Calibration and Validation of an Empirical Dissolved Oxygen Model

Omar I. Abdul-Aziz1; Bruce N. Wilson2; and John S. Gulliver3

Abstract: A common difficulty in stream health assessments is the scarcity of real-time dissolved oxygen (DO) data. Discrete DO measurements, collected at times often imposed by sampling constraints, are difficult to use in assessments because of diurnal variations. An empirical model is developed here to adjust these discrete measurements to a common time-reference value using an extended stochastic harmonic analysis (ESHA) algorithm, which was originally formulated with a fraction of DO saturation model by the authors. The model was calibrated and validated for different stream sites across Minnesota, incorporating effects of different ecoregions and variable drainage areas. Data were normalized to increase the general applicability of the fitted parameters. Model calibration for five long record stations accurately represented observed diurnal variations in DO. The root-mean-square error (RMSE) for predicting hourly DO ranged from 0.53 to 0.80 mg/L and for predicting DO at a standard time ranged from 0.44 to 0.91 mg/L. Estimated model parameters were robust in terms of both spatial and temporal variations. Analytical as well as numerical analyses of parameter uncertainties were performed using sensitivity coefficients. Model validation with independent data for eight different Minnesota streams was performed using three different approaches for estimating parameters. The best approach considered both ecoregional location and watershed size to select representative model parameters. The RMSE for predicting hourly DO and standard DO respectively ranged from 0.53 to 1.65 mg/L and 0.00 to 1.83 mg/L. The developed model is a useful tool for total maximum daily load assessment of aquatic ecosystem health across a range of temporal and spatial scales. It is more elegant and simpler than the application of the ESHA algorithm for the fraction of DO saturation model.

DOI: 10.1061/ASCE0733-9372(2007)133:7(698)

CE Database subject headings: Dissolved oxygen; Empirical models; Algorithms; Fourier series; Sensitivity analysis; Validation; Calibration.

Introduction

Dissolved oxygen (DO) is an important indicator for the general health of an aquatic ecosystem. Different processes (e.g., atmospheric diffusion and reaeration, photosynthesis and respiration, direct input from incoming tributaries or effluents, organic decomposition, and sediment oxygen demand, etc.) may result in a substantial diurnal variation in DO. Sampling constraints frequently limit the collection of discrete DO measurements to a single sample collected at almost any time during the daylight hours. To compare DO trends among different sites and/or different days, standard procedures are required to convert observations at different times to those corresponding to a reference or standard time.

Many studies have focused on DO processes and prediction in streams. Dyar and Alhadeff (2005) is an example of using a harmonic curve-fitting procedure to develop a statewide DO model by analyzing DO data for Georgia streams. By using Fourier transforms, Gallegos et al. (1977) demonstrated a method for calculating short-term variations in oxygen exchange based on frequent oxygen measurements. O’Connor and Di Toro (1970) analyzed DO balance in a stream by considering time-varying photosynthetic oxygen sources. Using power spectrum analyses and other time-series techniques, Thomann (1970) investigated waste treatment plant performance while indicating possible DO variations in rivers. Van Orden and Uchrin (1993) used harmonic analysis to simulate the variability of DO deficit within a stream. Other studies (Piasecki 2004; Adrian and Alshawabkeh 1997; Rounds 2002; Erdmann 1979a,b; Gelda et al. 2001; Atkinson et al. 1995; Butcher and Covington 1995) represented DO using methods that do not give particular reference to harmonic analysis. The diel-oxygen change, whole-stream and benthic-chambers, and open-channel methods are frequently used for investigating ecosystem metabolism in ecological research (McTammany et al. 2003; Mulholland et al. 2001; Young and Huryn 1996, 1999; Fellows et al. 2001). Interested readers are referred to Abdul-Aziz et al. (2006) for a more detailed review of recent literature on the measurement and modeling of DO.

Abdul-Aziz et al. (2006) extended the theory of classical harmonic analysis to formulate an extended stochastic harmonic analysis (ESHA) by incorporating a constraint that forces the Fourier series through a specified value. The ESHA was used to evaluate the diurnal variation in the fraction of DO saturation.
Although the algorithm was evaluated for different streams in Minnesota, no validation was performed with an independent data set therein. As the fraction of DO saturation was used, the model also requires both DO and water temperature data for its application. Diurnal temperature data are likely not available at many stream sites, which generally increases the difficulty in using the algorithm by Abdul-Aziz et al. (2006).

The primary objective of this paper is to formulate an ESHA algorithm to develop a direct empirical model that only requires DO data as inputs. The second objective is to calibrate the DO model for different stream locations in Minnesota. Validation of the model is done with independent data for other sites in Minnesota as the third objective. Finally, the paper ends with a discussion of the results, conclusions, and recommendations.

**Model Development**

The theory of the ESHA is described in details by Abdul-Aziz et al. (2006). This theory is used in this section to develop an empirical DO model. An important goal of such model is to convert measured DO at any time of the 24-h day to DO at a flexibly defined standard time.

The primary input requirement is real-time data for DO, termed as DO_{obs}. One hour data resolution was found to be adequate for the data sets of this study. DO_{obs} refers to the cross-sectional average concentration that reflects the zero physical dimensionality of the model under development. This is generally a reasonable assumption for streams and rivers, whereas possibly being less valid for some lakes.

Observed DO at a standard time (t_s) is defined as DO_s = DO_{obs}(t_s), where DO_s is labeled as the standard DO. Any hour (integer or fractional, depending on input data resolution) instant within the diurnal time domain may be taken as the value of t_s (Abdul-Aziz et al. 2006). The observed DO data is normalized by DO_s to improve the generality of fitted parameters. The normalized variable DO_{obs}^s is parameterized by taking time t as an independent variable and by applying the ESHA. The development steps are summarized as

1. Collect real-time data for DO_{obs}(t);
2. Choose a standard time, t_s;
3. Define DO_s = DO_{obs}(t_s);
4. Derive a normalized time series as DO_{obs}^s(t) = DO_{obs}(t)/DO_s, and
5. Parameterize DO_{obs}^s(t) by ESHA (Abdul-Aziz et al. 2006). As apparent from Step 4, a quantity DO_{s} is defined as DO_{s} = DO_{obs}^s(t=t_s)=1. This is used later as the special constraint of least-squares optimization in parameterization of DO_{obs}.

