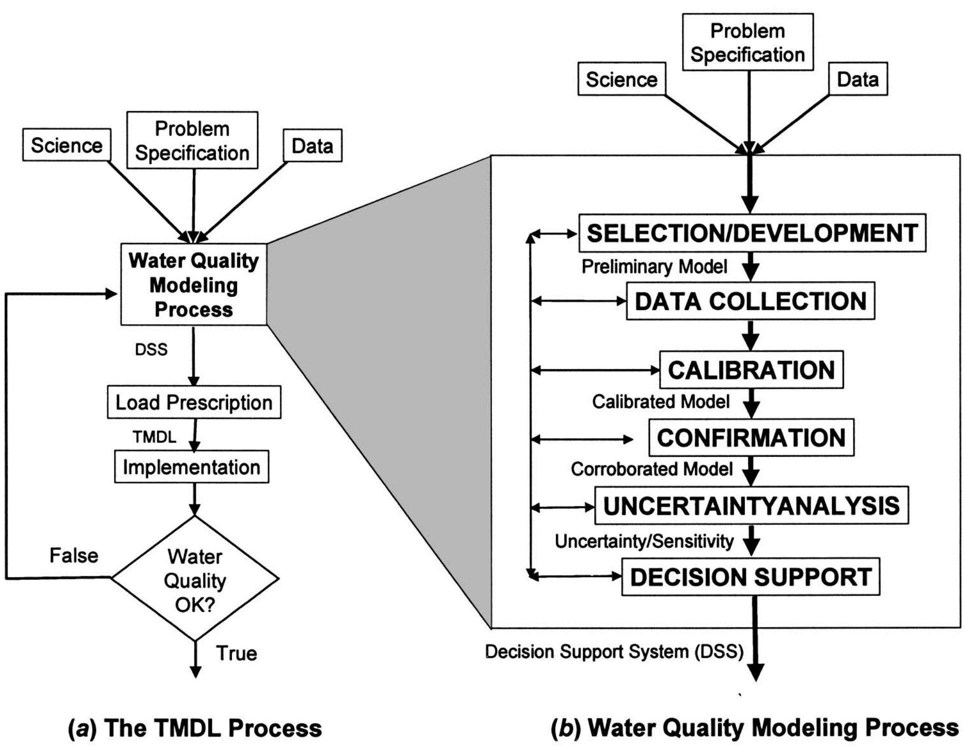

Fig. 1. Water-quality-modeling process (b) within context of TMDL process (a)

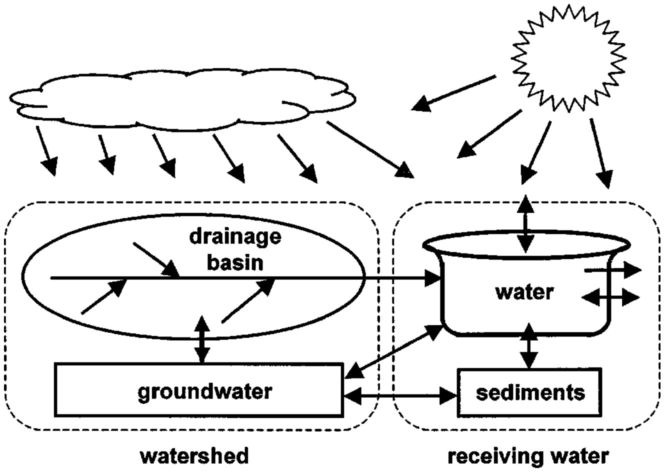

Fig. 2. Watershed/receiving water system

calibration processes. Finally, new models might be adopted, necessitating that the entire process be repeated.

This paper addresses six particular areas related to the process: model complexity, data and monitoring, reliability, uncertainty, decision support, and future investment. The discussion will be limited in certain ways. The focus will be on the watershed/receiving water framework depicted in Fig. 2. Note that the general conclusions are equally relevant when the framework is expanded to include other systems such as airsheds. The discussion will be further limited to the eutrophication/dissolved oxygen/heat problem. Again, many of the conclusions should be generally relevant to other problems such as toxic contaminants and pathogens. Most of the discussion is focused on process or mechanistic models based on momentum, heat, and mass balances. Some issues related to statistically based empirical models will be addressed in the section on uncertainty.

Model Complexity

Progress in science and computing along with changing environmental problems have allowed modelers to develop increasingly complex and comprehensive modeling frameworks (see reviews in Chapra 1997 and Thomann 1998). Unfortunately, this often leads to the common misconception that complex models are necessarily superior to simpler approaches. In fact, as illustrated in Fig. 3, the choice of a water quality model involves trade-offs among model complexity, required reliability, cost, and time. Complexity refers to the model equations and mathematical structure. Cost relates primarily to data collection and parameter estimation.

The straight line in Fig. 3(a) represents the underlying assumption that, if the budget is unlimited, a more complex model will be more reliable. In essence, as we add complexity to the model (that is, more equations with more parameters), we assume that sufficient funds are available to perform the necessary field and laboratory studies to adequately specify the additional parameters. In fact, this assumption itself may not be true, because there may be limits to our ability to mathematically characterize the complexity of nature (a kind of ecological uncertainty principle). However, though we may never be able to perfectly characterize a natural water system, we generally function under the notion that more and better information leads to more reliable models.

In reality, because we are invariably constrained by funding, we must make do with limited data. In such cases there are two extremes: (1) A very simple model will be so unrealistic that it will not yield reliable predictions, or (2) a very complex model will be so overparametrized that it outpaces available information and becomes equally unreliable because of parameter uncertainty. As in Fig. 3, there is an optimal model that is consistent with the available level of information.

At this point two other considerations must be imposed: the reliability and the complexity needed to solve the problem [Fig. 3(b)]. Clearly the model could be optimal, given the available data, yet not be sufficiently reliable for addressing the decision. For example, whereas a simple model might be adequate for assessing aesthetic impacts, a more complex model (and more supporting data) might be needed to assess problems dealing with public health.

