The modeled and observed response of Lake Spokane hypolimnetic dissolved oxygen concentrations to phosphorus inputs

Michael T. Brett, Suvi K. Ahopelto, Hunter K. Brown, Barton E. Brynestad, Titus W. Butcher, Emily E. Coba, Christina A. Curtis, Jebessa T. Dara, Kallie B. Doeden, Kevin R. Evans, Lu Fan, Josh D. Finley, Noah J. Garguilo, Senait M. Gebreeyesus, Marissa K. Goodman, Kenneth W. Gray, Crystal Grinnell, Kathryn L. Gross, Brian R.E. Hite, Andrew J. Jones, Philip T. Kenyon, Angela M. Klock, Raji E. Koshy, Amanda M. Lawler, Man Lu, Luke Martinkosky, Jeremy R. Miller-Schulze, Quyen T.N. Nguyen, Elizabeth R. Runde, Joanie M. Stultz, Songlin Wang, Francesca P. White, Corey H. Wilson, Andrea S. Wong, Shirley Y. Wu, Peter G. Wurden, Tessora R. Young & George B. Arhonditsis

Taylor & Francis
Taylor & Francis Group

The modeled and observed response of Lake Spokane hypolimnetic dissolved oxygen concentrations to phosphorus inputs

Michael T. Brett[a], Suvi K. Ahopelto[a], Hunter K. Brown[a], Barton E. Brynestad[a], Titus W. Butcher[a], Emily E. Coba[a], Christina A. Curtis[a], Jebessa T. Dara[a], Kallie B. Doeden[a], Kevin R. Evans[a], Lu Fan[a], Josh D. Finley[a], Noah J. Garguilo[a], Senait M. Gebreeyesus[a], Marissa K. Goodman[a], Kenneth W. Gray[a], Crystal Grinnell[a], Kathryn L. Gross[a], Brian R.E. Hite[a], Andrew J. Jones[a], Philip T. Kenyon[a], Angela M. Klock[b], Raji E. Koshy[a], Amanda M. Lawler[a], Man Lu[a], Luke Martinkosky[a], Jeremy R. Miller-Schulze[a], Quyen T.N. Nguyen[a], Elizabeth R. Runde[a], Joanie M. Stultz[a], Songlin Wang[a], Francesca P. White[a], Corey H. Wilson[a], Andrea S. Wong[a], Shirley Y. Wu[a], Peter G. Wurden[a], Tessora R. Young[a], and George B. Arhonditsis[c]

[a]Department of Civil and Environmental Engineering, University of Washington, Seattle, WA; [b]School of Environmental and Forest Sciences, University of Washington, Seattle, WA; [c]Ecological Modeling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario, Canada

ABSTRACT

Brett MT, Ahopelto SK, Brown HK, Brynestad BE, Butcher TW, Coba EE, Curtis CA, Dara JT, Doeden KB, Evans KR, Fan L, Finley JD, Garguilo NJ, Gebreeyesus SM, Goodman MK, Gray KW, Grinnell C, Gross KL, Hite BRE, Jones AJ, Kenyon PT, Klock AM, Koshy RE, Lawler AM, Lu M, Martinkosky L, Miller-Schulze JR, Nguyen QTN, Runde ER, Stultz JM, Wang S, White FP, Wilson CH, Wong AS, Wu SY, Wurden PG, Young TR, Arhonditsis GB. The modeled and observed response of Lake Spokane hypolimnetic dissolved oxygen concentrations to phosphorus inputs. Lake Reserv Manage. 00:1–13.

Lake Spokane, a reservoir in eastern Washington State, was previously hypereutrophic due to phosphorus discharges from the City of Spokane wastewater treatment plant (WWTP). This reservoir subsequently recovered to a meso-oligotrophic state after implementation of advanced phosphorus removal. The present study tested whether the mechanistic Lake Spokane water quality (WQ) model realistically represents the sensitivity of this reservoir's hypolimnetic oxygen concentrations to phosphorus inputs. We compared the observed relationship between the mean summer input total phosphorus concentration (TP_{IN}) and the minimum volume weighted hypolimnetic dissolved oxygen concentration (DO_{MIN}) to model values for conditions ranging from hypereutrophic to oligotrophic. Prior to advanced phosphorus removal, TP_{IN} and DO_{MIN} averaged 86 ± 37 (SD) µg/L and 1.4 ± 1.3 mg/L, respectively. Currently (2010–2014), these values average 14 ± 3 µg/L and 6.5 ± 0.8 mg/L, respectively. By contrast, the model's DO_{MIN} response for similar TP_{IN} concentrations was much less pronounced, with hypereutrophic and contemporary DO_{MIN} averaging 3.8 ± 0.4 and 4.7 ± 0.04 mg/L, respectively. The model also has a structural DO deficit (saturated DO − DO_{MIN}) of 5.3 mg/L that was evident when all TP inputs to the reservoir were set to zero. Similarly, when all WWTP effluent sources were set to $TP_{EFF} = 0$ µg/L, the reservoir epilimnetic TP concentrations were ≈8 µg/L higher than the Spokane River inputs. The water quality model indicates that even if effluent phosphorus concentrations are reduced to zero, the dissolved oxygen goals for Lake Spokane cannot be met.

Mechanical models are important tools for researching and managing water quality in lakes, reservoirs, and estuaries (Arhonditsis and Brett 2004). These models can be used to carry out detailed scenario analyses, including how changing nutrient inputs or climate conditions affect aquatic ecosystem properties. The validity of earth system models has been strongly challenged, however, and there have been several attempts to rigorously address issues pertaining to model structure and input error (Oreskes et al. 1994). Model input error mainly stems from the uncertainty underlying the specification of model parameters, initial conditions, and forcing functions as well as the realization that all models are drastic simplifications of reality that approximate the real processes with spatially and temporally averaged parameter values (Rykiel 1996, Arhonditsis and Brett 2004, NRC 2007).

CONTACT Michael T. Brett ✉ mtbrett@uw.edu

🅑 Supplemental data for this article can be accessed on the publisher's website.

Model practitioners also commonly encounter the problem that several distinct choices of model inputs can lead to the same model outputs, and therefore many parameter sets may fit the data equally well. This non-uniqueness of model solutions is known in the modeling literature as equifinality (Beven 2006). The main reason for the equifinality (or poor identifiability) problem is that the causal mechanisms and hypotheses used for understanding how the system works internally are of substantially higher order than the data that can be realistically collected from real systems. Therefore, following specific best-professional-practices is essential for evaluating environmental models, including proper calibration, verification, sensitivity analyses, and error quantification (Schladow and Hamilton 1997, Stow et al. 2003, Arhonditsis and Brett 2004, NRC 2007, EPA 2009, Wellen et al. 2015). Although it is impossible to show that any environmental model is a correct representation of actual processes (Beven 2006), rigorous model evaluation can be used to show a particular model is a poor representation of a specific system (Oreskes et al. 1994).