**Parameter Estimation**

Denoting the periodic dependent variable DO_{obs}^s by y and the independent variable time by t, a stochastic Fourier series y(t) is defined as (Priestly 1981)

\[
y(t) = a_0 + \sum_{k=1}^{W} \left[ a_k \cos(2\pi f_k t) + b_k \sin(2\pi f_k t) \right] + e(t)
\]

(1a)

where \( k \) is harmonic number; \( a_k \) and \( b_k \) are Fourier coefficients; \( a_0 \) is constant term for \( k=0 \); \( f_k = k/t' \) is appropriate discrete frequencies; \( n \) is total number of observations within one processing period \( t' \); \( \Delta t \) is sampling interval; \( W \) is appropriate maximum number of harmonics; and \( e(t) \) is zero mean random error sequence. An alternative and widely used phase-angle form of the Fourier series is

\[
y(t) = a_0 + \sum_{k=1}^{W} \alpha_k \cos(2\pi f_k t - \phi_k) + e(t)
\]

(1b)

\[
\alpha_k = \sqrt{a_k^2 + b_k^2}
\]

(1c)

\[
\phi_k = \arccos \left( \frac{a_k}{\sqrt{a_k^2 + b_k^2}} \right)
\]

(1d)

where \( \alpha_k \) and \( \phi_k \) are amplitudes and phase angles of the cosine only series, respectively. Selection of the appropriate maximum frequencies on the basis of sampling interval and number of observations is discussed by Priestley (1981).

A least-squares estimation procedure is applied to determine the estimated Fourier coefficients \( \hat{a}_k \) and \( \hat{b}_k \). The objective function, \( M \), is defined such that the sum of the squared errors (SSE) between the harmonic process \( h \) and observed \( y \) is minimized. The functional form of \( M \) is

\[
M = \text{SSE} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - h_i)^2
\]

(2)

Following classical harmonic analysis, \( \hat{a}_0, \hat{a}_j, \) and \( \hat{b}_j \) are obtained by setting the partial derivatives of \( M \) with respect to each of \( a_0, a_j, \) and \( b_j \) to zero. In the ESHA, however, the harmonic process \( h \) is required to pass through a known point, say, \( h(t_j) = \gamma_k \). Hence, following Eq. (1a), the term \( a_0 \) is defined as

\[
a_0 = \kappa - \sum_{k=1}^{W} [a_k \cos(2\pi f_k t_j) + b_k \sin(2\pi f_k t_j)]
\]

(3)

By applying appropriate mathematical manipulations, Abdul-Aziz et al. (2006) developed a system of linear equations and provided the solution for parameters \( \hat{a}_j \) and \( \hat{b}_j \) for \( j \neq 0 \) vector of length \( 2W \); and \( R \) vector of length \( 2W \) consisting of the terms associated with the observation \( y_j \) and fixed point \( h(t_j) = \gamma_k \). The parameter \( a_0 \) is calculated following Eq. (3) and the parameters \( \hat{a}_k \) and \( \phi_k \) are obtained from Eqs. (1c) and (1d).

**Application Methodology**

The model \( DO_{obs}^s \), namely \( DO_{obs}^s \), is used to convert measured DO to that at a standard time. Assume that at a stream location, a discrete DO, \( DO_{obs} \), is measured at any time \( t_j \) of a 24-h diurnal cycle (typically, within the 12-h daylight time). The steps of converting \( DO_{obs}(t) \) to \( DO_s = DO_{obs}(t) \) may be summarized as

1. Model, \( DO_{obs}^s(t) \) is obtained for \( DO_{obs}^s(t) = 1 = \kappa \) following parameterization with the ESHA of the previous section;
2. \( DO_{obs}(t) \) is known from discrete measurement;
3. Model, \( DO_s \) is estimated as \( DO_s = DO_{obs}(t) / DO_{obs}^s(t) \).
Study Sites

The state of Minnesota was chosen for calibration and validation of the DO model. Soils, geology, land use, and vegetation vary widely across Minnesota. Areas of relative homogeneity in land use, soil, topography, and natural vegetation are defined as ecoregions by the U.S. Environmental Protection Agency (USEPA 2006). Level-III classification divided Minnesota into seven major ecoregions. They are the Northern Minnesota Wetlands (NMW), Northern Lakes and Forests (NLF), North Central Hardwood Forests (NCHF), Northern Glaciated Plains (NGP), Western Corn Belt Plains (WCP), Red River Valley (RRV), and Driftless Area (DA). Readers are referred to Omernik (1987) for more details. Ecoregions provide a good framework for evaluating similarities and differences in Minnesota streams, particularly for identifying sites with different water quality characteristics.

Calibration-site selection was based on the availability of DO data to incorporate the seasonal characteristics of ESHA for a time period that included observation in May through August. Five stations were selected for model calibration purpose, representing mainly the four ecoregions of NLF, NCHF, WCP, and DA. For validation of the model, available independent data sets for two different stations from each of the four ecoregions of RRV, NLF, NCHF, and WCP were selected. Data were collected through personal communications from the United States Geological Survey (USGS), Minnesota Department of Agriculture (MDA), and Minnesota Pollution Control Agency (MPCA).

The geographic locations of seven ecoregions, as well as the selected five calibration and eight validation stations are shown in Fig. 1. Table 1 summarizes key characteristics and given identification numbers of the sites. One calibration station (ID 03) and four validation stations (IDs 002, 003, 007, and 008) represent relatively larger watersheds (drainage area >10,000 km²), four other calibration stations (IDs 01, 02, 04, and 05) represent...
smaller watersheds (drainage area \(\leq 1.000 \text{ km}^2\)), and all the remaining four validation stations (IDs 001, 004, 005, and 006) refer to the medium watersheds (1000 \(\text{km}^2 < \text{drainage area} \leq 10,000 \text{ km}^2\)). The calibration sites had full or significant partial records of DO for the May through August study period in the years from 2000 to 2005. The validation stations had around three to six days of data record for August, 2000. A day was defined by taking 24 consecutive hourly data values between 1 a.m. and midnight for relevant analyses.

Diurnal data sets were screened for missing data, for unexpectedly large changes in DO, and for unusually small or large observations likely caused by instrumentation or data recording errors. The filters are similar to those used by Abdul-Aziz et al. (2006). Data sets with missing data for more than three consecutive hours in a day were excluded from the analysis. A diurnal data set was also excluded if DO exceeded a rate of change of 2 mg/L per hour and/or if DO values were less than 1 mg/L or greater than 25 mg/L.