Further, the model might be consistent with the available data, but not sufficiently complex to address management questions. Regulatory end points often drive this question. For example, a model that simulates a single phytoplankton group as measured by chlorophyll *a* would not be adequate if nuisance algae such as cyanobacteria are the end point. Similarly, if nutrient export is to be managed on a farm-by-farm basis, a highly distributed drainage basin model might be required rather than a simpler lumped approach. In such cases, additional complexity (along with more data and information) would be necessary.

Finally, because they often involve legally mandated deadlines, TMDLs are extremely sensitive to the issue of time. This has major implications for both establishing model credibility and using the model to make effective decisions. As discussed below,

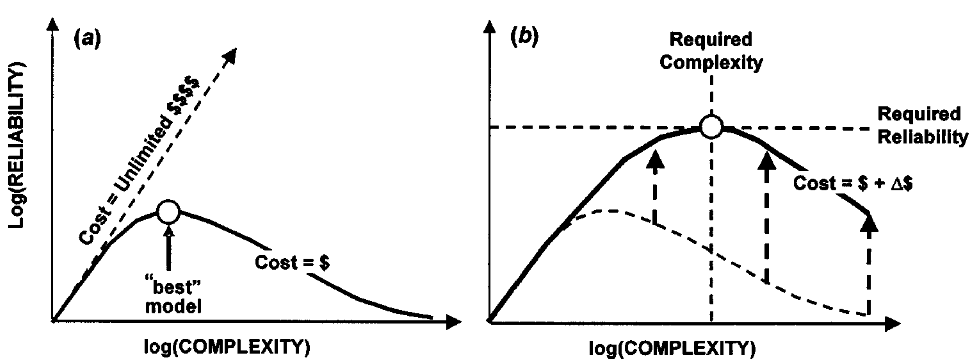

Fig. 3. Model reliability versus complexity (redrawn from Chapra 1997): (a) Modeling isolated from decisions; (b) modeling as influenced by decision context

many systems will have insufficient historical data to yield credible results. Significant time will then be required for additional data collection, and process studies required for model calibration and confirmation. Such a case recently occurred in the development of the Delaware Estuary PCB TMDL (Thomas J. Fikslin, Delaware River Basin Commission, personal communication, 2002). A major sampling effort was scheduled for the spring and summer of 2002 in order to characterize and model both high-flow and low-flow conditions. Because of drought conditions, a sufficiently high flow did not occur and the sampling effort has now been extended through spring 2003. Although a six-month extension might seem minimal, it looms large because the preliminary TMDL is to be developed and reviewed by December 2003.

Adding unnecessary model complexity also increases computer simulation time. This could have a deleterious effect on the ability to estimate the uncertainty of predictions using techniques such as Monte Carlo analyses. Further, less complex and faster models will expedite time-intensive tasks such as calibration, confirmation, sensitivity analysis, scenario analysis, optimization, and real-time control.

Because of all the above factors, it must be recognized that the problem needs and the available resources must drive the choice of model, rather than the pursuit of model complexity for its own sake. An important consideration in this regard is that all water bodies should not be modeled in a single fashion. Just as you do not need a high-performance racecar to go out and buy groceries, some systems will require coarser approaches whereas others will demand more complex methodologies.

Finally, modeling is a process, not an end. Given the present state of data, an adaptive approach to modeling would start with simpler models at the initial phases and then progress to more complex frameworks as additional data are collected and as more focused remedial measures are assessed.

Models and Data

Models are only as good as the information upon which they are based. For the purposes of the present discussion, this information can be divided into science and data.

Science represents numbers and mathematical constructs that reflect scientific understanding. These consist of default parameters (e.g., Bowie et al. 1985; Schnoor et al. 1987; EPA 1990), model constructs (the science embedded in the model equations), and modeled parameters. The latter are values that are calculated with formulations based on scientific studies (for example, evapotranspiration and reaeration formulas).

Data are numerical values that are site-specific and obtained by direct measurements. These consist of spatial data (topology, topography, soil types, land use, areas, morphometry), forcing functions (meteorology, point loads), state data (temperature, flow, concentration, water transparency), and rate data (direct measurement of model parameters such as settling velocity, SOD, etc.).

Although both science and data bear on model credibility, the present discussion focuses on data. In particular, the forcing function and state data that figure prominently in model calibration and confirmation are emphasized. This data can be broken into three categories: monitoring data, old model data, and new model data.

Monitoring refers to data that are measured on a routine basis for reasons other than modeling (e.g., compliance or detection).

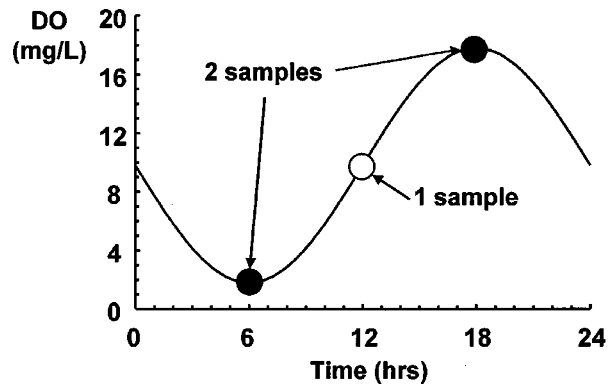

Fig. 4. An example of how model resolution influences sampling strategies. If diurnal variations of state variables are being simulated, adequate temporal sampling must be conducted to capture such variability. This may cause data collectors to install continuous sampling devices or to sample at several times of day to capture both the mean and the range.

Although such data can be of great value to TMDL modeling, there are two problems that can detract from their utility. First, because they are not aimed at supporting models, the wrong variables might be measured. For example, most current monitoring plans originated from the need to assess the performance of wastewater treatment plants. Hence, treatment-oriented parameters such as carbonaceous BOD (and in some cases, 5-day BOD) are frequently used to characterize organic carbon. As described by Chapra (1999), current models (e.g., Cerco and Cole 1993; Connolly and Coffin 1995) are carbon based and require direct measurement of organic carbon species (e.g., labile dissolved carbon, refractory dissolved carbon, and particulate organic carbon). In such cases, BOD measurements are inadequate.