The goal of this analysis was to determine whether the Lake Spokane water quality model (henceforth the Lake Spokane WQ model) addresses the key environmental attributes on which to base the Spokane River basin total maximum daily load (TMDL) decision. The Lake Spokane WQ model is a site-specific parameterization of CE-QUAL-W2 (Berger et al. 2009). To answer this question, we conducted a sensitivity analysis to quantify how well the Lake Spokane WQ model predicts the annual minimum hypolimnetic dissolved oxygen (DO) concentrations for a range of summer input total phosphorus (TP_{IN}) concentrations spanning the highest in the observational record ($TP_{IN} \approx$ 140 µg/L) to the future input target ($TP_{IN} \approx$ 10 µg/L). This study also quantified how well the Lake Spokane WQ model represents the mechanistic basis of the DO TMDL (i.e., TP_{IN} determines reservoir algal biomass, which in turn determines minimum annual hypolimnetic DO) based on model calibration data, as well as assessed the quality of the Lake Spokane phosphorus data on which the model was based. Furthermore, this study conducted a sensitivity analysis of the influence of a wide range of parameter values for the maximum phytoplankton growth rates, phytoplankton settling velocities, organic matter decomposition rates, and sediment oxygen demand on model DO outputs. Our analysis showed the Lake Spokane WQ model had a strong bias toward low late summer hypolimnetic DO

concentrations and mesotrophic epilimnetic TP concentrations, even when the wastewater treatment plant (WWTP) effluent TP and riverine input phosphorus concentrations were set to extremely low values. The model indicates that even if effluent phosphorus concentrations were reduced to zero, the water quality goals of the TMDL cannot be met.

Study site

Lake Spokane, Spokane County, Washington State, has a history of severe eutrophication and subsequent recovery. This case study is unique because there is also an observational record for this reservoir spanning extremely high historic to moderately low current TP input concentrations ($TP_{IN} \approx$ 140 and 12 µg/L, respectively) and corresponding summer minimum volume weighted hypolimnetic concentrations of $DO_{MIN} \approx$ 1.4 ± 1.3 (SD) and 6.5 ± 0.8 mg/L, respectively (Welch et al. 2015). During its hypereutrophic phase in the early 1970s, TP discharges from upstream conventional wastewater treatment plants (effluent $TP_{EFF} \approx$ 3500 µg/L) were identified as the critical driver of eutrophication in Lake Spokane (Soltero et al. 1974). In 1978, advanced phosphorus (P) removal was added to the City of Spokane WWTP, which reduced effluent TP concentrations to $TP_{EFF} \approx$ 500 µg/L and caused Lake Spokane to shift to a mesotrophic state (Soltero and Nichols 1984) and ultimately to meso-oligotropic (Welch et al. 2015). Currently, a total maximum daily load (TMDL) is in place to control TP discharges to Lake Spokane to meet a DO standard of "no measurable (0.2 mg/L) decrease from natural conditions when dissolved oxygen levels are lower than aquatic life criteria for core summer salmonid habitat (9.5 mg/L)" (Moore and Ross 2010). Approximately $900 million have been spent in the previous decade and budgeted to be spent in the coming decade to expand and modernize the wastewater, stormwater, and combined sewer overflow treatment infrastructure in the Spokane River Basin. This cost estimate reflects total capital costs for these facilities. Many, but not all, of these design upgrades are a component of strategies to reduce P discharges to meet the TMDL goals.

The TMDL model

The water quality model CE-QUAL-W2 has been used as the central analytical tool to determine which WWTP phosphorus removal targets are

needed to comply with the DO TMDL, as well as to determine "natural conditions" in Lake Spokane (Moore and Ross 2010). CE-QUAL-W2 is a 2-dimensional, laterally averaged, hydrodynamic and water quality model. Because this model assumes lateral homogeneity, it is commonly used to represent water quality characteristics of long and narrow reservoirs (Cole and Wells 2003). The present study is not designed to assess the utility of the CE-QUAL-W2 model in general; rather it is intended to assess whether the version of this model currently parameterized for the Lake Spokane DO TMDL (i.e., the Lake Spokane WQ model) is a robust predictor of eutrophication related phenomena in this reservoir. The field input P data were collected at the Spokane River input to Lake Spokane, which occurs at river mile 151 of the Spokane River. The field DO data were collected at a sampling site located near the deepest point in Lake Spokane adjacent to Long Lake dam. This site is known as LL0 in the field sampling literature and Segment 36 in the CE-QUAL-W2 model. Welch et al. (2015) provided a map of this reservoir system.

Data quality

An unpublished memo detailed concerns related to the validity of the field sample P determinations (see Appendix 1 in the online supplemental file). During the 2001 calibration period, P samples for the Spokane River/Lake Spokane system were processed by 2 different laboratories, the Columbia Analytical Services (CAS) and the Washington State Department of Ecology Manchester Analytical Laboratory (Manchester lab). The 2 labs each processed 32 samples in parallel, and the agreement between these 2 sets of samples was only modest ($r^2 = 0.61$), with the CAS lab finding concentrations 32% higher on average than the Manchester lab. In 25 of 32 cases, the duplicate samples analyzed by the 2 labs did not meet the Department of Ecology ±15% precision criteria for field replicates for this project (Cusimano and Carroll 1999). The CAS data were also mostly reported at 10 µg/L increments: 10, 20, 30, 40, and 50 µg/L. Furthermore, the Manchester lab changed its TP analysis methodology around this time, and the method employed during 2001 gave TP results 4–20 µg/L higher on average than subsequent, more rigorous methods (Hallock 2012). The vertical TP profiles to which the Lake Spokane WQ model was fit were based on samples processed by the CAS lab, and

the model on average gave TP concentrations 3.2 µg/L (or 19%) higher than the field data. From this series of analytical and computational errors, it seems plausible that the actual TP concentrations in the reservoir during the calibration year may have been considerably less than predicted by the model.

Model assessment

Because the Lake Spokane WQ model was designed as the decision support tool for the TMDL, the main emphasis in the model calibration assessment was placed on TP and closely related constituents, such as reservoir chlorophyll (Chl) and DO concentrations. Vertical profiles for model calibration were collected on 2 dates, 9 and 30 August 2001 (Berger et al. 2009). Temperature and DO data were collected at 6 stations, samples for Chl were collected from 3 stations, and TP samples were collected from 2 stations. Model verification data were also collected in 2000, but TP data were not reported for that year, so only the calibration results for 2001 are summarized in this analysis. All of the model calibration data were extracted directly (using the software DigitizeIt 1.6.1) from the vertical profile plots reported in the original model calibration report (Berger et al. 2009).

Similar to previous studies (Reckhow et al. 1990, Schladow and Hamilton 1997, Stow et a. 2003, Arhonditsis and Brett 2004), we used multiple statistics to characterize model performance. The Quality Assurance Project Plan for the model (Wells and Berger 2009) indicated scatter plots of model versus observed data and best-fit linear regressions would be provided, but these were not provided in the model calibration report (Berger et al. 2009). For this analysis, model performance was assessed using the Nash-Sutcliffe model efficiency index, the root mean square error (RMSE), the mean error (ME), the mean absolute error (MAE), and the relative error (RE). We also calculated a nonparametric r^2 value, which is simply the r^2 of ranked observed and modeled values. The Nash–Sutcliffe model efficiency index (McCuen et al. 2006) penalizes for both lack of fit and bias. It tests the model results against a one-to-one objective and is equivalent to the r^2 value of a regression equation with a slope of one and an intercept of zero. As with a conventional r^2, higher values suggest better fit, and 1.0 is indicative of perfect fit. A value of zero indicates a model that predicts the observations as well as their

corresponding average, whereas a negative value indicates a model that predicts more poorly than the average of the observations. The RMSE, ME, MAE, and RE are all measures of the magnitude of differences between the modeled and observed values, with values close to zero indicating agreement; however, ME values close to zero could also be due to large positive and negative errors canceling each other out. Model bias is indicated by ME values substantially smaller or larger than zero. The RMSE is analogous to a standard deviation of the model error.

