American Bar Association
Section of Environment, Energy, and Resources

The Use and Misuse of Models in Water Disputes

Surface Water Hydrodynamic and Water Quality Models: Use and Misuse

Scott A. Wells
Department of Civil and Environmental Engineering
Portland State University
Portland, Oregon

23rd Annual Water Law Conference
San Diego, CA
February, 24-25, 2005

Introduction

The use of computer simulation models in resolving water and water quality issues is becoming common in the state and federal regulatory environment and in litigation. Often models are developed to answer either management and/or research questions. The research model and management model distinction is becoming blurred, since even management models are expected to be up-to-date with the latest scientific theory and observations. These models are used to examine management or "what-if" scenarios. The models are useful in terms of their ability to integrate complex inter-relationships between state variables that are too difficult in some cases for human intuition to track in a quantitative sense.

This paper will outline several topics on the theme of model use:

- Types of surface water models
- Correlation or Deterministic Models?
- Examples of model misuse
- The fallacy of validation

Types of Water Quality and Hydrodynamic Models for Surface Water

There are numerous water quality and hydrodynamic models that have been used for surface water systems: rivers, lakes, reservoirs, and estuaries. Some of these are shown in Table 1. There is a saying: "There are no bad models, just bad modelers". This is absolutely true in that if a model is used for the purpose for which it was constructed, it can be an appropriate tool. But often "modelers" (defined as anyone who can use the interface, double-click the executable and graph the results) misapply a model. How? The misuse of a model can fall into the following categories:

- Applying a model to a system for which the model was not defined
- Calibrating a model for the wrong reasons
- Mistakenly overlooking something in the model set-up

In all model applications, the use of peer-review is an important element in ensuring that the modeling is done correctly according to acceptable standards. Many of these models are very complex and most model users have very little formal training in all aspects of the model development which includes expertise in the following fields: chemistry, biology, physics, computer science, and mathematics. Much of the effort in model development in the last 20 years has been to enable modelers to use models more easily. This has led to individuals with little training using complex models and treating the model as a "black box."

Table 1 List of several water quality and hydrodynamic models.

Model Name	**Dimensionality**	**Model Type**	**Surface Water system**	**Reference**
QUAL2EU	1-D longitudinal	steady-state water quality (including temperature) and hydrodynamics (Manning's equation or rating curve)	rivers with steady-state inputs	Brown and Barnwell (1987)
SNTEMP	1-D longitudinal	daily average temperature model including shading, steady-state hydrodynamics (Manning's equation)	rivers where model predicts daily average temperature and a statistical regression model is used to compute daily maximum	Theurer et al. (1984), Bartholow (2000)
QUAL2K	1-D longitudinal	steady-state water quality and temperature (except for diel dynamics), and hydrodynamics (Manning's equation or rating curve)	rivers with steady-state inputs, updated QUAL2E model, set-up in the Excel environment	Chapra and Pelletier (2004)
HSPF	1-D longitudinal	dynamic water quality and hydrodynamic	channel routing and runoff model	Donnigan et al. (1984)

Model Name	Dimensionality	Model Type	Surface Water system	Reference
WASP	3-D	dynamic toxics and eutrophication water quality model	rivers, lakes, estuaries, requires hydrodynamic model	Ambrose et al. (1988)
DYNHYD	1-D longitudinal	dynamic hydrodynamics	rivers and estuaries, no stratified flow	Ambrose, Wool, Conolly, and Schanz (1988)
CE-QUAL-W2	2-D longitudinal-vertical	dynamic, water quality, temperature and hydrodynamics	rivers, estuaries, reservoirs, lakes	Cole and Wells (2003)
EFDC	3-D	dynamic, water quality, temperature and hydrodynamics	rivers, estuaries, reservoirs, lakes	Hamrick (1996)

Statistical Correlation Models or Complex Deterministic Models?

There have been criticisms of complex deterministic models (shown in Table 1). A typical critique of these models is that they are too complex, they take a lot of time and effort to calibrate, and even though they are complex, they are not complex enough. Opponents claim that statistical correlation models are simpler and easier to parse out complex interrelationships between state variables and forcing functions. Also, there is the refrain that the complex deterministic models are easily misused.

Statistical correlation models have their place in modeling, but they can suffer from serious deficiencies that only a more complex tool can resolve. The following example illustrates some of the typical issues that can arise with statistical correlation models.