Second, the monitoring data may not be collected properly in space and time. For example, a traditional monitoring program might situate sampling locations below point-source inflows. Although such information is valuable, models evaluating nonpoint sources must be distributed more evenly to resolve more gradual spatial variations in water quality. The same holds for temporal monitoring. For example, dissolved oxygen might be traditionally sampled at a single time during daylight hours. For systems with strong diurnal variations, critical oxygen levels usually occur at dawn. If a model simulated diurnal changes, a more effective sampling strategy might be to (1) install a continuous monitoring device or (2) sample at dawn and late in the afternoon (see Fig. 4).

The use of monitoring data is also complicated by the fact that multiple agencies with differing mandates collect such data (Table 1). For example, drinking water utilities routinely measure quantities such as UV-254, turbidity, and TOC in order to assess disinfection byproducts and bacterial contamination. In contrast, many do not sample the standard limnological data necessary to model water quality (species of carbon, nutrients, algae, etc.). It should be noted that water utilities are increasingly looking to watershed controls as a means to improve the quality of their raw water. Thus, beyond their utility for TMDLs, water-quality-oriented data would prove directly useful in such efforts.

The second class of data is that collected to support past modeling studies. Such studies were often conducted to assess oxygen and/or eutrophication during the 1960s and 1970s. These data sets are often imperfect, in the sense that they might not be compatible with present analytical and modeling standards. However, they

Table 1. Different Measures of Organic Carbon Employed by Four Groups

Type	Measures of organic carbon
Water quality compliance	CBOD, TOC
Sewage treatment effluent	COD, CBOD
Drinking water	UV254, TOC
Modern WQ model	labile DOC, refractory POC, detrital POC, algal POC

Note: CBOD=carbonaceous biochemical oxygen demand; TOC=total organic carbon; COD=chemical oxygen demand; UV254=absorption of ultraviolet light at 254 nm; DOC=dissolved organic carbon; POC=particulate organic carbon.

can still have great value for model assessment. In particular, they might prove extremely useful in model confirmation, as discussed in the next section.

The final class is new data collected expressly for model support. Although these would seem to be the most reliable, problems can occur because modelers are sometimes not included in the sampling design process. In such cases, the correct variables and sampling frequencies required for the model may be omitted. This sometimes occurs because agency monitoring groups are employed to collect the data. Because they are set up to measure traditional variables, they are sometimes unaware that the modelers might require differing and additional data. It can also occur because the modeler is hired after the data collection effort is already underway.

It should be noted that the question of sampling design is a "two-way street." As mentioned above, monitoring groups should be sensitive to the needs of the modelers. On the other hand, modelers should not propose state variables that are difficult or impossible to measure. For example, it makes no sense to model six species of phosphorus if it is economically infeasible or operationally impossible for typical laboratories to measure them. It should be stressed that purely scientific models developed for exploratory purposes are not necessarily subject to this constraint.

It should also be stressed that along with adequate state data, accurate estimates of the model forcing functions are equally critical. If the loadings are highly inaccurate, model calibration and confirmation become meaningless exercises.

Some general conclusions regarding modeling and data collection:

- Modelers should be an integral part of the sampling design process.
- All modelers should use the same data. The most conflicted analyses occur when modelers use different data. Any data that bear on the TMDL decision process should be available to all parties involved in the process.
- Data should be archived with accompanying quality assurance information (precision, method, etc.).
- Data collection should be coordinated among different collectors. Although each entity must collect data that support its own mission, additional parameters might be included at marginal cost in order that the data sets are more broadly useful.

Model Calibration and Confirmation

Once adequate data sets are compiled, the model should be run and compared with state data. The initial runs can be made with default parameters and any site-specific measured process rates that are available (e.g., reaeration rates, settling velocities, etc.).

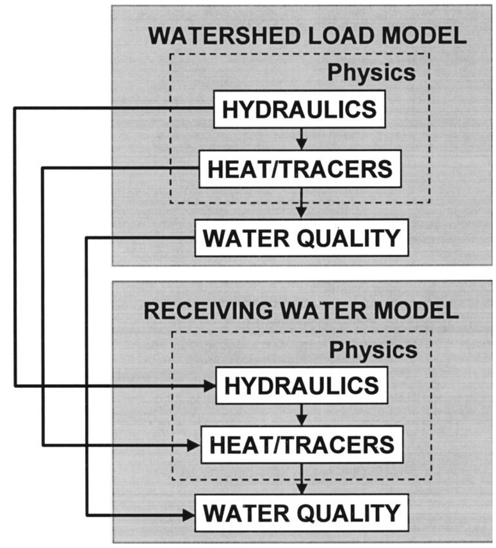

Fig. 5. Natural flow of model calibration process

Invariably this will result in a poor "fit" to the data. At this point, the model parameters are adjusted to optimize agreement between the model output and the state data. If the final fit is deemed adequate, the model is considered calibrated. Assessing the adequacy of a model fit involves graphical comparisons and statistical tests. Because the subject has been discussed in great depth elsewhere (Reckhow and Chapra 1983a; Berthoux and Brown 2002; McCuen and Snyder 1986; etc.), I will not cover it here.

When the submodels composing the framework are independent, the calibration process can be conducted sequentially (Fig. 5). This sequence is dictated by the natural information flow between the major submodels: watershed→receiving water. Similarly, within each submodel there is a hierarchy dictated by both information flow and the uncertainty of the estimates. First, the hydraulic model (flows) would be calibrated. Next, heat and conservative tracers can be calibrated to (1) provide an independent check on the hydraulics, and (2) demonstrate that the constituent transport is adequate. Finally, the least certain part of the process—the water quality component would be simulated. Note that in stratified systems, temperature and tracers can significantly affect the hydraulics, and hence the three must be calibrated simultaneously. The important point is that once the physical processes are calibrated, they should be not modified during the water quality calibration.

Three techniques can greatly enhance the model calibration process:

- **Sensitivity analysis**. The model parameters can be perturbed, and the variations in the state variables observed. A number of techniques are available for this, including first-order error analysis, Monte Carlo simulation, and generalized sensitivity analysis (Spear and Hornberger 1980). The objective is to identify which parameters have the greatest impact on key state variables. This information can improve manual calibration by guiding the modeler to focus on the most sensitive parameters. It can also influence decisions regarding direct rate measurements, as described next.
- **Direct rate measurements**. As stated above, the conventional approach to model calibration involves adjusting uncertain model parameters such that a model output time series compares favorably with state-data time series. Another approach is to directly measure model parameters so that they are esti-

mated with greater precision and accuracy. Common examples include reaeration rates, settling velocities, primary production rates, community respiration rates, and sediment oxygen and nutrient fluxes. If these rates can be ''pinned down'' via field or laboratory experiments, the model's degrees of freedom can be reduced. The intent is to reduce the degrees of freedom sufficiently that fewer parameters are subject to adjustment.