We also tested the sensitivity of the Lake Spokane WQ model to hypothetical reservoir input TP concentrations (TP_{IN}) across a gradient of values ranging from the future input TP goal of 10 µg/L up to 200 µg/L. In this study, TP_{IN} refers to the P concentration of the Spokane River where it enters Lake Spokane. Based on the model input files, the Spokane River accounts for ≈92% of the flow in Lake Spokane, with the Little Spokane River accounting for another ≈7%, and direct groundwater inputs contributing ≈1%. Eight water treatment facilities discharge upstream of Lake Spokane, including Coeur d'Alene WWTP, Hayden publicly owned treatment works (POTW), Post Falls sewage treatment plant (STP), Liberty Lake WWTP, Spokane City WWTP, Spokane County WWTP, Inland Empire Paper treatment plant, and Kaiser Aluminum treatment plant. The Spokane River P concentrations were varied by modifying the WWTP effluent files in the official TMDL version of CE-QUAL-W2 for 7 of the facilities. The file for the Kaiser Aluminum treatment plant was not changed because the effluents from this facility have historically low P concentrations (TP_{EFF} ≈ 25 µg/L). The 7 effluent files were modified so that their variable TP_{EFF} concentrations were 10, 35, 50, 75, 100, 250, 500, 750, 1000, 1500, 2000, 2500, 3000, and 3500 µg/L, with phosphate (PO_4) contributing 36% and P bound to organic matter (i.e., carbonaceous biochemical oxygen demand [CBODP]) contributing 64% of the TP_{EFF}. All other constituents in these files were left unchanged.

For the scenarios above, we calculated a mean summer (June–October) reservoir TP_{IN} value for the Spokane River output file CWO_151 that represented the boundary condition between the Spokane River and Lake Spokane portions of the TMDL model. The TP input concentration for the CWO_151 file was calculated as $TP_{IN} = PO_4 + \Sigma CBODP_{1\text{-}12}$, where PO_4 represents the modeled riverine phosphate

concentrations and $CBODP_{1\text{-}12}$ represents the particulate P associated with organic matter (i.e., CBOD) from each of the 12 nutrient inputs to the model (Berger et al. 2009). The minimum volume-weighted hypolimnetic DO concentration (DO_{MIN}) was then calculated for reservoir Segment 36 in the model output file SPR6 for depths below 15 m to be consistent with the observational record (Welch et al. 2015). Segment 36 represents the Lake Spokane station closest to the dam outlet and also tends to have the lowest DO in the historical record. A hypsographic curve was applied to the depth profiles of Lake Spokane to represent the volume of each depth strata. Finally, we also compared TP_{IN} to the mean observed epilimnetic (0–8 m) TP and Chl concentrations and phytoplankton biomass (i.e., diatoms, chlorophytes, and cyanobacteria) for the June to October period for Segments 9–36, which represent all of the model segments deep enough to have thermally stratified water columns in the summer.

We also carried out sensitivity analyses to elucidate how the uncertainty of the model outputs can be apportioned to the following parameters: maximum phytoplankton growth rate, phytoplankton settling velocity, organic matter decomposition, and sediment oxygen demand (SOD) on model DO_{MIN} outputs. These terms were chosen for this initial sensitivity analysis because previous research (Schladow and Hamilton 1997, Stow et al. 2003, Arhonditsis and Brett 2005) and the TMDL model itself (Berger et al. 2009) suggested they might be influential. For the sake of simplicity, we implemented a local or one-step-at-a-time (OAT) strategy in which output variations were evaluated with respect to fractional change of one input parameter while the other parameters were held constant. Specifically, these sensitivity analyses varied the parameter values up and down by a factor of ∼2 and 4 relative to the values used in the original Lake Spokane WQ model (Berger et al. 2009). These analyses were also run for 3 effluent TP_{EFF} concentrations: 50, 500, and 2000 µg/L. The varied TP_{EFF} corresponded with P input (TP_{IN}) concentrations for the Lake Spokane WQ model of 5.5, 28, and 112 µg/L, respectively. The maximum phytoplankton growth rates used in the original model averaged ≈1.6/d for the 3 phytoplankton groups. In the present study, maximum growth rates of 0.8, 1.6, 3.2, and 6.4/d were tested. The original phytoplankton settling velocities used in the model were 0.2 m/d for diatoms and green algae and −2.0 m/d for cyanobacteria (Berger et al. 2009), intended to represent a buoyant
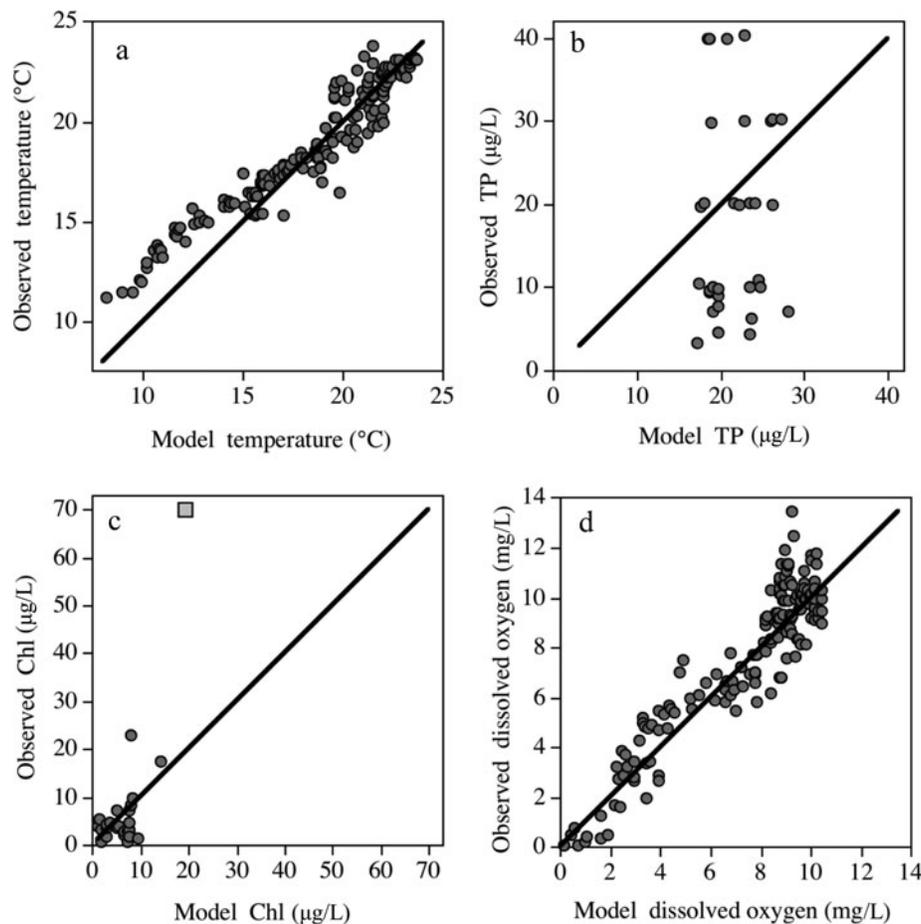
Figure 1. Bivariate plots of the model and field data for (a) temperature, (b) total phosphorus (TP), (c) chlorophyll (Chl), and (d) dissolved oxygen concentrations reported in the CE-QUAL-W2 model calibration report (Berger et al. 2009). The gray-filled square in the chlorophyll plot represents an observed surface cyanobacteria scum.

cyanobacteria scum. In the present study, uniform phytoplankton settling velocities of −2.0, 0.05, 0.1, 0.2, 0.4, and 0.7 m/d were tested. Uniform settling velocities were used to simplify the interpretation of these model sensitivity analyses. The original Lake Spokane WQ model used an average organic matter degradation rate (K_d) of ≈0.08/d for labile organic matter and 0.001/d for refractory organic matter. The present study tested K_d values of 0.005, 0.02, 0.08, 0.16, and 0.32/d for the labile pool. For this series of simulations, the K_d value for the refractory pool was always set to one-eightieth of the corresponding labile pool. The Lake Spokane WQ model used an assumed SOD of 0.25 g/m²/d of O_2, and this study tested SOD values of 0.0625, 0.125, 0.25, 0.5, and 1.0 g/m²/d.