In a two-year study of water quality in Brownlee Reservoir by Nurnberg and Brown & Caldwell (2001), a regression model approach was undertaken to predict the effect of management strategies on water quality. This project attempted to develop statistical correlations between water quality variables in order to predict the impact of management strategies. The over-riding concept was to minimize the complexity by ‘keeping it simple, stupid!” A typical result of this approach is shown in Figure 1, a statistical relationship between anoxic factor (how many days a year the hypolimnion would be anoxic) and the total phosphorus (TP) in the reservoir.

The underlying simplified view was that by controlling TP in the reservoir, the oxygen depletion in the reservoir could be controlled. In Figure 1, data relationships from 73 diverse lakes around the world were plotted with a best-fit regression line. When data from the reservoir were added to the correlation, field data from the riverine section of the reservoir and the lacustrine section of the reservoir did not agree with the “global” correlation model. But when the two data sets were averaged, they fit almost ‘exactly’ on the expected relationship (or “two wrongs make a right”). In this context, simple strategies were suggested to reduce TP in the reservoir with a predicted reduction in expected days of anoxia.

This particular reservoir was also studied using a deterministic model. A discussion of the deficiencies of the statistical correlation approach was discussed in Wells and Cole (2002). Besides developing a spurious statistical correlation (TP in the reservoir was not the controlling mechanism for hypolimnetic anoxia in this reservoir as it is in some lakes), the statistical model was fundamentally unable to answer any questions related to hydraulic management of the outflow from the reservoir (such as changing the discharge water level) or any situation not encapsulated within the range of the observed field data.

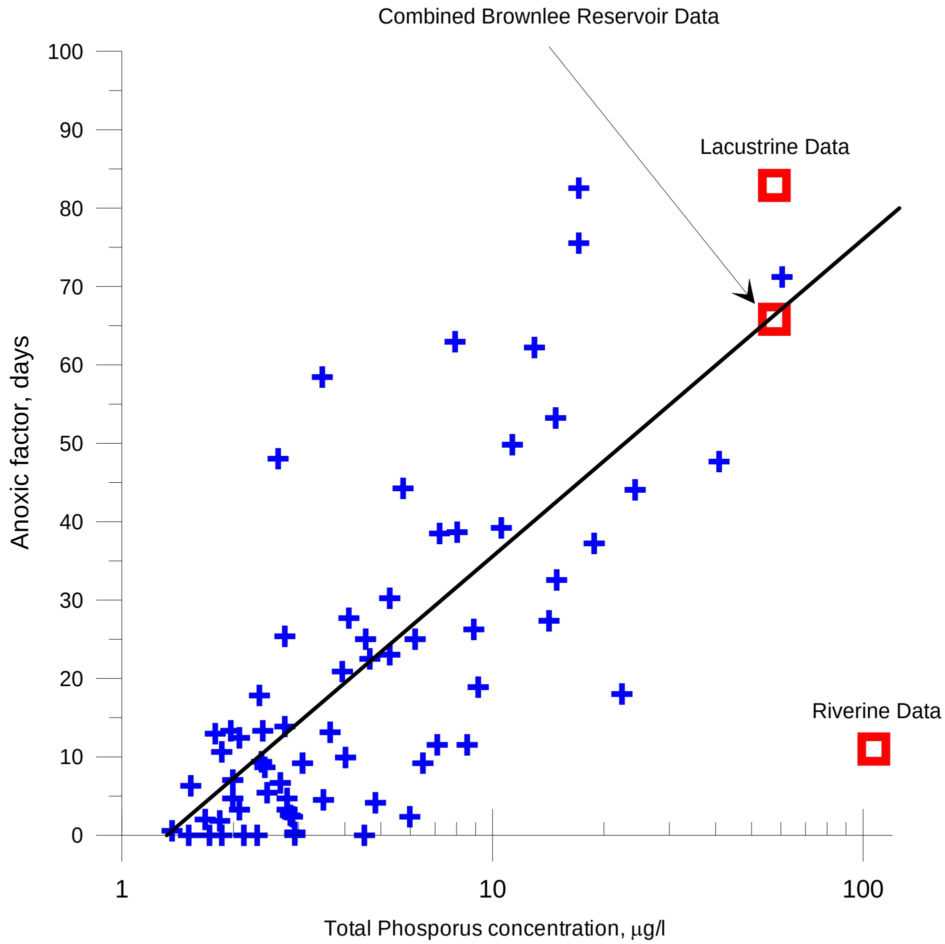

Figure 1. Correlation between AF and TP for Brownlee Reservoir and 73 other lakes (based on Nurnberg and B&C, 2001).