**Model Calibration**

**Criteria for Optimal Parameters**

Selection of a maximum number of harmonics is an important criterion for obtaining the optimal parameter estimates from application of the ESHA algorithm. Use of the Akaike information criterion (AIC) (Priestley 1981) and the sum of square errors in DO [SSE_DO, see Eq. (5b)] suggested an optimal number of harmonics \(W_{\text{opt}}\) of 2 to maintain both accuracy and simplicity in obtaining a parsimonious set of parameters. Flexibility of standard time \(t_r\) was also tested. The model was not highly sensitive to \(t_r\), whereas the time slot of 10–13 h (i.e., from 10 a.m. to 1 p.m., inclusive) seemed to be the preferable domain for \(t_r\) selection. As such, a \(t_r\) of 12 h (i.e., 12 p.m.) that represents the midpoint of a diurnal cycle was selected here for the study sites.

**Results**

Fourier coefficients \((\hat{a}_0, \hat{a}_k, \text{and } \hat{b}_k; k=1,2)\) for each day of the screened data sets were obtained using Eqs. (3) and (4). They were converted to daily estimates of amplitudes \((\hat{a}_k \text{ and } \hat{a}_k)\) and phase angles \((\hat{\phi}_1 \text{ and } \hat{\phi}_2)\) using Eqs. (1c) and (1d). The ensemble mean of estimated parameters and associated standard deviations over the study period of May–August are summarized in Table 2. The constant term \(\hat{a}_0\) revealed values from 0.9122 to 1.0283 with standard deviations of 0.0378–0.0658. First harmonic parameters take values from 0.0567 to 0.2155 with standard deviations of 0.0354–0.1027 for \(\hat{a}_1\) and 60.33–139.10° with standard deviations of 15.48–29.95° for \(\hat{\phi}_1\). Among the second harmonic parameters, \(\hat{a}_2\) ranged from 0.0249 to 0.0496 and \(\hat{\phi}_2\) ranged from 86.34 to 118.40°. The associated ranges of standard deviations were 0.0155–0.0347 and 29.06–44.33° for \(\hat{a}_2\) and \(\hat{\phi}_2\), respectively.

The equivalent ensemble-mean estimates of amplitudes \((\hat{a}_1^e \text{ and } \hat{a}_2^e)\) and phase angles \((\hat{\phi}_1^e \text{ and } \hat{\phi}_2^e)\) were obtained by direct conversion of the ensemble-mean estimates of Fourier coefficients (i.e., \(\hat{a}_k \text{ and } \hat{b}_k; k=1,2\)) using Eqs. (1c) and (1d). As shown in Table 2, equivalent estimates ranged from 0.0446 to 0.2050 for \(\hat{a}_1^e\), from 0.0136 to 0.0403 for \(\hat{a}_2^e\), from 65.89 to 140.58° for \(\hat{\phi}_1^e\), and from 86.08 to 137.97° for \(\hat{\phi}_2^e\). For \(k=1,\ldots,W\), note that \(\hat{a}_k \text{ and } \hat{b}_k\) are uncorrelated, whereas \(\hat{a}_k \text{ and } \hat{\phi}_k\) are interrelated by the Eqs. (1c) and (1d). In essence, statistics of daily estimates of amplitudes and phase angles may indicate some of the temporal as well as biogeochemical characteristics of estimated parameters. However, \(\hat{a}_1^e \text{ and } \hat{\phi}_1^e\) (or ensemble mean of \(\hat{a}_k \text{ and } \hat{\phi}_k\) for \(k=1,\ldots,W\)) should be used to reconstruct the actual Fourier signals.

The mean errors in regular DO, namely \(\text{ME}_{\text{DO}}\), were calculated by taking ensemble average of the errors in modeled DO (i.e., Error_{\text{DO}}) over the study period, where Error_{\text{DO}} is the difference between modeled and observed values. Table 2 revealed approximately zero values of \(\text{ME}_{\text{DO}}\) at all stations. The root-mean-square errors (RMSE) of modeled and observed values are
also shown in Table 2. RMSE$_{DO}$ was estimated by taking the square-root of sum of squared errors in DO (SSE$_{DO}$), where SSE$_{DO}$ was defined as

$$\text{DO}(t) = \text{DO}_{\text{obs}}(t)$$

where $\text{DO}_i$ is the ith observation of DO in the jth day; $\text{DO}_{ij}$ is corresponding model; $n$ = total number of observations within one-process period of $t'$ (24 h here); $\eta$ = total number of days after data screening; and $\text{DO}_{ij}^* = \text{DO}_{ij}$ obtained by using ensemble-mean estimates of parameters $\tilde{a}_k$ and $\tilde{b}_k$ or equivalent ensemble-mean estimates of amplitudes (i.e., $\tilde{a}_k^*$) and phase angles (i.e., $\tilde{\phi}_k$), where $k = 1, \ldots, W$. RMSE$_{DO}$ ranged from 0.48 to 0.80 mg/L among different stations. RMSE$_{DO}$, the RMSE in predicting standard DO (i.e., DO$_s$), were estimated by taking the square root of average SSE between observed and predicted DO$_s$. $\tilde{a}_k$ and $\tilde{b}_k$ were estimated for discrete data at 8 a.m. and 4 p.m., respectively. As shown in Table 2, $\tilde{a}_k$ ranged from 0.44 to 0.91 mg/L and $\tilde{b}_k$ ranged from 0.44 to 0.91 mg/L. The relatively small values of RMSE$_{DO}$ and RMSE$_{DO}$ within and across ecoregions demonstrate the good performance of the ESHA algorithm to predict observed values.

Fig. 2 reveals the daily variations of estimated errors in modeled DO (i.e., Error$_{DO}$) at Whitewater River site using box plots. Each box has solid lines at the lower quartile, median, and upper quartile values. Dashed lines are the whiskers extending from each end of the box and correspond to errors within 1.5 × IRQ, where IRQ = interquartile range. Positive (+) signs present individual outliers beyond the ends of the whiskers. As required by the constraint for ESHA, the Error$_{DO}$ at $t_s$ = 12 p.m. were zero and the magnitude of errors generally increased with distance from $t_s$. The slightly larger error-bars during the time span from 2 to 6 p.m. may be attributed to the more variable biochemical activities (e.g., differential photosynthesis and respiration rates) in the afternoon, leading to higher variances in observed DO data.