Results

The model calibration assessment showed that the Lake Spokane WQ model adequately simulated the

temperature and DO depth profiles in Lake Spokane (Fig. 1, Table 1). Compared to a large compilation of model fits for this class of models (Arhonditsis and Brett 2004), the r^2 value for temperature was between the 30th and 40th percentiles, and the r^2 for the modeled DO data was between the 60th and 70th percentiles. However, the model outputs had a strong bias toward colder temperatures when the observed temperatures were ≤15 C (ME = −2.5 C), but this bias disappeared for observed temperatures ≥17.5 C (ME = 0.0 C; Fig. 2). The modeled DO values had a strong systematic error, with the model tending to underestimate the observed DO values in Lake Spokane by ≈2.0 mg/L when the actual DO concentrations were ≥10.5 mg/L (Fig. 2).

Compared to other studies that have attempted to model nutrient dynamics, the model calibration for TP was poor, with a TP r^2 value <10th percentile (Arhonditsis and Brett 2004; Fig. 1, Table 1). In the present case, the calculated r^2 value for TP was

Table 1. The results of the model calibration analysis based on the observations reported in Berger et al. (2009). The Nash–Sutcliffe model efficiency index (r^2), nonparametric coefficient of determination (N-P r^2), the root mean square error (RMSE), mean error (ME), mean absolute error (MAE), and relative error (RE) are reported.

	Count	Mean ± SD	Units	MEI (r^2)	N-P r^2	RMSE	ME	MAE	RE
Temperature	176	18.9 ± 3.2	C	0.82	0.92	1.3	− 0.5	1.0	0.06
Total phosphorus	32	18.2 ± 11.8	µg/L	− 0.10	0.01	12.2	3.4	10.5	0.58
Chlorophyll	26	5.2 ± 5.0	µg/L	0.14	0.04	4.6	0.7	3.4	0.66
Dissolved O$_2$	169	7.7 ± 3.1	mg/L	0.85	0.69	1.2	− 0.4	0.9	0.12

negative because the model error sum of squares was larger than the total sum of squares for the observed TP data, indicating model predictions worse than the average of the observed values. It is also noteworthy that the RMSE value (i.e., ±12.2 µg/L) was larger than the standard deviation value for the observed data from Lake Spokane (i.e., ±11.8 µg/L), and the MAE (±10.5 µg/L) is only slightly less than the SD value. In addition, the model TP data (21.6 ± 3.2 µg/L) had much less dispersion than did the observed field data (18.2 ± 11.8 µg/L; Fig 1), and the modeled TP concentrations were on average higher than the observed values (ME = 3.4 µg/L; Table 1). The observed TP data also had a multi-modal distribution, with most interpolated data clustering within a few tenths of a microgram per liter from 10, 20, 30, or 40 µg/L. This finding suggests the observed TP data were either rounded to 10 µg/L increments or the method used to analyze the field samples had a 10 µg/L detection limit. The model calibration report (Berger et al. 2009) also left out a field sample with a TP concentration of 50 µg/L and a corresponding model value of 23 µg/L.

The modeled Chl data also poorly matched the observed data (Fig. 1, Table 1). The r^2 between modeled and observed Chl data ($r^2 = 0.14$) fell between the 10th and 20th percentiles for the compilation of model fits (Arhonditsis and Brett 2004). Additionally, the observed Chl data also included one extreme outlier value (from a surface cyanobacteria scum) excluded from our model performance evaluation. This single value (70 µg/L Chl) was 3 times higher than the next highest value in the observed dataset. If this single value were included in the model evaluation, the r^2 value increased to 0.35; however, this inclusion caused the RMSE to more than double to 10.8 µg/L (which is also twice the mean Chl concentration) and the AME to increase by 50% to 5.2 µg/L.

The relationship between summer TP$_{IN}$ and DO$_{MIN}$ concentrations for the field data (Welch et al. 2015) was compared with the same values from the model for conditions ranging from the hypereutrophic era
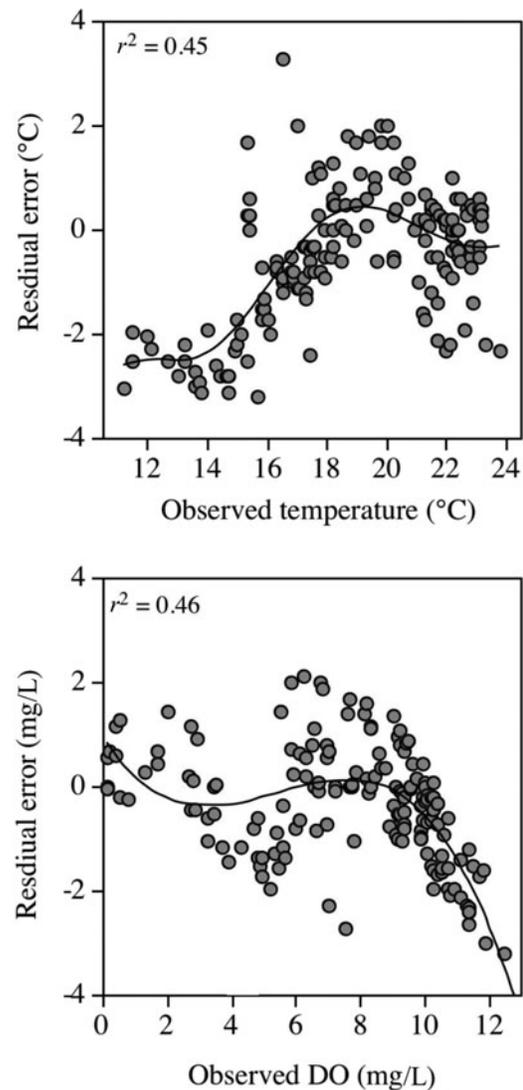

Figure 2. The residual error (model value − observed value) for the model predictions for temperature and dissolved oxygen (DO). The residual error is plotted against the observed data, and fit to a third-order polynomial regression.