Finally it should be noted that autocalibration might have some utility in guiding and informing the calibration process. This involves setting up bounds for automatically adjusting model parameters (within constrained ranges) to optimize some objective function that reflects goodness of fit (e.g., least squares). For systems with a large data uncertainty, one difficulty with such approaches is that the resulting parameters often bump up against the constraints, suggesting that the optimum lies outside the parameter space. Nevertheless, because of advances in computing speed, the value of such automatic calibration approaches should be explored.

Once the model is calibrated, its reliability must be established. This process has traditionally been mislabeled as verification or validation. As pointed out by Reckhow and Chapra (1983b), validation (the ascertainment of truth) is inconsistent with the logic of scientific research. The only real validation of a model is confirmation by independent observations (Anscombe 1967). The testing of scientific models is considered an inductive process, which means that, even with true premises we can at best assign high probability to the correctness of the model. Thus, the terms *confirmation* or *corroboration* are preferred (Reckhow and Chapra 1983b; Oreskes et al. 1994).

The purpose of confirmation is not to validate that the model is ''true'' but rather to ensure that the model predictions are considered sufficiently credible for decision making. The fact that models can never be absolutely verified has significant policy implications. By admitting that models are approximations, it negates stall tactics based on the premise that remedial action be indefinitely postponed because models can never be demonstrated to be absolutely true.

From the standpoint of practical applications, the issue of model confirmation then reduces to two considerations: assessment of ''goodness'' of fit and required tests. As noted previously, the former has been addressed in detail elsewhere. The latter will be discussed here.

In essence, there is a hierarchy of tests that can be applied:

- **Level 0**: Application to a case almost identical to the calibration case. This is merely additional calibration disguised as confirmation. It is next to useless, unless the new case fails. One possible explanation for a failure would be that the original model was highly influenced by its initial conditions. This can commonly occur for long residence time systems such as large lakes.
- **Level 1**: Application to a case with different meteorology than the calibration case—for example, a wet year versus a dry year or a cold year versus a warm year. Such confirmation usually yields adequate corroboration for the physical model. In addition, it may partly corroborate the water quality model. Such would be the case for systems dominated by meteorologically driven nonpoint loads, which are typically highly influenced by runoff.
- **Level 2**: Application to a case with significantly different loadings. This provides a means to corroborate the model's adaptive mechanisms (e.g., species shifts, long-term shifts in sediment oxygen, and nutrient fluxes) as a result of loading changes.

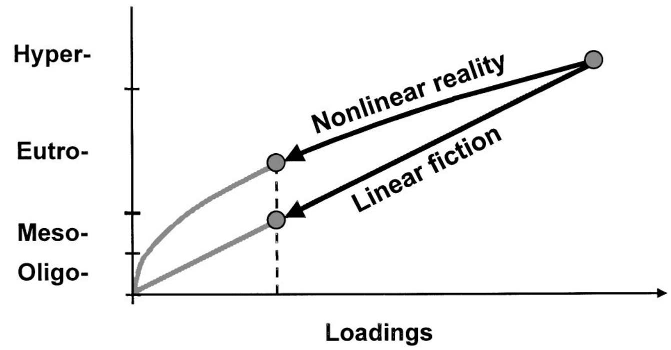

Fig. 6. Relationship between trophic state and loadings for conventional pollutants. More pristine systems tend to be more sensitive to load changes than highly degraded systems.

In practice, it is easier to obtain the proper data to confirm the physical model (Level 1). Given a three- to five-year observation period, it is likely that meteorological conditions would vary sufficiently to assess whether the hydraulics, tracers and temperature are simulated adequately.

Level 2 confirmation is more problematic. As with the physics, the intent is to simulate data sets that differ significantly from the calibration set. As noted, for systems dominated by nonpoint sources, it is possible that physically different years would exhibit different loadings. However, it is less likely that these meteorologically induced load variations would mimic the reductions needed to bring the system to the desired quality.

This is especially vexing when the water body is far from the desired target. This stems from the fact that the relationship between trophic state and loadings is nonlinear. If the trophic continuum were linear, one would expect that a 50% reduction in loadings would result in a 50% improvement in receiving water quality. In fact, nonlinear mechanisms, e.g., algal limitation (temperature, nutrients, and light), sediment-water interactions, and species shifts are nonlinear. And, as depicted in Fig. 6, the larger the load reduction, the more the curvature comes into play.

Today, nonlinear algal limitation is included in most models, and rational sediment-water submodels (e.g., Di Toro et al. 1990; Di Toro and Fitzpatrick 1993; Di Toro 2001) are increasingly being employed. The remaining issue is the inclusion of adequate constructs to predict species shifts. Although some models do include multiple phytoplankton groups (and a smaller number include attached plants), very few systematic tests have been conducted to corroborate whether these formulations adequately simulate species shifts across trophic states.

In rare cases, data are available for both the polluted case and either the prepolluted or postcleanup states. Lake Washington (in Seattle) represents a classic example (Edmondson 1994). The lake was enriched with increasing phosphorus loading from 1941 to 1963. Over this period, the abundance of algae increased severalfold. Further, the species composition became dominated by cyanobacteria (in particular, *Oscillatoria*), which were inedible by higher organisms. In the period 1963–1968, sewage was diverted around the lake. As a result, phosphorus concentration and phytoplankton abundance decreased. The *Oscillatoria* essentially disappeared and were supplanted by edible diatoms and green algae. The zooplankton assemblage became dominated by *Daphnia*, a filter-feeding zooplankton that is extremely efficient at clearing the water of edible phytoplankton. Hence, the water clarity improved to a greater degree than would be expected.