Temporal patterns of the estimated parameters were investigated by using scatter-plot analyses for the study period of May through August of the associated calendar years at different stations. As an example plot, Fig. 3 shows the variation of normalized $\tilde{a}_0$, $\tilde{a}_0/E(\tilde{a}_0)$, with normalized Julian days (Jday/TDY). Normalization was done in both axes in order to obtain appropriate scaling for different stations presented in the plot. $E(\tilde{a}_0)$ refers to the ensemble-mean estimates for daily values of $\tilde{a}_0$. Jday refers to the Julian days (or calendar days), and TDY refers to the total number of days in a year. Apart from no prevalent temporal trends apparent for $\tilde{a}_0$ in any station, relatively small variability in $\tilde{a}_0/E(\tilde{a}_0)$ was noticed for stations across ecoregions. However, the scatter-plot analyses revealed slight increasing trends in $\tilde{a}_1$ and $\tilde{a}_2$ and no notable trends in $\tilde{\phi}_1$ and $\tilde{\phi}_2$. These results are similar to those obtained by Abdul-Aziz et al. (2006).

Estimated model parameters were also investigated for possible trends with atmospheric solar radiation (SR) and flow rate ($Q$) using scatter-plot analyses. Because of the relative abundance of nutrients in Minnesota streams, energy is often an important limiting factor controlling algal growth. Flow rate is usually an available stream characteristic that captures possible effects of depth, width, roughness, and velocity on DO values. No notable trends of SR with $\tilde{a}_0$ as well as phase angles ($\tilde{\phi}_1$ and $\tilde{\phi}_2$) were apparent. Amplitudes (i.e., $\tilde{a}_1$ and $\tilde{a}_2$) decreased slightly with SR. No particular trends were evident between the estimated parameters and $Q$. However, a larger spread of estimated parameters for lower flows was apparent, reemphasizing the hypothesis of

---

Table 2. Ensemble-Mean Estimates of Parameters and Relevant Statistics for the Study Period of May–August of the Associated Calendar Years at Different Calibration Stations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>UM 5</th>
<th>Swan 12</th>
<th>Whitewater</th>
<th>Little Cobb</th>
<th>Minnesota Fort Snelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{a}_0$</td>
<td>0.9215</td>
<td>0.9711</td>
<td>0.9122</td>
<td>1.0072</td>
<td>1.0283</td>
</tr>
<tr>
<td></td>
<td>(0.0453)</td>
<td>(0.0658)</td>
<td>(0.0478)</td>
<td>(0.0475)</td>
<td>(0.0378)</td>
</tr>
<tr>
<td>$\tilde{a}_1$</td>
<td>0.1165</td>
<td>0.2155</td>
<td>0.1391</td>
<td>0.1331</td>
<td>0.0567</td>
</tr>
<tr>
<td></td>
<td>(0.0354)</td>
<td>(0.1027)</td>
<td>(0.0537)</td>
<td>(0.08)</td>
<td>(0.0438)</td>
</tr>
<tr>
<td>$\tilde{a}_2$</td>
<td>0.0249</td>
<td>0.0496</td>
<td>0.0400</td>
<td>0.0372</td>
<td>0.0274</td>
</tr>
<tr>
<td></td>
<td>(0.0155)</td>
<td>(0.0347)</td>
<td>(0.0195)</td>
<td>(0.0321)</td>
<td>(0.0267)</td>
</tr>
<tr>
<td>$\tilde{\phi}_1$ (degrees)</td>
<td>136.65</td>
<td>108.17</td>
<td>139.10</td>
<td>96.87</td>
<td>60.33</td>
</tr>
<tr>
<td></td>
<td>(26.77)</td>
<td>(18.10)</td>
<td>(19.96)</td>
<td>(15.48)</td>
<td>(29.95)</td>
</tr>
<tr>
<td>$\tilde{\phi}_2$ (degrees)</td>
<td>86.34</td>
<td>113.19</td>
<td>104.09</td>
<td>118.40</td>
<td>117.43</td>
</tr>
<tr>
<td></td>
<td>(33.32)</td>
<td>(29.06)</td>
<td>(30.44)</td>
<td>(35.45)</td>
<td>(44.33)</td>
</tr>
<tr>
<td>$\tilde{\phi}_1^*$ (degrees)</td>
<td>0.1002</td>
<td>0.2050</td>
<td>0.1326</td>
<td>0.1293</td>
<td>0.0446</td>
</tr>
<tr>
<td>$\tilde{\phi}_2^*$ (degrees)</td>
<td>0.0159</td>
<td>0.0403</td>
<td>0.0331</td>
<td>0.0326</td>
<td>0.0136</td>
</tr>
<tr>
<td>$\tilde{\phi}_1^* (degrees)$</td>
<td>140.58</td>
<td>103.86</td>
<td>138.40</td>
<td>95.43</td>
<td>65.89</td>
</tr>
<tr>
<td>$\tilde{\phi}_2^* (degrees)$</td>
<td>86.08</td>
<td>120.09</td>
<td>110.07</td>
<td>126.68</td>
<td>137.97</td>
</tr>
</tbody>
</table>

Note: Values in parentheses refer to the standard deviations of estimated parameters; a prime refers to the equivalent amplitudes and phase angles of ensemble-mean coefficients of a sine–cosine model; RMSE$_{DO}$ = root-mean-square error in DO; ME$_{DO}$ = average of errors in modeled DO; and $\tilde{a}_0$, $\tilde{a}_1$, $\tilde{a}_2$, $\tilde{\phi}_1$, $\tilde{\phi}_2$, $\tilde{\phi}_1^*$, $\tilde{\phi}_2^*$, $\tilde{\phi}_1^* (degrees)$, and $\tilde{\phi}_2^* (degrees)$ are used to estimate the RMSE in DO, RMSE in DO, ME DO, and errors in modeled DO, respectively.

---
Fig. 2. Box-plot revealing statistics of errors in modeled DO (ErrorDO) for each hour of the 24-h diurnal cycle (current day 1 a.m. to midnight) incorporating individual days of the study period (May–August) at Whitewater River site. Each box has solid lines at the lower quartile, median, and upper quartile values. Dashed lines are the whiskers extending from each end of the box to show the extent of the rest of the errors within 1.5 IRQ, where IRQ=interquartile range. Positive (+) signs present individual outliers beyond the ends of the whiskers.