to the present meso-oligotrophic state (Fig. 3). During the hypereutrophic era, TP$_{IN}$ and DO$_{MIN}$ in Lake Spokane averaged 86 ± 37 and 1.4 ± 2.7 mg/L, respectively. Currently (2010–2014), these values average 14 ± 3 and 6.5 ± 0.8 mg/L, respectively. By comparison, the Lake Spokane model's DO$_{MIN}$ response for
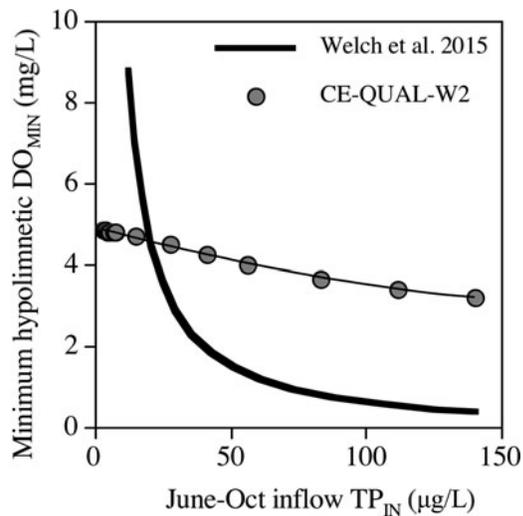
Figure 3. The modeled and observed relationship between input total phosphorus concentrations (TP$_{IN}$) and the minimum hypolimnetic dissolved oxygen concentration (DO$_{MIN}$) in Lake Spokane. The Welch et al. (2015) curve is based on field data collected over the last 40 years spanning the time from when Lake Spokane was hypereutrophic to the present. This statistical model has an r^2 of 0.87. The CE-QUAL-W2 curve represents the reservoir's predicted response to WWTP effluent TP concentrations ranging from 0 to 3500 µg/L.

similar TP$_{IN}$ concentrations averaged 3.8 ± 0.4 and 4.7 ± 0.04 mg/L, respectively, for analogous hypereutrophic and contemporary conditions.

The June to October average TP$_{IN}$ concentrations to Lake Spokane (from CWO_151) predicted by the Lake Spokane WQ model was compared with the average June to October epilimnetic (0–8 m) TP$_{EPI}$ concentration at Segments 9–36 (from SPR6). As expected, this comparison showed that when effluent TP$_{EFF}$ concentrations were at contemporary levels (≈500 µg/L) or higher, the reservoir TP$_{EPI}$ concentration was on average ≈30% lower than the river TP$_{IN}$ concentration (Fig. 4). The outcomes for effluent TP$_{EFF}$ concentrations ≤100 µg/L were unexpected, however. In those cases, the model predicted reservoir TP$_{EPI}$ concentrations were ~8 µg/L higher than the river TP$_{IN}$ concentrations. Furthermore, when all of the TP$_{EFF}$ concentrations were set to zero, the model predicted an average TP$_{EPI}$ concentration of 12 µg/L (Fig. 4).

The average summer epilimnetic TP and Chl concentrations predicted by the model were compared to a global compilation of TP versus Chl responses for real lakes (Brown et al. 2000): log(median Chl) = −0.44 + 1.10∗log(TP). The modeled Chl concentrations matched the global median values well. At a TP$_{EPI}$ concentration of 50 µg/L, CE-QUAL-W2

predicted a Chl concentration of 20 µg/L compared to the global median value of 25 µg/L. When TP$_{EPI}$ was 12 µg/L, the model and global median Chl values were 6.3 and 5.6 µg/L, respectively. We also compared the percentage cyanobacteria output by the model to the epilimnetic TP$_{EPI}$ concentration. Irrespective of TP concentrations, the model predicted the phytoplankton community was composed of 17 ± 1% diatoms, 69 ± 3% chlorophytes, and 14 ± 2% cyanobacteria. There was a slight tendency for a higher percentage of cyanobacteria at high and low P concentrations and a lower percentage of cyanobacteria at intermediate concentrations.

The sensitivity analyses showed the model was rather insensitive to the maximum phytoplankton growth rate assumed during the model run (Fig. 5). On average, the highest assumed growth rates had ≈0.5 mg/L lower DO$_{MIN}$ than the lowest assumed maximum growth rates. The model outputs were not sensitive to the assumed phytoplankton settling velocity within the range of −2.0–0.0 m/d because the phytoplankton do not settle to the hypolimnion in these cases, but the model was sensitive to this assumption within the range of 0.05–0.70 m/d. Within the higher range, larger settling velocities resulted in ≈1.4 mg/L lower DO$_{MIN}$. The model DO$_{MIN}$ outputs were sensitive to the assumed organic matter degradation rate within the range of 0.005–0.08/d and insensitive
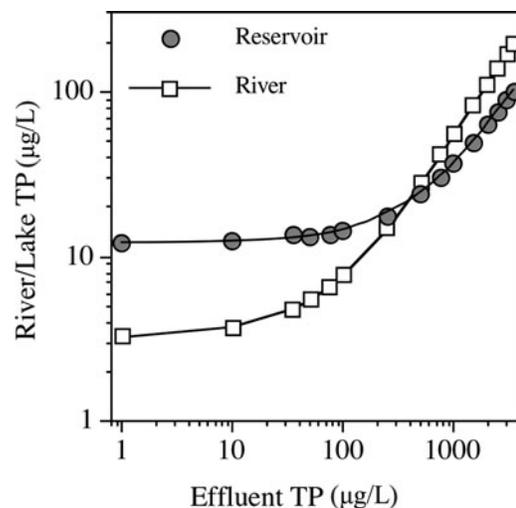

Figure 4. The relationship between the modeled Spokane River (CWO_151) and Lake Spokane epilimnetic (SPR6) TP concentrations for the June to October period. The epilimentic TP concentrations were the average 0–8 m values for model Segments 9–36. The results depicted for TP$_{EFF}$ = 1 µg/L actually represent the outcomes for the TP$_{EFF}$ = 0 µg/L simulations.
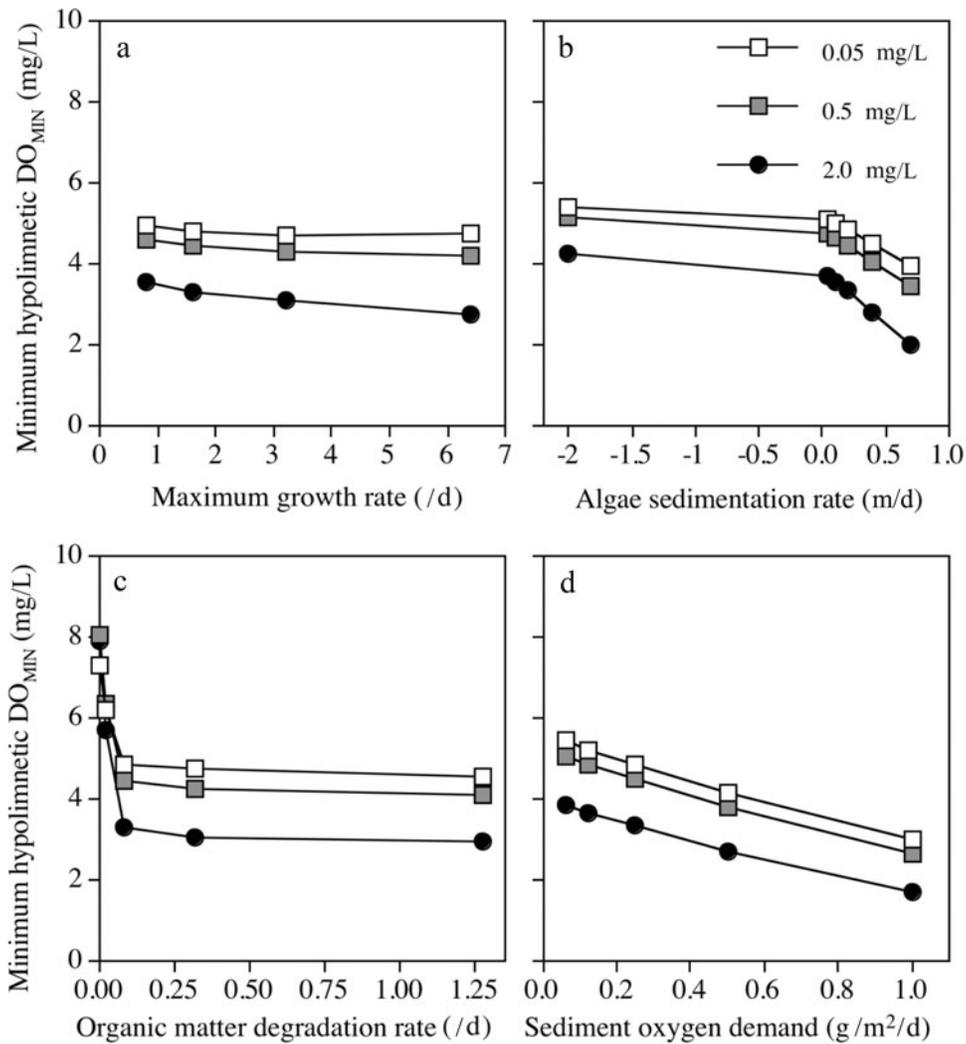
Figure 5. The results of sensitivity analyses of the influence of parameter values on model DO_{MIN} outputs for (a) the maximum phytoplankton growth rate, (b) phytoplankton settling velocity, (c) organic matter decomposition, and (d) sediment oxygen demand. The concentrations depicted in the upper right hand corner of panel b represent the 3 different effluent TP scenarios tested for the sensitivity analyses.