For calibrated model applications where the end points involve long-term water clarity and dominant species, applying them to systems like Lake Washington could provide a kind of cross-sectional confirmation. Such applications would involve Lake Washington's system-specific data, but with kinetic parameters taken from the calibration run. If the new run adequately predicts the observed changes, such an exercise greatly strengthens the model's credibility.

This type of cross-sectional confirmation of longitudinal models can be generalized to the notion of benchmark data sets. Such sets could be developed for the range of water bodies (e.g., lakes, impoundments, streams, rivers, estuaries) and water quality problems (eutrophication, pathogens, toxics, sediments, etc.) for which TMDLs are actively being pursued.

This idea can be further systematized by developing a confirmation portfolio for all modeling software used for TMDLs. Such a portfolio could comprise case studies (along with their input files) demonstrating that the model works adequately well for the systems and water quality problems for which it was designed. If peer reviewed, the portfolio could also serve as a source of benchmarks against which new models could be tested. Further, a documented track record of the model's general success would increase confidence in its application to cases where confirmation data were sparse.

As described next, water quality models should include estimates of model uncertainty. Thus, the confirmation might also include special cases/watersheds/waterbodies where error analysis was conducted for each of the large process models. This would at least provide some official, reported estimate of error. While this would not be quite the same as a site-specific error analysis, it would provide the model user with some sense for the uncertainty.

Finally, after the model has been calibrated and confirmed, the model can then be recalibrated to the entire data set to obtain an optimal fit (Reckhow and Chapra 1983a). By using all the available data, this pooling of information further strengthens the model's reliability in the actual TMDL prescription.

Models and Uncertainty

Several investigators have made persuasive arguments for including uncertainty as an essential and explicit part of the water quality modeling process, and the TMDL process in particular (e.g., Reckhow 1977, 2003; Reckhow and Chapra 1983a; NRC 2001; Beck 1987). Most engineers and scientists agree with this argument because we all know that (1) our models are imperfect and that (2) these imperfections are best expressed probabilistically. At minimum, this means that estimates of uncertainty accompany all model predictions and that the margin of safety (MOS) should be formulated probabilistically (Reckhow 2003).

That said, performing a complete error analysis of a process-oriented water quality model is not trivial. The fact that very few have been conducted (most notably by Di Toro and van Straten 1979 and van Straten 1983) supports this contention. Hence, because of the heavy data demands of a proper error analysis (i.e., considering both parameter and model error, as well as covariance) and the severe time constraints for TMDL development, it is simply unrealistic to require that such complete analyses be implemented as part of every TMDL process. Hence, Reckhow (2003) has suggested that, in the short term, more practical but incomplete uncertainty analyses could be conducted and incorporated into the decision process.

In the long term, he suggests, models might be restructured so that a relatively complete error analysis is feasible. Although such approaches could prove extremely useful, the suggestion to restructure modeling bears some scrutiny lest it be misinterpreted or misconstrued. At face value, restructuring might be taken to imply that simpler empirical/statistical approaches would supplant complex process-oriented/numerical models. Although the statistical approaches by nature greatly expedite a complete uncertainty analysis, something is lost in the bargain.

By their nature, empirical models are slaves to their training data sets. For cross-sectional models (based on data from many water bodies), this means that training data sets must span the entire range of decision alternatives. For example, a model intended to evaluate nutrient load should be trained on data spanning the full range of trophic states. Otherwise, predictions amount to highly uncertain extrapolations.

Although effective cross-sectional models can be developed in such a manner (particularly when developed regionally), prospects seem less sanguine for longitudinal models, that is, those based on time-series data from a single water body. Unless detailed historical data sets spanning polluted and unpolluted conditions are available, it would seem that the resulting predictions would again involve extrapolation.

Such models would also seem limited for highly distributed systems like rivers and estuaries with multiple inputs. In particular, the ability of such models to disaggregate the effects of individual point and nonpoint sources would seem to be problematic. The latter might be particularly difficult, because their dependence on precipitation might make them covary significantly.

These deficiencies suggest that whereas empirical approaches might yield more precise predictions (as reflected by lower uncertainty), they might be less accurate (as reflected by their ability to predict central tendency). Conversely, the complex process models might yield more accurate but more uncertain predictions.

Regardless, it is clear that both approaches have utility and will be important over the short term (three-year horizon). There will certainly be problem contexts where empirical approaches will be superior to process models, and vice versa. In fact, wherever they both have been developed, empirical approaches should be used in tandem with process-based models in a complementary rather than in competitive fashion. This occurred 25 years ago when empirical (Vollenweider et al. 1980), simple lumped mechanistic (Chapra 1980), and highly developed process models (Thomann and Segna 1980; Di Toro 1980) were used to develop a consensus regarding Great Lakes phosphorus control (Bierman 1980).

For the long term, research on model uncertainty should be increased with particular emphasis on rationalizing the margin of safety, facilitating uncertainty analysis through simpler models, and investigating how Bayesian approaches might be constructively employed. In addition, research should be directed toward practical and feasible approaches to incorporate uncertainty into process-oriented, mechanistic models. Finally, intermediate hybrid approaches could prove a useful means to capture the strengths of both types. These would consist of simpler process models that would account for key mechanisms, but which would be sufficiently simple to accommodate a more thorough yet feasible uncertainty analysis (e.g., Chapra 1977; Chapra and Robertson 1977; Walker 1985, 1986; Chapra and Canale 1991; Haith et al. 1992; Borsuk et al. 2001). Such models might be particularly useful in the sort of adaptive implementation scheme suggested in this volume by Reckhow (2003).

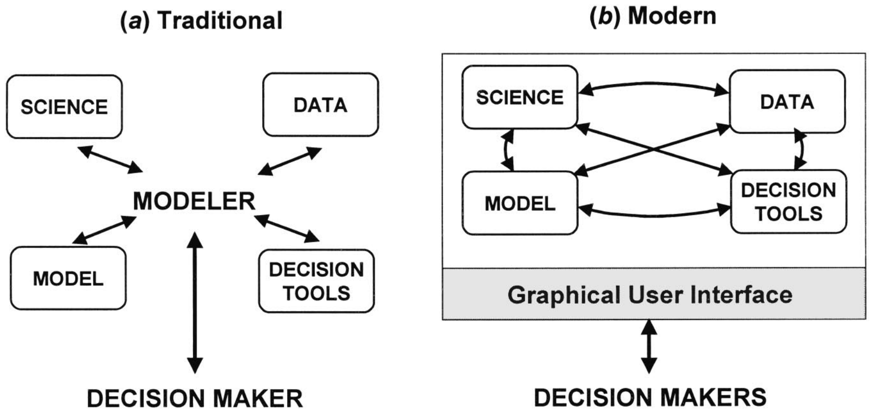

Fig. 7. Information flow between components in decision-making process: (a) Historical and (b) present. Note that the ''decision makers'' in (b) refer to both regulatory agencies and stakeholders.