Fig. 3. Variation of normalized \( \hat{a}_0 \), \( \hat{a}_0/E(\hat{a}_0) \), with normalized Julian days (Jday/TDY) within the study period of May–August of the associated calendar years at different stations. \( E(\hat{a}_0) \) refer to the ensemble-mean estimates for daily values of \( \hat{a}_0 \), Jday=Julian days, and TDY=total number of days in a year.
Abdul-Aziz et al. (2006) that above a threshold flow estimated parameters are largely indifferent of the prevailing flow rate.

**Sensitivity and Uncertainty Analysis**

The sensitivity coefficients $S_{\hat{a}_0}$, $S_{\hat{a}_s}$, and $S_{\hat{b}_s}$, respectively, for $\hat{a}_0$, $\hat{a}_s$, and $\hat{b}_s$ are defined analytically using Eqs. (1b), (3), and (5a) as

\[
S_{\hat{a}_0}(t_1) = \frac{\partial \text{DO}_{\hat{a}_0}(t_1)}{\partial \hat{a}_0} = \text{DO}_a
\]

\[
S_{\hat{a}_s}(t_1) = \frac{\partial \text{DO}_{\hat{a}_s}(t_1)}{\partial \hat{a}_s} = \text{DO}_a \left[ \cos(2\pi f_s t_s - \hat{b}_s) - \cos(2\pi f_s t_s - \hat{b}_s) \right]
\]

for $k = 1, 2, \ldots, W$

\[
S_{\hat{b}_s}(t_1) = \frac{\partial \text{DO}_{\hat{b}_s}(t_1)}{\partial \hat{b}_s} = \text{DO}_a \hat{a}_s \left[ \sin(2\pi f_s t_s - \hat{b}_s) - \sin(2\pi f_s t_s - \hat{b}_s) \right]
\]

for $k = 1, 2, \ldots, W$

where the equivalent ensemble-mean estimates (i.e., $\hat{a}_k$ and $\hat{b}_k$, $k = 1, 2$) for the study period are used to evaluate the sensitivity coefficients. Other notations are as previously defined. For a particular station, the coefficient $S_{\hat{a}_0}$ has a constant value of DO$_a$, whereas $S_{\hat{a}_s}$, $S_{\hat{b}_s}$, and $S_{\hat{b}_s}$ are periodic functions. The absolute values of the ratio $S_{\hat{a}_s}/\text{DO}_a$, $S_{\hat{a}_s}/\text{DO}_a$, $S_{\hat{b}_s}/(\text{DO}_a \hat{a}_s)$, and $S_{\hat{b}_s}/(\text{DO}_a \hat{a}_s)$ vary from a minimum of 0 to a maximum of 2.

Model uncertainties were estimated by multiplying sensitivity coefficients by the standard deviations of parameters. Fig. 4 presents the estimated model uncertainties for all five parameters for a small [Whitewater River, see Fig. 4(a)] and a large [Minnesota River near Fort Snelling, see Fig. 4(b)] watershed basins. As expected, the model showed zero uncertainties at 12 p.m. for both sites. Uncertainties associated with $\hat{a}_0$ were constant at approximately 0.5 mg/L and 0.3 mg/L for the small and large watershed stations, respectively. Uncertainties for $\hat{a}_s$ ranged from $-0.92$ to $0.13$ mg/L at Whitewater and $-0.19$ to $0.46$ mg/L at Minnesota Fort Snelling sites. The associated uncertainties in the phase angle $\hat{a}_s$ were, respectively, from $0.75$ to $0.15$ mg/L and from $-0.33$ to $0.01$ mg/L at those sites. Uncertainties associated with the second harmonic parameters (i.e., $\hat{a}_s$ and $\hat{b}_s$) were smaller than those for the first harmonics. Among the second harmonic parameters, $\hat{a}_s$ and $\hat{b}_s$ revealed model uncertainties from $-0.12$ to $0.25$ mg/L and from $-0.01$ to $0.33$ mg/L at the Whitewater, as well as from $-0.05$ to $0.34$ mg/L and from $-0.02$ to $0.13$ mg/L at the Minnesota Fort Snelling sites.

Careful observation of the earlier results indicates several important aspects. Modeled DO showed relatively more uncertainty for the small watershed station (Whitewater site) compared to that of large watershed one (Minnesota Fort Snelling site). This fact appears to be mainly related to the hydrology (i.e., lower or higher flow rates depending on the smaller or larger watershed size, respectively) of the corresponding stations. Model uncertainties showed similar diurnal patterns for the second harmonic parameters (i.e., $\hat{a}_s$ and $\hat{b}_s$) at both sites and for the first harmonic parameters (i.e., $\hat{a}_0$ and $\hat{b}_0$) only at the Whitewater site. The first harmonic parameters rather revealed slightly reverse patterns for the later station. These facts may also be partially attributed to the differential watershed size effect on DO processes, as well as to the mathematical formulations [Eqs. (6a)–(6c)] of estimating uncertainties that involve incorporation of phase-angle quantities for all parameters except $\hat{b}_0$. In essence, the smaller model sensitivities and uncertainties for the larger watershed station of Minnesota Fort Snelling seem to further support the hypothesis that above some upper threshold flow, the DO processes as well as the estimated parameters may be largely indifferent of the prevailing flow rates.

**Model Validation**

**Methods**

Model calibration was done with the existing short data records for all eight validation stations as the first step of the validation process. Results obtained with calibration are used as a reference for the performance of the three validation methods used herein. Validation Method 1 used the overall average of equivalent ensemble-mean estimates of amplitude and phase angles incorporating all five long record stations to predict DO and DO$_a$ at each of the validation stations. Validation Method 2 used ensemble parameters derived from the calibration station in Table 1 that has the same or most similar ecoregion and the closest watershed-size (the former criterion comes first in precedence) as those of the validation site. Validation Method 3 selects the calibration station in Table 1 based only on watershed size, i.e., ensemble parameters for the validation site are taken as those obtained from the calibration site that is the closest in watershed size. Ecoregions and watershed size were chosen as selection criteria because of process-driven considerations as well as calibration results obtained with the long record stations that indicated apparent correspondence between drainage size and DO processes.