Misapplication of Models

There are many ways in which models can be misapplied. Often this can be a result of a lack of proper understanding of the underlying model theory and limitations or an overly zealous desire to match field data "at all costs" so that the model will "appear" correct. Several examples are highlighted below.

Inappropriate calibration technique – calibrating for the wrong reasons

The calibration process is illustrated in Figure 2. In many cases, modelers attempt to improve model predictive ability by adjustment of model hydrodynamic, temperature or kinetic parameters. In many cases though, a default set of model parameters is adequate to attain reasonable model predictive ability. Much of the real detective work in calibration is trying to estimate the unmeasured input forcing or deriving boundary conditions for the model from other related field data.

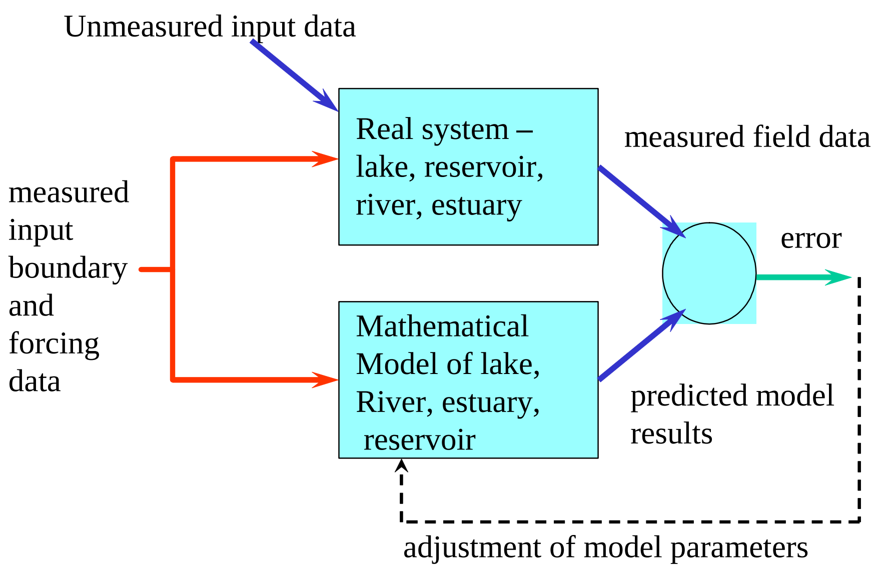

Figure 2. The calibration process.

Some modelers though have been 'caught' trying to adjust model coefficients to make up for a lack in proper boundary condition data. In one case of a modeling study for Coeur d'Alene Lake in Idaho, the modeling team chose model coefficients for nitrification that allowed nitrification to proceed even at temperatures near 4^{o}C. Their rationale: they needed to remove excess ammonia in the model predictions during the cooler months. In this case, the modelers calibrated the model for the wrong reason. Nitrification is a biological process that is very sensitive to temperature. In general, there is little or no nitrification occurring at temperatures near 4^{o}C. So what was wrong with their model? It was probably a poorly understood characterization of N loadings into the lake. A model set up in this way could not be used reliably to predict N dynamics.

Data are averaged – model results are averaged

In a paper by Tufford and McKellar (1999), a model was calibrated and verified for a reservoir in South Carolina. The authors applied WASP and DYNHYD to this lake assuming the water quality and hydrodynamic changes were in the horizontal direction (2-D horizontal model), ignoring lake stratification. Even though much of the reservoir is shallow, the reservoir lends itself to being stratified since the system had detention times of 2-3 months with a maximum depth of 24 m. Stratification affects the model hydrodynamics significantly. The DYNHYD model was not developed for modeling a stratified water body. Also, the model did not predict temperatures which are essential to all water quality and hydrodynamic predictions, especially in a stratified system. These issues were not addressed in the paper. Besides these issues, the authors compared average monthly monitoring data over 5 years (!) with monthly mean model results. So, the authors took instantaneous field data and averaged them; then they took instantaneous model predictions of water quality and averaged them. In the paper there were no comparisons of unaveraged field data with model predictions. The fact that some of the 5-year average data might be in the same ball-park as monthly average modeling data shows little about the model capability. It shows more about the expected range of water quality field data than about the model reliability. The authors also used 36-year mean observations of wind speed at 10-day intervals to determine an 'average' reaeration coefficient . "Flow measurements or estimates with daily frequency were entered into the model as 10-day arithmetic means…. because of a limitation with the DYHYD5 model." This consistent averaging of all input data, model results, and field data for calibration basically neutralizes the modeling effort and leads to a model that may look good but is severely compromised.