As stated previously, the use of models for TMDLs is not a ''one size fits all'' type of endeavor. Although the inclusion of uncertainty is a laudable goal, it would be tragic if the issue undermines the great value of process models for TMDL prescription. This is particularly vexing in light of history. Despite the fact that little or no uncertainty analyses were conducted, process-oriented models have been used effectively over the past 75 years to determine load allocations for a broad range of pollutants in a wide range of receiving waters. Although admittedly imperfect, they have been deemed as sufficiently sound engineering tools for rationalizing water quality management. It would seem ironic if healthy discussions of uncertainty were misconstrued as a negation of this historical fact.

Decision Support

When computers were not ubiquitous and water quality problems dealt primarily with point-source discharges, the modeler was usually at the center of the decision process [Fig. 7(a)]. Thus, the modeler acted as the interface between the model analysis and a single decision-making agency.

In the late 1980s, environmental and water resource engineers began developing decision support systems, or DSS (e.g., Loucks et al. 1985; Fedra and Loucks 1985; Loucks and Fedra 1987; Reitsma et al. 1996; Chapra and Canale 1987). Due primarily to computer advances, a DSS can now be developed to allow decision makers and stakeholders to interact more intimately and efficiently with the modeling environment [Fig. 7(b)]. Today, computer improvements allow much of the modeling process to be integrated electronically (e.g., Chen et al. 1999). By designing a graphical user interface expressly designed for decision making, the modeler is no longer at the center of the process. Rather, the decision makers (including stakeholders) can be empowered to explore the decision space in a more transparent and direct fashion.

Before proceeding, it should be stressed that the simple fact that a model has an interface does not mean that it is a decision support framework. In fact, model interfaces are often written for modelers rather than to support the decision process. Modeling interfaces are usually motivated by the need to perform simulations in order to calibrate and confirm the model. Thus, they might include tools to expedite the preparation of input files, display graphs and statistical comparisons of model output and versus data, and perform sensitivity analyses.

By contrast, decision support interfaces should be designed for the needs of decision makers. In particular, they should enhance exploration of the decision space. For example, they would allow users to readily modify such forcing functions as loads and weather. In contrast, users would not be permitted to change model parameters established during the calibration/confirmation phases. In the same vein, scenario generation tools would be included so that users could conveniently assess varieties of management options.

As a final observation, although there has been extensive research on the use of optimization algorithms for decision support in a variety of water quality management contexts, including wasteload allocation of point sources of pollution (Thomann 1972; Loucks et al. 1981), there has been very little attention given to the problem of integrating optimization algorithms into the TMDL problem. In other words, the very essence of the TMDL decision problem has yet to be cast as a decision problem. Including an optimization component in DSS to allow cost-effective management scenarios to be identified could do this.

Modeling Infrastructure

Thomann's (1998) ''Golden Age'' may not be realized if sound infrastructure is not in place to support the modeling process. To date, most modeling support has been directed toward software development (e.g., BASINS). Equally important are software support, modeling institutions, and human expertise infrastructure.

Software Support

Historically, many of our major water quality frameworks have been developed as spin-offs from high-profile modeling projects (e.g., Ambrose et al. 1988; Cerco and Cole 1993). Others were developed by consulting firms, academia, and government and subsequently moved into the public domain. Over the years, these models have been archived, maintained, and updated by various government entities, such as the U.S. Environmental Protection Agency's Center for Exposure Assessment Modeling (CEAM), in Athens, Georgia, and the Army Corps' Waterways Experiment Station (WES) in Vicksburg, Mississippi. Unfortunately, these centers have not always been adequately funded to support their missions. Increased use and software complexity of water quality models have exacerbated this problem. As a consequence, model updates and upgrades are presently not implemented quickly enough to meet user needs and to keep pace with scientific advances.

A specific example is the QUAL2E model, which has not been modified since 1987 (Brown and Barnwell 1987). Hence this valuable framework, which is part of BASINS, is unsuitable for modeling shallow streams dominated by attached plants. Similarly, although a viable framework for modeling sediment-water fluxes was published nearly 10 years ago (Di Toro and Fitzpatrick 1993), this mechanism is just now being integrated into the publicly available models.

No software company in the world would stay in business operating in such an ad hoc fashion. Government agencies should establish some centralized mechanism responsible for maintenance, upgrading, and quality assurance of their software products. These groups should consist of modelers as well as software engineers.

Institutions

As stated above, government agencies have decommissioned, deemphasized, or dispersed much of their water quality centers of excellence. The most notable of these has been the Center for Exposure Assessment Modeling (CEAM), in Athens, Ga. The government should reestablish and strengthen such centers. This would have a number of benefits, including quality control, standardization, and maintenance (including updates). In addition, a modeling center could have a research and development component that would allow scientific advances to be more rapidly integrated into modeling practice.

State agencies should assemble their own modeling teams. These teams should guide or implement models developed for their particular state's TMDLs. In addition, they should play the critical role of archiving models and data and maintaining and upgrading them for future applications.

Finally, the old idea of basin commissions might be revived. These arose in the 1960s to acknowledge that watersheds were the most rational vehicles to organize water quality management. In addition, they were extremely useful in managing interstate waters. Such entities are arising today in an ad hoc fashion. For example, the Charles River Watershed Association, in Massachusetts, has a technical stewardship function on that watershed. This group conducts sampling and modeling for the watershed. Most importantly, they serve as a vehicle to maintain the system's long-term institutional data and modeling memory.