to this assumption in the range of 0.08–1.28/d. Overall, higher degradation rates resulted in ≈3.9 mg/L lower DO_{MIN} concentrations. The model was insensitive to this assumption at ≥0.08/d because in these cases the degradation rate was substantially higher than the rate at which organic matter is advected from the lake during the summer period (i.e., ≈0.03/d). The model DO_{MIN} outputs were also sensitive to the SOD assumption within the range of values tested (0.0625–1.0 g/m²/d of O_2), and DO_{MIN} declined by ≈2.4 mg/L when higher SOD values were assumed.

Discussion

The Lake Spokane case presents a unique opportunity to test the ability of mechanistic models to

simulate natural conditions because it has a detailed observational record of water quality conditions in response to a wide range of input TP concentrations (i.e., 12–140 µg/L). Furthermore, when mechanistic water quality models are being used to establish important public policies, it is a critical best-professional-practice to quantify the extent to which a particular model is able to represent the key attributes of the system on which management decisions are based (Stow et al. 2003). Our analysis shows the current Lake Spokane parameterization of CE-QUAL-W2 is not able to sufficiently represent the TP–Chl–DO conditions in Lake Spokane to accurately predict the effect of the TMDL. Observed and modeled TP and Chl concentrations were not at all or only weakly correlated, and the model poorly represented the sensitivity of this reservoir's hypolimnetic DO concentrations to TP inputs.

The model also has a strong bias toward mesotrophic conditions (epilimnetic TP $\geq$ 12 µg/L), even when the effluent TP_{EFF} concentrations were set to $\leq$50 µg/L and the Spokane River concentrations (TP_{IN}) were <5 µg/L. The inability to reproduce the fundamental causal relationships underlying the eutrophication problem suggests that the model fails the most critical test of its capacity to be used in the extrapolation domain (Ramin et al. 2012).

Note, however, that our study has not shown that the water quality model CE-QUAL-W2 is flawed. This study was intended to test whether this model as currently parameterized for Lake Spokane is able to represent the key biogeochemical dynamics of that reservoir. Perhaps a different parameterization of CE-QUAL-W2 would result in an improved representation of the TP_{IN}, phytoplankton biomass, and hypolimnetic DO patterns in this reservoir. Furthermore, our sensitivity analyses suggest the model outputs are highly influenced by the assumptions used for several parameters.

Model performance

The Lake Spokane WQ model is based on the well-established cause and effect relationships between P input concentrations, lake or reservoir P concentrations, algal biomass and species composition, and the hypolimnetic oxygen concentrations of stratified systems (Cornett and Rigler 1979, Chapra and Canale 1991, Molot et al. 1992, Nürnberg 2004, Rucinski et al. 2014). Despite the extensive literature documenting the basis for the TMDL (Welch and Jacoby 2004, Cooke et al. 2005), the Lake Spokane WQ model was unable to represent these processes in a realistic manner. Conversely, Lake Spokane's recovery from severe eutrophication closely matched expectations based on research conducted in other aquatic systems (Welch et al. 2015). That is, when TP_{IN} decreased from an average of 86 $\pm$ 37 µg/L (during the hypereutrophic era) to 14 $\pm$ 3 µg/L (contemporary conditions), epilimnetic Chl concentrations decreased from 21 $\pm$ 4 to 4 $\pm$ 1 µg/L, and DO_{MIN} increased from 1.4 $\pm$ 1.3 to 6.5 $\pm$ 0.8 mg/L. In summary, Welch et al. (2015) showed the real reservoir responded as expected to changes in external nutrient inputs, but the present study showed the modeled reservoir did not.

Structural error

All numerical models in earth sciences have varying degrees of structural error embedded within their design (Oreskes et al. 1994). We conducted a series of experiments to quantify the magnitude of structural errors for the TP_{IN}, TP_{EPI}, and DO_{MIN} concentrations. In these experiments, we first set the TP_{EFF} values for all of the municipal dischargers as well as Inland Empire Paper to zero. This experiment yielded TP_{IN}, TP_{EPI}, and DO_{MIN} concentrations of 3.3 µg/L, 12.2 µg/L, and 4.9 mg/L, respectively. In the next experiment, we set all sources of P to the Spokane River to zero, including the Spokane River as it leaves Lake Coeur d'Alene, Hangman Creek, Kaiser Aluminum, stormwater, combined sewer overflows, groundwater, and various smaller discharges. This experiment yielded TP_{IN}, TP_{EPI}, and DO_{MIN} concentrations of 1.1 µg/L, 8.7 µg/L, and 5.1 mg/L, respectively, indicating a cryptic source of P in the model representation of Spokane River. This source had a pronounced diurnal sinusoidal pattern, which could indicate the model included a pool of P that exchanged with the river biofilm on a daily basis. Finally, we conducted an experiment where the P concentrations in the Little Spokane River and direct groundwater inputs to Lake Spokane were also set to zero, yielding TP_{EPI} and DO_{MIN} concentrations of 3.7 µg/L, and 5.1 mg/L, respectively. This experiment indicated the Little Spokane River contributes ≈5.0 µg/L to the modeled version of Lake Spokane and that the Lake Spokane model has a cryptic P source that contributed an additional ≈2.5 µg/L to the model TP_{EPI} concentrations.