As an example, Figure 3 shows a hypothetical comparison of field data and model predictions of temperature over 24 hours for both an instantaneous and averaged case. The model poorly predicts the instantaneous field data but is in exact agreement with the 24-hour average. One modeler during a peer-review conference call told me that he was hesitant to show the actual model predictions compared to field data. He preferred to average both. Why? As he stated, "we don't want others to get the wrong impression that the model is not correct."

<u>Physical data are ignored or purposely changed to obtain better model-data agreement</u>

In a study of the Willamette River, a consultant constructed a mathematical model of the hydrodynamics (flow and water level) in the tidally influenced part of the Willamette basin. In their report (City of Portland, 1997), the modelers had problems matching tidal water level and flow reversals using the model DYNHYD. What was their problem and solution?

> "The upstream boundary conditions for the Willamette and Columbia Rivers were described by a time series of measured flow data. However, inputting this flow into the model at the study area boundaries did not allow the solution technique to simulate the upstream movement of tides beyond the boundary. A non-tidal boundary condition numerically forces flow into the specified model unit and artificially diminishes or eliminates the tidal effect near the boundary. This input characteristic then artificially reduces the tidal effect throughout the entire model network.
>
> The artificial constraint caused by non-tidal boundaries was overcome by moving the flow input from RM 105.5 to RM 144.5 on the Columbia River and RM 25 to RM 100 on the Willamette River…. in the actual, physical system tidal effects will be completely dampened out at or near the Oregon City Falls. However, in order to properly simulate tidal effects up to the Oregon City Falls on the Willamette River and beyond Vancouver in the Columbia River, model boundary similitude with the physical system had to be relaxed."

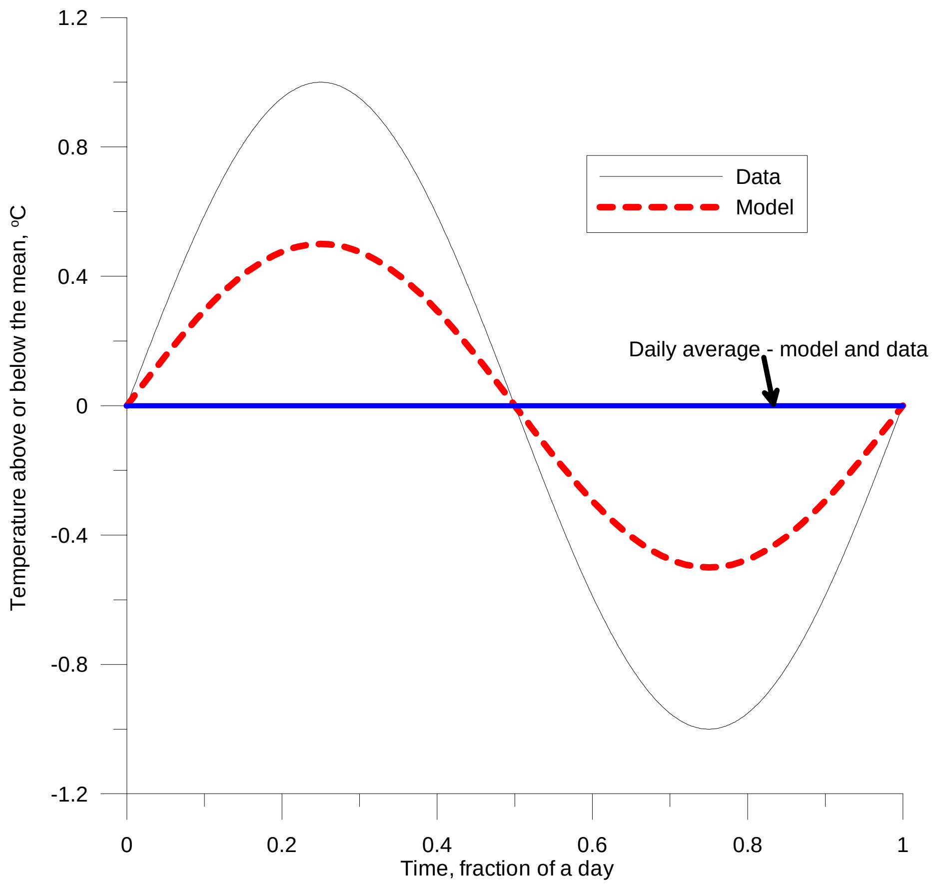

Figure 3. Hypothetical comparison of model predictions and field data of temperature for both instantaneous (poor agreement) and daily average (exact agreement) results.