Expertise

Water quality modeling is not a ''point-and-shoot'' endeavor. No matter how advanced the software, modelers must marshal skill, knowledge, experience, and good judgment in order to be effective. As Di Toro and Thuman (2001) put it:

> ... water-quality models are not simple, straightforward engineering calculations. The methodology has not progressed to the handbook stage, and perhaps the following analogy is useful: Models are less like a radio—plug it in, turn it on, and it produces beautiful music—and more like a violin. Only a talented and well-trained violinist can produce beautiful music.

Unfortunately, today there is a serious deficiency in water-quality-modeling expertise. The expertise deficiency is due to a number of factors. Because of a lack of funding of academic modeling research over the past 20 years, few universities offer graduate programs specializing in water quality modeling. Furthermore, many professionally oriented training courses emphasize the use of tools rather than the art and science of modeling. Because models are becoming easier to use, there are currently a large group of ''modelers'' who are essentially button-pushers.

In the short term, several suggestions might help reverse this trend:

- Modeling workshops should place increased emphasis on modeler education, rather than model operation. Sufficient time should be devoted to theory so that modelers understand the inner workings of the software implementations. Such application issues as data analysis, calibration/confirmation, and interpretation of model sensitivity and uncertainty analysis could be stressed.
- Guidelines for the assessment of models and specific model applications should be developed. These should be sufficiently flexible to allow different modeling approaches but structured enough to establish standards for assessing quality.

The long-term prospects depend upon the government and universities recognizing that environmental modeling goes well beyond software development. From the government's perspective, one idea would be to direct more graduate traineeships and fellowships toward environmental modeling. Increased funding in support of modeling science would begin to encourage universities to generate the modeling graduate students required to sustain the discipline.

Conclusion

An old joke (Chapra and Reckhow 1983) goes like this: A scientist, an engineer, and a lawyer were asked the question. ''What is two plus two?'' The scientist immediately answered: ''Two plus two equals four.'' The engineer shook her head and retorted: ''Approximately two plus approximately two equals approximately four.'' Both then turned to the lawyer and demanded, ''What is your answer? What is two plus two?'' The lawyer stared back and calmly replied: ''Well, what would you like it to be?''

The joke may be old, but the sentiment remains the same. As has always been the case, engineers (and their engineering TMDL models) find themselves as the moderators between truth-seeking scientists and answer-seeking policymakers and stakeholders. Over the next 10 to 15 years, modelers could make a major contribution toward helping disparate groups reach consensus regarding the quality of their watersheds.

As I hope this paper has made clear, water quality modeling should not be allowed to become a commodity industry where the ''low bid'' rules the day. Rather, it is an academics-based discipline with a long history of intellectual development and scholarship, one that requires long-term nurturing and investment to be sustainable. Modeling itself is an expertise that is acquired through education and experience—and merely generating numbers does not a modeler make. Only by recognizing these facts will the Golden Age be realized.

Acknowledgments

This paper is based in part on a keynote address at the American Water Resource Association Annual Conference, in Point Clear, Alabama, on November 16, 1998. I would like to thank the referees for several important insights and suggestions.

References

Ambrose, R. B., Jr., Wool, T. A., Connolly, J. P., and Schanz, R. B. (1988). ''WASP4, a hydrodynamic and water quality model-model theory, user's manual, and programmer's guide.'' *EPA/600/3-87/039*, Environmental Research Laboratory, U.S. Environmental Protection Agency, Athens, Ga.

Anscombe, P. J. (1967). ''Topics in the investigation of linear relations fitted by the method of least squares.'' *J. R. Stat. Soc. Ser. B. Methodol.*, 29, 1–52.

Beck, M. B. (1987). ''Water quality modeling: A review of the analysis of uncertainty.'' *Water Resour. Res.*, 23(8), 1393–1442.

Berthouex, P. M., and Brown, L. C. (2002). *Statistics for environmental engineers*, 2nd Ed., CRC, Boca Raton, Fla.

Bierman, V. J., Jr. (1980). ''A comparison of models developed for phosphorus management in the Great Lakes.'' *Phosphorus management strategies for lakes*, R. C. Loehr et al., eds., Ann Arbor Science, Ann Arbor, Mich.

Borsuk, M. E., Higdon, D., Stow, C. A., and Reckhow, K. H. (2001). ''A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones.'' *Ecol. Modell.*, 143, 165–181.

Bowie, G. L., et al. (1985). *Rates, constants, and kinetic formulations in surface water quality modeling*, EPN600/3-85/040, U.S. Environmental Protection Agency, ORD, ERL, Athens, Ga.

Brown, L. C., and Barnwell, T. O. (1987). ''The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS.'' *EPA/600/3-87-007*, U.S. Environmental Protection Agency, Athens, Ga.

Cerco, C. F., and Cole, T. (1993). ''Three-dimensional eutrophication model of Chesapeake Bay.'' *J. Environ. Eng.*, 119(6), 1006–1025.

Chapra, S. C. (1977). ''Total phosphorus model for the Great Lakes.'' *J. Environ. Eng. Div. (Am. Soc. Civ. Eng.)*, 103(2), 147–161.

Chapra, S. C. (1980). ''Application of the phosphorus loading concept to the Great Lakes.'' *Phosphorus management strategies for lakes*, R. C. Loehr et al., eds., Ann Arbor Science, Ann Arbor, Mich.

Chapra, S. C. (1997). *Surface water quality modeling*, McGraw-Hill, New York.

Chapra, S. C. (1999). ''Organic carbon and surface water quality modelling.'' *Progress in Environmental Science*, 1(1), 49–70.

Chapra, S. C., and Canale, R. P. (1987). ''Personal computers and environmental engineering, Part I—trends and perspectives.'' *Environ. Sci. Technol.*, 21(9), 832–837.

Chapra, S. C., and Canale, R. P. (1991). ''Long-term phenomenological model of phosphorus and oxygen in stratified lakes.'' *Water Res.*, 25(6), 707–715.

Chapra, S. C., and Reckhow, K. H. (1983). *Engineering approaches for lake management Vol. 2, Mechanistic modeling*, Ann Arbor Science, Ann Arbor, Mich.

Chapra, S. C., and Robertson, A. (1977). ''Great Lakes eutrophication: The effect of point source control of total phosphorus.'' *Science*, 196, 1448–1450.