As previously noted, the Lake Spokane field data showed a 5.1 mg/L increase in the minimum hypolimnetic dissolved DO concentration when this system shifted from hypereutrophy to meso-oligotrophy (Welch et al. 2015), whereas for the same TP_{IN} levels the model representation of Lake Spokane showed only a 1.0 mg/L improvement in DO_{MIN} (Fig. 3). When all P inputs to Lake Spokane were set equal to zero, the volume-weighted temperature <15 m and DO_{MIN} averaged 13.7 C and 5.1 mg/L, respectively. Freshwater at 13.7 C has a saturated oxygen concentration of 10.4 mg/L, which indicates the Lake Spokane model has a structural error of 5.3 mg/L relative to its sensitivity to P inputs. Additionally, these results indicate that according to the Lake Spokane WQ model, the TMDL goal of DO_{MIN} >9.3 mg/L cannot be met regardless of P inputs. This limitation is a critical flaw in the Lake Spokane model, especially because this model is the central analytical tool for the Spokane River basin TMDL. Furthermore, when the effluent

TP_{EFF} concentrations were set to zero, the model gave a mean epilimnetic TP_{EPI} concentration of 12 µg/L, a problematic outcome because the TMDL goal for Lake Spokane is TP_{IN} <10 µg/L. Conversely, the trends from Lake Spokane itself indicated continued reductions in TP_{IN} will likely lead to further improvements in hypolimnetic DO, although at the current low input concentrations the minimum hypolimnetic DO concentrations in Lake Spokane may be equally sensitive to water residence time (Welch et al. 2015).

Paradoxically, the 2001 calibration data for the Lake Spokane WQ model suggest it adequately simulated August vertical DO profiles (Fig. 1), whereas the TP_{IN} versus DO_{MIN} analysis showed this model poorly characterized Lake Spokane's sensitivity to TP_{IN} inputs (Fig. 3). These contradictory results and the outcomes of our sensitivity analyses suggest that the original DO calibration may have been achieved by adjusting the particle settling velocity and organic matter degradation submodels within CE-QUAL-W2 until the observed and modeled hypolimnetic DO concentrations matched. That is, the calibration process achieved the correct answer for the wrong reasons. Alternatively, we suggest the calibration process should attempt to match the observed DO profiles via the mechanistic basis of the TMDL (i.e., DO dependence on TP_{IN}), as the observed response showed. During calibration, equal effort should be devoted to matching the observed trends for the lake's TP, Chl, and DO concentrations.

In contrast to field studies showing that the percent cyanobacteria increases dramatically with average TP concentrations (Downing et al. 2001), the Lake Spokane model showed virtually no phytoplankton community composition dependency on TP_{EPI}. However, this model outcome and its relationship to Lake Spokane is further complicated because whereas Lake Spokane often has ephemeral cyanobacteria blooms (Welch et al. 2015), Lake Spokane was rarely dominated by cyanobacteria, even when hypereutrophic (Soltero and Nichols 1984). Based on the characterization postulated by the calibration vector, cyanobacteria can gain a competitive advantage because of their higher maximum growth rates, higher temperature optima for growth, neutral or even negative settling rates (buoyancy regulation capacity), and higher light saturation intensity at the maximum photosynthetic rate (Berger et al. 2009). Further, a *post hoc* adjustment assigned a lower P stoichiometry to cyanobacteria, but the

half-saturation constant for P is set to 3 µg/L for all 3 algal groups considered. Simply put, the model not only treats cyanobacteria as (practically) equal competitors for P uptake, but also the general designation of phytoplankton minimizes the likelihood of P limitation in the system.

Sensitivity analyses

Our sensitivity analyses for DO_{MIN} outputs show the model is surprisingly insensitive to the assumed maximum phytoplankton growth rate, sensitive to the assumed phytoplankton settling velocity, and sensitive to the organic matter decomposition rate and SOD. According to the Lake Spokane WQ model, a range of TP_{IN} values of 14–86 µg/L would give DO_{MIN} values that differed by 1.1 mg/L. By comparison, the observed sensitivity for the maximum phytoplankton growth rate was 0.5 mg/L, and the sensitivity for the phytoplankton settling velocity was 1.4 mg/L. The results for SOD and the organic matter decomposition rate (sensitivity = 2.3 and 3.9 mg/L, respectively) showed these submodels potentially have a much greater influence on model DO_{MIN} concentrations than do the modeled P inputs. This finding is important to consider during an eventual model recalibration. Also note that the local sensitivity analysis approach presented here may not always be able to fully elucidate which parameters are more influential because the individual effects of a particular parameter are also conditional on the parameter values assigned to additional elements of the calibration vector or other assumptions made regarding the model forcing functions (Arhonditsis and Brett 2005). Even if the parameters are independent of each other, the results may still be different due to nonlinearities in model behavior, which cannot necessarily be captured by single-parameter perturbations. Thus, our proposition in the next iteration of the model is to consider global sensitivity analysis methods that can more effectively scrutinize the parameter space and obtain a ranking of input parameters in terms of their importance, while accommodating any issues of nonlinearity and/or interactions with other model inputs.

Recommendations

The Lake Spokane WQ model should be recalibrated to a more appropriate dataset. Using only 2 sampling dates in close temporal proximity is far below the

norm for this type of model (Arhonditsis and Brett 2004). In a similar hypolimnetic hypoxia P loading response model, Rucinski et al. (2014) used 10+ vertical profiles per year over 19 separate years to calibrate their model. The new model calibration dataset should include at least monthly observations from at least 5 stations spanning at least 6 months preceding and during the stratified period when hypolimnetic hypoxia is of greatest concern. In this type of model, the SOD submodel can account for a large fraction of the overall hypolimnetic oxygen demand (Rucinski et al. 2014). SOD should also be dynamically coupled to P inputs so that higher nutrient inputs and phytoplankton production are associated with greater SOD and vice versa (Rucinski et al. 2014).

The model performance goals should be specified before initiating the model calibration process and should effectively emerge as synthesis of the following, sometimes conflicting, factors (Arhonditsis and Brett 2004, NRC 2007, EPA 2009): (1) the performance norms as reported in the modeling literature and the challenge to achieve balanced error among multiple model endpoints; (2) the credibility of the mechanistic understanding of the modeled system, including the available datasets; (3) the uncertainty characterizing the behavior of any natural system modulated by an inconceivably wide array of "ecological unknowns"; and (4) the management objectives being evaluated.

Considering these issues, we suggest goals of $r^2 >$ 0.5 against the Nash–Sutcliffe model efficiency index (McCuen et al. 2006) for both TP and Chl for normally distributed datasets. We also recommend an RMSE goal of <3 µg/L for TP and <2 µg/L for Chl. We further recommend r^2 values of >0.9 for temperature and >0.8 for DO, with RMSEs of <1 C and <1 mg/L, respectively. The parameter values used to calibrate the model should be based on representative literature values and not just legacy or "typical" values, as is currently the case (Berger et al. 2009). Recognizing that even the most articulate models are associated with some degree of uncertainty and error, we also advocate for the adoption of a probabilistic approach to TMDL assessment (Borsuk et al. 2002, Arhonditsis et al. 2011). Probabilistic strategies offer percentile-based predictions that explicitly incorporate residual variability into assessments, leading to more appropriate evaluation of the frequency of water quality standard violations. A rigorous assessment of model uncertainty provides decision-makers and stakeholders with a tangible measure of the degree of confidence in model results and provides an explicit basis for the choice of a margin of safety in setting a TMDL (Arhonditsis et al. 2011).