In this case, the modelers decided that the physical system was not correct. Apparently the modelers thought that it was better to move the head of tide, the Willamette Falls, 75 miles further upstream. Their explanation is full of incorrect theoretical reasoning. Ideally, the more accurate the model represents the physical system, the more accurate the model will be. If the model does not represent the physical system correctly and matches field data, there is something very wrong with the model set-up. The results of the above modeling study may have matched data better with this change but for all the wrong reasons. Results from this study were erroneous and were later redone with a proper representation of the physical system (Berger et al. 2002).

Model results do not show all model-data comparisons

As alluded to in many of the examples already shown, an honest modeling study will show all model-data comparisons of the model predictions and field data. Model results and field data can be averaged or processed in an aggregate way to show trends, but the accuracy of the model should be judged on the data-model comparisons without averaging if the model is a dynamic one. A well-thought out report would then have graphical and statistical summaries of all calibration field data.

Modelers are often afraid to show model-data errors

Modeling clients do not want their model to show serious defects especially if it will be subjected to scrutiny by non-modelers who expect “perfect” model-data agreement. As an example, Wells et al (2003) undertook a modeling study of hydrodynamics and temperature of the Willamette River basin – a system of over 1000 km in length. There were model comparisons to field data of temperature at close to 100 sites in the basin, with many of these being continuous temperature sites. For most of those sites the typical absolute mean error between model predictions and field measurements was less than 1^{o}C (and usually less than 0.5^{o}C). But for 1 site on the McKenzie River, the model errors were close to 2^{o}C. An exhaustive effort to calibrate the model was undertaken but with little success. Theoretical issue after theoretical issue was reviewed; boundary condition data were scrutinized; model coefficients were reviewed; but with no success in reducing the model error. At sampling sites above and below this one site, temperature errors were about 0.2^{o}C. It was finally determined that the temperature data were mislabeled. These temperature measurements came from a sampling site on a different river in a different part of Oregon. How embarrassing it would have been to have been able to ‘calibrate’ the model to this erroneous data! Often, showing a case of poor model-data agreement can spur further investigation into the real causes of the problem – which could be poor data or lack of inclusion of an important process (perhaps groundwater inflow). The real purpose of modeling is to refine our understanding of the system. Showing areas where model and field data do not agree can in the end improve the overall model.

The Validation Fallacy

How often modelers will claim: “The model is both calibrated and validated”. Or a reviewer will ask: “I know you calibrated the model, but did you validate it?” This concept can be misused. The typical definition of validation is that the model matches an independent data set (independent from the calibration time period) without adjustment of model parameters determined during the calibration period. Many regulatory agencies require that a model be both calibrated ***and*** validated. Also, this term is used interchangeably with other terms, such as verification, corroboration, authentication, substantiation, and confirmation. This validation step is performed in order to add credibility to the modeling process.

In a recent study of a river in the state of Washington by Payne and Associates and Parametrix (2002), the authors state:

> “Typically, to have confidence that a calibrated stream temperature model will predict accurately over a wide range of flows and climate conditions, the model will be validated. Validation is generally accomplished by applying the global calibration factors to another independent set of data, or by splitting the available data set into two equal-sized sets, and running the model as a test of the calibration. Statistics for the validation data that are comparable to those for the initial calibration provide confidence in the calibration.”

The authors took field data taken over a 2 month period from June 19-August 20, 2002 and split these data into a calibration set and a validation set. The term validation is not being applied correctly when the modeler takes a small time series of data and says half of it is calibration and the other half is validation. How could the model be said to be calibrated (much less validated) to weather and flow conditions that did not occur during the calibration or verification period? Could one say that field data taken hourly in the morning is good for the calibration period and data taken hourly in the afternoon is good as a validation set?

Besides the issue of whether the period of calibration or validation is statistically long enough, the real issue is that validation is not a valid concept. It is all calibration. For example, let us assume that a modeler applied his/her model without changing parameter values from the calibration period, and the model predictive ability was poor. Would the modeler be finished and say, “the model calibrated but was not validated”? No, the modeler would go back and re-examine his original calibration and investigate whether the model was calibrated correctly in the first place. If by changing the original calibration, the model then represented field data well during the validation period, the modeler has only calibrated the model to multiple time periods. If a model is applied to an independent data set and the model matches data well with the original parameter set, then one can say that the model was calibrated well to the 2 time periods under consideration. When the term validation is used, it makes others think that the model is “valid” and does not have serious weaknesses. This though can be an inappropriate label. Hence, discarding the term altogether would eliminate this misconception.