Chen, C. W., Herr, J., Ziemelis, L., Goldstein, R. A., and Olmsted, L. (1999). ''Decision support system for total maximum daily load.'' *J. Environ. Eng.*, 125(7), 653–659.

Connolly, J. P., and Coffin, R. B. (1995). ''Model of carbon cycling in planktonic food webs.'' *J. Environ. Eng.*, 121(10), 682–690.

Di Toro, D. M. (1980). ''The effect of phosphorus loading on dissolved oxygen in Lake Erie.'' *Phosphorus management strategies for lakes*, R. C. Loehr et al., eds., Ann Arbor Science, Ann Arbor, Mich.

Di Toro, D. M. (2001). *Sediment flux modeling*, Wiley-Interscience, New York.

Di Toro, D. M., and Fitzpatrick, J. J. (1993). ''Chesapeake Bay sediment flux model.'' *Technical Rep. EL-93-2*, U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, Miss.

Di Toro, D. M., and Thuman, A. (2001). ''It takes skill to make music.'' *Stormwater*, 2(5), 8.

Di Toro, D. M., and van Straten, G. (1979). ''Uncertainty in the parameters and predictions of phytoplankton models.'' *Working Paper WP-79-27*, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Di Toro, D. M., Paquin, P. R., Subburamu, K., and Gruber, D. A. (1990). ''Sediment oxygen demand model: Methane and ammonia oxidation.'' *J. Environ. Eng.*, 116(5), 945–986.

Edmondson. (1994). ''Sixty years of Lake Washington: A curriculum vitae.'' *Lake Reserv. Manage.*, 10(2), 75–84.

Environmental Protection Agency (EPA). (1990). *Technical guidance manual for performing waste load allocations*, *Book III: Estuaries*, *Part 2: Application of estuarine waste load allocation models*. J. L. Martin, R. B. Ambrose, Jr., and S. C. McCutcheon, eds., Office of Water, Washington, D.C.

Fedra, K., and Loucks, D. P. (1985). ''Interactive computer technology for planning and policy modeling.'' *Water Resour. Res.*, 21(2), 114–122.

Haith, D. R., Mandel, R., and Wu, R. S. (1992). *GWLF: Generalized watershed loading functions user's manual*, Version 2.0, Cornell University, Ithaca, N.Y.

Loucks, D. P., and Fedra, K. (1987). ''Impact of changing computer technology on hydrologic and water resource modeling.'' *Rev. Geophys.*, 25(2), March, 107–112.

Loucks, D. P., Kindler, J., and Fedra, K. (1985). ''Interactive water resources modeling and model use: An overview.'' *Water Resour. Res.*, 21(2), 95–102.

Loucks, D. P., Stedinger, J. R., and Haith, D. A. (1981). *Water resource systems planning and analysis*, Prentice-Hall, Englewood Cliffs, N.J.

McCuen, R. H., and Snyder, W. M. (1986). *Hydrologic modeling: Statistical methods and applications*, Prentice-Hall, Englewood Cliffs, N.J.

National Research Council (NRC). (2001). ''Assessing the TMDL approach to water quality management.'' *Rep. of the NRC Committee to Assess the Scientific Basis of the Total Maximum Daily Load Approach to Water Pollution Reduction*, National Academy Press, Washington, D.C.

Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994). ''Verification, validation, and confirmation of numerical models in the earth sciences.'' *Science*, 263, 641–646.

Reckhow, K. H. (1977). ''Phosphorus models for lake management.'' PhD dissertation, Harvard University, Cambridge, Mass.

Reckhow, K. H. (2003). ''On the need for uncertainty assessment in TMDL modeling and implementation.'' *J. Water Resour. Plan. Manage.* 129(4), 245–246.

Reckhow, K. H., and Chapra, S. C. (1983a). *Engineering approaches for lake management. Vol. 1, Data analysis and empirical modeling*, Ann Arbor Science, Ann Arbor, Mich.

Reckhow, K. H., and Chapra, S. C. (1983b). ''Confirmation of water quality models.'' *Ecol. Modell.*, 20, 111–133.

Reitsma, R., Zagona, E., Chapra, S. C., and Strzepek, K. M. (1996). ''Decision support systems (DDS) for water-resources management.'' *Handbook of water resources*, L. W. Mays, ed., McGraw–Hill, New York.

Schnoor, J. L., Sato, C., McKechnie, D., and Sahoo, D. (1987). ''Processes, coefficients, and models for simulating toxic organics and heavy metals in surface waters.'' *Rep. EPA/600/3-87/015*, Environmental Protection Agency, Athens, Ga.

Spear, R. C., and Hornberger, G. M. (1980). ''Eutrophication of Peel Inlet—II. Identification of critical uncertainties via generalized sensitivity analysis.'' *Water Res.*, 14, 43–49.

Thomann, R. V. (1972). *Systems analysis and water quality management*, McGraw-Hill, New York.

Thomann, R. V., and Segna, J. S. (1980). ''Dynamic phytoplankton-phosphorus model of Lake Ontario: Ten-year verification and simulations.'' *Phosphorus management strategies for lakes*, R. C. Loehr et al., eds., Ann Arbor Science, Ann Arbor, Mich.

Thomann, R. V. (1998). ''The future 'Golden Age' of predictive models for surface water quality and ecosystem management.'' *J. Environ. Eng.*, 124(2), 94–103.

van Straten, G. (1983). ''Maximum likelihood estimation of parameters and uncertainty in phytoplankton models.'' *Uncertainty and forecasting of water quality*, M. B. Beck and G. van Straten, eds., Springer, Berlin.

Vollenweider, R. A., Rast, W., and Kerekes, J. (1980). ‘‘The phosphorus loading concept and Great Lakes eutrophication.’’ *Phosphorus management strategies for lakes*, R. C. Loehr et al., eds., Ann Arbor Science, Ann Arbor, Mich.
Walker, W. W. (1985). ‘‘Empirical methods for predicting eutrophication in impoundments; Report 3, Phase III: Model refinements.’’ *Technical Rep. E-81-9*, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Miss.
Walker, W. W. (1986). ‘‘Empirical methods for predicting eutrophication in impoundments; Report 3, Phase III: Applications manual.’’ *Technical Rep. E-81-9*, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Miss.