Data at the validation stations were screened for possible instrumental and/or recording errors using the previously discussed filtering criteria. Using the screened data, the regular DO were predicted using Eq. (5a). The standard DO (i.e., DO$_a$) were obtained at all validation sites for a standard time ($t_s$) of 12 p.m. and discrete DO data for each hour of the diurnal cycle using the procedure demonstrated in the application methodology.

**Results**

Four statistics, namely RMSE$_{DO}$, ME$_{DO}$, RMSE$_{DO_a}$, and ME$_{DO_a}$, were used to summarize the results of model validation (Table 3). As previously defined, RMSE$_{DO}$ refers to the root-mean-square error (RMSE) in DO, ME$_{DO}$ refers to the average of errors in modeled DO, RMSE$_{DO_a}$ and ME$_{DO_a}$, respectively, refer to the overall RMSE and mean error in DO$_a$, where the predicted DO$_a$ was obtained for each hour within a day. Positive and negative mean errors (i.e., ME$_{DO_a}$ and ME$_{DO_a}$) correspondingly indicate underprediction and overprediction.

Among different stations, as shown in Table 3, ME$_{DO_a}$ ranged from $-0.11$ to $0.01$ mg/L with an average value of $-0.03$ mg/L for the reference method, from $-0.65$ to $0.72$ mg/L with an average of $0.14$ mg/L for Method 1, from $-0.51$ to $0.70$ mg/L with an average of $0.01$ mg/L for Method 2, from $-1.00$ to $0.39$ mg/L with an average of $-0.19$ mg/L for Method 3. For RMSE$_{DO_a}$, the ranges and average values were, respectively, from 0.09 to 0.91 mg/L and 0.38 mg/L for the reference method, from 0.35 to 1.73 mg/L and 0.95 mg/L for Method 1, from 0.53 to 1.65 mg/L and 0.94 mg/L for Method 2, from 0.53 to 1.65 mg/L and 0.98 mg/L for Method 3.
The statistics for the error between predicted and observed DO are also shown in Table 3. Values of ME\textsubscript{DO} ranged from −0.01 to 0.11 mg/L with an average of 0.03 mg/L for the Reference method, from −0.66 to 0.73 mg/L with an average of −0.12 mg/L for Method 1, from −0.65 to 0.51 mg/L with an average of 0.01 mg/L for Method 2, and from −0.28 to 1.03 mg/L with an average of 0.01 mg/L for Method 3. As well, the ranges and averages of RMSE\textsubscript{DO} were, respectively, from 0.01 to 1.19 mg/L and 0.32 mg/L using parameters for the reference method, from 0.02 to 1.71 mg/L and 0.76 mg/L for Method 1, from 0.00 to 1.83 mg/L and 0.88 mg/L for Method 2, and from 0.00 to 1.95 mg/L and 0.78 mg/L for Method 3.

Scatter plots of observed and predicted DO were also used to assess the accuracy of the different methods for estimating the parameters of the model. As shown in Table 1, the number of days of data varied among stations. Only the first two days were used for all stations to reduce the potential bias of result towards stations with longer data sets while improving overall legibility of the plots. The exception was for Station 005, where data of only one diurnal period were available after screening. Figs. 5(a–f) present the typical plots of observed and predicted standard DO (i.e., DO\textsubscript{o}) and associated RMSE for the reference method, respectively, for discrete data at 6 a.m., 8 a.m., 10 a.m., 2 p.m., 4 p.m., and 6 p.m. at all validation sites. In each plot, circles refer to the corresponding data points while the solid line refers to the perfect fit line. The smallest and largest RMSE\textsubscript{DO} were, respectively, 0.26 and 0.67 mg/L. The scatter-plots for Methods 1, 2, and 3 are correspondingly shown in Figs. 6(a–f), 7(a–f), and 8(a–f). The smallest and largest RMSE\textsubscript{DO} were, respectively, 0.75 and 1.63 mg/L for Method 1, 0.64 and 1.41 mg/L for Method 2, and 0.71 and 1.53 mg/L for Method 3.

In general, all methods resulted in reasonably good correspondence between observed and modeled data. As expected, the methods generally were less capable to represent the observed DO during the early morning hours (e.g., from 6 to 8 a.m.) or late afternoon hours (e.g., from 4 to 6 p.m.), respectively. However, as
Table 3. Summary of Validation Results at Different Stations