Conclusions

Computer simulation models are powerful tools when applied properly. As with all complicated tools, there are situations when models are misused. In order to ensure that a given model application is reasonable, the following suggestions are recommended:

- Peer review should be built into the modeling study from its inception.
- All model results from dynamic models should be animated in order to detect errors in the model or boundary condition data.
- All model boundary conditions should be presented graphically in a model report.
- All field data (that pass a QA/QC check) should be compared to model results.
- For a dynamic model, model predictions and field data should be shown graphically and statistically on an instantaneous basis. Averages of model predictions and field data should be viewed in the context of the instantaneous comparisons.
- Model reports should always have a list of items for improvement and suggestions for the collection of field data that could improve model performance.

References

Ambrose, R. B.; Wool, T.; Connolly, J. P.; and Schanz, R. W. (1988) "WASP4, A Hydrodynamic and Water Quality Model: Model Theory, User’s Manual, and Programmer’s Guide," Envir. Res. Lab., EPA 600/3-87/039, Athens, GA, 1988.

Bartholow, J.M. (2000) The Stream Segment and Stream Network Temperature Models: A Self-Study Course, Version 2.0 U.S. Geological Survey Open File Report 99-112. 276pp.

Berger, C.; Annear, R. and Wells, S. (2002) “Willamette River and Columbia River Waste Load Allocation Model,” Proceedings, 2nd Federal InterAgency Hydrologic Modeling Conference, Las Vegas, July 28-Aug 1, 2002.

Chapra, S.C. and Pelletier, G.J. (2004) QUAL2K: A Modeling Framework for Simulating River and

Stream Water Quality: Documentation and Users Manual. Civil and Environmental Engineering Dept., Tufts University, Medford, MA.

City of Portland (1997) Technical Memorandum T3.B.4. Willamette River Hydraulics Characterization, Willamette River CSO Predesign Project, Technical report, 33 pages.

Cole, T. and Wells, S.A. (2003) “CE-QUAL-W2: A Two-Dimensional, Laterally Averaged, Hydrodynamic and Water Quality Model, Version 3.2,” Instruction Report EL-2000-, USA Engineering and Research Development Center, Waterways Experiment Station, Vicksburg, MS.

Donigian, A.S., Jr., J.C. Imhoff, B.R. Bicknell and J.L. Kittle, Jr. (1984) "Application Guide for Hydrological Simulation Program Fortran (HSPF)," EPA-600/3-84-065, U.S. Envir. Prot. Agency, Athens, GA.

Hamrick, J.M. (1996) User's Manual for the Environmental Fluid Dynamics Computer Code, Special Report No. 331 in Applied Marine Science and Ocean Engineering, School of Marine Science, Virginia Institute of Marine Science, The College of William and Mary, Goucester Point, VA

Nurnberg, G. and Brown and Caldwell (2001) “Quantification of epilimnetic anoxia and hypoxia in Brownlee Reservoir,” report prepared for the City of Boise, Idaho.

Payne and Associates and Parametrix (2002) “Chelan River Stream Network Temperature Model, Lake Chelan Hydroelectric Project FERC Project No. 637,” prepared for Public Utility District No. 1 of Chelan County, Wenatchee, Washington.

Tufford, D. and McKellar, H. (1999) Spatial and temporal hydrodynamic and water quality modeling analysis of a large reservoir on the South Carolina (USA) coastal plain, Ecological Modeling, 114, 137-173.

Theurer, F.D., Voos, K.A., and Miller, W.J. (1984) Instream Water Temperature Model, Instream Flow Information Paper 16, U.S. Fish and Wildlife Service, FWS/OBS-84/15.

Wells, S. A. and Cole, T. M. (2002) ”TMDLs: Statistical Correlations or Mechanistic Modeling?”, Proceedings National TMDL Science and Policy Conference, Phoenix, AR, November 13-16, 2002.

Wells, S. A., Berger, C. J., Annear, R. L., McKillip, M. and Jamal, S. (2003) “Willamette River Basin Temperature TMDL Modeling Study,” Proceedings National TMDL Science and Policy Conference, Chicago, IL, November 16-19, 2003.