Conclusions

Lake Spokane was previously hypereutrophic with a strongly hypoxic hypolimnion due to P discharges from several WWTPs in the Spokane River basin. The Lake Spokane and Spokane River parameterization of the water quality model CE-QUAL-W2 is the sole decision support tool for the DO TMDL (Moore and Ross 2010), projected to cost several hundred million dollars in infrastructure upgrades for this region. Because of the unique long-term dataset that exists for this system, this case study is an excellent opportunity to develop and implement the most scientific and credible approach for eutrophication management. For example, the model should be recalibrated to a much more rigorous field dataset; model performance goals should be determined prior to calibration; and finally, the expected beneficial outcomes, if the TMDL goals are achieved, should be specified, and these benefits should be weighed against the greatly increased energy consumption and greenhouse gas emissions associated with advanced nutrient removal processes (Falk et al. 2013).

The Lake Spokane WQ model was poorly calibrated to the field TP and Chl data, and the quality of the field P data used in model calibration was problematic. The model's DO_{MIN} response to a wide range of TP_{IN} concentrations indicated only small differences in hypereutrophic and contemporary DO concentrations. The model also indicated that no WWTP phosphorus removal scenario will meet the 9.3 mg/L hypolimnetic DO goal for Lake Spokane. By contrast, the observational record for the lake showed a sharp functional response between TP_{IN} and DO_{MIN}, indicating that further reductions in TP_{IN} inputs will likely result in improved hypolimnetic DO. Recent evidence suggests, however, that Lake Spokane DO concentrations may now be equally sensitive to water residence time (Welch et al. 2015). The Lake Spokane case study is unusual because a detailed observational record of this system's deterioration and recovery from severe eutrophication exists. We suggest that the reservoir's actual response to TP inputs be given more credence in the TMDL process than the model representation of this system.

Acknowledgments

References

Arhonditsis GB, Brett MT. 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. Where are we? Mar Ecol Prog Ser. 271:13–26.

Arhonditsis GB, Brett MT. 2005. Eutrophication model for Lake Washington (USA): Part I. Model description and sensitivity analysis. Ecol Model. 187:140–178.

Arhonditsis GB, Stremilov S, Gudimov A, Ramin M, Zhang W. 2011. Integration of Bayesian inference techniques with mathematical modeling. In: Wolanski E, McLusky DS, editors. Treatise on estuarine and coastal science, Vol 9. Waltham (MA): Academic Press. p. 173–192.

Berger C, Wells S, Wells V. 2009. Spokane river modeling report 2009, model development and calibration check. Portland State University, Technical Report EWR-03-09.

Beven K. 2006. A manifesto for the equifinality thesis. J Hydrol. 320:18–36.

Borsuk ME, Stow CA, Reckhow KH. 2002. Predicting the frequency of water quality standard violations: a probabilistic approach for TMDL development. Environ Sci Technol. 36:2109–2115.

Brown CD, Hoyer MV, Bachmann RW, Canfield DE Jr. 2000. Nutrient–chlorophyll relationships: an evaluation of empirical nutrient–chlorophyll models using Florida and north-temperate lake data. Can J Fish Aquat Sci. 57:1574–1583.

Chapra SC, Canale RP. 1991. Long-term phenomenological model of phosphorus and oxygen for stratified lakes. Wat Res. 25:707–715.

Cole TM, Wells SA. 2003. CE-QUAL-W2: a two-dimensional, laterally averaged, hydrodynamic and water quality model. Version 3.1. Vicksburg (MS): US Army Engineering and Research Development Center, Instruction Report EL-03-1.

Cooke GD, Welch EB, Peterson S, Nichols SA. 2005. Restoration and management of lakes and reservoirs. Boca Raton (FL): CRC press.

Cornett RJ, Rigler FH. 1979. Hypolinimetic oxygen deficits: their prediction and interpretation. Science. 205:580–581.

Cusimano B, Carroll J. 1999. Quality Assurance Project Plan: Spokane River and Long Lake TMDL study for biochemical oxygen demand and update of the phosphorus attenuation model. Washington State Department of Ecology, Publication No. 99-03-101.

Downing JA, Watson SB, McCauley E. 2001. Predicting cyanobacteria dominance in lakes. Can J Fish Aquat Sci. 58:1905–1908.

[EPA] Environmental Protection Agency. 2009. Guidance on the development, evaluation, and application of environmental models. US Environmental Protection Agency, Council for regulatory environmental modeling. EPA/100/K-09/003.

Falk MW, Reardon DJ, Neethling JB, Clark DL, Pramanik A. 2013. Striking the balance between nutrient removal, greenhouse gas emissions, receiving water quality, and costs. Water Environ Res. 85:2307–2316.

Hallock D. 2012. River and stream water quality monitoring report water source 2011. Washington State Department of Ecology, Publication No. 12-03-039.

McCuen RH, Knight Z, Cutter AG. 2006. Evaluation of the Nash–Sutcliffe efficiency index. J Hydrol Eng. 11:597–602.

Molot LA, Dillon PJ, Clark BJ, Neary BP. 1992. Predicting end-of-summer oxygen profiles in stratified lakes. Can J Fish Aquat Sci. 49:2363–2372.

Moore DJ, Ross JD. 2010. Spokane River and Lake Spokane dissolved oxygen total maximum daily load: water quality improvement report. Department of Ecology, State of Washington.

[NRC] National Research Council. 2007. Models in environmental regulatory decision making. Committee on models in the regulatory decision process. National Research Council. ISBN: 978-0-309-10999-4.

Nürnberg GK. 2004. Quantified hypoxia and anoxia in lakes and reservoirs. Sci World J. 4:42–54.

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

Ramin M, Labencki T, Boyd D, Trolle D, Arhonditsis GB. 2012. A Bayesian synthesis of predictions from different models for setting water quality criteria. Ecol Model. 242:127–145.

Reckhow KH, Clements JT, Dodd RC. 1990. Statistical evaluation of mechanistic water-quality models. J Env Eng. 116:250–268.

Rucinski DK, DePinto JV, Scavia D, Beletsky D. 2014. Modeling Lake Erie's hypoxia response to nutrient loads and physical variability. J Great Lakes Res. 40:151–161.

Rykiel EJ. 1996. Testing ecological models: the meaning of validation. Ecol Model. 90:229–244.

Schladow SG, Hamilton DP. 1997. Prediction of water quality in lakes and reservoirs: Part II - Model calibration, sensitivity analysis and application. Ecol Model. 96:111–123.

Soltero RA, Gasperin AF, Graham WG. 1974. Chemical and physical characteristics of a eutrophic reservoir and its tributaries - Long Lake, Washington. Water Res. 8:419–431.

Soltero RA, Nichols DG. 1984. The improved water quality of Long Lake following advanced wastewater treatment by the City of Spokane, Washington. Lake Reserv Manage. 1:395–404.

Stow CA, Roessler C, Borsuk ME, Bowen JD, Reckhow KH. 2003. Comparison of estuarine water quality models for total maximum daily load development in Neuse River Estuary. J Water Res Pl-ASCE. 129:307–314.

Welch EB, Brattebo SK, Gibbons HL, Plotnikoff RW. 2015. A dramatic recovery of Lake Spokane water quality following wastewater phosphorus reduction. Lake Reserv Manage. 31:157–165.

Welch EB, Jacoby JM. 2004. Pollutant effects in freshwater: applied limnology. Spon Press.

Wellen C, Kamran-Disfani A-R, Arhonditsis GB. 2015. Evaluation of the current state of distributed nutrient watershed-water quality modeling. Environ Sci Technol. 49:3278–3290.

Wells SA, Berger C. 2009. Spokane River in Idaho and Washington TMDL water quality and hydrodynamic modeling Quality Assurance Project Plan - DRAFT. Portland State University, Department of Civil and Environmental Engineering.