<table>
<thead>
<tr>
<th>Method</th>
<th>Statistical parameters</th>
<th>001</th>
<th>002</th>
<th>003</th>
<th>004</th>
<th>005</th>
<th>006</th>
<th>007</th>
<th>008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>ME&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>−0.04</td>
<td>0.01</td>
<td>−0.05</td>
<td>−0.03</td>
<td>0.01</td>
<td>−0.11</td>
<td>−0.01</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>RMSE&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>0.48</td>
<td>0.13</td>
<td>0.58</td>
<td>0.47</td>
<td>0.09</td>
<td>0.91</td>
<td>0.23</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>ME&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>0.04</td>
<td>−0.01</td>
<td>0.06</td>
<td>0.04</td>
<td>−0.01</td>
<td>0.11</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>RMSE&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>0.42</td>
<td>0.12</td>
<td>1.19</td>
<td>0.36</td>
<td>0.09</td>
<td>0.04</td>
<td>0.29</td>
<td>0.01</td>
</tr>
<tr>
<td>1–Average</td>
<td>ME&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>−0.65</td>
<td>0.26</td>
<td>−0.07</td>
<td>0.72</td>
<td>0.51</td>
<td>0.02</td>
<td>0.06</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>RMSE&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>1.02</td>
<td>0.52</td>
<td>0.73</td>
<td>1.73</td>
<td>0.92</td>
<td>1.71</td>
<td>0.35</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>ME&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>0.73</td>
<td>−0.30</td>
<td>0.10</td>
<td>−0.66</td>
<td>−0.48</td>
<td>0.09</td>
<td>−0.08</td>
<td>−0.32</td>
</tr>
<tr>
<td></td>
<td>RMSE&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>1.51</td>
<td>0.43</td>
<td>1.00</td>
<td>1.71</td>
<td>0.55</td>
<td>0.05</td>
<td>0.02</td>
<td>0.78</td>
</tr>
<tr>
<td>2–Ecoregion and size</td>
<td>ME&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>−0.25</td>
<td>0.57</td>
<td>−0.51</td>
<td>0.70</td>
<td>0.20</td>
<td>−0.36</td>
<td>−0.26</td>
<td>−0.01</td>
</tr>
<tr>
<td></td>
<td>RMSE&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>0.94</td>
<td>0.71</td>
<td>1.17</td>
<td>1.27</td>
<td>0.63</td>
<td>1.65</td>
<td>0.53</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>ME&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>0.32</td>
<td>−0.65</td>
<td>0.51</td>
<td>−0.62</td>
<td>−0.15</td>
<td>0.48</td>
<td>0.22</td>
<td>−0.04</td>
</tr>
<tr>
<td></td>
<td>RMSE&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>0.97</td>
<td>0.98</td>
<td>1.79</td>
<td>1.83</td>
<td>0.00</td>
<td>0.65</td>
<td>0.53</td>
<td>0.25</td>
</tr>
<tr>
<td>3–Size</td>
<td>ME&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>−1.00</td>
<td>0.00</td>
<td>−0.51</td>
<td>0.39</td>
<td>0.20</td>
<td>−0.36</td>
<td>−0.26</td>
<td>−0.01</td>
</tr>
<tr>
<td></td>
<td>RMSE&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>1.28</td>
<td>0.54</td>
<td>1.17</td>
<td>1.43</td>
<td>0.63</td>
<td>1.65</td>
<td>0.53</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>ME&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>1.03</td>
<td>−0.04</td>
<td>0.51</td>
<td>−0.28</td>
<td>−0.15</td>
<td>0.48</td>
<td>0.22</td>
<td>−0.04</td>
</tr>
<tr>
<td></td>
<td>RMSE&lt;sub&gt;DO&lt;/sub&gt; (mg/L)</td>
<td>1.95</td>
<td>0.01</td>
<td>1.79</td>
<td>1.04</td>
<td>0.00</td>
<td>0.65</td>
<td>0.53</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Note: RMSE<sub>DO</sub>=root-mean-square error in DO; ME<sub>DO</sub>=average of errors in modeled DO; RMSE<sub>DO</sub> and ME<sub>DO</sub>=overall RMSE and mean error in DO, obtained, respectively, by averaging RMSE and mean errors of predicted DO<sub>i</sub> for discrete data at all 24 h; and positive and negative mean errors correspondingly indicate underprediction and overprediction.

Fig. 5. Plots of observed and predicted standard DO (i.e., DO<sub>s</sub>) in Calibration Method for input data at (a) 6 a.m.; (b) 8 a.m.; (c) 10 a.m.; (d) 2 p.m.; (e) 4 p.m.; and (f) 6 p.m. at all validation sites. In each plot, circles refer to the corresponding data points and the solid line refers to the perfect fit line.
the time of discrete data collection approaches the standard time of 12 p.m., the error between predicted and observed values decreases. In particular, for the discrete data collected during the 10 a.m. to 2 p.m. range, the model resulted in the best fits as shown by the minimum RMSEDO.

Discussion

The results of calibration with the five long-record stations (Table 2) lead to several observations. Relatively small standard deviations in the estimated parameters, as well as, small RMSEDO and RMSE\(_{DO}\) demonstrate robust parameter estimates and good modeling accuracy. The parameters, and therefore the model, at each station seemed to be nearly indifferent of Julian day. These results refer to the strength and power of ESHA algorithm for DO modeling. The success of the approach is likely related to the use of normalized data.

The results of model validation [Table 3; Figs. 5(a–f), 6(a–f), 7(a–f), and 8(a–f)] with independent data sets clearly demonstrated superior performance by actual calibration (i.e., reference method) over the other three methods for estimating parameters. All three validation methods were close in terms of consistency of modeling error (i.e., similar RMSE\(_{DO}\) and RMSE\(_{DO}\)). However, Method 2, using ecoregional location and closest watershed size criteria, showed the best modeling accuracy by obtaining the minimum average values for ME\(_{DO}\) and ME\(_{DO}\). This finding was consistent with the intuitive selection of the appropriate validation method on the basis of likely watershed attributes and processes that dominate DO in streams.

The simpler DO model of this study is roughly equivalent in terms of accuracy and consistency to the fraction of DO saturation model of Abdul-Aziz et al. (2006), which involves analyses of both DO and water temperature measurements. On an average, the RMSE\(_{DO}\), RMSE\(_{DO}\), and RMSE\(_{DO}\) for the simpler DO model were respectively ~4% greater, ~1% smaller, and ~2% greater than their counterpart values of the model by Abdul-Aziz et al. (2006). An optimal number of harmonics (\(W_{opt}\)) of 2 seemed appropriate in both cases. The standard time (\(t_s\)) were flexible, however it demonstrated a preferable time slot of 10 a.m. to 1 p.m., inclusive for obtaining lower RMSE\(_{DO}\). The trends of estimated parameters with time, flow, and atmospheric solar radiation were also similar in both models. These similarities were mainly caused by the generally identical diurnal patterns of DO and water temperature data for a study site since saturated DO data were derived from temperature in the fraction of DO saturation model. Such identical performance by the two models also refers to the robustness of ESHA algorithm in obtaining good modeling accuracy, particularly for forcing the Fourier series through a fixed point.

Numerous problems can arise in collecting observed diurnal DO data. The impacts of data uncertainty were first minimized by using data reported to be of good quality. Data filtering was done to further remove possible erroneous values. Selections of thresh-
old limits were reasonable and prudent given the historical data at the Minnesota streams. For example, a DO less than the minimum threshold of 1 mg/L generally reflects a highly impaired water body and diurnal trend is of little utility for water quality decisions there. A highly hypereutrophic water body or a stream dominated by point source inputs may show a more than 2 mg/L of hourly DO change. This possibility was not evaluated in the study. A stream dominated by point source inputs might, however, be better represented by process-based models. Nonetheless, it is not possible to be absolutely certain that unexpected observations were not caused by stream processes resulting from unusual conditions. As such, screening threshold limits may be viewed as limitations of the type of situations under which the model was developed and tested.

Conclusions and Recommendations

A zero-dimensional empirical DO model is developed by using an ESHA algorithm originally formulated by Abdul-Aziz et al. (2006) for a fraction of DO saturation model that requires simultaneous measurements for water temperature and DO. Given similar modeling accuracy and consistency with both approaches, as well as apparent diurnal independence of the estimated parameters, the empirical formulation has the advantage of a simpler framework and single parameter (i.e., DO) input data requirement.

The DO model was calibrated with data for May–August of different calendar years within 2000–2005 for five long-record stations that represent different watersheds in an ecoregional framework. Input data were screened for possible errors by applying two simple filters of allowable hourly slope and threshold limits. Estimated model parameters demonstrated notable evidence of robustness regarding both temporal and spatial variations. Both analytical and numerical analyses of sensitivity coefficients and associated model uncertainties were performed. The diurnal patterns of model uncertainties were obtained by multiplying the parameter uncertainties by the respective sensitivity coefficients. The model had noticeable uncertainties associated with estimated parameters, particularly to that in phase angles. Model sensitivity and associated uncertainties also seemed to be larger for stations that drain smaller watersheds.

Model validation was done using independent data for eight different stations across Minnesota. Three validation methods were used for selecting representative parameters from the five possible long-record calibration stations using the overall average as well as ecoregional locations, and size of drainage area. Model calibration was also performed with available short data-record at each validation site to compare with the results of different validation methods. While all the validation methods showed nearly equivalent performance, Validation Method 2, which involved selection of parameters following ecoregional location and the nearest sized drainage area, resulted in the best modeling accuracy. All the validation methods showed minimum RMSE in standard DO (i.e., DO) if it was predicted using discrete data measured at a time closer to the standard time (t). However, it was clearly
evident that direct calibration with appreciable data coverage generally resulted in the most desirable model in terms of accuracy and consistency of the predicted value.

The DO model developed from the ESHA algorithm would convert discrete DO data measured at any time of the diurnal cycle to those at a standard time given an appropriate set of Fourier coefficients. This would particularly aid in total maximum daily load (TMDL) studies of aquatic ecosystem health. A TMDL analysis may consider daily minimum and/or average DO as the criterion for assessing stream health. An example of a load allocation for DO impaired water is given by Gunderson and Klang (2004). Minimum DO values generally occur in the early morning hours. Since the method is more accurate when the sample is collected at time closer to the standard time, better accuracy can be obtained with minimum DO by shifting the standard time from noon to an early morning hour. The model can also be used to estimate the diurnal cycle of DO from a single observation given an appropriate set of Fourier coefficients. Clearly the best approach for estimating the Fourier coefficient is obtained from a long-record of diurnal data set at the site. However, regional coefficients also showed promise and increased the ease of using the model. Further improvement of the model could be obtained by investigating relationships between model parameters and other biogeochemical parameters (e.g., feeding watershed area, watershed slope, land use type etc.). A better understanding of such mechanisms may aid in unraveling spatial scaling relations among parameters.

Acknowledgments

The research described in this paper was funded by a total maximum daily load (TMDL) grant from the U.S. Environmental Protection Agency (USEPA). Data sets were collected from Mark Evenson (MPCA), James Klang (MPCA), Jason Elmert (MPCA), Paul Wotzka (MDA), and Kathy Lee (USGS). Important suggestions and comments from Dr. Heinz G. Stefan, Dr. Efi Foufoula-Georgiou, Dr. James A. Perry, and Dr. Bruce Vondracek from the University of Minnesota are gratefully acknowledged.

Notation

The following symbols are used in this paper:

- \( AIC \) = Akiak information criterion;
- \( \hat{a}_k, \hat{b}_k \) = estimated Fourier coefficients;
- \( a_k, b_k \) = theoretical Fourier coefficients;
- \( \text{DO}_{\text{obs}} \) = observed dissolved oxygen (DO) (mg/L);
- \( \text{DO}_{\text{obs}} \) = fraction of \( \text{DO}_s \), obtained by dividing \( \text{DO}_{\text{obs}} \) by \( \text{DO}_s \);
- \( \text{DO}_s \) = value of \( \text{DO}_{\text{obs}} \) at \( t_s \) (mg/L);
- \( \tilde{\text{DO}}_{\text{obs}}(t) \) = model \( \text{DO}_{\text{obs}}(t) \) obtained by using equivalent ensemble-mean parameter estimates;
- \( \text{Error}_{\text{DO}} \) = error in modeled DO (mg/L);
- \( f_k \) = discrete frequencies;
\( h(t) \) = a harmonic process;  
\( IRQ \) = interquartile range;  
\( J_{\text{day}} \) = Julian days;  
\( k \) = harmonic number;  
\( M \) = objective function for least squares minimization;  
\( \text{ME}_{\text{DO}} \) = mean error in modeled DO (mg/L);  
\( \text{ME}_{\text{DO}}^2 \) = mean error in modeled \( \text{DO}_j \) (mg/L);  
\( n \) = total number of observations within \( t' \);  
\( P \) = vector of length 2W consisting of parameters \( \hat{a}_j, \hat{b}_j \) (for \( j \neq 0 \));  
\( Q \) = \( 2W \times 2W \) nonsingular transition matrix;  
\( Q \) = flow rate (m\(^3\)/s);  
\( R \) = vector of length 2W consisting of the terms associated with the observation \( y_i \) and fixed point \( h(t_i) = \kappa \);  
\( \text{RMSE}_{\text{DO}} \) = root-mean-square error in DO (mg/L);  
\( \text{RMSE}_{\text{DO}}^2 \) = RMSE in predicting DO (mg/L);  
\( S_{\alpha_0}, S_{\alpha_3}, S_{\beta_k} \) = sensitivity coefficients, respectively, for \( \hat{a}_0, \hat{a}_3, \hat{\beta}_k \);  
\( \text{SR} \) = atmospheric solar radiation (cal/cm\(^2\));  
\( \text{SSE}_{\text{DO}} \) = total sum of squared errors (SSE) for DO (mg\(^2\)/L\(^2\));  
\( TDY \) = total number of days in a calendar year;  
\( t \) = time;  
\( t_i \) = \( i \)th time instant of the diurnal cycle (h);  
\( t_s \) = standard or reference time;  
\( t' \) = process period;  
\( W \) = appropriate maximum number of harmonics;  
\( W_{\text{opt}} \) = optimal number of harmonics;  
\( y(t) \) = a stochastic Fourier series;  
\( \hat{a}_k \) = estimated amplitudes;  
\( \Delta t \) = sampling interval;  
\( e(t) \) = a zero mean random error sequence;  
\( \eta \) = total number of 24-h days;  
\( \kappa \) = value of harmonic process \( h \) as a special constraint of least-squares optimization; and  
\( \hat{\phi}_k \) = estimated phase angles.

References


